



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.



Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

COVID-19 forecasting using new viral variants and vaccination effectiveness models

Essam A. Rashed^{a,*}, Sachiko Kodera^b, Akimasa Hirata^{b,c}^a Graduate School of Information Science, University of Hyogo, Kobe 650-0047, Japan^b Department of Electrical and Mechanical Engineering, Nagoya Institute of Technology, Nagoya 466-8555, Japan^c Center of Biomedical Physics and Information Technology, Nagoya Institute of Technology, Nagoya 466-8555, Japan

ARTICLE INFO

Keywords:

COVID-19
Forecasting
Deep learning
Vaccination effectiveness

ABSTRACT

Recently, a high number of daily positive COVID-19 cases have been reported in regions with relatively high vaccination rates; hence, booster vaccination has become necessary. In addition, infections caused by the different variants and correlated factors have not been discussed in depth. With large variabilities and different co-factors, it is difficult to use conventional mathematical models to forecast the incidence of COVID-19. Machine learning based on long short-term memory was applied to forecasting the time series of new daily positive cases (DPC), serious cases, hospitalized cases, and deaths. Data acquired from regions with high rates of vaccination, such as Israel, were blended with the current data of other regions in Japan such that the effect of vaccination was considered in efficient manner. The protection provided by symptomatic infection was also considered in terms of the population effectiveness of vaccination as well as the vaccination protection waning effect and ratio and infectivity of different viral variants. To represent changes in public behavior, public mobility and interactions through social media were also included in the analysis. Comparing the observed and estimated new DPC in Tel Aviv, Israel, the parameters characterizing vaccination effectiveness and the waning protection from infection were well estimated; the vaccination effectiveness of the second dose after 5 months and the third dose after two weeks from infection by the delta variant were 0.24 and 0.95, respectively. Using the extracted parameters regarding vaccination effectiveness, DPC in three major prefectures of Japan were replicated. The key factor influencing the prevention of COVID-19 transmission is the vaccination effectiveness at the population level, which considers the waning protection from vaccination rather than the percentage of fully vaccinated people. The threshold of the efficiency at the population level was estimated as 0.3 in Tel Aviv and 0.4 in Tokyo, Osaka, and Aichi. Moreover, a weighting scheme associated with infectivity results in more accurate forecasting by the infectivity model of viral variants. Results indicate that vaccination effectiveness and infectivity of viral variants are important factors in future forecasting of DPC. Moreover, this study demonstrate a feasible way to project the effect of vaccination using data obtained from other country.

1. Introduction

The emergence of Coronavirus Disease-2019 (COVID-19) in late 2019 resulted in several changes in the daily routine of people and has become a significant cause of mortality worldwide, causing more than 5.9 million deaths [1]. Due to vaccination, the number of daily positive cases (DPC) has decreased in several countries. Although some countries have achieved high vaccination rates [2], other countries are far behind, with only a small proportion of their respective populations being vaccinated. This is mainly due to the lack of resources [3], vaccination hesitancy [4,5], or other related issues.

One of the first countries to vaccinate its population was Israel; however, relatively high DPC were reported in August 2021 despite the country's vaccination rate being above 68% [6]. One reason for this upsurge was attributed to the high transmissibility of the Delta variant [7] and the waning protection from vaccination [8], especially for those who have been vaccinated very early during the pandemic [9]. A similar trend was observed in the United Kingdom [10] and the United States [11]. The data obtained from countries with high vaccination rates would be useful in predicting the future potential in follow-up regions. However, it is difficult to manipulate data acquired from other regions considering the variations of the different influencing factors.

* Corresponding author.

E-mail address: rashed@gsis.u-hyogo.ac.jp (E.A. Rashed).

<https://doi.org/10.1016/j.complbiomed.2022.105986>

Received 1 February 2022; Received in revised form 28 June 2022; Accepted 14 August 2022

Available online 17 August 2022

0010-4825/© 2022 Elsevier Ltd. All rights reserved.

Here, we propose an efficient method based on deep learning framework to forecast COVID-19 of one area based on viral variant modeling and vaccination effectiveness using a framework for data projection obtained from other regions. Considering that more than 65% of countries still have vaccination rates below 70% [12], the proposed approach can learn the experience of countries with high vaccination rates in a way that can provide useful insights for other countries with low vaccination rates. Moreover, the developed model can provide a clearer understanding of potential booster shot requirements and when they should be administered. The contribution of this work can be summarized as follows:

- Development of new model that enable projection of vaccination effectiveness at population level from one country to another.
- Construct viral infectivity model that enable different pandemic spread considering percentage of the variant and potential relative infectivity.
- Optimize model parameters using data from Tel Aviv, where vaccination rate is relative high compared with other countries.
- Validation study of different data inputs in the accuracy of DPC forecasting within different pandemic waves such as spread, peak, and decay phases for three prefectures in Japan.

The remainder of this article is structured as follows. Section 2 discuss related work with emphasis on machine learning/deep learning approaches. In Section 3, the proposed method is discussed while data description is presented in Section 4. Different scenarios are discussed in Section 5 and achieved results are Section 6. Discussion and conclusion are in Sections 7 and 8, respectively.

2. Related work

Several national and regional projects are currently in progress to predict the infection during the COVID-19 pandemic [12,13]. The aim of such projects is to understand the pandemic data to improve medical resource allocation, intervention, and policy settings. Susceptible, Exposed, Infectious, and Recovered (SIR or SEIR) models have often been used to solve these problems [14,15], and several recent studies have demonstrated the robust abilities of machine learning approaches to adjust for realistic scenarios without forming strict assumptions [16]. In contrast, earlier studies did not consider the public's mobility [17], which has been clarified as a dominant factor characterizing new cases as a surrogate marker for social distancing [18,19]. In addition, the forecasts were limited to only a few days [20]. Recently, several studies have considered the data pattern change due viral variants, and we have underlined the difficulty of predicting when new variants will appear [21]. It is critical for any successful model to be able to predict the beginning of a new wave of infections and its potential magnitude. The difficulty in modeling the upsurge of cases may also be attributed to behavioral changes when the DPC are low. Machine learning and deep learning models was used for knowledge discovery and forecasting of different aspects associated with the COVID-19 pandemic [22–26]. A systematic review that summerize recent work on COVID-19 data analysis is in [27].

Although conventional projection strategies did not thoroughly consider the effects of vaccination, some studies did [28]. With the emergence of messenger RNA vaccines, the efficiency of vaccination in protecting against infection will dramatically improve. The weekly incidence of COVID-19 since administering the first vaccine dose started to decrease after two weeks, which further decreased after four weeks [29]. After the second dose, vaccine effectiveness reached 75%–95% after a few weeks [30,31]. However, vaccine efficiency may depend on the dominant viral variant [30]. For the Pfizer–BioNtech vaccine, its efficacy for health-care workers has been extensively examined [29] and complete vaccination is defined as two doses given 21 days apart. Therefore, data from regions with high vaccination rates and new variants may serve as important guides for forecasting potential risks in other regions.

3. Methods

3.1. Forecasting model

The forecasting model was designed using a multi-path long short-term memory (LSTM) neural network based on our earlier study detailed in [21]. The main difference of LSTM from conventional methods such as SIR/SEIR is that in the deep learning model, the number of variables (network features) are extremely large and can handle data non-linearity in a more efficient manner. The network parameters are optimized based on an ablation study [21]. The main data stream of the forecasting model and fine details on the training and testing phases are shown in Fig. 1 and network detailed architecture is in Fig. 2. This architecture is implemented in Wolfram Mathematica (R) version 12.1 installed on a workstation with four Intel (R) Xeon CPUs running at 3.60 GHz, with 128 GB of memory and three NVIDIA GeForce 1080 GPUs. Training is implemented through a set of networks, with each network trained to forecast a single indicator (DPC, serious cases [SC], hospitalized cases [HC], daily death cases [DC], or daily hospital discharged cases [CC]). The current models are trained to forecast different COVID-19 incidences (DPC, SC, HC, DC and CC) for 14 future days. Then, the estimated values are used as input again for further 14 days (day 15 to day 28) forecasting (recurrent data), and so on. This feature is demonstrated in Fig. 3 (Testing) where big arrows indicate the normal flow of data testing to get the estimation of (day 1 to day 14). The estimated values are feedback (small arrows) as input for forecasting of (day 15 to day 28) and further future estimation.

