Statistical analysis on the COVID-19 infection spread in United State of America: A Prophet Forecasting Model

S.R. Mishra
Siksha O Anusandhan University
Bhubaneswar, India
satyaranjan_mshr@yahoo.co.in

Priya Mathur
Poornima Institute of Engineering & Technology
Jaipur, India
Drpriyamathur21@gmail.com

Amit Kumar Gupta*
ASET-Amity University Rajasthan
Jaipur, India
dramitkumargupta1983@gmail.com

Sushama Baag
College of Basic Science & Humanities OUAT
Bhubaneswar, India
sbaag22@gmail.com

Kapil Kumar Nagwanshi*
ASET-Amity University Rajasthan
Jaipur, India
dr.kapil@ieee.org

Shrinath Tailor
Sri Balaji College of Engineering & Technology
Jaipur, India
tailorshrinath@gmail.com

Ashwin Verma
Nirma University
Ahmedabad, Gujarat, India
ashwin.verma@nirmauni.ac.in

2021 Sixth International Conference on Image Information Processing (ICIIP) | 978-1-6654-3361-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIIP53038.2021.9702595

***Abstract*—In the current scenario, the pandemic COVID-19 spread globally starting from the end of 2019, in Wuhan, a city of China. As per the current data taken up to 26th of May 2020, globally there are a huge number of people are affected (Approximately 3 billions) by the pandemic. Though the entire data varies depending upon the several parameters like, population size, congestion of area, climate condition, awareness of peoples etc. we have only analyzes on the data of the country USA. The entire data is partitioned into various categories such as: infected rate, mortality rate. A statistical analysis is prepared to analyze or predict the future strategies of the infected rate as well as the removal (Death/cured) rate. The growth of both the infected and the removed can be predicted with the same observed data taken on daily basis from 15th February 2020. We retrieved these data from an authenticate source provided by ”Worldometer” (http://www.worldometers.info). However, Prophet Forecasting Model (PMF) is used to simulate and discussed for the prediction of the mortality rate, active rate due to pandemic COVID-19. The proposed method is also tested for accuracy of model via cross validation method.**

***Index Terms*—Coronavirus, infected and removed data, statistical analysis, Prophet Forecasting Model, Mortality Rate, Active Rate.**

I. Introduction

The global influenza pandemic is one of the severe pandemic in the history of last ten decades. Starting from the observation made in the 1918 influenza pandemic which was caused by H1N1 virus along with the genes of avian origin. During 1918-1919 the spreading was globally and initially located in the United States and identified in military personnel. The spreading of the influenza generally more active in the spring and the same also happens in the spring season in US. In estimation it is viewed that more than one third of the world's population (nearly 500 millions) was infected by the influenza virus. In US only the number of deaths was at least 6.75 lakhs and globally that was 50 millions. Statistically it was observed that the people younger than 5 years, between 20-40 years and more than 65 years old, the mortality rate is high. However, the H1N1 virus has been synthesized and evaluated evaluated properties are not so clear to understand. To protect the virus neither any vaccine nor antibiotics were developed. The control efforts globally were limited to ”non-pharmaceutical interventions such as isolation, quarantine, good personal hygiene, use of disinfectants, and limitations of public gatherings” [1]–[3].

Asian Flu, is an another pandemic caused by influenza A virus known as H2N2 in the starting of 1957, month of February. Initially it is emerged in East Asia and the virus was comprised of three types of genes originated from an avian influenza A virus along with H2 hemagglutinin and the N2 neuraminidase genes. In February 1957, the first infected people were reported from Singapore then in April 1957 at Hong Kong and in summer the peoples are infected in US. However, from the global death toll of 1.1 million only in US the number is nearly 1.16 lakhs. Another pandemic caused by influenza A, named as H3N2 virus was occurred in the year 1968. It was first originated in US in September 1968. From the statistical point of view the people from 65 years and older were infected more and a total of one million deaths are reported out of which from Us only the number is one lakh. From the above report it is concluded that every time the people from US are always affected with more numbers [4]–[7].

