



# Enhanced Gaussian process regression-based forecasting model for COVID-19 outbreak and significance of IoT for its detection

Shwet Ketu<sup>1</sup> · Pramod Kumar Mishra<sup>1</sup>

Published online: 28 September 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Virus based epidemic is one of the speedy and widely spread infectious disease which can affect the economy of the country as well as it is life-threatening too. So, there is a need to forecast the epidemic lifespan, which can help us in taking preventive measures and remedial action on time. These preventive measures and corrective action may consist of closing schools, closing malls, closing theaters, sealing of borders, suspension of public services, and suspension of traveling. Resuming such restrictions is depends upon the outbreak momentum and its decay rate. The accurate forecasting of the epidemic lifespan is one of the enormously essential and challenging tasks. It is a challenging task because the lack of knowledge about the novel virus-based diseases and its consequences with complicated societal-governmental factors can influence the widespread of this newly born disease. At this stage, any forecasting can play a vital role, and it will be reliable too. As we know, the novel virus-based diseases are in a growing phase, and we also do not have real-time data samples. Thus, the biggest challenge is to find out the machine learning-based best forecasting model, which could offer better forecasting with the limited training samples. In this paper, the Multi-Task Gaussian Process (MTGP) regression model with enhanced predictions of novel coronavirus (COVID-19) outbreak is proposed. The purpose of the proposed MTGP regression model is to predict the COVID-19 outbreak worldwide. It will help the countries in planning their preventive measures to reduce the overall impact of the speedy and widely spread infectious disease. The result of the proposed model has been compared with the other prediction model to find out its suitability and correctness. In subsequent analysis, the significance of IoT based devices in COVID-19 detection and prevention has been discussed.

**Keywords** IoT · Machine learning · Novel coronavirus (COVID-19) · HealthCare · Virus

## 1 Introduction

In the last two decades, we have seen various epidemics conditions. These conditions were started from SARS in 2002 and followed by SWINE FLU in 2009, after that EBOLA in 2013, then MARS in 2014, and now COVID-19 in 2019 [1–5]. These epidemic encounters can cause severe human and economic losses. If we talk about the latest COVID-19 disease, which came into the picture in late December of 2019 from china. The first case of this virus came from Wuhan city of China, which was new and never seen before. Initially, it was

known as the Wuhan virus, and after that, it was coded as COVID-19 or novel coronavirus or 2019-nCov [6]. According to the hypothesis, this virus had come from the fish market (probably from the bat) of Wuhan city. Initially, it started infecting the people of Wuhan city and spread in other cities in a short span [7]. In fifteen days, china was brutally affected, and within a month, it becomes a global epidemic. Initially, it was claimed that it is an air-borne disease, but after the proper assessment, the scientist told that it is a touched contagious disease, and it lasts from hours to days in different environments or surfaces. As we know, it is touched contagious disease, so we can say that it can be spread from the human to human interaction or human to surface contact. Due to the rapid spread rate of COVID-19 worldwide, the warning regarding global health emergency had been issued by the World Health Organization (WHO) [8] in the last week of January 2020. In the first week of March 2020, the WHO had declared the COVID-19 as a pandemic disease [9].

The novel coronavirus, or COVID-19, has come from the same virus family from where the Severe Acute Respiratory

---

✉ Shwet Ketu  
shwetiita@gmail.com

Pramod Kumar Mishra  
mishra@bhu.ac.in

<sup>1</sup> Department of Computer Science, Institute of Science, Banaras Hindu University, Varanasi, India

Syndrome (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV) came. It is a new disease for the human that came into the picture at the end of December 2019 [10]. Before 2019 it was not identified in humans, so we did not know much about it. The COVID-19 is a zoonotic disease, which means it was transmitted from animal to people. This hypothesis is correct because some known coronaviruses are circulated from animals but have not affected humans yet. From the studies, we found that the SARS-CoV and MERS-CoV were transmitted from the animals. The SARS-CoV came from the civet cats, and MERS-CoV came from camels [1–5]. These diseases were born from the mutation process, which means they had transformed their properties before going into the human body from the animals.

In Table 1, some virus-based epidemic diseases have been shown [1–5]. This table consists of, disease name, year of its occurrence, total cases, and total death caused by the virus, totally recovered people, and the number of countries where these viruses were caught.

In the novel coronavirus or COVID-19, the patient initially feels usual flu-like symptoms i.e., shortness of breath, cough, fever, and breathing difficulties, which can become more severe in a couple of days. It targets the human respiratory system, or we can say that it is a respiratory system-based disease. In severe cases, it can cause acute respiratory syndrome, pneumonia, kidney failure, other severe respiratory-related diseases, and even death too. To protect ourselves from COVID-19 and to stop its spread, we have to take precautions such as cover nose and mouth when sneezing and coughing, regular and proper handwashing, and eating those eggs and meat that is adequately cooked. Use hand sanitizer and mask on public places, don't spit on the road, avoid close contact with the ill person, avoid traveling with public transport and stay at home as much as possible [5, 8, 11].

In Fig. 1, the situation caused due to COVID-19 outbreak is represented with a bar graph. In the figure, the y-axis represents the number of people, and on the x-axis, the number of cases, the number of recovered cases, number of active cases, and deaths caused by COVID-19 have been presented [11].

In this epidemic situation, any small finding can help a lot. Therefore, in this challenging pandemic situation, the forecasting of coronavirus outbreak plays a vital role and gives an idea about its widespread in upcoming days, which will

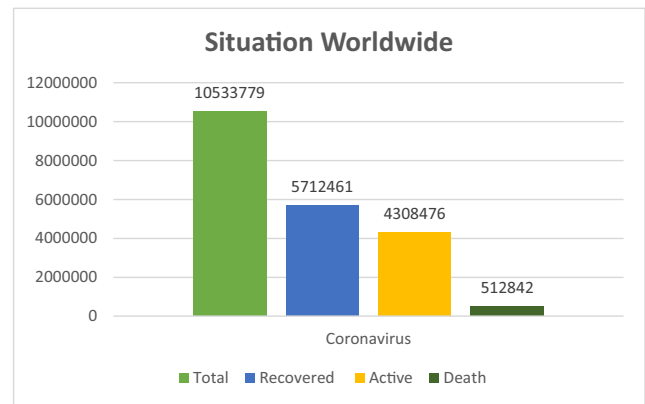


Fig. 1 COVID-19 Situation Worldwide

help the government to take preventive measures for minimizing its spread [12, 13]. In this case, we need such a model that can be accurate, efficient, and widely applicable. It is a challenging task, and it becomes more challenging because we don't have sufficient real-time data. Considering these limitations in our mind, we start testing with the basic forecasting models, which are not as accurate as we are looking for [14]. After several trials, we found that our proposed Enhanced Multi-Task Gaussian Process Regression (MTGP) model on the COVID-19 outbreak forecasting is better than the traditional forecasting model.

The primary aim of this study is:

- To develop such a model that can be capable of dealing with this epidemic situation and give the best forecasting results.
- To provide information on possible conditions in the upcoming days due to the outbreak of COVID-19, which will help in planning the preventive measures accordingly.
- Along with that, the significance of IoT in COVID-19 detection and IoT based possible solutions for minimizing the impact of the COVID-19.

The rest structure of this research paper is as follows. The related work has been presented in Section 2. Section 3 deliberate about the data sets and proposed methodology along with the four traditional forecasting models. The experiment results are based on performance evaluation of the COVID-19 dataset and which has been presented in Section 4. In

Table 1 Virus based epidemic diseases till date

Disease caused by virus	Year	Total cases	Total death	Total recovered	Target countries
SARS	2002–2003	8096	774	7322	29
SWINE FLU	2009–2010	6,724,149	19,654	6,704,495	58
EBOLA	2013–2016	28,646	11,323	17,323	10
MARS	2014–2015	1360	527	833	26
CORONA (COVID-19)	2019–2020	8,525,042	456,973	4,355,602	213+

Section 5, the detailed discussion about the results, and the significance of IoT in the detection of COVID-19 with IoT based possible solutions are shown. In Section 6, concluding remarks with the future scope is drawn.

## 2 Related work

Complexity in developing epidemiological models encourages the researchers to find out the machine learning-based model to deal with the epidemic situation. The primary aim of the machine learning models is to develop models with better prediction reliability and better generalization ability [15–19]. Various machine learning models were used to predict multiple pandemics such as Ebola, SARS, Swine flu, Cholera, H1N1 influenza, Zika, Dengue fever, Oyster norovirus, and MARS [20–30]. These machine learning methods are limited to some basic models such as Neural Networks, Random Forest Regression, Bayesian Networks, Genetic Programming, Naïve Bayes, Classification and Regression Tree, Linear Regression, and Support Vector Regression. However, over a long time, machine learning models are being used to serve in natural disasters and weather predictions [31, 32].

For dealing with this COVID-19 epidemic, various researchers and research groups have given the mathematical and algorithmic approaches. The sub-epidemic model, the Richards model, and the logistic model had been used by Roosa et al. [33] for the prediction of confirmed cases in upcoming days (in the next 5 and 10 days) of Zhejiang province and Guangdong province. The mathematical model for the propagation status simulation due to COVID-19 in China was used by Liu et al. [34]. This mathematical model will further help the government in minimizing the impact of the pandemic. The transmission dynamics' time-dependent compartmental model was adopted by Boldog et al. [35] for estimating the confirmed cases outside china's Hubei province. The improved version of the salp swarm algorithm and the flower pollination algorithm based on ANFIS (Adaptive Neuro-fuzzy Inference System) had been proposed by Al-Qaness et al. [36], for the prediction of confirmed cases in the next ten days. For evaluating the performance of the proposed algorithm, GA (Genetic Algorithm), PSO (Particle Swarm Optimization), ABC (Artificial Bee Colony), and FPA (Flower Pollination Algorithm), had been taken. The statistical induction and delayed distribution estimation based mathematical model was proposed by Jung et al. [37] for predicting COVID-19 cases. An analytical method based mathematical model was proposed by Fan et al. [38] for predicting the floating population of Wuhan city. In this paper, the authors had established the correlation between the floating population and daily confirmed cases.

