Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Projecting the criticality of COVID-19 transmission in India using GIS and machine learning methods

Farhan Mohammad Khan a,*, Akshay Kumar a, Harish Puppala b, Gaurav Kumar a, Rajiv Gupta a

a Department of Civil Engineering, BITS Pilani, Pilani, Rajasthan, India
b BML Munjal University, Gurgaon, India

ARTICLE INFO

Keywords:
COVID-19
Machine learning
Transmission
Lockdown
Gaussian process regression
Support vector machine
Decision tree

ABSTRACT

There is a new public health catastrophe forbidding the world. With the advent and spread of 2019 novel coronavirus (2019-ncov). Learning from the experiences of various countries and the World Health Organization (WHO) guidelines, social distancing, use of sanitizers, thermal screening, quarantining, and provision of lockdown in the cities being the effective measure that can contain the spread of the pandemic. Though complete lockdown helps in containing the spread, it generates complexity by breaking the economic activity chain. Besides, laborers, farmers, and workers may lose their daily earnings. Owing to these detrimental effects, the government has to open the lockdown strategically. Prediction of the COVID-19 spread and analyzing when the cases would stop increasing helps in developing a strategy. An attempt is made in this paper to predict the time after which the number of new cases stops rising, considering the strong implementation of lockdown conditions using three different techniques such as Decision Tree, Support Vector Machine, and Gaussian Process Regression algorithm are used to project the number of cases. Thus, the projections are used in identifying inflection points, which would help in planning the easing of lockdown in a few of the areas strategically. The criticality in a region is evaluated using the criticality index (CI), which is proposed by authors in one of the past of research works. This research work is made available in a dashboard to enable the decision-makers to combat the pandemic.

1. Introduction

Coronavirus disease (COVID-19) is a new and contagious disease caused by a new virus, known as novel coronavirus. The disease affects the lungs and causes a respiratory illness with symptoms such as cold, throat inflammation, cough, fever, and trouble breathing in severe cases. Public health authorities recommended that one can protect herself/himself by frequently washing and/or sanitizing hands, avoiding touching the nose, ears, and face, and by maintaining social distancing with other people. Considering the global footprint of this pandemic, the World Health Organization (WHO) declared COVID-19 to be a pandemic on 30th March 2020 and set out guidelines to help countries safeguard critical health services during the COVID-19 epidemic. Action plans were in place by different countries to contain the spread of this pandemic [1].

In order to control the situation, it is essential to have a plan in place, which depends on the prediction of new cases due to COVID-19. This would help hospitals and administrations to take necessary measures in advance [2]. In the context of an emerging infectious disease outbreak, predicting the trend of the epidemic is of paramount importance to plan effective control strategies and determine how said strategies impact the course of the epidemic.

Gaussian process regression (GPR) is a nonparametric, Bayesian approach to regression, which is widely used in the area of machine learning. A significant benefit of GPR is that it can work well on small datasets and have the ability to provide uncertainty measurements on the required predictions [3]. Like other supervised machine learning algorithms that learn exact values for every parameter in a function, the Bayesian approach deduces a probability distribution over all the possible values. GPR is not limited by a functional form; it calculates the probability distribution of parameters for a specific function by distributing the probabilities over all admissible functions that fit the data. A prior function space is specified that predicts the pattern observed using the training data and computed the predictive posterior distribution based on the classifiers. The mean function and covariance kernel function is selected based on model architecture and tuned during model selection. The mean function is typically constant, either zero or the mean of the training dataset. There are many options for the covariance kernel function: it can have many forms as long as it follows the properties of a kernel (i.e., semi-positive definite and symmetric). Some common kernel functions include constant, linear, exponential, square exponen-

* Corresponding author.
E-mail address: p20180411@pilani.bits-pilani.ac.in (F.M. Khan).

https://doi.org/10.1016/j.jnlars.2021.05.001

Received 25 June 2020; Received in revised form 24 May 2021; Accepted 24 May 2021

© 2021 China Science Publishing & Media Ltd. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Daily Cases

![Cumulative Cases](Image)

Fig. 1. Time series plot of COVID-19 cases as of 10th June 2020.

Fig. 2. Response plot of the GPR model.

