*The Importance of COVID-19 Analysis and Prediction using Machine Learning*

Hussein Abdulsada   
*Bachelor of Technology (Automation Engineering – Smart Systems)*  
*McMaster University*  
Hamilton, Canada [abdulsah@mcmaster.ca](mailto:abdulsah@mcmaster.ca)

Tony (Dongha) Kim  
*Bachelor of Technology (Automation Engineering – Smart Systems)*  
*McMaster University*  
Hamilton, Canada [kimd19@mcmaster.ca](mailto:kimd19@mcmaster.ca)

*Abstract*— This project allows us to use existing data collected regarding covid cases in Ontario to predict future covid cases. Global data sets are also explored to generate a predictive model that can aid in analyzing and visualizing how Canada is doing relative to the rest of the world. The generated graphs can help visualize the different aspects with covid 19 analysis, from visualizing the amount of covid 19 testing centers and where they are located in Ontario, this helps show just how much the virus was impacting the province in providing testing kits and facilities to the public as the demand highly increased. Such data not only allows the individuals to verify their health status, but it is useful for data analysis and smart cities applications such as the case of this project's scope – determine the future trend based on machine learning algorithms. In this project, the focus is on using recorded data to create a linear regression model that aids in predicting the future total cases in Ontario as well as the total number of covid related deaths around the world.

# Introduction

Since the beginning of covid 19, people have always turned to news channels and going on various sites to see the changes in covid 19 cases, this meant people could see the numbers of the day of and before. What is needed for the medical side and for the governments to know is an accurate prediction of how the future would most likely look based on current trends. This is to aid in figuring out the number of health care workers, types of laws to enforce and a number of products that may need to be distributed like masks vaccines etc. We decided to look at the data provided by the government of Canada and use the linear regression model to analyze and predict the total confirmed cases of COVID-19 in Ontario and the total deaths in the world. This information would help the government and people set strategies to adapt to the situation and further discuss tactics to minimize the cases based on the prediction. Statistically speaking, it is also important to see how covid is affecting the world to learn from other countries and see how the changes they make have affected the impact of covid on their country.

# Literature Review

Canada had to react quickly to the rapid spread of covid 19, back in 2019 and 2020, they did that by increasing their efforts in testing and contact tracing. “The provincial public health authorities worked closely with regional public health officers and local governments to set policies and recommendations and to implement services such as testing and contact tracing.” [1] It is clear to see from figures 1 and 2 how many testing centers were established by 2022 to keep up with the testing requirements, this map is great for the visualization of how many centers are available to the public to get tested.

The government was encouraging people to get tested when they felt they had a symptom, that way preventative measures could be taken sooner. This also assisted in gathering data to find the status of the virus in Canada. Having this system in place required that more testing kits and facilities be available “The capacity for COVID-19 polymerase chain reaction testing was limited in many regions early on but was gradually expanded and by June was widely available across the country. Public health agencies hired and trained people to expand contact tracing efforts.” [1] In this report, we focused on showing the availability of these facilities in Ontario rather than all of Canada as it would not be as useful for our scope but also challenging to plot without making it seem crowded

The use of smart city technologies and data analysis was not only implemented by Canada, but other countries also had great success with contact tracing and the use of smart technology to reduce the spread. “Canada’s efforts contrast sharply with those of Taiwan, Singapore, and South Korea, which relied more heavily on modern communication technology, including rapid integration of personal databases, more rapid expansion of tests and tracing, and wide-scale use of masks.” [1]

There have been doubts and ethical questions regarding the collection of data and keeping track of the infected and overall covid cases. The usefulness of collecting this data is evident, experts have also created algorithms and projects to aid in the analysis of covid 19 data. “Since the beginning of COVID-19, more than 13,036,550 people have been infected, and 571,574 died because of the disease by July 13, 2020. Developing new methodologies to predict the COVID-19 pandemic will help policymakers plan to contain the spread of the virus. In this research, we develop a Stochastic Fractal Search algorithm combined with a mathematical model to forecast the pandemic. To enhance the algorithm, we employed a design of the experiment's approach for tuning. We applied our algorithm to public datasets to model the COVID-19 pandemic in Canada in the upcoming months.” [2]