The ongoing coronavirus disease 2019, named COVID-19 pandemic occurred in the end of 2019 i.e. in December 2019 originated in Wuhan a city of China. The current outbreak is something different in symptoms in comparison to earlier influenza pandemics. The investigations on the novel coronavirus is going on. Scientists are racing on to find out more about COVID-19. Initially, the patients of Wuhan presenting with pneumonia of unknown origin. With a great damage in China the virus is spread globally and nearly eight lakhs people were infected globally till the date of 28th of March 2019 [7], [8]. It is noted that, US is one of the countries in which the rate is again more as compared to the pandemics reported earlier [9]. As per our collected report till 28th of March 2020 [10]–[12], the total number of infected people in US is 1, 23,578 and active cases are 1, 18,127. The number of people removed (death and cured) is 5,451 out of which 2,220 peoples are died.

In the current investigation it is aimed that, statistical analysis is made using the proposed model with the daily data provided for the United States. Starting from 15th February 2020 (earlier data not traced) data for the total infected cases, cured/deaths cases are present up to 28th March 2020 in daily basis. The simulation is carried out for the fitting of the data both for infected and mortality rate and further, Prophet Forecasting Model is used to predict the Mortality rate and active rate (infection rate-Cured rate- Mortality rate) and accuracy of Prophet Model is also checked by ”Cross Validation Method” on the basis of 50 days horizon. The experimental results clearly show that the USA will face serious disaster due to COVID-19. It is also shows by simulated prediction and trend of COVID-19 data set the pandemic may continue till the end of this year. A good sign in this prediction model is that the active rate is decreasing [13], [14].

II. Earlier statistical data for mortality rate, age-sex and pre-existing case

In a report conducted by WHO-China joint Mission on Novel coronavirus 2019 it is observed that the mortality rate (number of deaths/number of cases) i.e. the probability of dying if infected by the virus (%) is nearly 27% from the people of age between 50-80 years (80+ it is 14.8%, 70-79 it 8%, 60-69 it is 3.6%, 50-59 it is 1.3%) however, no fatalities are tolled for the people of 0-9 years till dated presented in Table I. If we compared the sex ratio, it is observed that the confirmed and all cases of death ratio for male is 4.7% and 2.8% respectively whereas for female people it is 2.8% and 1.7% respectively displayed in Table II. The statistics of pre-existing conditions for the mortality rates like by coronavirus disease, diabetes, chronic respiratory diseases, Hypertension and cancer are also reported with both confirmed and all cases. These rates are 13.2% and 10.5% respectively for the diseases by coronavirus, 9.2%, and 7.3% respectively by the diabetes, 8.0%and 6.3% from chronic respiratory diseases, 8.4% and 6.0% by Hypertension, 7.6% and 5.6% by cancer. However, the mortality rate for all cases with no pre-existing conditions is only 0.9% which is presented in Table III.

TABLE I: Mortality rate both for confirmed and all cases.

Age	Mortality rate (Confirmed Cases)	Mortality rate (All Cases)
80+ Year Old	21.9%	14.8%
70-79 Years Old		8.0%
60-69 Years Old		3.6%
50-59 Years Old		1.3%
40-49 Years Old		0.4%
30-39 Years Old		0.2%
20-29 Years Old		0.2%
10-19 Years Old		0.2%
0-9 Years Old		No fatalities

TABLE II: The sex ratio for both confirmed and all cases.

SEX	**Mortality rate** (Confirmed Cases)	**Mortality rate** (All Cases)
Male	4.7%	2.8%
Female	2.8%	1.7%

TABLE III: Pre-existing conditions for both confirmed and all cases.