Fig. 3. Response plot of the SVM model.

tial, and Matern kernel. A prior function space and covariance kernel functions have been used for GPR [4]. The Gaussian process regression method is the Octave and MATLAB implementation of several localized regression methods, like the domain decomposition method [5], partial independent conditional [6], localized probabilistic regression (Urtasun and Darrell, 2008, LPR), and bagging for Gaussian process regression [7]. For general machine learning problems, most localized regression methods can be applied, although the domain decomposition method is only applicable for spatial data sets. The GPR method also offers two parallel versions of the DDM for the computation. The ease of parallelization is one of the advantages of localized regression, and the two parallel implementations can provide proper guidance on how this benefit can be realized as software [8]. Gaussian process regression has the advantage of being able to combine different kernels, creating a rich set of interpretable and reusable building blocks [9]. For example, adding two kernels together models the data as an independent functional superposition. Multiplying a kernel with a function kernel with a radial basis smooths the first kernel’s predictions locally. Gaussian process model learns all functions efficiently. Even if the inputs are sampled at random, the error for a Gaussian process regression always goes down [10].

A support vector machine is a supervised learning algorithm that can be used for classification or regression. An SVM classifies data by finding the best hyperplane that separates data points of one class from points of the other class [11]. Support vectors refer to a small subset of training data that is used as support for the optimal location of the decision surface. SVM contains linear, quadratic, cubic, fine gaussian medium gaussian, coarse gaussian SVM for classification and predictions in which quadratic SVM has medium model flexibility, hard interpretability, and it uses binary medium and large multiclass for memory usage [12].

In machine learning, a decision tree can be used to represent decisions and decision-making for deriving a strategy to predict future data. Decision trees are easy to interpret, fast for fitting and predictions and, low on memory usage. It contains a coarse tree, medium tree, fine tree to increase the model flexibility with the maximum number of splits settings. A fine tree has many large leaves to make many subtle distinctions between different classes [13]. For predictions, the decisions follow in the tree from the beginning node down to a leaf node. A fine tree with many leaves is usually highly accurate on the training data. A leafy tree tends to over train, and its validation and accuracy often higher than its training. A fine DT is a greedy algorithm that performs a binary classification of the feature space to maximize the information gained at a tree node [14].

Predictive mathematical models have been formulated over the years [15–18], including the SIR model, a widely-used and straightforward de-
terministic model for human-to-human transmission that describes the flow of individuals through three mutually exclusive stages of infection: Susceptible (S), Infected (I) and Recovered (R) [19]. Manikandan et al. [20] estimated the number of cases of the Ebola virus using autoregressive integrated moving average (ARIMA) models in western African countries. The data were collected from March 2014 through December 2014, and estimates were made until December 2016. Results showed an increasing trend in the actual and forecast number of cases with the Ebola virus using ARIMA (1,1,0) based on Bayesian Information Criteria (BIC) values. Chowell et al. [21] analyzed the prediction trend of the Zika epidemic and estimated the spread of disease based on a generalized Richards model. The study predicted the number of reproductions from the growth phase over an increased duration. Results showed that the generalized Richards model outperformed the logistic growth model and concluded that the phenomenological models were able to predict the epidemic based on the interval distribution assumptions of generation. Batista [22] used the logistical model to predict the total number of cases and peak periods of the coronavirus epidemic in China, South Korea, and the rest of the world, providing a reasonable description of the outbreak in those countries. Menon et al. [23] used a genetic algorithm method to predict the final size of the COVID-19 pandemic in India and showed that the initial doubling time for confirmed cases was 3.6 days. The study concluded that the reported cases expected will be 111,185, based on the initial estimate. Dehesh et al. [24] used autoregressive integrated moving average (ARIMA) models to estimate the confirmed cases of COVID-19 in Italy, China, South Korea, Iran, and Thailand. The study recommended taking preventive measures, such as social distancing and quarantining, to control the COVID-19 pandemic. Hu et al. [25] proposed artificial intelligence-based methods for predicting COVID-19 cases in China in real-time. For estimating the confirmed cases, a modified auto-encoder and clustering algorithms were used. The authors concluded that AI-based methods are accurate in determining the final COVID-19 projection and policy-making process.