3.2. Adaptation model

The projection of the epidemic tendency in one country to other countries is not always successful as epidemic parameters and associated factors in different countries may not be consistent. The two models work well especially during the early stages of vaccination when the effects of vaccination are still unclear. The adaptation model is trained using a simplified combination of data wherein the target must learn the effects of vaccination within different stages of the pandemic. With this design, we can think of the *forecasting model* as the local scope network and the *adaptation model* as the global scope network. This strategy can efficiently enable the use of data of countries with high vaccination rates without considering local features. The data flow between the forecasting and adaptation models is shown in Fig. 1(c).

3.3. Vaccination effectiveness at population level

As the vaccination rate may not reflect the actual efficiency due to the variations in vaccines and waning protection over time [9], we proposed a representation of vaccination protection, which is defined as the vaccination effectiveness at the population level that considers the waning protection and is used as a metric for herd immunity. The vaccination effectiveness at the population level in each city or prefecture was assumed to be as follows:

$$E(d) = \sum_{i=0}^d \frac{\sum_{t=1}^T (N_t(d-i)e_t(i))}{P}, \quad (1)$$

$$e_t(i) = \max(0, \bar{e}_t(i)), \quad \bar{e}_t(i) = \begin{cases} a_t \frac{i}{K} & (i \leq K) \\ a_t - s(i - K) & (i > K) \end{cases} \quad (2)$$

where d is the day index and $N_t(d-i)$ denotes the number of people newly administrated with the t^{th} vaccine dose on day $(d-i)$. $e_t(i)$ is the non-negative individual vaccination effectiveness of the t^{th} dose on i days after inoculation. Parameters a and s are adjusted to reach an individual vaccination effectiveness peak within K days after inoculation, which then decrease linearly due to waning effect [32,33]. The waning

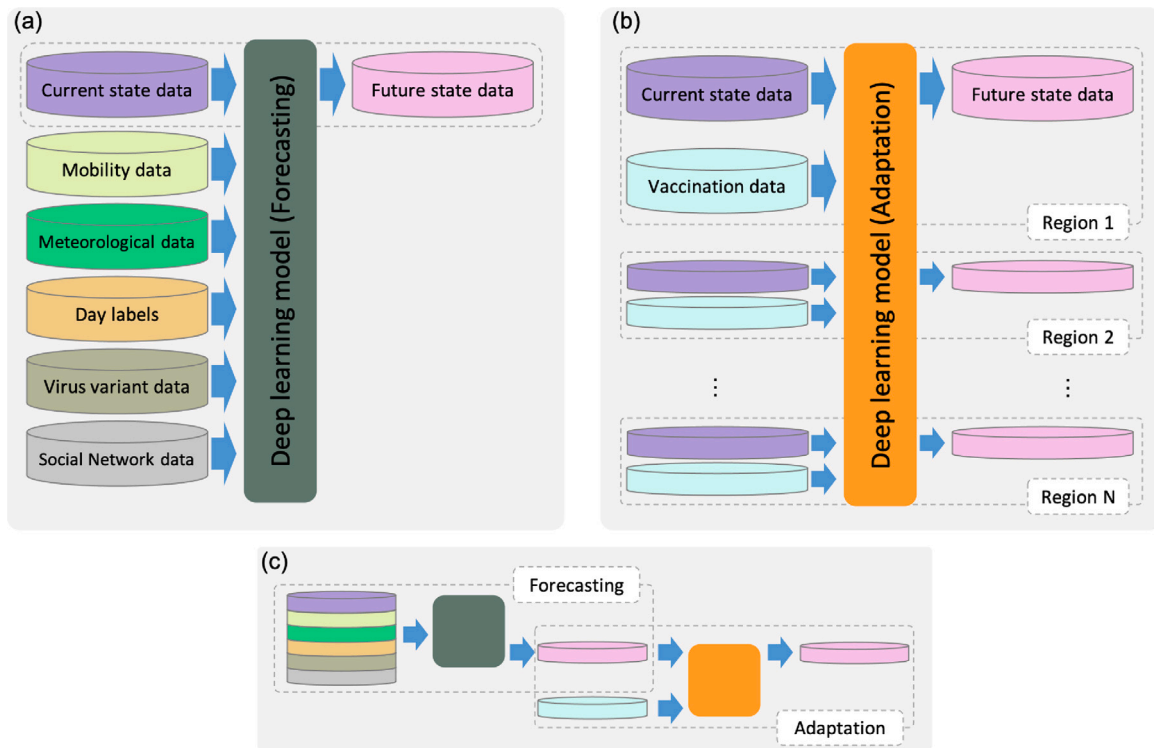


Fig. 1. Outline of proposed model for COVID-19 forecasting with vaccination effectiveness. (a) Initial forecasting is computed using a blend of time-series data; (b) the vaccination effect is computed using data acquired from different regions; and (c) the full model includes steps in (a) forecasting and (b) adaptation. Network detailed architecture is in Fig. 2.

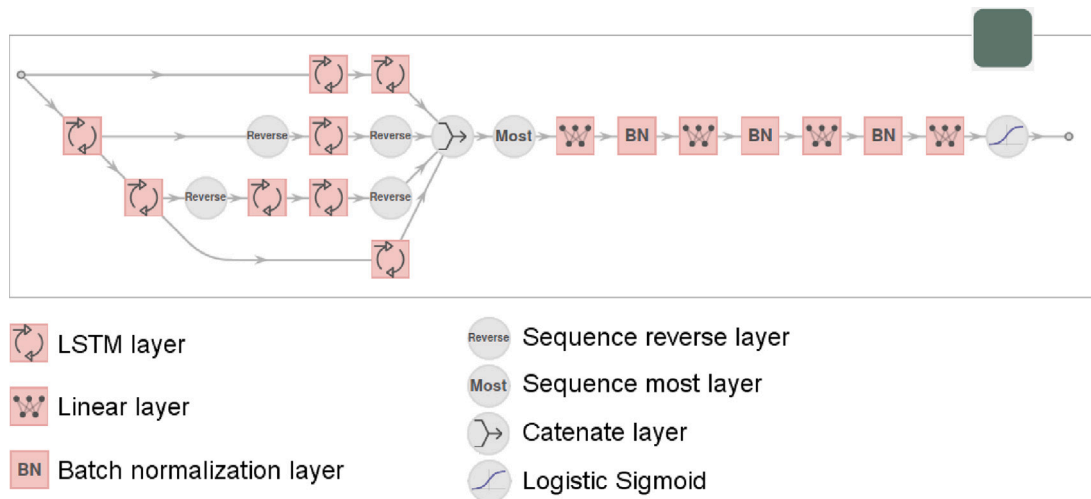


Fig. 2. LSTM network architecture.

effect was adjusted when people inoculated the second or third dose by considering the number of people vaccinated in the past (e.g., the number of second shot vaccinated people in past was reduced with increasing the number of people with third shot). To highlight this point an illustration demonstrate a simple example of population vaccination with different status of potential subjects is in Fig. 4. P is the population within urban region under consideration (P is considered a constant value during the time frame of this study).

We assumed $T = 3$ to demonstrate the number of vaccine doses (vaccines with a single dose, such as that of Johnson & Johnson, was not considered here) and $K = 14$ for the two-week latency period of the vaccination effect [34]. The parameters of a_1 and a_2 , characterizing the individual effectiveness of vaccination, for the Delta variant were chosen as 0.605 and 0.756, respectively, both of which are based on

a meta-analysis of systematic review (11 study groups) as detailed in [30]. These parameters coincide with the reported real-world vaccination effectiveness [35] and also consistent with computational estimation in Japan [36]. The parameters of s and a_3 are computed as explained below (please refer to Section 6.1). The antibody levels of infected people are comparable to or somewhat lower than those of fully vaccinated persons [37]; thus, the DPC is counted as additional value of fully vaccinated people.