Now, though there is an importance of increasing the healthcare facilities related to covid testing and treatment, it is important to see how the implementations of smart cities technologies will affect this, mainly, how is the internet and phone calls changing the physical examinations. “Compared with 2019, total primary care visits between March and July 2020 decreased by 28.0%, from 7.66 to 5.51 per 1000 people/day. The smallest declines were among patients with the highest expected healthcare use (8.3%), those who could not be attributed to a primary care physician (10.2%), and older adults (19.1%). In contrast, total visits in rural areas increased by 6.4%. Office visits declined by 79.1% and virtual care increased 56-fold, comprising 71.1% of primary care physician visits. The lowest uptake of virtual care was among children (57.6%), rural residents (60.6%) and physicians with panels of ≥ 2500 patients (66.0%).” [3] From this, we understand the necessity for the health care system and government to collect the data and reduce the severity of the spread of the corona virus. From 2019 to 2020 there was a big improvement, and this is due to the data collection and preventative measures carried out.

Machine learning can also help researchers and practitioners analyze large volumes of data to forecast the spread of COVID-19, in order to act as an early warning system for future pandemics and to identify vulnerable populations. By examining social determinants of health information and incorporating patient data into algorithms, machine learning can assist in identifying patients who are more likely to need medical resources. “Additionally, policymakers can use algorithms to inform state and national public health decisions regarding mask mandates and reopening based on population health data and community transmission.” [5]

# Infographic Analysis and Solutions

## Covid-19 data analysis

Map

Description automatically generated  
Fig. 1. Assessment center location map

This infographic gives Ontarians an idea of where the testing centers for covid 19 are so they can have an idea of where to go and the number of open areas, this can also help other provinces compare with other provinces and see how well they are doing in comparison.

Map

Description automatically generated  
Fig. 2. Assessment center location - real map

Figure 2 is the same as figure 1 but with a detailed map to show the regions relative to the other provinces. It might not be necessary to include but it gives an idea of the possible ways to visualize the data.

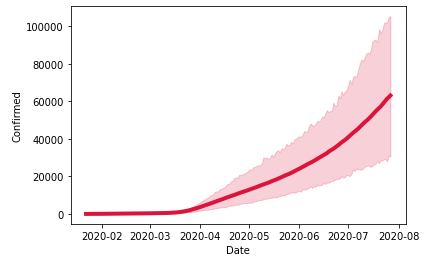


Fig. 3. Confirmed cases of COVID-19 in the world from January 2020 to July 2020

Figure 3 lets shows the increasing numbers of covid 19 confirmed cases globally, this graph is useful because it lets analysts be aware of the presence of a certain factor that is causing an exponential increase in the number of covid 19 cases around the world. From this graph, hypothesis could be made, and further investigations and research could help find potential causes and it also helps set new strategies that to replace previous failed ones.

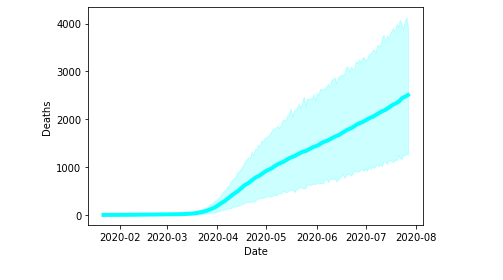


Fig. 4. Deaths by COVID-19 in the world from January 2020 to July 2020

This graph provides a similar benefit as figure 3 with the main focus being death cases caused by covid-19 instead of the number of confirmed cases.

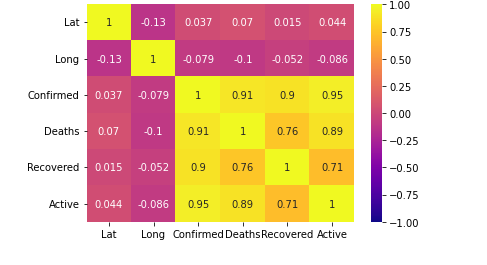


Fig. 5. Correlation of Each factor

Ignoring latitude and longitude in figure 5, the chart shows the strong correlation between confirmed, deaths, recovered and active cases. The heat map draws out the correlation, these 4 mentioned factors were features for the created model to predict number of covid related deaths in each country.

