*Corresponding author:
tspark@stats.snu.ac.kr

Taewan Goo and Catherine Apio
contributed equally to this work.

Forecasting of the COVID-19 pandemic situation of Korea

Taewan Goo[1], Catherine Apio[1], Gyujin Heo[1], Doeun Lee[1],
Jong Hyeok Lee[2], Jisun Lim[3], Kyulhee Han[1], Taesung Park[1,2*]

[1]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 08826, Korea
[2]Department of Statistics, Seoul National University, Seoul 08826, Korea
[3]The Research Institute of Basic Sciences, Seoul National University, Seoul 08826, Korea

For the novel coronavirus disease 2019 (COVID-19), predictive modeling, in the literature, uses broadly susceptible exposed infected recoverd (SEIR)/SIR, agent-based, curve-fitting models. Governments and legislative bodies rely on insights from prediction models to suggest new policies and to assess the effectiveness of enforced policies. Therefore, access to accurate outbreak prediction models is essential to obtain insights into the likely spread and consequences of infectious diseases. The objective of this study is to predict the future COVID-19 situation of Korea. Here, we employed 5 models for this analysis; SEIR, local linear regression (LLR), negative binomial (NB) regression, segment Poisson, deep-learning based long short-term memory models (LSTM) and tree based gradient boosting machine (GBM). After prediction, model performance comparison was evelauated using relative mean squared errors (RMSE) for two sets of train (January 20, 2020–December 31, 2020 and January 20, 2020–January 31, 2021) and testing data (January 1, 2021–February 28, 2021 and February 1, 2021–February 28, 2021) . Except for segmented Poisson model, the other models predicted a decline in the daily confirmed cases in the country for the coming future. RMSE values' comparison showed that LLR, GBM, SEIR, NB, and LSTM respectively, performed well in the forecasting of the pandemic situation of the country. A good understanding of the epidemic dynamics would greatly enhance the control and prevention of COVID-19 and other infectious diseases. Therefore, with increasing daily confirmed cases since this year, these results could help in the pandemic response by informing decisions about planning, resource allocation, and decision concerning social distancing policies.

Keywords: COVID-19, deep learning, disease transmission, mathematical model, pandemics, statistical model

Introduction

The novel coronavirus disease 2019 (COVID-19) presents an important and urgent threat to global health. Since the outbreak in early December 2019 in the Hubei province of the People's Republic of China, the number of patients confirmed to have the disease has exceeded 118 million as the disease spread globally, and the number of people infected is probably much higher [1]. More than 2.6 million people have died from COVID-19 (up to 11 March 2021) [2]. Despite public health responses aimed at containing the disease and delaying the spread [3,4], several countries have been confronted with a critical care crisis, and more countries could follow [5].

To mitigate and suppress the burden of COVID-19 on the healthcare system, while also protecting the general public, especially the highly susceptible group of people, ro-

bust models that predict the prognosis of COVID-19 were urgently needed to support decisions about shielding, hospital admission, treatment, and population level interventions [6]. In this situation, prediction tools can help project different scenarios such as (1) number of possible confirmed (new) cases, (2) number of possible hospitalized cases, (3) number of possible death cases and so forth. As a consequence, prediction tools are useful for several different purposes [7].

Other features, such as social distancing, stay-at-home orders, use of facemasks or self-quarantine, travel restriction, and contact tracing could help predict what comes next. For better understanding, prediction models are important for better estimation about the disease and its possible threats such as the number of cases based on the level of severity can help determine the need of numbers of ventilators and other sophisticated medical equipment. Furthermore, countries need to shape their health system responses in accordance with the need [8]. Therefore, access to accurate outbreak prediction models is essential to obtain insights into the likely spread and consequences of infectious diseases. Governments and other legislative bodies rely on insights from prediction models to suggest new policies and to assess the effectiveness of the enforced policies [9].