Bhatnagar [26] used a decision tree algorithm for COVID-19 prediction of confirmed, recovered, and deceased cases in India, France, Italy, and the USA. The study considered the prediction lockdown phase and concluded that lockdown plays a vital role in controlling the spread of COVID-19. The confirmed cases predicted by Arti et al. [27] in India using the tree-based model. The study showed that infection spread was 2.3 per day, and disease spread was 0.15 per day in the lockdown conditions. The study assumed the forecast recovery rates and concluded that it takes 54 days of complete lockdown to control the spread of the pandemic. Perc et al. [28] used a simple predictive iteration method for COVID-19 confirmed cases in USA, Slovenia, Iran, and Germany. Results indicated that the daily rate of growth should be kept at least below 5%. Kumar et al. [29] used decision trees, support vector machine, and Gaussian progress regression to project the spread of COVID-19 in India and the further use of order-preference similarity technique (TOPSIS), a multi criteria-based analysis to generate the criticality indices at district level datasets.

Pinter et al. [30] proposed a hybrid method of MLP-ICA and ANFIS for the forecasting of confirmed cases and mortality rate of COVID-19 pandemic. Results for the MLP-ICA algorithm showed that the highest value of the coefficient of determination (R²) for confirmed cases
obtained was 0.9971 and for mortality rate was 0.9986. Onovo et al. [31] studied the pattern and correlation of the COVID-19 disease outbreak in sub-Saharan Africa (SSA) using Lasso Regression with the lowest MSE of $2.72 \times 10^{-28}$ and the highest $R^2$ value of 1.0. M.A.M.T. et al. [32] predicted the confirmed cases of the COVID-19 outbreak in Senegal using the SIR model. Still, there was no clear numerical description in the study. Sattler et al. [33] forecasted the risk of COVID-19 transmission from the BLE signal strength measurements using linear regression. Results from the study showed the area under the curve (AUC) obtained was 0.96. Many researchers have predicted confirmed, recovered and deceased cases of COVID-19 disease using machine learning approach [34], ODE Solver and SIRD model ([35]), Linear regression and Multilayer perceptron network ([36]), SIR model with time-dependent parameters and deep learning approach [37], A-SIR model [38], Support vector machine (SVM), Polynomial regression (PR), Standard deep neural network (DNN) and Recurrent neural network (RNN) using long short term memory [LSTM] [39], SEIR and LSTM model [40], Augmented ARGoNet model [41], Statistical machine learning model [42], SIR and prophet model [43], Autoregressive Integrated Moving Average (ARIMA) model [44], Unsupervised machine learning algorithms [45]. The studies were based on the prediction of cases using various machine learning algorithms. Still, there is a research gap observed in past studies as no clear numerical description was reported. With various improvements to the past methodology [46] developed by authors and by considering the COVID-19 database, the spread of COVID-19 is predicted at a national level, which is further distributed among the districts in terms of cumulative cases. In view of the findings of past research work [46], the Gaussian Process Regression method is adopted in this study to project the number of cumulative cases that are likely to arise. Additionally, the inflection points at a district level, which was not addressed in the past study [46], is discussed.

2. Methodology

The need of the hour is to understand the spatial footprint of COVID-19 and to predict the possible spread along with analyzing the risk in a region. In this regard, a framework is designed to facilitate the decision-maker in forecasting the likely scenarios of its extent. Besides mapping the cumulative cases which are evaluated using the data set available as of 10th June 2020, the attributes of the parameters were projected.
for the next two months using three Machine Learning Techniques such as Gaussian Process Regression (GPR), Support Vector Machine (SVM) and Decision Tree (DT).

Gaussian process regression (GPR) is a nonparametric, Bayesian approach to regression, which is widely used in the area of machine learning. A significant benefit of GPR is that it can work well on small datasets and have the ability to provide uncertainty measurements on the required predictions. GPR taking exponential kernel function was used in forecasting considering the worst-case scenario. Exponential GPR is identical to the Squared Exponential GPR except that the Euclidean distance is not squared. Exponential GPR replaces inner products of basis functions with kernels slower than the Squared Exponential GPR. The Exponential GPR handles smooth functions well with minimal errors. A support vector machine is a supervised learning algorithm that can be used for classification or regression. An SVM classifies data by finding the best hyperplane that separates data points of one class from points of the other class. SVM contains linear, quadratic, cubic, fine gaussian medium gaussian, coarse gaussian SVM for classification and predictions. The quadratic SVM has medium model flexibility, hard interpretability, and it uses binary medium and large multiclass for memory usage. Decision trees are easy to interpret, fast for fitting and predictions and, low on memory usage. It contains a coarse tree, medium tree, fine tree to increase the model flexibility with the maximum number of splits settings.

