



BIG DATA ANALYTICS FOR COVID-19: DATA VISUALIZATION, TREND ANALYSIS, AND MACHINE LEARNING APPLICATIONS

¹Mohammad Naeem Kanyar

¹Independent Researcher

¹ Punjab, India

¹naeem.kanyar@gmail.com

Abstract: The COVID-19 pandemic generated large volumes of epidemiological data that can be analyzed to understand disease trends, monitor reported cases, and support public-health decision-making. This study examines the application of big data analytics, data visualization, and machine-learning techniques to historical COVID-19 data. The analysis uses a dataset containing reported confirmed cases, deaths, and recoveries across countries and regions. Descriptive analysis and visualization are used to examine changes in reported cases over time, while mortality and recovery proportions are calculated to summarize selected epidemiological indicators. The study also discusses the potential applications of machine learning in case prediction and pandemic monitoring. However, the reliability of these analyses depends on data completeness, consistency in reporting, and the suitability of the analytical methods. The limitations of simple linear regression for representing complex epidemic trends are also considered. This study highlights the importance of transparent data preprocessing, appropriate model evaluation, and careful interpretation of epidemiological indicators when applying data analytics to pandemic-related research.

Index Terms – Big data analytics, COVID-19, data visualization, epidemiological data, machine learning, mortality analysis, time-series analysis.

1 Introduction

The COVID-19 outbreak became a global public-health emergency in 2020, creating significant challenges for healthcare systems, governments, and communities worldwide. The rapid spread of SARS-CoV-2 highlighted the importance of timely epidemiological surveillance, reliable data collection, and evidence-based decision-making. Reported case counts, deaths, and recoveries provided important indicators for monitoring the development of the outbreak, although differences in testing capacity and reporting practices affected comparisons between regions. Analyzing these data using statistical methods, visualization techniques, and computational tools can help researchers identify historical patterns and communicate changes in the pandemic over time [1].

Timely identification and monitoring of potential COVID-19 cases can support public-health surveillance and resource planning. These activities generate data from multiple sources, including laboratory reports, hospital records, public-health databases, and digital monitoring systems. Big data analytics provides methods for processing and examining large, diverse, and frequently updated datasets. Commonly discussed characteristics of big data include volume, velocity, variety, veracity, and value; additional characteristics, such as variability, may also be considered depending on the framework used. In the context of COVID-19, analytical methods can support descriptive statistics, visualization, trend analysis, and the exploration of possible forecasting approaches. Nevertheless, data quality, privacy, reporting inconsistencies, and methodological limitations must be considered when interpreting results [2], [3].

Wearable devices and mobile health applications can provide additional information for remote monitoring, subject to appropriate consent, privacy safeguards, and data-quality controls. When combined with other epidemiological data, such information may support research into symptom monitoring and disease surveillance. However, the usefulness of these technologies depends on the representativeness of the collected data, the reliability of the measurements, and the suitability of the analytical methods. Estimates of healthcare-resource requirements should therefore be supported by clearly described data and validated analytical procedures rather than inferred from the availability of wearable-device information alone, during pandemics or similar critical situations.

2 APPLICATIONS OF BIG DATA ANALYTICS FOR COVID-19

The COVID-19 pandemic generated epidemiological, clinical, and public-health data from a variety of sources. Depending on the availability and quality of the data, analytical methods can support several activities, including diagnostic decision support, risk assessment, healthcare planning, and pharmaceutical research. Figure 1 summarizes these potential application areas. The figure is conceptual: it identifies possible uses of data analytics and does not imply that every application was implemented or evaluated in this study.

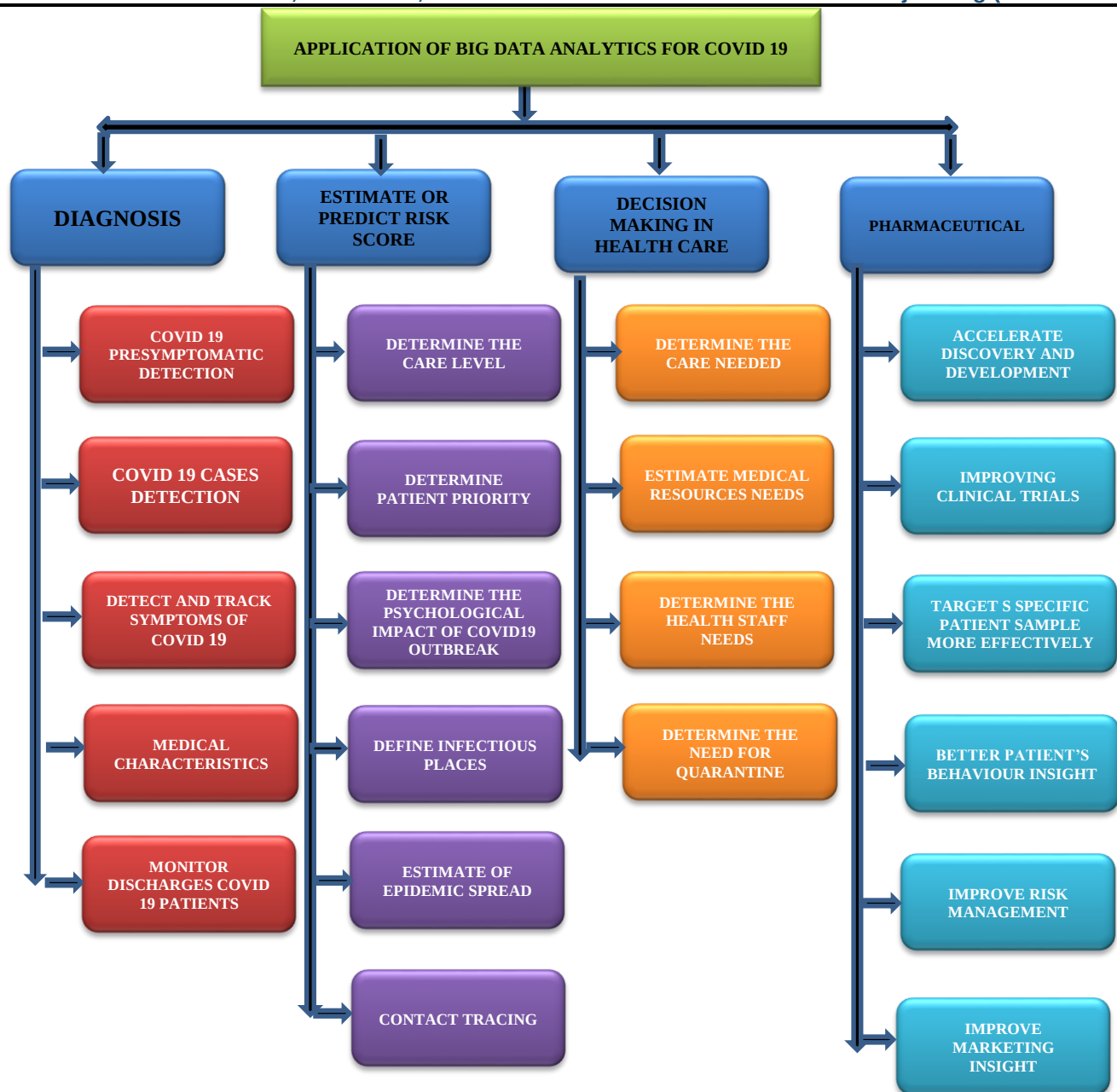


Figure 1. Areas where big data analytics could be used for COVID-19

By primarily concentrating on the evaluation of research works that have presented remedies for the management of the COVID-19 pandemic and specifically fall into one of the three specified domains, namely (1) diagnosis (Section 2.1), (2) risk score estimation or prediction (Section 2.2), and (3) healthcare decision-making, we proceed to illustrate various instances of COVID-19 data application derived from existing scholarly sources within the subsequent subsections.

2.1 DATASET ANALYSIS

This analysis uses the COVID-19 dataset identified in the manuscript as covid_19_data.csv, obtained from Kaggle. The dataset was published by Sudalai Rajkumar (SRK) and contains daily records of reported COVID-19 cases, including observation date, province or state, country or region, last-update information, confirmed cases, deaths, and recovered cases. These fields support descriptive analysis of reported COVID-19 outcomes over time and across geographical areas. The dataset is available at: <https://www.kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset>. The dataset's exact contents, observation period, and coverage should be verified against the version used in the analysis before drawing conclusion, [18].

The dataset contains the following fields:

- SNo: Unique serial number or record identifier, as provided in the original dataset.
- Observation Date: Date associated with the reported observation.
- Province/State: Province or state associated with the record, where available.
- Country/Region: Country or region associated with the record.
- Last Update: Last-update timestamp supplied by the dataset, where available.
- Confirmed: Reported cumulative number of confirmed COVID-19 cases.
- Deaths: Reported cumulative number of deaths attributed to COVID-19.
- Recovered: Reported cumulative number of recovered cases.