PRE-EXISTING CONDITION	**Mortality rate** (Confirmed Cases)	**Mortality rate** (All Cases)
Cardiovascular Disease	13.2%	10.5%
Diabetes	9.2%	7.3%
Chronic respiratory dieses	8.0%	6.3%
Hypertension	8.4%	6.0%
Cancer	7.6%	5.6%
No-Preexisting Conditions	–	0.9%

III. Statistical analysis for daily infected cases, mortality rate, trends for pandemic COVID-19

We have performed the Statistical analysis for daily infected cases, mortality rate, trends due to pandemic COVID-19. The Polynomial curve fitting method has been use to analysis of daily infected from 15th February to 28th March 2020 data set [11], [12]. The data has been fitted in to 5th order polynomial fitting which is shown in Figure 1. The Type of trend in changing the daily infected people of US is also analyzed which is shown in figure 2. The trends is calculated for the identification of nature of increasing or decreasing the infection rate which is require to correct prediction of data. The figure 3 showed the increments in infection rate on daily basis which shows on which rate the infection rate is increasing. The figure 2 shows from 15th February to 28th March 2020 with 6th order polynomial fitting [15]–[17].

Form the above statistical analysis it is found that the number of infected people and death cases is increasing with the higher percentages. The prediction Model for mortality rate and active rate in US due to pandemic Novel COVID-19 is discussed in the section 4.

IV. Prediction of mortality rate and active rate in US due to pandemic Novel COVID-19

In this study we have adopted the Prophet Model for forecasting the mortality rate and active rate in US due to Novel COVID-19. The below section briefly describe the prophet Model.

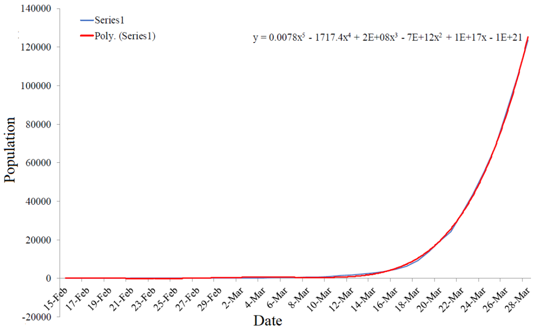

Fig. 1: Daily infected cases from 15th February to 28th March 2020 with 5th order polynomial fitting.

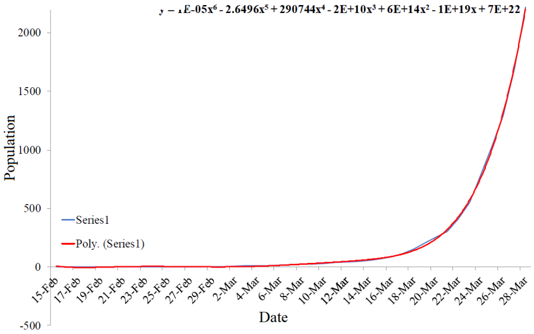

Fig. 2: Daily death cases from 15th February to 28th March 2020 with 6th order polynomial fitting.

A. *Prophet Forecasting Model*

The Prophet model [18], [19] has been designed by the Facebook for forecasting the time series data. The prophet model used the approach of additive model in which the trends (Non-Linear) are fitted with seasonality (yearly, weekly, and daily) and the effects of holidays. The prophet model has combined trend, seasonality and holidays in the form of linear Eq. 1.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{1}$$

Here $g(t)$ represents the trends which observed the aperiodic changes in time series values, $s(t)$ represents the seasonality observed the periodic changes in time series values (weekly, daily and yearly), $h(t)$ represents the holiday changes in time series values which may irregularly arise for few days and ϵ_t represents the error.

B. *Trends $g(t)$*

The trends $g(t)$ can be described by two models; first is saturating growth model, and a piecewise linear model.

