An Investigation of COVID-19 Spreading Factors with Explainable AI Techniques

Xiuyi Fan[1], Siyuan Liu[1], Jiarong Chen[2,3,4], Matthew Williams[4], and Thomas C. Henderson[5]

[1]Computer Science Department, Swansea University, United Kingdom
[2]Clinical Experimental Center, Jiangmen Key Laboratory of Clinical Biobanks and Translational Research, Jiangmen Central Hospital, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen 529030, China
[3]Department of Oncology, Jiangmen Central Hospital, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen 529039, China
[4]Computational Oncology Group, Imperial College London, United Kingdom
5School of Computing, University of Utah, USA

xiuyi.fan@seansea.ac.uk

Abstract

In this work, we adapt and apply Explainable AI (XAI) methods to conduct an assessment of the relative effectiveness of COVID-19 control measures implemented in 18 countries and regions using data from 23/01/2020 to 02/04/2020. Specifically, we model classification problems with the instantaneous reproduction number (R_t) as the prediction target and non-pharmaceutical control measures as model features; we then apply two XAI techniques, SHAP and ECPI, to analyze feature importance in our models. Our results show that city lockdown and contact tracing are the two most effective measures for reducing R_t, while measures such as mass testing, school closure, and public wearing face masks all have some effect in reducing R_t. Warm temperature also contributes to reducing the transmission. These results suggest that to prevent resurgent disease after countries or regions lift city lockdown, effort should be put to develop privacy preserving, practical and effective contact tracing techniques.

Keyword: COVID-19; Explainable AI; Non-pharmaceutical Control.

I. Introduction

Since COVID-19 was first identified in December 2019, various public health interventions have been implemented across the world. For example, Wuhan was locked down on January 23, 2020; France closed its schools on March 16, 2020; South Korea banned international travelers from Hubei on February 02, 2020; Singapore started contact tracing on January 23, 2020, etc. Though countries who have implemented these measures are seeing a reduced rate of number of confirmed cases growth, different countermeasures implemented by different countries at different time still pose the question: *which control measures are effective?* One way to answer this question is by looking at impact of control measures on R_t, the average number of secondary cases generated by one primary case with symptom onset on day t. R_t is one of the most important quantities used to measure the epidemic spread. If $R_t > 1$, then the epidemic is expanding at time t, whereas if $R_t < 1$, then it is shrinking at time t. Thus, factors that are most influential for controlling R_t are effective.

This paper presents an Explainable AI (XAI) study on understanding the relation between COVID-19 control measures and R_t, using Shapley additive explanations (SHAP) [1] and Explainable Classification with Probabilistic Inferences (ECPI) [2]. XAI is a rising field in AI. In addition to developing AI systems that make accurate predictions, XAI systems "explain" their predictions [3, 4, 5, 6, 7, 8]. The development of XAI is motivated by building trustworthy systems and revealing insights from data. Both SHAP and ECPI are designed to identify *decisive features* in prediction tasks. They are both data-driven; and they both can "explain" a prediction by pinpointing factors that are most important for the prediction based on the data provided. They are based on different underlying computation mechanisms with SHAP being a model-agnostic method that only computes feature importance and ECPI both making predictions as well as explanations.

Our study takes a two-step approach. In the first step, we estimate R_t from time series data (daily confirmed cases), then we produce a dataset containing both estimated R_t and control measures. In the second step, we build classification models from the constructed dataset for predicting R_t and explaining the predictions. In this way, the constructed classification models serve as surrogates to the real world; and identifying effective factors in controlling R_t becomes explaining the classifications.

From our results, we identify city lockdown and contact tracing as the two most effective countermeasures. Mass testing, mask use by the public, and warm weather are also useful for controlling transmission. For countries that are implementing city lockdown, once that is lifted, alternative measures should be in place to combat any possible resurgent spread. Contact tracing has been most successfully implemented with technology-based approaches such as tracking mobile phones, developing policies and technologies that enable contact tracing while providing privacy protection should be considered; promoting mask use and ensuring its supplies should also be considered.

