

INTERNATIONAL REAL ESTATE REVIEW

202 Vol. 25 No. 5: pp. 525 – 542

Forecasting COVID-19 Infection Rates with Artificial Intelligence Model

Jesse Yang Jingye

Institute of Real Estate and Urban Studies, National University of Singapore
E-mail: jesse@nus.edu.sg

This study applies an artificial intelligence (AI) based model to predict the infection rate of coronavirus disease 2019 (COVID-19). The results provide information for managing public and global health risks regarding pandemic controls, disease diagnosis, vaccine development, and socio-economic responses. The machine learning algorithm is developed with the Python program to analyze pathways and evolutions of infection. The finding is robust in predicting the virus spread situation. The machine learning algorithms predict the rate of spread of COVID - 19 with an accuracy of nearly 90%. The algorithms simulate the virus spread distance and coverage. We find that self-isolation for suspected cases plays an important role in containing the pandemic. The COVID-19 virus could spread asymptotically (silent spreader); therefore, earlier doctor consultation and testing of the virus could reduce its spread in local communities.

Keywords

COVID-19, Artificial Intelligence, Forecasting

1. Introduction

The world has been fighting a new public health enemy, the Coronavirus (COVID-19 virus), since the virus first appeared in Wuhan, China, in December 2019. COVID-19 has spread rapidly worldwide, and the number of confirmed infected cases worldwide has exceeded 2.0 million as of 15 April 2020 (Bennhold et al., 2020). The U.S. has the highest number of COVID -19 infections in the world, and the number of death cases spiked in a short time in April 2020. On 3 April 2020, Singapore's Prime Minister Lee Hsien Loong announced in a televised speech that the government had introduced the "circuit breaker" for one month till 4 May 2020. Workplaces were closed, schools were converted into home-based learnings, and movements and gatherings in the community were restricted. As of 20 April 2020, a total of 8014 infected cases in Singapore were reported, with a record high of 1426 new cases, 768 cured and discharged, and 11 died. The COVID -19 pandemic has spread widely worldwide, and the number of new patients is increasing sharply every day. In response to the Covid-19 pandemic, China calls on people to wear masks, formulate preventive measures in due time, and close public entertainment places when necessary to reduce the spread of the COVID -19 pandemic. By December 2022, the number of confirmed cases of the COVID -19 pandemic has exceeded 630 million according to COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) of Johns Hopkins University. The invisible "enemy" has caused unprecedented economic and social impact since the last global recession in the 1930s. The COVID-19 pandemic has caused disruptions in economic, social, and cultural activities across the world. The international travel bans imposed by many countries have crippled the airline and travel-related industries, which dire need rescue packages from various governments. For example, the U.S. travel and tourism industries seek the government's relief of up to US\$150 billion to revive the grounded businesses (Hirsch, 2020). Singapore Airlines (SIA), the national carriage of Singapore, also plans to raise USD\$ 11.3 billion with the support of Temasek, the government's investment arm, to help tide over the pandemic that has dented the global air travel businesses (Kaur, 2020).

1.1 Applications of Artificial Intelligence in Predicting Pandemic

Early detection and issuing pandemic warnings are important in effectively preventing disasters and preventing any virus from spreading in local communities and across borders. BlueDot, a Canadian artificial intelligence (AI) health monitoring platform, issued its customers the first warning of the COVID-19 outbreak as early as 31 December 2019. The alert was issued much earlier than the outbreak warning of the World Health Organization which was issued on 9 January 2020 and the US Centers for Disease Control and Prevention on 6 January 2020. BlueDot issued the outbreak warning earlier with more accurate information than any of the other major health organizations worldwide because of its use of advanced data analytics. BlueDot uses AI

technology, natural language processing, and web crawler technology to continuously collect and analyze data from reports and comments on diseases and major public health events worldwide. The machine learning algorithms extract key data efficiently from massive volumes of data and intelligently forecast the possibility of a global outbreak.

However, the steps involved in developing a robust predictive model for the pandemic can be elaborate and complex, including collecting and processing big data, effective screening of information, and automatic removal of false information. At the early stages of the pandemic, the information was mixed with various kinds of "noises" and trivial information, which, if not cleaned properly, distort the predictive outcomes. Moreover, an "artificial" and unsystematic approach to processing the data can be inefficient and makes it difficult to ensure the validity, reliability, and timeliness of the information. AI technology is an expert system that can process massive amounts of data. The technology trains computers with learning capabilities in selected fields and applies the trained knowledge to mimic and simulate human decision-making processes in real-world scenarios.