3.4. Infectivity of viral variant

Different viral variants and mutations have been reported during the recent waves of infection of COVID-19. In addition, different variants have different rates of spread, infectivity, and resistance to the

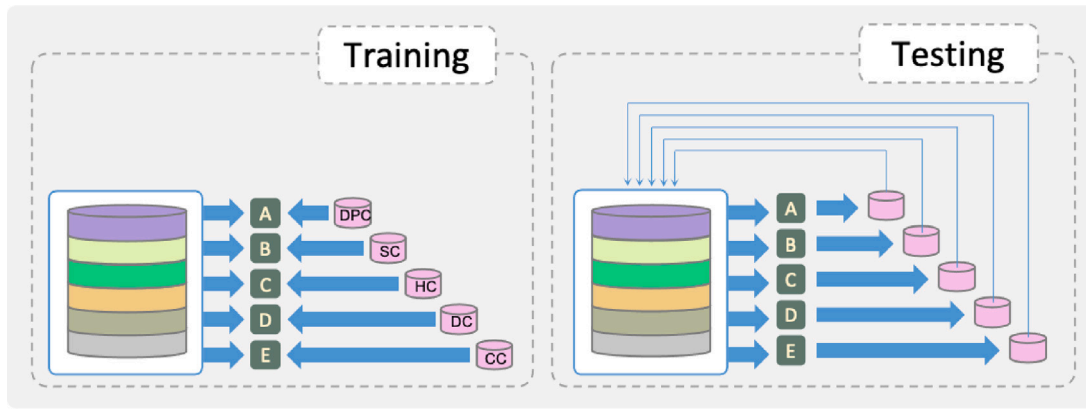


Fig. 3. Training and testing phases for the COVID-19 forecasting model. In training, different networks (A–E) are trained to forecast specific indicators. Long-term forecasting is achieved in the testing phase with concurrent data updates.

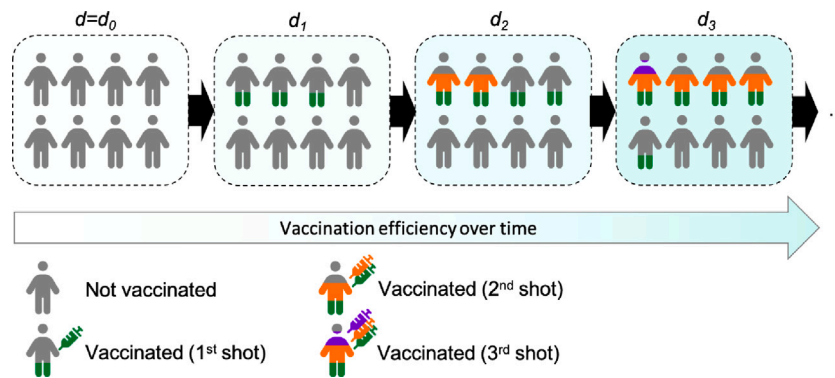


Fig. 4. Schematic explanation of the change in vaccination status with a sample population ($P=8$) over time. At $d = d_0$ $N_1 = N_2 = N_3=0$. At d_1 , $N_1=3$ and $N_2 = N_3=0$, at d_2 , $N_1 = N_2=2$ and $N_3=0$. Finally, at d_3 , $N_1=1$, $N_2=3$ and $N_3=1$.

currently administered vaccines. This effect has become significant since the emergence of the Delta and Omicron variants. Therefore, developing an infectivity model based on the dominant (or partially spreading) variant is necessary. The normalized infectivity index (\tilde{f}) is computed using the following equation:

$$\tilde{f}(d) = (\beta - \alpha) \frac{f(d) - \min(f)}{\max(f) - \min(f)} + \alpha \tag{3}$$

$$f(d) = \sum_{j=1}^N \omega_j v_{dj} \quad \forall d \tag{4}$$

where v_{dj} is the percentage of variant j at day d , ω_j is a weighting parameter assigned to each variant that demonstrates its relative infectivity, and parameters α and β are scaling parameters and $\max(f)/\min(f)$ are global values computed using all available measured data. As variant data were reported weekly, daily values were computed using linear interpolation. Therefore the normalized index of infectivity is an indicator of infectivity risk considering the percentage of different variant within study area.

3.5. Validation measurements

For quantitative evaluation, the average relative error over a period of N days was computed as follows:

$$Error = \frac{1}{N} \sum_{d=1}^N \frac{|y_d - \hat{y}_d|}{y_d} \tag{5}$$

where y_d and \hat{y}_d are the real and estimated positive cases on day d , respectively.

4. Data

The data combination used in the forecasting model includes the (i) current COVID-19 parameters (DPC, SC, HC, DC and CC), (ii) mobility data (retail & recreation, grocery & pharmacy, parks, transit stations, workplaces, and residences), (iii) meteorological data (daily maximum and minimum temperatures and daily average humidity), (iv) day labels (working days or holidays [i.e., weekends or national holidays]), (v) viral variant infectivity, and (vi) vaccination effectiveness. The main difference of this model from that from our previous study is that serious cases, hospitalized cases, and deaths were considered in item (i) in addition to items (v) and (vi). In addition, in the analysis of Tokyo, Osaka, and Aichi, the number of tweets and population at night were considered, which are potentially related to changes in public behavior. The effectiveness of the latter can be reported in [38]. The breakdown and definition of each dataset are listed in Table 1. Vaccination data, along with current COVID-19 data, were collected from external regions (Tel Aviv, Israel).

4.1. Input data for Japan

The COVID-19 data of Tokyo were obtained from online open sources maintained by the Japanese Ministry of Health, Labor, and Welfare (MHLW).¹ Mobility data were downloaded from the global Google mobility reports.² Meteorological data were obtained from the

¹ https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000164708_00079.html

² <https://www.google.com/covid19/mobility/>



Fig. 5. Map of Japan with study areas and regions used to represent downtown districts.

Japan Meteorological Agency.³ Day labels were based on the Japanese calendar, which were assigned as 1/0 for working/vacation days, respectively. Official state-of-emergency declarations by the Japanese government were assigned as 1/0 for active/inactive status, respectively. Information about the dominant variant was obtained from official MHLW reports.⁴ Vaccination rates were obtained from the Government CIO's Portal, Japan.⁵

In several earlier studies, public mobility was used as a key indicator for public interaction and social distancing (e.g. [18,19,39]). However, mobility data was criticized as it may not clearly indicate the social behavior, such as drinking parties, that is reported to be a potential major source of infection in Japan. Social networking data were obtained from Twitter, and mobility at downtown areas were computed as the nighttime population who stayed near restaurants and bars. Twitter data were used as it may indicate social activities where close contact occurs, and the downtown population was considered as several domestic reports have indicated that the main infection clusters may be due to close contacts in these areas. Tweets with keywords *BBQ*, *drinking party*, and *karaoke* (in Japanese) were chosen as risk-related terms. Data were obtained from NTT Data, INC. and processed by Toyoda Lab., University of Tokyo and shared through the Cabinet Secretariat COVID-19 AI Simulation Project [12]. When determining the number of tweeted keywords, those completed during the day or the previous day, or those planned until the next day, were extracted. While it is difficult to confirm if these gathering events are actually hold or not, recorded data can clearly indicate time frames where these

events are more popular. For corresponding tweets, information on the prefecture was extracted from the user's address. Note that due to the limited number of tweets, we only focused on three prefectures (Tokyo, Osaka, and Aichi); the number of tweets in other prefectures was generally lower, and the required number of tweets from other prefectures was not obtained. This is one reason why the analysis focused on these prefectures only. Three (Tokyo), two (Osaka), and one (Aichi) metropolitan areas were selected to represent the downtown districts with restaurants and bars (mesh size of 500m×500 m area) (see, Fig. 5).

4.2. Input data for Tel Aviv

The COVID-19 data and vaccination rate in Tel Aviv were obtained from online open sources,⁶ and mobility data were obtained from Google mobility reports. The average interval between the administration of the first and second doses was assumed as three weeks.