1) Saturating Growth Model: In saturating growth model, on the basis of data nature it identifies that how the data has grown and how it is expected to continue growing. This growth model is similar to natural ecosystem, which defines that nonlinear growth is saturated at the carrying capacity. The saturating growth model define the trends g(t) in the form of Eq. 2.

$$g(t) = \frac{C}{1 + e^{-k(t-m)}} \tag{2}$$

where C represents the carrying capacity, and k represents the growth rate and m is offset.

2) Piecewise Linear Model: In piecewise linear model, in saturating growth model the carrying capacity and growth rate is kept constant but in the data it may be changed or variable in nature. In this model the changing in growth rates are measured at points that are called change points and carrying capacity is changed from constant to variable with respect to time denoted by C(t). The final equation for trends g(t) in piecewise linear model given by Eq. 3.

$$g(t) = \frac{C(t)}{1 + e^{-k + a(t)^T \delta}(t - (m + a(t)^T \gamma)} \tag{3}$$

where k represents the base growth rate, m is offset, δ represents the vector of rate adjustment, $a(t)^T$ represents the vector of changing in the rate for all change points and γ represents the rate of adjustment.

C. *seasonality $s(t)$*

b. The seasonality $s(t)$ is calculated on the basis of Fourier transformation which provide the better result on periodic effects. The prophet model has measured the seasonality on the basis of Eq. 4.

$$s(t) = \sum_{n=1}^{N} \left(a_n cos \left(\frac{2\pi \times n \times t}{p} \right) + a_n sin \left(\frac{2\pi \times n \times t}{p} \right) \right) \tag{4}$$

where the P represents the regular period.

D. *Holiday $h(t)$*

The holiday $h(t)$ are assumed independent events that are added in prophet model which has effect the time series. The prophet model automatically adds the effect of holiday if it comes under the date range in particular data set.

Considering these advantages of Prophet Model we have used the prophet Model for predicting the mortality rate in US due to novel COVID-19.

V. Experimental Setup and Results

We retrieved these data from an authenticate source provided by "Worldometer" ($http : //www.worldometers.info$) from date January 22, 2020 to May 25, 2020. We have applied the prophet method to predict the mortality rate in US and also predicted the mortality rate for coming two months. The actual mortality rate on date May 25, 2020 was approximately 6% while our model has predicted the approximately 7%. The trends of mortality rate and active rate in US due to COVID-19 as shown by Fig.5 and Fig.6 respectively. The horizontal shows the dates and vertical shows the predicted trends.

The model is also cross validated on the basis of 50 days horizon for performance measurements of the predicted data. The results are shown in Table IV. The table had shown all

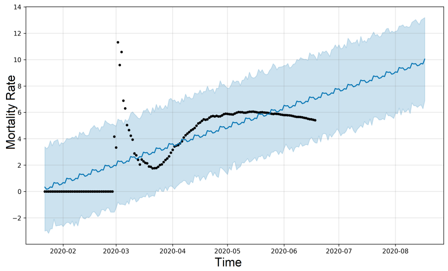

Fig. 3: Showing the trends of mortality rate in US (COVID-19).

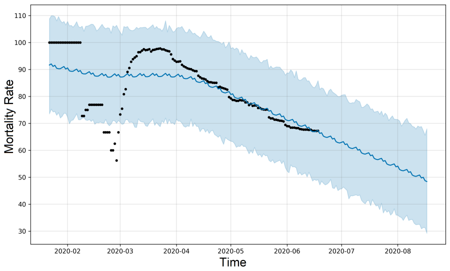

Fig. 4: Showing the trends of active rate in US (COVID-19).

the performance parameter for the predicted model in terms of Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Median Absolute Percentage Error (MAPE). It is very clear from the table that the proposed model predict the most accurate results. The performance matrix parameter is also shown by Fig.5-14 for active rate and mortality rate of proposed method.