The deep learning algorithms are being widely used in the genome-based prediction of COVID-19 propagation. The

SEIR (Susceptible Exposed Infectious Removed) model with LSTM (Long Short-Term Memory) model had been proposed by Yang et al. [39] for the integration of the epidemic curve. Hu et al. [40] had suggested the improved version of the stacked autoencoder and clustering algorithm for the grouping of promptly confirmed cases in each province. In this paper, the authors had also found that the AI-based methods achieve higher accuracy for the prediction of COVID-19 trajectory. The deep learning algorithm for the gene sequence-based virus prediction had been used by Guo et al. [41]. In this paper, the authors had compared the gene sequence of SARS-CoV (Severe Acute Respiratory Syndrome), bat SARS coronavirus, and MERS-COV (Middle East Respiratory Syndrome) with the COVID-19 to find out the similarity among the viruses.

The early outbreak trajectories were predicted by Riou et al. [42] to determine the reproduction rate of COVID-19. Besides that, the authors had also told the possibility of frequent infections among people due to COVID-19. The improved version of the SIR (Susceptible Infectious Recovered) model for the prediction of infection cases due to COVID-2019, was proposed by Ming et al. [43]. The authors also predicted the actual load on the ICU (Intensive Care Units) under the various public health intervention efficacies and diagnosis rates. Zhao et al. [44] had proposed a probability model for the accurate prediction of the infection nodes by the snapshots of spreading. The machine learning-based pathogen-risk models were introduced by Fountain-Jones et al. [45] for predicting the COVID-19 outbreak over the various machine learning models. The ARIMA (Autoregressive Integrated Moving Average) model had been used by Benvenuto et al. [46] for the prediction of the trend and morbidity of the COVID-19 outbreak. In this paper, the authors also cited that, if the virus does not mutate, then the number of cases will reach a plateau.

## 3 Materials and methods

In this section, we are going to discuss the materials and methods which have been used for result finding. This section is divided into four subsections, i.e., dataset description, proposed methodology, model comparison, and statistical analysis. In the first subsection, the detailed discussion about the COVID-19 dataset has been presented. In the second subsection, the proposed MTGP model has been discussed in detail. In the third subsection, a brief discussion with the mathematical foundation of the four-prediction algorithm has been presented. In the last subsection, the performance evaluation metrics have been discussed.

### 3.1 Data

For this research work, we had gone through the health bulletin of the WHO. Based on the WHO's situation reports, the

dataset has been earned, and all the experimental evaluation task has been performed on that. Due to COVID-19, approximately 213 countries have been affected. The data we took for making the forecasting model was from 31/12/2019 to 25/06/2020 time stamp. This COVID-19 dataset is consisting of the nine columns which are: country/territory/area name, date, the total number of confirmed cases, the total number of deaths, the total number of new confirmed cases, total number new deaths, the total number of recovered patients, the total number of active patients, and transmission classification [11]. Based on these nine columns, the data of all the affected countries has been recorded.

The visualization of the dataset is shown in Fig. 2. The total number of confirmed cases in the period of 31/12/2019 to 25/06/2020 has been drawn on the map. The radius of the circle is depended on the exposer in the respective country. For the visualization on the map, we had taken the country's geographical location, which is mapped with the number of confirmed cases.

### 3.2 Proposed methodology

Gaussian Process Regression (GBR) is the non-parametric based regression model that is most commonly used in the predictive analysis [14, 47]. It can also be used for a complex system like an arbitrary system. Multi-Task

Gaussian Process Regression (MTGP) is the extensive model of the basic GPR model. It is also known as the advance or special case of a typical GPR model [48]. It is applicable when we want multiple outputs using a standard GPR model. This MTGP model was proposed by Bonilla et al. in 2008 [49], and its advanced version was introduced by Dürichen et al. in time series analysis of multivariate psychological application [50]. More recently, this method was utilized by Richardson et al. to predict battery capacity with improved results [51]. In respect to the prediction of the COVID-19 outbreak using the proposed MTGP model, the COVID-19 time series data has been given at the input side, and in the output side, the historical and reference series has been received. In the MTGP model for the prediction of time series, we follow the same testing and training mechanism as we were using in the traditional GPR model, excluding the kernel matrix. For a better understanding of the kernel matrix in the MTGP model, the two cases of output tasks are considered as an example.

The standard GPR model is described in Eq. (1)

$$a = f(b) \sim N(m(b), i(b, b')) \quad (1)$$

Where the output and input have been denoted by  $a$  and  $b$ , the mean function has been defined as  $m(b)$ ,  $f(b)$  denotes the latent variable, and  $i(b, b')$  denotes the covariance function.

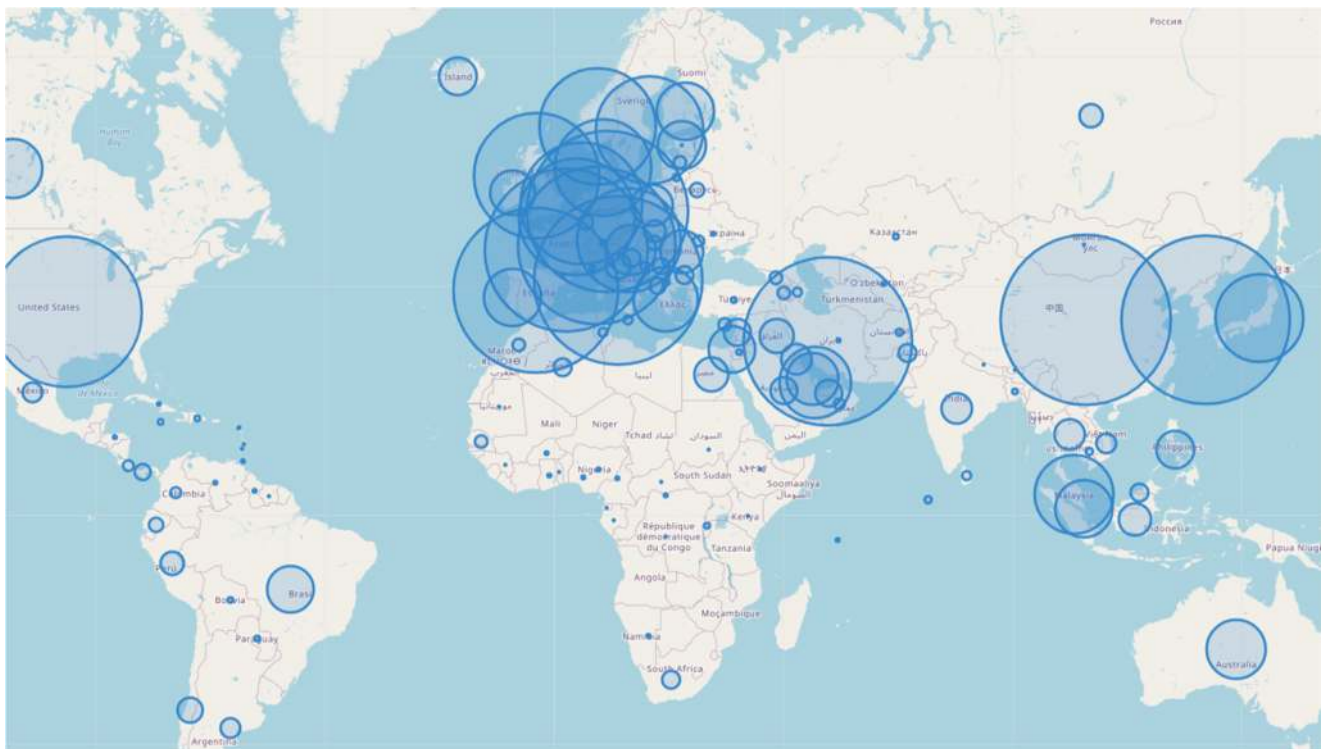


Fig. 2 COVID-19 Outbreak Worldwide



The squared exponential kernel (SEK) function is defined as,

$$i_{SEK} = \theta_1^2 \exp\left(-\frac{d^2}{\theta_2}\right) \tag{2}$$

Where  $d$  denotes the Euclidian distance, and  $\theta_1, \theta_2$  denotes the hyper-parameter, which needs to be optimized.

Thus, the hyper-parameter of the kernel matrix  $I$  have been estimated by minimizing the NLML (Negative Log Marginalized Likelihood) has been shown in Eq. (3)

$$NLML = -\log(p(a|b, \theta))$$

$$NLML = -\frac{1}{2} \log|I + \sigma_n^2 J| - \frac{1}{2} a^T (I + \sigma_n^2 J)^{-1} a - \frac{n}{2} \log(2\pi) \tag{3}$$

The mathematical implementation of the MTGP regression model has been delivered in Eq. (4).

$$\begin{pmatrix} a_r \\ a_p \end{pmatrix} = N \left[ 0, \begin{pmatrix} \theta_{rr} I(b_r, b_r) \theta_{pr} I(b_p, b_r) \\ \theta_{rp} I(b_r, b_p) \theta_{pp} I(b_p, b_p) \end{pmatrix} \right] \tag{4}$$

$$\begin{pmatrix} a_r \\ a_p \end{pmatrix} = N(0, I_{MTGP}(b, b)) \tag{5}$$

Where,  $a_r, a_p$  and  $b_r, b_p$  = two input and two output having with the same dimensionality (In the case of 1D the  $b_r$  and  $b_p$  are sever as time indicators for two time series).

