Results in Physics 40 (2022) 105855

Contents lists available at ScienceDirect

Results in Physics

journal homepage: www.elsevier.com/locate/rinp

Modified SIR model for COVID-19 transmission dynamics: Simulation with case study of UK, US and India

Pranati Rakshit [a], Soumen Kumar [b], Samad Noeiaghdam [c,d], Unai Fernandez-Gamiz [e,*], Mohamed Altanji [f], Shyam Sundar Santra [g,h]

[a] Department of Computer Science and Engineering, JIS College of Engineering, Kalyani, West Bengal, India
[b] Associate Consultant and Data Scientist in Tata Consultancy Services Ltd., Kolkata, West Bengal, India
[c] Industrial Mathematics Laboratory, Baikal School of BRICS, Irkutsk National Research Technical University, Irkutsk, 664074, Russia
[d] Department of Applied Mathematics and Programming, South Ural State University, Lenin prospect 76, Chelyabinsk, 454080, Russia
[e] Nuclear Engineering and Fluid Mechanics Department, University of the Basque Country UPV/EHU, Nieves Cano 12, 01006 Vitoria-Gasteiz, Spain
[f] Department of Mathematics, College of Science, King Khalid University, Abha 61413, Saudi Arabia
[g] Department of Mathematics, JIS College of Engineering, Kalyani, West Bengal 741235, India
[h] Department of Mathematics, Applied Science Cluster, University of Petroleum and Energy Studies (UPES), Dehradun, Uttarakhand 248007, India

ARTICLE INFO

Keywords:
COVID-19
SIR model
Prediction
Asymptomatic
R-Square score

ABSTRACT

Corona virus disease 2019 (COVID-19) is an infectious disease and has spread over more than 200 countries since its outbreak in December 2019. This pandemic has posed the greatest threat to global public health and seems to have changing characteristics with altering variants, hence various epidemiological and statistical models are getting developed to predict the infection spread, mortality rate and calibrating various impacting factors. But the aysmptomatic patient counts and demographical factors needs to be considered in model evaluation. Here we have proposed a new seven compartmental model, Susceptible- Exposed-Infected–Asymptomatic–Quarantined–Fatal–Recovered (SEIAQFR) which is based on classical Susceptible-Infected-Recovered (SIR) model dynamic of infectious disease, and considered factors like asymptomatic transmission and quarantine of patients. We have taken UK, US and India as a case study for model evaluation purpose. In our analysis, it is found that the Reproductive Rate (R_0) of the disease is dynamic over a long period and provides better results in model performance (> 0.98 R-square score) when model is fitted across smaller time period. On an average $40\% - 50\%$ cases are asymptomatic and have contributed to model accuracy. The model is employed to show accuracy in correspondence with different geographic data in both wave of disease spread. Different disease spreading factors like infection rate, recovery rate and mortality rate are well analyzed with best fit of real world data. Performance evaluation of this model has achieved good R-Square score which is $0.95 - 0.99$ for infection prediction and $0.90 - 0.99$ for death prediction and an average $1\% - 5\%$ MAPE in different wave of the disease in UK, US and India.

Introduction

Virus causes infections which later turns into disease in human & animal body and are therefore have an impact on medical, social, and economical life. Novel Corona virus (COVID-19) caused an outbreak of a typical pneumonia first in Wuhan, China in December 2019 [1] and then extended its deadly characteristics throughout the whole world infecting more than 210 million population worldwide taking a death toll of 4.47 million [2] till date as of 27th Aug 2021. After assessing it is amazing spreading power and significant harm towards human civilization, the World Health Organization (WHO) announced this new pneumonia outbreak a "global pandemic" [3]. Though fever with coughing and respiratory problems are the main symptoms of COVID-19 which resemble the primary indications of simple flu, but later complexity arises resulting very low oxygen saturation in body. Primarily it was apprehended that Corona viruses are transmitted in two modes: droplets during coughing or sneezing and contact transmission (contaminated hands). Disinfectants like hydrogen peroxide and sodium hypochlorite, etc. [4] can destroy this virus. It is difficult to find out the spreading mechanism of COVID-19 because it can be silently carried by and transmitted from people without showing any

* Corresponding author.
E-mail addresses: pranati.rakshit@jiscollege.ac.in (P. Rakshit), soumen.kumar@tcs.com (S. Kumar), snoei@istu.edu, noiagdams@susu.ru (S. Noeiaghdam), unai.fernandez@ehu.eus (U. Fernandez-Gamiz), maltenji@kku.edu.sa (M. Altanji), shyamsundar.santra@jiscollege.ac.in, shyam01.math@gmail.com (S.S. Santra).

https://doi.org/10.1016/j.rinp.2022.105855
Received 1 April 2022; Received in revised form 22 July 2022; Accepted 22 July 2022
Available online 27 July 2022

symptoms [5], but this is critical to prevent the pandemic outbreak. The spread of the disease is frightening and a threat to human civilization. Though the administration of vaccination started across several countries, WHO still recommends a guideline like social distancing, sanitization of hands, usage of face masks to avoid it is successive waves. Government of each country has also taken various strategies like complete lockdown, banning international physical transport, and quarantining the suspected persons to control the epidemic progression at its early stages. As an example, countries such as South Korea [6,7] quickly started to trace contacts of infected persons, hence identifying the epicenter of the outbreak, and preventing huge impact.

Compartmental models which divide the total human population into classes according to the state of the disease are the most popular mathematical models [8]. Among them, the classic SEIR model is the most widely adopted for pandemic disease and also characterizes the COVID-19 outbreak [9]. In simple SEIR model, the number of people in different category like infected or recovered or dead are measured on the basis of different parameters of disease dynamics like the number of contacts per unit time, the transmission rate of the disease, the time period of disease incubation, rate of recovery of the sick person to be cured and the rate of death.

We know that due to the lack of proper detection technique, the asymptomatic patients who have mild or no symptom, are not considered in the confirmed case count, and this is the untouched area in basic SEIR model. To estimate the untraceable contacts some methods have been discussed in [10]. In this work, we aim to consider the "Asymptomatic" class as a vital factor in epidemic volume determination and has introduced new compartment in existing model. If the infected identified case is considered the tip of an iceberg, there is a large hidden population carrying the virus without any symptoms due to lack of public health surveillance and effective testing. We have measured the dynamic nature of the different key epidemic parameters which contribute to the spread of COVID-19, in a relatively effective way in different periods of disease outbreak across multiple geographical areas. The effectiveness of latency period, infectivity rate, recovery rate and several factors like human immunity, social awareness had been considered while calibrating the model in different phases of outbreak in India, UK and US.

In Section "Background" we have discussed the background of this present work, Section ‘‘Material and methods’’ describes about the material and method in which the proposed model, model parameter etc. are discussed, the results, outcome analysis have been discussed in Section "Results and Discussion" while Section "Future Work" deals with the future scope, conclusion.

Background

In different countries various mathematical modeling has been proposed and developed to understand the dynamics of Covid 19 spread. The epidemiological behavior of this virus can be evaluated as a complex patterns/system through mathematical models. This statistical data regarding the disease spread can be analyzed and interpreted to a model evolution which will help to predict the disease characteristics, to impose containment, to control and mitigate probable outbreak. The already known parameters about the virus spread like the number of infected count, the number of recovered people and the death count can be fed into the model to refine and calibrate it more accurately. The dynamic equations of mathematical models can provide powerful insights into the dynamics. These methods hold good for the disease where a large number of unknown, uncertain factors contribute to the disease growth.