Dataset source: [Novel Corona Virus 2019 Dataset — Kaggle](https://www.kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset)

The dataset can be used to examine the distribution of reported confirmed cases, deaths, and recoveries across countries or regions and to visualise how these counts change over the observation period. Comparisons between countries should account for differences in population size, testing practices, reporting systems, and data completeness.

The dataset can therefore support descriptive exploration of the pandemic, but the resulting patterns should not be interpreted as direct evidence of the effectiveness of a country's healthcare system.



```
covid=pd.read_csv("../input/novel-corona-virus-2019-dataset/covid_19_data.csv")
covid.head()
```

[2]:

	SNo	ObservationDate	Province/State	Country/Region	Last Update	Confirmed	Deaths	Recovered
0	1	01/22/2020	Anhui	Mainland China	1/22/2020 17:00	1.0	0.0	0.0
1	2	01/22/2020	Beijing	Mainland China	1/22/2020 17:00	14.0	0.0	0.0
2	3	01/22/2020	Chongqing	Mainland China	1/22/2020 17:00	6.0	0.0	0.0
3	4	01/22/2020	Fujian	Mainland China	1/22/2020 17:00	1.0	0.0	0.0
4	5	01/22/2020	Gansu	Mainland China	1/22/2020 17:00	0.0	0.0	0.0

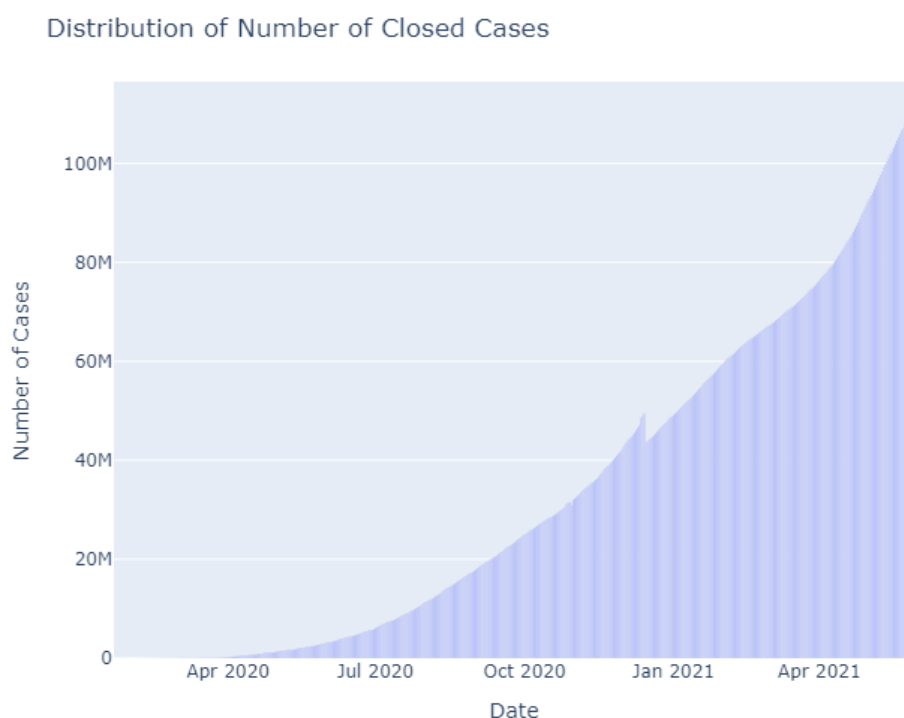
2.2 DIAGNOSIS

Laboratory testing, including reverse-transcription polymerase chain reaction (RT-PCR), has been an important method for detecting SARS-CoV-2 infection. The time required to obtain results can vary according to laboratory capacity, sample transport, testing procedures, and demand. Researchers have also investigated computational approaches that use clinical information, physiological measurements, and other available data to support diagnosis and patient monitoring. Such approaches may combine sensor data with analytical models to identify patterns associated with illness. However, computational systems should be evaluated against appropriate clinical reference standards and must not be presented as replacements for validated diagnostic procedures without adequate evidence. Previous research has explored the potential roles of artificial intelligence, big data analytics, and feature-selection methods in COVID-19 detection and surveillance [3], [4].

A flexible and affordable medical gadget that may be used to identify and monitor COVID-19 symptoms was developed by Radovan Stojanovi et al. [5]. Utilizing a mobile device and earphones, the identification of respiratory ailments is facilitated. The mobile application facilitates the collection and storage of signals in the audio file format. Subsequently, the MATLAB software is employed to scrutinize the signals and ascertain the respiratory symptoms associated with COVID-19. Richardson S. et al. devised a remote monitoring system for discharged COVID-19 patients, incorporating a program. Each registered patient receives a thermometer and pulse oximeter, enabling them to self-report their daily symptoms, oxygen saturation levels, and body temperature. [6]. A team of nursing professionals will assess the abnormal vital signs and symptoms subsequent to receiving an alert. The patient's potential re-admittance to the emergency room will depend on the outcomes of this assessment. Upon a patient's discharge from the hospital, the software provides remote monitoring capabilities that can be adjusted according to needs, thereby contributing to a reduction in emergency room utilization. Tejaswini Mishra and her colleagues conducted an extensive review of various studies pertaining to mathematical models aimed at enhancing the efficacy of COVID-19 detection and prediction. Based on their investigation, the utilization of artificial intelligence for identifying COVID-19 cases, big data for tracking these cases, and nature-inspired computing (NIC) for selecting relevant attributes would significantly enhance the accuracy of detection. [7]. Fan-Yun Lan et al. conducted a COVID-19 examination on a cohort of healthcare professionals with the objective of identifying the symptoms associated with positive outcomes in HCWs. Initial evaluation was performed via telephonic communication, and each HCW underwent a COVID-19 polymerase chain reaction (PCR) test to ascertain the symptoms linked to their respective cases. [8]. The investigation revealed that typical symptoms observed in positive cases of COVID-19 were fever, myalgia, and anosmia/ageusia. Conversely, the majority of negative cases either displayed no symptoms at all or exhibited very mild symptoms such as sore throat and nasal congestion. To examine the distribution of active COVID-19 cases from April 2020 to April 2021, an analysis of the provided data columns is necessary. These columns include SNo, ObservationDate, Province/State, Country/Region, Last Update, Confirmed, Deaths, and Recovered. Firstly, it is important to establish a clear definition of "number of active cases." Active cases refer to individuals presently infected with COVID-19 who have not yet recovered or passed away. Therefore, the number of active cases is determined by the difference between confirmed cases and the sum of deaths and recovered cases. The analysis can commence by grouping the data based on the Observation Date and Country/Region columns. This grouping will enable observation of how the number of active cases fluctuates over time in each country. Subsequently, the average number of active cases per day can be calculated for each country within the specified period. Furthermore, a line chart can be created to visually represent the trend of active cases over time in each country. This visualization will aid in identifying any notable increases or decreases in active cases within the specified period. Additionally, the total number of active cases for each country during the specified period can be computed, allowing for a ranking of countries based on their respective total active cases. This ranking will highlight which countries had the highest number of active cases during the specified period. Moreover, it is possible to calculate the mortality rate (deaths per confirmed cases) and the recovery rate (recovered cases per confirmed cases) for each country during the specified period. These calculations provide insights into the effectiveness of each country's healthcare system in managing the COVID-19 pandemic. In summary, analyzing the distribution of active cases between April 2020 and April 2021 necessitates employing data cleaning, grouping, and visualization techniques. The resulting insights from this analysis can inform policymakers and healthcare professionals in making well-informed decisions regarding the control of COVID-19 transmission. An increase in the number of active cases could indicate a declining ratio of recovered or deceased cases to confirmed cases. The relationship between changes in

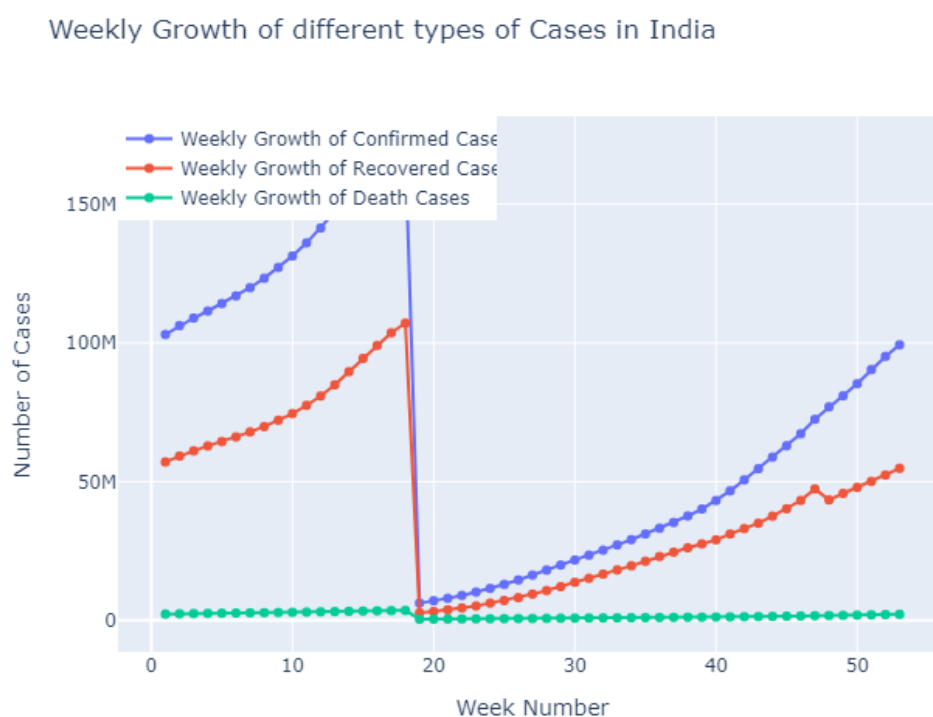
estimated active cases and the proportions of reported recoveries and deaths should be evaluated using the underlying dataset. Any conclusions should be based on the results of the completed analysis rather than on an assumed relationship.