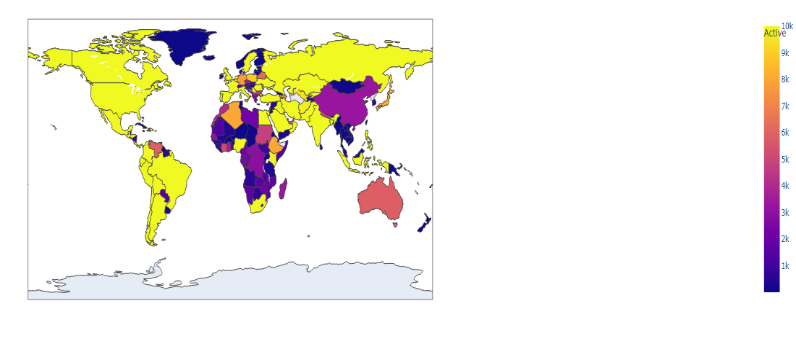


Fig. 6. Active Cases until 7/27/2020 based on a dataset provided

Figure 6 shows the active cases of covid 19 globally since the start of the pandemic until July 2020, where the hot zones where the lighter the color the higher the covid cases, refer to heat map on the right side.

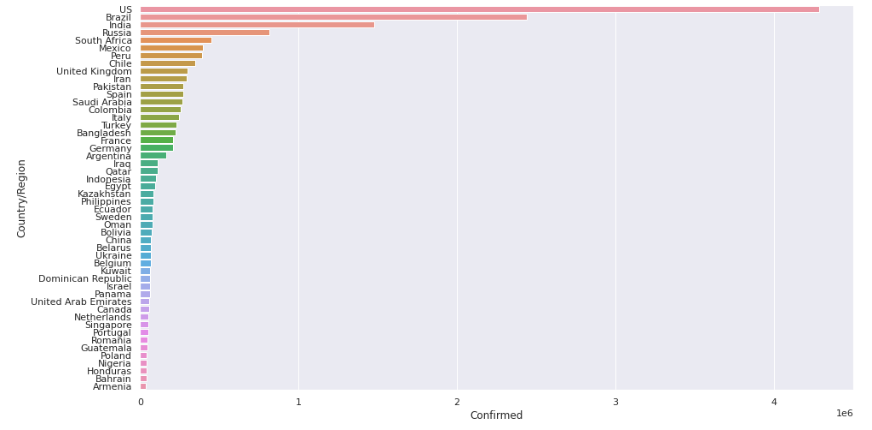


Fig. 7. Top 50 countries that have the most confirmed cases in the world - 7/27/2020

Top 50 countries in the world with most confirmed cases in the world until 27th of July 2022. This graph gives a very good and easy to understand visual of the state of covid-19 and which countries are affected the most and how they relate to each other.

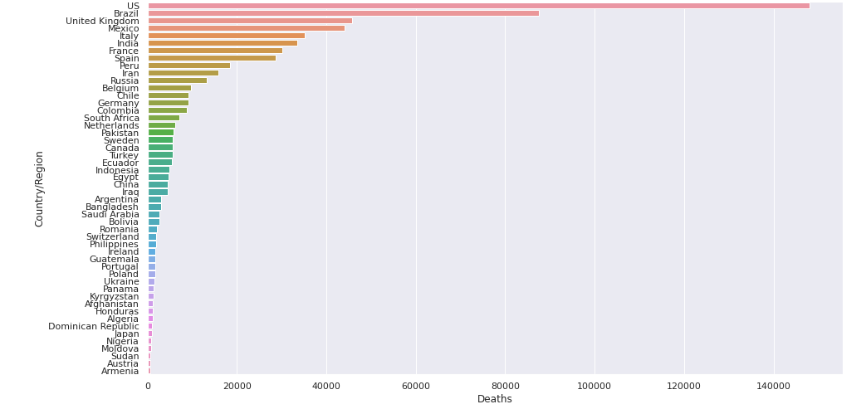


Fig. 8. Top 50 countries that have the greatest number of deaths in the world - 7/27/2020

Figure 8 is similar to figure 7 with the difference being death count instead of confirmed cases. This graph can be compared to the graph from figure 7 to get an idea of the relationship between the number of cases and death count related to the corona virus.