5. Scenarios

5.1. Optimize vaccination effectiveness using Tel Aviv data

The vaccination effectiveness calculated from Eq. (1) represents the model of protection from infection resulting from vaccination. With a variety of vaccines and other policy variables, parameters a_i and s should be adjusted based on real local data. For this purpose, we replicated the DPC in Tel Aviv and adjusted the parameters. Tel Aviv was selected as it has a high vaccination rate and shared similarities

³ <https://www.jma.go.jp/jma/indexe.html>

⁴ https://www.mhlw.go.jp/stf/seisakunitsuite/newpage_00054.html

⁵ https://cio.go.jp/c19vaccine_opendata

⁶ <https://info.data.gov.il/datagov/home/>

Table 1
Datasets used in the forecasting/adaptation models shown in Fig. 1.

#	Dataset	Items	Scale
1	Current state data	1-1 Positive cases 1-2 Serious cases 1-3 Hospitalized cases 1-4 Deaths 1-5 Hospital discharged cases	Daily number of cases
2	Mobility data	2-1 Retail & recreation 2-2 Grocery & pharmacy 2-3 Parks 2-4 Transit stations 2-5 Work places 2-6 Residents 2-7 Downtown area population	Percent change from baseline (pre-pandemic) Daily number of persons
3	Meteorological data	3-1 Maximum temperature 3-2 Minimum temperature 3-3 Average humidity	Daily value
4	Day labels	4-1 Working/holiday/ext. holiday 4-2 Normal/State of emergency	Labels (0/1/2)
5	Variant infectivity	5-1 $\tilde{f}(d)$	Computed using Eq. (3)
6	Behavior	6-1 Tweets (<i>nomikai</i>) 6-2 Tweets (<i>karaoke</i>) 6-3 Tweets (<i>BBQ</i>) 6-4 Downtown area population	Daily tweets using keywords Daily number of persons
7	Vaccination effectiveness	7-1 $E(d)$	Computed using Eq. (1)

with the vaccines used in Tokyo (BNT162b2). We then investigated three values for s (0.21, 0.24, and 0.27), which characterizes the waning protection from vaccines, and the efficiency of the third dose (booster) was represented by a_3 (0.75, 0.85, and 0.95). A verification study using training data from August 1, 2020 to July 30, 2021 and testing data from August 1 to September 23, 2021 was conducted to estimate s and a_3 . The optimum a_3 value was 0.95, which is consistent with that in the report of Pfizer and BioNTech⁷. Also, the slope of the waning protection was 0.24, which agrees with large scale study [11].

5.2. Exploring different input parameters for Tokyo, Osaka, and Aichi

We then explored parameters which will correlate well with the new DPC. The main factors that might potentially influence the DPC in the future are listed in Table 1. The viral variant is essential and can be extracted from the data in different countries, and the day label of “holiday” is potentially related to behavioral changes; both of which can be easily defined with any uncertainty, thus their use as the default parameters. There are different categories for mobility, including those in different urban regions. In our previous study, we have shown that in most prefectures, mobility at transit stations is an essential parameter, whereas remaining is also related to public activities in different urban regions. In addition, the nighttime population can identify social activities in regions where infection clusters were reported, which we compared using a new set of input parameters. Although weather was suggested to correlate with the number of DPC in some studies, others have denied this [40]. In this study, temperature and humidity, which are also related to the absolute humidity, were considered simultaneously. The vaccination effect was considered in the adaption of the neural network. To demonstrate the effectiveness of our proposed forecasting system, we applied the same scenarios for Tokyo to Osaka and Aichi. Input parameter optimization was then conducted to validate the accuracy of forecasts.

⁷ Pfizer and BioNTech Announce Phase 3 Trial Data (<https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-announce-phase-3-trial-data-showing>) [Accessed, March 5, 2022]

6. Results

6.1. Extraction of parameters characterizing the waning protection from vaccination and third dose

The parameters in Eq. (1) were revised to replicate new DPC in Tel Aviv. From Fig. 6, the observed and replicated DPC were in agreement when s and a_3 were set to 0.27 and 0.95, respectively. The average residual error of DPC from August 27 to November 21, 2021 was 0.289. Considering the incubation period (7–10 days), the efficiency of vaccination at the population level ranged from 0.29 to 0.32. The duration of vaccine effectiveness is plotted in Fig. 6(a). It is clear from the data presented in Fig. 6(c) that different vaccination models will lead to different estimations of the DPC.

6.2. Input parameters for DPC

After considering all input parameters, an extensive sensitivity study was conducted to minimize the input datasets. First, a single item was selected from each data set to minimize the total forecasting error. The seven selected inputs were then optimized to minimize possible number based on error. It was found that the optimized data inputs are those corresponded to the current DPC, mobility (transit station), model of the viral variant, twitter data *nomikai*, and vaccination effectiveness. The variant infectivity computed for Tokyo is shown in Fig. 7, with different weighting factors assigned to different viral variants. Fig. 8 shows the estimated DPC for the given parameters in Tokyo. The first set was generated using all datasets whereas the second estimation was obtained using optimized data which matched with the most accurate observed values. As shown in Table 2, the estimated DPC in Tokyo had error values of 0.23, 0.09, 0.78, and 0.36 for the spread, peak, decay, and all phases, respectively. The corresponding values for Osaka and Aichi were 0.24, 0.09, 0.41, 0.24, and 0.13, 0.16, 0.21, 0.17, respectively, which are highly consistent with the data of Tokyo. These results demonstrate that the viral variants model plays an important role during the spread phase in terms of starting time and peak value. In addition, the vaccination effectiveness model clearly contributes to the decay phase and can correctly forecast the rapid decay presented in the fifth COVID-19 wave in Japan.