Active Cases are calculated as follows: Estimated active cases = cumulative confirmed cases – cumulative recovered cases – cumulative deaths.



Closed cases = cumulative recovered cases + cumulative deaths.

An increase in closed cases indicates an increase in the cumulative number of reported recoveries and deaths. This measure alone does not indicate whether the increase is primarily due to recoveries or deaths; the two outcomes should be examined separately.



Growth of different types of cases

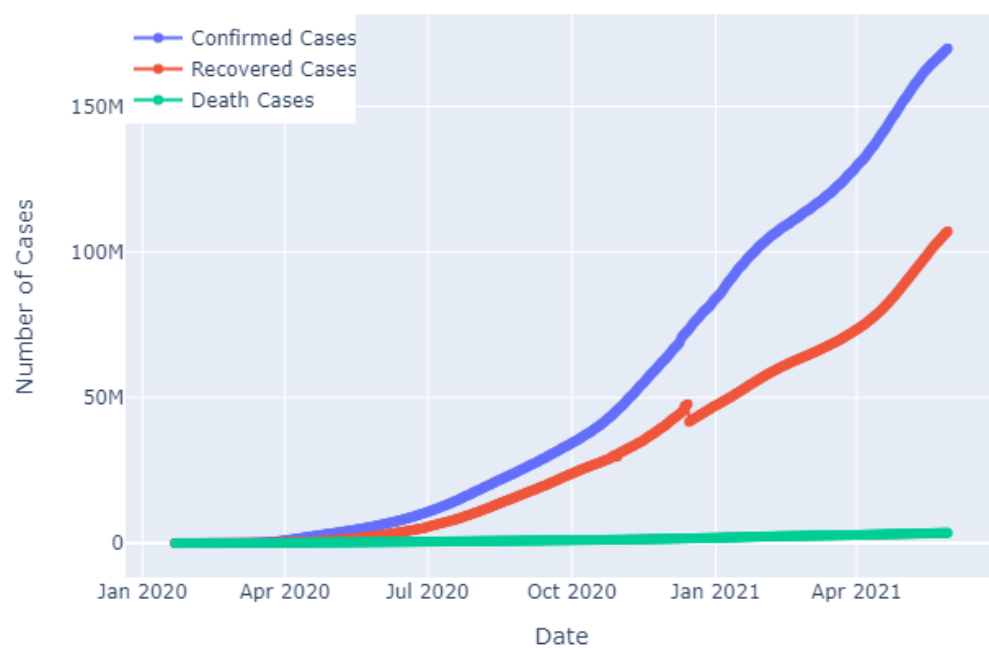
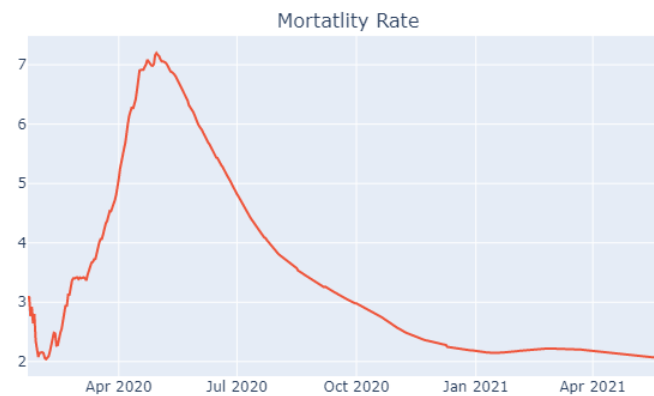
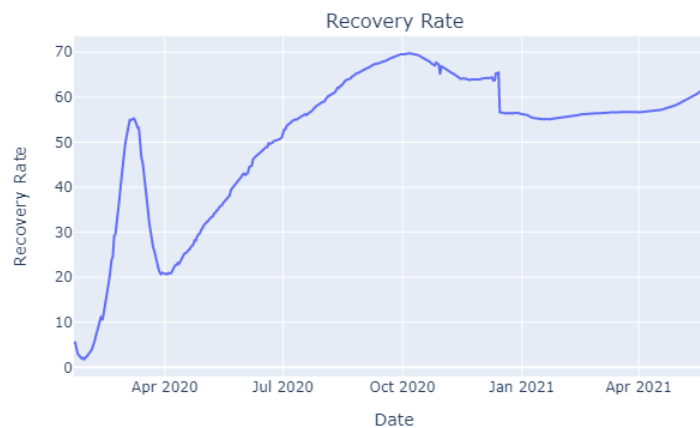


Figure 2. Reported cumulative confirmed cases, recoveries, and deaths over the observation period.

Analysis of Mortality and Recovery Rates Worldwide

— Recovery Rate
— Mortality Rate



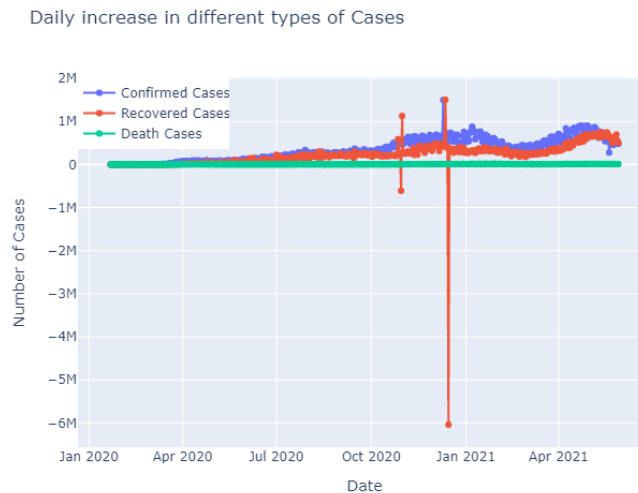
The reported case-fatality ratio is calculated as:

$$\text{Case-fatality ratio (\%)} = \frac{\text{Cumulative deaths}}{\text{Cumulative confirmed cases}} \times 100$$

The reported recovery proportion is calculated as:

$$\text{Recovery proportion (\%)} = \frac{\text{Cumulative recovered cases}}{\text{Cumulative confirmed cases}} \times 100$$

Changes in the reported case-fatality ratio and recovery proportion can describe how recorded deaths and recoveries evolve relative to confirmed cases over time. However, these trends must be interpreted cautiously because they may be affected by reporting delays, changes in testing, differences in case definitions, and incomplete outcome data.



2.3 DECISION MAKING IN HEALTHCARE

The COVID-19 pandemic placed substantial demands on emergency departments, hospital beds, ventilators, personal protective equipment, and healthcare personnel. Researchers explored mathematical models and digital applications to support public-health planning and healthcare decision-making. For example, compartmental models were used to investigate how interventions and changes in contact patterns could affect the spread of infection. Mobile applications and symptom-reporting systems were also investigated as ways to collect information that could support surveillance. Such models and applications depend on the quality of their input data and the assumptions used in their design. Their projections should therefore be interpreted as model-based estimates rather than guaranteed outcomes [1], [2], [9].

3 DATA ANALYTIC TOOLS

Data analytics can support the examination of COVID-19 data through data cleaning, statistical summaries, visualisation, geographic analysis, and predictive modelling. The choice of tool depends on the research question, data format, available computing resources, and required level of reproducibility. The following subsections describe examples of tool categories that may be used in epidemiological data analysis.

- Statistical computing: Tools such as Python and R can be used for data cleaning, statistical summaries, and visualisation.
- Geographic information systems (GIS): Tools such as QGIS and ArcGIS can map geographical distributions when location data are available and sufficiently reliable.
- Machine learning: Classification, regression, and clustering methods can be investigated for suitable prediction or pattern-recognition tasks. Their performance must be evaluated using appropriate validation methods.
- Social network analysis: Network-analysis methods can examine contact structures or patterns of interaction when relevant network data are available.
- Business intelligence and dashboards: Dashboard tools can present reported case counts and trends in a form that supports monitoring and communication.

