Coronatimes: Covid-19 Prediction using Time Series Forecasting

Bhargav Teja Jakku - #G01334185
Data Anlytics Engineering
George Mason University
Fairfax,VA
bjakku@gmu.edu

Naresh Kukkapalli - #G01340317
Data Anlytics Engineering
George Mason University
Fairfax,VA
nkukkap@gmu.edu

Abstract - The coronavirus family has significant human and animal pathogens. A new coronavirus was found as the cause of a group of pneumonia cases with an unknown etiology in Wuhan, a city in China's Hubei Province, around the end of December 2019. The new coronavirus has quickly spread throughout China, resulting in an epidemic, a pandemic, and a rising number of cases in other nations throughout the world. Coronavirus disease 2019 (COVID-19) is transferred by huge droplets produced by symptomatic patients coughing and sneezing, as well as asymptomatic people before they develop symptoms. The incubation period for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection is thought to be 14 days after exposure, with most cases lasting four to five days. SARS-CoV-2 infection can affect people of all ages; however, it is most common in middle-aged and older people. Fever, dry cough, tiredness, sore throat, rhinorrhea, conjunctivitis headache, myalgia, dyspnea, nausea, vomiting, and diarrhea are common clinical symptoms. As a result, there are no distinct clinical characteristics that reliably distinguish COVID-19 infection from other upper/lower airway viral infections. COVID-19 illness can progress to pneumonia, lung failure, and mortality in a subset of individuals before the end of the first week.

Keywords: *Covid-19, Pandemic, Social distancing, Self-quarantine*

The goal of this paper is to examine the COVID-19 illness, starting with virology and epidemiology and progressing through clinical signs, diagnosis, consequences, and finally potential treatment choices and conclusion.

I. Introduction

The term corona refers to the crown-like appearance that coronaviruses have due to the spike proteins that protrude from them. These spike proteins are critical to the virus's biology. The spike protein is the component of the virus that attaches to a human cell to infect it, allowing the virus to replicate within the cell and spread to other cells. By targeting these spike proteins, some antibodies can protect you from SARS-CoV-2. Because of the importance of this specific part of the virus, scientists who sequence the virus for research use a process known as genomic surveillance to constantly monitor mutations that cause changes to the spike protein.

The seventh human coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was found in Wuhan, Hubei Province, China, during a recent pneumonia outbreak in January 2020. The virus has since spread across the world very quickly. There is no specific line of treatment that has been licensed or a vaccination that is currently available. Various forms of possible vaccines and treatments, on the other hand, have been examined and/or are being tested in clinical studies against COVID-19.

COVID-19 is caused by the SARS-CoV-2 virus. It is a member of the coronavirus family, which includes common viruses that cause everything from head or chest colds to more severe (but rare) diseases like SARS and Middle East respiratory syndrome (MERS).

Coronaviruses, like many other respiratory viruses, spread quickly through droplets that you expel from your mouth or nose when you breathe, cough, sneeze, or speak.

The SARS-CoV-2 virus begins to form genetic lineages as genetic changes to the virus occur over time. The SARS-CoV-2 virus, like any other, has a family tree that can be traced. Sometimes different branches of that tree have different characteristics that affect how quickly the virus spreads, the severity of illness it causes, or the effectiveness of antiviral treatments. Scientists refer to these viruses as "variants." They are still SARS-CoV-2, but their behavior may differ.

In this project we are focusing on COVID-19 confirmed cases and deaths in ten severely infected nations, including the United States, Spain, Italy, the United Kingdom, France, Germany, Russia, Turkey, Iran, and China, as well as the pandemic response on how it effects these countries.

II. Problem statement

Covid-19 has had and continues to have an influence on people all around the world. The major goal is to examine how these nations have been affected and forecast if Covid-19 will rise or reduce in the future. As we know that the covid has affected the entire world it is impossible to predict for all the nations. Due to this we limit ourselves to take three biggest nations and predict the future situations.

III. Project Literature

Since 2020, Covid has impacted everyone's lives as it spreads at a faster rate than anybody could have predicted. To prevent being affected by the virus, everyone began to stay in quarantine and take the appropriate precautions. Each country has suffered its own setbacks, although few appear to be suffering more than others. The development of point-of-care diagnostics, optimizing PPE, and determining the utility of facemasks are among the research priorities. Other priorities include identifying the animal reservoir to prevent further spillover, accelerating the evaluation of therapeutics, particularly remdesivir and Kaletra, which are currently undergoing trials in China, and vaccines, which may prove vital in the long run. All of this necessitates higher financial commitments for both the epidemic response and research.

The excess figures in younger individuals has implications in terms of community transmission and infection of more vulnerable population groups. Many young people live with older people such as parents and grandparents, increasing the possibility of passing on infection to those likely to have severe disease. However, the findings might also be of concern for the younger individuals themselves.[1]

Several studies have found that diabetes patients are more susceptible to some infectious infections, notably those of bacterial origin, due to a dysregulated immune response. Diabetes patients make up a large percentage of COVID-19 patients that

are hospitalize. Chronic comorbidities, in addition to advanced age, enhance the risk of severe COVID-19 infection and death. COVID-19's total worldwide fatality rate of roughly 6.9% is already much greater than the seasonal flu mortality rate. This is especially problematic for patients with pre-existing cardiovascular illness, who have a poorer prognosis and, according to some statistics, a death rate of above 10%.[2]