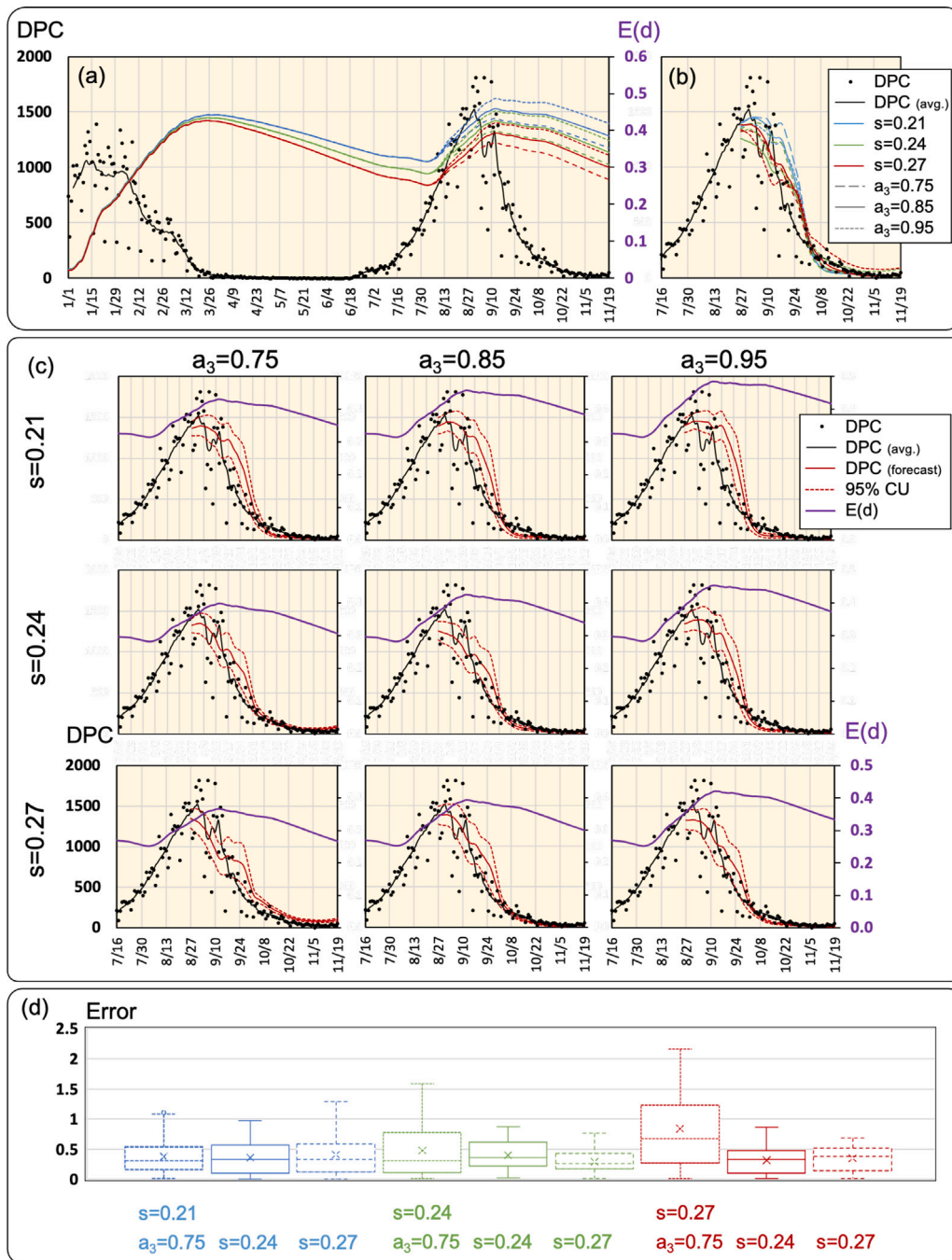


Fig. 6. (a) Vaccination effectiveness model (Eq. (1)) in Tel Aviv with different values of s and a_3 along with DPC. (b) Forecasted DPC (7-day average) using different vaccination effectiveness models during the decay of the COVID-19 wave. (c) Detailed forecasted DPC data including the 95% confidence interval and associated vaccination effectiveness model. (d) Error associated with the forecasts using different vaccination effectiveness models.

6.3. Adaptation of vaccination effectiveness modeling

The estimation of new DPC using the combination of forecasting and adaptation models is shown in Fig. 9. With the forecasting model alone, the spreading phase was highly consistent with real data; however, the decay phase was not. On the other hand, the combination of the two models resulted in more accurate results, especially in the decay phase, due to the application of the vaccination effectiveness acquired from the Tel Aviv data. The difference between the two models in the spread phase was marginal; however, it was significant in the decay phase. This tendency is similar to that of our previous study wherein the adaptation of new viral variants was discussed [21]. We found that different combinations of data as well as different time frames would

lead to significant changes in the results, especially those for the long-term forecasting. For example, in the early stages of the pandemic and prior to the emergence of viral variants, the meteorological data was suggested to highly correlate with the incidence of infection [40]. However, when new factors, such as vaccination effectiveness and viral variant infectivity, were considered, the importance of meteorological data was lessened. Table 2 demonstrates a brief assessment wherein a single dataset is excluded at a given time. This assessment was conducted in Tokyo, Osaka, and Aichi, and the learning period was from April 1, 2020 to June 30, 2021. Three time periods were included to demonstrate the spread (cases are increasing), peak (cases are at high values), and decay (cases are decreasing) of cases. Fig. 10 demonstrates forecasting during a pandemic state wherein positives, serious cases,

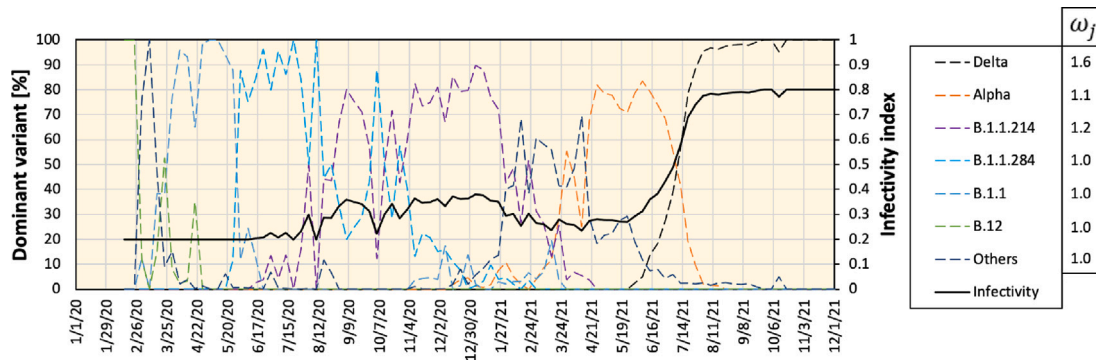


Fig. 7. Example of a variant infectivity index computed using data of viral variant measures in Tokyo with associated weight values representing relative infectivity.

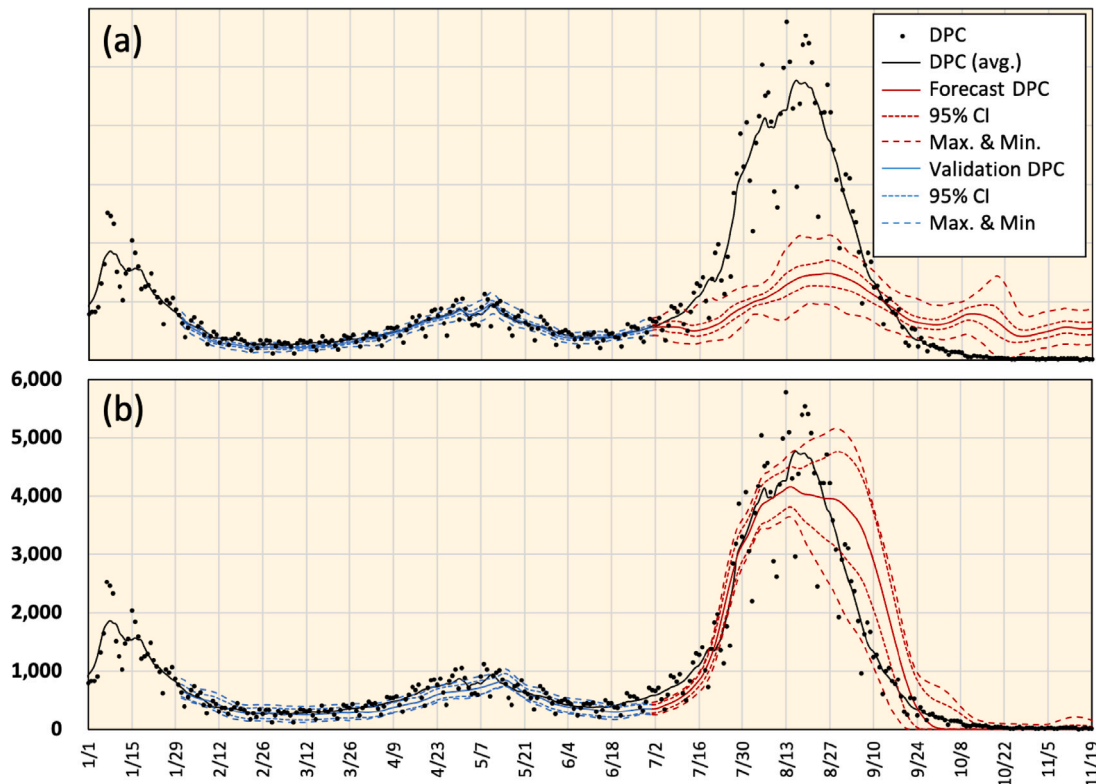


Fig. 8. Predicted DPC in Tokyo for the fifth wave with (a) all datasets (included in Table 1) and (b) optimized datasets (only values of 1-1, 2-4, 5-1, 6-1, 7-1 from Table 1) along with true values. Training data are from August 1, 2020 to July 30, 2021.

and deaths, for Tokyo, Osaka, and Aichi are forecasted. These results indicate that DPC and serious cases can be estimated with high accuracy while deaths are not. This might be attributable to the large variability of time-series data, which make it difficult for the LSTM network to learn the data pattern.