Related works

Batista et al. [11] proposed susceptible infectious recovered (SIR) model which tried to forecast the number of Corona virus pandemic cases and final size, which seemed to be very accurate considering at the month of February 2020 data, but the current trend has proved the prediction not so effective. Huang Y et al. [12] utilized the SIR model equations and estimated the future infected count due to Corona virus in Japan, South Korea, Italy, and Iran. Incubation period has not been considered while constructing this model. Fotios et al. [13] has adopted techniques which follow simple time series forecasting. They predicts the future with non convergence growth. Leonardo et al. [14], considered the latent period in a compartmental model of SEIR where the infection spreads with changeable containments. The exponential curve data represents a good prediction in deaths and recovered count. Carcione et al. [15] has used the modified SEIR model and varied different initial parameters values to calibrate the model. While analyzing the results, their model is refined with number of dead individuals. They found a range of incubation period and Reproduction number to fit the real-world data and calculate the infection fatality rate (IFR). FaïçalNdaïrou et al. [16] has proposed a new compartmental model considering super–spreaders. Moreover, they consider a death compartment. They have mentioned the "Asymptomatic" class in their model. They have done a good sensitivity analysis of their model and numerical simulation with the case study of Wuhan. Eugene B. Postnikov et al. [17] propose a simple SIR model with predictive estimation considering correlation of epidemic dynamics. Peng et al. [18] has introduced quarantined state to propose a new SEIR model and also done a sensitivity analysis. Mukesh Jakhar et al. [19] has made a state-wise analysis of India. They have made a study and forecast based on the demographical diversity. Also, the phase wise Lockdown population density is considered while predicting the growth of the epidemic. A detailed analysis and comparative study of the different forecasting methodology has been found in [20]. Subhas Khajanchi et al. [21] has introduced mathematical modeling considering different Indian states and they calculated R0 value and other sensitive parameters using the partial rank correlation coefficients techniques.

Motivation of the research

Though all the above works throw light on the respective areas during analysis, but in most of the cases the parameters estimation and prediction is largely based on the statistics of confirmed cases, whereas we felt that in a developing country like India with its variety of population, there are large number of undetected cases which will contribute to future spread significantly. Also, we have represented the model with UK and US cases to ascertain the acceptability across various geographical data. In this study, we have analyzed the number of asymptomatic cases and unreported infectious cases as impacting factors, whereas the number of reported infectious cases and death counts are considered to fit the model parameter with the real-world statistical data. While all the other previous researchers have focused on a particular Geographical area with a specific time span for their model calibration and evaluation, we have simulated the model behavior with respect to different demographic data like UK, US and India in both first wave and second wave of the disease peak with varying parameter values.

Material and methods

At first, we will briefly discuss the properties of basic Susceptible-Exposed-Infected-Removed (SEIR) system with four states of the disease. Susceptible are healthy individuals but have a chance and vulnerable to get infected. Exposed are individuals who have had contact with an infected person but not yet considered as infected. Infected are

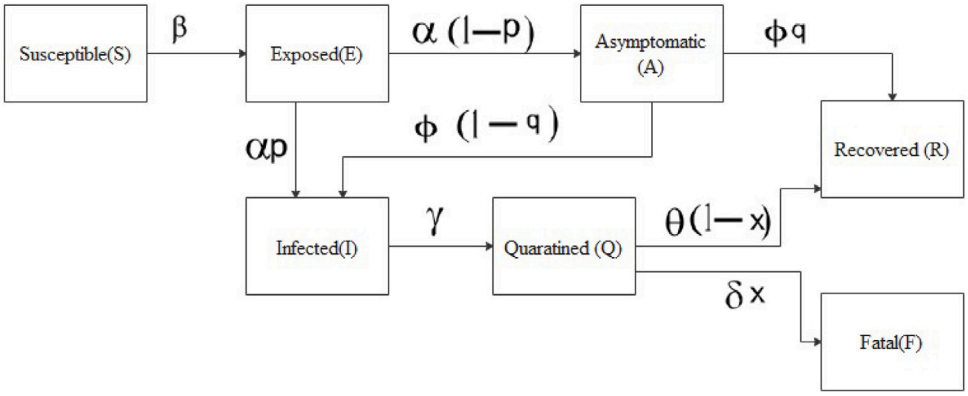

Fig. 1. Proposed SEIAQFR model with seven compartments/classes.

individuals who tested positive for infection and recovered is individuals who are declared cured. We have employed model analysis with existing data of different countries by varying the parameters and initial assumptions and considered the reported fatality number and infected count as reference while establishing the model dynamics. Here we have emphasized on current and predictive trend of the outbreak with real world data along and hence the model calibration becomes complex and flexible to accommodate impacting features.

Proposed model (SEIAQFR)

Now, let us describe our proposed model which considers 7-compartment/classes of the disease and the elaboration of state transformation is depicted in Fig. 7. Susceptible (S): They are the individuals of the total populations (N) who are exposed to the disease spread. Exposed (E): The individuals that have been exposed or came into contact of Infective persons. At this stage person is infected but not yet developed any symptoms, hence not identified as infectious to spread the disease. Disease is in a latent stage. People goes from S to E depending on the number of contacts with Infected (I) individuals and the rate of infection spreading.

Infected (I): The individuals those are revealing symptoms of Covid-19, identified and symptomatically infected. They are also infective, so they can spread the disease(see Fig. 1).

Asymptomatic (A): These groups are infected and infective, they may act as a hidden carrier of the disease, but not identified as they are not showing any disease symptoms and due to lack of rapid tests conducted in mass, they are not tested positive.

Quarantined (Q): When infected people are confirmed with Covid-19 after proper medical test and either hospitalized or home quarantined.

Recovered(R): Persons recovered from epidemic disease, so asymptomatic individuals may get automatically recovered without their disease detection & treatment, whereas Quarantined persons get well after proper treatment.

Fatal (F): This group of people failed to survive after COVID-19 diagnosis.

Model parameters and state transitions

Different model parameters contributing to this model is described below:

β: This denotes average contact frequency i.e., Number of people an infected person can infect/day. This controls the rate of spread, which represents the probability of transmitting disease between a susceptible and infectious individual. In other words, given a population N, among them initially susceptible is S and I is infected people, then infected people start impacting susceptible population at a rate of β to make them exposed COVID-19 virus.

α: Exposed (E) becomes infectious at the rate of α i.e. $(\frac{1}{\alpha})$ is the average time required for an exposed individual between his initial contact of an infectious person and showing the symptoms of COVID-19. Therefore $(\frac{1}{\alpha})$ or α^{-} is the incubation period of the disease. Assuming that among Exposed (E) persons, fraction p develops symptoms and become symptomatically infected (I) whereas (1-p) remains Asymptomatic (A) after the incubation period.

γ: The symptomatic(I) are moved to Quarantined(Q) state either in hospital or at home at a rate of γ so $(\frac{1}{\gamma}$ or $\gamma^{-})$ is the infectious period during which a symptomatic infected person can spread infection. And obviously, after being quarantined he is isolated and cannot spread infection. We assume that all identified symptomatic persons will be quarantined, hence probability is 1.

