Research Article

Graphical Trends and Machine Learning Forecast of COVID19 Around the Globe

Keshav Kumar[1]*, PG Naveen Kumar[2], Dharmendra Kumar[3], Mausam Kumar[4], Vaishali[5] and Kunal Krishn[6]

[1]*Faculty of Dental Sciences, Institute of Medical Sciences, Banaras Hindu University, Varanasi, UP, India*

[2]*Professor, Department of Public Health Dentistry, Faculty of Dental Sciences, Institute of Medical Sciences, Banaras Hindu University, Varanasi, UP, India*

[3]*Indian Statistical Institute Bangalore, India*

[4]*Department of Metallurgical Engineering, Indian Institute of Technology, Banaras Hindu University, Varanasi, UP, India*

[5]*Faculty of Dental Sciences, Institute of Medical Sciences, Banaras Hindu University, Varanasi, UP, India*

[6]*School of Language, Literature and Culture Studies, Jawaharlal Nehru University, New Delhi, India*

*Corresponding Author: Keshav Kumar, Faculty of Dental Sciences, Institute of Medical Sciences, Banaras Hindu University, Varanasi, UP, India.

Abstract

Background: An epidemic disease is contagious and have the capability of spreading into entire nation if timely adequate measures are not taken to curb the epidemic. Coronavirus outbreak from Wuhan in South China is engulfing approximately entire world within nearly four month of time span. Till date nearly 10 lakh people are infected and around 50 thousand died due to this disease around the world. In this article we are going to represent the graphical trends of infected cases around the WHO six zone, calculation of Basic Reproduction Number, Using Machine Learning tried to find out the near future scenario.

Method: we collected number of infectious cases data from various official source. We complied the data according to the need of our study and performed data validation from IBM SPSS software. Further, we plotted various graph using Microsoft Excel tools. Studied the nature of graphical trend and discussed. Using RStudio we calculated the Basic reproduction Number of WHO zones and countries. Lastly, we used Machine learning in forecasting the near future trend of infected number of cases in near future.

Findings: Our finding in the graphical trend is we found that except China all other studied areas have Exponential phase of increase in number of new cases. Whereas, china is in plateau phase. Basic Reproduction Number is highest for USA and Regions of Americas 2.11, 2.08 respectively. Thses results are adding an extra information add on for the policy makers and healthcare providers to act accordingly and also adding new dimension to the existing Knowledge data base.

Interpretaion: our study shows that almost all WHO zone is facing an exponential increase in infected cases. As well as its trend is increasing, thus it the time to act more aggresivley and work collabroatevely to curb the disease and stop the spread.

Keywords: Trend; Pandemic; Machine Learning; Basic Reproduction Number; Forecast; Poisson Regression; COVID-19; SARS COV

Introduction

Epidemic diseases are contagious and have capability of spreading into entire nation if the contagion measurement touches the outbreak level and have enough strength to wipe out entire population [1]. Coronaviruses are spherical or pleomorphic shaped with surface having petal or club shaped peplomer projection [2]. These viruses are enveloped single stranded RNA genome with a positive sense. Its genome ranges from 26 to 32 kilobases, amongst RNA

viruses coronavirus carries the largest genome [3]. Coronaviruses are broadly classified into two groups: Acid labile (causing common cold like illness), Acid stable (Causing human and animal gastro-enteritis) [2]. Coronavirus typically cause mild respiratory disease in humans [4]. However, in November 2002 an outbreak of respiratory tract infection of unknown origin broke out in Guangdong province of South China [5]. This new disease was later named severe acute respiratory syndrome caused by Coronaviruses. this virus appears to be different from other coronavirus and categorized into Alpha beta and gamma [3], which cause lethal human disease [2]. Newly recognized coronavirus is assumed to have originated as a virus that has mutated or ecologically adapted to cross animal human species barrier [3,6,7]. In July 2003 SARS COV caused 8437 infected cases in 26 countries with 774 deaths [5]. Experiments and laboratory testing evidence says reservoir for this RNA virus may involve a range of animal species [5]. It is seen that masked palm civet cats are associated with animal-to-human transmission [8]. SARS coronavirus discovered in November 2002 in China and spread internationally to Toronto, Canada is now named SARS COV-1 [4,9]. June 2012 an epidemic broke out in Saudi Arabia causing respiratory infection to more than 2400 cases were later named as Middle East Respiratory Syndrome (MERS) [4]. Dromedary Camel was found responsible for transmission from animal to human [8]. December 2019 a cluster of patients with respiratory infection reported in Wuhan China which on sequencing clinical samples identified as Severe Acute Respiratory Syndrome COV-2 (COVID19 Disease) [10].