4 Global Trends in Reported Confirmed COVID-19 Cases

The rate of exponential growth of confirmed COVID-19 cases worldwide refers to the duration it takes for the number of confirmed cases to multiply by two. This metric holds significant importance as it offers insight into the extent and gravity of the virus's transmission. The rate of doubling can vary due to numerous factors, including the efficacy of implemented measures to impede the virus's spread, the availability and effectiveness of vaccines, and the prevalence of novel virus variants. Initially, during the onset of the pandemic, the rate of doubling for confirmed cases was considerably high, with instances of cases doubling within a few days in certain countries. Nonetheless, as more stringent measures like lockdowns and social distancing protocols were enforced to curb the virus's propagation, the rate of doubling gradually diminished. Moreover, the accessibility of vaccines has played a pivotal role in significantly reducing the rate of doubling in numerous countries. Vaccination efforts have proven effective in diminishing the virus's transmission, as well as mitigating severe illness and fatalities. However, the emergence of new virus variants has presented fresh obstacles in controlling the rate of doubling. Notably, some variants, such as the Delta variant, exhibit significantly higher transmissibility than earlier strains, resulting in a substantial upsurge in the rate of doubling within certain countries. Overall, the rate of doubling confirmed cases of COVID-19 worldwide represents a multifaceted and ever-evolving metric that relies on a diverse range of factors. It remains imperative for policymakers and health officials to continually monitor the rate of doubling and implement measures aimed at curbing the virus's spread and safeguarding public health.

[49]:

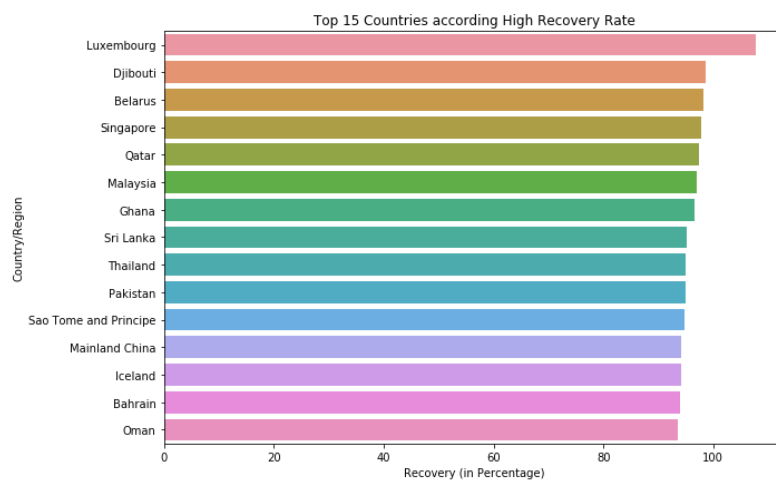
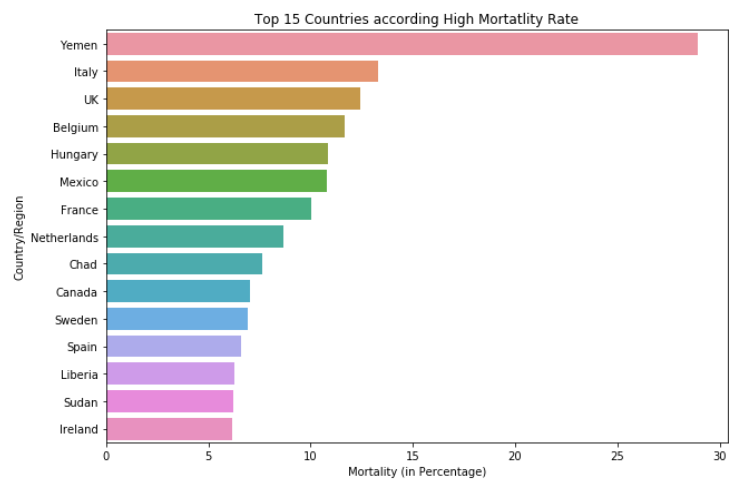
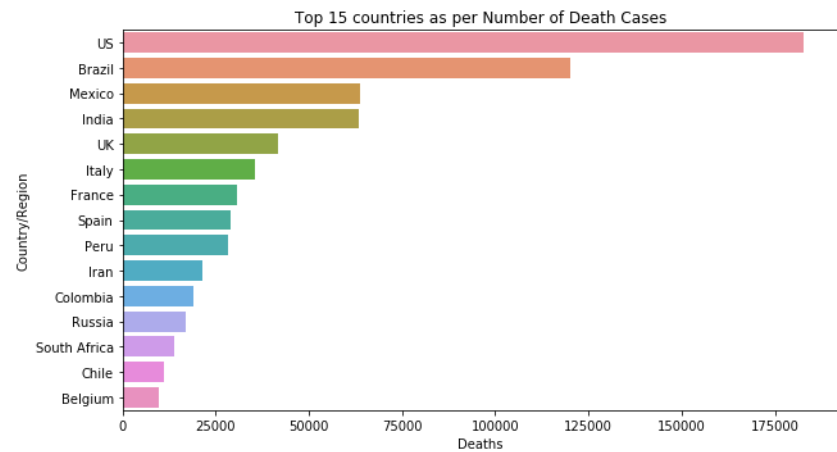
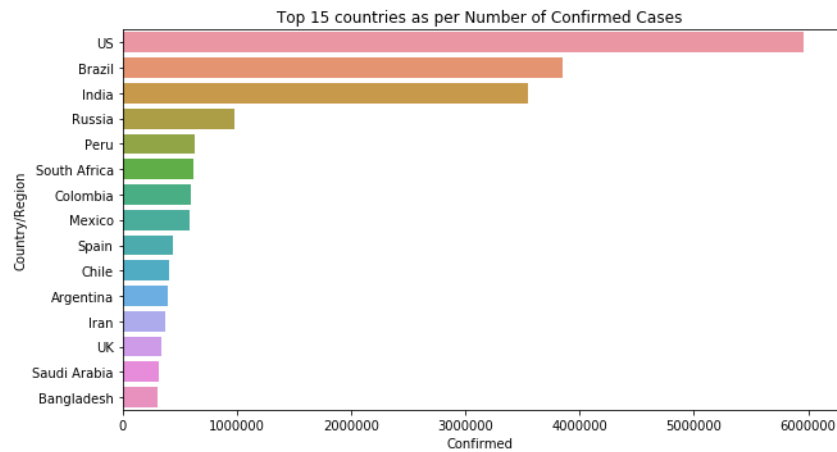
	No. of cases	Days since first Case	Number of days for doubling
0	560	0 days	0 days
1	1120	2 days	2 days
2	2240	4 days	2 days
3	4480	5 days	1 days
4	8960	8 days	3 days
5	17920	11 days	3 days
6	35840	16 days	5 days
7	71680	25 days	9 days
8	143360	50 days	25 days
9	286720	58 days	8 days
10	573440	64 days	6 days
11	1146880	72 days	8 days
12	2293760	86 days	14 days
13	4587520	114 days	28 days

4.1 PROPORTION OF COUNTRIES IN CASES OF DEATH, RECOVERY, AND CONFIRMATION

[55]:

	Country Name	Proportion of Confirmed	Proportion of Recovered	Proportion of Deaths
0	US	49.012	0	33.1365
2	Brazil	11.8072	6.84634	10.2266
5	UK	11.6524	0.0128054	9.18873
3	Russia	5.68365	5.56678	5.51324
7	Italy	4.88255	3.43203	5.28471
4	France	4.16817	0.231678	2.39002
9	Germany	4.02778	0.936958	5.34184
8	Spain	3.75326	0	1.40926
16	South Africa	3.74349	2.37864	4.15159
20	Czech Republic	3.54363	1.59535	1.43782
10	Colombia	3.39285	2.21414	2.89469
6	Turkey	2.99063	4.33826	2.27576
13	Poland	2.80013	2.01911	5.09427

