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.
Original Article

Can naive Bayes classifier predict infection in a close contact of COVID-19?
A comparative test for predictability of the predictive model and healthcare workers in Japan

Hideo Yoshikawa

Division in Chief, Health Promotion Division, Chuo Public Health Centre, 12-1, Akashichou, Chuo City, Tokyo 104-0044, Japan

ARTICLE INFO

Keywords:
COVID-19
Machine learning model
Naive Bayes classifier
Predictive model
Close contact

ABSTRACT

Background: Those who are found in close contact with COVID-19 patients and are also negative by polymerase chain reaction (PCR) test may act without waiting for the incubation period to elapse, can become infectious and spread the infection.

Methods: A machine learning model that can evaluate the risk of infection in close contact with COVID-19 patients within the incubation period from the contact status reported from the index case was created using posterior probabilities. To confirm actual predictability, a verification test was conducted on 169 new close contacts, and the machine learning model was compared with four experienced healthcare workers for the predictability.

Results: In a verification test, 33 of the 169 contacts were infected with COVID-19 during the incubation period, and 13 of 33 were negative on initial PCR test, after that the disease developed and their PCR test became positive. The machine learning model predicted the eventual infection in 12 of 13 patients who had negative results on the initial PCR test. In the verification test, the sensitivity of the machine learning model was 0.879 and the specificity was 0.588. The mean standard deviation of the sensitivity and the specificity of the four health care workers was 0.568 (0.230) for sensitivity and 0.689 (0.103) for specificity.

Conclusion: If it is possible to convey that individual risk of infection, the close contact may take suppressive action during the incubation period regardless of the result of the initial PCR test, thereby preventing secondary spread of infection.

1. Introduction

The COVID-19 virus, which was first confirmed in Wuhan, China in 2019, has spread worldwide, and as of September 2021, more than 230 million cases and 4.72 million deaths have been confirmed. Since the end of 2020, several vaccines for COVID-19 have been developed. As of September 2021, 5.8 billion have been vaccinated worldwide, but still, millions of people are newly affected every day around the world [1].

Moore et al. say that even if the most optimistic scenario for vaccination (85% effectiveness) is given, considering the spread of B.1.1.7 which is one of the variants of concern (VOC), the reproduction rate will remain at 1.58 (95% CI, 1.36–1.83) [2] without non-pharmaceutical interventions (NPI). New VOC are constantly emerging, and as of September 2021, B.1.617.2 is on the rise worldwide.

The Israeli Ministry of Health said that the BNT162b2 vaccine, provided by Pfizer-BioNTech, had 94.3% effectiveness in May. However, in late June, when above 90% of new cases were B.1.617.2, the effectiveness of the vaccine decreased to 64% [3], and in July, it further decreased to 39%; the effectiveness after six months of vaccination was 16% [4]. The vaccination rate was over 80% as of August 2021 in Israel; however, new cases have increased sharply since August, >500 times in three months [1].

Therefore, suppressing the spread of infection only through vaccination is challenging and the development of oral antiviral drugs in mild cases should be promoted. In addition to pharmaceutical interventions, continuing NPIs are necessary.

Chan et al. investigated the effectiveness of NPIs internationally, and the most effective countermeasures are risk communication, followed by personal protective measures, such as social distancing, and thirdly, national measures with the allocation of financial and human resources to non-health systems such as aid, taxation, and lockdown; the fourth measure is related to case identification and contact tracing [5].

E-mail address: kenko_m@city.chuo.lg.jp.
2. Materials and methods

In May 2021, a machine learning model was created using the naïve Bayes classifier, which can evaluate the risk of an individual having close contact within the incubation period from the contact status reported by a COVID-19 patient.

According to the Infectious Diseases Control Law, all close contacts were interviewed immediately after the confirmation of the index case and underwent an initial PCR test within a few days; however, after the incubation period, repeated PCR tests were not performed to confirm negative results. These data were converted into unlinkable anonymous data that could not be used to identify specific individuals, and these data were used in this study.