$\theta_{rr}, \theta_{rp}, \theta_{pr}$  and  $\theta_{pp}$  = correlation coefficients of output series

$$F_p^* = I_{MTGP}(b, b_p^*) (I_{MTGP}(b, b) + \sigma_n^2 J)^{-1} \begin{pmatrix} a_r \\ a_p \end{pmatrix} = X_{MTGP} \begin{pmatrix} a_r \\ a_p \end{pmatrix} \tag{6}$$

Where,

$$I_{MTGP}(b, b) = \begin{pmatrix} \theta_{rr} I(b_r, b_r) \theta_{pr} I(b_p, b_r) \\ \theta_{rp} I(b_r, b_p) \theta_{pp} I(b_p, b_p) \end{pmatrix}_{(r+p) \times (r+p)}$$

$$I_{MTGP}(b, b_p^*) = \begin{pmatrix} \theta_{pr} I(b_p^*, b_r) \theta_{pp} I(b_p^*, b_p) \end{pmatrix}_{15 \times (r+p)}$$

The algorithm for the proposed MTGP for COVID-19 outbreak forecasting has been deliberated below in Algorithm 1.

**Algorithm 1.** Enhanced MTGP for COVID-19 outbreak forecasting

At time  $t$ ,

**Step 1:** Initialize the length of time chunk  $l$  and  $a_r, a_p, b_r, b_p, b_r^*$  as:

$b_r = \{T_{t-l}, \dots, T_t, \dots, T_{t+15}\}$ ,  $b_p = \{T_{t-l}, \dots, T_t\}$ ,  $b_r^* = \{T_{t+1}, \dots, T_{t+15}\}$ ,  $a_r \in R^{r \times 1}$  are the prediction of COVID-19 outbreak at the time  $b_r, a_p \in R^{p \times 1}$  is the total number of patients in time  $b_p$ .

**Step 2:** MTGP is constructed as Eq. 4 with the optimized kernel hyper-parameters  $\theta_{rr}, \theta_{pp}, \theta_{rp}$  and other hyper-parameters. The minimized NLML (Negative Log Marginalized Likelihood) is used to get optimal hyper-parameters.

**Step 3:** Obtained the MTGP predictive distribution.

**Step 4:** Transmit it to time  $t + 1$  and repeat Step 1 to Step 3.

### 3.3 Model comparison

In this section, a brief discussion with the mathematical foundation of the prediction algorithm used in the experimental evaluation has been presented.

#### 3.3.1 Linear regression (LR)

Linear Regression (LR) is one of the most commonly used regression models for predictive analysis. The LR model uses two-variables, where the first one is dependent or reliant, and the second one is explanatory or descriptive variable. It is being used for the establishment of the

relationship between the two variables. It is denoted by the slope formulary [52].

The LR model has been represented in Eq. 7, where slope in the line has been denoted by  $b$ ,  $y$ -intercept has been indicated by  $a$ , the reliant or dependent variable has been denoted by  $y$ , and the descriptive or explanatory variable has been denoted by  $x$ .

The LR (Linear Regression) model with the mathematical formulation is defined in the below equation:

$$y = a + bx \tag{7}$$

For calculation of  $a$  and  $b$  the following formulas have been used:

$$b(\text{slope}) = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} \tag{8}$$

$$a(\text{intercept}) = \frac{n\sum y - b(\sum x)}{n} \tag{9}$$

### 3.3.2 Support vector regression (SVR)

Support Vector Regression (SVR) is one of the most commonly used regression models for predictive analysis.

For the prediction using the SVR model, Eq. 10 is used [53].

$$f(x) = \omega\phi(x) + b \tag{10}$$

$$\phi : R^n \rightarrow F, \omega \in F$$

Where, high dimensional feature space has been denoted by  $\phi$  and coefficient of  $\phi(x)$  are  $\omega$  and  $b$ .

$$R_{SVR}(c) = R_{emp} + \frac{1}{2}\|\omega\|^2 = c \times \frac{1}{n} \sum_{i=1}^n L_\varepsilon(d_i, y_i) + \frac{1}{2}\|\omega\|^2 \tag{11}$$

Where empirical risk has been denoted by  $R_{emp}$ , Euclidian norms have been denoted by  $\frac{1}{2}\|\omega\|^2$ , the empirical error has been denoted by  $c \times \frac{1}{n} \sum_{i=1}^n L_\varepsilon(d_i, y_i)$ , and empirical risk cost has been denoted by  $c$ .

$$L_\varepsilon(d, y) = \begin{cases} |d-y|-\varepsilon & |d-y| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

In Eq. (12) the loss function has been shown, and it is obtained from Eq. (11).

Minimize,

$$R_{SVR}(\omega, \xi^*) = \frac{1}{2}\|\omega\|^2 + c \sum_{i=1}^n (\xi_i + \xi_i^*) \tag{13}$$

$$d_i - \omega\phi(x_i) - b_i \leq \varepsilon + \xi_i$$

Subjected to,

$$\omega\phi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^* \xi_i^* \geq 0$$

The testing error minimization has been shown in the Eq. (13), where slack variable for ups and downsides have been denoted by  $\xi^*$  and  $\xi_i^*$ .

### 3.3.3 Random Forest regression (RF)

Random Forest Regression (RF) comes from the family of ensemble learning. It is an ensemble learning-based regression model that has been used for predictive analysis. It was formulated in 1995 by TinKam Ho [54].

In Eq. (14) mathematical RF model has been shown

$$\bar{h}(X) = \left(1/x\right) \sum_{k=1}^K h(x; \theta_k) \tag{14}$$

Where a large number of predictions have been ensured by  $k \rightarrow \infty$ .

$$E_{x,y} \left( Y - \bar{h}(x) \right)^2 \rightarrow E_{x,y} \left( Y - E_\theta h(X; \theta) \right)^2 \tag{15}$$

The random forest prediction error has been defined in the Eq. 16, and the selection of  $pE_t^*$  has been defined in the Eq. 15 to protect the model from overfitting. The average error of an individual tree  $h(X; \theta)$  has been shown in Eq. (16)

$$pE_t^* = E_\theta E_{x,y} (Y - h(x; \theta))^2 \tag{16}$$

Here we adopt that all  $\theta$  are unbiased for each tree, i.e.,

$$EY = E_x h(x; \theta)$$

Then,

$$PE_f^* \leq \bar{p}PE_t^* \tag{17}$$

Where weighted correlation has been defined by  $\bar{p}$  with independent of  $\theta$  and  $\theta'$ , lies between residuals  $y - h(x; \theta)$  and  $y - h(x; \theta')$ .

### 3.3.4 Long short-term memory (LSTM)

The Long Short-Term Memory (LSTM), which is an evolutionary model of Recurrent Neural Network (RNN), was proposed by Hökreiter and Schmiduber [55]. It had been developed to deal with the deficiencies of RNN's anterior with the help of additional interactions cell or module. In other words, we can say that the LSTM is a special type of RNN capable of remembering past information and learning long-term dependencies. Olah [56], quoted that the LSTM model can be organized as a series structure. It consists of four interaction layers instead of a single layer, such as in standard RNN. It also follows the unique communication method for interaction between the four segments. The basic structure of the LSTM model has been shown in Fig. 3.

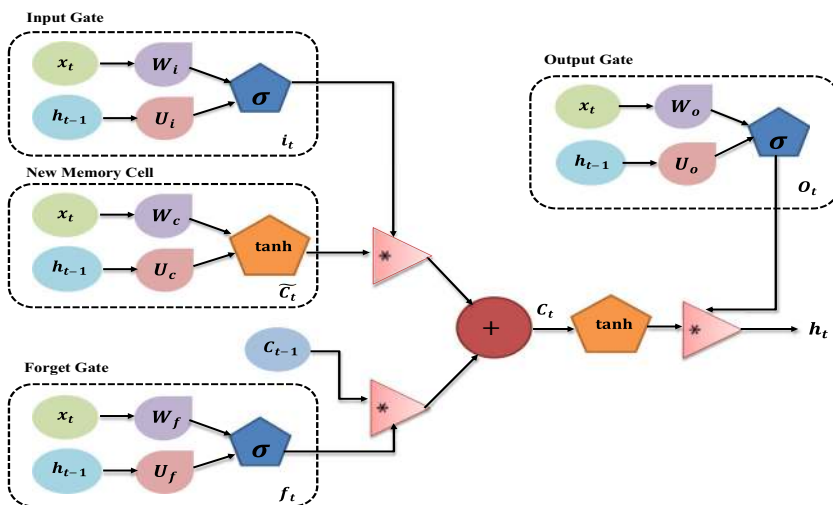
The LSTM model with the mathematical formulation has been defined in the below equations:

The Input gate is calculated by Eq. 18:

$$i_t = \sigma(W_i x_t + X_i h_{t-1}) \tag{18}$$

Where sigmoid function has been denoted by  $\sigma$ , the output of the previous unit has been denoted by  $h_{t-1}$ , and  $W_i$  indicates the weight matrix.

Fig. 3 LSTM Model Structure



The Forget gate is calculated by Eq. 19:

$$f_t = \sigma(W_f x_t + X_f h_{t-1}) \tag{19}$$

Where the output of the previous unit has been indicated by  $h_{t-1}$ , the sigmoid function has been denoted by  $\sigma$ , and  $W_f$  denotes the weight matrix.

The Output /Exposure gate is calculated by Eq. 20:

$$o_t = \sigma(W_o x_t + X_o h_{t-1}) \tag{20}$$

Where the sigmoid function has been denoted by  $\sigma$ , the output of the previous unit has been indicated by  $h_{t-1}$ , and  $W_o$  denotes the weight matrix.

The New memory cell is calculated by Eq. 21:

$$\tilde{c}_t = \tanh(W_c x_t + X_c h_{t-1}) \tag{21}$$

Where the output of the previous unit has been indicated by  $h_{t-1}$ , and  $W_c$  denotes the weight matrix.

The Final memory cell is calculated by Eq. 22:

$$c_t = f_t \times \tilde{c}_{t-1} + i_t \times \tilde{c}_t \tag{22}$$

Forget gate has been denoted by  $f_t$ ,  $\tilde{c}_{t-1}$ , and  $\tilde{c}_t$  indicates cell states at time  $t-1$  and  $t$  respectively.