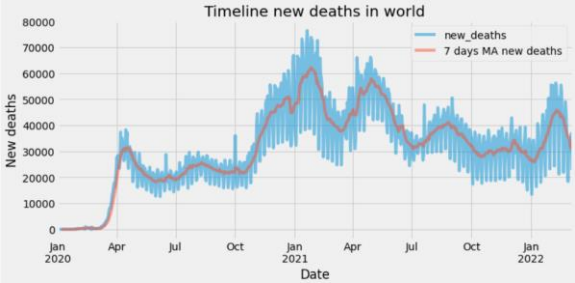


Fig1. New Death cases

According to sources, the COVID-19 has been seen in other Chinese cities as well as 198 other countries. COVID-19 is transmitted from person to person, according to the Centers for Disease Control and Prevention. COVID-19 is transferred via touching surfaces, close contact, air, or things that carry virus particles, according to the CDC. COVID-19 is a hazardous virus since it has a 14-day incubation period and may transmit to others during that time. According to a recent study, the median age and incubation time of confirmed cases are 3 days and 47.0 years, respectively.

The number of confirmed cases is rising every day in several nations, including the United States of America, Italy, Spain, Germany, Iran, China, and others. The spread of COVID-19 poses several risks, necessitating the implementation of stringent particular plans and procedures. As a result, predicting and projecting future verified instances is crucial when considering strategies and regulations. To estimate the number of unreported COVID-19 cases in China, they calculated that there are 469 unreported COVID-19 instances in China between January 1 and 15, 2020, using data-driven research.[4]

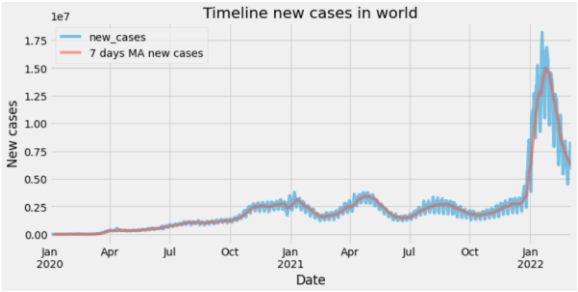


Fig2. New covid cases

With approximately 1000-fold variability between children and senior nursing care residents, COVID-19 mortality risk reveals substantial risk stratification. In Germany, life expectancy is 81 years old, 84 (82) in Italy, 81(85) in the United Kingdom, and 79 (77) in the United States. Divergence may be greater in some low-income countries, such as India29, since many weak persons live to old age in high-income countries but not in low-income nations. A few of the first million fatalities were likely caused by mistakes and mismanagement that might be avoided in the future. For example, some health-care systems were caught off guard3; frequently prescribed hydroxychloroquine has been linked to an increase in mortality40; and inadequate mechanical ventilation management has been linked to worsening outcomes. Some strategic decisions, such as sending COVID-19-infected patients to be nursing homes, are likely to have resulted in many unnecessary deaths41, and nosocomial infections are likely to have contributed to many deaths in some hard-hit areas, such as Lombardy. 3 Many of these issues, hopefully, may be avoided in the future. [5]

The spread of COVID-19 has resulted in a considerable slowdown in economic activities. According to an early forecast of The World Bank, global GDP in 2020 relative to 2019 is forecasted to fall by 5.2%. Similarly, the OECD forecasts a fall in global GDP by 6 to 7.6%, depending on whether a second wave of COVID-19 emerges. In its latest forecast, the International Monetary Fund projected a contraction of 4.4% considering the stronger than expected recoveries in advanced economies which lifted lockdowns during May and June of 2020. This was mainly the result of the unprecedented fiscal, monetary, and regulatory responses in these countries that helped to maintain household disposable income, protect cash flows for firms, and support credit provisions.

The economic implications will be wide ranging and uncertain, with different effects expected on labor markets, production supply chains, financial markets, and GDP levels. The negative effects may vary by the stringency of the social distancing measures (e.g., lockdowns and related restrictions), their length of implementation, and the degree of compliance with them. In addition, the pandemic and the subsequent interventions may well lead to higher levels of mental health distress, increased economic inequality, and particularly harsh effects on certain socio-demographic groups. [6]

As of June 29, 2021, more than 181 million infections had been documented, with approximately 4 million fatalities from COVID-191. The 73rd World Health Assembly passed a resolution in May 2020 acknowledging the need of widespread vaccination as a worldwide public-health priority for preventing, controlling, and terminating SARS-CoV-22 transmission. There are presently more than 125 vaccine candidates worldwide, 365 vaccination studies in progress, and 18 COVID-19 vaccines authorized by at least one country3. Concerns regarding the safety of COVID-19 vaccines, notably the quick pace of vaccine development, have been cited as one of the key reasons for reticence in published studies, however evidence from low- and middle-income countries has been scarce.

More individuals in LMICs can now access the internet and social media thanks to the widespread availability of smartphones. Although this can be a useful tool for self-education, which is an important part of the vaccination decision-making process, it also has its drawbacks, such as misinformation and incomplete information, as well as inconsistent and complicated scientific information that can be difficult to understand. The causes behind COVID-19 vaccination acceptance and skepticism are yet unknown. As additional SARS-CoV-2 variations emerge, adding to the complexity11, and new vaccines enter the market, it will be critical to strike a fine balance between sharing what is known and admitting the unknowns. Researchers and pharmaceutical companies should be as open as possible, with research data on COVID-19 vaccinations publicly accessible. International medical journals should guarantee that the use of 'expedited reviews' does not jeopardize the peer-review process for important articles on vaccination safety and efficacy, as well as associated research findings. Governments should be open about their COVID-19 response initiatives and vaccine availability, as well as the processes used to make crucial choices. [8]

Before AI frameworks can be transmitted in the COVID-19 application, they must be

trained using data derived from medical trials, such as screening diagnosis, virus analysis, and treatment work, among other things. The link between items emphasizes and outputs the premium, in addition to the purpose of being able to learn comparable groups of things.

