Covid-19 Containment Measure Analysis of Global & Indian Data with Predictive Model Comparison

Debojyoti Saha  
Electronics and Communication Engineering  
Institute of Engineering and Management  
debjojotsaha5@gmail.com

Debanjan Ghosh  
Electronics and Communication Engineering  
Institute of Engineering and Management  
dghosh533@gmail.com

Pranita Ganguly  
Electrical and Electronics Engineering  
Institute of Engineering and Management  
pranitaganguly17@gmail.com

*Anish Banerjee  
Information Technology  
Institute of Engineering and Management  
*anish.banerjee1998@gmail.com

Malay Gangopadhyaya  
Electronics and Communication Engineering  
Institute of Engineering and Management  
malay.ganguly@iemcal.com

Abstract

In the constant fight with the uncertainty of life and a broken economic and mental stature, machine learning and analysis is striving to identify probable remedies and even trying to predict the future of this hostile situation. Visualisation of treatment procedures, health data and economic data yields an in-depth analysis of the scenario. Remedial measures can be practised based on the predictive outcome. We performed predictive and statistical models like the SEIR for building highly accurate and analytical data outputs and plots for better visualisation of the data. This work aims to analyse the spread of the virus across the world and different regions in India and predict the near future of this pandemic in social, health and economic sectors.

Introduction

Coronavirus sickness 2019 (COVID-19) is an infectious illness accompanied by serious respiratory issue coronavirus 2 (SARS-CoV-2). This virus has a spot with a gathering of diseases that is liable for affliction reaching out from essential infection to deadly illnesses as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS) which were first found in China [2002] and Saudi Arabia [2012].

This disease has a high demise pace of about 8.0% for elderly people(aged 70-79), for those who are more than 80, the passing rate is even higher(14.8%). Side effects for this ailment can take 2-14 days to show up and can extend from fever, hack, windedness to pneumonia, kidney disappointment, and even untimely demise[1]. The transmission is individual to individual by means of respiratory beads among close contact with the normal number of individuals tainted by a patient being 1.5 - 3.5, however, the infection isn't airborne [2].

An abundance of live and verified data of different sectors including health and economic sections from government and non-government units have made it possible to visualise the trends in the data and make a predictive model based on the information generated upon the thorough analysis of it. Extensive study reveals about the interrelationship between data and machine learning models help in constructing a predictive model based on it. The outcome can be useful in taking proper measures for global benefits. AI and mobile computing are the key components for the
accomplishment of innovation in medicinal services frameworks [3]. In the realm of smart gadgets, information is being created in a remarkable path than at any other time and advanced the job of AI in medicinal services [3].

**Background Study**

Pneumonia of unknown cause first reported on 31st December 2019 in Wuhan, China, the virus turned into a viral outbreak and was issued as a Public Health Emergency of International Concern by the World Health Organization (WHO) on 30th January 2020 [4] and was named Covid-19 by the WHO on 11th February 2020. But, on 11th March 2020, with the exponential growth of inhuman mortality and infected cases with the statistical data of 1,18,000 cases in 114 countries and 4,291 confirmed death cases, with China, Korea, Italy, Iran contributing to around 90% of the total cases, Covid-19 was declared a global pandemic by the WHO. The World Health Organisation reported that as of 25th June 2020, the total number of confirmed cases are 292,77,214 and total confirmed deaths reported adds up to 4,78,691 in a total of 216 countries, with USA, Brazil, Russia and India leading the charts.

Based on the confirmed cases, fatality rate is lesser than other respiratory diseases: study of 72,000 COVID-19 patients finds 2.3% death rate [5].

India, with one of the lowest number of cases in the early stages, has now shown exponential growth in the spread of this epidemic with a current statistic of 4,74,585 confirmed and recorded cases, 14,915 recorded deaths. Being an extremely contagious virus, Covid-19 has shown this massive outbreak in India due to its extremely dense population. India tried to exercise potential control over the outbreak by executing one of the largest lock-downs of 1.3 billion people over almost 95 days and was quite successful in its early stages. Following 21st March 2020, the rate of cases fluctuated from, 41% to 14%, but, on the next day again to 30%.