Thus, the final output is opted by,

$$h_t = o_t \times \tanh(c_t) \tag{23}$$

Where,

$o_t$  and  $c_t$  denotes output /exposure gate and final memory cell, respectively.

### 3.4 Statistical analysis

For the statistical analysis, two performance evaluation metrics have been taken: Mean Absolute Percentage Error (MAPE) and

Root Mean Squared Error (RMSE). Based on these results, we can justify the performance, accuracy, and suitability of forecasting models. In this section, we have discussed the mathematical foundation of the evaluation metrics [57].

#### 3.4.1 Root mean square error (RMSE)

The RMSE (Root Mean Absolute Error) is one of the critical statistical measures which is based on the concept of standard derivation for residuals. It is used for the validation of the predicted result. The measurement of distances between the regression line and data points is known as residuals mean error prediction. In the Eq. 24, the number of errors has been indicated by, the squares of errors have been denoted by  $(\hat{y}_i - y_i)^2$ ,  $\hat{y}_i$  &  $y_i$  denotes the observed values and forecasted values, respectively.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \tag{24}$$

#### 3.4.2 Mean absolute percentage error (MAPE)

The RMSE (Root Mean Absolute Error) is one of the significant statistical measures used to calculate the accuracy of the forecasted system. In the Eq. 25, the number of predicted values have been denoted by  $N$ ,  $X_i$  &  $Y_i$  indicates the predicted value and actual value, respectively.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i - X_i}{Y_i} \right| \times 100 \tag{25}$$

## 4 Results

In this section, we are going to discuss the prediction result of the COVID-19 outbreak based on the five prediction models,

i.e., Linear Regression (LR), Random Forest Regression (RFR), Support Vector Regression (SVR), Long Short-Term Memory (LSTM), and our proposed Enhanced Multi-Task Gaussian Process Regression (MTGP).

#### 4.1 Model identification and hyperparameter selection

Finding out the accurate forecasting model, which will be very useful for predicting the COVID-19 outbreak worldwide, is a very complex task. This study's fundamental purpose is to build such a forecasting model that can correctly forecast the epidemic caused by COVID-19 across the world. The correct forecasting is an essential need because all the preventive measures depend upon the forecasting results. With the help of an accurate forecasting model, we can reduce the COVID-19 outbreak's exposure by making the plan accordingly.

In Fig. 4, four traditional regression models and one proposed model i.e., Linear Regression (LR), Random Forest Regression (RFR), Support Vector Regression (SVR), Long Short-Term Memory (LSTM), and Enhanced Multi-Task Gaussian Process Regression (MTGP) has been shown. These models have been used in the forecasting of the COVID-19 outbreak. This study aims to determine the suitability and accuracy of the proposed model among the traditional models.

In Table 2, the hyperparameter setting, which has been used in the experimental evaluation, has been shown. This table consists of the information such as Prediction Model, Hyperparameter, Parameter Selection, and Best Hyperparameter Used. Before performing the experimental evaluation, we have hyper-tuned each model under the various selection criteria to determine the best parameter for the model. That best hyper-tuned parameter has been used in this experimental evaluation.

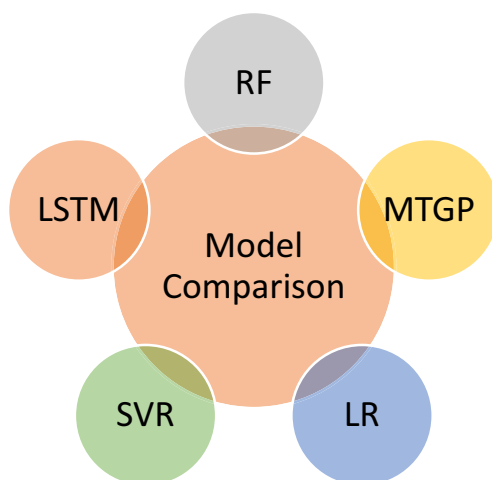


Fig. 4 Prediction Models a Quick Lookup

#### 4.2 Model performance evaluation on COVID-19 outbreak worldwide

Table 3 shows the result of four regression models based on the performance evaluation to forecast the COVID-19 outbreak worldwide. The forecasting result has been divided into five parts. In the first part, the forecasting of COVID-19 confirmed cases worldwide has evaluated. In the remaining part, the country-specific (China, India, Italy, and the USA) forecasting has been performed. China, India, Italy, and the USA has been taken for the prediction of COVID-19 confirmed cases. The forecasted result has been evaluated using two performance measures: MAPE (should be low) and RMSE (should be low). The performance of each model has been calculated using performance measures to find out the best suitable forecasting model among them. All the experiment has been performed using five selection or prediction criteria which are 1-day ahead, 3-day ahead, 5-day ahead, 10-day ahead, and 15-day ahead. The proposed MTGP model wins the battle with the lowest MAPE and RMSE throughout the experiment under different selection criteria (various countries).

### 5 Discussion

In the process of result finding, we have made the dataset with the help of the WHO health bulletin. Day to day bulletin report has been taken into consideration from the timestamp of 31/12/2019 to 25/06/2020. The data is consisting of approximately 213 affected countries with the country/territory/area name, date, the total number of confirmed cases, the total number of deaths, the total number of new confirmed cases, total number new deaths, the total number of recovered patients, the total number of active patients, and transmission classification. The default settings have been used in four machine learning models i.e., Linear Regression (LR), Random Forest Regression (RFR), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM). But in the MTGP model, the improved kernel matrix has been used. We followed the same testing and training mechanism for the prediction of time series as we were using in the traditional GPR model, excluding the kernel matrix. In the proposed MTGP forecasting model, the dataset is given at the input side, and in the output side, the historical and reference series is received.

In Figs. 5 and 6, the performance measures-based prediction outcomes of China, India, Italy, USA, and worldwide using various forecasting model has been shown. Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) have been used as performance measures. The performance of each model has been calculated using these performance measures to determine the best suitable forecasting model among them. All the experiment has been



**Table 2** Hyperparameter selection

Prediction model	Hyperparameter	Parameter selection	Best hyperparameter used
LR	dual	[True, False]	False
	max_iter	[100,110,120,130,140]	100
	C	[1,1.5,2,2.5]	2
SVR	kernel	rbf	rbf
	C	[0.1, 1, 100, 1000]	100
	Gamma	[0.0001, 0.001,0.01 0.005, 0.1, 1, 3, 5]	0.1
	Epsilon	[0.0001, 0.001, 0.01, 0.05, 0.1, 0.5, 1, 5, 10]	0.0001
	n_estimators	[100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]	400
RFR	bootstrap	[True, False]	True
	max_depth	[10, 20, 30, 40, 50, 60, 70, 80, 90, 100, None]	70
	max_features	['auto', 'sqrt']	auto
	min_samples_leaf	[1, 2, 4]	4
	min_samples_split	[2, 5, 10]	10
	n_estimators	[100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000]	400
LSTM	neurons	[1–5]	1
	batches	[1, 2, 4]	4
	epochs	[500, 1000, 2000, 4000, 6000]	1000

performed using five selection or prediction criteria: 1-day ahead, 3-day ahead, 5-day ahead, 10-day ahead, and 15-day ahead. With the help of these measures results, we have found that our proposed model's performance is better than all other

**Table 3** Performance evaluation of the prediction algorithms (confirmed cases)

Country	Methods	Predict days									
		1-day ahead		3-day ahead		5-day ahead		10-day ahead		15-day ahead	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
Worldwide	LR	7.22	2092.62	6.56	2102.11	6.01	2162.48	6.86	2189.33	7.08	2199.49
	SVR	3.65	465.50	6.63	483.35	3.18	528.18	3.56	546.71	3.93	596.33
	RFR	6.80	1762.12	6.43	1792.36	5.39	1820.11	6.44	1840.24	5.40	1863.75
	LSTM	2.06	293.96	2.86	324.73	2.18	366.16	2.66	411.32	2.91	460.18
	MTGP (Proposed)	1.16	122.97	1.27	166.18	1.15	206.23	1.45	286.36	1.48	331.27
China	LR	3.5	688.21	3.57	713.79	3.81	723.53	3.63	774.18	3.51	798.90
	SVR	1.50	220.92	1.56	241.42	1.78	285.16	1.75	334.94	1.83	359.89
	RFR	2.60	457.70	2.59	461.93	2.69	512.35	2.80	551.37	2.90	595.40
	LSTM	1.10	178.93	1.19	198.24	1.28	212.75	1.31	241.13	1.37	256.5
	MTGP (Proposed)	0.70	119.84	0.82	129.81	0.85	139.10	0.88	148.86	0.92	154.74
India	LR	6.40	2125.89	6.77	2151.64	6.98	2175.23	7.32	2267.69	7.50	2297.94
	SVR	3.72	537.20	3.80	557.08	4.27	582.32	4.38	608.1	4.54	619.73
	RFR	5.29	1837.49	5.54	1873.55	5.97	1908.22	6.07	1958.13	6.39	2060.97
	LSTM	2.91	367.36	3.09	389.50	3.45	407.83	3.59	435.97	3.65	440.86
	MTGP (Proposed)	2.10	197.35	2.38	222.13	2.54	232.85	2.74	262.92	2.77	280.19
Italy	LR	7.25	1706.76	7.43	1754.17	7.59	1782.82	7.65	1805.70	7.97	1822.05
	SVR	2.93	393.16	3.14	405.75	3.50	425.67	3.68	449.53	3.83	486.87
	RFR	5.46	1542.48	5.49	1584.83	5.7	1604.12	5.95	1659.67	6.03	1665.71
	LSTM	2.11	287.50	2.20	297.13	2.53	318.87	2.75	337.5	2.92	376.14
	MTGP (Proposed)	1.27	182.38	1.42	190.13	1.65	215.73	1.74	225.51	1.88	257.24
USA	LR	7.21	2033.71	7.69	2070.22	7.87	2190.84	8.18	2250.66	8.36	2183.56
	SVR	3.48	408.4	3.84	467.32	3.97	483.36	4.08	497.20	4.21	513.02
	RFR	6.05	1679.66	6.32	1772.70	6.57	1785.79	6.71	1862.87	6.95	1894.36
	LSTM	2.80	302.64	3.05	336.47	3.19	360.12	3.31	378.43	3.45	395.50
	MTGP (Proposed)	1.89	196.85	2.07	203.79	2.35	235.01	2.59	258.50	2.65	274.12

**Fig. 5** Bar Graph of the Results based on Mean Absolute Percentage Error (MAPE) of Confirmed Cases: **a** Worldwide **b** China **c** India **d** Italy **e** USA



**Fig. 6** Bar Graph of the Results based on Root Mean Square Error (RMSE) of Confirmed Cases: **a** Worldwide **b** China **c** India **d** Italy **e** USA



models. The proposed model achieves the lowest MAPE and RMSE throughout the experiment under various selection criteria (multiple countries).