This clinical data is frequently available, but it is never limited to demographics, clinical notes, and electronic records from clinical gadgets, physical evaluations, and clinical research facilities, as well as photos In particular, one of the key aspects of the AI data writing from determination photos is the finding stage, which also includes hereditary testing and electrodiagnosis. [9]

COVID-19, which has been labeled a worldwide pandemic by the World Health Organization, has emerged as the most aggressive illness, affecting more than 90% of the world's countries. The virus, which began with a single human being in China and has since spread internationally at a pace of 3% to 5% each day, is now a never-ending process. According to some research, the virus will be with us indefinitely. India, the world's second most populous country, is similarly unprotected, and the virus is spreading at a communal level. As a result, it's critical to assess COVID-19's potential influence in India and estimate how it will behave in the next days.

When the overall number of instances grows by a factor of ten, there is cause for alarm. GEP models must be optimized in these situations. As a result, state-of-the-art evolutionary algorithms are used to optimize mathematical equations produced from GEP models. These equations can be deduced in the future, and methods like the Krill herd algorithm, the naked mole-rat algorithm, and others can be employed to optimize prediction models. Furthermore, the forecast models for the next 10 days suggest CC and DC, indicating that rigorous efforts are required to keep the virus under control. Lockdown and social separation should be severely enforced in this situation so that the infection can be contained and limited to certain locations. [10]

IV. Dataset Review

It is difficult to identify the countries that are making the most progress against a rapidly evolving pandemic.

Each profile has four sections:

1. **Cases:** How many new cases are confirmed every day? How many confirmed cases have there been since the pandemic began? What is the trend in the number of cases?
2. **Deaths:** How many COVID-19-related deaths have been reported? Is the number of deaths increasing or decreasing? What is the death rate in comparison to other countries?
3. **Vaccinations:** How many vaccine doses are given out each day? How many doses have been given in total? What percentage of the population has been immunized?
4. **Testing:** How much coronavirus testing do countries do? How many tests did a country perform to find one COVID-19 case?

We need to know what characteristics there in order are to better comprehend the dataset. The dataset has about 160k records and 67 columns, however it needs be preprocessed to remove empty columns and values.

The attributes that the dataset contain are:

S.No	Attribute Name	Type	Example
1	Iso Code	Categorical	AFG
2	Continent	Categorical	Asia
3	Location	Categorical	India
4	Date	Date	24/02/20
5	Total Cases	Numeric	5
6	New Cases	Numeric	5
7	Total Cases Per Million	Float	0.126
8	New Cases Per Million	Float	0.075
9	Stringency Index	Float	8.33
10	Population	Float	39835428
11	Population Density	Float	54.22
12	Median Age	Float	18.6
13	Aged 65+	Float	2.581
14	Aged 70+	Float	1.337
15	GDP Per capita	Float	1803.987
16	Cardiovascular death	Float	597.029
17	Diabetes Prevalence	Float	9.59
18	Handwashing Facilities	Float	37.746
19	Hospital Beds Per Thousand	Float	37.746
20	Life Expectancy	Float	64.83
21	Human Development Index	Float	0.511
22	Hospital Beds Per Thousand	Float	0.5

V. Proposed Methodology

The main purpose is to forecast the trend of nations impacted by the covid using the factors available. This aids in determining which countries are most affected and taking preparations. This may be done by forecasting the trend using time series. Furthermore, we must use regression to identify the best fit for the model to forecast fatalities.

The information was gathered from the popular website 'our world in data.' This page covers all the data for all the nations affected by the Covid-19. The dataset has 67 columns and over 160k data rows (which is considerably large).

VI. Exploratory Analysis

We can obviously go into some EDA because we have such a large dataset. We start by looking at the data to acquire a better understanding of the factors.

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...	female_s
0	AFG	Asia	Afghanistan	2020-02-24	5.0	5.0	NaN	NaN	NaN	NaN	...	
1	AFG	Asia	Afghanistan	2020-02-25	5.0	0.0	NaN	NaN	NaN	NaN	...	
2	AFG	Asia	Afghanistan	2020-02-26	5.0	0.0	NaN	NaN	NaN	NaN	...	
3	AFG	Asia	Afghanistan	2020-02-27	5.0	0.0	NaN	NaN	NaN	NaN	...	
4	AFG	Asia	Afghanistan	2020-02-28	5.0	0.0	NaN	NaN	NaN	NaN	...	

5 rows × 67 columns

Fig.3 Head of the dataset

Anyone with no prior understanding of analytics may notice the locations, number of fatalities, new cases, and other simple factors at first sight. They simply want to know which country has the highest and lowest. However, those who understand how to manipulate variables will be able to produce

some astonishing results. We did our best to demonstrate what we discovered utilizing the approaches.