Based on the criterion of WHO in March 2020, the criteria for close contacts were those who had contact with a COVID-19 patient within 1 m for more than 15 min without using proper personal protective equipment. However, even if the criteria were not completely met, irrefutable close contact, such as spending 10 min with a lover, was targeted.

Microsoft Excel 2016 was used to create a spreadsheet so that people engaged in contact tracing can easily handle the data. The free statistics software ‘R 4.0.3’ was used for plotting the receiver operating characteristic (ROC) curve. To build the machine learning model, the data of 1130 close contacts in Chuo City, Tokyo in 2020 were used.

Six explanatory variables used for predictions are listed in Table 1. The age of the contact was introduced because the sensitivity to COVID-19 in children and adolescents was low (0.56) compared to that in adults [12]; the age was set in increments of 10 years. Regarding the relationship between COVID-19 patients and the contacts, and contact situation, this information is an index to measure the contact density. Additionally, it is considered that this information affects the contact frequency and the amount of virus exposure, which was used by Shams et al. [10] Regarding the relationships, there were seven items: family living together/spouse, relatives, colleague, friend, supplier and customer, medical/nursing care provider and recipient and other, in consideration of intimacy and the type of contact. Regarding the contact

Table 1

| Variable description | Variable options |
|----------------------|------------------|
| 1 Contact’s age (yrs. old) | 1. 0–9 2. 10–19 3. 20–29 4. 30–39 5. 40–49 6. 50–59 7. 60–69 8. >70 |
| 2 Relationship to patient | 1. Family living together or spouse 2. Relatives 3. Colleague 4. Friend 5. Supplier and customer 6. Medical/nursing care provider and recipient 7. Other |
| 3 Contact’s situation | 1. Living together 2. Eating and drinking (with alcohol) 3. Eating and drinking (without alcohol) 4. Office/meeting room 5. Break room 6. School 7. Sports or recreations 8. Other |
| 4 Wearing a mask | 1. Yes 2. No |
| 5 Distance to the patient | 1. Within 1 m 2. Farther than 1 m |
| 6 Contact time | 1. More than 15 min 2. Within 15 min |
situation, there are eight items, such as living together, eating and drinking (with alcohol), eating and drinking (without alcohol), office/meeting room, break room, school, sports/recreation and others. Whether a mask is worn, contact distance and contact time are set based on the WHO criterion for close contact [9] (March 20, 2020). There are two choices for each criterion: the mask is used or not, the contact distance is within 1 m or > 1 m and the contact time is within 15 min or > 15 min.

Location was also examined; however, location was not adopted as an explanatory variable because the difference was not large compared to other variables, and it was unclear whether it was universal.

This study is based on Bayes' theorem, and from the explanatory variable data, (D) is obtained from those who finally infected (H). In the comparative test for predictability between the machine learning model and the four healthcare workers in 2021, 33 of the 169 close contacts were infected with COVID-19 during the incubation period. Thirteen of 33 were negative according to the PCR test performed within a few days of the last contact with a COVID-19 patient but were infected with COVID-19 within the incubation period of 2 weeks and finally had a positive PCR test result. The machine learning model predicted the final infection in 12 of 13 patients who were negative according to the initial PCR test. The final sensitivity of the machine learning model in the verification test was 0.879, specificity 0.588 and AUC 0.790 (95% CI, 0.715–0.864). The (sensitivity, specificity) of the four healthcare workers are (0.667, 0.559), (0.182, 0.868), (0.636, 0.706) and (0.788, 0.625) (Fig. 1), respectively, and the mean–standard deviation of the four healthcare workers is 0.568 (0.230) for sensitivity and 0.689 (0.103) for specificity. The positive predictive value for each risk category was ‘very high’ 0.667 (2/3), ‘high’ 0.338 (27/80), ‘middle’ 0.075 (3/40), and ‘low’ 0.022 (1/46).