Due to an abundance of live time-stamped data available on the internet, it was possible to analyse the trends in each sector, monitor minute changes and even draw analogies in similar behaviour through data visualisation. Data Science and Machine Learning can develop predictive models on extracting and learning hidden patterns and interrelationships of data and helps in predicting the near future trend based on the study of the present and past trends and decisions can be taken based on this outcome generated measures. An in-depth analysis of different areas of business and prediction of possible sectors of entrepreneurship is possible, including lucrative areas for investors plus new possibilities of expansion for companies to generate handsome revenue.

![Rate of Active Cases vs Death from 15th March to 31st March](image)

**Fig 1**: The rate of Active cases and Death cases

**Data Overview**

The time-series data used to analyse the global outbreak and spread of the Novel Coronavirus throughout the world was taken from John Hopkins University. The dataset used to analyse the Indian scenario exclusively is sourced from
covid19india.org and the state selective data was collected from the Ministry of Health and Family Welfare. The data we worked on was from 30th January 2020 till present date (25th June 2020).

The Data consists of a daily time series summary table focussing on confirmed cases, deaths and recovery cases and active cases. The data is retrieved from daily reports and properly reported cases.

The other predictions or reports generated on the economic factors are based on the data collected on stock prices and gold rate, from the National Stock Exchange, others from news, articles and magazines about startups and other ventures and their official revenue reports.

**Pandemic Data Analysis**

**WORLD DATA ANALYSIS**

![Fig 2: World pandemic overview heat plot](image)

Fig, above, which is the dynamic plot of the live global data represents the trend of confirmed cases, deaths and recovery cases across the globe. This representation depicts the rate of confirmed cases increases exponentially from 16.03.2020. The rate of increase in the number of deaths across the globe is very low but the trend seems increasing which is alarming. Although, the recovery rate seems to be increasing as well which is a positive sign. Thus we can infer that the virus is not too deadly but very contagious.
The following figure represents the Global data distribution during the pandemic, based on the attributes: total confirmed cases, total deaths, total active cases, incident rates and mortality rates across the globe. From this plot we get to see the detailed distribution and present standing of each country is based on the total number of Confirmed cases.

| Country | Confirmed | Deaths | Recovered | Active | Incident Rate | Mortality Rate |
|---------|-----------|--------|-----------|--------|---------------|---------------|
| US      | 20,816,096,000 | 0 | 15,107,000,000 | 5,619,096,000 | 12,409,096,000 | 600,924 |
| Brazil  | 19,822,000,000 | 0 | 15,930,000,000 | 3,922,000,000 | 12,490,000,000 | 600,924 |
| Russia  | 17,140,000,000 | 0 | 15,930,000,000 | 1,743,000,000 | 12,490,000,000 | 600,924 |
| India   | 17,105,000,000 | 0 | 15,930,000,000 | 1,743,000,000 | 12,490,000,000 | 600,924 |
| UK      | 17,140,000,000 | 0 | 15,930,000,000 | 1,743,000,000 | 12,490,000,000 | 600,924 |
| Peru    | 5,945,000,000  | 0 | 4,066,000,000  | 1,863,000,000 | 12,490,000,000 | 600,924 |
| Chile   | 15,305,000,000 | 0 | 12,239,000,000 | 3,066,000,000 | 12,490,000,000 | 600,924 |
| Spain   | 2,748,000,000  | 0 | 2,078,000,000  | 670,000,000   | 12,490,000,000 | 600,924 |
| Italy   | 2,397,000,000  | 0 | 1,823,000,000  | 574,000,000   | 12,490,000,000 | 600,924 |
| Iran    | 2,310,000,000  | 0 | 1,753,000,000  | 557,000,000   | 12,490,000,000 | 600,924 |

The data shows a uniform positive outbreak for all the categories, for the USA, which justifies its position in the list. Mortality rate shows an abnormal outbreak in countries like the UK, Spain and Italy. Russia has shown the least mortality rate. In the case of Brazil, the total number of confirmed cases is much less than the USA, but the recovery rate is almost greater than that of the USA, thus Brazil has got better control over the situation.