φ: The asymptomatic people(A) also contribute to Symptomatic Infected (I) group by developing symptoms after $(1/\varphi)$ days i.e., at a rate of φ, otherwise they get automatically recovered from infection without showing any symptoms and hence without undergoing any treatment. The asymptomatic (A) are removed at a rate φ, a fraction q of them going in automatic remission and fraction $(1-q)$ becomes symptomatic (I).

θ: This denotes Recovery Rate after treatment i.e., the Quarantined people(Q) get well at a rate of θ. i.e. $(1/\theta$ or $\theta^{-})$ is the average quarantined period in hospital/home for the recovered person during treatment.

δ: This denotes Mortality/Fatality Rate. Unfortunately, quarantined persons (Q) get deceased at δ rate with probability x, then $(1/\delta$ or $\delta^{-})$ is the average days spent by any patient before death since treatment has been started and he is quarantined. So, quarantined persons (Q) get recovered at a probability of $(1-x)$. Basic Reproduction Number $(R) = \frac{\beta}{\gamma}$ determines the severity of spread at any time and a key factor in this model. We have proposed the virus transmission by the following nonlinear ordinary differential equations (ODE) as SEIAQFR model:

$$\frac{dS}{dt} = -\frac{\beta S(t)I(t)}{N},$$

$$\frac{dE}{dt} = \frac{\beta S(t)I(t)}{N} - \alpha E(t),$$

$$\frac{dA}{dt} = \alpha(1-p)E(t) - \Phi A(t),$$

$$\frac{dI}{dt} = \alpha pE(t) + \Phi(1-q)A(t) - \gamma I(t),$$

$$\frac{dQ}{dt} = \gamma I(t) - \theta(1-x)Q(t) - \delta xQ(t),$$

$$\frac{dR}{dt} = \Phi qA(t) + \theta(1-x)Q(t),$$

$$\frac{dF}{dt} = \delta xQ(t).$$

In this study, our model is used with the following assumptions:

- Constant population numbers (an equal number of deaths balances the number of births)
- Homogeneous population which means each individual has the same opportunity to make contact with other individuals.
- The spread of the corona virus only occurs between human to human.
- Individuals affected by COVID-19 can either recover or dies.
- Individuals who have recovered cannot be infected again.
- Infection can only spread when the person is in infectious stage, all other possibilities like Quarantined patient in Hospital or home cannot spread infection, so they can either recover or die.

In this study, we will analyze the number of asymptomatic cases and unreported infectious cases, as well as the number of reported infectious cases to fit the model parameter with the real-world data. The model building and simulation has been developed in python.

Parameter estimation

According to present scenario, the number of Deaths (F) and the number of Recovered (R) from Quarantined (from hospitalized or home) and the tests positive (Infected-I) cases are available in Government published statistics. Since the values of the above parameters vary on several time dependent factors, we have carefully estimated them according to the demographic features, periods of infection spread, socio-economic structure and public health. The initial values of Infected, Recovered and Dead (I_0, R_0, F_0) are available from the dataset of respective population, whereas the (E_0, Q_0, A_0) are calculated according to the initial spread of disease in that region and also dependent on the time reference of the model. In this optimization problem of determining the correct coefficient values of all parameters are very critical, either over fitting or under estimation happen due very small variation of time dependent model dynamics. To fix this, we firstly determine some parameter values from the previous analysis of disease dynamics and then fit others to achieve the best possible results. During the analysis of different data of the literatures [15,19,21] we decided to simulate our model with data of India during its first and second surge, as well as Country with less huge population like US and UK where there was an evident surge of Covid-19 with high infection rate . Though there are a lot of other factors like Lockdown period, social distancing awareness, preventive measures of face mask and hand washing, and the UnLock-1 have impacts on the time series data to fit the model, but the analysis from medical bulletin [22] and discussion of health professionals put the actual picture of the disease spread. Incubation Period (α^-) and Infectious period (γ^-) are two important parameters to determine the disease spreading and they vary in different country in different time phase of their pandemic outbreak. Normally there is a linear proportional relationship i.e., $\alpha^- \propto \alpha^-$. Again, when the severity of disease spreading is faster i.e., the R_0 is increasing to higher value, and then α^- and γ^- tend to the lower side which helps the disease to spread more sharply. The α^- has gained a range of 5–20 days depending upon the locality and infection peak, so on average, it takes 10–15 days for an infectious individual to get symptoms and become (I) from (E) state.

Now according to the data from medical research all over the world, the COVID-19 virus has a typical contagious period of around 5–7 days after infecting a person, so γ^- is considered. This γ^- value holds good for US, whereas UK have a range of 11–20 and India have 5–16 days depending upon the infecting strain of the disease and the contamination velocity. These parameters (α^- and γ^-) are getting changed with proportional to every changing virus characteristic and important severity determinant. The other parameters like δ^- (the average day spent before death), φ^- (the average day to get recovered from asymptomatic condition) and θ^- (the average day to get recovered from after proper treatment) are dependent on the patient's immunity power and medical facility of that geography. So, in our model optimization, these three parameters values are kept almost constant through the analysis period of Peak or Off-Peak period for a specific region, referenced in Table 1. While fitting the model, observation period for UK, US and India is considered at the peak outbreak of the disease at that Geography: September 2020–February 2021 i.e. 180 days for UK, March 2020–May 2020 i.e., 90 days for US and for India we have considered two periods May 2020–Sep 2020(Corona first wave), March 2021–May 2021(Corona second wave). Also, these spans are divided into shorter time span to fit the model with more accuracy and varying parameter values. For Initial values of ODE equations: S_0 is the total population, whereas I_0, R_0 and F_0 are taken from the actual data [23] and E_0, A_0 and Q_0 is assumed to be in initial values for that region and taken as a fixed percentage of total population according to the population density. While fitting the curve in smaller span or during the Covid-19 s wave, the initial parameters are considered as corresponding dataset synchronization with outcome of the previous time span in our Model e.g., the May month Curve fitting starts with the data values generated as a result of April month curve fit.

Table 1
Statistics for SEIAQFR model parameter values to best fit for current actual data of US, UK and India in different phases of disease spread.

Country	Span	γ^-	α^-	R	δ^-	Φ^-	θ^-	p	q	x
UK	Sep2020–Feb2021	20	12.5	2.7	18	16	12	0.6	0.6	0.009
UK	Sep2020–Dec2020	22	21	2.9	18	16	12	0.6	0.55	0.0095
UK	Dec2020–Feb2021	11.6	14	3	18	14	12	0.75	0.74	0.007
US	March2020–May2020	7	6	2.8	14	10	14	0.68	0.65	0.032
US	March20–April2020	5	5	3.45	12	10	10	0.77	0.65	0.045
US	April2020–May2020	7	7	3.45	12	10	10	0.77	0.65	0.016
India	March 2021	12	11.5	2.59	14	9.35	14	0.56	0.61	0.0003
India	April 2021	11.5	12	3.65	12	9	14	0.56	0.6	0.0045
India	May 2021	16	10	3.4	10	9.2	14	0.56	0.61	0.0034
India	May2020–Sep2020	4.5	5	1.76	18	15	10	0.75	0.8	0.004
India	March2021–May2021	11.5	13.5	2.6	14	14	12	0.52	0.57	0.0015

While fitting the model, several simulations result the value of p, q and R, hence β. The x is varying for different countries which resembles to the actual data of this country. The reproduction rate (R) plays a vital role in reaching the peak, while the large value reflects a sharp rise and pandemic outbreak, but the smaller value represents the slow spread, but for a longer duration. For India, our analysis period covers the Lock down period in a controlled environment and as well as the Unlock of Phase 1 nationwide. During the initial period of Lockdown and the Virus spread, as the medical diagnosis and tests results were doubtful in India whereas for US and UK, we were confident about the reported outcome, so periods were chosen accordingly. Also the second wave is considered while predicting the India data from March 2021 to May 2021. Also, we have taken a longer period to fit our model and compare the result with actual data as compared to other available model under research. We have plotted the curves to fit our model with the actual death and infected number as per the statistics [22] and the other curves are derived.