4.2 Top 25 Countries with more than 500 confirmed cases based on mortality rate and recovery rate



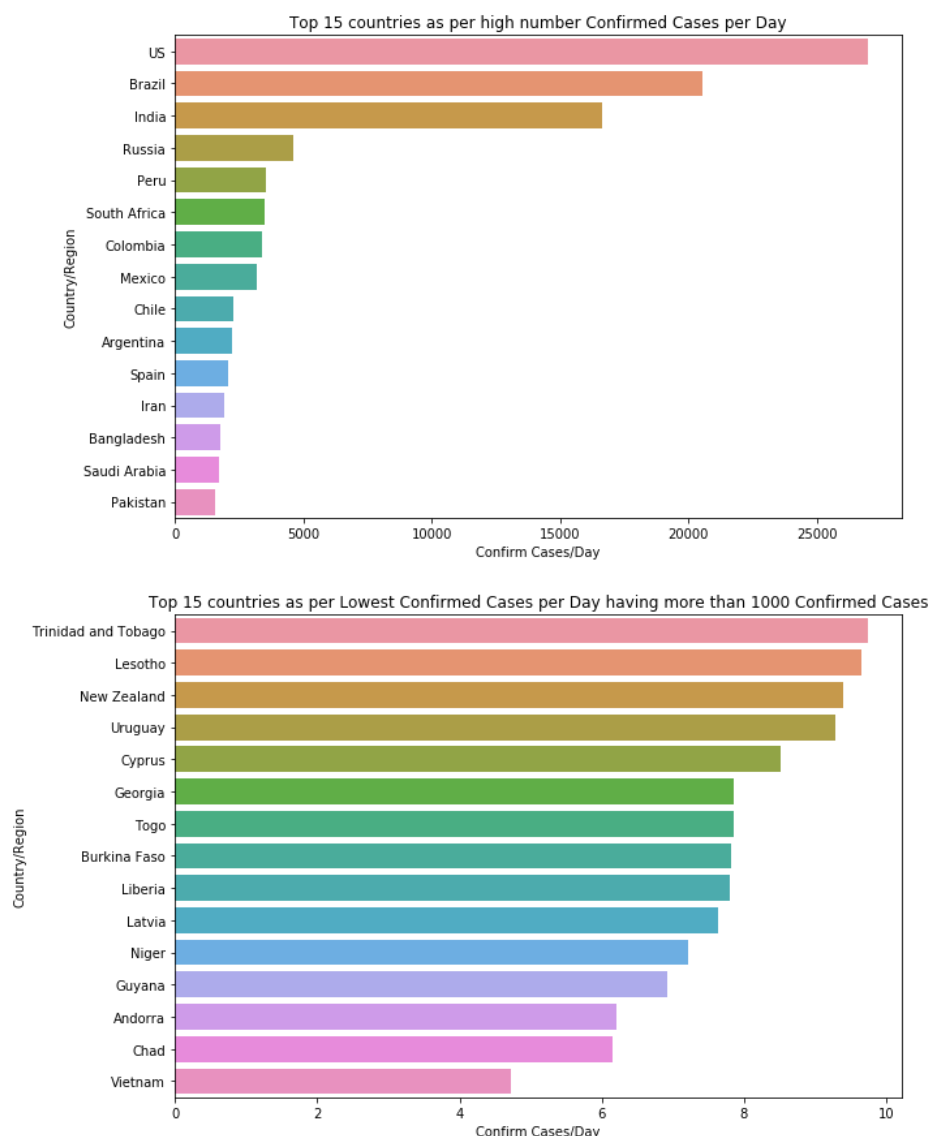
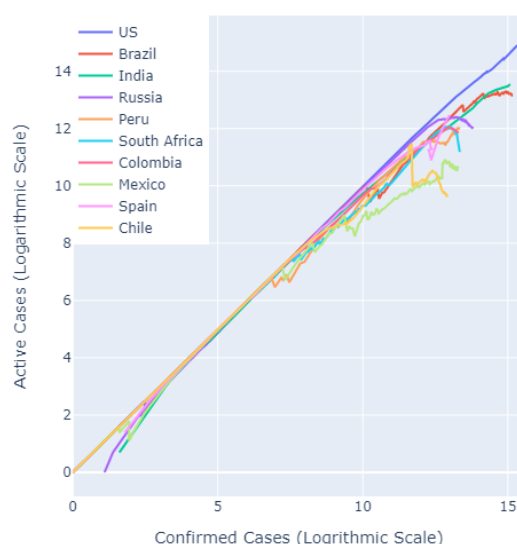


Figure 4. Comparison of reported case-fatality ratios and recovery proportions among selected countries with more than 500 confirmed cases.

4.3 JOURNEY OF DIFFERENT COUNTRIES IN COVID-19

Reported COVID-19 case counts changed rapidly during the pandemic, making it difficult to interpret country-level trends from isolated daily reports. Time-series visualisations can help show how reported cases developed over a selected period. However, comparisons must account for differences in population size, testing availability, reporting practices, and the dates covered by each dataset. A graph can describe a trend, but it cannot by itself establish why the trend occurred.

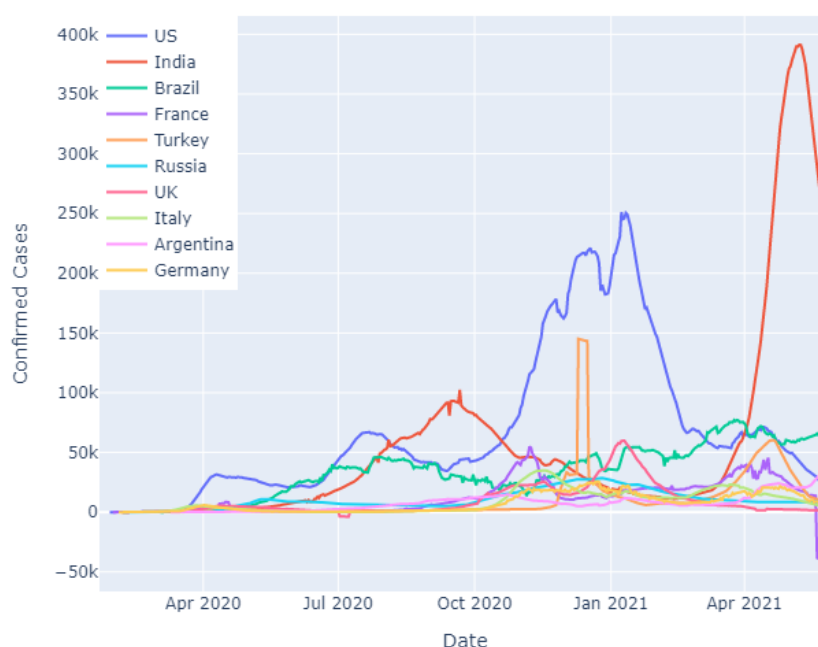
COVID-19 Journey of Top 15 Worst Affected Countries



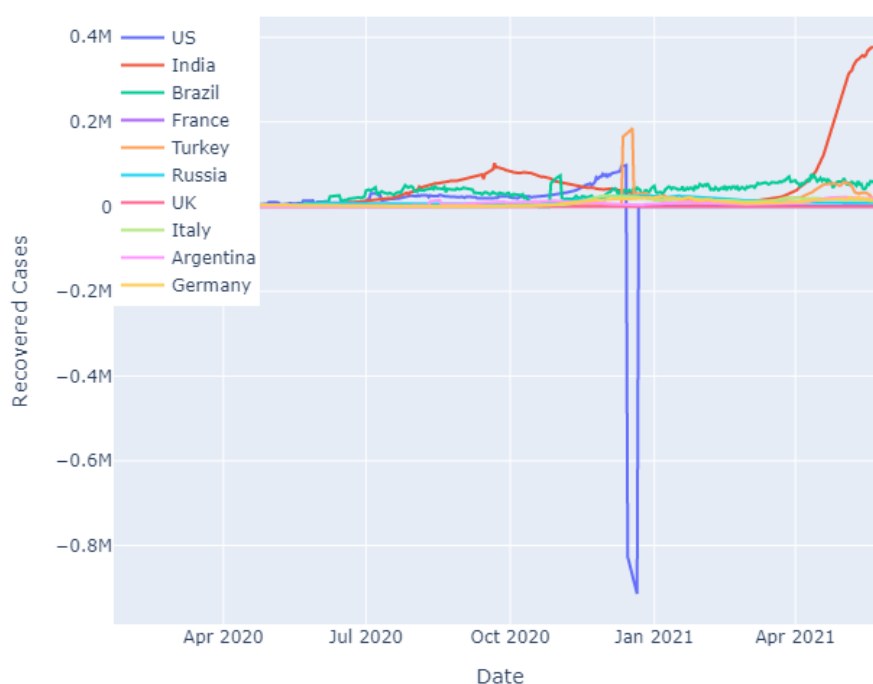
The plotted trends illustrate differences in reported COVID-19 case trajectories across the countries included in the analysis. Some curves may show sustained increases, while others may show periods of slower growth or decline. These differences should be interpreted cautiously: reported case counts alone cannot establish whether a country adhered to public-health guidance or whether a particular policy caused a change in transmission. Such conclusions would require additional information about interventions, testing, reporting practices, and other relevant factors.

Country-level comparisons should therefore be presented as descriptive observations rather than definitive evaluations of policy Effectiveness.