The model performance was evaluated using the value of the coefficient of determination ($R^2$), mean square error (MSE), root mean square error (RMSE), and mean absolute error (MAE). The mean squared error (MSE) is one of several ways of quantifying the difference between the predicted values and the actual values of the measured quantity. The mean squared error quantifies the difference between the predicted values and the actual values of the measured quantity. Both MSE and RMSE are integral components in regression models. The mean absolute error is an average of the absolute errors. Lesser values of these measures show more precisely predicted output. The higher value of $R^2$ indicates that the model explains variation in the dependent parameter. The value of $R^2$, MSE, RMSE, and MAE can be calculated using Eqs. (1)–(4) respectively:

$$R^2 = \frac{\sum_{i=1}^{m}(y_i - \bar{y}_i' )}{\sqrt{\sum_{i=1}^{m} (y_i - \bar{y}_i')^2}}$$

(1)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y}_i)^2$$

(2)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y}_i)^2}$$

(3)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \bar{y}_i|$$

(4)

Where,

- $t_{pj}$ = Target or real value p.
- $y_{pj}$ = $i^{th}$ output of the final layer or predicted value.
- $\bar{y}_{pj}$ = mean of targeted or real value.
- $\bar{y}_{pj}'$ = mean of the predicted value.
- $n$ = number of datasets.

Data regarding the number of cases reported in India till 10th June 2020, were collected from the Ministry of Health and Family Welfare (MoHFW) [47] and https://www.COVID19india.org, and was compiled in a repository. The time series plot of cumulative confirmed, recovery, and deceased cases reported in India till 10th June 2020 is shown in Fig. 1.

---

**Fig. 9.** Forecast of Cumulative Confirmed Cases using Quadratic SVM method.

**Fig. 10.** Forecast of Cumulative Confirmed Cases using Fine DT method.
The different parameters considered are daily positive new cases, population, population density, deceased cases, and recovered cases. The considered positive cases define the intensity of COVID-19 in a region. Population data reflects the possible asymptomatic carriers. Population density, on the other hand, defines the risk of spread/community transmission. The number of deceased and recovered cases in the region determines the quality of health care facilities. The daily data corresponding to each of these variables was collected from various sources and is compiled in a repository. The attribute related to the criteria discussed in the aforementioned sections was collected from 30th Jan, 2020. Subsequently, the possible variation for the next two months forecasted using the three different techniques, as discussed above. The forecasted attributes consequently help in estimating the cumulative score of the region, which is determined using TOPSIS, a Multi-criteria decision-making technique. The obtained cumulative score is introduced as the criticality index. Based on this distribution of criticality index, the regions can be classified into clusters of low risk, moderate risk, and high risk.

It is recommended that districts that are in the green zone can be completely released. However, travel to another adjacent district can be restricted if they fall into another zone. Lockdown can alternatively be opened and closed in these regions with continuous monitoring of the new positive cases. The importance of the work lies in identifying the districts which are falling in the more severe zone in the following weeks. For such districts, the policy of partial release is recommended with various preventive actions in place. Due to the nature of the problem, it is recommended that maps should be updated daily, and changes of district from one critical zone to another should be identified. Though the findings help to plan the combat strategies, updating the database would help to forecast which could mimic the reality. The following are a few of the assumptions in the current study that may test the capabilities of the decision-maker in planning the essential strategies since there always be possibility for a certain deviation. For better clarity, the assumptions are listed below-

Assumption 1. For district-wise analysis, the dataset used in this study for forecasting was taken from various sources. Due to the sudden mass outbreak of this pandemic, there is a discrepancy in data over various sources.

Assumption 2. It may have to be noted that the time series analysis for the total number of daily cases is considered from the date when the first infection was registered in the state. Also, those days were considered as zero values when no case appeared in the district and vice versa for recoveries and deaths for all the districts, respectively.

Assumption 3. For missing data for the district, a proportionate value to the nearest hotspot in the state is taken. In forecasting, if the daily increase in the number of cases is less than 0.4, then we have considered it as zero, and if the daily increase in the number of cases is greater than 0.4, then we have considered it as one.

Assumption 4. District-wise deceased cases and recovered cases are based on the new cases which arose 15 days before.