To begin, we attempted to determine what kind of variables exist. Are they discrete or continuous? Do they have one or more variables? So I just used Python's 'info' function. It displays the dataset's summary information.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3970 entries, 0 to 3969
Data columns (total 67 columns):
 #   Column                                  Non-Null Count  Dtype
---  ------                                  --------------  -----
 0   iso_code                                3970 non-null   object
 1   continent                               3152 non-null   object
 2   location                                3970 non-null   object
 3   date                                    3970 non-null   object
 4   total_cases                             3956 non-null   float64
 5   new_cases                               3955 non-null   float64
 6   new_cases_smoothed                      3930 non-null   float64
 7   total_deaths                            3867 non-null   float64
 8   new_deaths                              3891 non-null   float64
 9   new_deaths_smoothed                     3866 non-null   float64
 10  total_cases_per_million                 3956 non-null   float64
 11  new_cases_per_million                   3955 non-null   float64
 12  new_cases_smoothed_per_million          3930 non-null   float64
 13  total_deaths_per_million                3867 non-null   float64
 14  new_deaths_per_million                  3891 non-null   float64
 15  new_deaths_smoothed_per_million         3866 non-null   float64
 16  reproduction_rate                       3027 non-null   float64
 17  icu_patients                            598 non-null    float64
 18  icu_patients_per_million                598 non-null    float64
 19  hosp_patients                           0 non-null      float64
 20  hosp_patients_per_million               0 non-null      float64
```

Fig.4 Information stats of dataset

We're working with real-world data, and it's evident that null values should be included. We can't just remove the null values since it affects the efficiency of the output if we're doing time series forecasting. As a result, we use the 'ffill' function to fill in the null values with the value adjacent to it. We started by looking at the null values. After that, we filled the values and examined how many of the null values were filled.

```
iso_code                                    0
continent                               10814
location                                    0
date                                        0
total_cases                              6866
                                           ...
human_development_index                 36122
excess_mortality_cumulative_absolute   178790
excess_mortality_cumulative            178790
excess_mortality                       178790
excess_mortality_cumulative_per_million    178790
Length: 67, dtype: int64
```

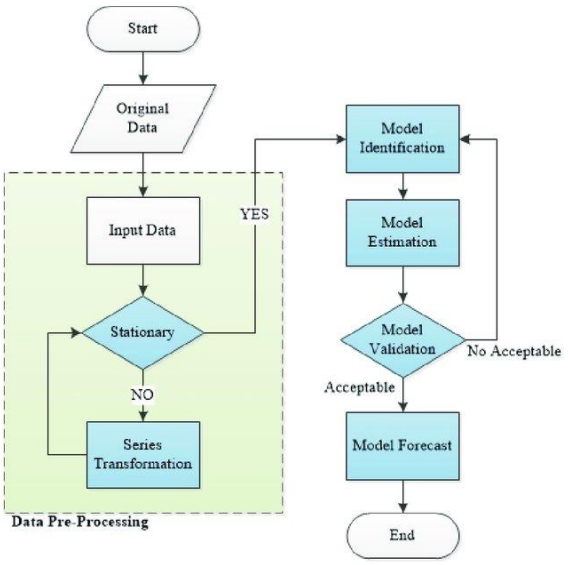
Fig.5 Null values before removing them