In Figs. 7, 8, 9, 10 and 11, on the y-axis, the number of infected patients has been taken, and, on the x-axis, the date has been represented. In this experiment, the 15-days advance forecasting of confirmed cases Worldwide, China, India, Italy, and the USA, have been performed, and their results have been plotted in Figs. 7, 8, 9 and 10, respectively. Four forecasting models have been used to perform the experiments. The models we have taken are LR, SVR, RF, LSTM, and proposed MTGP model, and based on these models' results; the graph has been plotted. The prediction results for Worldwide has been presented in Fig. 7a to e. The prediction results for China, India, Italy, and the USA has been shown in the graph 8(a) to 8(e), 9(a) to 9(e), 10(a) to 10(e), and 11(a) to 11(e) respectively. All the models simulate the prediction of the COVID-19 outbreak relatively well. In the context of performance, the proposed enhanced MTGP model performs better among all other forecasting models.

## 5.1 Significance of IoT in the detection of COVID-19

The bitter truth about the technologies is that we cannot replace humans with the technology because the human is enriched with decision-making authority and technologies have limited power. However, it is also a truth that humans can opt for the latest technology for making their life simple and smooth. The keywords which will define IoT for everything are efficiency, convenience, and automation. Emerging of IoT technology is a game-changer not only in the industrial perspective but also in the healthcare domain because it connects to various heterogeneous devices using wired or wireless connections and sends or receives the data to cloud base stations. The beginning of IoT in healthcare has been seen, such as the use of smart sensors, remote monitoring of patients, and the integration of various medical devices. Apart from this, the use of applications like wearable biometric sensors, activity recognition via sensors, medication dispensers, glucose monitors, and smart beds gives the wing to IoT technology and emergence to the IoHT (Internet of Healthcare Things). The significance of the IoT devices in disease detection has been discussed below.

### 5.1.1 In directly turning data into action

Data should always ready into action, and if we are using the IoT device for health monitoring, so we don't have to worry about the data at an epidemic stage. In the case of the COVID-19 outbreak, we face data-related issues because we don't have the correct numbers about infected patients across the world. We have some data where such smart healthcare infrastructure has been developed, but it is very lesser. We are only

dependent on the WHO, which is somehow managing informative data of all countries with respect to current scenarios. That is one of the most significant issues to date. Thus, neither can we control the virus spread rate nor make an efficient prevention plan. If we have such infrastructure in the future, we will never suffer like this, because we will have enough data and could make plans accordingly.

### 5.1.2 In improving the health of patients

With the help of wearable IoT devices, humans can track their health and take precautions to improve their health. If we talk about the significance of IoT devices in detecting the COVID-19 outbreak, we can follow each ill person and take appropriate action to minimize its spread. Suppose we have collected data that are coming from various tracking devices. In that case, we can identify the maximum exposed area in terms of disease spread as well as the people who are escaping from the hospitals or isolation wards can also be tracked. We can also keep on eyes at the suspects. So overall, with the help of IoT devices, we can efficiently control the exposure rate as well as improve the health of patients.

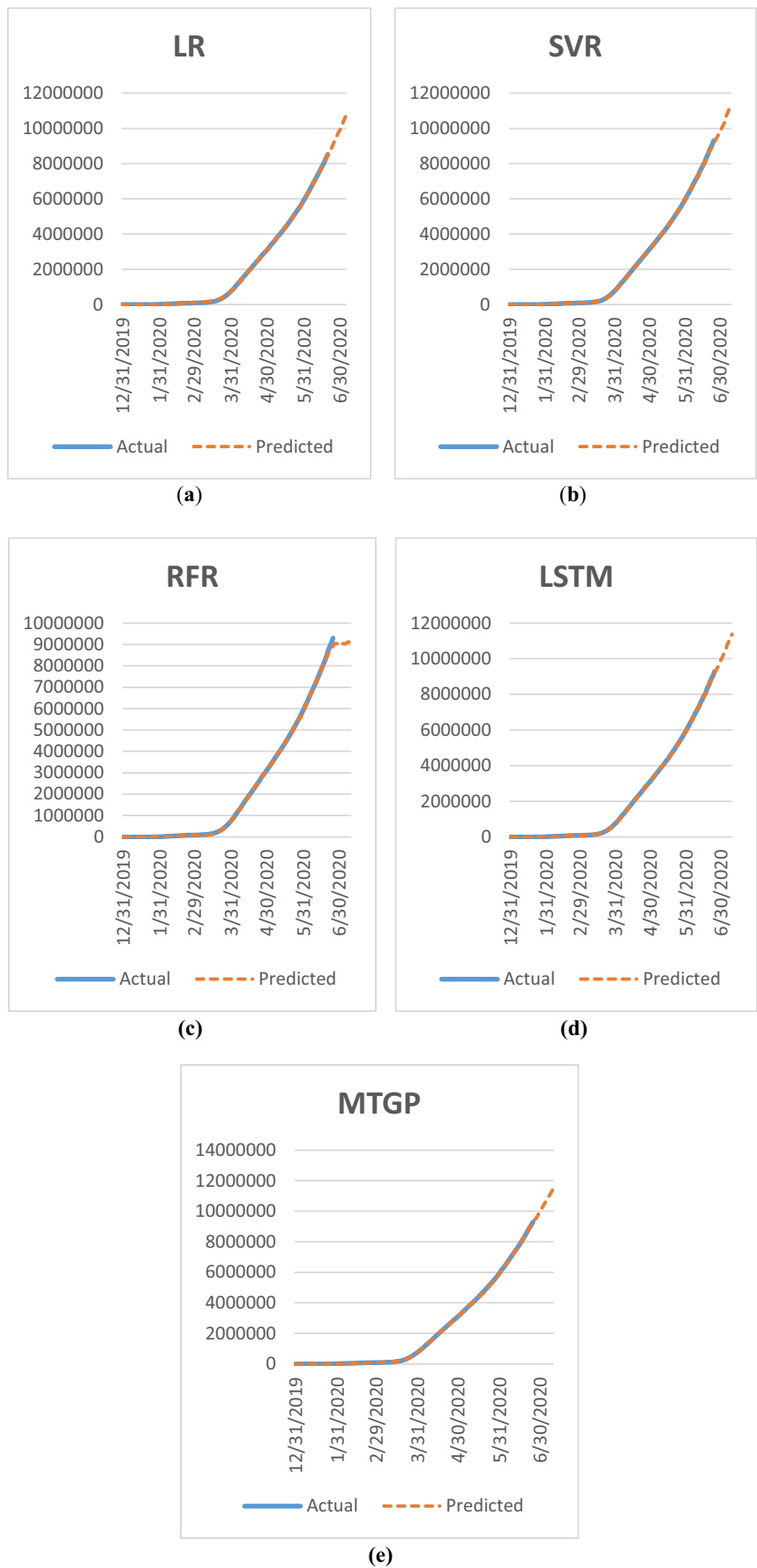
### 5.1.3 In the promotion of preventative measures

With the help of IoT infrastructure, we can circulate the prevention method for the public and aware of them with its consequences and effects on human health. As we know that IoT devices may generate the data in real-time thus, we can share this information or data among the mobs so they can keep track of the epidemic situation. Apart from it, we can warn people not to go in such areas where the maximum exposure has been being taken place. Thus, with the help of a real-time system like IoT for healthcare, we can promote and spread awareness among people. We can also guide them on how to deal with epidemic situations.