The USA conducted 30,110,061 tests over a population of 331,002,651 which means a test has been conducted for every 11 people. India, on the contrary, has conducted a total of 2,07,871 tests over a population of 1,387,297,452 which implies a test for every 6,674 people.

The above plot represents the availability of hospital beds per one thousand population. The low availability of healthcare facilities by countries like Iran, Peru, Chile and India but their high number of active cases and a positive exponential growth rate is something to take under consideration seriously in order to exercise control over the pandemic.
India has shown an abnormal outburst of population density and an extremely low count of hospital beds and virus tests being conducted. This is highly contrasted by the statistics of Russia depicting one of the least population densities amongst the countries and with extremely well-planned healthcare facilities and rigorous testing.

Coronavirus, an extremely contagious disease has thus shown excessive outbreak in India in spite of the strict lockdown measures. The figures represent the number of tests conducted per one thousand people which is alarmingly high for Russia, which, along with better healthcare facilities and a low population density counts for its low mortality rate.

The plot shows the rate of deaths due to this virus outbreak for different countries over a specified time period (June 2020) where Russia and Chile have the least rate of death. In spite of holding the second position in the global list with respect to maximum cases Russia, with its health facilities and immediate action has been able to exercise control over the death rate.

Italy and the United Kingdom, on the other hand, have shown an alarmingly high death rate.
Italy, with an extremely high death rate, has still registered a comparatively lesser number of deaths than what is registered in the United States with a lower death rate. One of the reasons can be that affected cases in the USA are much higher than Italy, even the highest in the world.

The global recovery rate (Recovered cases/Total cases) shows an extreme fall marked by the United Kingdom which accounts for its second-highest death rate. China and Germany, on the other hand, have shown an extremely high recovery rate and were amongst the first countries to be declared safe.
The pioneer countries have shown a very positive recovery rate especially Taiwan and New Zealand at a very early stage.

The testing distribution shows that most of these countries started with rigorous testing at an early stage. Also, these countries are known to have proper medical health facilities. So a combined effort of these combined with proper citizen responsibility and rapid action has made these countries shine above the pandemic at an early stage with lesser casualty count.

**INDIAN DATA ANALYSIS**

Indian scenario represents that the confirmed cases and recovery cases show exponential growth. The rate of both these parameters seems to be almost equal. Death cases are very low reflecting the general global trend.
The Covid-19 cases across the nation on 24th June is shown in Fig 13.

India at the beginning of this pandemic went into a severe lockdown from 21st March 2020 which has continued into several phases during which this Fig(above) shows the outbursts of newer cases rising with time. Sudden spikes of new cases started during the month of April but were still in the considerable range.

From the month of June, the outbreak of newer cases was beyond control. We can see over here that the increase of newer cases rose beyond 16,000 within no time.
The figure above (Fig) represents the state-wise scenario of the country. Maharashtra, Delhi and Tamil Nadu are at the top of this list with around 1,40,000 cases, 66,000 cases and 64,000 cases respectively. Gujarat takes the next place. Andaman and Nicobar Islands, Arunachal Pradesh, Chandigarh, Ladakh, Mizoram, Nagaland, Meghalaya, Himachal Pradesh, Chandigarh, Goa, Dadra Nagar Haveli and Daman & Diu are the states with the least registered cases.
The plot of Cured versus Death of Indian states depicts a uniform trend. The number of cured cases in every state is much higher than the number of death cases. This counters for the low mortality rate and the high recovery rate in India.
The plot of total active cases versus cured distributed country-wise shows an almost uniform pattern of the number of cured cases being more than the number of active cases for most of the states except Ladakh, Arunachal Pradesh, Goa, Telangana, Manipur where the scenario is just opposite.