3. Analysis, results, and discussions

In this study, three machine learning-based algorithms are proposed to observe the transmission pattern of COVID-19 by changing the learning coefficients from 0.1 to 0.001. Performance evaluation of the forecast is estimated in terms of Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE), and Coefficient of Determination ($R^2$) values, the statistics are found to be within acceptable limits. The model architecture is optimized using Exponential GPR that showed the best performance with the least RMSE 124.38 and the highest coefficient of determination values ($R^2$) of 0.95, as given in Table 1. After the training of regression models, the performance is compared based on model statistics, and the results are visualized in response plot by plotting actual versus predicted response, and evaluation of the model is performed using the response plot. The model statistics are computed using the observations in the k-fold cross-validation. The model makes predictions on the observations in the validation folds, and the plot shows the forecast. It also computes the residuals on the observations in the validation folds. The validation of the model was done by 10-fold cross-validation on daily confirmed cases. The output of the model response of the GPR, SVM, and DT model is shown in Fig. 2, Fig. 3, and Fig. 4, respectively. The predicted and observed pattern of the GPR, SVM, and DT model is shown in Fig. 5, Fig. 6, and Fig. 7, respectively.

Based on available data as of 10th June 2020, the daily confirmed, recovered, and deceased cases of COVID-19 cases in India are forecasted using the three machine learning techniques such as GPR, SVM, and DT model for the next two months, as Figs. 8–10, represents the forecast of daily new confirmed cases by taking the value of learning coefficient as 0.001, 0.01 and 0.1, respectively. The prediction using the GPR method has been made considering that the lockdown will remain active, and it
is expected that confirmed cases will start declining from the first week of August. It can be seen clearly from the curve that at the end of July, the daily new confirmed will start declining if the conditions remain the same in the country. Figs. 11, Fig. 12, and 13 represent the prediction for recovered cases, and Figs. 14, Fig. 15, and 16 depicts the prediction for deceased cases, which is based on the assumption that today’s confirmed cases either will be changed to recovered or deceased cases after around 15 days. The results are validated with existing data. However, the percentage of each state may be different as far as recovered and deceased cases are concerned.

These predictions are made with existing conditions. However, it can be improved by taking a few preventive steps. We intend to further improve our model by collecting more data in the upcoming days. Hence it is proposed to update the site weekly. The projected statistics are distributed among the districts. Although this exercise is associated with deviation, it gives a holistic picture and also helps decision-makers to foresee the intensity of COVID-19 transmission at a district level. The mapping helps to find the surrounding districts to take further precautions. The distributed confirmed cases, recovery cases, and deceased cases are shown in Fig. 17, Fig. 18, and Fig. 19, respectively.
Fig. 15. Forecast of Cumulative Deceased Cases using Quadratic SVM method.

Fig. 16. Forecast of Cumulative Deceased Cases using Fine DT method.

Fig. 17. Projected confirmed cases in each district for next 4 weeks.
Fig. 18. Projected recovery cases in each district for the next 4 weeks.

Fig. 19. Projected deceased cases in each district for the next 4 weeks.
Table 2
Proposed lockdown opening and closing strategy as per criticality index.

|          | Open                          | Close                          | Open                          | Close                          | Open                          |
|----------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| Low CI   | From 16th June to 29th June; all local market only | From 30th June to 13th July; partial lockdown | From 14th July to 28th July; Intrastate Transport | From 29th July to 11th August, close Transport | All from 12th August |
| Moderate CI | From 23rd June to 7th July; all local market only | From 8th July to 15th July; partial lockdown | From 15th July to 14th August; Intrastate Transport | From 15th August to 21th August, close Transport | All from 22nd August |
| High CI  | From 30th July to 14st July; all local market only | From 15th July to 23th July; partial lockdown | From 24th July to 23rd August; Intrastate Transport | From 24th August to 30th August, close Transport | All from 1st September |