### 5.1.4 In the advancements of management strategies for healthcare

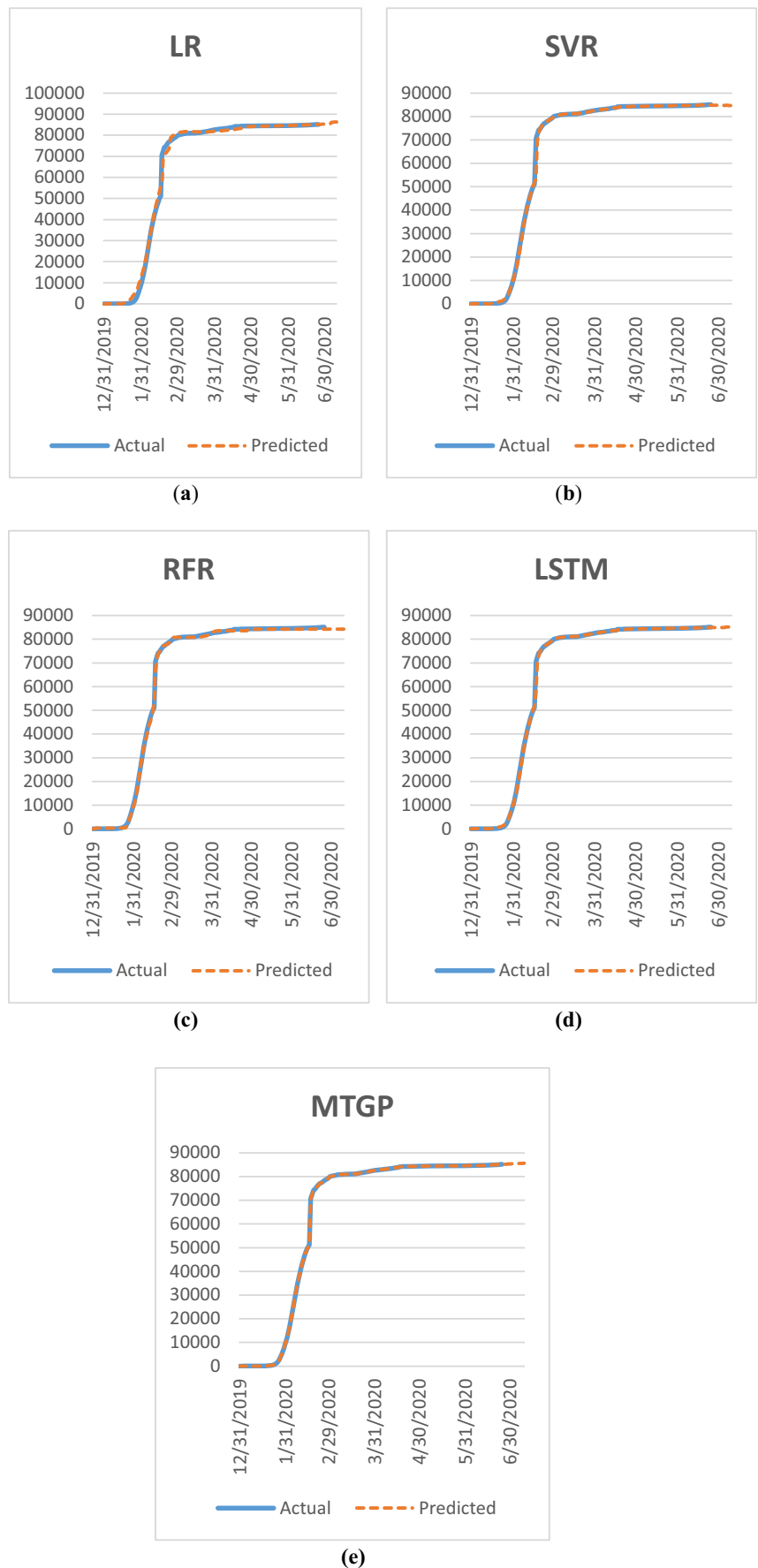
Due to the COVID-19, the people who are suffering from respiratory diseases need extensive care. Thus, with the help of a health dataset, we can identify and track the persons who are having such problems. The prevention strategies can be made using the real-time data of disease spread and applying efficient management strategies to deal with it. It was earlier not possible because we didn't have real data. So, advanced management strategies need real-time data to deal with any epidemic diseases like COVID-19. That is possible only if, when we are using IoT infrastructure for the healthcare system.

**Fig. 7** Graph of 15-Days Advance Forecasting of Confirmed Cases in COVID-19 Outbreak for Worldwide: **a** LR Model **b** SVR Model **c** RFR Model **d** LSTM Model **e** Proposed MTGP Model

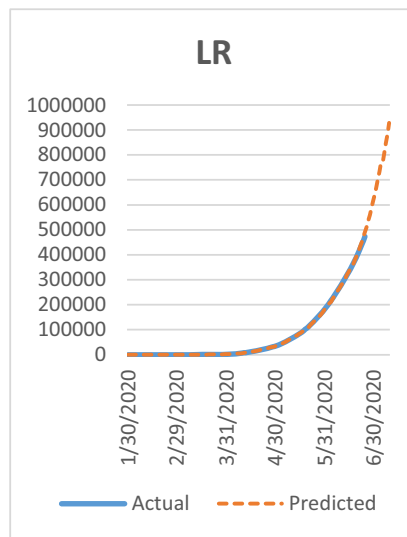




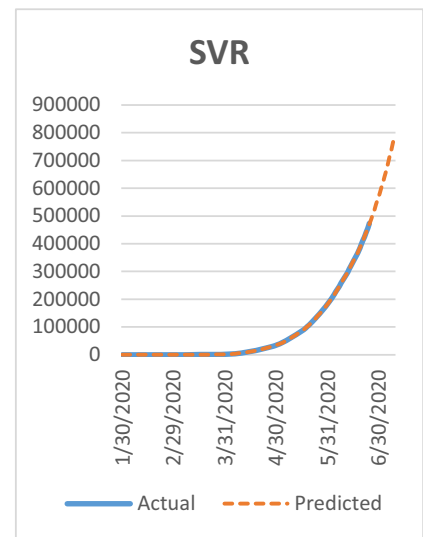
**Fig. 8** Graph of 15-Days Advance Forecasting of Confirmed Cases in COVID-19 Outbreak for China: **a** LR Model **b** SVR Model **c** RFR Model **d** LSTM Model **e** Proposed MTGP Model



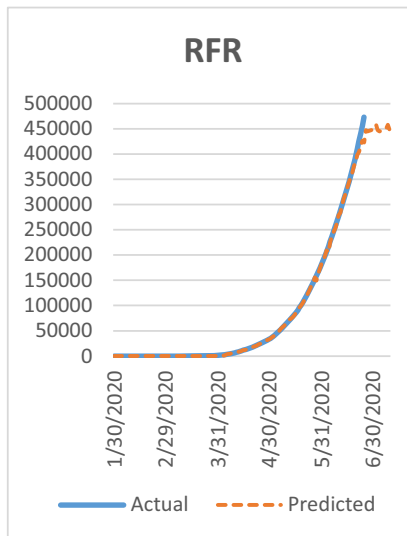
**Fig. 9** Graph of 15-Days Advance Forecasting of Confirmed Cases in COVID-19 Outbreak for India: **a** LR Model **b** SVR Model **c** RFR Model **d** LSTM Model **e** Proposed MTGP Model



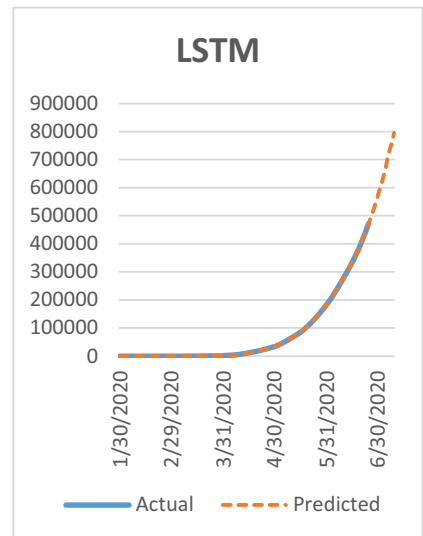
(a)



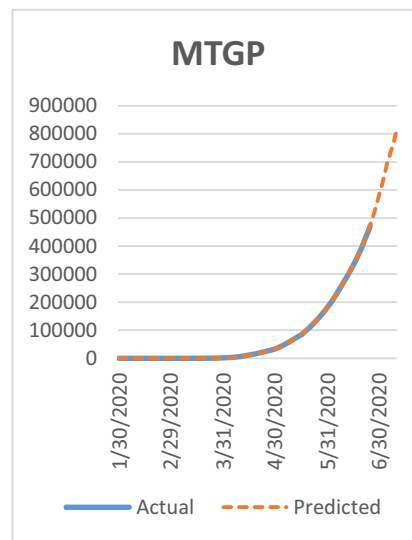
(b)



(c)

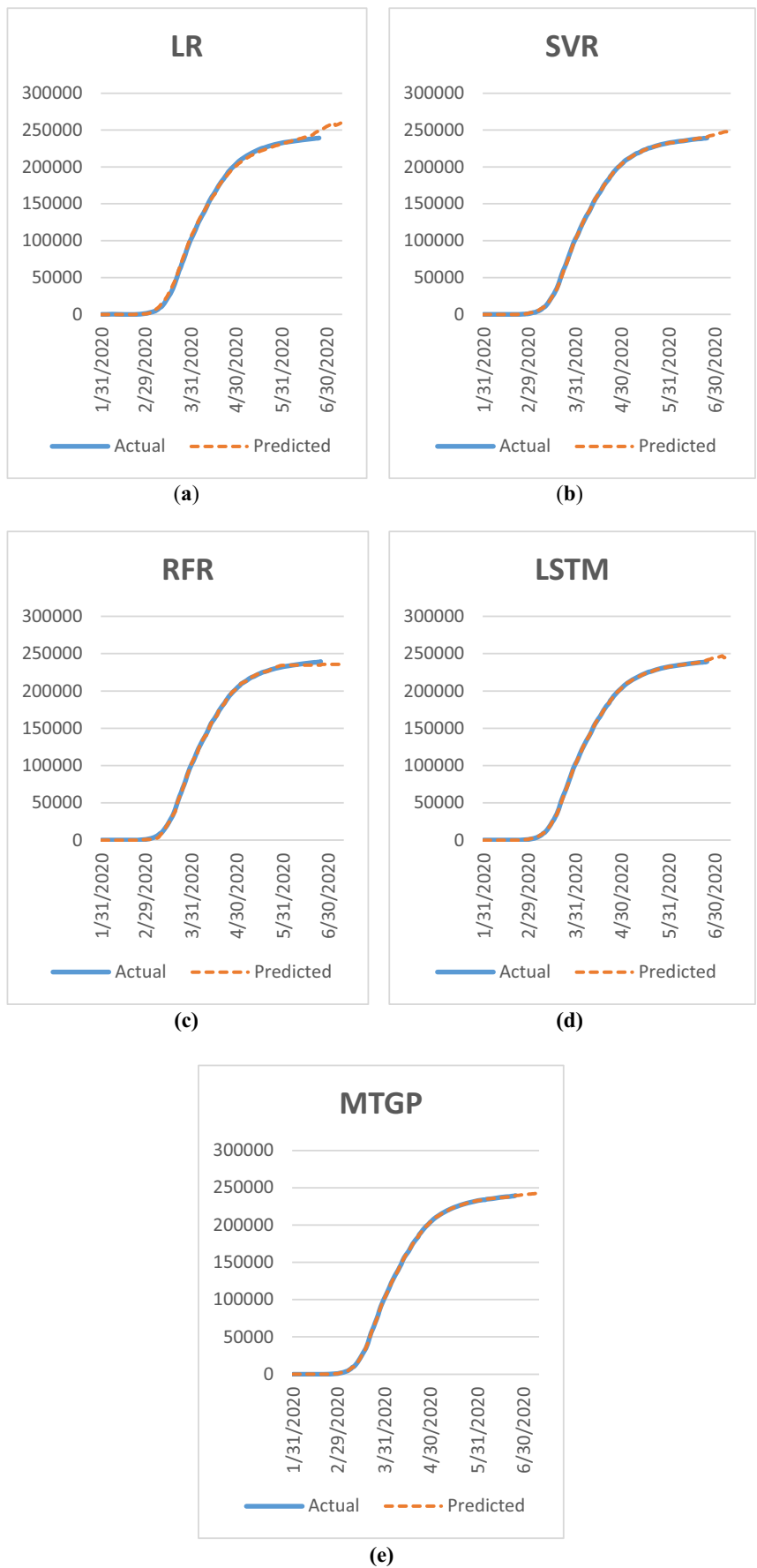


(d)