```
iso_code                                    0
continent                                   0
location                                    0
date                                        0
total_cases                                 0
                                           ...
human_development_index                     0
excess_mortality_cumulative_absolute     1629
excess_mortality_cumulative              1629
excess_mortality                         1629
excess_mortality_cumulative_per_million  1629
Length: 67, dtype: int64
```

Fig.6 Null values after removing them

VII. Data Preprocessing

A time series is a set of data that is gathered at regular intervals and is uniformly spaced and arranged. As a result, there is the possibility of correlation between the response variables. For the preprocessing, we converted the date format into data and time format for the flexibility that offers us to perform forecasting.



We tried to plot the graph using the data column as the index. We ended up with a jumbled graph. There is so much going on in

the graph, with swings, ups and downs, that it is impossible to analyze anything. We used the 7 days Rolling function of the needed variable to smooth the graph. This procedure is repeated for new instances, deaths, and so on. Figures 1 and 2 illustrate the smoothed graph.

We decided to construct a new function named 'covid' to further preprocess and make things easier. This is used to obtain a smoothed sample graph of the relevant location. We focused on new cases per million, new cases, new fatalities, new deaths per million, and the percentage of vaccinated and completely vaccinated persons in this section. To do this, we built a data frame and dumped four empty columns, which will be utilized to fill the rolling mean values of the variables indicated above. We used the plot function to draw the graph to examine the difference between the original and sampled graphs.

VIII. Methods and Approach

We wanted to observe how Covid -19 has influenced a few countries. Because it is hard to check all the nations, we chose to focus on four and see if any trends emerge. As a result, the countries are the United Kingdom, the United States of America, India, and Russia.

The virus arrived in the country in early 2020, primarily from travel elsewhere in Europe. Various sectors responded, with more widespread public health measures being phased in beginning in March 2020. The first wave was one of the world's largest outbreaks at the time. The peak had passed by mid-April, and restrictions were gradually eased. A second wave began in the autumn and peaked in mid-January 2021, with a new variant that originated in the UK becoming dominant. It was deadlier than the first. In early December 2020, the United Kingdom began a COVID-19 vaccination program. General restrictions were gradually lifted, and most were lifted by August 2021. A third wave, fueled by the new Delta variant, began in July 2021, but the rate of deaths and hospitalizations was lower than in the previous two waves, owing to the mass vaccination program. The Omicron variant had arrived by early December 2021, causing record infection levels.

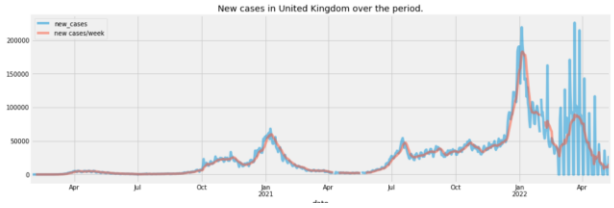


Fig.7 New cases in UK

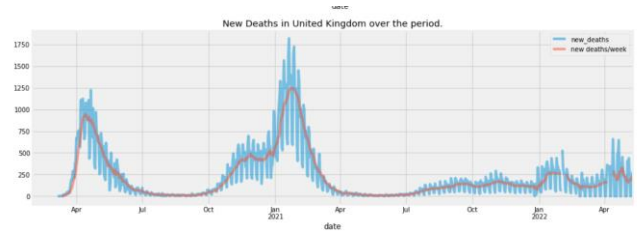


Fig.8 New Deaths in UK

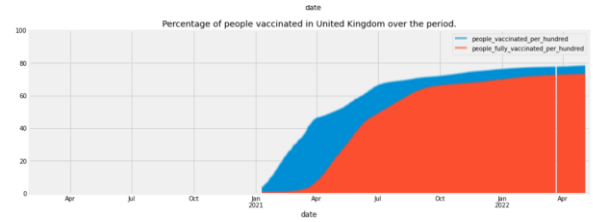


Fig.9 Percentage of partially vaccinated to fully vaccinated in UK

We may detect a nearly linear slope in the beginning of April in Fig.7 since we know that covid reaches us in early January. Because Covid began in January, the number of deaths began to rise in April, which is a proven fact. And, as seen in Fig.8, there is a significant increase in 2022 January, which is due to the new variation known as 'omicron.' In contrary to the initial logic, the number of

deaths is lower than before. This is seen in the following graph, which demonstrates that by the end of March 2021, most individuals will have been partially or completely vaccinated. As a result, the number of deaths was undoubtedly fewer.