The forecasted attributes consequently help in estimating the cumulative score of the region, which is determined using TOPSIS, a Multi-criteria decision-making technique. The obtained cumulative score is introduced as the criticality index. Based on this distribution of criticality index, the regions can be classified into clusters of low risk, moderate risk, and high risk. Lockdown should be imposed in the area of high risk, whereas red zones can be identified in the regions of moderate-risk, and restriction to movement can be imposed. Lockdown in the region of low risk can be released with some precautionary measures. The probable variation of the criticality index for the next four weeks is mapped using the three different techniques. The variation obtained from TOPSIS using the data forecasted using DT, SVM, and GPR. The criticality of COVID-19 in a region is evaluated using the Criticality Index. Three crucial parameters are considered that capture the CI are - intensity, preparedness, and the probability of transmission. The intensity of COVID-19 in a region is evaluated using the total number of confirmed cases; preparedness in a region is evaluated using the hospitals to population ratio, and population density is taken as a measure to evaluate the probability of transmission. The attributes of these parameters corresponding to each of the regions are collected, and the cumulative risk associated with each of the districts is evaluated.

Based on this distribution of criticality index, the regions can be classified into clusters of low risk, moderate risk, and high risk. Lockdown should be imposed in the area of high risk, whereas red zones can be identified in the regions of moderate risk, and restriction to movement can be imposed. Lockdown in the region of low risk can be released with some precautionary measures. The probable variation of the criticality index for the next two months from 10th June is mapped using the GPR method. The criticality index mapped using the forecasted data obtained from GPR is shown in Fig. 20. From the evaluated criticality index, it is apparent that the intensity would increase in the districts geographically located in Rajasthan, Andhra Pradesh, Maharashtra, Bihar, Delhi, Punjab, and Chhattisgarh.

Keeping in view the projected scenario, current duration of lockdown, degrading economic activity, a lockdown opening strategy is proposed, as shown in Table 2.

The framework of opening the lockdown is based on the fact that relaxing the activities may raise the number of cases. Hence a subsequent lockdown is essential to identify the critical hidden districts, stop the resurgence of the cases, and finally to update the resources. A web GIS-based platform has been developed to portray the real-time statistics of the lockdown estimations as per the current number of active cases. It is under development and can be checked at http://rainwater.bits-pilani.ac.in:8081/COVID19/.

The findings of the study are further used in identifying the total number of days after which the curve flattens. With the day-wise projected data set, the inflection points are identified, and the districts are grouped into five different groups, as shown in Fig. 21. The lockdown can be released sequentially as per the groups.

From Fig. 21, it can be noted that for the regions lying in zone 1 i.e.0 days from 1st week of August; it can be allowed to resume the economic activity. Similarly, the region in each of the clusters can be permitted after the respective inflection points. In general, it can be conclusively stated for a few of the districts, and lockdown should be extended, which is found to be in line with the decision taken by Gov. Besides this, the findings help for the sequential release of lockdown.

Figs. 22–25 shows the time series plot of COVID-19 cases in China, Italy, USA, and United Kingdom as of 10th June 2020. Currently, since people infected with COVID-19 will travel to countries or regions with low ongoing transmission, attempts should be made to stop disease transmission, deter future outbreaks, and reduce second and subsequent COVID-19 epidemic waves. Reduced export rates could slow the import

---

1 https://www.ideasforindia.in/topics/trade/harnessing-technology-to-fight-COVID-19-a-web-enabled-dashboard-to-visualise-risk.html
of cases into cities or countries that are unaffected or have a low number of COVID-19 cases, giving time to organize an adequate public health response [48,49].

Wells et al. [50] since the bulk of transmission happened after symptom initiation, the severe acute respiratory syndrome epidemic in southern China in 2003 was able to be monitored by tracing contacts of cases [51]. These measures are also crucial in outbreaks where signs and infectiousness appear at the same time, such as Ebola virus disease [52,53], MERS [54,55], and other viral infections [56,57]. The isolation of cases and touch tracking, according to Kucharski et al. [58], could be less successful for COVID-19 because infectivity begins before symptoms appear [59]. According to Hellewell et al. [60], adequate touch tracing and case isolation are sufficient to contain a new COVID-19 epidemic within three months, although the likelihood of control declines with lengthy delays between symptom initiation and isolation, which increases dissemination until symptoms. In the context of COVID-19 outbreaks, it's critical to understand the factors that influence the infectious disease's dissemination dynamics to develop strategies for halting or slowing its spread and empowering health policy through fiscal, social, and environmental interventions.

**Comparison of models**

In order to demonstrate the performance of the detection result through the proposed method, various benchmark methods are used to perform comparison experiments. Table 3 shows that the proposed method in this paper outperforms other methods in the prediction with an accuracy of 95%.