The article is structured as follows. Section II presents several existing research on impact of non-pharmaceutical control measures. Section III presents our methods. Section IV presents the two XAI techniques used in this study and our results. We conclude in Section V.

II. Related Work

Intense effort has been put to study the effectiveness of control measures in containing the COVID-19 pandemic. In [9], the authors estimated the instantaneous reproduction number (R_t) of COVID-19 in 4 Chinese cities and 10 provinces, and the confirmed case-fatality risk (cCFR) in 4 Chinese cities and 31 Chinese provinces. They found that though aggressive non-pharmaceutical interventions (e.g., city lockdown) have made the first wave of COVID-19 outside of Hubei abated, control measures

should be relaxed gradually. Close monitoring of R_t and cCFR is needed to achieve an optimal balance between health and economic protection.

A quantitative analysis was carried out in [10] to study the effectiveness of travel restrictions and transmission control measures in preventing the spread of infection during the period from 31 December 2019 to 19 February 2020 in China. It was found that Wuhan travel ban and a few elements of national emergency response, e.g., suspending intracity public transport, closing entertainment venues, and banning public gatherings, were strongly associated with a delay in epidemic growth and a reduction in case numbers during the first 50 days of the COVID-19 epidemic in China. The impact of other factors as parts of national emergency response, such as the isolation of suspected and confirmed patients and their contact, is not yet clear. In addition, the analysis provided no evidence with respect to impact of the prohibition of travel between cities on reducing the number of cases in other cities across China. Another quantitative analysis to explore the contribution of travel restrictions on limiting the spread of COVID-19 in China was conducted in [11]. It was suggested that travel restrictions are particularly useful in the early stage of an outbreak, while it may be less effective once the outbreak is more widespread. The roles of each of other control measures, such as improved rates of diagnostic testing; clinical management; rapid isolation of suspected cases, confirmed cases, and contacts are yet not determined although the combination of them was successful in mitigating spread and reducing local transmission of COVID-19. The impact of lockdown (i.e., travel restriction, closing of public places as well as schools and universities, and limiting people's outside activities) in Wuhan was further investigated [12] through exploring the China domestic air traffic and passenger throughput data and the number of confirmed cases of COVID-19. The analysis suggested that stringent containment measures should be taken in seriously affected regions. In [13], a cohort study was conducted to investigate the of impact of the control measures (i.e., cordons sanitaire, traffic restriction, social distancing, home quarantine, centralized quarantine, and universal symptom survey) on the disease transmission in Wuhan. It was found that the institutions of the control measures were

associated with reduced effective reproduction number, suggesting that the measures were effective in improving control of the COVID-19 outbreak in Wuhan. It was also found that there was a short period of the increase of reproduction number although implementing the control measures, which may be due to a shortage of medical resources.

The effect of physical distancing measures, such as extended school closures and workplace distancing, on the progression of the COVID-19 epidemic was explored in [14]. An extensive simulation based on age-structured susceptible-exposed-infected-removed (SEIR) model [15, 16] was carried out. The simulation results show that sustained physical distancing measures have a potential in reducing the magnitude of the epidemic peak of COVID-19, while the timing to lift interventions is particularly important. Premature and sudden lifting could lead to secondary peak.

A stochastic transmission model was developed to assess the effect of isolation of cases and contact tracing in controlling onwards transmission from imported cases of COVID-19 in [17]. Through simulations parameterized by initial number of cases, the basic reproduction number, the delay from symptom onset to isolation, and the proportion of subclinical infections, it was found that case isolation or contact tracing alone is insufficient to control outbreaks, suggesting that further interventions would be required to achieve control.