Based on WHO guidelines while referring to the virus or infection it is called SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) and for referring to disease or symptoms it is known by COVID-19 (coronavirus disease 2019) [11]. WHO declared a public health emergency of international concern on 30th January [12], later on WHO declared COVID19 as pandemic on 11th March 2020 [13].

Clinical symptoms of COVID19 include fever, cough, myalgia, fatigue, headache, gastrointestinal symptoms, moderate to severe dyspnea [7]. Median incubation period is 5.1 days (95% Confidence Interval 4.5, 5.8) days [14]. SARS-COV-2 remain viable in aerosol for up to 3 hours with reduction in titer, it remains more stable on plastic surfaces up to 72 hours as compared to cooper and cardboard. Therefore, transmission by aerosol and fomites cannot be ruled out now [15].

Need of this review

We are trying to bring out a comprehensive picture of COVID19 spread through this study, which will also help us in understanding the course of disease propagation as well as help us to take better preventive measures and precautions as required.

Delay in Publication

Final touch to this paper was completed on 4 April 2020, which was quite ahead of time as per forecasting is concerned and were duly submitted to few Eminent journals like Nature, Lancet, MMWR, IJMR, Unfortunately two of them rejected on ground of finding it not suitable for their journal space while IJMR expressed their inability in publication and Nature denied on the ground of degree of advance provided, timeliness and the breadth of potential interest to the wider research community. However, there is less published articles in journal using possion regression model in Machine learning for forecasting COVID19 cases. Publishing, it now will strengthen the methodology as we can verify the prediction data output with the original data set.

Data source

For the data collection we referred to the official websites of (1).WHO Situation report page was the prime source of data regarding epidemic spread (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/) Other sources of data associated to epidemic are (2)John Hopkins Bloomberg school of public health COVID19 Situation reports.(http://www.centerforhealthsecurity.org/resources/COVID-19/index.html/), (3) John Hopkins coronavirus resource center Dashboard(https://coronavirus.jhu.edu/map.html), (4) Centre for Disease Control and Prevention (https://www.cdc.gov/coronavirus/2019-ncov/index.html), (5) National Health commission of the Peoples Republic of China (http://en.nhc.gov.cn/). (6) For coding we have used the data from GitHub (https://github.com/midas-network/COVID-19/tree/master/parameter_estimates/2019_novel_coronavirus/).

Methodology

All the data collected from data source and plotted in Microsoft Excel Sheet. Data validation is done by IBM SPSS Statistics 20 software. Using Excel tool graph were plotted for selected data. Nature of graph studies from Slope of Curve and each graph is explained in Trends section.

Basic Reproduction Number (R0): Calculation is done using RStudio Version 1.2.5033 software. We used gamma distributed generation interval for an exponential growth. Where the mean generation interval time is taken from the previous study of COVID19 which is 3.3 (95% Confidence Interval as 2.3, 4.3) and variance as 3.1 (95% Confidence Interval as 1.0, 8.0).[21] Implementing same value, we calculated R0 for various zone and countries.

Machine Learning Forecast: Using Anaconda Navigator 1.9.12 platform we coded in Jupyter Notebook 6.0.3 software for performing machine learning. We divided whole data set into two parts- Training Data and Test Data. 80 percent of the data is taken as training data and 20 percent data is put for checking the machine efficiency. For training the machine Poisson Regression is used. firstly, trained the machine then executing forecast function, predicted the future number of infected cases. Coding for RStudio and JUPYTER Notebook is provided in the Supplementary Material(RStudio Code, Machine Learning Code).