Table 2 shows the breakdown of each explanatory variable in both the training data of the machine learning model and subjects in the verification test. The items of explanatory variables with different relationships and situations were set in consideration of the degree of contact, but since infection prevention measures were thoroughly implemented among medical workers, few were targeted as close contacts. In addition, in society, messages to avoid dinner and recreation were repeatedly broadcasted, and as a result of people restraining their behaviour, the number of relatives and customers was small.

Table 3 shows a comparison of sensitivity, specificity, infection risk groups and AUC on the ROC curve between the training data and verification test. During the verification test, the Predict Score at the time of creating the prediction model was used as the cut-off value without updating the data to prevent the criteria from changing.

The ROC curve in the demonstration test is shown in Fig. 2.

4. Discussion

The created machine learning model showed equal or better predictability with experienced healthcare workers’ involvement in contact tracing. Along with the explanatory variables, the healthcare workers grasped the surrounding situation of each close contact through contact tracing, and while there was a lot of information in making a judgement, it may have been a factor that created a cognitive bias.

The reason for not adopting a simple comparison of the posterior probabilities of infection and non-infection as the criteria for the machine learning model is that it was necessary to maximise the sensitivity and specificity to improve the accuracy of the machine learning model. Therefore, using the posterior probabilities of infection and non-infection, an index called ‘Predict Score’ was created, and based on this result, the ROC curve was created, and the optimum cut-off value was obtained from the Youden index.

Further, the machine learning models predicted the infection of most of the close contacts who had a negative initial PCR test and subsequently infected during the incubation period. This difference is due to the difference in purpose; the PCR test examined the presence or absence of the viral gene at that time, whereas the machine learning model aimed at the presence or absence of infection within the two-week incubation period.

The reason why the sensitivity and the specificity obtained in the verification test were superior to that of the training data was that the machine learning model was created in May 2021 from the close contact data in 2020. Since it was created after the fact, it is a summary of the notations that are similar in classification from the survey results that have already been conducted, and thus the strict target conditions were applied to 169 new close contacts from 26 May to July 20, and the cut-off value calculated at the time of model creation was used between the machine learning model and the four healthcare workers who had at least one year of experience in the contact tracing of COVID-19. After interviewing a patient and before the PCR testing of the close contacts, the healthcare workers shared the information of the contact situation and each healthcare worker predicted whether each close contact would be infected during the incubation period. After all healthcare workers had completed their predictions, the predictive model made the predictions, and after the incubation period of the close contacts, the results were compared.

The created machine learning model was based on the start of vaccination, so vaccinated persons were excluded from the verification test. Note that this study was approved by the Ethics Committee of Chuo Public Centre.

The created machine learning model is sequentially updated to optimise the judgement criteria when the data is added; however, to prevent changes in the criteria, the data was not updated during the verification test. The cut-off value calculated at the time of model creation was used. The subjects selected in the verification test were within one close contact per patient to prevent a large number of contacts with the same patient in large-scale cluster cases.

3. Results

The training data of the machine learning model created from the close contact data in 2020 had the sensitivity of 0.782 and the specificity of 0.413 at the Youden index on the ROC curve, and the area under the curve (AUC) was 0.642 (95%CI, 0.607–0.678). The positive predictive values in the four risk categories were ‘very high’ 0.474 (37/78), high 0.290 (180/620), middle 0.242 (56/231) and low 0.084 (17/201). In the comparative test for predictability between the machine learning model and the four healthcare workers in 2021, 33 of the 169 close contacts were infected with COVID-19 during the incubation period. Thirteen of 33 were negative according to the PCR test performed within a few days of the last contact with a COVID-19 patient but were infected with COVID-19 within the incubation period of 2 weeks and finally had a positive PCR test result. The machine learning model predicted the final infection in 12 of 13 patients who were negative according to the initial PCR test. The final sensitivity of the machine learning model in the verification test was 0.879, specificity 0.588 and AUC 0.790 (95% CI, 0.715–0.864). The (sensitivity, specificity) of the four healthcare workers are (0.667, 0.559), (0.182, 0.868), (0.636, 0.706) and (0.788, 0.625) (Fig. 1), respectively, and the mean–standard deviation of the four healthcare workers is 0.568 (0.230) for sensitivity and 0.689 (0.103) for specificity. The positive predictive value for each risk category was ‘very high’ 0.667 (2/3), ‘high’ 0.338 (27/80), ‘middle’ 0.075 (3/40), and ‘low’ 0.022 (1/46).