AI technology has shown good performance when applied to predict virus outbreaks. As early as 2009, researchers at University College London (UCL) in the United Kingdom applied AI technology to process internet information generated by real-time comments and reports from social media channels, such as Twitter. They detected and provided early warnings before the spike of the swine flu pandemic (Szomszor et al., 2010). Researchers from UCL and scientists worldwide continued to develop various AI models and used data from the Internet (such as search engine indexes) to enhance early surveillance and prevent the spread of epidemics. Methodologies, including regression models, neural networks, and AI, are some of the big data analytic tools applied to predict the trends of epidemics and pandemics, with inputs of propagation data and dynamic infection data.

1.2 The Approach

In predicting the spread and tracing the root cause of the pandemic, we took advantage of emerging AI technologies, such as deep learning. Based on data of high-risk populations and newly diagnosed, suspected, dead, and cured cases, we applied AI technology to create heat distributions at the onset of the risk for close contacts. We also predicted the peak and inflection point of the pandemic outbreak. The model uses multiple sources of big data, including travel trajectory flow information, social information, consumption data, and exposure history, in developing a predictive model of the pandemic. We follow the diagnosis sequence of the patient, close contact personnel, and other information to locate spatio-temporal collision points and then calculate the path of disease transmission, thus providing a theoretical basis for the traceability analysis of infectious diseases.

The traditional ways of mapping the potential spread of viruses usually took several weeks to alert the relevant authorities. The government may sometimes lag in releasing the data, which could sometimes be delayed by weeks or months. In the interconnected and mobile world, we conduct big data mining by using deep learning and other AI technologies combined with cloud processing technologies. The inputs of information generated from various sources include keyword search data, locations where people click on Wikipedia pages, etc. AI and big data could significantly improve the tracking of infectious disease outbreak paths and provide real-time prediction and early warning of viruses. The early detection and real-time prediction of the pandemic situation have been critical in the war against the pandemic, such that swift anti-pandemic measures could be implemented, and targeted pandemic clusters could be detected earlier and contained.

1.3 Objectives of the Proposed Study

There are two objectives of this study:

- To develop an AI-based pandemic predictive model that can predict the rate of spread of the pandemic and the distance of the spread, and to test the power and reliability of the proposed AI algorithms in predicting the pandemic based on the existing data; and
- To simulate the impact of infection on various scenarios when different containment and isolation measures are implemented for the pandemic; and to simulate useful predictive outcomes that can be used to manage the evolving pandemic situation.

1.4 Data source

The data used in this article is obtained from github.com, which publishes COVID-19 detection and diagnosis data extracted from the reports of health departments worldwide. In addition, the latitude and longitude information data of each province in China are derived from COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) of Johns Hopkins University (JHU; https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports).

2. Methodology

In this study, we provide a robust methodology that can be used to predict the COVID-19 pandemic situation by accurately using AI technology. The study results could have significant implications for public health management with regard to pandemic control, disease diagnosis, vaccine development, and socio-

economic aspects. The results can also be easily extended to prevent the occurrence of other pandemics in the future.

In this report, we develop preliminary machine learning algorithms to predict the rate of spread of the COVID-19 virus and simulate the distance of its spread. Built on preliminary knowledge of the AI technology in this study, we hope to improve and refine the model, which could be adopted in real-life situations to support the policy-making process.

2.1 Assumptions

As the COVID-19 situation in mainland China has been contained at the time of writing of this paper, we use only the data of the new COVID-19 virus cycles in mainland China (including the number of confirmed cases from 1 February 2020 to 1 December 2022) for illustration purposes in this preliminary study. We could expand the study to other countries by including data on their new COVID-19 cases and various reference verification groups. There are three sources of data that are used in this study: (a) confirmed cases; (b) suspected cases; and (c) infected areas.

Since this study is limited to China, we use various regional indicators to track the number of deaths over time. Death rates vary significantly at different stages of the pandemic, including the early, middle and late stages. Suspected cases vary depending on the testing and consultation capabilities of different places. For example, Hong Kong reported many suspected cases, but few were formally diagnosed; so the "diagnose" indicators are not used for Hong Kong, but used as the reference index for other regions of China.

Suppose the spread of the virus is heterogeneous by region, climate conditions, and characteristics of the virus. In that case, we could develop algorithms that predict the diagnosed and confirmed cases of the virus by connecting a virus of similar properties to the three properties (region, climate, and new coronavirus characteristics). The algorithm can also predict and test for new coronavirus outcomes. The data on COVID-19 are collected from the CSSE at the JHU.