Trends and Results

The graphical trend is analyzed from comparative graph plotted in Excel sheet. In comparative graph the nature of graph varies according to the other data in the graph. WHO has divided the world into six zones, namely West Pacific region, south-East Asian region, Eastern Mediterranean region, European Region, African Region and Region of Americas. Thus, all six zone is selected and from each zone three countries were randomly selected and analyzed.

West Pacific Region: In figure 1 graph of the total number of infected cases of West Pacific region along with China, Republic of Korea and Japan are plotted. COVID19 first outbreak reported in Wuhan, South China and from there it spread all around the world. Therefore, the curve of china and west pacific region go hand in hand overlapping each other approximately till 21 Feb. But as soon as significant infected cases transmitted across the globe the curve of total cases of WPR region started to follow an increasing trend. The curve of China can be studied in 3 phases.

Lag Phase- increase in reported cases were gradual

Exponential phase- Infected cases were reported exponentially. There is a sharp slope of curve between 15 and 16 February as the clinically diagnosed cases were included from along with laboratory confirmed cases from 16 Feb. looking at the curve we can make out that from 26 January to 18 February for a period of approximately 23 days there were exponential increase of infected cases thereafter trend begin to follow the declining phase.

Plateau Phase- reported cases following a decreasing trend:

Overall trend of china curve is giving a S shaped curve which is a logistic model. The WPR curve is overlapping china curve grossly till 21 February thereafter a gradual increasing trend appears which signifies the cases are increasing in other countries of WPR region and its territory (Figure 1).

Figure 1: Total number of infected cases in Western Pacific Region, China, Korea and Japan.

South-East-Asia region: This graph contains infected cases of South East Asian region along with 3 randomly selected countries Indonesia, Thailand and India. In this graph there is a long lag phase which signifies gradual increase of reported cases. it took time for cases to import from China and converting it into community spread, this indicates there was a window period for epidemic preparedness. Rise in reported cases started from 5 to 8 March, it includes Rise in cases along with rise in number of countries having COVID19 cases in South East Asia region (Figure 2).

Figure 2: Total number of Infected cases in South-East Asia Region, Indonesia, Thailand and India.

Region of the Americas: This graph shows total number of infected in Regions of Americas and countries like USA, Canada and Brazil. Lag phase followed by exponential having steep slope which signifies reporting of large number of new cases. The curve of USA having same trend as that the curve of Regions of Americas. This signifies that the number of cases in USA is much larger than other countries of this zone (Figure 3).
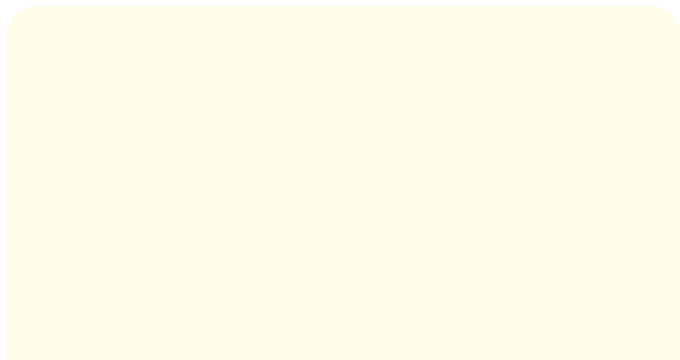
Figure 3: Total number of Infected cases in Region of Americas, USA, Canada and Brazil.

European region: The graph is plotted for European region and countries like Italy, Spain and France. curve of European region is increasing exponentially. In this region Italy has highest number of cases than other countries. However the curve of European region lie above the curve of Italy which signifies there is significant amount of cases in other countries too in this region (Figure 4).
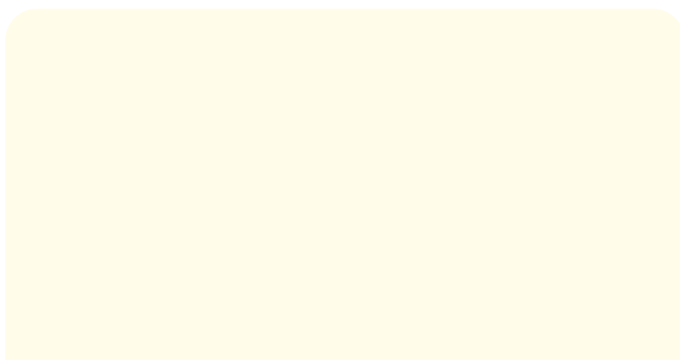

Figure 4: Total number of Infected cases in European Region, Italy, Spain and France.