Results and discussion

Considering the existing data of the Infected and Death [23], the following Table 1 and corresponding curves (Figs. 2–7) are obtained for US, UK and India. As mentioned above, the model has been calibrated on the basis of the total number of infected and casualties for a certain time period. In this model, Reproduction Rate (R) acts as threshold to determine whether the infection disease spread is coming out as an epidemic or going to die down. Theoretically, at initial the infected count increases exponentially and at a peak the rate of change of infected population is zero i.e., $\frac{dI(t)}{dt} = 0$ and the curve start declining, thus when $t \to \infty$, S and I $\to \infty$; hence the prediction of t is important. But the hidden disease spread by asymptomatic carrier (A) and sudden surge of infection cause the increase in R value and thus impacting the curve of disease rise and die down. Moreover, the Vaccination process has also accelerated the disease controlling timeline.

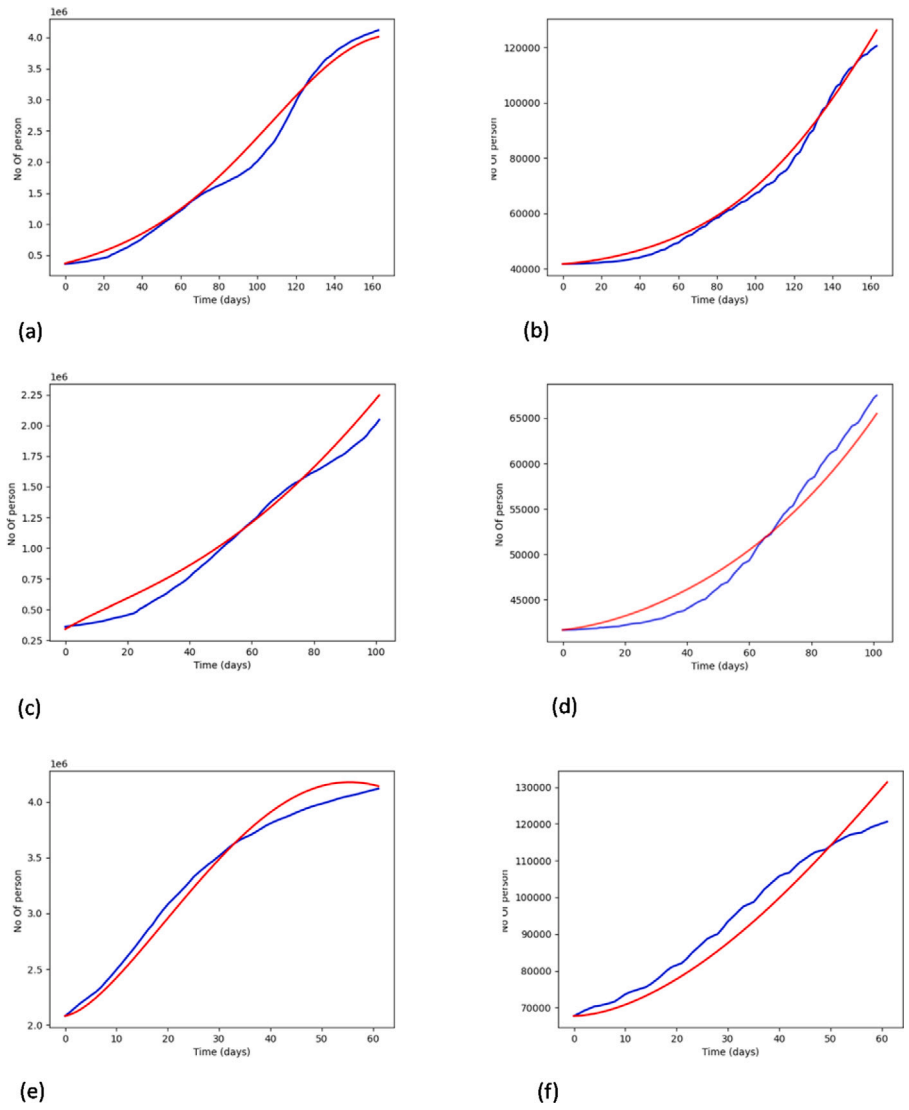

Fig. 2. UK: Actual and Model predicted Curve value of Infected and Death count in different time phases of disease (a) Infected Count: period: Sep 2020–Feb 2021 (b) Death Count: period: Sep 2020–Feb 2021 (c) Infected Count: period: Sep 2020–Dec 2020 (d) Death Count: period: Sep 2020–Dec 2020 (e) Infected Count: period: Dec 2020–Feb 2021 (f) Death Count: period: Dec 2020–Feb 2021. Red line: Predictid value, Blue line: Actual value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

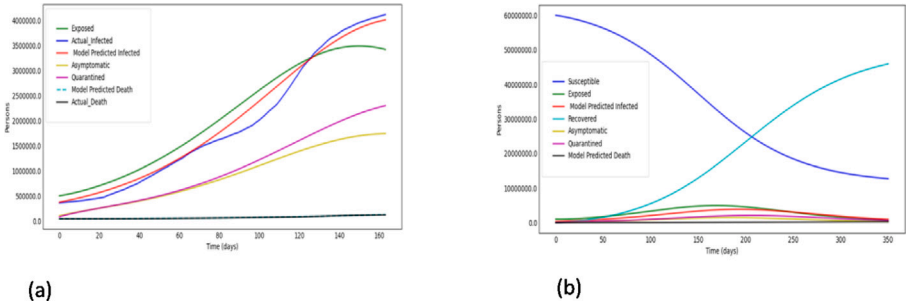

Fig. 3. UK: (a) Model SEIAQFR predicted parameters trajectories during peak period Sep 2020 - Feb 2021 (b) Model SEIAQFR Predicted Parameters value for pandemic's predicted lifespan.

Interpretation of existing data

With the following parameters, we have implemented the proposed model using python. Here we have analyzed on the data of UK, US and India with their response to COVID 19. As mentioned earlier, the actual infected and death number are compared with the model predicted number with varying parameter values. Different countries actual Death and Infected count is compared with the model prediction for phases of the disease outbreak.