The effectiveness of surveillance and containment measures for the first 100 patients with COVID-19 in Singapore was evaluated in [18]. The multipronged surveillance strategy adopted in Singapore includes applying the case definition at medical consults, tracing contacts of patients with laboratory-confirmed COVID-19, enhancing surveillance among different patient groups and allowing clinician discretion. The containment measures include patient isolation and quarantine, active monitoring of contacts, border controls, and community education and precautions. It was found that rapid

identification and isolation of cases, quarantine of close contacts, and active monitoring of other contacts have been effective in suppressing expansion of the outbreak.

A statistical analysis was conducted to assess the effect of community-wide mask usage to control COVID-19 in Hong Kong [19]. Based on sampling result, the compliance of face usage by HK general public was 96.6%. Given the situation that the incidence of COVID-19 cases in Hong Kong were significantly less than that of the counties that are with well-established health-care system and having over 100 confirmed cases, it was suggested that community-wide mask wearing may contribute to contain COVID-19 transmission in a densely populated city like HK.

The impact of physical distance measures in the UK was evaluated through comparing the contact patterns during the "lockdown" to patterns of social contact made before the epidemic period [20]. It was found that the estimated change in reproduction number significantly decrease, suggesting that the physical distancing measures adopted by the UK public will probably lead to a decline in cases though the decline will not happen immediately.

A raid review was conducted to assess the effects of quarantine of individuals [21]. The review suggested the existing modelling studies consistently showed the simulated quarantine measures alone are effective in containing COVID-19. When combining quarantine with other control measures, such as school closures, travel restrictions, and social distancing, more effects will be achieved in reducing new cases, transmissions and deaths.

The effect of different containment measures for mobility restrictions and contact reduction (e.g., mass testing and lockdown) in Italy was assessed through a spatially explicit SEIR -like transmission model [22]. The simulation resulted suggested that the restrictions posed to mobility and human-to-human interactions are effective in reducing disease transmission.

The effect of an individual of the control measures (i.e., case isolation in the home, voluntary home quarantine, social distancing of those over 70 years of age, social distancing of entire population, closure of schools and universities) and the combinations of them was simulated in [23]. It was found that the most effective combination to achieve epidemic suppression is population-wide social distancing combined with home isolation of cases and school and university closure. To achieve epidemic mitigation, an optimal policy combination will be case isolation, household quarantine, and social distancing of those at the higher risk.

An analysis about the necessities of wearing face marks as precautionary principle to contain COVID-19 was conducted in [24]. Some literatures (including preprints) were reviewed. Though these literatures did not present explicit results on whether wearing face masks will contain the transmission of COVID-19, the authors suggested that wearing face masks in the home and outside the home could have a substantial impact on containing COVID-19.

A list of the non-pharmaceutical control factors studied by the existing literatures is summarized in Table 1.

Table 1. Summary of the literatures studying the non- pharmaceutical control factors. GA = Government Advocation, MU = Mask Use, SC = School Closure, CL = City Lockdown, MT = Mass Testing, ITB = International Travel Ban, CT = Contact Tracing, T = Temperature, and H = Humidity.

Measures & Factors	**GA**	**MU**	**SC**	**CL**	**MT**	**ITB**	**CT**	**T**	**H**
Leung et al. [9]				✓					
Tian et al. [10]			✓	✓			✓		
Kraemer et al. [11]				✓	✓		✓		
Lau et al. [12]			✓	✓		✓			
Pan et al. [13]				✓	✓	✓			
Prem et al. [14]			✓						
Hellewell et al. [17]							✓		

Ng et al. [18]						✓	✓		
Cheng et al. [19]		✓							
Jarvis et al. [20]				✓					
Nussbaumer-Streit et al. [21]			✓	✓		✓	✓		
Gatto et al. [22]				✓	✓				
Ferguson et al. [23]			✓	✓					
Greenhalgh et al. [24]		✓							

III. Methodology

Our XAI-COVID framework is shown in Figure 1. A pre-processor is used to remove noise from data, estimate daily R_t, and discretize the data set. Outputs of the pre-processor are sent to several standard classifiers, ECPI and SHAP for prediction and explanation computation. Our XAI-COVID framework builds a model for COVID spreading factors in that it contains components to predict R_t as well as explains its prediction. We use the prediction results to validate the model and so that explanation results from ECPI and SHAP can be trusted.