Eastern Mediterranean region: This graph is plotted for Eastern Mediterranean region including countries like Qatar, Iran and UAE. Curve of Iran lie close to the curve of East Mediterranean region which is an indicator that the highest number of cases are there in Iran as compared to other countries of this region, later on the curve of East Mediterranean region seems to be increasing more than the curve of Iran buy this weekend we can make out that the number of cases in other countries of this region hav also increased significantly (Figure 5).
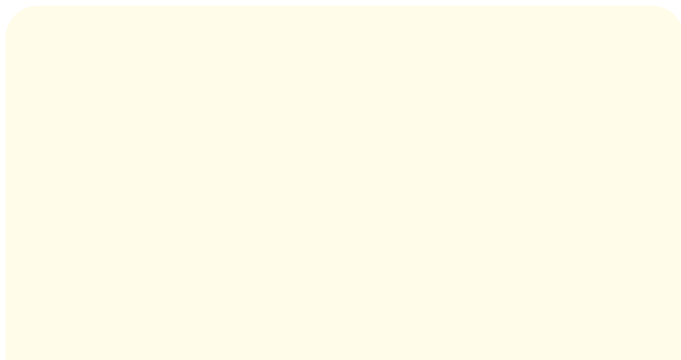

Figure 5: Total number of Infected cases in Eastern Mediterranean Region, Iran, Qatar and UAE.

African region: This graph is plotted for African region. Graphical curve is increasing exponentially approximately after 14 March, as South Africa shows highest increase in reported cases as compared to other countries of this region. Number of countries are also increasing along with number of case surge. This region has lowest number of cases as compared to other regions of WHO (Figure 6).
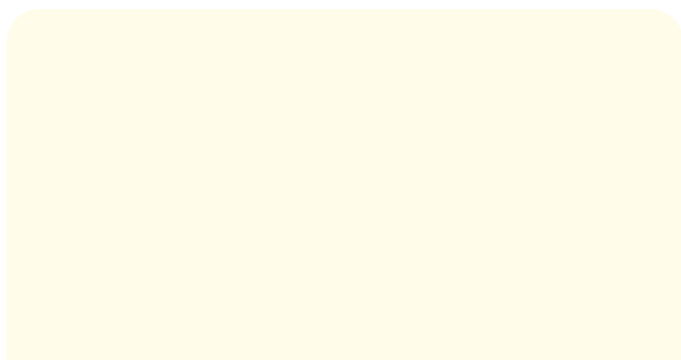

Figure 6: Total number of Infected cases in African Region, South Africa, Algeria and Senegal.

Globally infected cases: This graph shows trend of infected cases across the globe. Global trend is increasing exponentially, current infected cases is around 10 lakhs. Earlier in January the global graphical trend was similar to West pacific region which later on upon increase in cases graph started to increase exponentially. This signifies that there is increase in number of countries also increasing around the world along with increased cases. Graphical trend may change upon increase in cases (Figure 7).
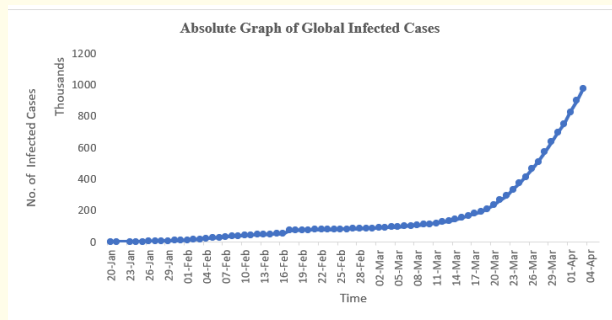
Figure 7: Total number of infected cases globally.