2.2 Evaluation algorithm

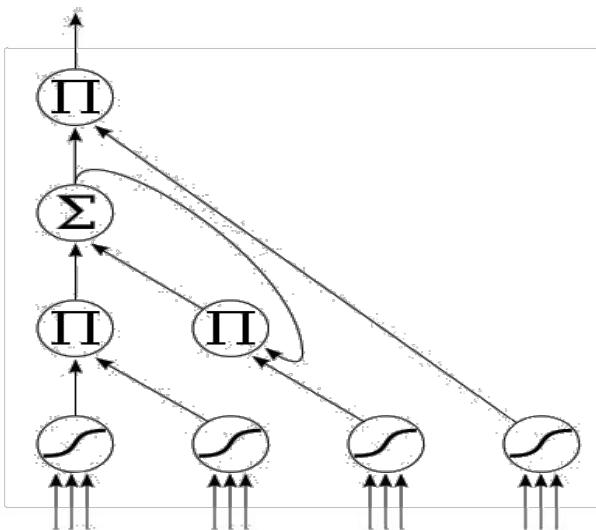
We build the proposed model based on the long short-term memory (LSTM) network (LSTM; Sundermeyer et al., 2012). The LSTM network is a classic algorithm often used to quantify investment and good at forecasting time series. The adjustment algorithm in the model is based on past data, and the latest data can be used for out-of-sample forecasting and testing. The LSTM network is a type of recurrent neural network (RNN), also known as RNN sequences. The neural network structure is based on time series and character sequences (depending on the application scenarios). After expansion, the sequence becomes a common three-layer neural network commonly used in speech recognition.

The feed-forward neural network has proven to produce effective outcomes, but cannot simulate a time-dependent relationship, given that the data points have fixed-length vectors. A recursive neural network is created to overcome such limitations. Recursion is a method that calls itself repeatedly. The difference between a RNN and other networks is that its hidden layer is a self-connected layer that can span across different time points. This layer allows access to the hidden layer in the next step. At each step, the hidden layer retains information and deduces the latter state according to the previous state.

LSTM networks are a kind of neural network that contains LSTM blocks, described as intelligent network units in the literature. The blocks can remember values in each block for an indefinite time. The gate will determine whether an input is important enough to be remembered and the value can be an output.

Figure 1 illustrates a structure with four "S" function units. The leftmost function is the input of the block according to the situation. The three on the right determine whether the input can be passed into the block through the gate. The second node on the left is the input gate, such that "zero" will block the value here and will not go to the next layer. The third node on the left is the "forget" gate, so a value close to zero in the block will be forgotten. The fourth node is the output gate, which decides whether an input in the block memory can be turned into an output. According to Google's tests, the forget gate is the most important gate in LSTM networks, followed by the input gate and then the output gate. There are many versions of LSTM networks, one of which is the gated recurrent unit (GRU).

Figure 1 Long-Short-Term Memory Network Neural Network Model (Hochreiter & Schmidhuber, 1997)



2.3 Preliminary Model

In this study, we use only the COVID-19 data from mainland China in our preliminary experiment and devise strategies to build up the machine learning models for our predictive purposes. The model is divided into the virus infection probability and individual movement models.

2.3.1 Individual virus infection probability model

Point P (x_i, y_i) on the coordinate plane system describes the individual. Assuming that the distance between any two individuals P (x_i, y_i) and Q (x_j, y_j) is denoted as d_{ij} , if Q is a virus carrier, the direction of potential virus spread is only from Q to P. The probability of the virus spreading that causes P to be infected is denoted as p_{ij} . The relationship between the two is described by the formula (Gers & Schmidhuber, 1999):

$$p_{ij}(d_{ij}) = \begin{cases} (1 + w_{ij})p_{0i} & \text{if } d_{ij} \leq d_{0i} \\ (1 + w_{ij})p_{0i}\alpha_i^{-(d_{ij}-d_{0i})} & \text{if } d_{ij} > d_{0i} \end{cases} \quad (1)$$

where d_{0i} is the critical distance of individual P (x_i, y_i) ; w_{ij} reflects the intimacy of individuals P and Q, and its value is $0 \sim 1$; p_{0i} is the probability of P being infected within the critical distance d_{0i} when $w_{ij} = 0$, which is also known as the probability of basic infection; and α_i is a coefficient greater than 1. The critical distance d_{0i} of an individual and its probability of being infected within the critical distance r_{0i} reflect the resistance of an individual; w_{ij} reflects the intimacy index between the individuals; and α_i reflects the range of influence.

Equation (1) captures the distance between the individuals and the probability of being infected by the virus. It should be noted that distance here refers to a generalized distance, and not actual physical distance between two individuals. It can be considered the comprehensive distance between two individuals in various aspects, including space-time and intimacy relationships, resistance, etc., where $p_{ij}(d_{ij}) \neq p_{ji}(d_{ij})$.