7. Discussion

In the early months of the COVID-19 pandemic, it was possible to estimate the DPC using only a small number of factors; however, after considering large-scale vaccination campaigns and the emergence of different variants, DPC estimation has become complicated. Regarding vaccination effectiveness, the effect of vaccination is not direct; hence, it needs to be carefully modeled by considering the variations in the vaccine type and potential waning of protection. Here, we present a method where the vaccination effect in one country can be projected to another country. Specifically, we have used vaccination data from Israel to train adaptation model (Fig. 1(a)), which is used to adjust forecasting

results of Japan obtained from forecasting model (Fig. 1(b)). For new viral variants, it is crucial to model its infectivity to correctly estimate such that the trigger of new wave and potential peak can be correctly estimated. Several deep learning approaches have been developed for COVID-19 forecasting using different factors such as current pandemic state, meteorological data, public mobility, and others. However, the trend of forecasting is changing with the wide administration of the vaccines and the potential higher infectivity of new variants. These new factors are hardly being used in previous models due to the data limitations. We studied forecasting using different set of inputs that demonstrate varied factors discussed in the literature such as public mobility and behavior, meteorological data, vaccination effectiveness, and viral variant infectivity. The results demonstrate that different parameters have different implications along a given time course. Therefore, the training data should be carefully selected to obtain highly accurate long-term forecasts. We presented a feasible method to project the vaccination effectiveness obtained from another country and a model to manage the change in the infectivity of viral variants. As

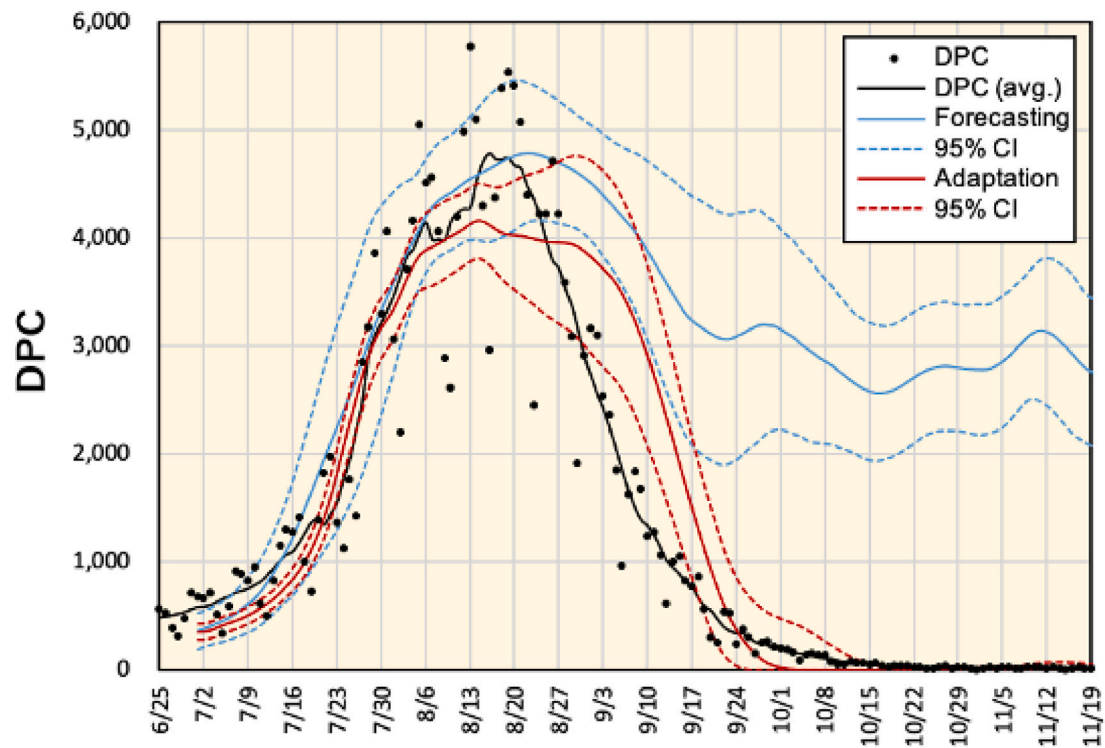


Fig. 9. Comparison of forecasting and adaptation models (shown in Fig. 1) for DPC in Tokyo. This demonstrates the effect of vaccination, which shows a weak spread phase and prolonged decay phase of the fifth COVID-19 wave.

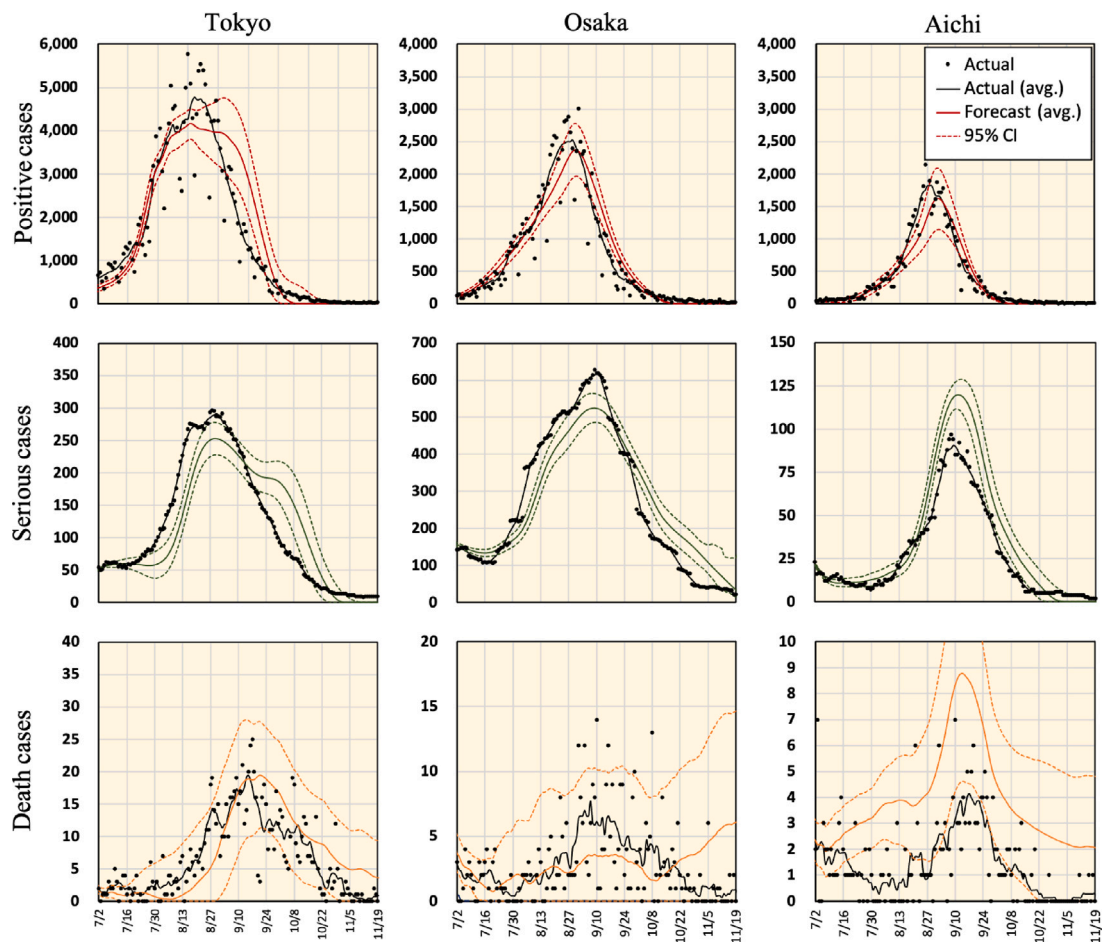


Fig. 10. Forecasting of DPC, serious cases, and deaths in Tokyo, Osaka, and Aichi with considering optimized input data and both forecasting and adaptation models.

Table 2

Errors computed in the forecasts of separate phases of the fifth COVID-19 wave in Tokyo, Osaka, and Aichi with different data sets. Each experiment was conducted by excluding a single dataset (1–7) in Table 1. None demonstrates the case wherein all datasets are used, and *Optimized* demonstrates the best scenario. Green and red colors are the lowest and highest error values, respectively.