For COVID-19, predictive modeling, in the literature, uses broadly susceptible exposed infected recoverd (SEIR/SIR), agent-based, curve-fitting models. Besides, machine-learning models that are built on statistical tools have widely been used too [7]. Here, we employ statistical models; segmented Poisson, negative binomial (NB), and local likelihood regression (LLR), mathematical model SEIR, deep-learning based model long short-term memory (LSTM), and tree based gradient boosting machine (GBM) for prediction of future COVID-19 pandemic situation of Korea. The COVID-19 daily confirmed cases of Korea was divided into two regions: capital area (Seoul metropolitan area) and non-capital area (non Seoul metropolitan area). Domestic which is the sum of Capital and non-capital areas was also analyzed (Fig. 1). The daily confirmed cases of these regions were then split into train (January 20, 2020–December 31, 2020 and January 20, 2020–January 31, 2021) and test (January 1, 2021–Febraury 28, 2021 and Febraury 1, 2021–February 28, 2021) datasets. The prediction performance of the above models were tested using relative mean square error (RMSE). RMSE takes the total squared error and normalizes it by dividing by the total squared error of a simple predictor. Thus, the smaller the RMSE value, the better the prediction performance of the model.

Therefore, with increasing daily confirmed cases since the beginning of 2021, in Korea and elsewhere, such models could help in the response to pandemic by informing decisions about planning, resource allocation, and decision of the social distancing.

Methods

COVID-19 confirmed cases data
The daily series of confirmed cases of COVID-19 for South Korea from January 20, 2020 to Febraury 28, 2021 was obtained from
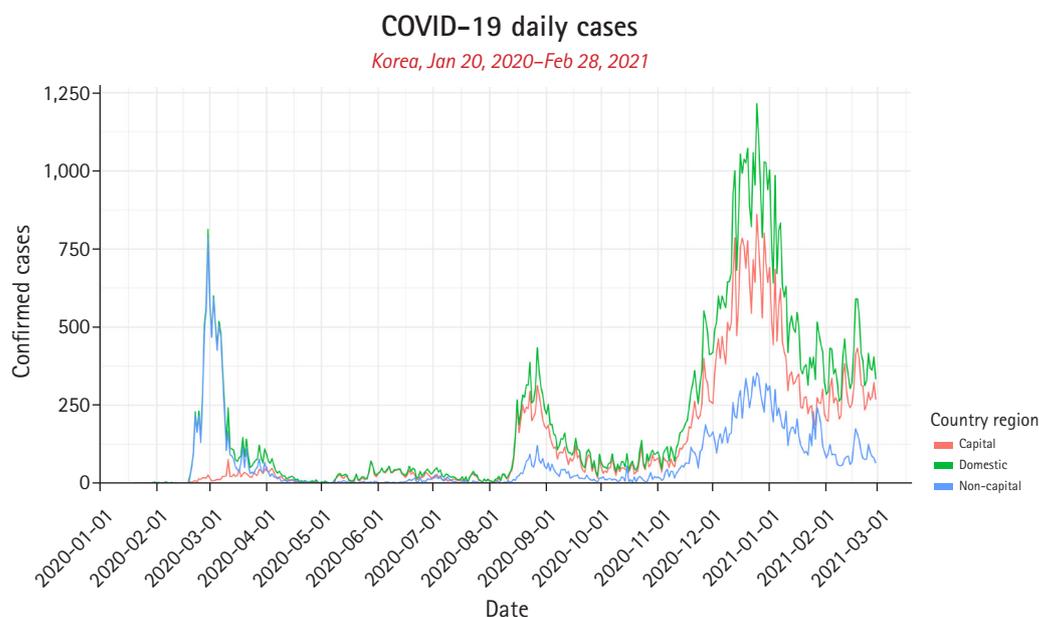

Fig. 1. Daily confirmed cases of South Korea. Daily confirmed cases of capital, non-capital, and domestic is represented in red, blue, and green, respectively.