A higher intimacy between individuals means a higher chance of infection. The model assumes that the probability of infection of an individual within a critical range d_{0i} is constant; the probability of infection gradually decreases after exceeding this range. After reaching a certain value, the probability of infection rapidly decreases to 0.

2.4 Individual mobility model

Suppose any individual P (x_i, y_i) walks randomly in a plane (random walk). The step of movement of an individual, d_{step} , is used to describe the activity of an individual. A larger value indicates that the individual is more active. The

random walk-related knowledge is not repeated here. We assume that the distance between two individuals is a fixed value d_f , and $p(d_f) \approx 0$ is satisfied.

2.5 Solution process

The solution process is described in 4 sequential steps as follows:

Step 1: The model parameters are determined which mainly include:

1. The number of individuals NI ;
2. Individual initialization distance d_i ; and
3. Individual critical distance matrix

$$D_0 = (d_{01} \quad d_{02} \quad \dots \quad d_{0NI})$$

4. The 0 intimacy probability matrix of individuals infected by carriers within a critical distance, also known as the basic infection probability matrix

$$P_0 = (P_{01} \quad P_{02} \quad \dots \quad P_{0NI})$$

5. Intimacy matrix of each individual and other individuals

$$W = \begin{pmatrix} w_{00} & \dots & w_{0NI} \\ \vdots & \ddots & \vdots \\ w_{NIO} & \dots & w_{NINI} \end{pmatrix}$$

6. Base matrix of individual probability calculation model

$$A = (a_{01} \quad a_{02} \quad \dots \quad a_{0NI})$$

7. Random walk time step t_{step} ;
8. Moving distance step size matrix d_{step} ; and
9. Individual walk duration t_1 .

Step 2: The position of the individual is initialized;

Step 3: The individual of source of infection is placed; and

Step 4: The position of the individual within each time step t_{step} is updated and the infected individual updated according to the infection probability model until time t_1 is reached.

2.6 Sample Analysis

Next, we simulate many different situations one by one. Due to the randomness of many parameters in the model, we constantly adjust the parameters to achieve better output results.

2.7 Simulating normal activity propagation

This situation simulates the early stage of virus development. Some people have been infected with the virus, but this has not attracted attention yet. Everyone, including those infected, are normally active. The propagation process is defined by setting different parameters which are listed in Table 1.

Table 1 Basic parameters

1	$N = 20$
2	$NI = N * N$ #number of individuals
3	$D_0 = 2.$ #critical distance
4	$P_0 = 0.3$ #Individual basic infection probability
5	$W = 0$ #Intimacy
6	$d_f = 6.$ #initial distance
7	$t_step = 0.01$ #Time step
8	$d_step = 0.05$ #moving step
9	$m =$ Diffusion (N)
10	$m.init_pos (d_f)$ #initialization position
11	$m.init_para (D_0, P_0, W, t_step = t_step, d_step = d_step)$ #Model parameter setting
12	$m.place_source (153)$ #Place of the infection source individual
13	for t in [1,2,3,4,5,6,7]:
14	$m.walk (t)$
15	$m.plot ()$

2.8 Preliminary results

The total number of infected individuals at various points under the different parameters is graphically represented in Figures 2 to 5.

It is shown that the total number of infections at time $t = 5$ is the same. With an increase in time, the rate of growth of infections is faster in the scenario with a high basic infection rate $p_0 = 0.3$ and a high moving step $d_{step} = 0.1$. This means that the increase in diffusion at the early stages does not have a relationship with parameter size, timely discovery, and isolation.

If the assumption is not rejected with actual data, the later stage of propagation is reached. We use different models for verification in different periods of time. Here, we only use mainland China data in the testing. The curve shows the number of new COVID-19 cases diagnosed from 1 February 2020 to 1 December 2022. We verify the prediction of our models with newly diagnosed COVID-19 infection data. The first 70% of the data are used to train the model, and the remaining 30% is used in the model for out-of-sample estimation and verification purposes.

Figure 2 Actual number of infected cases from 1 February 2020 to 1 December 2022 and definition of infection stages.

