SIR-RL: Reinforcement Learning for Optimized Policy Control during Epidemiological Outbreaks

Maeghal Jain[1,*] **and Ziya Uddin**[1]

[1]BML Munjal University
[*]e-mail: maeghaljain@gmail.com

ABSTRACT

The outbreak of COVID-19 has highlighted the intricate interplay between public health and economic stability on a global scale. This paper aims to develop a reinforcement learning framework to balance health and economic outcomes during infectious disease outbreaks. The framework utilizes the SIR model without vital dynamics and incorporates globally comparable government responses. The study acknowledges the limitations of deterministic models and proposes the use of deep reinforcement learning to reduce input dimensionality and normalize input. While the model offers transparency in terms of its dependency on the reward policy defined, it recognizes the need for a comprehensive consideration of decision factors beyond pure reinforcement learning results.

Introduction

In the past, global spread of infectious diseases was largely due to colonization, slavery, and war, leading to widespread illness and death from diseases like tuberculosis, polio, smallpox, and diphtheria. Medical advancements, better access to health care, and improved sanitation have worked towards improving the situation of mortality and morbidity linked to infectious diseases in the past twenty years. However, in low and lower-middle income countries the burden of infectious diseases still persists. The rapid pace of urbanization in low and middle-income countries, along with the rise in populations living in crowded, poor-quality homes, has led to new conditions that favor the emergence of infectious diseases[1,2].

Recently, the COVID-19 pandemic caused a havoc worlwide. Till date there have been 772 million cases and more than 6 million deaths[3]. The pandemic triggered the sharpest economic recession in modern history with a 3% decline, much worse than during the 2008–09 financial crisis[4]. As nations grappled with the immediate health crisis, the economic fallout disproportionately affected vulnerable populations and exacerbated existing inequalities. Lockdowns and restrictions imposed to curb the spread of the virus led to widespread unemployment, business closures, and disruptions in global supply chains[5]. The challenges faced by low and lower-middle income countries were particularly acute, highlighting the intricate interplay between public health and economic stability on a global scale[6].

The need for a nuanced understanding of how interventions impact both health outcomes and economic indicators became increasingly evident, prompting a comprehensive examination by epidemiologists to assist policy makers[7]. The outbreak of COVID-19 has prompted epidemiologists to research on various aspects, including mobility control[8,9], vaccination strategies[10,11], stringency measures/non-pharmaceutical interventions (NPIs)[12–14], and financial considerations[15]. Despite the numerous studies conducted, very few explore how common interventions meet multiple policy objectives or how a precise articulation of the main policy goals directs the selection of the most effective interventions in terms of health and economic

results[8,16–19]. The economic impact of the COVID-19 pandemic varied between rich and poor countries. Although COVID-19 deaths had a slightly larger negative effect on the Gross Domestic Product (GDP) in advanced economies, this difference was not statistically significant. However, lockdown restrictions were found to have a more damaging impact on economic activity in emerging and developing economies[6].

Many economists have studied the effect of COVID-19 on the economy of nations[6,20–22]. In advanced economies like Korea, where the stringency index was below the median the recession was milder than other advanced economies like the United Kingdom where the stringency was much higher[20], they achieved it mostly with very aggressive testing, contact tracing, and enforced quarantines and isolations[23,24]. In India, social distancing and containment measures have been effective in reducing the number of COVID-19 cases but have come with economic costs. Social distancing had the most adverse effect on the economy in areas with high urbanization[21].

In this paper we want to optimize the government policies regarding stringency. Therefore, alongside epidemiological data, we use the measures of globally comparable government responses[25]. We use the simple SIR model without vital dynamics[26–28] as it is assumed that the time scale is short enough so that can be neglected[29]. By lesioning the model, as opposed to proposing a new mathematical model with more specialized compartments to more accurately represent the actual environment[30,31], we can effectively address the real-world conditions and propose a solution that is both effective and extendable. However, the study has limitations; the deterministic SIR model fails to account for chance in disease spread and lacks confidence intervals on results and while stochastic models incorporate chance, they are typically more challenging to analyze than their deterministic counterparts[27].