Kaggle (from January 20 to June 30, 2020) [10] and Korea public data portal of the Ministry of Health and Welfare (from July 1, 2020 to February 28, 2021) [11]. The combined data was divided into two regions: capital or Seoul Metropolitan area (Capital; Seoul, Incheon, and Gyeonggi-do) and non-capital or non-Seoul Metropolitan area (non-capital; other cities beside Seoul, Incheon, and Gyeonggi-do). The analysis was conducted on the Domestic area (capital and non-capital), Seoul Metropolitan area, and non-Seoul Metropolitan area data, respectively. The data was split into two subsets. First subset is composed of training (January 20, 2020–December 31, 2020) and test data (the last 59 days, January 1, 2021–Febraury 28, 2021). And second subset is composed of training (January 20, 2021–January 31, 2021) and test data (the last 28 days, February 1, 2021–Febraury 28, 2021) for downstream analysis with the test data used for prediction analysis.

Prediction models

As one model may not give the best prediction of the COVID-19 situation of Korea, we present prediction results estimated by different models that can apply the above data. Many models are available and have been implemented for forecasting the pandemic situation of many countries and look at afew. In this section, we introduce segmented Poisson model, LLR model, deep-learning based LSTM model, NB model, SEIR model, and GBM model used for predicting the COVID-19 situation of Korea.

Segmented Poisson model

Here, we regarded the confirmed cases as a function of time t based on a segmented Poisson model. Let Y_t be the confirmed cases at day t which is the number of days since the first case occurred. Poisson model is defined as;

$$Y_t \sim poisson(\mu_t),$$

where μ_t is the expectation of Y_t with segments.

Breakpoints were considered in the daily confirmed cases during the analysis by splitting the daily confirmed cases into segments (Supplemantary Table 1). These breakpoints were decided using some of the aforementioned significant events linked to the spread of COVID-19 in South Korea. Since there are three breakpoints, four segments are defined as follows:

$$\begin{cases} \log(\mu_t) = \\ \beta_0 + \beta_{11}t + \beta_{21}\log(t+1), (t = 0, 1, ..., c_1 - 1) \\ \beta_0 + \beta_{11}t + \beta_{21}\log(t+1) + \beta_{12}(t-c) + \beta_{22}\log(t-c+1), (t = c_1, ..., c_2 - 1) \\ \beta_0 + \beta_{11}t + \beta_{21}\log(t+1) + ... + \beta_{13}(t-c) + \beta_{23}\log(t-c+1), (t = c_2, ..., c_3 - 1) \\ \beta_0 + \beta_{11}t + \beta_{21}\log(t+1) + ... + \beta_{14}(t-c) + \beta_{24}\log(t-c+1), (t = c_3, ..., n) \end{cases}$$

where c_i ($i = 1,2,3$) are breakpoints.

NB model

NB model is defined as [12];

$$Y_t | F_{t-1} \sim NB(\lambda_t, \phi),$$

where λ_t is the conditional expectation of Y_t given F_{t-1} as the history of the joint process $\{Y_t, \lambda_t : t \in \mathbb{N}\}$. Conditional mean and variance of Y_t are defined as;

$$E(Y_t | F_{t-1}) = \lambda_t$$
$$VAR(Y_t | F_{t-1}) = \lambda_t + \lambda_t^2 / \phi,$$

where ϕ is the dispersion parameter. And overdispersion parameter σ^2 is defined as $\sigma^2 = 1/\phi$. NBdistribution is defined as;

$$P\left(Y_t = y : |F_{t-1}\right) = \frac{\Gamma(\phi + y)}{\Gamma(y+1)\,\Gamma(\phi)} \left(\frac{\phi}{\phi + \lambda_t}\right)^\phi \left(\frac{\lambda_t}{\phi + \lambda_t}\right)^y$$

where $y = 0, 1, ... n$. For estimating λ_t, $l = \{1, 7, 21\}$ were used as lagged confirmed cases and $\log(\lambda_t) = \beta_0 + \Sigma_{i=1}^{L} \beta_i^* \log(Y_{t-li} + 1)$, were used for the model. For NB model, package 'tscount' were used to analyze the confirmed cases of South Korea as time series count data [13].