**4. Conclusion**

This study considered the data related to daily confirmed, deceased, and recovered cases of COVID-19 in India. Besides evaluating the present data, efforts are made in projecting the cases using Machine Learning Techniques to forecast the possible scenario of the future of COVID-19 in India. Additionally, this study proposed a criticality index that can help to quantify the risk in a region, which is further used to classify the regions into zones of high risk, low risk, and moderate risk. Developing maps by considering the updated data and mapping the risk for the following weeks by integrating machine learning tools and GIS would certainly help to combat the COVID-19 transmission. The critical contribution of the work lies in preparedness and mitigation of the disaster with at least some projection, which is based on scientific methods.
Even if the predictions may not be very accurate, it does give a systematic plan to combat an unknown enemy. It has to be noted that the findings of this study will match reality if the lockdown continues. For future modifications, the new methodology will be used for forecasting the scenario of COVID-19 after removing the complete lockdown.

Declaration of Competing Interest

The authors have declared no competing interest.

References

[1] G. Pinter, G. Felde, I. Moseavi, A. Ghamisi, P., & Golgouen, R. (2020). COVID-19 pandemic prediction for Hungary: a hybrid machine learning approach. Hybrid Machine Learning Approach (May 2, 2020).

[2] P. Tariq, A. Ashatele, A., Kaliyaw, A., Oanunci, C., James, E., Gado, P., & Russell, M. (2020). Using supervised machine learning and empirical bayesian kriging to reveal correlates and patterns of covid-19 disease outbreak in sub-saharan africa: exploratory data analysis. Available at SSRN 3580721.

[3] A.S. Nanda, M.A.M. T. (2020). Fitting SIR model to COVID-19 pandemic data and comparative forecasting with machine learning. medRxiv.

[4] F. Satlari, J. Ma, P. Wagner, D. Neumann, M. Wenzel, R. Schäfer, W. Samek, Klaus-Robert Müller, T. Wiegand, Risk estimation of SARS-CoV-2 transmission from blueprints of low energy measurements.

[5] S. Tiwari, S. Kumar, K. Guleria, Outbreak trends of Coronavirus (COVID-19) in India: a prediction, Disaster Med. Public Health Prep. 14 (5) (2020) 1–9.

[6] M. Margi, L., & Doan, N.A. K. (2020). First-principles machine learning modelling of COVID-19. arXiv preprint arXiv:2004.09478.

[7] Sujatha, R., & Chatterjee, J. (2020). A machine learning methodology for forecasting of the COVID-19 cases in India.

[8] Jo, H., Son, H., Jung, S.Y., & Hwang, J.H. (2020). Analysis of COVID-19 spread in South Korea using the SIR model with time-dependent parameters and deep learning. medRxiv.

[9] Paggi, M. (2020). Simulation of Covid-19 epidemic evolution: are compartmental models really predictive?. arXiv preprint arXiv:2004.08207.

[10] Ponn, N.S., Soobhahad, S.K., & Agarwal, S. (2020). COVID-19 epidemic analysis using machine learning and deep learning algorithms. medRxiv.

[11] Z. Yang, Z. Zeng, K. Wang, S.S. Wong, W. Liang, M. Zanin, J. Liang, Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions, J. Thorac. Dis. 12 (3) (2020) 165.

[12] Liu, D., Clemente, L., Poirier, C., Ding, X., Chinnazii, M., Davis, J.T., & Santillana, M. (2020). A machine learning methodology for real-time forecasting of the 2019-2020 covid-19 outbreak using Internet searches, news alerts, and estimates from meta-predictive models. arXiv preprint arXiv:2004.04015.

[13] Das, S. (2020). Prediction of COVID-19 disease progression in India: under the effect of national lockdown. arXiv preprint arXiv:2004.03147.

[14] Ndaye, B.M., Tendler, L., & Seck, D. (2020). Analysis of the COVID-19 pandemic by SIR model and machine learning techniques for forecasting. arXiv preprint arXiv:2004.01574.P.

[15] Kumar, P., Kailta, H., Patairya, S., Sharma, Y.D., Nanda, C., Rani, M., & Bhagavathula, A.S. (2020). Forecasting the spread of COVID-19 pandemic in Top 15 countries in April 2020: ARIMA model with machine learning approach. medRxiv.