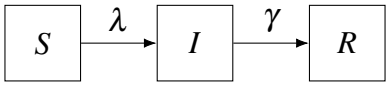
In order to capture competing costs within the environment and achieve a balance between health and economic outcomes, we intend to employ reinforcement learning[8,19,32,33]. When we conceptualize our problem as a reinforcement learning task, an agent is tasked with making decisions in an environment with the aim of optimizing cumulative rewards. This approach relies on trial and error, in contrast to dynamic programming, which assumes complete knowledge of the environment[34]. We make the use of deep reinforcement learning, which is an advancement to reinforcement learning as it allows to reduce the input dimensionality and normalize the input[34–36]. While our model is more trasparent in terms of its dependency on the reward policy defined, it has its limitations. A universal optimal policy may not suit diverse socioeconomic contexts due to variations in healthcare resources and economic vulnerabilities across countries, regions, or cities and a comprehensive consideration of decision factors, extending beyond pure reinforcement learning results is needed[8,37,38].

Results

Some results and their discussion here; will add once everything is finalized after the meeting on Wednesday.

Methodology

In SIR-RL, we use a commpartmental model to model infectious disease. Here we divide people of the population, based on whether they are yet to come into contact with an infected person (Susceptible), are infectious themselves (Infectious), or have recovered from the infection (Recovered). These compartments create the SIR model which can be represented as follows:



$$\frac{dS}{dt} = -\lambda S \tag{1}$$

$$\frac{dI}{dt} = \lambda S - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

Here, λ is the force of infection, it is the rate at which susceptible individuals acquire an infectious disease[39]. It depends on other factors:

$$\lambda = pc\frac{I}{N} \tag{4}$$

Here, c is the average number of contacts a susceptible makes per day. p is the probability of the susceptible person becomes infectious after coming into contact with an infectious person. $\frac{I}{N}$ is the proportion of the contacts that are infectious.

And, β the effective transmission rate is defined as:

$$\beta = pc \tag{5}$$

During an epidemic, the fundamental drivers of an epidemic growth is the rate of infection β i.e. the average number of infections per infected case and the infectious period $1/\gamma$ i.e. the average period for which the infected case is infected for. Epidemics can only happen if the case is infectious enough for long enough and this defined by $R_0 = \beta/\gamma$. Here, R_0 is The average number of secondary infections caused by each infected case, in an otherwise fully susceptible population.

At the peak of an epidemic R_0 declines as there are no more susceptible people left in the pool, therefore, R_{eff} comes into play. R_{eff} is defined as the average number of secondary cases arising from an infected

case, at a given point in an epidemic - therefore, it takes into account the existing immunity of the system.

$$R_{eff} = R_0 \frac{S}{N} \tag{6}$$

S is the number of susceptible people, N is the total population. At the start of an epidemic when everyone is susceptible, $R_{eff} = R_0$ as, $S = N$ (i.e. the whole population is susceptible). β and γ are also used to define probablity of and infectious individual infecting another individual $\beta/(\beta+\gamma)$ and the probability of recovery, $\gamma/(\beta+\gamma)$.

Most government policies look at the value of R_{eff} to come us with an effective strategy to combat the disease as the fate of the evolution of the disease depends upon it. When R_{eff} is less than one, the infected population I will steadily decline to zero. Conversely, if R_{eff} is greater than one, the infected population will increase. In other words, when $\frac{dI(t)}{dt} < 0 \Rightarrow R_{eff} < 1$ and $\frac{dI(t)}{dt} > 0 \Rightarrow R_{eff} > 1$, the effective reproductive rate R_{eff} serves as a critical threshold that determines whether an infectious disease will rapidly extinguish or escalate into an epidemic.

At the onset of an epidemic, when $R_{eff} > 1$ and $S \approx 1$, the rate of increase in the infected population is approximated by $\frac{dI(t)}{dt} \approx (\beta - \gamma)I(t)$. Consequently, the infected population I will experience initial exponential growth, described by $I(t) = I(0)e^{(\beta-\gamma)t}$. The peak of the infected population occurs when the rate of change becomes zero, $dI(t)/dt = 0$, which corresponds to $R_{eff} = 1$. Following the peak, the infected population starts decreasing exponentially, following the pattern $I(t) \propto e^{-\gamma t}$. Eventually, as time approaches infinity ($t \to \infty$), the system converges toward $S \to 0$ and $I \to 0$. Interestingly, the existence of a threshold for infection may not be evident from recorded data but can be elucidated through the model. This distinction is crucial for identifying potential second waves, wherein a sudden increase in the susceptible population S results in $R_{eff} > 1$, leading to another exponential surge in the number of infections I[29].