LLR model

Our LLR model is based on Poisson model which previously mentioned. For this model, local quadratic approximation is fitted within a smoothing window of bandwidth h, which is the number of the nearest past observations to be used in the local fit. We use tricube kernel of weight $W(u) = (1 - |u|^3)^3$ for each point. The local quadratic log-likelihood is defined as;

$$L_t(a) = \sum_{t=1}^{h} w_i(t)\, l\left(Y_t, a_0 + a_1\left(t_i - t\right) + a_2\left(t_i - t\right)^2\right),$$

where $w_i(t) = W\left(\frac{h/2}{t_i - t}\right)$ and l are the log-likelihood function based on Poisson distribution assumption. The local likelihood estimate is made by maximizing over the parameter $a = (a_0, a_1, a_2)^t$.

We utilize a rolling origin cross validation to select optimal bandwidth of the smoothing window [14]. Validation sets are divided at the local peaks of counts. Validation MSE is cumulated by each validation set's counts being predicted using past validations sets. The bandwidth with the smallest validation MSE is selected as optimal bandwith. And then using optimal bandwith we finally fitted LLR model. For LLR model [15], package 'locfit' was used [16].

Long short-term memory

Here, LSTM network is considered [17]. Let the input data X_t be as a set of vector consisting of Y_{t-h} to Y_{t-1} according to day t, where h is the bandwidth having values {7, 14, 21, 28, 35, 42, 49, 56}. Among these, the optimal h is selected using validation set, which is last 7 days of the training period. The data is normalized using minmax normalizer to transform data to be in the range of 0 to 1.

The LSTM architecture is described in Fig. 2 [18]. Each blocks in the model use current input value X_t with C_{t-1} and h_{t-1} to be trained. The C_{t-1} and h_{t-1} are the state and output of the last block, respectively. We assume four blocks with 64 units, each with a 0.2 dropout layer. The optimization is held using the adam optimizer to minimize the MSE during the training process.

Once the optimal model is built, it is applied to the test data for prediction. The prediction is performed sequentially by using the current output as the part of input of the next prediction. The analysis was performed using Python version 3.7.6, and 'keras' library.

SEIR model with least squares

The infectious disease dynamic can be formulated with a mathematical model. We consider the SEIR model to fit the dataset of COVID-19 daily confirmed cases and predict the incidence of COVID-19 epidemic in Korea. In SEIR model, population is divided in four groups: susceptible (S), exposed (E), symptomatic and infectious (I) and recovered (R) individuals. This model includes the spread of infection during the latent period. The latency of COVID-19 infection is biologically realistic. The SEIR model is defined by the following the system of ordinary differential equations [19-22]:

$$\frac{dS}{dt} = -\beta S \frac{I}{N}$$

$$\frac{dE}{dt} = -\beta S \frac{I}{N} - \kappa E$$

$$\frac{dI}{dt} = -\kappa E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

where β is the transmission rate, γ is the recovery rate, and $1/\kappa$ is the average incubation period. The initial condition of this model $S(0)$, $E(0)$, $I(0)$, $R(0)$ must satisfy the condition $S(0) + E(0) + I(0) + R(0) = N$, where N is the total population size. In data fitting, the unknown parameters in model were estimated by a least squares algorithm. The numerical simulation and analysis were performed in MATLAB 2020a.

Gradient boosting machine

GBM is a tree based machine learning algorithm that can be used for regression and classification problems. GBM consist of weak regression learner and decision trees. The decision tree uses the input value to determine which regression learner is best to make predictions.