(e)

**Fig. 10** Graph of 15-Days Advance Forecasting of Confirmed Cases in COVID-19 Outbreak for Italy: **a** LR Model **b** SVR Model **c** RFR Model **d** LSTM Model **e** Proposed MTGP Model



**Fig. 11** Graph of 15-Days Advance Forecasting of Confirmed Cases in COVID-19 Outbreak for the USA: **a** LR Model **b** SVR Model **c** RFR Model **d** LSTM Model **e** Proposed MTGP Model



## 5.2 Person who needs extensive care from COVID-19

As we know, the COVID-19 is a virus-based contagious disease that has been widely spread nowadays. It can affect all the age group person but deadliest for the persons who are suffering from the below-mentioned diseases. Such a person needs extensive care from COVID-19. In Table 4, the persons who need extensive care has been discussed [11, 58, 59].

## 5.3 Possible solution

For the detection of novel infectious diseases by existing resources and definition is the step by step process. The real-time data and their analytics are needed for taking preventive measures. With the help of accurate information and quick action, we can reduce the overall impact of the novel diseases worldwide, which can also reduce the social and economic factors of countries. The real-time tracking and detecting of the novel diseases are an arduous task, but by using the IoT, it is possible too. IoT can play a vital role in detecting and controlling the outbreak of virus-based infectious diseases in real-time.

### 5.3.1 Real-time tracking of disease spread

The emergence of technologies such as big data and IoT in healthcare gives researchers an area to work on. With the help of IoT, we can get the data from the various places in real-time where earlier, it was either possible thru manually or not possible at all. The device like a smart thermometer and smart thermal cameras plays a vital role in the people's initial screening so that we can detect the people suffering from COVID-19. Intelligent devices such as thermal cameras can be installed at every place like airports, railways, roadways, and home doors. Thus, we can easily identify the healthy person and ill person. All systems should be integrated with the base stations (not in a case of home automation), where we

can monitor the data and plan the desired action accordingly. These things are real, which means the data has been collected and analyzed in real-time. Thus, with the help of IoT based systems, the real-time tracking of the disease spread rate and spread density can be achieved. As we know, the IoT is an interconnection of heterogeneous devices or systems, and with the help of this, we can also collect the data from remote areas. Integrating these data globally for participating in the global health system will not only help in the real-time tracking of the diseases but also in the prediction mechanism for the prevention of disease spread.

### 5.3.2 IoT based responding and monitoring for public health

The use of IoT can change individuals' lives by integrating smart clinics. Smart clinics mean the clinics which are using the various IoT devices and collecting the data in real-time. It can also communicate among the other smart clinics and also share information for medical studies in real-time. These smart clinics are connected to the central server, where all the data is being collected and monitored in real-time. All the records about the patient should be available at the primary database so, it will easy for the government to take the preventive measure when the epidemic situation comes into the picture. The primary aim of such infrastructure is to enable IoT devices and enable the associated health clinics to manage public health crises. It will also be helpful in detecting suspicion outbreaks as well as in the detection of a confirmed outbreak caused by virus-based diseases. In the situation of growing suspicion of an explosion, it can use the IoT devices to obtain focused data by locating the epidemic source. Similarly, the same network can be used for providing medication and health services in the affected areas.

### 5.3.3 Employing efficient mechanisms for prevention of infection

The lack of data availability and relevant data is a big hurdle toward making prevention policies. Data is one of the essential requirements for making appropriate approaches and test the hypothesis. It also plays a vital role in controlling infectious diseases. With the help of IoT and advanced technology in the proficient healthcare system, we can overcome these limitations or challenges. With the help of IoT based systems, we can collect the data from various locations and resources. After that, these data are sent into the central health server, which will serve as an input to the researcher for the analysis. Based on the analysis result, we can find the impact of the outbreak and make the preventive roadmap accordingly to minimize the overall effect. As we know, the whole world is suffering from the COVID-19, where preventive measures are significant to minimize the overall impact. With the help of

**Table 4** The person who needs extensive care

Diseases	Severe health problems for
COVID-19	<ul style="list-style-type: none"> <li>✓ The person suffering from heart disease</li> <li>✓ The person suffering from congestive heart failure</li> <li>✓ The person suffering from coronary artery syndrome</li> <li>✓ The person suffering from asthma,</li> <li>✓ The person suffering from Emphysema</li> <li>✓ The person suffering from COPD (Chronic Obstructive Pulmonary Disease)</li> <li>✓ Women with pregnancy</li> <li>✓ Outdoor labors</li> <li>✓ Old age peoples</li> <li>✓ Sportsperson who exercise strongly in outdoors</li> </ul>



IoT based smart system, we can easily find highly infected areas. Based on IoT-based intelligent systems, we can make effective solutions such as where to place the sanitizing points, which area should be a lockdown, and where we need massive health support.

## 6 Conclusions

In an epidemic situation, any small finding can help a lot. In this challenging pandemic situation, the forecasting of the COVID-19 outbreak plays an important role. It gives an idea about its widespread in upcoming days, which will help the government to take preventive measures for minimizing its spread. Correct and efficient forecasting of outbreaks in any epidemic situation is a very complex but novel task. It becomes more complicated when preventive measures depend upon the prediction, and we don't have real-time data. In this paper, a COVID-19 outbreak forecasting method has been proposed to deal with all limitations compared to other traditional models. For the experimental evaluation, we have compared the performance of four traditional forecasting models, i.e., Linear Regression (LR), Random Forest Regression (RFR), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) with our proposed MTGP forecasting model to find out its suitability and correctness. Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) have been used as performance measures. The performance of each model has been calculated using these performance measures to determine the best suitable forecasting model among them. All the experiment has been performed using five selection or prediction criteria: 1-day ahead, 3-day ahead, 5-day ahead, 10-day ahead, and 15-day ahead. With the help of these measures results, we have found that our proposed model's performance is better than all other models. The proposed model achieves the lowest MAPE and RMSE throughout the experiment under various selection criteria. Apart from that, we have found the significance of IoT in healthcare as well as the importance of IoT for COVID-19 detection and IoT based possible solutions for minimizing the impact of the COVID-19.

In the future, we again analyze the proposed model with different datasets and find out the further boosting techniques which can boost the model efficiency. The Sparse Gaussian Process method is one of the probable solutions.

**Acknowledgments** We would like to thank Prof. Moonis Ali and Prof. Hamido, Editor-in-Chief of Applied Intelligence for their critical reviews and suggestions. We are also very thankful for giving us the opportunity to revise our manuscript. We show our sincere gratitude to the anonymous reviewer(s) whose valuable comments helped us greatly to improve the presentation.