Table 2 shows the breakdown of each explanatory variable in both the training data of the machine learning model and subjects in the verification test. The items of explanatory variables with different relationships and situations were set in consideration of the degree of contact, but since infection prevention measures were thoroughly implemented among medical workers, few were targeted as close contacts. In addition, in society, messages to avoid dinner and recreation were repeatedly broadcasted, and as a result of people restraining their behaviour, the number of relatives and customers was small.

Table 3 shows a comparison of sensitivity, specificity, infection risk groups and AUC on the ROC curve between the training data and verification test. During the verification test, the Predict Score at the time of creating the prediction model was used as the cut-off value without updating the data to prevent the criteria from changing.

The ROC curve in the demonstration test is shown in Fig. 2.
not applied. Conversely, since the data to be collected was clearly
decided at the time of the verification test, it is considered that it was
strictly applied whether it meets the target requirements, which led to
greater accuracy. Even when newly collecting and operating data in
different countries, it is expected that the sensitivity and specificity will
be higher in the actual operation than in the training data.

Repeated PCR testing is not performed after the incubation period
ends; thus, capturing asymptomatic contacts who subsequently develop
asymptomatic infections among the initial PCR-negative contacts is
impossible. However, asymptomatic infections can be divided into pre-
symptomatic infections and those that remain asymptomatic, and it is
becoming clear that not many infected people remain asymptomatic.
Asymptomatic infections are also infectious; however, their infectivity is
much lower than that of symptomatic infections. Current studies [13]
have shown that the combined percentage of patients who are asympto-
matic at the time of initial testing is 15.6%, of which approximately
48.9% are infected with prodromal symptoms. The secondary clinical

Table 2
Data composition of the training data and verification test.

|                | Training data (n = 1130) | Verification test (n = 169) |
|----------------|--------------------------|----------------------------|
|                | Infection | Non-infection | Total | Infection | Non-infection | Total |
| Age            |           |              |       |           |              |       |
| 1. <10         | 39        | 134          | 173   | 4         | 20           | 24    |
| 2. 10–19       | 22        | 122          | 144   | 1         | 22           | 23    |
| 3. 20–29       | 43        | 123          | 166   | 6         | 23           | 29    |
| 4. 30–39       | 53        | 167          | 220   | 10        | 31           | 41    |
| 5. 40–49       | 54        | 123          | 177   | 5         | 22           | 27    |
| 6. 50–59       | 38        | 94           | 132   | 4         | 12           | 16    |
| 7. 60–69       | 15        | 43           | 58    | 2         | 5            | 7     |
| 8. >70         | 26        | 34           | 60    | 1         | 1            | 2     |
| Relation       |           |              |       |           |              |       |
| 1. Living togetherspouse | 237     | 543          | 780   | 29        | 60           | 89    |
| 2. Relatives   | 2         | 6            | 8     | 0         | 5            | 5     |
| 3. Colleague   | 20        | 122          | 142   | 1         | 24           | 25    |
| 4. Friend      | 28        | 80           | 108   | 3         | 36           | 39    |
| 5. Supplier and customer | 1      | 10           | 11    | 0         | 3            | 3     |
| 6. Medical/nursing care provider and recipient | 1 | 2 | 3 | 0 | 0 | 0 |
| 7. Other       | 1         | 77           | 78    | 0         | 8            | 8     |
| Situation      |           |              |       |           |              |       |
| 1. Living together | 227     | 536          | 763   | 30        | 64           | 94    |
| 2. Eating and drinking (with alcohol) | 23 | 69 | 92 | 0 | 16 | 16 |
| 3. Eating and drinking (without alcohol) | 8 | 61 | 69 | 2 | 31 | 33 |
| 4. Office/meeting room | 4 | 22 | 26 | 0 | 2 | 2 |
| 5. Break room  | 2         | 5            | 7     | 0         | 1            | 1     |
| 6. School      | 1         | 57           | 58    | 0         | 10           | 10    |
| 7. Sports/recreations | 1 | 6 | 7 | 0 | 3 | 3 |
| 8. Other       | 24        | 84           | 108   | 1         | 9            | 10    |
| Mask of index case |           |              |       |           |              |       |
| 1. Used        | 3         | 80           | 83    | 0         | 19           | 19    |
| 2. Unused      | 287       | 760          | 1047  | 33        | 117          | 150   |
| Distance       |           |              |       |           |              |       |
| 1. Within 1 m  | 267       | 691          | 958   | 33        | 116          | 149   |
| 2. Farther than 1 m | 23 | 149 | 172 | 0 | 20 | 20 |
| Time           |           |              |       |           |              |       |
| 1. More than 15 min | 272 | 817 | 1089 | 32 | 129 | 161 |
| 2. Within 15 min | 18 | 23 | 41 | 1 | 7 | 8 |
returning positive test. Of those who had a negative initial PCR test and a positive repeated PCR test, it is likely that some of them represent unconfirmed patients. Therefore, the impact of this is unlikely to be significant, and the positive rate of asymptomatic index cases was approximately one-seventh of the symptomatic index cases and that of attack rate and observed reproduction of asymptomatic index cases was positively correlated with the number of very high-risk cases and the number of high-risk cases. The positive rate among close contacts conducted at the Chuo Public Health Centre was 16.6% (297/1794) in 2020 and 17.8% (582/3265) in 2021, and no significant difference was found according to a Fisher’s exact test (p = 0.26). In the training data, a significant difference was confirmed in the positive rate among whether masks are worn, with a positive rate of 3.6% (3/83) among mask wearers and a positive rate of 27.4% (287/1047) among non-mask wearers (P = 0.00001). It is possible that the effect of epidemic strains is relatively small compared to the independent variables, such as whether masks are worn.