Next to UK, we analyzed United states covid spread and where it peaks? What causes the peak.? Was it effective enough to damage the economic system.? The COVID-19 pandemic in the United States is part of the 2019 global coronavirus pandemic caused by the severe acute respiratory syndrome coronavirus. It has resulted in 82,223,174 confirmed cases and 998,997 all-time deaths in the United States, the most of any country and the nineteenth-highest per capita worldwide. The COVID-19 pandemic ranks first in terms of death toll in the United States; it was the third-leading cause of death in the United States in 2020, trailing only heart disease and cancer. Life expectancy in the United States fell by three years for Hispanic Americans, 2.9 years for African Americans, and 1.2 years for white Americans between 2019 and 2020. These effects have persisted, with COVID-19-related deaths in the United States exceeding those in 2020 in 2021.

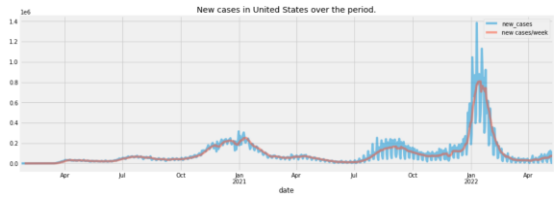


Fig.10 New cases in USA

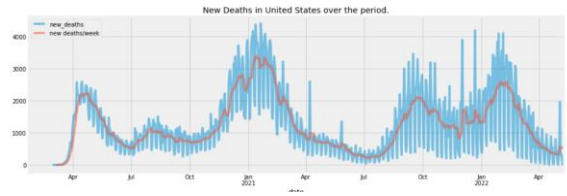


Fig.11 New cases in USA

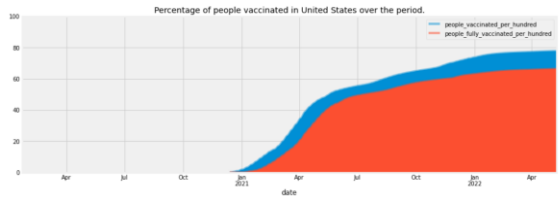


*Fig.12 Percentage of partially vaccinated to fully vaccinated in USA***.**

The CDC discovered that unvaccinated people were 5 to nearly 30 times more likely to become infected or hospitalized than fully vaccinated people. Vaccine skepticism persists for a variety of reasons, even though side effects are uncommon. There have also been numerous reports that unvaccinated COVID-19 patients have strained hospital capacity across the country, forcing many to turn away patients with potentially fatal diseases.

The first COVID-19 cases in India were reported on January 30, 2020, in three towns in Kerala, among three Indian medical students who had returned from Wuhan, the pandemic's epicenter. Lockdowns were declared in Kerala on March 23, and the rest of the country on March 25. For the first time on June 10, India's recoveries exceeded active cases. In September, infection rates began to fall, as did the number of new and active cases. Daily cases peaked in mid-September, with over 90,000 reported per day, before falling to less than 15,000 in January 2021. The second wave, which began in March 2021, was far more devastating than the first, with parts of the country experiencing shortages of vaccines, hospital beds, oxygen cylinders, and other medical supplies. By late April, India had surpassed the rest of the world in terms of new and active cases. It became the first country to report over 400,000 new cases in a 24-hour period on April 30, 2021. According to experts, the virus may reach an endemic

stage in India rather than completely disappear; in late August 2021, Soumya Swaminathan stated that India may be in an endemic stage where the country learns to live with the virus. By March 2022, India had only 22,487 cases nationwide. With 58.8 percent of the population fully vaccinated and 70% having received at least one dose, the post-pandemic recovery has been steady.

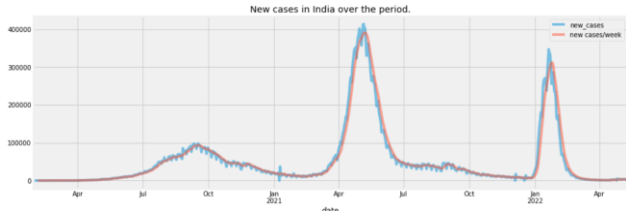


Fig.13 New Cases in India.

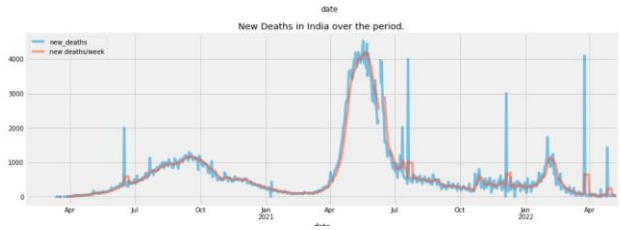


Fig.14 New Deaths in India

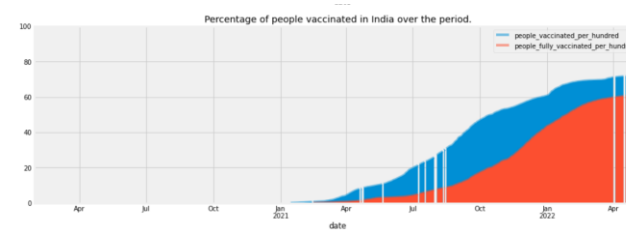


*Fig.15 Percentage of partially vaccinated to fully vaccinated in India***.**

Russia ranks seventh in the world in terms of confirmed cases, trailing only the United States, India, Brazil, France, Germany, and the United Kingdom. Between April and November 2020, 114,268 people with COVID-19 died, according to detailed data published by the Federal State Statistics Service. Over 300,000 extra deaths were reported during the same time, implying that the official pandemic death tally greatly underestimated the true number of COVID-19-related deaths. Excess deaths from official government demographic statistics, based on births and deaths but excluding migration, revealed that Russia experienced its largest annual population drop in peacetime, with the population declining by 997,000 between October 2020 and September 2021, which demographer Alexei Raksha attributed primarily to the COVID-19.

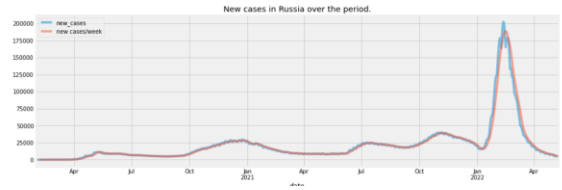


Fig.16 New Cases in Russia

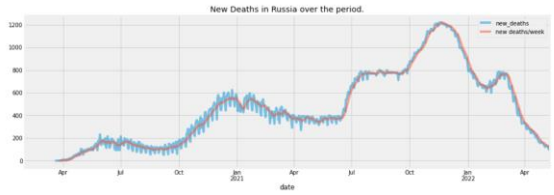


Fig.17 New Deaths in Russia

We can see that there are two peaks in new cases in India, indicating that there have been moments when many cases have been reported. When Delta and Omicron Variant are present on the timeline, this occurs. As previously said, most patients have received their first and second doses, thus the curve will undoubtedly remain steep. However, we must realize that there were more deaths in the beginning, maybe because of people failing to recognize the power and value of vaccinations.

We've spoken about how covid acts in different nations so far. We now want to know how many new cases and fatalities are occurring in the nations stated above.