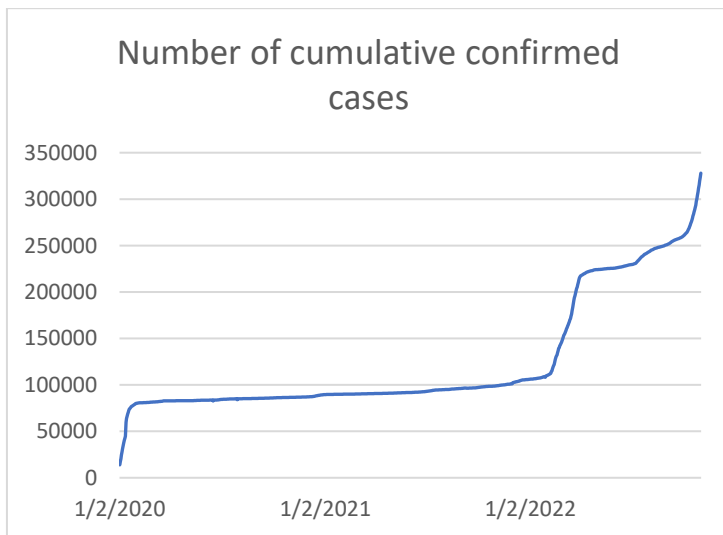


Figure 3 Plotted changes in modelled data numbers (forecasted vs. actual) in early stages of pandemic

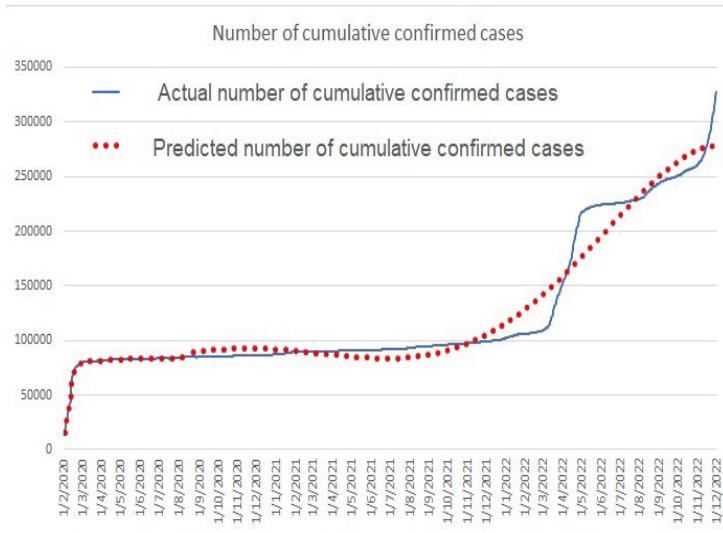


Figure 4 Plotted changes in modelled data numbers (forecasted vs. actual) in middle stage of pandemic

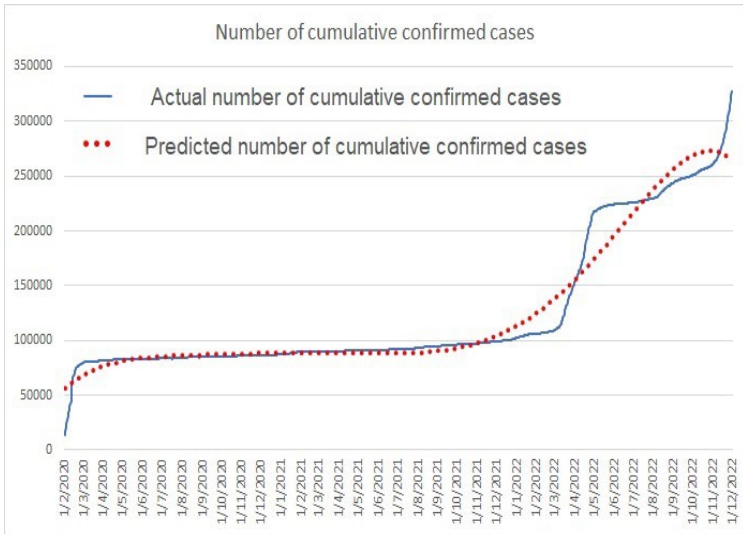
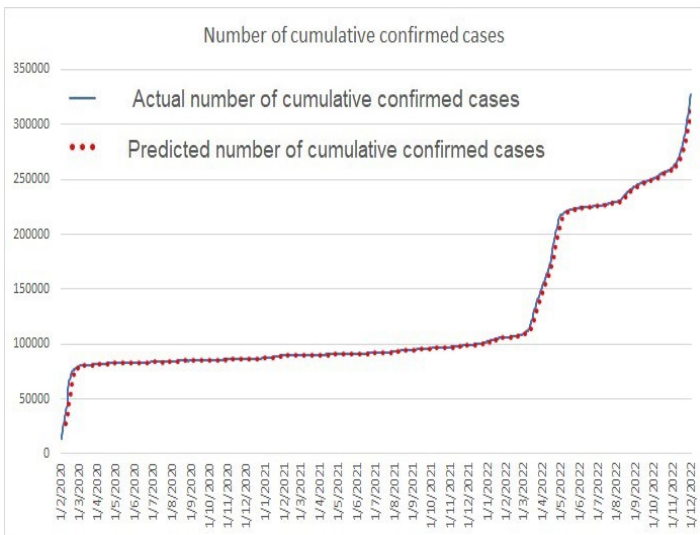