7 Days Rolling Average of Daily increase of Confirmed Cases

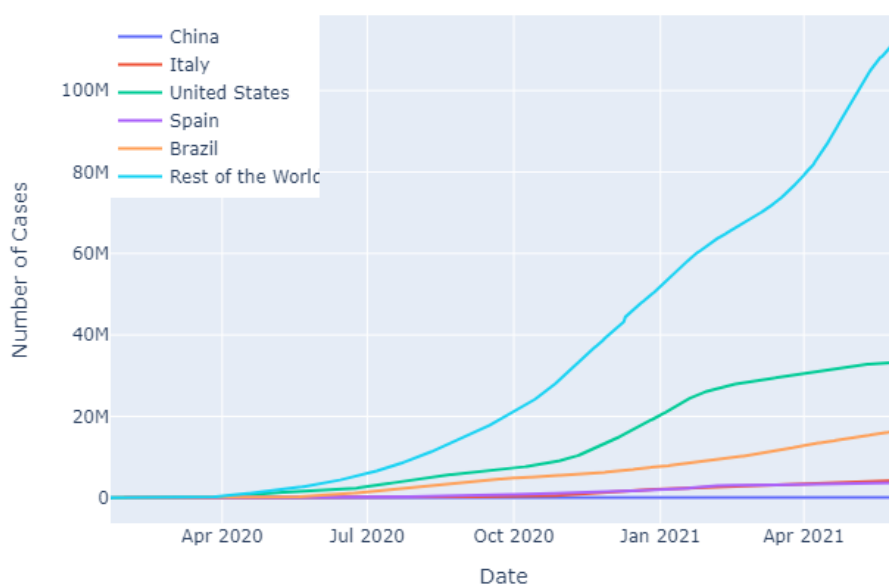


7 Days Rolling Average of Daily increase of Recovered Cases

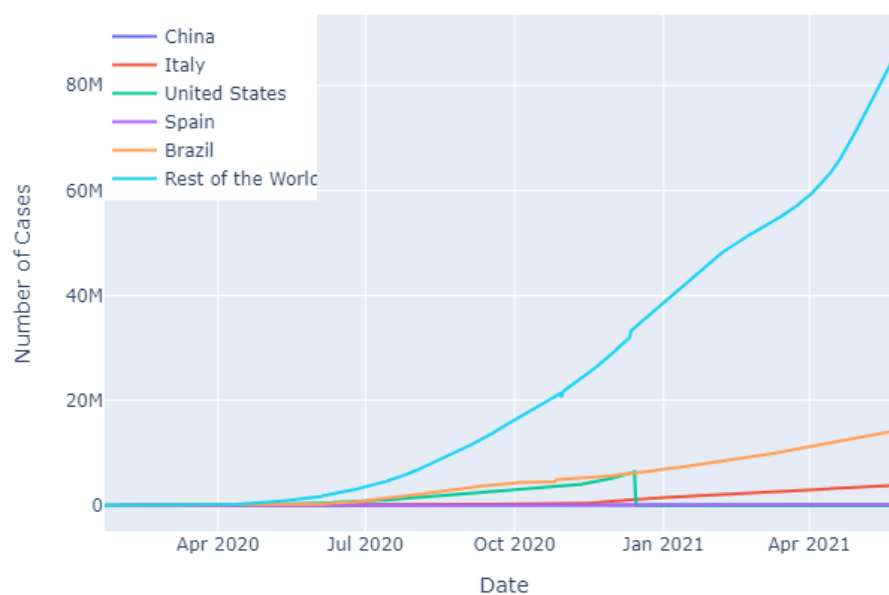


4.4 DIVERSE COUNTRIES' JOURNEY IN COVID-19

Confirmed Cases plot

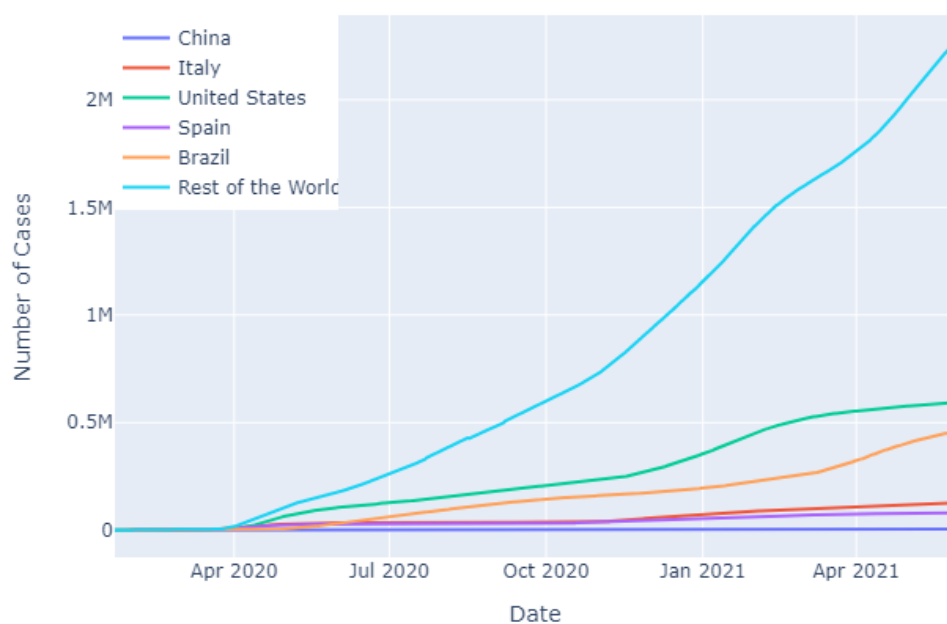


Recovered Cases plot

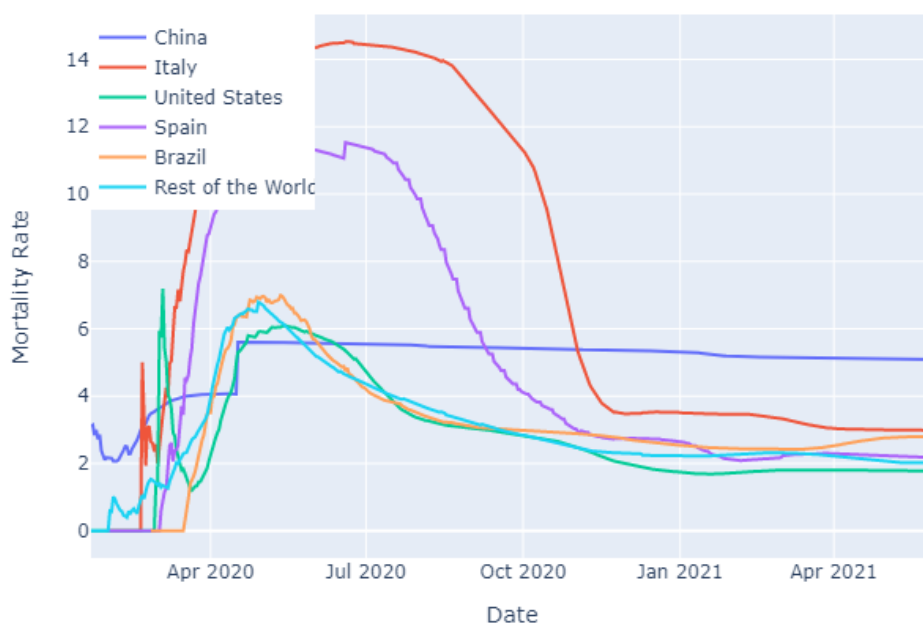


The plotted data may show changes in the reported numbers of confirmed cases, deaths, and recoveries across countries. A slower increase in reported cases may be described as a flattening of the observed curve, but the interpretation depends on the period and measure being plotted. A high recovery proportion or a low reported case-fatality ratio should not, on its own, be taken as proof of an effective healthcare system. Meaningful comparisons require consistent definitions, comparable observation periods, and consideration of reporting limitations.