date	uk new cases per million	us new cases per million	india new cases per million	uk new deaths per million	us new deaths per million	india new deaths per million
2020-01-31	NaN	0.002571	NaN	NaN	NaN	NaN
2020-02-01	NaN	0.002571	NaN	NaN	NaN	NaN
2020-02-02	NaN	0.001286	NaN	NaN	NaN	NaN
2020-02-03	NaN	0.002571	NaN	NaN	NaN	NaN
2020-02-04	NaN	0.002571	NaN	NaN	NaN	NaN

Fig.18 Dataset of Rolling means

The area under the curve, which corresponds to the number of cases per million in the whole pandemic era, in other words, the total number of cases per million, is an intriguing approach to quantify the meaning of these curves. We did it for both new cases per million and new deaths per million.

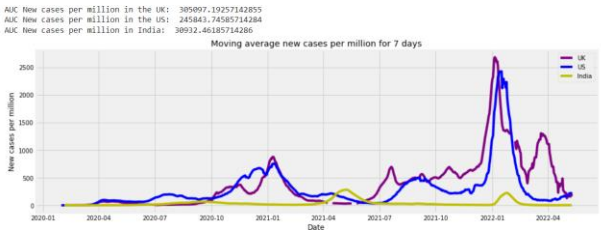


Fig.19 Moving Avg new cases per million

In Fig.19 we can see that omicron has a big impact for all the countries. There is a considerable raise for UK, USA, and India. Out of three countries UK and USA begins with almost same new cases while in the end USA has recorded less no of cases when compared to UK. Anyways India did a great job in maintain low new cases. To get a better picture we print down the new cases per million for three countries. UK tops with 305097 cases followed by USA with 245843 cases and India has the lowest cases with 30932 cases.

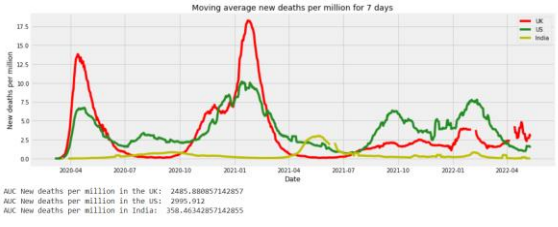
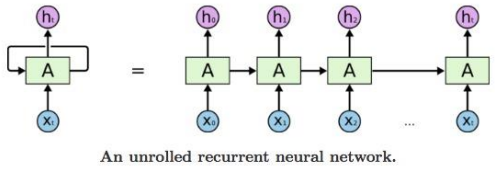


Fig.20 Moving Avg new deaths per million

In fig20 the deaths were at their peak in the beginning phase of covid and when delta variant hits. But there isn't a good raise in deaths in the Jan 2022 because people got vaccinated.

IX. Time Series Forecasting Model

A Recurrent Neural Network (RNN) is a type of Artificial Neural Network in which the connections between nodes form a directed graph and the behavior is temporally dynamic. It aids in the modeling of feedforward network-derived sequential data. It predicts outcomes in the same way that human brains do. A recurrent neural network resembles a regular neural network with the exception that the neurons have a memory state. The calculation to incorporate a memory is straightforward.



An unrolled recurrent neural network.

The following forecast was mostly accomplished in two steps. We developed a function named 'Covid model' in the first stage. This method takes an empty dataset and fills it with the rolling mean of new cases and deaths. We partitioned the dataset into 80 percent training and 20% test sets, using a 10 percent window. To anticipate the covid's projected cases, we used basic Rnn with three thick layers and 100 epochs. The model was then used to forecast the values, which were then dumped into a new data frame alongside the actual test data points. We put it all together to check if the validation points corresponded to the expected sites.

We developed another function containing the dataset, model, and the number of days for which the outcome should be forecasted to predict future events. The projected outputs were plotted using the plot function. We opted to acquire the root mean square log error of each prediction to verify confidence in addition to receiving the anticipated points.

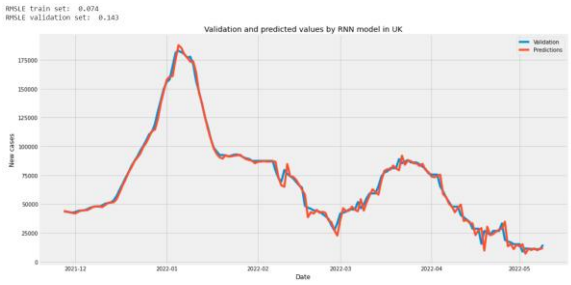


Fig.21 Validation and predicted values in UK

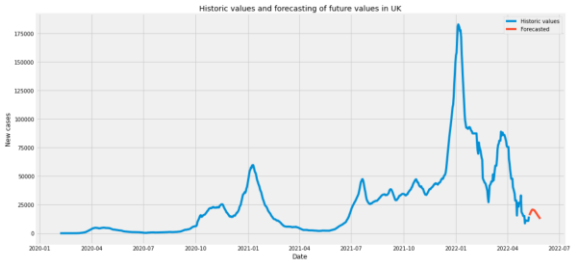


Fig.22 Forecasted outcomes in UK

As two graphs are comparable, our model works well in forecasting new cases in the United Kingdom. The function uses the prior data to get the needed anticipated output for the next 20 days by watching the Forecasted new instances. Figure 22 depicts the projected values, which indicate a little increase followed by a sharp decline. In addition, the train and validation RMSLEs are 0.074 and 0.143, respectively. The lower the RMSLE, the more confidence you get.

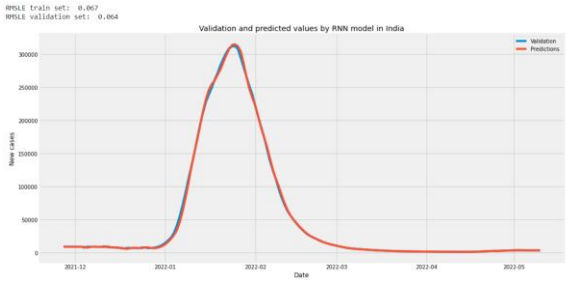


Fig.23 Validation and predicted values in INDIA

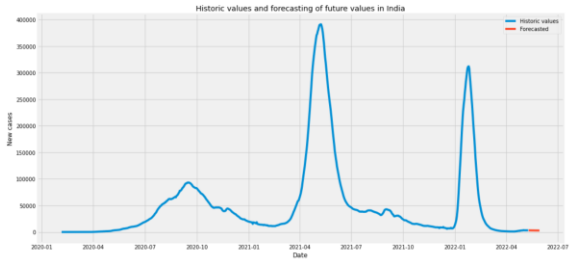


Fig.24 Forecasted outcomes in INDIA.

We are quite impressed with the model, which appears to have only one curve in Figure 23. Validation and prediction values are nearly identical. The predicted values appear to be in line with the preceding graph. The train and validation RMSLE values are 0.067 and 0.064, respectively. The lower the RMSLE, the more confident you get.

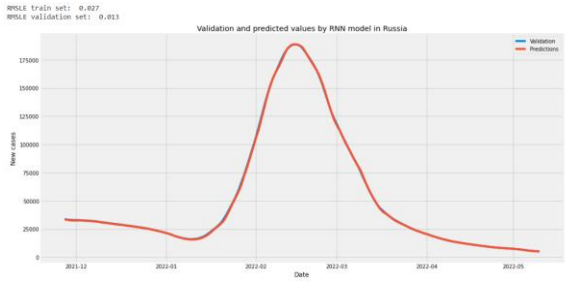


Fig.25 Validation and predicted values in Russia

Fig.26 Forecasted outcomes in Russia.