Figure 5 Plotted changes in modelled data numbers (forecasted vs. actual) in late stage of pandemic



China has always adhered to the principle of "early detection, early reporting, early isolation, and early treatment". As illustrated by Figures 2-5, at the early and mid-term stages, the predictive data are larger than the actual data, implying

that the measures implemented in China have effectively curbed the spread of COVID-19. First, free nationwide nucleic acid testing should be carried out in areas where COVID-19 has occurred to find potential positive patients as soon as possible and narrow the spread of the pandemic. Secondly, reduce public gathering activities to avoid the occurrence of a gathering pandemic. Finally, we should strengthen medical security and constantly increase the number of medical personnel and pandemic prevention workers for pandemic prevention and control. The infection rate at the later stage is lower, as shown by the flattening slope in the predicted line (at the time of writing of this paper).

2.9 Isolation simulation

The next analysis is conducted with minor changes to the basic parameters. The number of individuals in the model described above is 400. The critical distance d_0 of each individual is the same. The probability of basic infection is also the same at $p_0 = 0.3$, and intimacy is $w = 0$ among all individuals. The movement steps of each individual are the same. The default infection probability of the base model is e , and the mathematical constant is approximately equal to 2.71828. The base of the natural logarithm is equal to one. It is the limit of $(1 + 1/n)^n$ as n approaches infinity, an expression that arises in the study of compound interest.

The moving step size $d_{\text{step}} = 0.05$ and $d_{\text{step}} = 0.1$ are calculated respectively, with the probability of basic infection $p_0 = 0.3$, and time increment of putting one pathogen $t = [1, 2, 3, 4, 5, 6, 7]$ infection process on each node: white dots represent uninfected individuals; yellow dots represent suspected cases; and red dots represent confirmed cases.

With self-quarantine near the source of infection and fixing its step length at $d_{\text{step}} = 0$ (green dot), the basic infection rate is still set at $p_0 = 0.3$, and the basic infection rate of other individuals is also $p_0 = 0.3$. The moving step length, $d_{\text{step}} = 0.1$. Figures 6 shows the infection process diagrams.

Figure 6 Infection process diagrams without self-isolation
 (a) GIF of dynamics of the infection rate and distance (Left)
 (b) Static image of the final stage of infection (Right)