To estimate the parameters β and γ for India, based on the data from May, 2021 to December, 2022, we simply define the cost to calibrate the model as a the mean absolute error:

$$cost = \frac{1}{T}\sum_{i=1}^{T} |(S - \hat{S})| + \frac{1}{T}\sum_{i=1}^{T} (I - \hat{I}) + \frac{1}{T}\sum_{i=1}^{T} (R - \hat{R}) \tag{7}$$

Where, S is the number of susceptible people and $\hat{S}$ is the predicted number of susceptible people, similarly, I for infected and $\hat{I}$ for the predicted number of infected people, and R for recovered. Minimizing this cost function using the Nelder-Mead method[40], we get:

$$\beta_{optimal} = 0.03925422815833437 \tag{8}$$

$$\gamma_{optimal} = 0.022659519392619114 \tag{9}$$

$$R_{0_{optimal}} = \frac{\beta_{optimal}}{\gamma_{optimal}} = 1.7323504297765755 \tag{10}$$

$$cost = \frac{1}{T}\sum_{i=1}^{T}|(S-\hat{S})| + \frac{1}{T}\sum_{i=1}^{T}(I-\hat{I}) + \frac{1}{T}\sum_{i=1}^{T}(R-\hat{R}) = 99920319.37922898 \tag{11}$$

Now that, we have a model - we want to say β is a time-varying parameter controlled by the stringency index, s. The stringency index is a composite measure based on nine response indicators including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (100 = strictest) [7]. This index simply records the strictness of government policies. It does not measure or imply the appropriateness or effectiveness of a country's response. A higher score does not necessarily mean that a country's response is 'better' than others lower on the index.

To define the new time-varying beta that is dependent on the current stringency index s_i, in accordance with our current optimal beta ($\beta_{optimal} = 0.03925422815833437$), we say,

$$\frac{dS}{dt} = -\beta(t)\frac{SI}{N} \tag{12}$$

$$\frac{dI}{dt} = \beta(t)\frac{SI}{N} - \gamma I \tag{13}$$

$$\frac{dR}{dt} = \gamma I \tag{14}$$

$$\beta(t) = \beta_{optimal} + (s_w * s(t)) \tag{15}$$

Where, $s(t)$ is the stringency index at time t and s_w is the stringency weight that we want to optimize. Optimizing this using the Nelder-Mead method we get:

$$s_{w_{optimal}} = -1.7905485677020872\text{e}{-7} \tag{16}$$

A negative weight implying the fact that a higher stringency index negatively affects β. The model's fit is better, as the $cost = 99898867.66702047$ which is 21451.712208509445 less than the previous model.

Now, that we have set up the following relation between β and s.

$$\beta_i = \beta_{optimal} + (s_{w_{optimal}} * s_i) \tag{17}$$

We investigate how stringency index affects the normalized Gross domestic product (GDP). Below are scatter plots with pearson correlation, for three countries from May, 2020 to December 2021.

We observe that stringency has a negative impact on the normalized Gross domestic product (GDP) of each country. Therefore, in some countries policies made during an epidemic have competing costs. For India after fitting a 4 degree polynomial we get the following equation:

$$\begin{aligned} GDP = 1.86904950e-06s^4 - 4.93161348e-04s^3 + 4.47470261e-02s^2 \\ -1.71334921e+00s^1 + 1.23142928e+02 \end{aligned} \tag{18}$$