Based on adaptive boosting algorithm, GBM can build a strong regression learner by iteratively combining a set of weak regression leaners. GBM use gradient descent for minimizing loss function of a strong regression learner. Like other boosting algorithms, GBM adds models into the tree using greedy style [23]:

$$F_m(x) = F_{m-1}(x) + \rho_m h_m(x),$$

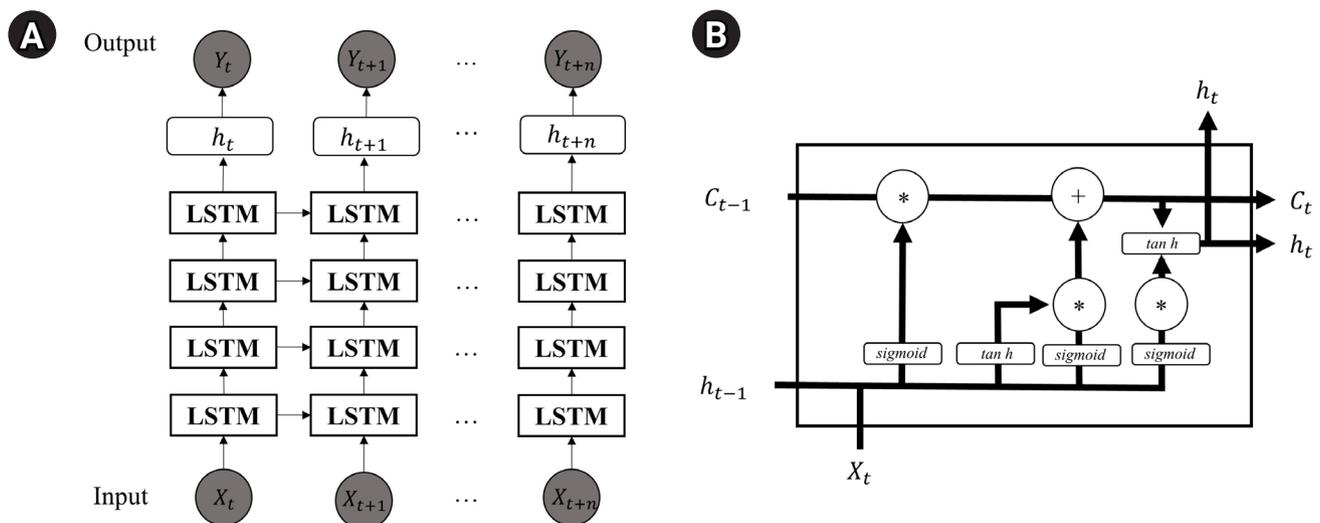where F_m is the updated model, F_{m-1} is previous model and $\rho_m h_m$ is

Fig. 2. Long short-term memory (LSTM) model architecture. (A) Overall architecture of LSTM. (B) The LSTM block architecture.

the newly added model. h_m is the trained base learner which minimizes the loss function L and ρ is the multiplier which is found by solving one dimensional optimization problem.

$$\rho_m = \underset{\rho}{argmin} \sum_{t=1}^{n} L\left(y_i, F_{m-1}\left(x_i\right) + \rho h_m\left(x_i\right)\right),$$

To build GBM, 'LightGBM' library was used [24].

Model assessment
To evaluate the above models, RMSEs for the train and test datasets for each of the fitted models were calculated as follows:

$$RMSE = \sum_{t=1,\dots,n} \frac{(\hat{\mu t} - yt)^2}{(y - yt)^2},$$

where n is the number of data points, y_t is the observed values, $\hat{\mu_t}$ is the predicted values from a fitted model and y^3 is the mean of observed values. To compare models predicting different regions, having different scale of confirmed cases, RMSE measure was chosen.

Results

The COVID-19 daily confirmed cases of the country were divided into two regions (non-capital and capital) with the total being domestic and analysed using the above models. The data was split into two subsets and used in the training and prediction analysis of the models.