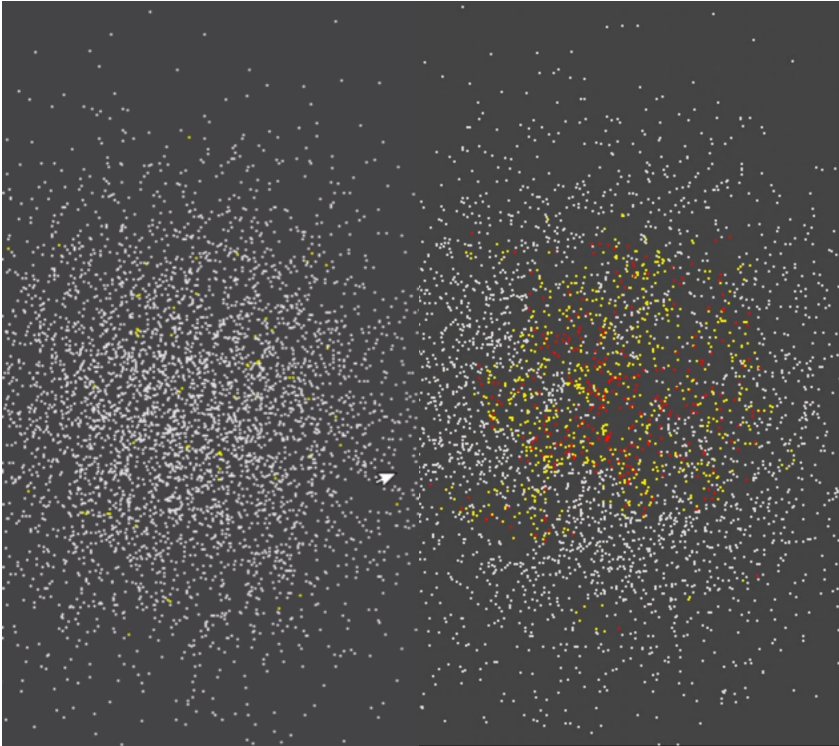


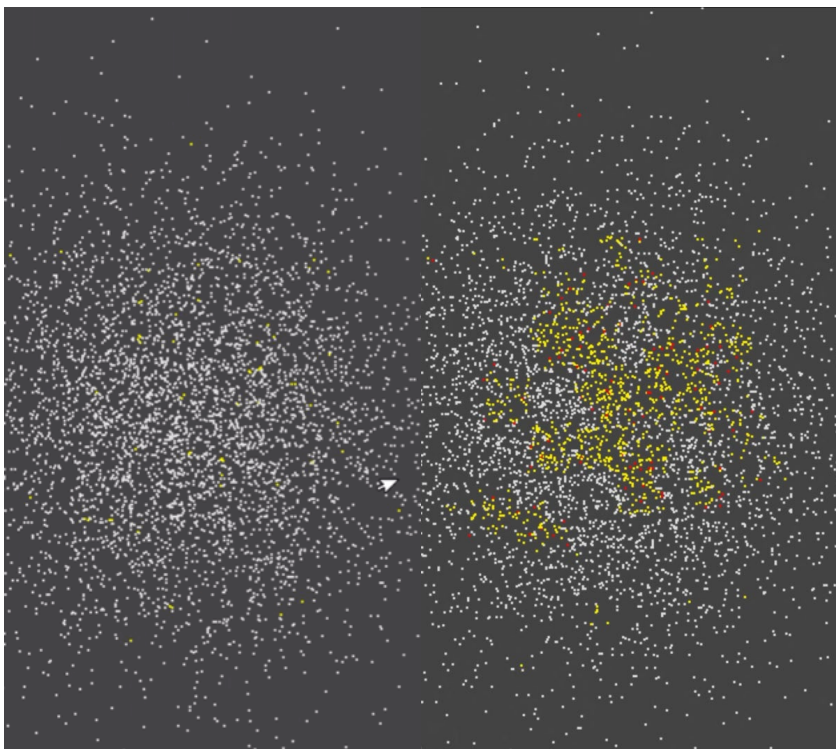
Figure 7 shows that the chance of infection of self-isolated individuals is significantly reduced.

If the probability of basic infection during quarantine is set at $p_0 = 0.0$, there will be no infection. The propagation rate is low because quarantine effectively reduces the propagation path. This means that self-isolation is an important way to reduce virus transmission.

The above results show that our proposed machine learning algorithm can forecast the rate of spread of the COVID -19 virus in mainland China. The actual rate of spread can be influenced by other factors, such as change in slope, errors, number of turning points, etc. It can be seen that there are one or more turning points in each period (not counting the end of the pandemic). The error rate at the closest turning point is directly calculated. During the outbreak period, the error value at the closest turning point is estimated to be 0.21. The error value at the closest turning point during the control period is estimated to

be 0.14. The average value is estimated to be 0.175. The change in error rate means that the predicting model may vary at different stages of the COVID-19 pandemic. Different countries may experience different peak case days, peak death days, and peak patterns. The error rate in predictions during the early days of COVID-19 in China may provide policymakers with sufficient evidence to understand the health status of the local diagnosed population and propose accurate and timely public health interventions to control and prevent similar infectious diseases.

Figure 7 Infection process diagram with self-isolation:
(a) dynamics of the infection rate and distance (Left)
(b) static image showing the final stage of infection (at time of writing) (Right)



On 20 January 2020, China's pandemic expert Zhong Nanshan explained the existence of person-to-person transmission of COVID-19 for the first time during an interview with China Central Television after 14 medical staff were infected. This was the starting point that attracted the attention of public agencies and media to the COVID-19 virus. Then they began to report updates and responses to the pandemic situation. However, as such, earlier data are

missing, and the model may not be able to produce reliable predictions for the period before 20 January 2020. Thus, we truncate the study period to 1 February 2020, which is the starting point of our model.

It has been indicated that the virus has become stronger. According to the model estimations, the number of people infected with COVID-19 would increase sharply in March 2020. This was exactly the case in China. In March 2020, COVID-19 occurred in all parts of China, and the number of infected people continued to increase. There was the urgent need to build a system for enhanced surveillance, early detection, and pandemic prediction. This study illustrates the potential and applications of AI predictive models of pandemics. The model depends on the data inputs in the predictions. The model outputs could be further improved if more data could be obtained, including the number of medical staff, accuracy of detection methods, time, virus incubation period, etc. More resources including manpower and material resources could be required to improve the model such that more reliable models could be developed to produce useful information in the fight against the COVID-19 pandemic.

There is still room for further improvement in the model construction process and collection of relevant data on the pandemic situation, where more information is unveiled daily. For example, statistics on the supported items, such as masks, ventilators, hospital beds, the number of patients, demand for protective items, material statistics needed in each period of the pandemic, etc., could further enhance the results. If available, more detailed data on this pandemic could be used as input to the system. With these data, we can create a more powerful pandemic monitoring and evaluation system, which could be extended and adapted to other future event monitoring and evaluation purposes. When a pandemic or disaster occurs again next time, nationwide resources can be quickly dispatched for treatment, prevention, and containment efforts.