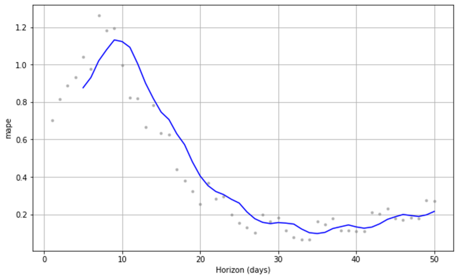

Fig. 5: MAPE for mortality rate prediction (On the basis of 50 days horizon)

We have predicted the mortality rate in US due to COVID-19 and the proposed model has predicted the data for commence two months. The prediction has given that the mortality rate becomes nearly 10.40% and active rate becomes nearly 66.90% on the date July 24, 2020. This prediction shows that the mortality rate due to COVID-19 in US will be one of height mortality rate than other disease i.e. shown in Table IV. The figure 17 shows the actual mortality rate and predicted mortality rate in US due to COVID-19 using our proposed model. The figure 18 shows the actual active rate and predicted active rate in US due to COVID-19 using our proposed model [20]–[22].

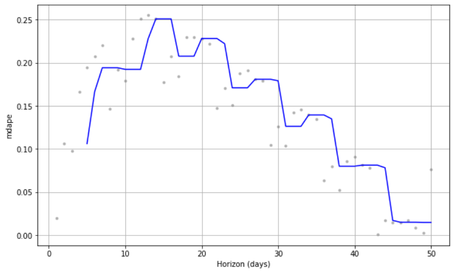

Fig. 6: PMADPE for mortality rate prediction (On the basis of 50 days horizon)

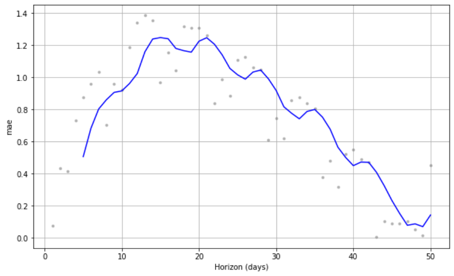

Fig. 7: MAE for mortality rate prediction (On the basis of 50 days horizon)

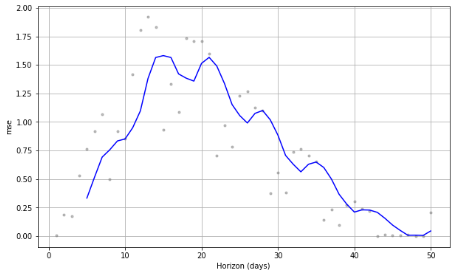

Fig. 8: MSE for mortality rate prediction (On the basis of 50 days horizon)

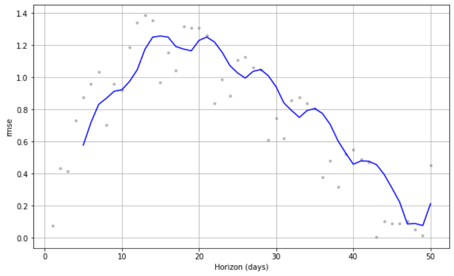

Fig. 9: RMSE for mortality rate prediction (On the basis of 50 days horizon)

VI. Conclusion

The present paper provides a statistical analysis on the current pandemic novel COVID-19 for the United State of

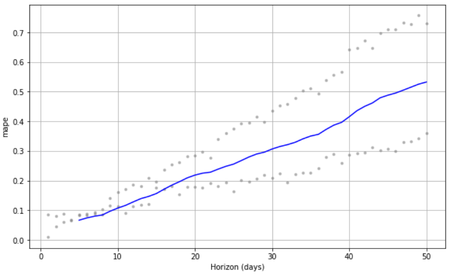

Fig. 10: MAPE for active rate prediction (On the basis of 50 days horizon)

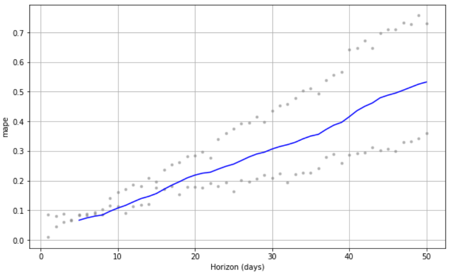

Fig. 11: MADPE for active rate prediction (On the basis of 50 days horizon)