## References

1. Severe Acute Respiratory Syndrome (SARS). Available online: [https://en.wikipedia.org/wiki/Severe\\_acute\\_respiratory\\_syndrome](https://en.wikipedia.org/wiki/Severe_acute_respiratory_syndrome). Accessed on 10 Mar 2020
2. SWINE FLU. Available online: [https://en.wikipedia.org/wiki/2009\\_flu\\_pandemic\\_by\\_country](https://en.wikipedia.org/wiki/2009_flu_pandemic_by_country). Accessed on 10 Mar 2020
3. Ebola. Available online: [https://en.wikipedia.org/wiki/Western\\_African\\_Ebola\\_virus\\_epidemic](https://en.wikipedia.org/wiki/Western_African_Ebola_virus_epidemic). Accessed on 10 Mar 2020
4. Middle East respiratory syndrome (MERS). Available online: [https://en.wikipedia.org/wiki/Middle\\_East\\_respiratory\\_syndrome](https://en.wikipedia.org/wiki/Middle_East_respiratory_syndrome). Accessed on 10 Mar 2020
5. Novel Coronavirus(COVID-19). Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed on 10 Mar 2020 to 24 Mar 2020
6. "WHO | Novel Coronavirus – China". WHO. Archived from the original on 23 January 2020. Accessed on 5 Mar 2020
7. Cohen J (2020) Wuhan seafood market may not be source of novel virus spreading globally. *Science* 10
8. Coronavirus(COVID-19). Available online: [https://www.who.int/health-topics/coronavirus#tab=tab\\_1](https://www.who.int/health-topics/coronavirus#tab=tab_1). Accessed on 5 Mar 2020
9. WHO Emergency Committee (2005) Statement on the second meeting of the international health regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (COVID-19). WHO, Geneva, p 2020
10. Sparrow A (2020) How China's coronavirus is spreading—and how to stop it. *Foreign Policy*. Archived from the original on, 31
11. Coronavirus disease (COVID-2019) situation reports. Available online <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>. Accessed on 23 Jan to 21 June 2020
12. Croda RMC, Romero DEG, Morales SOC (2019) Sales prediction through neural networks for a small dataset. *IJIMAI* 5(4):35–41
13. Hyndman RJ, Kostenko AV (2007) Minimum sample size requirements for seasonal forecasting models. *foresight* 6(Spring):12–15
14. Liu H, Ong YS, Shen X, Cai J (2020) When Gaussian process meets big data: a review of scalable GPs. *IEEE Transactions on Neural Networks and Learning Systems*
15. Burke RM, Shah MP, Wikswa ME, Barclay L, Kambhampati A, Marsh Z, Cannon JL, Parashar UD, Vinjé J, Hall AJ (2019) The norovirus epidemiologic triad: predictors of severe outcomes in US norovirus outbreaks, 2009–2016. *J Infect Dis* 219(9):1364–1372
16. Carlson CJ, Dougherty E, Boots M, Getz W, Ryan SJ (2018) Consensus and conflict among ecological forecasts of Zika virus outbreaks in the United States. *Sci Rep* 8(1):1–15
17. Kleiven EF, Henden JA, Ims RA, Yoccoz NG (2018) Seasonal difference in temporal transferability of an ecological model: near-term predictions of lemming outbreak abundances. *Sci Rep* 8(1):1–6
18. Rivers-Moore NA, Hill TR (2018) A predictive management tool for blackfly outbreaks on the Orange River, South Africa. *River Res Appl* 34(9):1197–1207
19. Yin R, Tran VH, Zhou X, Zheng J, Kwok CK (2018) Predicting antigenic variants of H1N1 influenza virus based on epidemics and pandemics using a stacking model. *PLoS One* 13(12):e0207777
20. Koike F, Morimoto N (2018) Supervised forecasting of the range expansion of novel non-indigenous organisms: alien pest organisms and the 2009 H1N1 flu pandemic. *Glob Ecol Biogeogr* 27(8):991–1000
21. Agarwal N, Koti SR, Saran S, Senthil Kumar A (2018) Data mining techniques for predicting dengue outbreak in geospatial domain using weather parameters for New Delhi, India. *Curr Sci* 114(11): 2281–2291
22. Anno S, Hara T, Kai H, Lee MA, Chang Y, Oyoshi K, Mizukami Y, Tadono T (2019) Spatiotemporal dengue fever hotspots

- associated with climatic factors in Taiwan including outbreak predictions based on machine-learning. *Geospat Health* 14(2)
23. Chenar SS, Deng Z (2018) Development of artificial intelligence approach to forecasting oyster norovirus outbreaks along Gulf of Mexico coast. *Environ Int* 111:212–223
  24. Chenar SS, Deng Z (2018) Development of genetic programming-based model for predicting oyster norovirus outbreak risks. *Water Res* 128:20–37
  25. Iqbal N, Islam M (2019) Machine learning for dengue outbreak prediction: a performance evaluation of different prominent classifiers. *Informatica* 43(3)
  26. Liang R, Lu Y, Qu X, Su Q, Li C, Xia S, Liu Y, Zhang Q, Cao X, Chen Q, Niu B (2020) Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data. *Transbound Emerg Dis* 67(2):935–946
  27. Mezzatesta S, Torino C, De Meo P, Fiumara G, Vilasi A (2019) A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis. *Comput Methods Prog Biomed* 177:9–15
  28. Raja DB, Mallol R, Ting CY, Kamaludin F, Ahmad R, Ismail S et al (2019) Artificial intelligence model as predictor for dengue outbreaks. *Malaysian Journal of Public Health Medicine* 19(2): 103–108
  29. Tapak L, Hamidi O, Fathian M, Karami M (2019) Comparative evaluation of time series models for predicting influenza outbreaks: application of influenza-like illness data from sentinel sites of healthcare centers in Iran. *BMC research notes* 12(1):353
  30. Muurlink OT, Stephenson P, Islam MZ, Taylor-Robinson AW (2018) Long-term predictors of dengue outbreaks in Bangladesh: a data mining approach. *Infectious Disease Modelling* 3:322–330
  31. Mosavi A, Ghamisi P, Faghan Y, Duan P (2020) Comprehensive review of deep reinforcement learning methods and applications in economics. *arXiv preprint arXiv:2004.01509*
  32. Nádai L, Imre F, Ardabili S, Gundoshmian TM, Gergo P, Mosavi A (2020) Performance analysis of combine harvester using hybrid model of artificial neural networks particle swarm optimization. *arXiv preprint arXiv:2002.11041*
  33. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, Yan P, Chowell G (2020) Short-term forecasts of the COVID-19 epidemic in Guangdong and Zhejiang, China: February 13–23, 2020. *J Clin Med* 9(2):596
  34. Liu Z, Magal P, Seydi O, Webb G (2020) Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. *arXiv preprint arXiv:2002.12298*
  35. Boldog P, Tekeli T, Vizi Z, Dénes A, Bartha FA, Röst G (2020) Risk assessment of novel coronavirus COVID-19 outbreaks outside China. *J Clin Med* 9(2):571
  36. Al-Qaness MA, Ewees AA, Fan H, Abd El Aziz M (2020) Optimization method for forecasting confirmed cases of COVID-19 in China. *J Clin Med* 9(3):674
  37. Jung SM, Akhmetzhanov AR, Hayashi K, Linton NM, Yang Y, Yuan B, Kobayashi T, Kinoshita R, Nishiura H (2020) Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: inference using exported cases. *J Clin Med* 9(2):523
  38. Fan C, Liu L, Guo W, Yang A, Ye C, Jilili M, Ren M, Xu P, Long H, Wang Y (2020) Prediction of epidemic spread of the 2019 novel coronavirus driven by spring festival transportation in China: a population-based study. *Int J Environ Res Public Health* 17(5):1679
  39. Yang W, Cao Q, Qin L, Wang X, Cheng Z, Pan A et al (2020) Clinical characteristics and imaging manifestations of the 2019 novel coronavirus disease (COVID-19): a multi-center study in Wenzhou city, Zhejiang, China. *Journal of Infection*
  40. Huang CJ, Chen YH, Ma Y, Kuo PH (2020) Multiple-input deep convolutional neural network model for covid-19 forecasting in China. *medRxiv*
  41. Zhu H, Guo Q, Li M, Wang C, Fang Z, Wang P, ... Xiao Y (2020) Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm. *BioRxiv*
  42. Riou J, Althaus CL (2020) Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* 25(4):2000058
  43. Ming WK, Huang J, Zhang CJ (2020) Breaking down of healthcare system: mathematical modelling for controlling the novel coronavirus (2019-nCoV) outbreak in Wuhan, China. *bioRxiv*
  44. Chen DB, Zhao N, Wang J, Yu Y, Zhao J (2020) Spreading predictability in complex networks. *bioRxiv*
  45. Fountain-Jones NM, Machado G, Carver S, Packer C, Recamonde-Mendoza M, Craft ME (2019) How to make more from exposure data? An integrated machine learning pipeline to predict pathogen exposure. *J Anim Ecol* 88(10):1447–1461
  46. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M (2020) Application of the ARIMA model on the COVID-2019 epidemic dataset. *Data in brief*, 105340
  47. Rasmussen CE (2003) Gaussian processes in machine learning. In: *Summer school on machine learning*. Springer, Berlin, pp 63–71
  48. Liu H, Cai J, Ong YS (2018) Remarks on multi-output Gaussian process regression. *Knowl-Based Syst* 144:102–121
  49. Bonilla EV, Chai KM, Williams C (2008) Multi-task Gaussian process prediction. In: *Advances in neural information processing systems*, pp 153–160
  50. Dürrichen R, Pimentel MA, Clifton L, Schweikard A, Clifton DA (2014) Multitask Gaussian processes for multivariate physiological time-series analysis. *IEEE Trans Biomed Eng* 62(1):314–322
  51. Richardson RR, Osborne MA, Howey DA (2017) Gaussian process regression for forecasting battery state of health. *J Power Sources* 357:209–219
  52. Weisberg S (2005) *Applied linear regression*, vol 528. Wiley
  53. Drucker H, Burges CJ, Kaufman L, Smola AJ, Vapnik V (1997) Support vector regression machines. In: *Advances in neural information processing systems*, pp 155–161
  54. Liaw A, Wiener M (2002) Classification and regression by randomForest. *R news* 2(3):18–22
  55. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
  56. Olah C (2018) Understanding lstm networks, August 2015. URL <https://colah.github.io/posts/2015-08-Understanding-LSTMs>. Accessed on 10 June 2020
  57. Coe\_cient of Determination. Available online: [https://en.wikipedia.org/wiki/Coe\\_cient\\_of\\_determination](https://en.wikipedia.org/wiki/Coe_cient_of_determination). Accessed on 9 Sept 2019
  58. Jin YH, Cai L, Cheng ZS, Cheng H, Deng T, Fan YP et al (2020) A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). *Military Medical Research* 7(1):4
  59. Schwartz DA, Graham AL (2020) Potential maternal and infant outcomes from (Wuhan) coronavirus 2019-nCoV infecting pregnant women: lessons from SARS, MERS, and other human coronavirus infections. *Viruses* 12(2):194

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Shwet Ketu** received the B.Tech. degree in Computer Science and Engineering from B.B.S. College of Engineering and Technology, India, in 2012, and M.Tech. degree in Information Technology from Indian Institute of Information Technology, India, in 2015. Currently, he is pursuing Ph.D. in Computer Science at Department of Computer Science, Institute of Science, Banaras Hindu University, Varanasi, India. His research interests include Internet of Things

(IoT), Big Data Analytics, Machine learning, and Data Science.



**Pramod Kumar Mishra** is Professor at Department of Computer Science, Institute of Science, Banaras Hindu University, India. He is a Fellow of Institute of Mathematical Analysis (UK) and a senior member of IEEE (USA). His research interests include Computational Complexity, Data Mining, IoT, High Performance Computing and VLSI Algorithms.