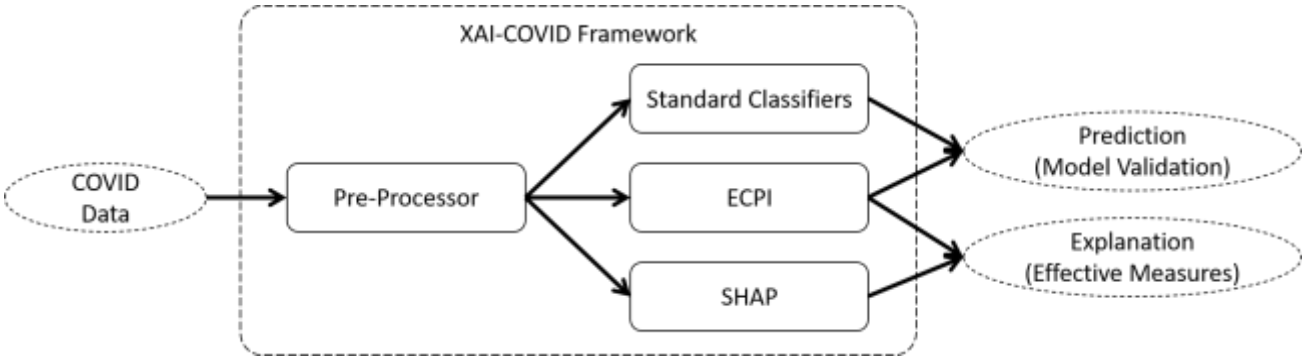


Figure 1. Our XAI framework for analyzing COVID spreading factors.

Our analysis is based on the following information:

- Implementation dates of control measures as shown in Table 2.
- The daily number of confirmed cases from 23/01/2020 to 02/04/2020 in countries and regions shown in Table 2.

- Temperature and humidity during our study period at these countries and regions.

Table 2. Implementation dates of control measures at 18 countries and regions. GA = Government Advocation, MU = Mask Use, SC = School Closure, CL = City Lockdown, MT = Mass Testing, ITB = International Travel Ban, CT = Contact Tracing.

Countries and Regions	**GA**	**MU**	**SC**	**CL**	**MT**	**ITB**	**CT**
Australia	13/03/2020					01/02/2020	
France	12/03/2020		16/03/2020	17/03/2020		16/03/2020	
Germany	28/01/2020		26/02/2020	16/03/2020		28/01/2020	
Italy	31/01/2020		04/03/2020	08/03/2020		31/01/2020	
Japan	24/01/2020	22/01/2020	02/03/2020			01/02/2020	25/02/2020
Singapore	22/01/2020	01/02/2020				29/01/2020	23/01/2020
South Korea	22/01/2020	22/01/2020	22/01/2020		31/01/2020	02/02/2020	22/01/2020
Spain	14/03/2020		12/03/2020	14/03/2020		10/03/2020	
United Kingdom	01/03/2020		20/03/2020	21/03/2020			
Beijing	24/01/2020	07/02/2020	22/01/2020	24/01/2020	24/01/2020	28/03/2020	24/01/2020
California	04/03/2020		13/03/2020	19/03/2020		02/02/2020	
Gurangdong	23/01/2020	26/01/2020	22/01/2020	24/01/2020	23/01/2020	28/03/2020	23/01/2020
Hong Kong	04/01/2020	08/01/2020	22/01/2020		04/01/2020	27/01/2020	04/01/2020
Hubei	20/01/2020	22/01/2020	22/01/2020	23/01/2020	05/02/2020	23/01/2020	03/02/2020
Macua	31/12/2019	03/02/2020	22/01/2020		20/02/2020	28/01/2020	
New York	07/03/2020		15/03/2020	20/03/2020	13/03/2020	02/02/2020	
Taiwan	20/01/2020	31/01/2020	22/01/2020		01/02/2020	23/01/2020	27/01/2020
Washington	29/02/2020		13/03/2020	23/03/2020	17/03/2020	02/02/2020	