Of the explanatory variables, relationships and contact situations may vary greatly depending on the culture and customs of the group to which they belong, so the data must be reconsidered if their country or ethnicity is different. In such a case, it will be necessary to recollect the data as much as possible.

The reason why we set four categories for each risk of infection, and not just the simple choice of presence or absence of the infection, is that in the case of the survey subjects, such as a group of students, the contact situation with the index case is the same, yielding the same prediction result. However, in the actual contact tracing, some people become infected while others do not. For this reason, the risk grouping based on the infection probability was closer to the actual results in the field of contact tracing than the two choices of infection and non-infection.

Regarding the outcome when the specificity is not high, this machine learning model is not for the general public but for close contacts with a high risk of infection, and it is evident that false negatives cause problems rather than false positives. It seems reasonable to determine that sensitivity takes precedence over specificity.

In the contact tracing of COVID-19, VOC appeared one after another, epidemic strains changed and many new vaccines were developed. It is expected that the infection tendency of people will change over time. In such a situation, a machine learning model by Bayesian statistics that takes in the acquired data, updates it sequentially and changes the calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable. Moreover, in situations that change from moment to moment, to reflect situations calculated posterior probability is considered to be suitable.
that can be implemented in the field. Therefore, one of the methods proposed is to communicate the individual infection risk to close contacts so that people will consider the impact on those around them and act in a restrained manner.

Because time advantage is a priority relative to spreading infection, the machine learning model was constructed to predict the risk of infection in the field when patient surveys are conducted. In contact tracing, many people are chasing greater numbers of close contacts. The prediction models should be versatile, require no advanced knowledge and return clear results with simple input. Thus, this machine learning model is considered to be highly suitable for on-site operation.

Furthermore, considering that the data with an uncertainty of the reported survey is used as the explanatory variable, using an advanced prediction model method may not lead to a highly accurate result.

5. Conclusion

If it is possible to convey the risk of infection during the incubation period at the time of the first contact with a close contact of COVID-19, the close contact may take suppressive action during the incubation period regardless of the result of the initial PCR test, thereby preventing a secondary spread of infection.