Total death cases: This graph is about the total number of death cases globally according to WHO zone and 18 randomly selected countries. Global death cases have exponential trend followed by European region, Italy and Spain curve. This indicates there is highest number of death reported in European region among all six zone till now. However, among selected countries Italy faces highest number of death followed by Spain till now (Figure 8).

Figure 8: Total number of death cases globally, WHO six zone and randomly selected eighteen countries.

Global infected, cured and death cases: This graph describe about the total infected, death and cured cases globally. By the graph we can make out that there is marked increase in recovered cases. This comparative graph shows exponential increase in infected cases and linear increase in recovered cases (Figure 9).
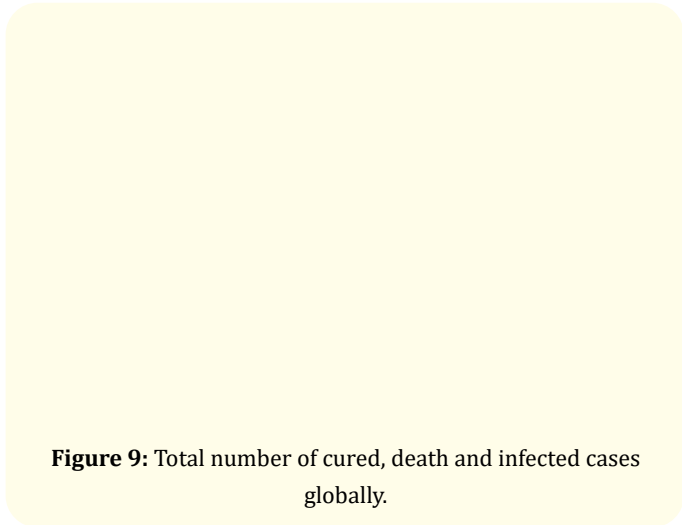
Figure 9: Total number of cured, death and infected cases globally.

Basic reproduction number (R0): It is a statistical analytical tool for an ecological model and signifies the rate of spread of infection at the Individual level. It is defined as an average number of secondary cases arises from the primary cases in a very large population. This is a non-negative, unitless quantity. Basic reproduction number is calculated by gamma distributed generation interval time. In this article generation interval is taken from (Meili Li., *et al*.) study 3.3 (95% confidence Interval- 2.3, 4.3) and variance as 3.1 (95% confidence Interval 1.0, 8.0). Considering same value calculated the R0 value of different countries and WHO zone based on the Infected number of cases from 1 March 2020 to 03 April 2020 data. Below mentioned table is plotted giving R0 value by two methods, Exponential growth and Maximum Likelihood Method. This table gives an estimate about what more steps are needed to decrease the R0 value. R0 value is highest for Region of Americas 2.08 among all the six WHO zone. Whereas R0 for USA is 2.11 which is highest among all the selected countries for study followed by Spain 1.77, India 1.76 (Table 1).

Machine learning forecasting

Forecast is the analytical capability using appropraite model to predict the future course of action. Epidemic forecasting enables to predict extent of disease spread as well as cases counts to give an insight to public health and policy makers for taking effective measures. Appling Poisson Regression method where y_p is predicted

count, is predicted event rate for p^{th} sample, is regressors for p^{th} sample, is regressor coefficient vector. forecast is done based on above model,graph of Mean Infected number of cases is plotted, so that there can be effective prepardness to combat the situation. This forecasting is done using machine learning. Firstly trained the machine using poisson regression about the infected number of cases rising trend and tested machine for the same thereafter predicted the future increase of cases. This output by machine is plotted in terms of graph. Red line depicts the actual count graph and the green one is the predicted graph. However, change in parameters will make the results vary from this predicted value.Empirical forecast involve lot of parameters to be considered as there are less number of data available which can vary the forecast result (Figure 10).