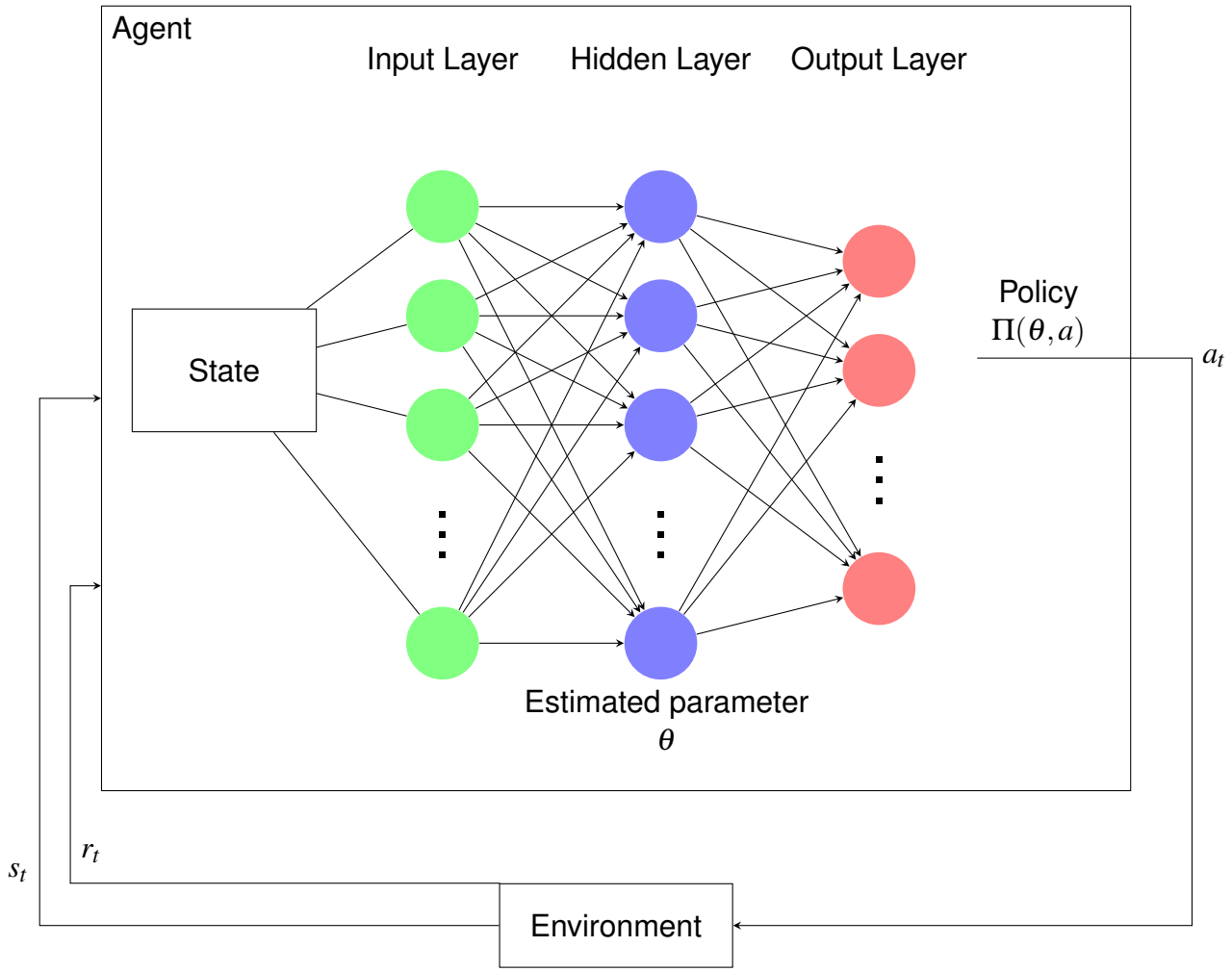
Given that the government is an agent that takes decisions in a deterministic environment defined above, we use reinforcement learning to model the competing costs of the environment. This environment is formally as a Markov decision process, and can be described as follows:

- https://jmlr.org/papers/volume22/20-1364/20-1364.pdf RL Paper that references https://araffin.github.io/slides/rl-tuto-jnrr19/#/13 RL ppt (for our RL notes).
- Set of States $\mathscr{S}$: The state of the enviroment are described through the descriptors like the normalized GDP $((GDP_{predicted} - GDP_{min})/(GDP_{max} - GDP_{min}))$, R_{eff}, a list of all the previous actions and the proportion of the population that was susceptible, infected and recovered. The starting states are simply these values at the starting date and no previous actions.
- Actions $\mathscr{A}$: The stringency index in the data from May, 2021 to December, 2022, has ranges from [28.7, 81.94], and the absolute differences between consecutive time points ranges from [0.2648768472906403, 13.420000000000002]. Based on this we define the discrete action space. There are 7 actions for the agent, it can keep the stringency index same, reduce/increase by 2.5, reduce/increase by 5, and reduce/increase by 10 given that the stringency index doesn't exceed 100 or go below 0.
- Transition dynamics $\mathscr{T}(\mathbf{s}_{t+1} \mid \mathbf{s}_t, \mathbf{a}_t)$ that map a state-action pair at time t onto a distribution of states at time $t+1$. This state transition is defined by the SIR model and the model of how stringency index affects the GDP.

- Immediate reward $\mathscr{R}(\mathbf{s}_t, \mathbf{a}_t, \mathbf{s}_{t+1})$. We compare several reward strategies. (Check rescaling of rewards - to do or not to do read: https://arxiv.org/pdf/1709.06560.pdf)
- Discount Factor $\gamma \in [0,1]$, where lower values place more emphasis on immediate rewards.

In general, the policy π is a mapping from states to a probability distribution over actions: $\pi : \mathscr{S} \to p(\mathscr{A} = \mathbf{a} \mid \mathscr{S})$. If the MDP is episodic, i.e., the state is reset after each episode of length T, then the sequence of states, actions and rewards in an episode constitutes a trajectory or rollout of the policy. Every rollout of a policy accumulates rewards from the environment, resulting in the return $R = \sum_{t=0}^{T-1} \gamma^t r_{t+1}$. The goal of RL is to find an optimal policy, π^*, which achieves the maximum expected return from all states.

Given that the stringency index plays a role in both controlling the GDP as well as, controlling the spread of infection, the agent can choose to increase or decrease the stringency index. The agent observes the percentage of the population that is susceptible, infected, recovered, and the GDP which extrapolated from the stringency index and all the past moves played, i.e., all the past stringency indexes decided by the agent.



Use this for the RL formulae: https://arxiv.org/pdf/1708.05866.pdf

Discussion

The paper seeks to inspire epidemiologists by highlighting the advancements achieved through the application of reinforcement learning in policy making during the pandemic. We introduce a virtual

environment that closely simulates a pandemic scenario and thoroughly explore innovative strategies for disease mitigation using reinforcement learning. Our proposed approach demonstrates compelling efficacy in achieving optimal decision-making, effectively balancing the formidable challenges posed by the pandemic and economic considerations. We are confident that this research contribution will forge a connection between epidemic studies and reinforcement learning, offering valuable insights to help humanity better defend against the ongoing pandemic crisis.

Experiment Settings

0.1 Dataset

Quaterly GDP data can be obtained from the Organisation for Economic Co-operation and Development: https://www.oecd-ilibrary.org/economics/data/main-economic-indicators/main-economic-indicators-complete-databas data-00052-en The population-level epidemiological data can be obtained from the Our World In Data COVID-19 dataset: https://ourworldindata.org/coronavirus

0.2 Code

We used stable-baseline3[41], Pytorch, Scipy, Pandas, Matplotlib, Python[42]. Code: https://github.com/psymbio/sir_rl