## Solutions: Prediction Models

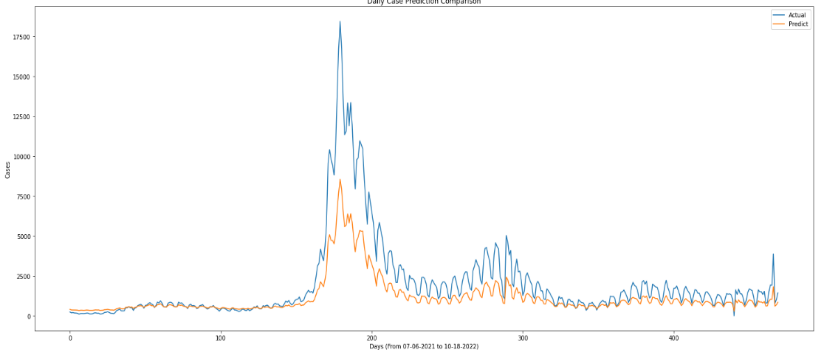


Fig. 9. Actual & Predicted numbers of daily confirmed cases in Ontario from July 2021 – October 2022

We have created two prediction models that predict the total confirmed cases in Ontario from July 2021 to October 2022 and the total number of deaths caused by COVID-19. It is crucial to have such predictions to anticipate and estimate the maximum healthcare capacity that can handle the outbreak since there is a limitation in terms of capacity that the health systems in Canada and the world can manage. The prediction models can be very useful in aiding the healthcare industry set their strategies to deal with the pandemic. Figure 9 demonstrates the accuracy of the obtained model from the dataset of confirmed number of cases in Ontario retrieved from the government's website. The orange line is the model generated using linear regression and the blue line is the actual cases. This model is supposed to help predict the future trends, to know how good these predictions are, this figure lets us see how much the predicted model is compared to the actual data for reliability purposes. It seems from the two lines that linear regression does not yield a very accurate result, but it is reasonably close.

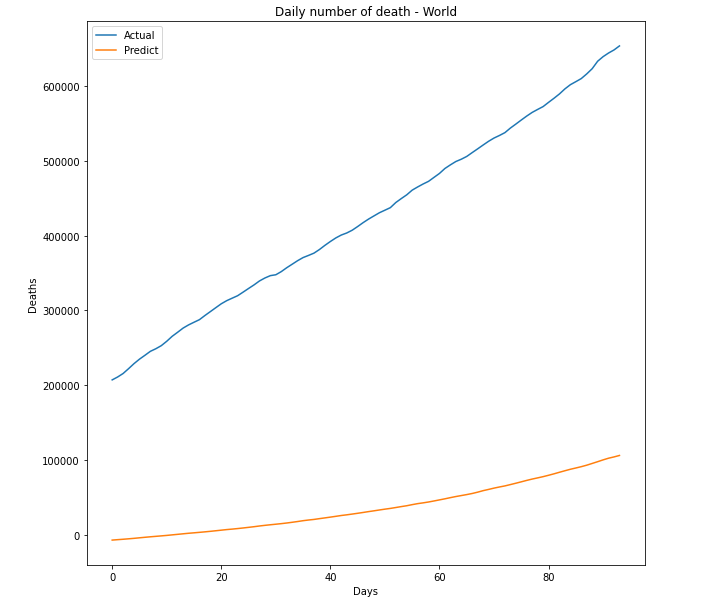


Fig. 10. – Daily number of covid related deaths in the world from 4/25/2020 to 7/27/2020

This model uses a different dataset to cover the death counts in each country. This model was created with a linear regression as multiple regression and classifications did not provide as high a score as the linear regression showed, though the score was not great due to the lack of features in the dataset. Using a different dataset with more features would give a higher accuracy score than the model shown in Figure 10, however, the point of building a prediction model was to provide several clues for people to understand the situation and suggest some of the solutions they could put into action. According to figure 10, there would be an incremental number of deaths around the world daily.