ICMJE statement

Hideo Yoshikawa is the sole author.

Yoshikawa was fully responsible for the research design, building the predictive model, determining the data to collect, analysing and interpreting the data, writing the paper and journal selection.

Declaration of competing interest

The author declares that there are no conflicts of interest with respect to this research study and paper.

Acknowledgements

It has become possible to create this machine learning model and conduct a verification test with the cooperation of various people who entered data and healthcare workers who participated in the verification test. And the author would like to thank Enago (www.enago.jp) for the English language review, I would like to express my deepest gratitude and ask for your continued cooperation in the future.

References

[1] WHO Coronavirus (COVID-19) Situation Dashboard. https://covid19.who.int/. [Accessed 25 September 2021].
[2] Moore S, Hill EM, Tildesley MJ, Dyson L, Keeling MJ. Vaccination and non-pharmaceutical interventions for COVID-19: a mathematical modelling study. Lancet Infect Dis 2021;21(6):793–802.
[3] Israel confirms vaccine less effective against Delta variant, eyes third dose. Times of Israel. https://www.timesofisrael.com/israel-confirms-vaccine-less-effective-against-Delta-variant-eyes-third-dose/. [Accessed 6 July 2021].
[4] Health Ministry says COVID vaccine is only 40% effective at halting transmission. Times of Israel. https://www.timesofisrael.com/liveblog_entry/health-ministry-says-covid-vaccine-is-only-40-effective-at-halting-transmission/. [Accessed 28 July 2021].
[5] Chan LYH, Yuan B, Convertino M. COVID-19 non-pharmaceutical intervention portfolio effectiveness and risk communication predominance. Sci Rep 2021;11:10605.
[6] To solitude: learning from other countries on how to improve compliance with self-isolation. http://www.nuffieldtrust.org.uk/news-item/to-solitude-learning-from-other-countries-on-how-to-improve-compliance-with-self-isolation/. [Accessed 31 December 2021].
[7] Lewis B. Why many countries failed at COVID contact-tracing – but some got it right. Nature 2020;588:384–7.
[8] Smith LE, Potts HW, Amlot R, Fear NT, Mitchie S, Rubin GJ. Adherence to the test, trace, and isolate system in the UK: results from 37 nationally representative surveys. BMJ 2021;372:n608.
[9] Global surveillance for COVID-19 caused by human infection with COVID-19 virus Interim guidance 20 March 2020. World Health Organization (WHO). https://apps.who.int/iris/bitstream/handle/10665/331506/WHO-2019-nCoV-Surveillance-2020.6-eng.pdf. [Accessed 14 July 2021].
[10] Shams H, Wei S, Khwar P, Lalvani A, Moxon FR, Pogoda JM, et al. Enzyme-linked immunosorbent and tuberculin skin testing to detect latent tuberculosis infection. Am J Respir Crit Care Med 2005;172:1161–8.
[11] Yoshikawa H. Trial of onset prediction by Naive Bayes classifier in tuberculosis contact. J Natl Inst Public Health 2014;63(2).
[12] Viner RM, Mytton OT, Bollini C, Menezes-Torres GJ, Ward J, Hudson L, et al. Susceptibility to SARS-CoV-2 infection among children and adolescents compared with adults: a systematic review and meta-analysis. JAMA Pediatr 2021;175(2):143–56.
[13] You Y, Yang X, Hung D, Yang Q, Wu T, Deng M, et al. Asymptomatic COVID-19 infection: diagnosis, transmission, population characteristics. BMJ J. https://spc.annals.org/content/early/2021/07/29/bmjpcare-2020-002815. [Accessed 31 December 2021].
[14] Reid MJ, Prado P, Bronson H, Ernst A, Spindler H, Celestino J, et al. Assessing testing strategies and duration of quarantine in contact tracing for SARS-CoV-2: a retrospective study of San Francisco’s COVID-19 contact tracing program, June–August 2020. Open Forum Infect Dis 2021;8(7):171. https://doi.org/10.1093/ofid/ofab171.