UK: The total Population of this is 68 000 000 and the actual data considered to fit with our model is the peak span of the disease spread September 2020–February 2021 (160 days). Fig. 2.(a) and (b)

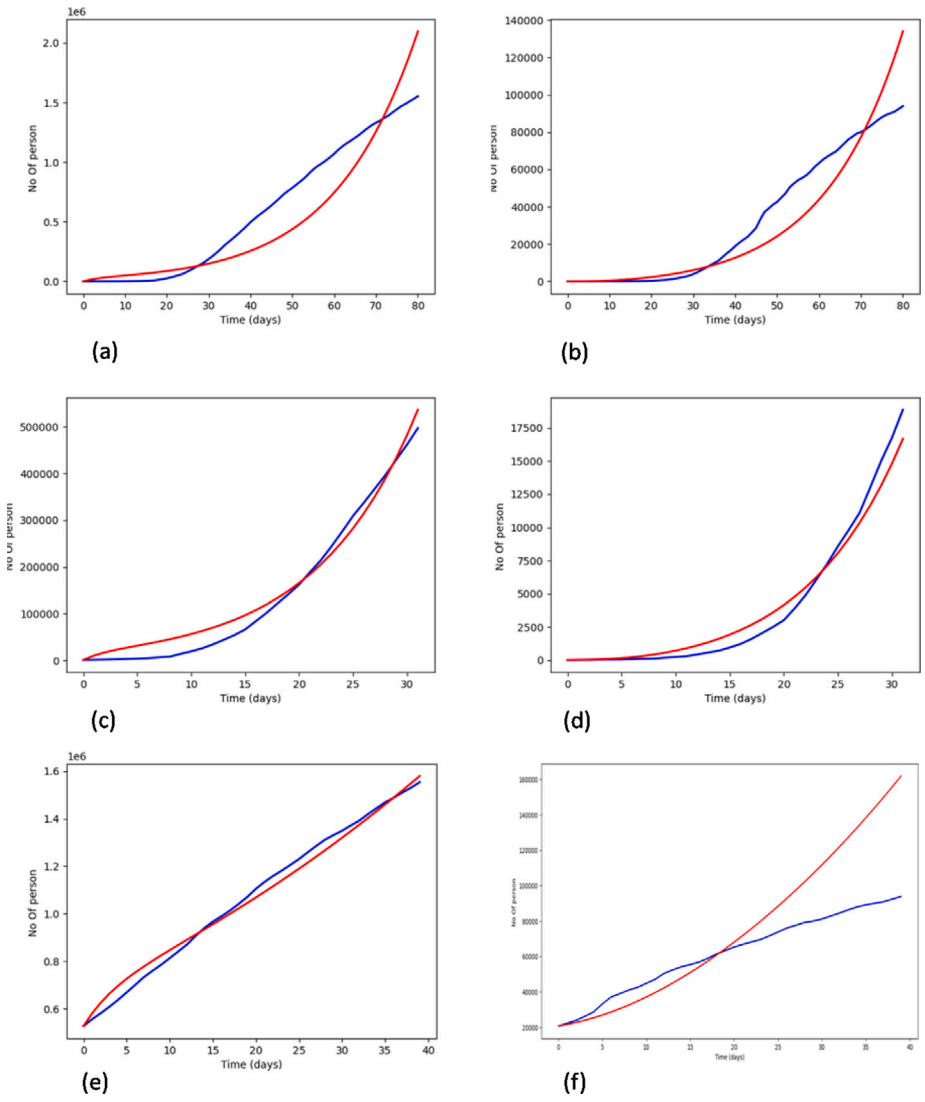

Fig. 4. US Actual and Model predicted Curve value of Infected and Death count in different time phases of disease (a) Infected Count: period: Mar 2020–May 2020 (b) Death Count: period: Mar 2020–May 2020 (c) Infected Count: period: Mar 2020–Apr 2020 (d) Death Count: period: Mar 2020–Apr 2020 (e) Infected Count: period: Apr 2020–May 2020 (f) Death Count: period: Apr 2020–May 2020.

has matched the model prediction with actual data for the 160 days for the death and infected count, whereas the (c), (d) & (e), (f) analyzes the pattern with small span of interval by dividing the 160 days into smaller span of 100 and 60 days. When the disease was in its early stage of spreading, the incubation and infectious period was longer (20–22 days), but in 2nd half they gets shortened with an increased R value, 3 compared to previous 2.7.

The value of δ^-, φ^- and θ^- are almost kept constant to reflect that infrastructure and human immunity are responding uniformly through the whole period. As the model is calibrated with the initial assumptions of 1 000 000 as exposed, 80 000 asymptomatic and 70 000 quarantined, and the infected, recovered and death count are 340 546, 1847 and 41 697 respectively from the available data. While the 1st time period (Sep 20 to Dec 20) uses the above said initial assumption, the 2-nd period is plotted against the outcome of the 1st period . In Fig. 3.(a) and (b) depicts the curves related to the all types of affected counts for the period of 160 days and the total span(approx 350 days) of pandemic end respectively. Fig. 3. proves that due to the best medical system in the world in UK, though the exposed and infected curve is higher than the asymptomatic and quarantine curve, but the spread is controlled.

US: The same analysis of our model fit with respect to prediction of death and infected is done during Peak infectious period of US. Here the total span is taken from Mar 2020–May 2020 of 80 days and then smaller time period of March 2020–April 2020 and April 2020–May 2020 is considered. The results are captured in Fig. 4.(a)–(f). The population is 33 000 000 and initial values of E_0, A_0 and Q_0 are 900 000, 700 000 and 300 000... The shorter period curves for Fig. 4.(c), (d) and Fig. 4.(e), (f) shows better results compared to the larger period curve of Fig. 4.(a) and (b). Analytically, life saving treatment significantly controls the death rate compared to prediction in Fig. 4.(f) while the infection curve Fig. 4.(e) is plotted with same parameter calibration. Fig. 5(a) and (b) provides a probable indication of end of this pandemic, the higher R and smaller γ^- and α^- in US has shortened the pandemic lifespan to 250 days compared to UK.

India: India being a developing country and having a population of 1 300 000 000 showed a good controlling over first phase, but second wave left a bad impact. In both the cases, asymptomatic cases play a vital role as to simulate the model performance. Fig. 6(a), (b) and (c) have simulated the pandemic first wave time period of 120 days (May 2020 to Sep 2020). Initial values of E_0, A_0 and Q_0 are 100 000. 60 000 and 30 000 and the trend indicate a large number of quarantined after the end of first wave.

The second phase outbreak in India in Fig. 7(a)–(h) has a total period of 72 days (10th March 2021 to 20th May 2021) and also their month wise individual curve.

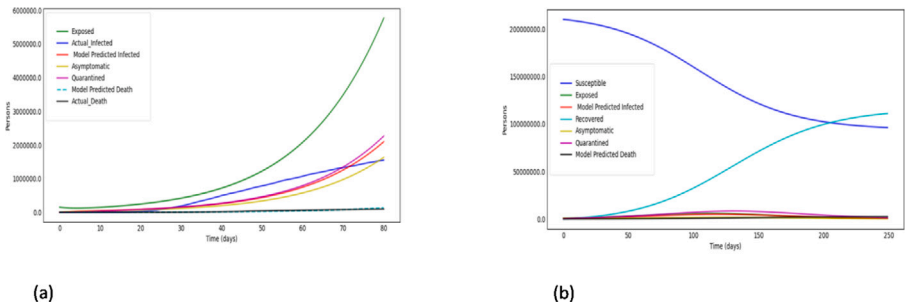

Fig. 5. US (a) Model SEIAQFR predicted parameters trajectories during peak period Mar 2020–May 2020 (b) Model SEIAQFR Predicted Parameters value for pandemic's predicted lifespan.