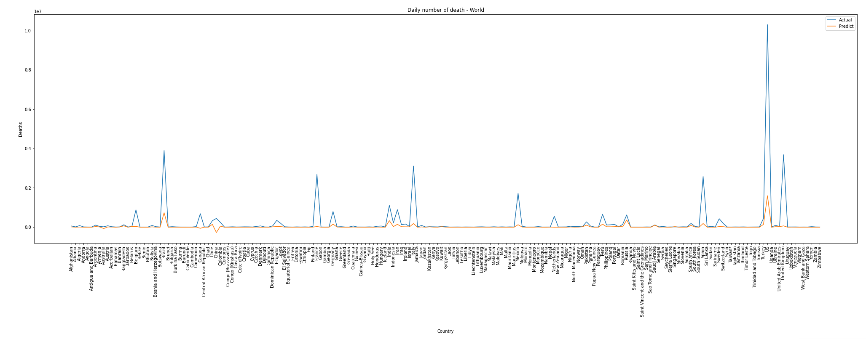


Fig. 11. – Actual and predicted number of covid related deaths in each country on 7/27/2020

Figure 11 shows the total predicted and actual number of covid related deaths in each country. The negative values represent the number of recovered individuals. As mentioned above, the accuracy of the model is not quite high, however, this information suggests to the governments what they should and need to do to further reduce the impact of COVID-19.

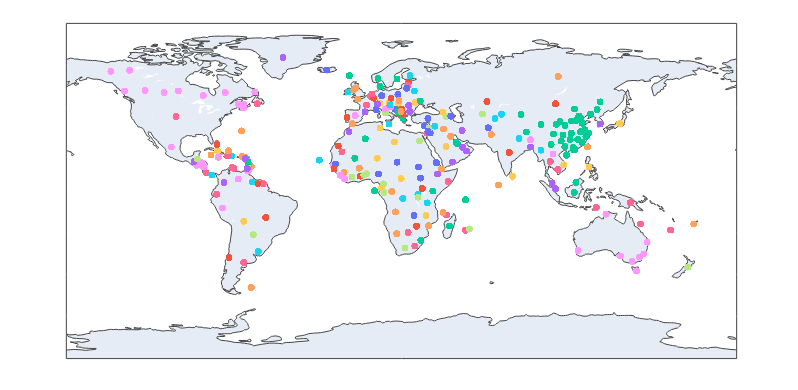


Fig.12. Actual & Predicted numbers of covid related deaths - World Map

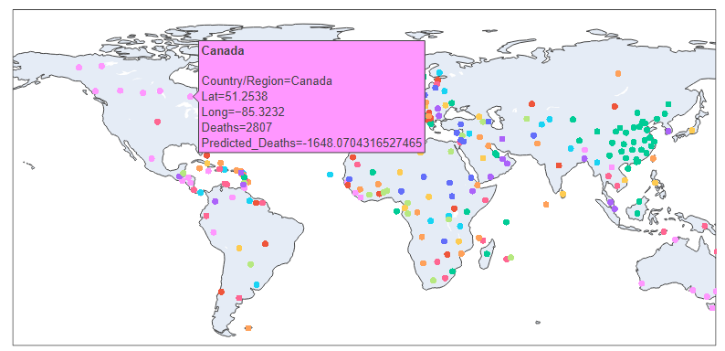


Fig.13. Actual & Predicted numbers of covid related deaths - World Map

The infographics above show the data obtained from the dataset and prediction model 2. The data points on the map represent information of the total actual deaths and predicted total deaths for that country.

# Discussion

Since the Covid-19 pandemic started, it has been very important to track the trend of the pandemic and its impacts on our society, not limited to communities, but also from cities to nations and the entire world. Analyses based on data collected from each country played a key role in minimizing the impact of the pandemic such as fatalities, collapsing medical systems, and exceeding the limit of capacity the healthcare facilities can possibly handle.

The main purpose of having the infographics and analysis on them is to help visualize the different aspects of COVID-19. Visualizing the amount of covid 19 testing centers and where they are in Ontario helps show how much the virus was impacting the province in providing testing kits and facilities to the public as the demand highly increased. Anticipating the total cases could also help plan the strategies to deal with the pandemic by understanding the prediction, estimating the ability of healthcare systems in Ontario to reduce the spread of covid-19. It is clear that the goal is not just limited to one city or one province, but it is also crucial to analyze how other countries perform and manage the pandemic with their own strategies, and to compare the analysis of future trend based on the data and adapt the best strategies depending on the circumstances of each country.