3. Research conclusions and deficiencies

3.1 Conclusion

Our proposed machine learning algorithms can be used to predict the rate of spread and probability of infection of COVID-19 with an acceptable level of accuracy with the use of data from mainland China in the analysis. We can simulate the spreading distance and coverage of the virus. We can also evaluate the effectiveness of self-isolation for suspected cases and the effectiveness of a strategy in containing the pandemic. If there are signs of infection, it will be helpful to control the virus spread by seeing a doctor earlier.

We could further enhance the predictive power of the model with more time and effort in collecting more data and adjusting the various parameters. There are some recommendations for future works that can include:

1. evaluating the infection probability model, including adjustment of critical distance and model base;
2. studying the impact of the transmission path of super-infected persons;
3. simulating various community communication channels and their impact (setting intimacy coefficient between individuals);
4. simulating and calculating the effects of cross-city transmissions (in combination with data from trains and airlines); and
5. since the calculation results are sensitive to the input parameters as a probabilistic model, we could reduce the gap to a controllable range and use the actual data to more accurately calibrate and optimize the parameters for different countries.

3.2 Limitations of the Methodologies

AI technology is still immature, so it requires some degree of manual intervention to predict pandemic situations. The current AI technologies may generate false positive results when the predictive outcome is influenced and triggered by unanticipated shocks that are not captured by the data. Therefore, human intervention for some of the stages is still required, and experts with in-depth knowledge of the pandemic could be invited to review the outcomes to minimize the prediction errors, which the predictive model could have neglected.

John Brownstein (2008), the founder of "HealthMap", which is the company behind BlueDot, indicated that big data will constantly change based on the behavior of people on the Internet, so we must constantly re-adjust our algorithm. BlueDot's pandemic prediction system uses other data from geographic information systems, spatial analyses, and computer science, clinical infectious disease, tourism, and tropical medicine sources. BlueDot also invites public health medical experts to review information on AI markers produced by the system.

More advanced AI algorithms based on data adaptation will be available with continuous improvements in computing power. To some extent, the use of AI technology for pandemic prediction depends on the quality and availability of data. For example, once the pandemic was known, governments began to limit or ban inter-regional movement and inter-country travel. Therefore, if the model could use the historical data of airlines, we could further compare the pandemic infection, to consider the connectivity between countries.

Using AI to process real-time geographic data of smart devices can better analyze the gathering and spreading trajectory of a virus. The Israeli government uses anti-terrorism-level technical surveillance to track the COVID-19 outbreak. Closely monitoring smartphones, biometrics and body

temperature reports can quickly identify virus carriers and remind people to distance themselves from infected persons. However, in Western countries that value data privacy, the availability of such data would limit and affect the performance of AI technology.

AI technology has had a major role in predicting the spread of the COVID-19 virus, but there is still ample room for improving AI technology and application. Currently, the pandemic data obtained through public channels are inconsistent, and the quality is mixed across different countries with ambiguity for some, which may cause difficulties when applying AI models in cross-country analyses.

3.3 Future research

Based on existing research, there is the need to further compare and analyze the changes in the COVID-19 pandemic infection rate among different countries with government strategies and their effectiveness in curbing and slowing the spread of the virus.

References

- Bennhold, K., Turkewitz, J., & González, J. (2020). Worldwide confirmed coronavirus cases top 2 million. *New York Times*, 15.
- Brownstein, J. S., Freifeld, C. C., Reis, B. Y., & Mandl, K. D. (2008). Surveillance Sans Frontiers: Internet-based emerging infectious disease intelligence and the HealthMap project. *PLoS medicine*, 5(7), e151. doi:<https://doi.org/10.1371/journal.pmed.0050151>
- Gers, F., & Schmidhuber, J. (1999). *Learning to forget: Continual prediction with LSTM*. Paper presented at the IEEE Conference Publication.
- Hirsch, L. (2020). Travel industry pushes for \$150 billion in aid as coronavirus cripples business. *CNBC*.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780. doi:<https://doi.org/10.1162/neco.1997.9.8.1735>
- Kaur, K. (2020). SIA aims to raise \$15 billion with support from Temasek to combat Covid-19 fallout. *Straits Times*.
- Sundermeyer, M., Schlüter, R., & Ney, H. (2012). *LSTM neural networks for language modeling*. Paper presented at the Thirteenth annual conference of the international speech communication association.
- Szomszor, M., Kostkova, P., & Quincey, E. d. (2010). # *Swineflu: Twitter predicts swine flu outbreak in 2009*. Paper presented at the International conference on electronic healthcare.