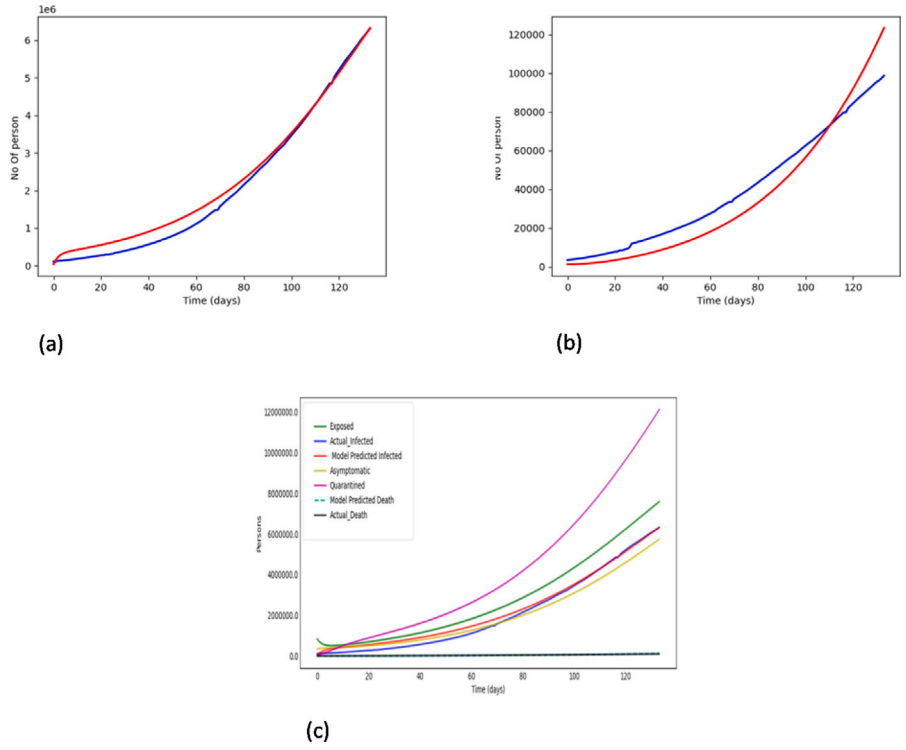

Fig. 6. India Actual and Model predicted Curve value of Infected and Death count in different time phases of disease during its outbreak in first wave (a) Infected Count: period: May 2020–Sep 2020 (b) Death Count: period: May 2020–Sep 2020 (c) Model SEIAQFR Predicted Parameters value for pandemic's predicted lifespan during first wave in India.

Fig. 6(a), (b) shows an improvement of medical treatment in the last span of September 2020, so that the expected death rate supersedes the actual value. In Fig. 6(c), the quarantined count is greatest and obviously a large number of asymptomatic has a contribution in that, Fig. 7(a)–(h) shows that the individual curve for individual month performs better than the whole period of second wave (March 21 to May 21). As an example, In May first week, we have seen that the Hospital stay becomes longer due to infection(increasing γ) and the demise of people in shorter period of time(reducing δ) due to delta variant attack, so adjusting the parameters the R is found to be smaller than previous month, but with a increased mortality rate. The overall R score is impacted due to the health system betterment and the infection & death rate decreases.

Analysis and discussion

Analyzing Figs. 2–7 and Table 1, we have the following observation while fitting the infected and death count with the actual value:

- When the Reproduction Factor(R) is changed from smaller to higher value, the rising of infection curve is sharper and the curves get flattened with the lower values of R. The value of the reproduction factor (R) and the incubation period (α^-) determines the spread of epidemics, R affects the intensity of spread and α^- affects the speed of transmission directly. We can observe from Table 1 and Figs. 2–7 that high R and less α^- values makes an higher infected rate in US in the first phase of COVID and UK remained in best position with high α^- value and less R value, while India stands in between that means as early as the exposed persons becomes infectious and their contact with other persons are more, then the growth of disease is also higher. The smaller infectious period (γ^-) also slows down the spread.
- Longer δ^- value flattens the death curve whereas the φ and θ also contribute to rise of the curve value. UK and India performs better to combat the disease with respect to mortality, so longer hospital stay, but better survival rate. from q value, The immunity of all countries UK, US and India are responding almost in similar proportion.
- p, q monitors the infected count curve, while changing the x value has a direct impact on the death curve. From the p value variation, the asymptomatic cases are following the pattern of India > UK > US which reflects the same relation with the countries population