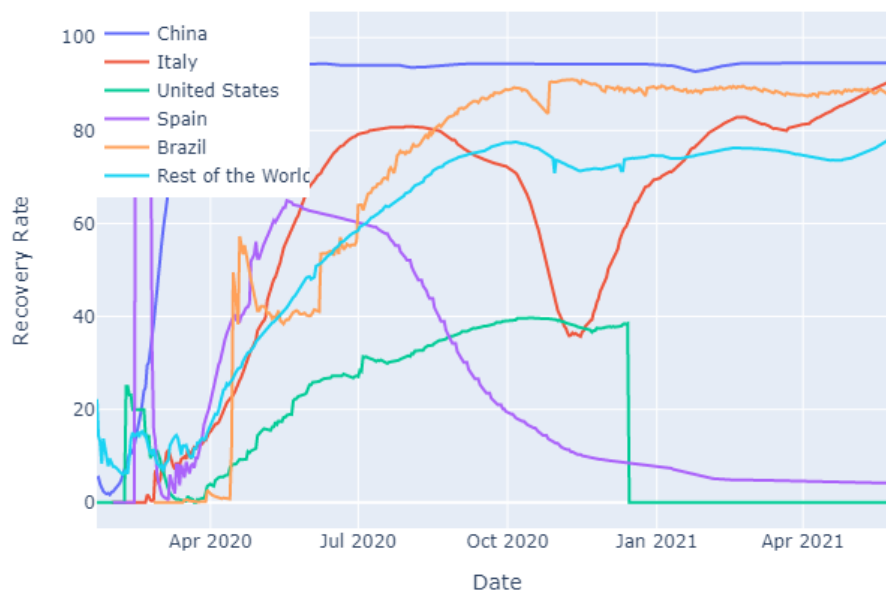
Death Cases plot



Mortality Rate comparison plot

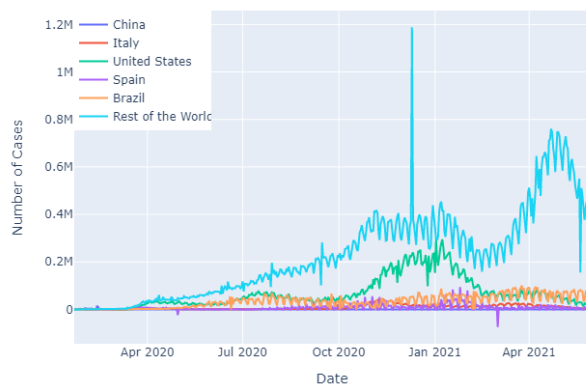


Recovery Rate comparison plot

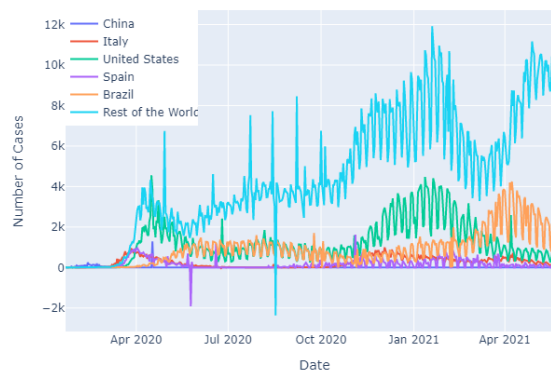


The figures compare reported recovery proportions and case-fatality ratios for the countries included in the analysis. Differences between these measures may reflect variations in reporting completeness, the timing of case outcomes, and the observation period. The figures should therefore be interpreted as descriptive comparisons of reported data rather than as definitive rankings of national healthcare performance.

Daily increase in Number of Confirmed Cases



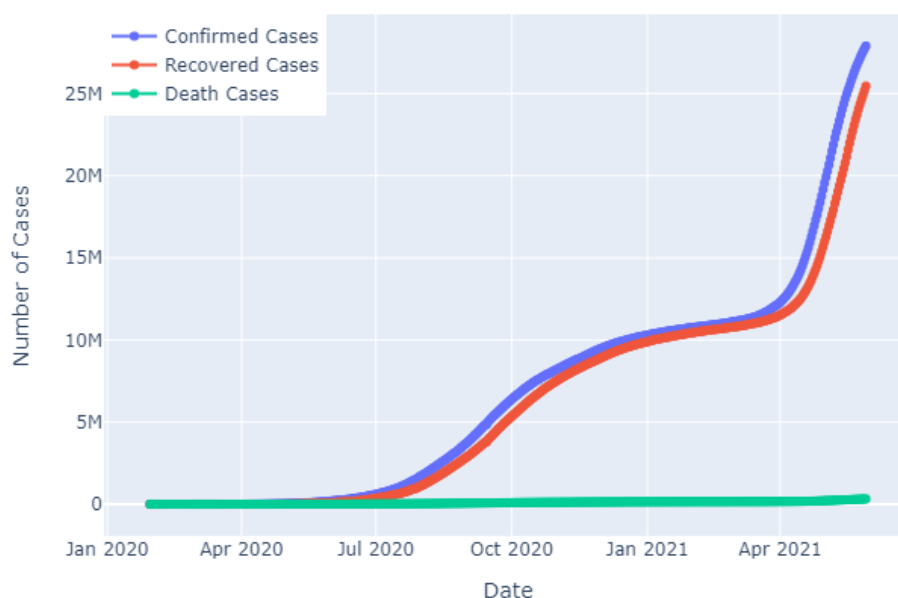
Daily increase in Number of Death Cases



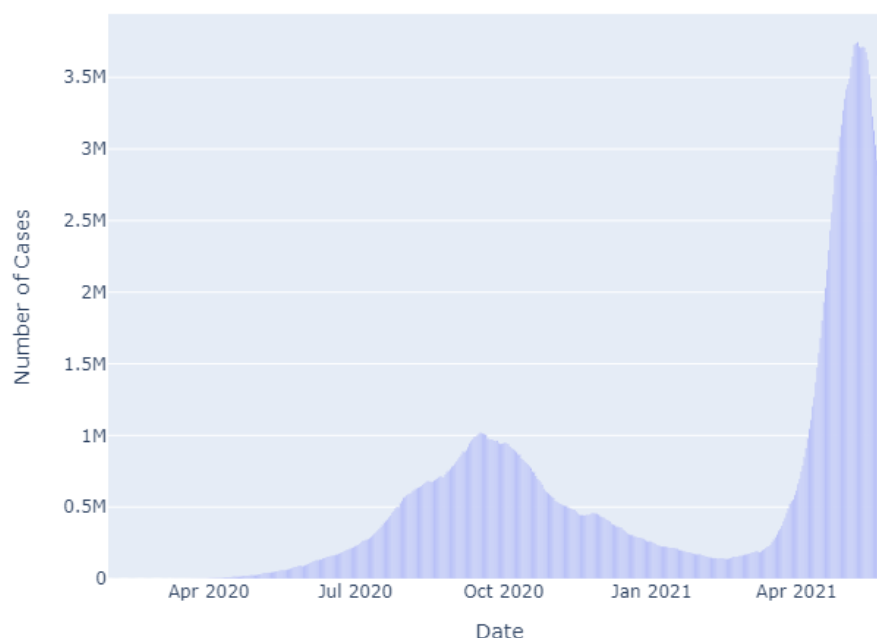
4.5 DATA ANALYSIS FOR INDIA

This section examines the reported COVID-19 data for India and compares selected trends with those of other countries. The analysis aims to describe changes in confirmed cases and related reported outcomes over the observation period. Comparisons should use consistent dates and clearly defined measures. If machine-learning or forecasting models are developed, their methods, training period, evaluation procedure, and results should be reported separately from descriptive visualisations. Historical government announcements may provide context, but they should be included only when they are directly relevant to the analysis and supported by appropriate references.

Growth of different types of cases in India

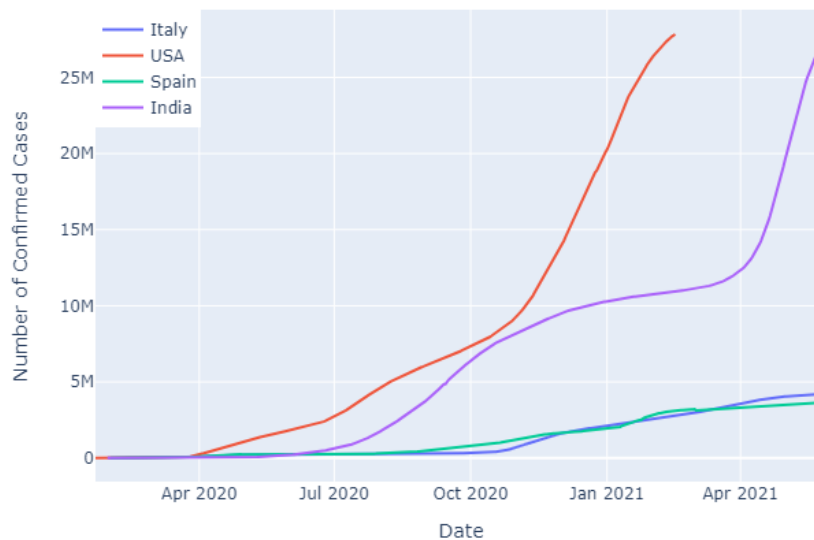


Distribution of Number of Active Cases in India



Using the selected dataset and the specified case-count threshold, the analysis compared the time taken by India and selected countries to reach the same reported confirmed-case level. The calculated intervals were 485 days for Italy, 392 days for the United States, and 484 days for Spain. India reached the reported total of 27,894,800 confirmed cases on June 1, 2021, according to the dataset used. These comparisons describe the recorded data and may be affected by differences in reporting practices and observation dates.

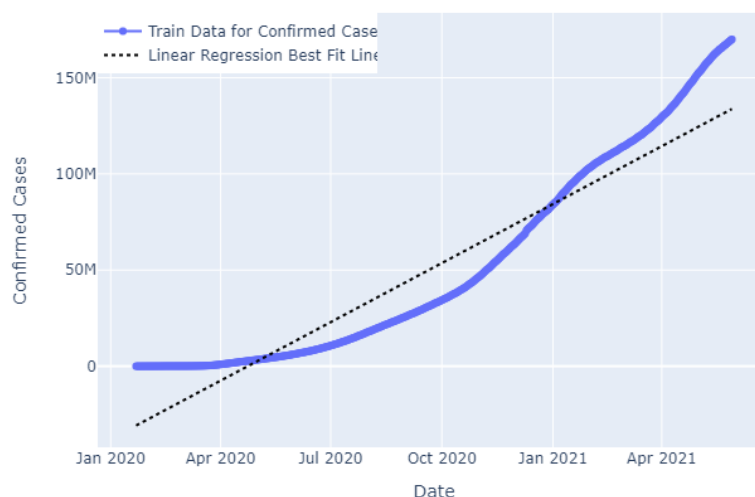
Growth of Recovered Cases with respect to India



4.6 PREDICTION USING A MODEL OF MACHINE LEARNING

Confirm Cases Prediction Using a Linear Regression Model.