Algorithm Analysis

In comparison to latent learning (ordinary AI classifiers), dynamic learning is used as a learning issue, where the observer has some activity in settling on what data it will be readied. Right when it is an emergency (COVID-19), it requires extraordinary consideration so data analysis and its corresponding decision can be made reliably without hanging tight for a long data collection. Likewise, utilizing consistent data (on-the-fly) is the must since one can hardly wait that long to prepare the machine and addition from them nor manual remark/assessment is possible. This suggests instead of having a customary course of action of train, endorsement, and test set, we need AI-driven gadgets that can learn after some time without having complete information about the information, which we call Active Learning (AL) [6].

Likewise, access to exact outbreak prediction models is basic to get bits of knowledge into the presumably spread and outcomes of irresistible diseases. Governments and other authoritative bodies depend on experiences from forecast models to propose new arrangements and to survey the viability of the upheld approaches. [7]
**SIR Model:** In comparison to numerical modeling based on dynamic equations are paid less attention to, though they can provide more elaborate revelations into the characteristics of epidemic. Among them, the classic susceptible exposed infectious recovered model also known as the (SIR) model is the most widely adopted one for characterizing the epidemic of COVID-19 outbreak in both China and other countries [8].

SIR statistical model is used to predict the pandemic situation future trend and was thus the most preferred algorithm for this current pandemic COVID-19.

![SIR Model](image)

**SIRD Model:** In compartmental models the populace under examination is separated into segments and, with the suspicion that each person in a similar division has similar attributes, the developments of individuals inside the compartments can be depicted. Specifically, the "SIRD" model represents the four classes "Susceptible", "Infected", "Recovered", and "Death" [9].

To examine the Covid19 disease in India we investigate a standard epidemiological model, the SIRD model. So as to actualize this model

- We utilize consistently combined contamination of different nations
- The quantity of contamination versus the number of tests completed.

The contamination rate, which is a period subordinate boundary, is set in the structure to assess the best fit with the given information. The model is reenacted with an intend to feature the likely component of the contamination in the nation.
Polynomial Bayesian Ridge Regression model: Bayesian regression helps a natural mechanism to survive insufficient data or poorly distributed data by formulating linear regression using probability distributors rather than point estimates. The output or response ‘y’ is assumed to be drawn from a probability distribution rather than estimated as a single value.

Mathematically, to obtain a fully probabilistic model the response y is assumed to be Gaussian distributed around Xw as follows

\[ p(y|X,w,\alpha) = N(y|Xw,\alpha) \]

One of the most useful types of Bayesian regression is Bayesian Ridge regression which estimates a probabilistic model of the regression problem. Here the prior for the coefficient w is given by spherical Gaussian as follows

\[ p(w|\lambda) = N(w|0,\lambda^{-1}I_p) \]
Facebook Prophet Model: The fb prophet algorithm is an open-source library distributed by Facebook that depends on decomposable (trend+seasonality+holidays) models. It gives us the capacity to make time-series expectations with great exactness utilizing basic natural boundaries. We have applied to foresee the situation on both Indian and word information. The relative plots are shown beneath. It is additionally intended to have natural boundaries that can be balanced without knowing the subtleties of the basic model. This is vital for the investigator to successfully tune the model [10].
The monthly and weekly trends are predicted and shown below in figures 28 and 29. The weekly trend shows that the number of cases reduced on weekends may be due to holidays and shutdowns and people tend to go out less. But it rises exponentially from Thursday to Friday.

**Linear Regression Model:** Linear regression model was trained over this accumulated and processed data and the following results were collected. The plot in fig 30 shows that the model nearly predicts the actual scenario very well and also is able to get a hint of the pattern of spread. The error of prediction is also considerably low.