As for model evaluation, in Table 1, for comparison of models in the whole country and the two regions, we observe that the train RMSE is always lower than the test RMSE, with the domestic region producing the highest RMSE values for all models. Also, the segmented Poisson model gives higher RMSE values when compared with other methods. With the first data subset: in the whole country (domestic), SEIR model and GBM had the lowest train RMSE values while NB and LLR had the lowest test RMSE values. In the Capital region, GBM and LLR have the lowest train RMSE while LLR and SEIR have the lowest test RMSE values, respectively. The non-capital region showed that SEIR and GBM have the lowest train RMSE while GBM and LLR have the lowest test RMSE values, respectively.

With the second data subset: in the country, LLR and SEIR had the lowest RMSE while NB and GBM had the lowest train RMSE values, respectively. Capital region showed that LLR and NB had the lowest train RMSE while NB and GBM had the lowest test RMSE values. In the non-capital region, SEIR and LLR had the lowest train RMSE while NB and GBM had the lowest test RMSE values.

Therefore, taking into lower train and test RMSE values for all

Table 1. RMSE for the regions and models following the the two data subsets

Region	Model	RMSE of data split 1		RMSE of data split 2	
		Train	Test	Train	Test
		(Jan 20, 2020–Dec 31, 2021)	(Jan 1, 2021–Feb 28, 2021)	(Jan 20, 2020–Jan 31, 2021)	(Feb 1, 2021–Feb 28, 2021)
Domestic	Segmented Poisson	0.088	1194.103	0.251	16.415
	Negative binomial	0.057	0.409	0.063	2.088
	Local regression	0.037	0.793	0.039	14.856
	LSTM	0.051	23.117	0.083	6.327
	SEIR	0.033	0.956	0.035	2.658
	GBM	0.022	1.507	0.082	0.591
Capital	Segmented Poisson	0.075	668.199	0.235	5.312
	Negative binomial	0.061	1.311	0.064	3.078
	Local regression	0.042	1.135	0.046	3.846
	LSTM	0.054	14.8	0.074	3.934
	SEIR	0.073	0.410	0.072	3.109
	GBM	0.021	1.960	0.095	0.892
Non–capital	Segmented Poisson	0.118	1131.838	0.195	34.935
	Negative binomial	0.097	0.912	0.103	1.157
	Local regression	0.074	0.522	0.076	33.232
	LSTM	0.087	15.207	0.119	4.774
	SEIR	0.036	0.610	0.036	1.964
	GBM	0.015	0.607	0.083	0.855

RMSE, relative mean squared error; LSTM, long short–term memory; SEIR, susceptible exposed infected recoverd; GBM, gradient boosting machine.

region and both data subsets, we can conclude that LLR model, GBM, SEIR model and then NB model were the best prediction models for forecasting of the COVID-19 situation of Korea. Segmented Poisson model tended to have the highest test RMSE values in all scenarios.

A look that the prediction plots of the these models shows that the daily COVID-19 confirmed cases will decline in the country (domestic), Seoul metropolitan (capital) and non-Seoul metropolitan (non-capital) areas using LLR, and NB models using the first data subset, while it will increase and stay constant using the segmented Poisson and LSTM models, respectively (Supplementary Figs. 1–3). With the second data subset, daily COVID-19 confirmed cases will decline in the three regions as predicted by NB, segmented Poisson and LSTM models, while it will increase in the country and non metropolitan areas but will decline in the metropolitan areas, using LLR model (Supplementary Figs. 4–6). The SEIR and GBM models shows a decrease in daily confirmed cases in the country and the two regions for all data subsets (Supplementary Figs. 7 and 8).

Discussion

The objective of our analysis was to predict the future COVID-19 situation of South Korea using daily confirmed cases. We employed six different models in this analysis and all models gave some different prediction results for different data subsets and regions. The evaluation of the models using RMSE showed that local likelihood rgression, GBM, SEIR and NB models had the lowest RMSE values, making them the best models, though LSTM gave better RMSE values compared to segmented Poisson model. LLR, GBM, SEIR, NB and LSTM models mainly predicted a decline in COVID-19 daily confirmed cases in the country and the two regions of Korea. We can reasonably take that these results portray the future situation of the country.

