# THE GLOBAL SPREAD OF COVID-19

**Student Name: Andre McQuaid (**S00216958**)**

## Introduction

Covid-19 has been with us for 2 years now and Johns Hopkins University has become an important and trusted data source for many analyses of Covid-19 related data. The Center for Systems Science and Engineering (CSEE, 2020) data, provided and maintained by JHU (John Hopkins University), is a Github project where John Hopkins makes collated Covid-19 data publicly available in one location. Having access to a central Covid-19 data repository allows many organisations to focus on leveraging the data to help fight the pandemic, leaving the task of managing this data to this trusted organisation.

The data itself is sourced from several locations, including the World Health Organization (WHO, n.d.), The European Centre for Disease Prevention and Control (ECDC, n.d.) and The United States Centers for Disease Control (US CDC, n.d.). The data sources are vetted as much as possible to ensure the accuracy and reliability of the data. Failure to do so could jeopardise the validity of the data and any reports and dashboards that take a dependency on the data.

For Irish data, the project consumes The Government of Ireland’s Covid-19 Data Hub (Ireland's COVID-19 Data Hub, n.d.).

## Literature Review

Relevant research will be discussed in the literature review section. This literature review serves to inform the forthcoming mini-project that proposes the development of a Covid-19 dashboard leveraging the JHU Covid dataset.

## Covid-19 Data Analysis

(Podder & Mondal, 2020) discuss techniques used to predict Covid-19 cases and Intensive Care Unit (ICU) requirements based on Covid-19 data from Brazil. A number of machine learning techniques were trialled, and their results compared. Results showed that in this case COVID-19 detection can be predicted with an accuracy of 94.39% and that ICU requirements can be predicted with an accuracy of 98.13%. It is clear that accurate predictive techniques, built on sound data, can help with targeted resource allocation during the pandemic, ultimately saving lives.

(Darapaneni et al., 2021) discuss the prediction of hospital bed utilization for COVID-19 in Telangana, India. Based on India's first Covid-19 wave, this paper evaluates the predictive power of Covid-related data, incorporating time series analysis, to predict the requirement of isolation beds, oxygen-enabled beds, and ICU beds in both public and private hospitals for COVID-19 patients in this region. The paper concluded that there was sufficient hospital capacity for the first wave. The paper also concluded that the techniques employed could be used in the analysis of future Covid waves in this same region, but also potentially beyond and even extended to other countries. Given that India was subject to subsequent Covid waves it would be interesting to see how this model held up, that is, how accurate it would have been in predicting hospital bed utilization for the Covid waves that followed.

(Kurniawan et al., 2020) analysed clustering and correlation methods for predicting COVID-19 risk in affected countries. Unsupervised techniques, namely K-means clustering and the analysis of correlation matrices, were applied to a Covid-19 data set to examine both the presence of clusters and the relationships between the features captured in the data. The system uncovered a set of five clusters, each containing a set of countries, as well as uncovering correlations between features in the data. For example, a strong positive correlation was found between total deaths and patients categorized as critical. The hope being that these findings could be built upon to help provide organisations, such as the US CDC, with better tools to help drive effective data-driven decision making.

The literature review unearthed a broad range of statistical and machine learning modelling techniques that were applied to the problems facing the world in the battle against Covid-19. This illustrates the broad arsenal of tools available to data scientists and related professionals and shows how much investigation is required across these various techniques to determine the most applicable ones to use.

One such example comes from (Nikhil; Saini; Panday; Gupta, 2021) where polynomial based linear regression models were employed to predict COVID-19 cases. This model outperformed the others that were also evaluated in this study. The paper also evaluated the potential application of machine learning and deep learning to aspects of the pandemic, including vaccine development.

## Application of Deep Learning

Machine learning, and deep learning in particular, are also prevalent in the Covid-19 literature. These techniques are very well suited to the area of Covid-19 diagnosis from scan images, such as chest x-rays. The National Institute of Health (NIH, 2017) Clinical Center provides one of the largest publicly available chest x-ray datasets to the scientific community and it is one that features regularly in the literature.

CovFrameNet is an example of a system that makes extensive use of the NIH scan dataset. CovFrameNet (Oyelade; Ezugwu; Chiroma, 2021) is an enhanced deep learning framework for Covid-19 detection. The system employed a deep learning Convolutional Neural Network (CNN) to attempt to detect and classify the presence of Covid-19 in x-rays and Computed Tomography (CT) scans. The system employed some novel pre-processing techniques and achieved encouraging results. The ability to quickly and accurately make Covid-19 diagnoses enables practitioners to make important decisions much earlier in the process. For example, determining whether contact tracing should be initiated for a newly diagnosed patient or whether they should be placed in a dedicated Covid ward.