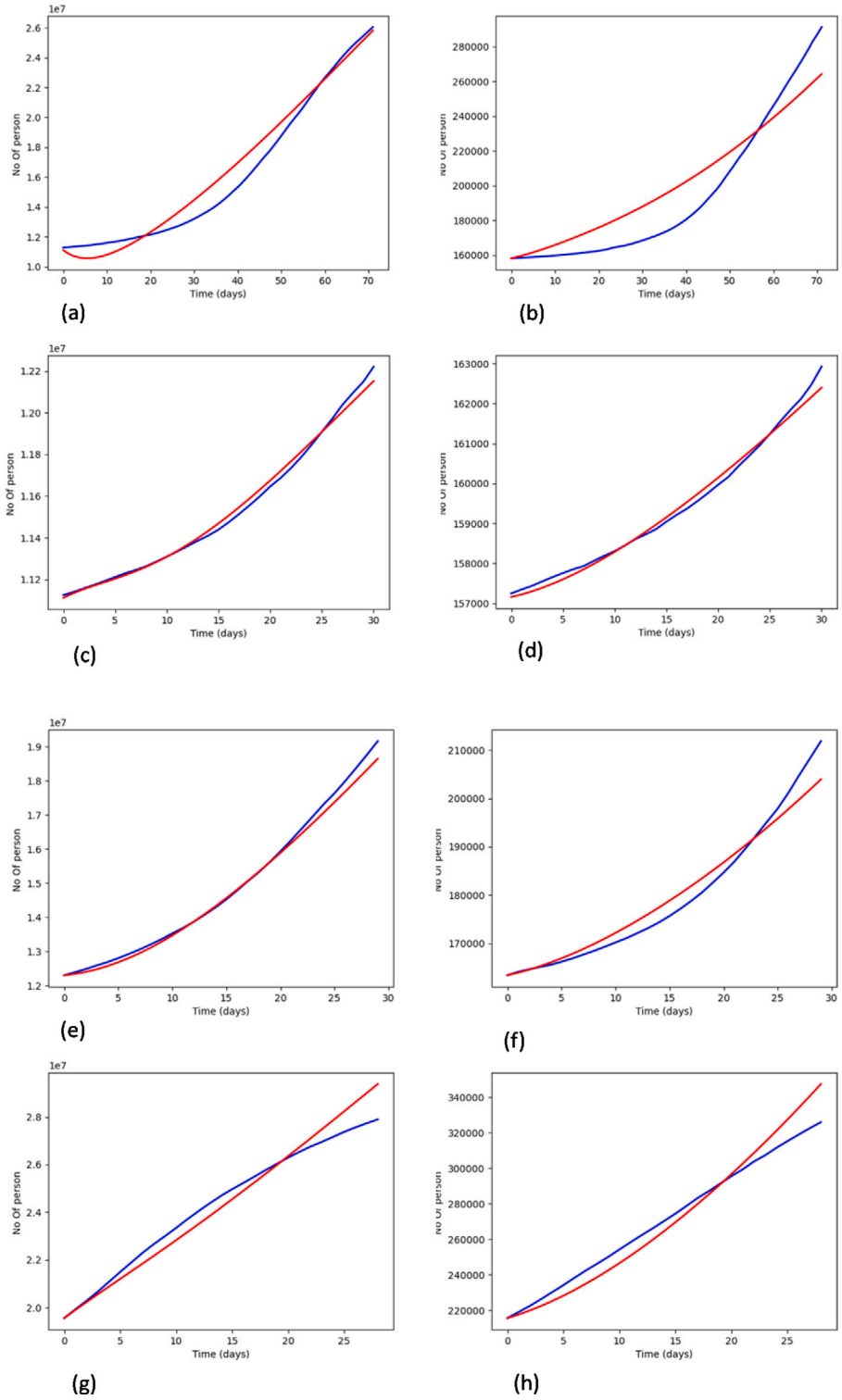

Fig. 7. India Actual and Model predicted Curve value of Infected and Death count in different time phases of disease during its outbreak in second wave in 2021 (a) Infected Count: period: Mar 2021–May 2021 (b) Death Count: period: Mar 2021–May 2021 (c) Infected Count: period: Mar 2021 (d) Death Count: period: Mar 2021 (e) Infected Count: period: Apr 2021 (f) Death Count: period: Apr 2021 (g) Infected Count: period: May 2021 (h) Death Count: period: May 2021.

density. So Asymptomatic case $\propto$ Population Density, which is very true for such contagious disease.

- In our model the asymptomatic cases (p value derivation) are getting evaluated to be 30%–40%, which is inline with different statistical analysis in India [24].
- As in this model, quarantined means the patient is either kept in home isolation or hospitalized for treatment, so from all the epidemic forecasting, it is evident that the number of quarantined persons are greater than the infected persons at any time, because of the slow rate of getting recovered or death compared to the infection rate. The observed cases in different states where the number of home quarantined are getting increased due to shortage of hospital bed, as a result the serious persons are dying without proper arrangement of ventilation & ICCU bed. In short,

there is a positive correlation between economic development and mortality rate in India as compared to US and UK. The hospital beds and infected correlation is established in [25].
- All the curves in Fig. 5, clearly depict the fact of very high recovery number as compared to the data that have been published [23, 24]. This is because of the reason that there is a large number of asymptomatic cases who are getting recovered automatically in shorter time compared to a person who are getting quarantined and treated.

If we denote from Exposed state, asymptomatic cases time to recover as ta and infected person time to recover during proper treatment & quarantine as ti, then

$$t_a = incubation\ period + infectious\ period + quarantined\ period = \alpha^- + \gamma^- + \theta^-$$

and

$$t_i = incubation\ period + time\ for\ automatic\ recovery\ of\ patient = \alpha^- + \varphi^-$$

From both the study of all countries we find that $t_a > t_i$. For this time difference, at any point of time (t) in our derived model, almost 80% asymptomatic cases gets recovered in shorter period of time. So, we can say that the actually infected persons are much higher than the confirmed cases, because of the presence of many asymptotic individuals or the less number of testings.

- Lockdown has a positive impact on the growth of Corona. From all figure, the initial fatality rate is at par to our model prediction whereas the effective social awareness and Lockdown restriction along with expertise to handle Covid patients have enabled a 10% difference between predicted and actual death count of each state.
- US, UK, India's first wave basic reproduction number is on the lower side as compared to Brazil [25] and Italy [26], and this may be because of the strict lockdown and immediate isolation of the infected, but the second wave in India has shown a rise in R value.
- As these data are dynamic ones affecting the curves in trends, therefore errors in positive case detection, reporting or other large uncertainties associated with these values may dramatically alter calibration parameters and the conclusive inference in terms of number of infected and associated mortality.
- The traditional concept of SIR model with a specific velocity of infection and recovery rate has been enhanced with a varying nature of disease with a correlation among different factors. The varying values of α, γ, R establish the dynamic characteristic of SIR model over time period and demographic orientation. It reflects the strength of medical infrastructure of the country, human immune system of specific ethnic, likewise the US, UK and India has different varying parameter values to have a best fit, so these parameters are not constants for any epidemic disease evolution.

To measure the performance of our model we have calculated some relevant parameter like the MAPE and R-Squared metrics and the results are shown in Table 2.

MAPE (mean absolute percentage error) measures model accuracy as a percentage, and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values.

$$MAPE = \frac{1}{N}\sum_{x=1}^{N}\frac{|y_x - \hat{y}|}{y_x} \times 100\%.$$

R-squared (Coefficient of determination) represents the coefficient of how well the values fit compared to the original values. The value from 0 to 1 interpreted as percentages. The higher the value is, the better the model is.

The above metrics can be expressed,

$$R^2 = 1 - \frac{\sum_{x=1}^{N}(y_x - \hat{y})^2}{\sum_{x=1}^{N}(y_x - \bar{y})^2},$$

where

$$y_x = Actual\ value,$$

$$\hat{y} = Predicted\ value,$$

$$\bar{y} = Mean\ value,$$

The value of R-square measure and corresponding MAPE reflects a significant performance of our Model prediction. Minimizing the Mean Absolute Error and Maximizing the R-square value gives an optimum model parameter to best fit for the actual live data.

- In case Studying the Infected count of the actual and the model predicted value we found an accuracy (R-square score) of $> .98$ and error (MAPE) of $< 5\%$. for most of the periods. Whereas for the Death count of the actual and the model predicted value accuracy (R-square score) and error (MAPE) are average 0.95 and 10% respectively.
- In India, the model performs best in 2-nd wave with an 99% accuracy for infected and death (R-Square $\approx$ 0.99 and MAPE $\approx$ 1%).
- During analysis, it is observed that when the change of p value is 0.11%, it is affecting the R-Square value with a variation of 60%.
- The R-Square and the MAPE value has a greater performance in the shorter period of time of model fit compared to longer time span.
- While making the best fit Curve, often to achieve optimized MAPE value, the accuracy of R-Square is getting sacrificed.

Future work

Model fit indicates that the pandemic end of UK and US after 350 and 250 days time period respectively, but the changing behaviour of virus variant and endurance of acquired resistance to virus is a challenging goal today and alongside the recurrence of the infection despite vaccination is affecting the prediction in long term. The immunization process and the efficacy of that against a particular strain of virus have proposed a second thought in our mind to improve the model considering different fuelling factors as well as prevention techniques. Unlike traditional statistical analysis, our model properly tries to assess and incorporate impacts of hidden asymptomatic and infectious cases on the entire procedure of epidemic. In our further study, we are aiming to incorporate the following points like considering a certain quantity of population with strict confinement, different age group, sex and people with co morbidity, introducing the preventive measures taken by different administration, the viral load of the infections [27], the strictness of lockdown & the migration of people among different places. Though we have considered the latent period of asymptomatic person as constant, it is varying nature is another major contributing factor of the virus spread which should be taken care of.

Conclusions

The present analysis of the proposed model indicates a very high number of peak infections in India by comparing the existing data to the model simulation. However, the parameters chosen for the model calibration are not unique, as different choices of parameter values with the same good fit curve, data can result to very different model predictions. For example, The reported number of deceased people could possibly be underestimated due to undeclared cases and can also be over estimated due to the comorbidity factors of elderly where the patients could have survived from COVID-19, but other illness like existing lungs-kidney problem, diabetes has accelerated the fatality rate. In a word, the age group, weather conditions of different states (as the humid and rainy season in India may put a break in the outbreak of the epidemic) can be considered as other impacting factors.