With the first dataset, NB, SEIR, and LLR, respectively showed the best test performance in domestic, capital, and non-capital areas, while with second dataset, GBM showed the best test performance for all regions. In case of NB model, we found that the coefficient of confirmed cases of the day before had the largest value. This means that confirmed cases of the day before can affect the prediction of most future confirmed cases. In case of GBM, we could obtain feature importance plots of the model (Supplementary Fig. 9). We discovered that the confirmed cases of the last day was the most important feature for all regions. Thus, the models using the confirmed cases of the past days seemed to perform better than other models without using such data. The parameters of the mathematical model SEIR can be easily interpreted as a rate or transition parameter. However, the local regression does have too many parameters to assign meaningful interpretation.

Note that the number of daily confirmed cases varied across regions of Korea, so we first fitted these models for each of the regions and compared them with the number of observed cases. However, the comparison of prediction models based on regional data was not convincing due to the small sample sizes. Instead of fitting the models for each region, we considered combining the two regions of capital and non-capital areas, which provided enough sample sizes. We also tried predicitng the number of deaths due to COVID-19. However, the data was not large enough (with only 1,669 deaths as of March 14, 2021) [25] to provide reliable fitted results from the models.

In our study, there is one challenge that predictions made reflect interventions in place at the time the model was developed. So, one can argue the influence of government intervention policies in the above observed results. However, our comparison result is still valid because all models reflected the same intervention effects. Actually, the Korean government has maintained a high level of social distancing with a ban in gatherings of more than five persons [26], in their efforts to lower the triple-digit number of daily confirmed cases that has been observed since the start of this year. According to Heo et al.'s study [27] on the COVID-19 situation of Korea, the Korean government social distancing policy was predicted to lead to a decrease in the daily confirmed cases observed in the country but with only segmented Poisson model which according RMSE value, did not perform well as compared to the other models. In future, we hope to control for the influence of government interventions using the other models, to give a whole picture of the future COVID-19 situation of the country.

A good understanding of the epidemic dynamic would greatly enhance the control and prevention of COVID-19 as well as other infectious diseases. Therefore, taking precaution when using prediction to support a decision, for example, return to work or lowering of the social distancing level, is highly encouraged too.

ORCID

Taewan Goo: https://orcid.org/0000-0001-9427-2290
Catherine Apio: https://orcid.org/0000-0003-0240-794X
Gyujin Heo: https://orcid.org/0000-0003-3499-2848
Doeun Lee: https://orcid.org/0000-0002-2791-8431
Jong Hyeok Lee: https://orcid.org/0000-0002-3573-1853
Jisun Lim: https://orcid.org/0000-0002-1058-3292
Kyulhee Han: https://orcid.org/0000-0002-7243-5711
Taesung Park: https://orcid.org/0000-0002-8294-590X

Authors' Contribution

Conceptualization: TP. Data curation: KH, GH. Formal analysis: TG, GH, DL, JHL, JL. Funding acquisition: TP. Methodology: TP. Writing - original draft: TG, CA. Writing - review & editing: TP, TG, CA.

Conflicts of Interest

Taesung Park serves as an editor of the Genomics and Informatics, but has no role in the decision to publish this article. All remaining authors have declared no conflicts of interest.

Supplementary Materials

Supplementary data can be found with this article online at http://www.genominfo.org.

References

1. Zhao H, Lu X, Deng Y, Tang Y, Lu J. COVID-19: asymptomatic carrier transmission is an underestimated problem. Epidemiol Infect 2020;148:e116.

2. Worldometer. United States: coronavirus cases. Worldometer, 2021. Accessed 2021 Mar 11. Available from: https://www.worldometers.info/coronavirus/.

3. Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. Nature 2020;584:262-267.

4. Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the effectiveness of worldwide COVID-19 government interventions. Nat Hum Behav 2020;4:1303-1312.

5. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328.