Estimate R_t from Data

From daily number of confirmed cases, we estimate R_t. As presented in [25]. A *serial interval distribution* is used to model the time between a person getting infected and he/she subsequently infecting another person. Following the research in [25] and [26], we let this distribution be a *Gamma* distribution g with mean 7 and standard deviation 4.5 for all countries at all time. A plot of this function is shown in Figure 2.

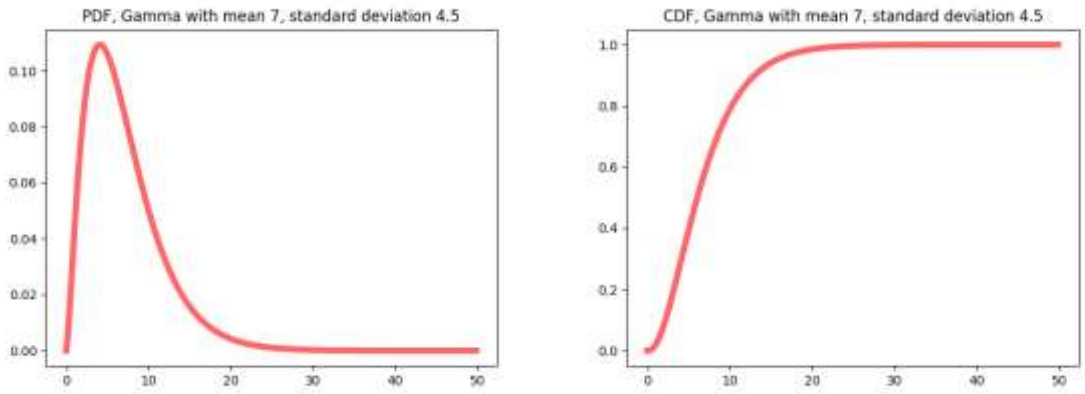


Figure 2. Plots of the Gamma PDF and CDF.

The number of new infections c_t on a given day t is given by the following discrete convolution function:

$$c_t = R_t \sum_{\tau=0}^{t-1} c_\tau g_{t-\tau}, \qquad (1)$$

where c_τ is the number of new infections on day τ,

$$g_s = \int_{\tau=s-0.5}^{s+0.5} g(\tau)d\tau,$$

for $s = 2, 3, \ldots$ and

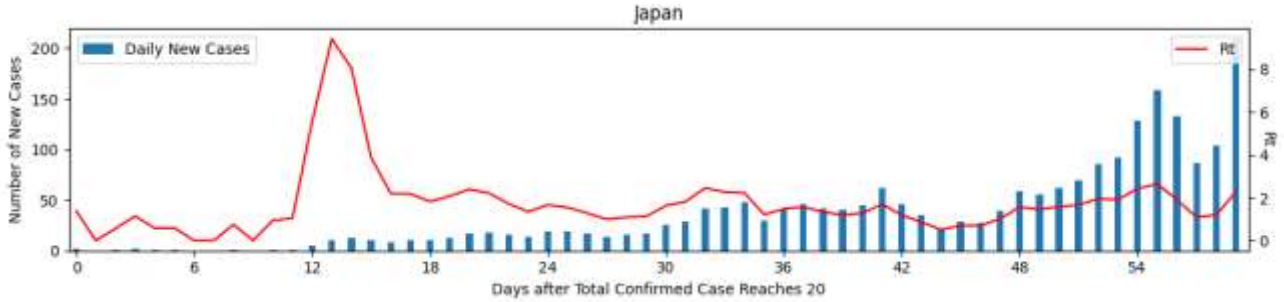
$$g_1 = \int_0^{1.5} g(\tau)d\tau.$$

We can see that new infections identified on day t depend on the number of new infections in days prior to t, weighted by the discretized serial interval distribution.