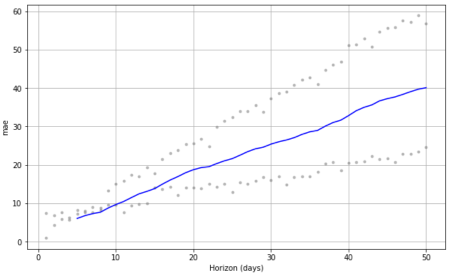

Fig. 12: MAE for active rate prediction (On the basis of 50 days horizon)

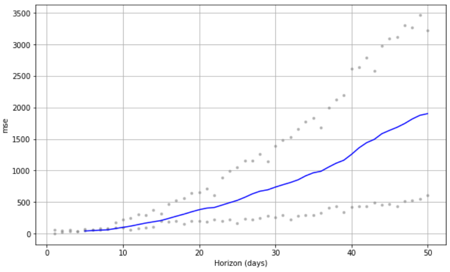

Fig. 13: MSE error for active rate prediction (On the basis of 50 days horizon)

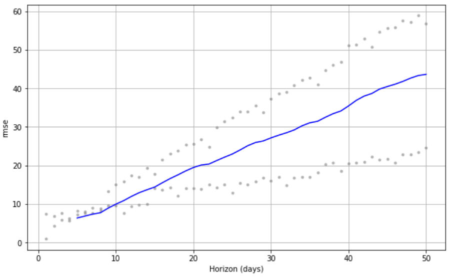

Fig. 14: RMSE error for active rate prediction (On the basis of 50 days horizon)

TABLE IV: Showing the performance matrix of proposed Method.

Horizon	MSE	RMSE	MAE	MAPE	MDAPE
5 days	2.787647	1.669625	1.665824	0.876204	0.889901
6 days	2.95452	1.718872	1.718299	0.930822	0.930466
7 days	3.387509	1.840519	1.827141	1.020113	0.977394
8 days	3.816131	1.953492	1.937297	1.078926	1.041522
9 days	4.390813	2.095427	2.078324	1.131757	1.183962
10 days	4.588444	2.142065	2.129617	1.123058	1.183962
11 days	4.668811	2.160743	2.152221	1.092227	1.183962
12 days	4.325564	2.079799	2.068588	1.003911	0.998026
13 days	3.865071	1.965978	1.948751	0.900755	0.823237
14 days	3.546925	1.883328	1.876776	0.818578	0.819707
15 days	3.358615	1.832652	1.828065	0.745975	0.783742
16 days	3.41432	1.847788	1.842948	0.706541	0.668178
17 days	3.143236	1.772917	1.760092	0.630485	0.635011
18 days	2.940423	1.714766	1.689986	0.572568	0.626065
19 days	2.376507	1.541592	1.512163	0.480709	0.439429
20 days	1.897597	1.377533	1.338161	0.404936	0.378593
21 days	1.557777	1.248109	1.235275	0.353241	0.367589
22 days	1.415324	1.189674	1.180094	0.321907	0.324447
23 days	1.410514	1.187651	1.178195	0.305393	0.296024
24 days	1.301444	1.140808	1.123956	0.280252	0.282759
25 days	1.221652	1.105284	1.075269	0.260048	0.282759
26 days	0.904952	0.95129	0.916484	0.212493	0.198743
27 days	0.68609	0.828305	0.782812	0.176568	0.155126
28 days	0.549597	0.741348	0.7212	0.157098	0.155126
29 days	0.531603	0.729111	0.710529	0.150141	0.155126
30 days	0.613393	0.783194	0.760275	0.155914	0.163957
31 days	0.612472	0.782606	0.759505	0.152817	0.163957
32 days	0.60004	0.774623	0.74569	0.147729	0.163957
33 days	0.444343	0.66659	0.627337	0.121325	0.114332
34 days	0.335876	0.579548	0.534562	0.101604	0.077693
35 days	0.312077	0.558638	0.521541	0.097215	0.077693
36 days	0.372054	0.609962	0.564194	0.103509	0.077693
37 days	0.54364	0.73732	0.68414	0.123956	0.145805
38 days	0.6043	0.777367	0.743561	0.133663	0.145805
39 days	0.661626	0.813404	0.801249	0.143136	0.145805
40 days	0.585116	0.764929	0.750554	0.132643	0.115189
41 days	0.534038	0.730779	0.715141	0.125748	0.112715
42 days	0.612959	0.782917	0.75087	0.131701	0.112715
43 days	0.798343	0.8935	0.852412	0.149152	0.112715
44 days	1.086567	1.042385	0.996381	0.173035	0.202446
45 days	1.230656	1.109349	1.081962	0.186965	0.202446
46 days	1.354569	1.163859	1.157404	0.199016	0.202446
47 days	1.296025	1.138431	1.131475	0.193396	0.181586
48 days	1.238959	1.113085	1.105583	0.188401	0.17923
49 days	1.389837	1.178913	1.156291	0.197274	0.17923
50 days	1.676132	1.294655	1.264268	0.215513	0.181586