		Ex.1	Ex.2	Ex.3	Ex.4	Ex.5	Ex.6	Ex.7	None	Opt. ^a
Tokyo	Spread ^b	0.48	0.30	0.26	0.18	0.41	0.53	0.43	0.44	0.23
	Peak	0.64	0.61	0.45	0.56	0.74	0.65	0.63	0.68	0.09
	Decay	0.79	1.06	1.88	0.63	0.39	0.61	0.78	0.53	0.78
	Full wave	0.63	0.65	0.85	0.46	0.52	0.62	0.61	0.55	0.36
Osaka	Spread	0.21	0.39	0.25	0.26	0.38	0.41	0.35	0.38	0.24
	Peak	0.81	0.76	0.69	0.56	0.82	0.75	0.73	0.75	0.09
	Decay	0.55	0.44	0.42	0.52	0.36	0.37	0.33	0.63	0.41
	Full wave	0.51	0.53	0.45	0.45	0.52	0.53	0.47	0.58	0.24
Aichi	Spread	0.83	0.15	0.71	0.20	0.71	0.48	0.14	0.17	0.13
	Peak	0.86	0.75	0.55	0.79	0.55	0.44	0.49	0.49	0.16
	Decay	0.95	0.85	0.96	0.65	0.85	0.92	1.08	0.60	0.21
	Full wave	0.88	0.58	0.74	0.54	0.70	0.60	0.57	0.42	0.17

^aOpt.: Optimized inputs are: 1-1, 2-4, 5-1,6-1 and 7-1 (see, Table 1).

^bSpread (July), peak (Aug.), decay (Sept.), and all wave (July–Sept.).

many countries have vaccination rates still below the target threshold for herd immunity set by the World Health Organization, it would be useful to validate the potential risks and forecast future waves of infection using data from other regions with high vaccination rates.

While the present study demonstrate a method to model vaccination effectiveness and viral variants for accurate forecasting of DPC, it has several limitations to be listed. The data used here are obtained from two countries where mRNA-based vaccines are used. The performance of the proposed model is unknown if the vaccine development technology is different. Also, the percentage of viral variant used in this study is based on relatively small number of samples that might not be a representative of real distribution.

8. Conclusion

In this study, we demonstrate a new framework for including infectivity variation caused by different viral variants and potential protection caused by vaccination effectiveness. These factors in addition to other known correlated data such as meteorological data, public mobility and others are combined in two successive LSTM models. The first model is a local scope model (forecasting) that is trained by local data measurements. The second model is a global scope model (adaptation) that can be trained using external data and used to adjust forecasting results. This approach demonstrate high accuracy results to forecast DPC in three regions of Japan with vaccination data obtained from Israel. This approach can be used to forecast DPC in countries with low vaccination rate using measurements at other countries with high vaccination rate. Therefore, the scope of potential usage is large as average global vaccination rate is still low in most countries.

Declaration of competing interest

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

Acknowledgments

The authors would like to thank Professor Masashi Toyoda (Institute of Industrial Science, The University of Tokyo, Japan) for generating processed Twitter data based on raw data provided by NTT Data, INC (Japan).

References

- [1] WHO, 2022, Online <https://covid19.who.int> (Accessed 5-Mar-2022).
- [2] N. Lurie, M. Saville, R. Hatchett, J. Halton, Developing Covid-19 vaccines at pandemic speed, *N. Engl. J. Med.* 382 (21) (2020) 1969–1973, <http://dx.doi.org/10.1056/NEJMp2005630>.
- [3] O.J. Wouters, K.C. Shadlen, M. Salcher-Konrad, A.J. Pollard, H.J. Larson, Y. Teerawattananon, M. Jit, Challenges in ensuring global access to COVID-19 vaccines: production, affordability, allocation, and deployment, *Lancet* (2021) [http://dx.doi.org/10.1016/S0140-6736\(21\)00306-8](http://dx.doi.org/10.1016/S0140-6736(21)00306-8).
- [4] S. Machingaidze, C.S. Wiysonge, Understanding COVID-19 vaccine hesitancy, *Nat. Med.* 27 (8) (2021) 1338–1339, <http://dx.doi.org/10.1038/s41591-021-01459-7>.
- [5] A. Alamoodi, B. Zaidan, M. Al-Masawa, S.M. Taresh, S. Noman, I.Y. Ahmaro, S. Garfan, J. Chen, M. Ahmed, A. Zaidan, O. Albahri, U. Aickelin, N.N. Thamir, J.A. Fadhil, A. Salahaldin, Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy, *Comput. Biol. Med.* 139 (2021) 104957, <http://dx.doi.org/10.1016/j.compbiomed.2021.104957>.
- [6] B. Mizrahi, R. Lotan, N. Kalkstein, A. Peretz, G. Perez, A. Ben-Tov, G. Chodick, S. Gazit, T. Patalon, Correlation of SARS-CoV-2-breakthrough infections to time-from-vaccine, *Nature Commun.* 12 (1) (2021) 1–5, <http://dx.doi.org/10.1038/s41467-021-26672-3>.
- [7] M. Wadman, A grim warning from Israel: Vaccination blunts, but does not defeat delta, 2021, <http://dx.doi.org/10.1126/science.ab19630>, [Online; Accessed 5-Mar-2022].
- [8] D.S. Khoury, D. Cromer, A. Reynaldi, T.E. Schlub, A.K. Wheatley, J.A. Juno, K. Subbarao, S.J. Kent, J.A. Triccas, M.P. Davenport, Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection, *Nat. Med.* (2021) 1–7, <http://dx.doi.org/10.1038/s41591-021-01377-8>.
- [9] K. Sanderson, COVID vaccines protect against delta, but their effectiveness wanes, *Nature* (2021) <http://dx.doi.org/10.1038/d41586-021-02261-8>.
- [10] K.B. Pouwels, E. Pritchard, P.C. Matthews, N. Stoesser, D.W. Eyre, K.-D. Vihta, T. House, J. Hay, J.I. Bell, J.N. Newton, et al., Effect of delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK, *Nat. Med.* 27 (12) (2021) 2127–2135, <http://dx.doi.org/10.1038/s41591-021-01548-7>.
- [11] B.A. Cohn, P.M. Cirillo, C.C. Murphy, N.Y. Krigbaum, A.W. Wallace, Breakthrough SARS-CoV-2 infections in 620,000 US veterans, february 1, 2021 to august 13, 2021, 2021, <http://dx.doi.org/10.1101/2021.10.13.21264966>, MedRxiv.
- [12] Japanese cabinet secretariat. COVID-19 AI & simulation project, 2021, Online <https://www.covid19-ai.jp/en-us/> (Accessed 10-Jan-2022).
- [13] I.C.-. forecasting team, Modeling COVID-19 scenarios for the United States, *Nat. Med.* (2020) <http://dx.doi.org/10.1038/s41591-020-1132-9>.
- [14] S. He, Y. Peng, K. Sun, SEIR modeling of the COVID-19 and its dynamics, *Nonlinear Dynam.* 101 (3) (2020) 1667–1680, <http://dx.doi.org/10.1007/s11071-020-05743-y>.
- [15] J.M. Carcione, J.E. Santos, C. Bagaini, J. Ba, A simulation of a COVID-19 epidemic based on a deterministic SEIR model, *Front. Public Health* 8 (2020) 230, <http://dx.doi.org/10.3389/fpubh.2020.00230>.
- [16] A. Zeroual, F. Harrou, A. Dairi, Y. Sun, Deep learning methods for forecasting COVID-19 time-series data: A comparative study, *Chaos Solitons Fractals* 140 (2020) 110121, <http://dx.doi.org/10.1016/j.chaos.2020.110121>.
- [17] S.F. Ardabili, A. Mosavi, P. Ghamisi, F. Ferdinand, A.R. Varkonyi-Koczy, U. Reuter, T. Rabczuk, P.M. Atkinson, COVID-19 outbreak prediction with machine learning, *Algorithms* 13 (10) (2020) <http://dx.doi.org/10.3390/a13100249>.