Further technologies and exploring more datasets could have given more precise models that are more capable of predicting that could be even used to show what the trends could look like years from now. “Machine-learning algorithms must be adequately trained using diverse data to serve the patient population best and avoid algorithm bias. Before moving the technology into clinical practice, the research team must ensure the technology is working accurately and equitably.” [5]

As mentioned above, the main challenges we had in creating our predictive models were finding an appropriate dataset to train our model and selecting a perfect regression or classification that would fit for the dataset. To find a perfect matching model, we spent a significant amount of time to obtain the highest score for the accuracy of model and this was done by trial and error. However, due to decisions we made on the datasets and model selection, we did not have a high score in accuracy. By exploring better datasets and model selection, we would be able to achieve an appropriate score and use the models in real world applications.

# Findings

Did you know that more than 182 countries had 0 confirmed cases of covid 19 on July 27th, 2020, out of the 187 countries?

Did you know that the top ranked country with the most reported covid cases was the United States of America although they are an advanced country? A question might be why it is so when they have an established health care system and resources. A logical assumption would be because of slow reaction time from the United States government, another factor would be the fact that other countries might not have as a robust system to keep track of the number of covid cases compared to advanced countries like the United States due to lack of resources and technology advancement.

Did you know that Canada placed 20th in the highest ranking when it comes to covid related deaths in 2020 with a total number of deaths of around 5700 deaths compared to the United States with a total of about 150000 deaths?

Did you know that a Canadian startup was first in the world to use AI to discover an odd signal showing a cluster of unusual Pneumonia cases happening in a market in Wuhan China and flagged it? BlueDot had spotted the virus nine days before the WHO released its statement alerting people to the emergence of a novel coronavirus. [6]

References

[1] A. Detsky and I. Bogoch, “COVID-19 in Canada Experience and Response,” Jama Network, 10-Aug-2020. [Online]. Available: https://jamanetwork.com/journals/jama/article-abstract/2769439. [Accessed: 29-Nov-2022].

[2] S. Khalilpourazari and H. H. Doulabi, “Robust modelling and prediction of the COVID-19 pandemic in Canada,” Production of Healthcare under Epidemic Outbreaks, 14-Jun-2021. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/00207543.2021.1936261?scroll=top&amp;needAccess=true. [Accessed: 29-Nov-2022].

[3] R. Glazier, M. Green, F. Wu, E. Frymire, A. Kopp, and T. Kiran, “Shifts in office and virtual primary care during the early COVID-19 pandemic in Ontario, Canada,” CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 08-Feb-2021. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33558406/. [Accessed: 30-Nov-2022].

[4] S. Sivasubramanian, “How AI and Machine Learning are helping to tackle COVID-19,” World Economic Forum. [Online]. Available: https://www.weforum.org/agenda/2020/05/how-ai-and-machine-learning-are-helping-to-fight-covid-19/. [Accessed: 30-Nov-2022].

[5] M. P. A. Erin McNemar, “How machine learning can guide covid-19 decision-making,” HealthITAnalytics, 07-Feb-2022. [Online]. Available: https://healthitanalytics.com/features/how-machine-learning-can-guide-covid-19-decision-making. [Accessed: 30-Nov-2022].

[6] C. Stieg, “How this Canadian start-up spotted coronavirus before everyone else knew about it,” CNBC, 06-Mar-2020. [Online]. Available: https://www.cnbc.com/2020/03/03/bluedot-used-artificial-intelligence-to-predict-coronavirus-spread.html. [Accessed: 30-Nov-2022].

1. Datasets

[1] [COVID-19 testing locations - Datasets - Ontario Data Catalogue](https://data.ontario.ca/en/dataset/covid-19-assessment-centre-locations)

[2] [Status of COVID-19 cases in Ontario - Open Government Portal (canada.ca)](https://open.canada.ca/data/en/dataset/f4f86e54-872d-43f8-8a86-3892fd3cb5e6)

[3] [COVID-19 Dataset | Kaggle](https://www.kaggle.com/datasets/imdevskp/corona-virus-report)

1. Google collab link

https://colab.research.google.com/drive/143LZbKj8tpz4XTiN0IF-xVG\_0bqOWMg-#scrollTo=MLfHZonPPg4z

1. Github Link

https://github.com/kimd19/4SC3project