America using the daily data of both active and mortality retrieved from the useful sourced cited earlier. Nonlinear polynomial fitting is used to predict the future situation and then the error analysis is obtained considering the Prophet Forecasting Model, the time series data. The proposed method predicted that the mortality rate becomes nearly 10.40% and active rate becomes nearly 66.90% on the date July 24, 2020. A negative sign from whole prediction is that the mortality rate is continuous increasing and the positive sign that active rate is continuously decreasing. From the aforesaid prediction, it is clearly shown that the pandemic COVID-19 will create a serious disaster and from the trend of the current situation the pandemic may continue till the end of this year. Therefore, to

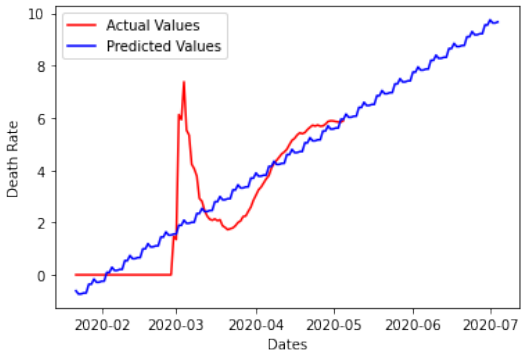

Fig. 15: Actual and Predicted Mortality rate in US due to COVID-19.

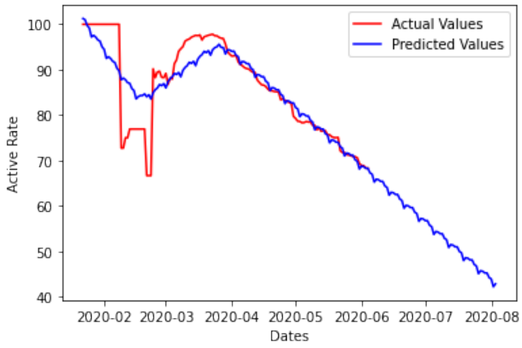

Fig. 16: Actual and Predicted Active rate in US due to COVID-19.

avoid the same situation proper care should to take care that provided by the government. There are the proper sanitizations, social distancing, and wearing of mask may be suggested to break the chain.

References

[1] L.-J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Transactions on neural networks*, vol. 14, no. 6, pp. 1506–1518, 2003.

[2] G. W. Flake and S. Lawrence, "Efficient svm regression training with smo," in *Machine Learning*. Citeseer, 2000.