Our model operates well in Russia since the validation and prediction values are nearly identical. The expected output shows that the number of instances will decrease. The train and validation RMSLE values are 0.027 and 0.013, respectively. The lower the RMSLE, the more confidence you get.

Country	RMSLE Training Set	RMSLE Validation Set
UK	0.074	0.143
India	0.067	0.064
Russia	0.027	0.013

X. Conclusion

In conclusion, we started off by finding some insights about the number of cases that are getting recorded. Then we preprocess the dataset to our requirement. In the meanwhile, we created three function which helps us in getting the sampled graph, to predict the cases and to forecast the future cases. Our model is working exceptionally as it predicts the values with the lowest root mean sq log error values. Thus, with the functions and the models we used in the project we can easily find out the stats and predict the upcoming event values. Out of all the countries we have taken into consideration the situation of covid in USA is much worse than the UK and India.

XI. Future Studies

We've been using the dataset to evaluate validation scores and anticipate future occurrences so far. In this study, we focused primarily on the number of new instances or fatalities that have emerged or may arise. To learn more about this subject, we'd like to learn more about how many ICU patients are reported and what the hospital rate is (which are not properly mentioned in the dataset.

References

[1] An RWJF Collection. (2021). *The Impact of Coronavirus on Households Across America*. Retrieved from Robert wood johnson Foundation: https://www.rwjf.org/en/library/research/2020/09/the-impact-of-coronavirus-on-households-across-america.html

[2] Abel Brodeur, D. G. (2021, April 18). *A literature review of the economics of COVID-19*. Retrieved from Wiley Online Library: https://onlinelibrary.wiley.com/doi/full

[3] ADEDOYIN AHMED HUSSAIN, O. B.-T. (2020, june 25). *AI Techniques for COVID-19*. Retrieved from IEEE Access: https://ieeexplore.ieee.org/stamp/

[4]cdc.gov. (2022). *COVID Data Tracker*. Retrieved from Centers for Disease control and prevention: https://covid.cdc.gov/covid-data-tracker/#datatracker-home

[5] Conger, K. (2022, March 11). *Two years into the pandemic, medical scientists consider what the future may hold*. Retrieved from Standford Medicine news center: https://med.stanford.edu/news/all-news/2022/03/scientists-pandemic-future.html

[6] Dagmar Kutsar, K.-K. (2021, aug 25). *The Impact of the COVID-19 Pandemic on Families*. Retrieved from frontiers in Sociology : https://www.frontiersin.org/articles/10.3389/fsoc.2021.732984/full

[7] Hannah Ritchie, E. M.-G.-O. (2022). *Our World in Data*. Retrieved from

Coronavirus Pandemic (COVID-19): https://ourworldindata.org/coronavirus

[8] Ioannidis, J. P. (2020, October 07). *Global perspective of COVID-19 epidemiology for a full-cycle pandemic*. Retrieved from Wiley Online Library : https://onlinelibrary.wiley.com/doi/full

[9] Johnson, D. C. (2021, March 24). *Past is future for the era of COVID-19 research in the social sciences* . Retrieved from PNAS: https://www.pnas.org/doi/10.1073/pnas.2104155118

[10] MD, S. P. (2020, May 20). *Diabetes and COVID-19*. Retrieved from Springer Link: https://link.springer.com/article

[11]MohsenMaleki, M. R.-H. (2020, November). *Modeling and forecasting the spread and death rate of coronavirus (COVID-19) in the world using time series models*. Retrieved from sciencedirect: https://www.sciencedirect.com/science/article

[12] NCBI. (2022, Jan 01). *PMC*. Retrieved from National library of Medicine: https://www.ncbi.nlm.nih.gov/pmc/

[13] RohitSalgotraaMostaf, G. A. (September, 2020). *Time Series Analysis and Forecast of the COVID-19 Pandemic in India using Genetic Programming*. Retrieved from Science Direct: https://www.sciencedirect.com/science/article/pii

[14] Thomson, B. (2020, April 23). *Circulation*. Retrieved from The COVID-19 Pandemic A Global Natural Experiment: https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA

[15] wikipedia. (2022, March 01). *Covid*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/COVID-19

[16] Wiysonge, S. M. (2021, july 16). *Understanding COVID-19 vaccine hesitancy*. Retrieved from Nature Medicine: https://www.nature.com/articles/s41591-021-01459-7

[17] Worldometer . (2022, May 13). *COVID-19 CORONAVIRUS PANDEMIC*. Retrieved from Worldometer : https://www.worldometers.info/coronavirus/