[16] R.M. Carrillo-Larch, M. Castillo-Cara, Using country-level variables to classify countries according to the number of confirmed COVID-19 cases: An unsupervised machine learning approach, Well-Being Open Res. 5 (2020) 56.

[17] Akshay Kumar, Farhan Mohammad Khan, Rajiv Gupta, Harish Puppala (2020) “Preparedness and Mitigation by projecting the risk against COVID-19 transmission Using Machine Learning Techniques’ medRxiv 2020.04.26.20080655.; doi: 10.1101/2020.04.26.20080655.

[18] Ministry of Health and Family Welfare, https://www.mohfw.gov.in/ and https://COVID19india.org portal, last access: 25th April 2020.

[19] European Centre for Disease Prevention and ControlPublic Health Management of Pandemics Having Had Compact With Novel Coronavirus in the European Union, European Centre for Disease Prevention and Control, 2020.

[20] B.J. Quilty, S. Clifford, S. Flasche, R.M. Eggo, Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV), Eurosurveillance 25 (5) (2020) 200088.

[21] C.R. Wells, P. Sah, S.M. Moghadas, A. Pandey, A. Shokut, Y. Wang, A.P. Galvani, Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak, Proc. Natl. Acad. Sci. 117 (13) (2020) 7504–7509.

[22] J.W. Glasser, N. Hupert, M.M. McCaigey, R. Hatchett, Modeling and public health emergency responses: lessons from SARS, Epidemics 3 (1) (2011) 32–37.

[23] WHO, C. (2015). Implementation and management of contact tracing for Ebola virus disease. WHO/EVD/Guidance/Contact/15.1. Setembro de.

[24] K.C. Swanson, C. Altare, C.S. Weshe, T. Nynsens, T. Ahmed, N. Elyman, N. Altmann, Contact tracing performance during the Ebola epidemic in Liberia, 2014-2015, PLoS Negl. Trop. Dis. 12 (9) (2018) e0006762.

[25] Public Health England MERS-CoV close contact algorithm. Public health investigation and management of close contacts of Middle East Respiratory Coronavirus (MERS-CoV) cases (v17 29 January 2019) 2019. https://assets.publishing.service. gov.uk/government/uploads/system/uploads/attachment_data/file/776218/ MERS-CoV_Close_contacts_algorithm.pdf.19.

[26] M. Kang, T. Song, H. Zhong, J. Hou, J. Wang, J. Li, Y. Zhang, Contact tracing for imported case of Middle East respiratory syndrome, China, 2015. Emerg. Infect. Dis. 22 (9) (2016) 1644.

[27] T.T. Hoang, V.N. Nguyen, N.S. Dinh, G. Thwaites, T.A. Nguyen, H.R. van Doorn, H.F. Herwurm, Active contact tracing beyond the household in multidrug resistant tuberculosis in Vietnam: a cohort study, BMC Public Health 19 (1) (2019) 1–8.

[28] H.C. Johnson, C.M. Gosner, E. Colzani, J. Kinnsman, L. Aleskini, J. Beant, K. Eidahl, Point of care scenarios for the progression of a COVID-19 epidemic in the European union and the European economic area, March 2020, Eurosurveillance 25 (9) (2020) 2000202.

[29] A.J. Kucharski, A. Camacho, F. Checchi, R. Waldman, R.F. Grais, J.C. Cabrol, W.J. Edmunds, Evaluation of the benefits and risks of introducing Ebola community care centers, Sierra Leone, Emerg. Infect. Dis. 21 (3) (2015) 393.
[59] C. Fraser, S. Riley, R.M. Anderson, N.M Ferguson, Factors that make an infectious disease outbreak controllable, Proc. Natl. Acad. Sci. (2004) Peak, C. M., Childs, L. M., Grad, Y. H., & Buckee, C. O. (2017). Comparing nonpharmaceutical interventions for containing emerging epidemics. Proceedings of the National Academy of Sciences, 114(15), 4023-4028. nces, 101(16), 6146-6151.

[60] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, R. M. Eggo, Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts, Lancet Glob. Health 8 (4) (2020) e488–e496.

[61] F.M. Khan, R. Gupta, Arima and nar based prediction model for time series analysis of covid-19 cases in india, J. Saf. Sci. Resil. 1 (1) (2020) 12–18.