[3] C. Zhao, H. Zhang, X. Zhang, M. Liu, Z. Hu, and B. Fan, "Application of support vector machine (svm) for prediction toxic activity of different data sets," *Toxicology*, vol. 217, no. 2-3, pp. 105–119, 2006.

[4] V. Kumar, J. M N, R. R, A. C M, and R. S, "Statistical analysis on novel corona virus: Covid-19," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 1, pp. 95–103, 2020.

[5] A. Hiranya . S, J. Priya, and V. P. andLakshminarayanan Arivarasu, "Challenges faced by china on covid-19," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 1, pp. 2230–2235, 2020.

[6] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, and R. Agha, "World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19)," *International journal of surgery*, vol. 76, pp. 71–76, 2020.

[7] Anonymous, "Timeline of who's response to covid-19, 2020," *URL: https://www. who. int/news-room/detail/29-06-2020-covidtimeline [accessed 2020-07-14]*, 2020.

[8] M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, "Features, evaluation and treatment coronavirus (covid-19)," *Statpearls [internet]*, 2020.

[9] B. Singh, V. Kumar, and S. Tripathi, "A review of covid-19 based on current evidences," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 2, no. 8, pp. 1449–1459, 2020.

[10] W. Commons, "File:3d medical animation corona virus.jpg — wikimedia commons, the free media repository," 2020, [Online; accessed 26-February-2021]. [Online]. Available: https://commons.wikimedia.org/w/index.php?title=File:3D_medical_animation_corona_virus.jpg&oldid=479016449

[11] WHO, "Who situation report," 2020, [Online; accessed 26-February-2021]. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters

[12] ——, "Coronavirus disease (covid-19) advice for the public: Mythbusters," 2020, [Online; accessed 26-February-2021]. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters

[13] S. Whitelaw, M. A. Mamas, E. Topol, and H. G. C. Van Spall, "Applications of digital technology in covid-19 pandemic planning and response," *The Lancet Digital Health*, vol. 2, no. 8, pp. e435–e440, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2589750020301424

[14] S. Basheer, K. K. Nagwanshi, S. Bhatia, S. Dubey, and G. R. Sinha, "Fesd: An approach for biometric human footprint matching using fuzzy ensemble learning," *IEEE Access*, vol. 9, pp. 26 641–26 663, 2021.

[15] Q. Li, W. Feng, and Y.-H. Quan, "Trend and forecasting of the covid-19 outbreak in china," *Journal of Infection*, vol. 80, no. 4, pp. 469–496, 2020.

[16] R. Samsudin, A. Shabri, and P. Saad, "A comparison of time series forecasting using support vector machine and artificial neural network model," *Journal of applied sciences*, vol. 10, no. 11, pp. 950–958, 2010.

[17] K. K. Nagwanshi and S. Dubey, "Estimation of centroid, ensembles, anomaly and association for the uniqueness of human footprint features," *International Journal of Intelligent Engineering Informatics*, vol. 8, no. 2, pp. 117–137, 2020.

[18] S. J. Taylor and B. Letham, "Prophet: forecasting at scale," 2017, [Online; accessed 24-February-2021]. [Online]. Available: https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale/

[19] E. O. Nsoesie, R. J. Beckman, S. Shashaani, K. S. Nagaraj, and M. V. Marathe, "A simulation optimization approach to epidemic forecasting," *PloS one*, vol. 8, no. 6, p. e67164, 2013.

[20] S. Swayamsiddha and C. Mohanty, "Application of cognitive internet of medical things for covid-19 pandemic," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 2020.

[21] E. J. Suba, W. J. Frable, and S. S. Raab, "Cost-effectiveness of cervical-cancer screening in developing countries." *The New England journal of medicine*, vol. 354, no. 14, pp. 1535–6, 2006.

[22] K. Liu, Y. Chen, R. Lin, and K. Han, "Clinical features of covid-19 in elderly patients: A comparison with young and middle-aged patients," *Journal of Infection*, vol. 80, no. 6, pp. e14–e18, 2020.