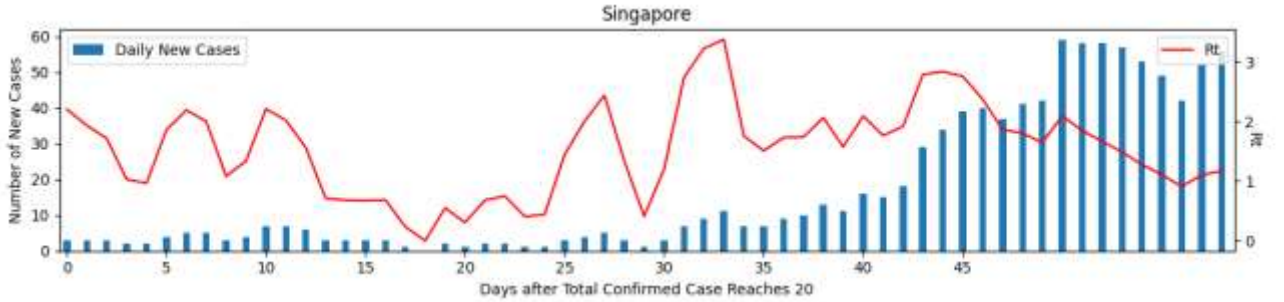
From Equation 1, solving for R_t, we have:

$$R_t = \frac{c_t}{\sum_{\tau=0}^{t-1} c_\tau g_{t-\tau}} \qquad (2)$$

c_t and c_τ are available from our data directly. For $x = t, \tau$, c_x is the difference between the confirmed case on day x and the confirmed case on day $x-1$. $g_{t-\tau}$ is obtained by integrating the Gamma distribution. Figure 3 illustrates the estimated R_t for Japan, Singapore, Australia and Hubei.

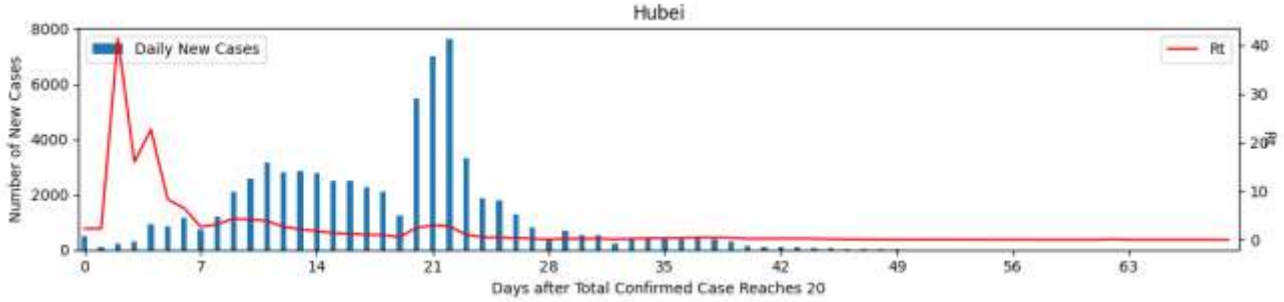


(a)



(b)

(c)



(d)

Figure 3. Daily new cases and computed R_t for selected countries and regions: (a) Japan; (b) Singapore; (c) Australia; (d) Hubei. We can see that R_t is a function of both the number of new infections on day t, and the numbers of new infections on the days prior to t.

A. *Data Pre-Processing for Noise Removal*

Since R_t is sensitive to noise in the number of new infection cases and the data set we use contains imperfection, e.g., for the United Kingdom, both 14/03/2020 and 15/03/2020 have 1140 confirmed cases so there is no increase on 15/03/2020, we thus run a sliding-window mean filter with radius 1 on the data to remove noise. , i.e., for a day t, let its confirmed case and filtered confirmed case be C_t and F_t, respectively, then $F_t = (C_{t-1} + C_t + C_{t+1})/3$.