(Panwar; Yadav; Mishra; Gupta, 2021) describes other deep learning techniques for the real time detection of Covid-19 (and pneumonia) using chest radiographs. X-rays were classified as either Covid-19 positive, pneumonia positive or normal. Transfer learning was used as there was a limited training set available in this case and it also accelerated the training process. Several CNN (Convolutional Neural Network) architectures were also evaluated for best results. The paper concludes that although high accuracy can be achieved with such systems (using an Inception V3 model and neural network classifier, an accuracy of 88.8% was obtained), they cannot be used in isolation. They can however be used as an additional aid in a clinical diagnosis setting. With a large influx of potentially Covid-19 positive patients, along with admissions due to other conditions, throughout the numerous waves of Covid, such systems can certainly help in streamlining and speeding up the path to clinical diagnosis. This can help relieve some of the additional pressure being exerted on hospital settings during the pandemic.

## Covid-19 Data Visualization

Visualizations related to Covid-19 have become a mainstay on media news programs, providing easier ways to consume what the data tells us about the previous, current, and potentially future situation regarding the virus. This speaks to the power of visualization. This has also become a hot topic in the research world also.

(Zuo et al., 2020) researched mobility and sociability trends during Covid-19, producing an interactive data visualization. The dashboard employed extensive data mining and powerful cloud computing techniques. This facilitated more insights to be derived from the data and visualised appropriately in the dashboard for consumption by end users. It was obvious from the dashboard, for example, that commuters were dropping public transport in favour of private vehicular transport. These visualizations help inform the relevant authorities of changing commuter patterns allowing them to make informed decisions in this regard. These decisions can be further honed over time as more recent data is processed and the dashboard refreshed accordingly.

(Comba, 2020) performs an interesting analysis of various visualisation techniques used to communicate information about the pandemic to a wide audience. This effort was undertaken with a view to motivating more work in this space and the provision of ever more powerful and insightful reports and dashboards that make the data around Covid-19, and the insights derived from this same data, more easily communicated to the intended audience. For example, the use of heatmap matrices to compare Covid death time series data was examined. Coupled with column alignment by date, this hybrid visualisation allowed country comparisons around when certain milestones were reached.

## Conclusion

The literature review has identified that many of the techniques at the disposal of data scientists and related professionals have been, and continue to be, applied to the problems posed by the pandemic. Analysis of data provided by trusted organisations such as John Hopkins has facilitated novel and effective solutions that have helped in our efforts to combat the virus. The death toll in this pandemic, despite being considerable, is significantly less than those that humanity has faced previously and at least some of that is due to the advances in technology such as data analytics, artificial intelligence, and machine learning.

As this body of knowledge is built up, we are arming our defensive arsenal with more and more weapons that can be used to fight the virus, including newer variants of concern such as Omicron. This new facet to the pandemic will no doubt also become the focus of more research.

# References


alia, N. D. (2021). Predicting Hospital Beds Utilization For COVID-19 In Telangana, India. *IEEE.* https://ieeexplore.ieee.org/document/9422598.

alia, R. K. (2020). Clustering and Correlation Methods for Predicting Coronavirus COVID-19 Risk Analysis in Pandemic Countries. *International Conference on Cyber and IT Service Management (CITSM).* https://ieeexplore.ieee.org/document/9268920.

Comba, J. L. (2020). Data Visualization for the Understanding of COVID-19. *Computing in Science & Engineering*, https://ieeexplore.ieee.org/abstract/document/9222822.

*ECDC*. (n.d.). Retrieved from European Centre for Disease Prevention and Control (ECDC): https://www.ecdc.europa.eu/en

*Ireland's COVID-19 Data Hub*. (n.d.). Retrieved from Ireland's COVID-19 Data Hub: https://covid19ireland-geohive.hub.arcgis.com

NIH. (2017). *News Releases*. Retrieved from National Institutes of Health: https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community

Nikhil, Saini, Panday, & Gupta. (2021). Polynomial Based Linear Regression Model to Predict COVID-19 Cases. *IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT).* https://ieeexplore.ieee.org/document/9574032.

Oyelade, Ezugwu, & Chiroma. (2021). CovFrameNet: An Enhanced Deep Learning Framework for COVID-19 Detection. *IEEE*, https://ieeexplore.ieee.org/abstract/document/9440437.

Panwar, Yadav, Mishra, & Gupta. (2021). Deep Learning Techniques for the Real Time Detection of Covid19 and Pneumonia using Chest Radiographs. *IEEE EUROCON 2021 - 19th International Conference on Smart Technologies.* https://ieeexplore.ieee.org/document/9535604.

Podder, P., & Mondal, M. R. (2020). Machine Learning to Predict COVID-19 and ICU Requirement. *ieeexplore.ieee.org*, https://ieeexplore.ieee.org/document/9393123. Retrieved from https://ieeexplore.ieee.org: https://ieeexplore.ieee.org/document/9393123

University, C. f. (2020). *https://github.com/CSSEGISandData/COVID-19*. Retrieved from https://github.com.

*US CDC*. (n.d.). Retrieved from Centers for Disease Control and Prevention: https://www.cdc.gov/coronavirus/2019-ncov/index.html

WHO. (n.d.). *World Health Organization*. Retrieved from World Health Organization: https://www.who.int

Zuo, W. G. (2020). An Interactive Data Visualization and Analytics Tool to Evaluate Mobility and Sociability Trends During COVID-19. *arXiv.org*, https://arxiv.org/pdf/2006.14882.pdf.