Confirmed Cases Linear Regression Prediction



A linear regression model was considered for analysing reported confirmed COVID-19 cases. However, the observed time series may not follow a linear pattern throughout the selected observation period. The suitability of a linear model should therefore be assessed using appropriate evaluation measures and a comparison between predicted and observed values. Before drawing conclusions about model performance, the analysis should report the input variables, training and testing periods, evaluation metrics, and any preprocessing steps. If the model performs poorly, alternative approaches may be investigated, but their performance must also be evaluated using an appropriate validation procedure.

5 FINDINGS, CHALLENGES, AND FUTURE DIRECTIONS

COVID-19 research has used epidemiological data, statistical methods, and computational approaches to examine the spread of SARS-CoV-2 and its effects on public health. The usefulness of these approaches depends on the quality of the underlying data, the research question, and the methods used to analyse the results. This section summarises relevant findings from the literature, discusses challenges in using COVID-19 data, and identifies potential directions for further research.

5.1 FINDINGS

Published research has documented a range of symptoms associated with COVID-19, including fever, cough, and fatigue. Some patients develop more severe complications, while others experience mild or no noticeable symptoms. Symptom presentation and disease severity vary across individuals and populations. These differences complicate surveillance and reinforce the importance of appropriate testing, clinical assessment, and reliable reporting. The findings summarised here are drawn from published literature and should not be interpreted as results independently established by the dataset analysis in this paper [10], [11].

5.2 CHALLENGES

COVID-19 created substantial challenges for public health systems and society. These included pressure on healthcare capacity, shortages of medical supplies in some settings, disruptions to economic activity, and social consequences associated with public-health restrictions. Data-driven research faced additional challenges, including incomplete records, inconsistent reporting across locations, missing values, changes in case definitions, and delays between infection, diagnosis, and recorded outcomes. These limitations can affect the reliability of statistical comparisons and predictive models. Researchers should document data-processing decisions, assess the sensitivity of their findings to data limitations, and communicate uncertainty when presenting results [12], [13].

5.3 FUTURE DIRECTIONS

Future research can examine how epidemiological data, clinical information, and computational methods can be used responsibly to support public-health planning. Potential areas include improving data quality, developing reproducible analytical workflows, evaluating forecasting methods, and assessing how well models generalize across locations and time periods. Research should also address privacy, data governance, representativeness, and transparency. Predictive models should be evaluated against suitable baselines and validated on data not used for training. Claims about the effectiveness of treatments, vaccines, or public-health interventions should rely on appropriate clinical or epidemiological evidence rather than descriptive graphs alone [14], [17].

6 CONCLUSION

This paper examined potential applications of big data analytics and machine learning in the analysis of COVID-19, including diagnostic support, risk assessment, healthcare decision-making, and pharmaceutical research. It also discussed a dataset containing reported confirmed cases, deaths, and recoveries and considered how these measures can be examined through descriptive statistics and visualization. The analysis illustrates the importance of defining measures consistently, documenting data-processing steps, and interpreting trends in light of reporting limitations. Although machine-learning methods may offer opportunities for epidemiological analysis, reliable predictive conclusions require clearly described models, appropriate validation, and transparent reporting of results. Future work should focus on reproducibility, data quality, model evaluation, privacy, and responsible interpretation of epidemiological evidence. Overall, data analytics can support the study of pandemic trends, but its conclusions must remain proportionate to the data and methods used.

7 REFERENCES

- [1] S. J. Alsunaidi *et al.*, “Applications of Big Data Analytics to Control COVID-19 Pandemic,” *Sensors* 2021, Vol. 21, Page 2282, vol. 21, no. 7, p. 2282, Mar. 2021, doi: 10.3390/S21072282.
- [2] J. Wu, J. Wang, S. Nicholas, E. Maitland, and Q. Fan, “Application of Big Data Technology for COVID-19 Prevention and Control in China: Lessons and Recommendations,” *J Med Internet Res* 2020;22(10):e21980 <https://www.jmir.org/2020/10/e21980>, vol. 22, no. 10, p. e21980, Oct. 2020, doi: 10.2196/21980.
- [3] I. E. Agbehadji, B. O. Awuzie, A. B. Ngowi, and R. C. Millham, “Review of Big Data Analytics, Artificial Intelligence and Nature-Inspired Computing Models towards Accurate Detection of COVID-19 Pandemic Cases and Contact Tracing,” *International Journal of Environmental Research and Public Health* 2020, Vol. 17, Page 5330, vol. 17, no. 15, p. 5330, Jul. 2020, doi: 10.3390/IJERPH17155330.
- [4] Q. Jia, Y. Guo, G. Wang, and S. J. Barnes, “Big Data Analytics in the Fight against Major Public Health Incidents (Including COVID-19): A Conceptual Framework,” *International Journal of Environmental Research and Public Health* 2020, Vol. 17, Page 6161, vol. 17, no. 17, p. 6161, Aug. 2020, doi: 10.3390/IJERPH17176161.
- [5] R. Stojanovic, A. Skraba, and B. Lutovac, “A Headset Like Wearable Device to Track COVID-19 Symptoms,” *2020 9th Mediterranean Conference on Embedded Computing, MECO 2020*, Jun. 2020, doi: 10.1109/MECO49872.2020.9134211.
- [6] S. Richardson *et al.*, “Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area,” *JAMA*, vol. 323, no. 20, pp. 2052–2059, May 2020, doi: 10.1001/JAMA.2020.6775.
- [7] T. Mishra *et al.*, “Pre-symptomatic detection of COVID-19 from smartwatch data,” *Nature Biomedical Engineering* 2020 4:12, vol. 4, no. 12, pp. 1208–1220, Nov. 2020, doi: 10.1038/s41551-020-00640-6.
- [8] F. Y. Lan *et al.*, “COVID-19 symptoms predictive of healthcare workers’ SARS-CoV-2 PCR results,” *PLoS One*, vol. 15, no. 6, p. e0235460, Jun. 2020, doi: 10.1371/JOURNAL.PONE.0235460.
- [9] G. Kricke *et al.*, “Rapid Implementation of an Outpatient Covid-19 Monitoring Program,” 2020, doi: 10.1056/CAT.20.0214.
- [10] “A virus that respects no borders: protecting refugees and migrants during COVID-19.” <https://www.who.int/news-room/feature-stories/detail/a-virus-that-respects-no-borders-protecting-refugees-and-migrants-during-covid-19> (accessed Mar. 04, 2023).
- [11] “How Coronavirus Spreads | CDC.” <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html> (accessed Mar. 04, 2023).
- [12] B. Chen *et al.*, “Visual Data Analysis and Simulation Prediction for COVID-19,” *International Journal of Educational Excellence*, vol. 6, no. 1, pp. 95–114, Feb. 2020, doi: 10.48550/arxiv.2002.07096.
- [13] N. Krawczyk, B. D. Rivera, E. Levin, and B. C. E. Dooling, “Synthesising evidence of the effects of COVID-19 regulatory changes on methadone treatment for opioid use disorder: implications for policy,” *Lancet Public Health*, vol. 8, no. 3, pp. e238–e246, Mar. 2023, doi: 10.1016/S2468-2667(23)00023-3.
- [14] Z.-R. Yang *et al.*, “Efficacy of SARS-CoV-2 vaccines and the dose–response relationship with three major antibodies: a systematic review and meta-analysis of randomised controlled trials,” *Lancet Microbe*, Feb. 2023, doi: 10.1016/S2666-5247(22)00390-1.
- [15] “COVID-19 Treatment Guidelines.” <https://www.covid19treatmentguidelines.nih.gov/> (accessed Mar. 04, 2023).

- [16] "COVID-19 Map - Johns Hopkins Coronavirus Resource Center." <https://coronavirus.jhu.edu/map.html> (accessed Mar. 04, 2023).
- [17] A. Amstutz *et al.*, "Effects of remdesivir in patients hospitalized with COVID-19: a systematic review and individual patient data meta-analysis of randomized controlled trials," *Lancet Respir Med*, Feb. 2023, doi: 10.1016/S2213-2600(22)00528-8.
- [18] S. Rajkumar, "Novel Corona Virus 2019 Dataset," Kaggle. [Online]. Available: <https://www.kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset>. [Accessed: Oct. 13, 2026].