Table 2
R-Square and MAPE value of SEIAQFR model evaluation in different phases of disease in US, UK and India.

Country	Population	Span	R-square (Infected)	R-square (Death)	MAPE (Infected)	MAPE (Death)
UK	68 000 000	Sep2020–Feb2021	0.98147292	0.99171328	8.95695898	3.08919093
UK	68 000 000	Sep2020–Dec2020	0.97551659	0.96716050	9.81254144	2.52670454
UK	68 000 000	Dec2020–Feb2021	0.98173659	0.93222153	2.61654960	4.14764188
US	330 000 000	March2020–May2020	0.91527276	0.87960577	131.35154014	44.88586172
US	330 000 000	March20–April2020	0.97167661	0.97287443	215.70701729	76.92727340
US	330 000 000	April2020–May2020	0.98648816	0.67740774	3.13226779	17.01848430
India	1 300 000 000	March 2021	0.99457064	0.98999972	0.15241638	0.08202740
India	1 300 000 000	April 2021	0.99653035	0.96073521	0.61232281	1.16277449
India	1 300 000 000	May 2021	0.95044058	0.94106317	1.74359144	2.34275027
India	1 300 000 000	May2020–Sep2020	0.98241572	0.90590407	40.10811585	33.09705949
India	1 300 000 000	March2021–May2021	0.96764717	0.88452368	5.03041917	5.03041917

Moreover, with governmental steps like Lockdowns and quarantines will reduce the contact between people drastically, consequently the impact of epidemic and the peak would reduce significantly. This model is purely based on available patient data and mathematical prediction, but considering the mutation of the virus gene in different environment is more complex biological and epidemiological problem. In the present acute crisis of delta strain and verge of community transmission, it is likely that researchers from various disciplines will collaborately work together with health care professionals, politicians, and administrators to combat against future pandemic.

CRediT authorship contribution statement

Pranati Rakshit: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Soumen Kumar:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing – original draft, Writing – review & editing. **Samad Noeiaghdam:** Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing. **Unai Fernandez-Gamiz:** Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Writing – review & editing. **Mohamed Altanji:** Conceptualization, Data curation, Formal analysis, Software, Supervision, Validation, Writing – review & editing. **Shyam Sundar Santra:** Conceptualization, Formal analysis, Methodology, Resources, Software, Validation, Visualization, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Funding

The work of U.F.-G. has been supported by the government of the Basque Country for the ELKARTEK21/10 KK–2021/00014 and ELKARTEK20/78 KK–2020/00114 research programs, respectively and the work of M. Altangi has been supported by the Deanship of Scientific Research at King Khalid University through the Research Group program with Grant Number R.G.P2/150/43. All authors approved the final version of the manuscript.

References

[1] WHO. Statement on the second meeting of the international health regulations (2005)emergency committee regarding the outbreak of novel coronavirus (2019-nCoV). 2020, https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov). Accessed 10 2021.

[2] WHO coronavirus (COVID-19) dashboard(2020). 2021, https://covid19.who.int Accessed 10 2021.

[3] COVID-19 coronavirus pandemic. 2021, https://www.worldometers.info, Accessed March 26 2021.

[4] Kampf G, et al. Persistence of coronaviruses on inanimate surfaces and its inactivation with biocidal agents. J Hosp Infect 2020;104(3):246–51.

[5] Rothe C, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany, New England. J Med 2020;382(10).

[6] Fang Y, et al. Transmission dynamics of the covid-19 outbreak and effectiveness of government interventions: A data-driven analysis. J Med Virol 2020;92(6):645–59.

[7] Nesteruk I. Nesteruk I estimations of the coronavirus epidemic dynamics in south korea with the use of SIR model. In: Investigation of dynamics of coronavirus spreading researchgate. 2020, http://dx.doi.org/10.13140/RG.2.2.15489.40807.

[8] Hethcote HW. The mathematics of infectious diseases. SIAM Rev 2000;42:599–653.

[9] Labadin J, Hong BH. Transmission dynamics of 2019-nCoV in Malaysia. 2020, http://dx.doi.org/10.1101/2020.02.07.20021188, medRxiv.

[10] Nishiura H, et al. The extent of transmission of novel coronavirus in Wuhan. China J Clin Med 2020;9:330.

[11] Batista M. Batista M Estimation of the final size of the coronavirus epidemic by the logistic model. 2020.

[12] Huang Y, et al. Epidemic situation and forecasting of covid-19 in and outside china, [preprint], vol. 16. Bull World Health Organ. E-pub; 2020, http://dx.doi.org/10.2471/BLT.20.255158.

[13] Petropoulos F, et al. Forecasting the novel coronavirus COVID-19. PLOS ONE 2020;15(3).

[14] Lopez L, et al. A modified SEIR model to predict the covid-19 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics. Results Phys 2021;21.

[15] Carcione JM. A simulation of a COVID-19 epidemic based on a deterministic SEIR model. 2020, http://dx.doi.org/10.1101/2020.04.20.20072272, medRxiv preprint.

[16] Ndaïrou F. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. Chaos Solitons Fractals 2020;135:109846, 2020.

[17] Postnikov EB. Estimation of COVID-19 dynamics on a back-of-envelope: Does the simplest SIR model provide quantitative parameters and predictions? Chaos Solitons Fractals 2020;135:109841.

[18] Peng L, et al. Epidemic analysis of COVID-19 in China by dynamical modeling. 2020, http://dx.doi.org/10.1101/2020.02.16.20023465, 2020.

[19] Jakhar M. COVID-19 epidemic forecast in different states of India using SIR model. 2020, http://dx.doi.org/10.1101/2020.05.14.20101725.

[20] hinde GR, et al. Forecasting models for coronavirus disease (COVID-19): A surveyof the state-of-the-art. SN Comput Sci 2020.

[21] Khajanchi S, et al. Mathematical modeling of the COVID-19 pandemic with intervention strategies. Results Phys 2021;25:104285. http://dx.doi.org/10.1016/j.rinp.2021.104285.

[22] Ghosh A. How far can someone without symptoms spread the Covid-19 infection. 2021, https://indianexpress.com/article/explained/how-far-can-someone-without-symptoms-spread-the-covid-19-infection-6376380/ Accessed 20 Aug.

[23] SRK. Novel corona virus 2019 dataset.day level information on covid-19 affected cases. 2021, https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset Accessed 25 Aug.

[24] IANS. Almost 80% of COVID cases in India asymptomatic: Union health minister. 2021, https://timesofindia.indiatimes.com/life-style/health-fitness/health-news/almost-80-of-covid-cases-in-india-asymptomatic-union-health-minister/articleshow/76069643.cms Accessed 20 Aug.

[25] Ghosh A. Asymptomatic transmission of Covid-19: Why it matters, where evidence stands. 2021, https://indianexpress.com/article/explained/asymptomatic-cases-covid-19-coronavirus-6454677/ Accessed 20 Aug.

[26] Lyra1 W, et al. COVID-19 pandemics modeling with SEIR(+CAQH), social distancing, and age stratification. In: The effect of vertical confinement and release in brazil. 2020, http://dx.doi.org/10.1101/2020.04.09.20060053.

[27] Pedersen MG, Meneghini M. Quantifying undetected COVID-19 cases and effects of containment measures in Italy: Predicting phase 2 dynamics. 2020, http://dx.doi.org/10.13140/RG.2.2.11753.85600.