**Conclusion**

Needless to say, the pandemic has been a tremendous blow in the face of the global economy and its kin. Despite the fact that the spread of the infection was very delayed, but because of the absence of adequate health care units, testing kits and PPEs and earlier prudent steps the spread has taken such a large number. But on the brighter side, the casualty rate in this nation stands quite low, consequently asserting much fewer lives when contrasted with the U.K, Spain and Italy. Another concerning reason being the tests conducted in this nation is truly low which gives out a vulnerability in regards to who has the infection, and which region ought to be hindered for acceptable. Now, the main thing we as residents of India can embrace is the carefully holding fast to the rules as said by the clinical experts, to stop going out pointlessly, to keep up social distancing, and get tested whenever exhorted by the specialist, maintain utmost hygiene in and around us, only then we can hope of somewhat flattening the curve. As of recent studies, the outbreak of the Covid-19 has manifested in over 210 nations, 22.3 million people across the globe has been affected and the fatality has totalled over 7,84 thousand. The US, Brazil along with India has been the most affected nations. SIR, SIRD models reflect a statistical analysis of the data whereas Facebook prophet, Bayesian Ridge regression and Linear Regression show the machine learning approach. The data visualisation and predictive modelling has been performed.
in such a way that one can understand the outbreak of the epidemic as well as the severity of the disease. Frameworks like SIR and SIRD, which are the statistical models portray the trend of the infection outbreak with respect to the population of that particular nation. On the other hand, predictive modelling like facebook prophet model gives us a detailed understanding of the spread of the disease on weekly basis, wherein we find that the spread of the disease attains a peak during the weekdays, while the spread remains considerably low during the weekends. Other machine learning models give us the idea of the pattern of the pandemic outbreak in India and around the world.

From the above analytical studies we can conclude that the pandemic is taking a very dangerous turn by claiming lives and destroying the economy. With this paper we offer to provide a better visualisation of the pandemic situation to the people and the authorities, so that they may adopt better measures to curb the outbreak, therefore might have a chance of adapting more accurate measures of control. In future, one can compare with the ARIMA and SARIMA time series modelling.

Reference

1. World Health Organization, 2020. Coronavirus disease 2019 (COVID-19): situation report, 67.2.
2. Wu, Z. and McGoogan, J.M., 2020. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72,314 cases from the Chinese Center for Disease Control and Prevention. Jama.
3. T. Panch, H. Mattie, L.A. CeliThe inconvenient truth about ai in healthcare Npj Digital Med, 2 (1) (2019), pp. 1-3
4. Medscape Medical News, The WHO declares public health emergency for novel coronavirus (January 30, 2020). https://www.medscape.com/viewarticle/924596.
5. Wu, Z., and McGoogan, J.M., Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72,314 cases from the Chinese center for disease control and prevention. JAMA, 2020. https://doi.org/10.1001/jama.2020.2648.
6. Bouguelia, M., Nowaczyk, S., Santosh, K. C., and Verikas, A., Agreeing to disagree: Active learning with noisy labels without crowdsourcing. Int. J. Machine Learning & Cybernetics 9(8):1307–1319, 2018. https://doi.org/10.1007/s13042-017-0645-0 Return to ref 18 in article
7. Remuzzi, A.; Remuzzi, G. COVID-19 and Italy: what next? Lancet 2020.
8. Labadin, J. & Hong, B. H. Transmission dynamics of 2019-nCoV in Malaysia. medRxiv https://www.medrxiv.org/content/early/2020/02/11/2020.02.07.20021188.full.pdf (2020).
9. Caccavo, Diego. (2020). Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model. 10.13140/RG.2.2.24485.86243.
10. Taylor SJ, Letham B. 2017. Forecasting at scale. PeerJ Preprints 5:e3190v2 https://doi.org/10.7287/peerj.preprints.3190v2