WHO zone/Country	Mean Exponential Growth Method	Mean Maximum Likelihood Method
China	1.00	1.07
West Pacific Region	1.02	1.09
USA	2.11	1.76
Region of Americas	2.08	1.75
India	1.76	1.61
South-East-Asia Region	1.70	1.51
Spain	1.77	1.52
European Region	1.65	1.46
Iran	1.37	1.31
Eastern Mediterranean Region	1.40	1.33
African Region	1.90	1.61

Table 1: Basic reproduction number (R0) (Data interval from 1 March 2020 to 03 April 2020).
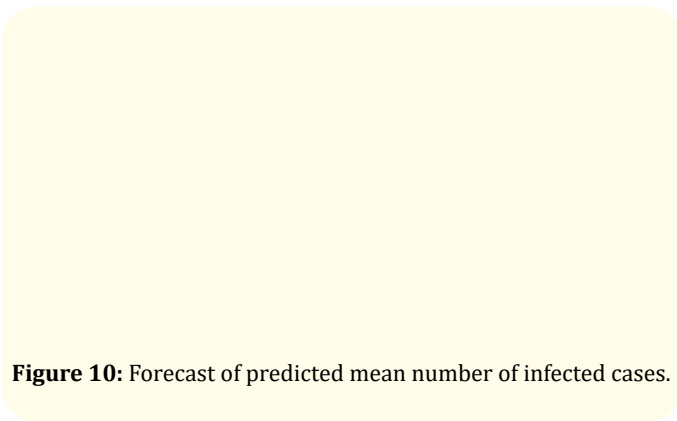

Figure 10: Forecast of predicted mean number of infected cases.

Discussion

In this article author wants to bring the attention of the people, administrator, epidemologist, policymakers towards the trends of epidemic effect on the basis of data collected from ongoing epidemic till the writing of this article. As graphs plotted are self explanatory about the trend of increasing cases among WHO six zones as well as 18 randomly selected countries. We have plotted a comparative graph for infected cases trends zone wise, total death among the the six zones and 18 countries. Total global infected cases, total cured, infected and death cases in a comparative graph. The nature of graphical curve and its significance is described in Trend section below each graph. Basic Reproduction Rate (R0): It is a non negative, unitless value first introduced by Alfred Lotka and Ronald Ross [16]. George Macdonald in (1957) was the first to call it as Basic Reproduction Rate and denoted by Z_0. Anderson and May popularized the name and gave symbol R0 [17]. It is an statistical analytical tool for an ecological model, widely used in Infection model. Basic Reproduction Number is defined as average number of secondary cases arising from a single primary case in a very large susceptible population. There are various method of calculating R0 such as Survival function method, final size equation, Next generation Method, existence of an epidemic equilibrium, constant term of characteristic polynomial, etc [16]. Value of R0 and its Interpretation [17]. R0 value signifies the amount of efforts needed to eradicate the epidemic. If R0 < 1: Infection is declining in population and will die out. If R0 = 1: infection is sustained in the population. If R0 >1: Infection is rapidly involving the susceptible population and have higher probability of epidemic outbreak.

There are realistic models such as Empirical and SEIR model which may provide a more realistic estimate than other non realistic models but peeping into other side of this model we can find that it requires more amount of data to interpret and are harder to calculate too, in an ongoing epidemic it is very difficult to get so much of data [18]. In ongoing epidemics have information of transmitability in terms of increase in new cases and the increase is exponential in initial phase of epidemics [19] (Pybus., *et al.* 2001) used linear equation value of reproductive number is related to growth rate R= 1+rTc [20]. Where r is defined as rate of exponential growth and T_c is the mean generation interval also called as mean duration between time of infection of a secondary infectee to the time of infection of its primary infector. Using gamma distributed mean generation interval we have calculated R0 for various regions of WHO Zone and few countries. However we have not gone

into the calculation of generation interval for each demographic zone. We imported the data of mean generation interval from the (Meili Li., *et al.*) study of COVID19(*21*). Mean generation interval estimated as 3.3 (95% confidence Interval- 2.3, 4.3) and variance as 3.1 (95% confidence Interval 1.0, 8.0) [21]. we have considered the data from the point the graph is entering into exponential phase on absolute graph till the data of last writing the article. we coded the data on RStudio software using previosly calculated generation interval.