- [18] S. Chang, E. Pierson, P.W. Koh, J. Gerardin, B. Redbird, D. Grusky, J. Leskovec, Mobility network models of COVID-19 explain inequities and inform reopening, *Nature* 589 (7840) (2021) 82–87, <http://dx.doi.org/10.1038/s41586-020-2923-3>.
- [19] E.A. Rashed, A. Hirata, One-year lesson: Machine learning prediction of COVID-19 positive cases with meteorological data and mobility estimate in Japan, *Int. J. Environ. Res. Public Health* 18 (11) (2021) <http://dx.doi.org/10.3390/ijerph18115736>.
- [20] B. Nikparvar, M. Rahman, F. Hatami, J.-C. Thill, et al., Spatio-temporal prediction of the COVID-19 pandemic in US counties: modeling with a deep LSTM neural network, *Sci. Rep.* 11 (1) (2021) 1–12, <http://dx.doi.org/10.1038/s41598-021-01119-3>.
- [21] E.A. Rashed, A. Hirata, Infectivity upsurge by COVID-19 viral variants in Japan: Evidence from deep learning modeling, *Int. J. Environ. Res. Public Health* 18 (15) (2021) <http://dx.doi.org/10.3390/ijerph18157799>.
- [22] L. Xu, R. Magar, A. Barati Farimani, Forecasting COVID-19 new cases using deep learning methods, *Comput. Biol. Med.* 144 (2022) 105342, <http://dx.doi.org/10.1016/j.combiomed.2022.105342>.
- [23] E.A. Rashed, S. Kodera, H. Shirakami, R. Kawaguchi, K. Watanabe, A. Hirata, Knowledge discovery from emergency ambulance dispatch during COVID-19: A case study of Nagoya city, Japan, *J. Biomed. Inform.* 117 (2021) 103743, <http://dx.doi.org/10.1016/j.jbi.2021.103743>.
- [24] F. Rustam, A.A. Reshi, A. Mehmood, S. Ullah, B.-W. On, W. Aslam, G.S. Choi, COVID-19 future forecasting using supervised machine learning models, *IEEE Access* 8 (2020) 101489–101499, <http://dx.doi.org/10.1109/ACCESS.2020.2997311>.
- [25] V.K.R. Chimmula, L. Zhang, Time series forecasting of COVID-19 transmission in Canada using LSTM networks, *Chaos Solitons Fractals* 135 (2020) 109864, <http://dx.doi.org/10.1016/j.chaos.2020.109864>.
- [26] V. La Gatta, V. Moscato, M. Postiglione, G. Sperli, An epidemiological neural network exploiting dynamic graph structured data applied to the COVID-19 outbreak, *IEEE Trans. Big Data* 7 (1) (2021) 45–55, <http://dx.doi.org/10.1109/TBDATA.2020.3032755>.
- [27] A. Heidari, N. Jafari Navimipour, M. Unal, S. Toumaj, The COVID-19 epidemic analysis and diagnosis using deep learning: A systematic literature review and future directions, *Comput. Biol. Med.* 141 (2022) 105141, <http://dx.doi.org/10.1016/j.combiomed.2021.105141>.
- [28] P. Wintachai, K. Prathom, Stability analysis of SEIR model related to efficiency of vaccines for COVID-19 situation, *Heliyon* 7 (4) (2021) e06812, <http://dx.doi.org/10.1016/j.heliyon.2021.e06812>.
- [29] S. Benenson, Y. Oster, M.J. Cohen, R. Nir-Paz, BNT162b2 mRNA Covid-19 vaccine effectiveness among health care workers, *N. Engl. J. Med.* 384 (18) (2021) 1775–1777, <http://dx.doi.org/10.1056/NEJMc2101951>.
- [30] B. Zeng, L. Gao, Q. Zhou, K. Yu, F. Sun, Effectiveness of COVID-19 vaccines against SARS-CoV-2 variants of concern: a systematic review and meta-analysis, *BMC Med.* 20 (2022) 200, <http://dx.doi.org/10.1186/s12916-022-02397-y>.
- [31] B.Y. Reis, N. Barda, M. Leshchinsky, E. Kepten, M.A. Hernán, M. Lipsitch, N. Dagan, R.D. Balicer, Effectiveness of BNT162b2 vaccine against Delta variant in adolescents, *N. Engl. J. Med.* 385 (22) (2021) 2101–2103, <http://dx.doi.org/10.1056/NEJMc2114290>.
- [32] S.Y. Tartof, J.M. Slezak, H. Fischer, V. Hong, B.K. Ackerson, O.N. Ransinghe, T.B. Frankland, O.A. Ogun, J.M. Zamparro, S. Gray, et al., Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study, *Lancet* 398 (10309) (2021) 1407–1416, [http://dx.doi.org/10.1016/S0140-6736\(21\)02183-8](http://dx.doi.org/10.1016/S0140-6736(21)02183-8).
- [33] N. Andrews, J. Stowe, F. Kirsebom, S. Toffa, T. Rieckard, E. Gallagher, C. Gower, M. Kall, N. Groves, A.-M. O'Connell, D. Simons, P.B. Blomquist, A. Zaidi, S. Nash, N. Iwani Binti Abdul Aziz, S. Thelwall, G. Dabrera, R. Myers, G. Amirthalingam, S. Gharbia, J.C. Barrett, R. Elson, S.N. Ladhani, N. Ferguson, M. Zambon, C.N. Campbell, K. Brown, S. Hopkins, M. Chand, M. Ramsay, J. Lopez Bernal, Covid-19 vaccine effectiveness against the Omicron (b.1.1.529) variant, *N. Engl. J. Med.* 386 (16) (2022) 1532–1546, <http://dx.doi.org/10.1056/NEJMoa2119451>.
- [34] J. Lopez Bernal, N. Andrews, C. Gower, E. Gallagher, R. Simmons, S. Thelwall, J. Stowe, E. Tessier, N. Groves, G. Dabrera, R. Myers, C.N. Campbell, G. Amirthalingam, M. Edmunds, M. Zambon, K.E. Brown, S. Hopkins, M. Chand, M. Ramsay, Effectiveness of Covid-19 vaccines against the b.1.617.2 (Delta) variant, *N. Engl. J. Med.* 385 (7) (2021) 585–594, <http://dx.doi.org/10.1056/NEJMoa2108891>.
- [35] D. Aran, Estimating real-world COVID-19 vaccine effectiveness in Israel using aggregated counts, 2021, <http://dx.doi.org/10.1101/2021.02.05.21251139>, MedRxiv.
- [36] S. Kodera, E.A. Rashed, A. Hirata, Estimation of real-world vaccination effectiveness of mRNA COVID-19 vaccines against delta and Omicron variants in Japan, *Vaccines* 10 (3) (2022) <http://dx.doi.org/10.3390/vaccines10030430>.
- [37] M. Bergwerk, T. Gonen, Y. Lustig, S. Amit, M. Lipsitch, C. Cohen, M. Mandelboim, E.G. Levin, C. Rubin, V. Indenbaum, I. Tal, M. Zavitan, N. Zuckerman, A. Bar-Chaim, Y. Kreiss, G. Regev-Yochay, Covid-19 breakthrough infections in vaccinated health care workers, *N. Engl. J. Med.* 385 (16) (2021) 1474–1484, <http://dx.doi.org/10.1056/NEJMoa2109072>.
- [38] M. Nakanishi, R. Shibasaki, S. Yamasaki, S. Miyazawa, S. Usami, H. Nishiura, A. Nishida, et al., On-site dining in Tokyo during the COVID-19 pandemic: Time series analysis using mobile phone location data, *JMIR MHealth UHealth* 9 (5) (2021) e27342, <http://dx.doi.org/10.2196/27342>.
- [39] R.B. Noland, Mobility and the effective reproduction rate of COVID-19, *J. Transp. Health* 20 (2021) 101016, <http://dx.doi.org/10.1016/j.jth.2021.101016>.
- [40] P. Majumder, P.P. Ray, A systematic review and meta-analysis on correlation of weather with COVID-19, *Sci. Rep.* 11 (1) (2021) 10746, <http://dx.doi.org/10.1038/s41598-021-90300-9>.