References

1. Baker, R. E. *et al.* Infectious disease in an era of global change. *Nat. Rev. Microbiol.* **20**, 193–205, DOI: 10.1038/s41579-021-00639-z (2022).
2. Tan, M. K. I. COVID-19 in an inequitable world: the last, the lost and the least. *Int. Heal.* **13**, 493–496, DOI: 10.1093/inthealth/ihab057 (2021). https://academic.oup.com/inthealth/article-pdf/13/6/493/41430650/ihab057.pdf.
3. Who coronavirus (covid-19) dashboard. https://covid19.who.int/. Accessed: 2023-11-30.
4. World economic outlook, april 2020: The great lockdown. Accessed: 2023-11-30.
5. Nicola, M. *et al.* The socio-economic implications of the coronavirus pandemic (covid-19): A review. *Int. J. Surg.* **78**, 185–193, DOI: 10.1016/j.ijsu.2020.04.018 (2020).
6. Gagnon, J. E., Kamin, S. B. & Kearns, J. The impact of the covid-19 pandemic on global gdp growth. *J. Jpn. Int. Econ.* **68**, 101258, DOI: 10.1016/j.jjie.2023.101258 (2023).
7. Anderson, R. M., Heesterbeek, H., Klinkenberg, D. & Hollingsworth, T. D. How will country-based mitigation measures influence the course of the covid-19 epidemic? *The Lancet* **395**, 931–934, DOI: 10.1016/s0140-6736(20)30567-5 (2020).
8. Song, S., Liu, X., Li, Y. & Yu, Y. Pandemic policy assessment by artificial intelligence. *Sci. Reports* **12**, 13843, DOI: 10.1038/s41598-022-17892-8 (2022).
9. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science* **368**, 395–400, DOI: 10.1126/science.aba9757 (2020).
10. Nguyen, T. *et al.* Covid-19 vaccine strategies for aotearoa new zealand: a mathematical modelling study. *The Lancet Reg. Heal. - West. Pac.* **15**, 100256, DOI: 10.1016/j.lanwpc.2021.100256 (2021).
11. Kim, D., Keskinocak, P., Pekgün, P. & Yildirim, The balancing role of distribution speed against varying efficacy levels of covid-19 vaccines under variants. *Sci. Reports* **12**, DOI: 10.1038/s41598-022-11060-8 (2022).

12. Jalloh, M. F. *et al.* Drivers of covid-19 policy stringency in 175 countries and territories: Covid-19 cases and deaths, gross domestic products per capita, and health expenditures. *J. Glob. Heal.* **12**, DOI: 10.7189/jogh.12.05049 (2022).

13. Caldwell, J. M. *et al.* Understanding covid-19 dynamics and the effects of interventions in the philippines: A mathematical modelling study. *The Lancet Reg. Heal. - West. Pac.* **14**, 100211, DOI: 10.1016/j.lanwpc.2021.100211 (2021).

14. Ferguson, N. *et al.* Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. DOI: 10.25561/77482 (2020).

15. De Foo, C. *et al.* Health financing policies during the covid-19 pandemic and implications for universal health care: a case study of 15 countries. *The Lancet Glob. Heal.* **11**, e1964–e1977, DOI: 10.1016/s2214-109x(23)00448-5 (2023).

16. Hollingsworth, T. D., Klinkenberg, D., Heesterbeek, H. & Anderson, R. M. Mitigation strategies for pandemic influenza a: Balancing conflicting policy objectives. *PLoS Comput. Biol.* **7**, e1001076, DOI: 10.1371/journal.pcbi.1001076 (2011).

17. Pangallo, M. *et al.* The unequal effects of the health–economy trade-off during the covid-19 pandemic. *Nat. Hum. Behav.* DOI: 10.1038/s41562-023-01747-x (2023).

18. Ash, T., Bento, A. M., Kaffine, D., Rao, A. & Bento, A. I. Disease-economy trade-offs under alternative epidemic control strategies. *Nat. Commun.* **13**, DOI: 10.1038/s41467-022-30642-8 (2022).

19. Ohi, A. Q., Mridha, M. F., Monowar, M. M. & Hamid, M. A. Exploring optimal control of epidemic spread using reinforcement learning. *Sci. Reports* **10**, DOI: 10.1038/s41598-020-79147-8 (2020).

20. Gagnon, J. E. & Rose, A. 23-8 how did korea's fiscal accounts fare during the covid-19 pandemic? Tech. Rep., Peterson Institute for International Economics (2023).

21. Deb, P., Furceri, D., Ostry, J. & Tawk, N. The economic effects of covid-19 containment measures. *IMF Work. Pap.* **20**, DOI: 10.5089/9781513550251.001 (2020).

22. Eichenbaum, M. S., Rebelo, S. & Trabandt, M. The macroeconomics of epidemics. *The Rev. Financial Stud.* **34**, 5149–5187, DOI: 10.1093/rfs/hhab040 (2021).