Forecast using Machine Learning: Epidemic disease classical analysis is there from ages before the intervention of technology [1]. with the advancement of technnology the data scientists came up with various models to describe the epidemic pattern, predicting the future outbreak or its impact via data acquisition. These models use ths data to train the machine for more dynamic input which are cross validated, this model is known as Machine Learning Process [1]. Forecast is the analytical capability using appropraite model to predict the future course of action [22]. Epidemic forecasting enables to predict extent of disease spread as well as cases counts to give an insight to public health and policy makers for taking effective measures. Predictive modelling can be an important tool for the epidemic forecasting [23]. Basically forecasting is done in Six steps namely data collection, data cleaning, data storage and sharing, analytics:training model, analytics: forecasting [22]. (A) Data collection: This step involves collection of raw data which is a slow step and involves lot of resources. It becomes a challenge to collect the data accurately and precisely. (B) Data cleaning: This step involves sorting and making raw data in terms of scientific terms and machine readable format. (C) Data storage and sharing: This step is about storage of structured data and sharing it for the analytic steps. (D) Analytics: Training model: Here with the available data we train the machine to know the changing pattern with appling appropriate tools. (E) Analytics: Forecasting: Trained machine is used to forecast the data of near future.In this article 80 percent of data is used to train the machine whereas with the rest 20 percent data machine is tested. Poisson regression model is applied because of added advantage over Ordianry Linear Regression. This model include a Skewed data, Discrete distribution, and restriction on predicted value to non-negative numbers. Poisson model asssumes that the error follow poission not a normal distribution [24].Forecasting requires a lot of parameteres and some are changing parameters as well as it need lot of extensive resesarch and artificial intelligence to have empirical result. Therefore only handfull of data across the globe are analysed and forecasted in the form of Graph. Machine Learning Forecast data is calculated graphically. We forecasted the Predicted mean number of Covid infected cases of 5 April, 10 April, 15 April and 20 April 2020 on 4th April 2020. As the time elaspsed of forecast, thus a comparative table of Forecasted cases and acutal cases is prepared to check the effectivity of methodology. (Table 2).

Date	Region/Country	South-East-Asia Region	Eastern Mediterranean Region	African Region	India	Canada	Iran
05 April 2020	Mean Predicted Cases	9255.55 ± 160.32	113841.62 ± 356.48	17840.09 ± 395.37	2974.61 ± 99.02	22003.83 ± 436.72	100068.46 ± 348.55
	Actual Cases	7816	70293	6420	3374	12938	55743
10 April 2020	Mean Predicted Cases	18902.50 ± 418.24	203181.196 ± 811.02	56590.24 ± 1785.65	6626.06 ± 294.81	62654.97 ± 1692.52	181035.89 ± 797.84
	Actual Cases	12978	88657	8789	6412	19759	66220
15 April 2020	Mean Predicted Cases	30619.76 ± 394.35	260829.2 ± 825.4	179508.93 ± 7390.83	13255.28 ± 379.70	178383.07 ± 6127.96	327515.71 ± 1759.41
	Actual Cases	20287	107389	11367	11439	26146	74877
20 April 2020	Mean Predicted Cases	59268.64 ± 941.60	443582.9 ± 1693.4	569417.17 ± 28977.26	28988.81 ± 1037.36	507869.03 ± 21215.29	592515.31 ± 3767.91
	Actual Cases	29576	129433	14760	17265	33909	82211

Table 2: Comparative data of forecasted cases v/s actual cases.

Conclusion

We have plotted few graphs which are describing the rising trends as well as current scenario of pandemic COVID19 disease across WHO zone and few randomly selected countries. Using the generation interval data we have calcualted few R0 value of WHO zone and few countries between 1 March to 03 April 2020 data. Highest value of R0 is seen in USA i.e. 2.11 and Regions of Americas as 2.08 which signifies that one primary case in USA is spreading to 2.11 other susceptible person. Whereas China having least R0 value 1.00 which signifies that the spread is gradually going to be halt down in coming days. This paper is highly significant as it is compilation of real time needed information for international community. Introduction of Machine Learning for forecasting is more accurate than linear models this can be a highly used for future pre-pardness to mitigate epidemic. Rather than foccusing on calculaion of parameters this paper is concentrated on application of methodology for maximum result for mobilising communities.