6. Sperrin M, McMillan B. Prediction models for COVID-19 outcomes. BMJ 2020;371:m3777.

7. Santosh KC. COVID-19 prediction models and unexploited data. J Med Syst 2020;44:170.

8. Centers for Disease Control and Preventions. COVID-19 mathematical modeling. Source: National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. Atlanta: Centers for Disease Control and Preventions, 2020. Accessed 2020 May 26. Available from: https://www.cdc.gov/coronavirus/2019-ncov/coviddata/mathematical-modeling.htm.

9. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, et al. COVID-19 outbreak prediction with machine learning. Algorithms 2020;13:249.

10. NeurIPS 2020: data science for COVID-19 (DS4C). DS4C: data science for COVID-19 in South Korea. San Francisco: Kagle, 2020. Accessed 2021 Mar 11. Available from: https://www.kaggle.com/kimjihoo/coronavirusdataset.

11. Korea Information Society Agency. Ministry of Health and WelfareCorona 19 City/ProvinceStatus. Daegu: Korea Information Society Agency, 2021. Accessed 2021 Mar 11. Available from: https://data.go.kr/data/15043378/openapi.do.

12. DeGroot MH. Probability and Statistics. 2nd ed. Reading: Addison-Wesley, 1986. pp. 258-259.

13. tscount: analysis of count time series. Comprehensive R Archive Network, 2021. Accessed 2021 Mar 11. Available from: https://cran.r-project.org/web/packages/tscount/index.html.

14. Tashman LJ. Out-of-sample tests of forecasting accuracy: an analysis and review. Int J Forecasting 2000;16:437-450.

15. Cleveland WS, Loader C. Smoothing by local regression: principles and methods. In: Statistical Theory and Computational Aspects of Smoothing. Contributions to Statistics (Hardle W, Schimek MG, eds.). Heidelberg: Physica-Verlag HD, 1996. pp. 10-49.

16. locfit: local regression, likelihood and density estimation. Comprehensive R Archive Network, 2021. Accessed 2021 Mar 11. Available from: https://cran.r-project.org/web/packages/locfit/index.html.

17. Chimmula VK, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos Soliton Fract 2020;135:109864.

18. Chandra R, Jain A, Chauhan D. Deep learning via LSTM models for COVID-19 infection forecasting in India. Preprint at https://arxiv.org/abs/2101.11881 (2021).

19. Baily NT. The Mathematical Theory of Infectious Disease and Its Applications. 2nd ed. London: Griffin, 1975.

20. Hethcote HW. The mathematics of infectious diseases. SIAM Rev 2000;42:599-653.

21. Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton: Princeton University Press, 2008.

22. Diekmann O, Heesterbeek H, Britton T. Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton Series in Theoretical and Computational Biology. Princeton, NJ: Princeton University Press, 2013.

23. Gumaei A, Al-Rakhami M, Al Rahhal MM, Albogamy FR, Al Maghayreh E, et al. Prediction of COVID-19 confirmed cases using gradient boosting regression method. Comput Mater Continua 2021;66:315-329.

24. lightgbm: Light Gradient Boosting Machine. San Francisan: GitHub, 2021. Accessed 2021 Mar 22. Available from: https://

github.com/microsoft/LightGBM.

25. Korea Disease Control and Prevention Agency. Cheongju: Korea Disease Control and Prevention Agency, 2021. Accessed 2021 Mar 14. Available from: http://www.kdca.go.kr/cdc_eng/.

26. Yonhap News Agency. (3rd LD) S. Korea to impose nationwide ban on gatherings of 5 or more people in virus fight: PM. Seoul: Yonhap News Agency, 2020. Accessed 2021 Mar 11. Available from: https://en.yna.co.kr/view/AEN20201222001553315.

27. Heo G, Apio C, Han K, Goo T, Chung HW, Kim T, et al. Statistical estimation of effects of implemented government policies on COVID-19 situation in South Korea. Int J Environ Res Public Health 2021;18:2144.