23. Lim, S. & Sohn, M. How to cope with emerging viral diseases: lessons from south korea's strategy for covid-19, and collateral damage to cardiometabolic health. *The Lancet Reg. Heal. - West. Pac.* **30**, 100581, DOI: 10.1016/j.lanwpc.2022.100581 (2023).

24. Coronavirus: South korea seeing a 'stabilising trend'. https://www.bbc.com/news/av/world-asia-51897979. Accessed: 2023-11-30.

25. Hale, T. *et al.* A global panel database of pandemic policies (oxford covid-19 government response tracker). *Nat. Hum. Behav.* **5**, 529–538, DOI: 10.1038/s41562-021-01079-8 (2021).

26. Hethcote, H. W. *Three Basic Epidemiological Models*, 119–144 (Springer Berlin Heidelberg, 1989).

27. Hethcote, H. W. *The Basic Epidemiology Models: Models, Expressions for R0, Parameter Estimation, and Applications*, 1–61 (WORLD SCIENTIFIC, 2008).

28. Allen, L. J. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infect. Dis. Model.* **2**, 128–142, DOI: https://doi.org/10.1016/j.idm.2017.03.001 (2017).

29. Cooper, I., Mondal, A. & Antonopoulos, C. G. A sir model assumption for the spread of covid-19 in different communities. *Chaos, Solitons amp; Fractals* **139**, 110057, DOI: 10.1016/j.chaos.2020.110057 (2020).

30. Bjørnstad, O. N., Shea, K., Krzywinski, M. & Altman, N. The seirs model for infectious disease dynamics. *Nat. Methods* **17**, 557–558, DOI: 10.1038/s41592-020-0856-2 (2020).

31. Mwalili, S., Kimathi, M., Ojiambo, V., Gathungu, D. & Mbogo, R. Seir model for covid-19 dynamics incorporating the environment and social distancing. *BMC Res. Notes* **13**, DOI: 10.1186/s13104-020-05192-1 (2020).

32. Nguyen, Q. D. & Prokopenko, M. A general framework for optimising cost-effectiveness of pandemic response under partial intervention measures. *Sci. Reports* **12**, DOI: 10.1038/s41598-022-23668-x (2022).

33. Bastani, H. *et al.* Efficient and targeted covid-19 border testing via reinforcement learning. *Nature* **599**, 108–113, DOI: 10.1038/s41586-021-04014-z (2021).

34. Francois-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G. & Pineau, J. An introduction to deep reinforcement learning. DOI: 10.48550/ARXIV.1811.12560 (2018).

35. Arulkumaran, K., Deisenroth, M. P., Brundage, M. & Bharath, A. A. Deep reinforcement learning: A brief survey. *IEEE Signal Process. Mag.* **34**, 26–38, DOI: 10.1109/msp.2017.2743240 (2017).

36. Henderson, P. *et al.* Deep reinforcement learning that matters. *Proc. AAAI Conf. on Artif. Intell.* **32**, DOI: 10.1609/aaai.v32i1.11694 (2018).

37. Dunn, W. N. *Public policy analysis* (Routledge, London, England, 2017), 6 edn.

38. Demir, T. & Miller, H. Policy communities. In *Handbook of Public Policy Analysis*, 137–147 (CRC Press, 2006).

39. HENS, N. *et al.* Seventy-five years of estimating the force of infection from current status data. *Epidemiol. Infect.* **138**, 802–812, DOI: 10.1017/S0950268809990781 (2010).

40. Gao, F. & Han, L. Implementing the nelder-mead simplex algorithm with adaptive parameters. *Comput. Optim. Appl.* **51**, 259–277, DOI: 10.1007/s10589-010-9329-3 (2010).

41. Raffin, A. *et al.* Stable-baselines3: Reliable reinforcement learning implementations. *J. Mach. Learn. Res.* **22**, 1–8 (2021).

42. Oliphant, T. E. Python for scientific computing. *Comput. Sci. Eng.* **9**, 10–20, DOI: 10.1109/MCSE.2007.58 (2007).