Scope for Future Research

This paper has got limitations to address in future communication and research work. In the calculation of Basic Reproduction Number we have not calculated the Generation Interval for each geographical region and countries as it becomes harder to calculate data to interpret as well as invovles lot of parameters to consider which becomes very difficult to get in an ongoing epidemic. In Forecast of epidemics we have used Machine Learning using handfull of data. Empirical forecasting involves a lot of changing and constant parameters to include and use of Artificial Intelligence can take the results near to ideal expected value. These are the fields which need to be address and arises prime consern for future research for an early warning and effecting policy decisions to curb such pandemic.

Acknowledgement

Bibliography

1. Ibrahim N., *et al*. "Predictive analysis effectiveness in determining the epidemic disease infected area". *AIP Conference Proceedings* 1891.1 (2017): 20064.

2. Kapil A Ananthnarayan and Paniker's Textbook of Microbiology. University Press (India) Private Limited, Hyderabad (2015): 559-560.

3. Schoeman D and Fielding BC. "Coronavirus envelope protein: Current knowledge". *Virology Journal* 16.1 (2019): 1-22.

4. Johns Hopkins Center for Health Security. Coronaviruses: SARS, MERS, and 2019-nCoV.

5. World Health Organization. WHO guidelines for the global surveillance of severe acute respiratory syndrome (SARS).

6. Park SW., *et al*. "A practical generation-interval-based approach to inferring the strength of epidemics from their speed". *Epidemics* 27.1 (2019): 12-18.

7. Elsevier's Novel Coronavirus Information Center.

8. World Health Organization. Coronavirus.

9. Raoult D., *et al*. "Coronavirus infections: Epidemiological, clinical and immunological features and hypotheses". *Cell Stress* 4.4 (2020): 66-75.

10. Dong E., *et al*. "An interactive web-based dashboard to track COVID-19 in real time". *Lancet Infection Disease* 3099.20 (2020): 19-20.

11. World Health Organization. Naming the coronavirus disease (COVID-19) and the virus that causes it.

12. WHO. Coronavirus disease 2019 (COVID-19). Situation report 11. January 31, 2020. Geneva: World Health Organization (2020).

13. WHO. Coronavirus disease 2019 (COVID-19). Situation report 51. March 11, 2020. Geneva: World Health Organization (2020).

14. Lauer SA., *et al*. "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application". *Annals of Internal Medicine* (2020): 1-15.

15. N van Doremalen., *et al*. "Aerosol and surface stability of HCoV-19 (SARS-CoV-2) compared to SARS-CoV-1". *The New England Journal of Medicine* (2020).

16. Mikucki Michael. "Sensitivity analysis of the basic reproduction number and other quantities for infectious disease models". (2012).

17. Heesterbeek JAP and Dietz K. "The concept of R_o in epidemic theory". *Statistica Neerlandica* 50 (1996): 89-110.

18. Park SW., *et al*. "A practical generation-interval-based approach to inferring the strength of epidemics from their speed". *Epidemics* 27 (2019): 12-18.

19. Wallinga J and Lipsitch M. "How generation intervals shape the relationship between growth rates and reproductive numbers". *Proceedings B is the Royal Society's* 274.1609 (2007): 599-604.

20. Pybus OG., *et al*. "The Epidemic Behavior of the Hepatitis C Virus". *Science* 292.5525 (2001): 2323-2325.

21. Li M., *et al*. "Transmission characteristics of the COVID-19 outbreak in China: a study driven by data". *medRxiv* (2020): 2020.

22. George DB., *et al*. "Technology to advance infectious disease forecasting for outbreak management". *Nature Communication* 10 (2019): 3932.

23. Desai AN., *et al*. "Real-time Epidemic Forecasting: Challenges and Opportunities". *Health Security* 17.4 (2019): 268-275.

24. The Analysis Factor. Regression Models for Count Data.