We then compose a data set in a tabular form where each row describes information for one country/region on a day, containing the number of new confirmed cases on that day, days since each of the control measures that have been implemented, and the temperature as well as humidity of that day. R_t is added to every row in the data set and later used as the target for prediction.

As R_t calculated in Equation 2 assumes a reasonably large t, (otherwise both c_τ and $g_{t-\tau}$ would be too small, resulting in an artificially large R_t, which is unlikely to be correct), we drop entries with confirmed case less than 20. In other words, we only use data where there are more than 20 accumulated confirmed cases in that country/region; and as the number of confirmed cases is monotonically increasing, there is no "skipped" date. For instance, our Singapore cases start on 03/02/2020 and Japan cases start on 02/02/2020. The full starting dates used for all countries and regions are shown in Table 3.

A fraction of this data set is shown in Table 4 for illustration. The data in the first row records that R_t is 0.31, with 78 new confirmed cases on that day, 55 days after government advocation, 55 days after mass use of face mask, school closure, and city lockdown not implemented, 46 days after mass testing, 44 days after international travel ban, 55 days after contact tracing, as well as temperature and humidity being 3.73 and 48.47, respectively. The data set contains 800 entries in total.

Table 3. Case starting dates for countries and regions used in this study.

Countries	Dates	Regions	Dates
Australia	03/03/2020	Beijing	23/01/2020
France	26/02/2020	California	03/03/2020
Germany	25/02/2020	Guangdong	23/01/2020
Italy	21/02/2020	Hong Kong	05/02/2020
Japan	02/02/2020	Hubei	23/01/2020
Singapore	03/02/2020	Macau	22/03/2020
South Korea	06/02/2020	New York	05/03/2020
Spain	27/02/2020	Taiwan	16/02/2020
United Kingdom	29/02/2020	Washington	03/03/2020

Table 4. An illustration of the data set with four data entries (Singapore, 12/02/2020, Japan, 26/03/2020, Germany, 26/03/2020, South Korea, 16/03/2020, and Guangdong, 08/02/2020). NC = New Case, GA = Government Advocation, MU = Mask Use, SC = School Closure, CL = City Lockdown, MT = Mass Testing, ITB = International Travel Ban, CT = Contact Tracing, T = Temperature, and H = Humidity.

R_t	NC	GA	MU	SC	CL	MT	ITB	CT	T	H
0.31	78	55	55	55	0	46	44	55	3.73	48.47
0.72	53	17	14	18	16	17	0	18	15.89	62.66
1.34	4	22	12	0	0	20	15	21	27.86	83.86
1.91	92	63	65	25	0	0	55	31	17.375	32.75
2.14	5962	59	0	30	11	0	12	0	6.19	39.35

B. Data Discretization

To obtain interpretable qualitative results and to further mask noises in data, we discretize our data into the following intervals.

- NC (Number of New Cases): $[0,10), [10, 100), [100, \infty)$.
    - The number of new cases is put into 3 intervals, represented with integers 0, 1, and 2, respectively. For instance, given a record with $\mathrm{NC} = 78$, since $10 \leq 78 < 100$ and $[10, 100)$ is the second interval for discretizing NC, $\mathrm{NC} = 78$ is mapped to 1; given

another record with$\mathrm{NC} = 4$, since $0 \leq 4 < 10$, and $[0, 10)$ is the first interval, $\mathrm{NC} = 4$ is mapped to 0.

- GA (Government Advocation), MU (Mask Use), SC (School Closure), CL (City Lockdown), MT (Massive Testing), ITB (International Travel Ban), CT (Contact Tracing): $[0,1), [1, 5), [5, 10), [10,15), [15, \infty)$.
    - Each of GA, MU, SC, CL, MT, ITB and CT is discretized into 5 intervals, represented by integers 0, 1, 2, 3, and 4, respectively. For instance, with $\mathrm{GA} = 55$, since $15 \leq 55 < \infty$, $\mathrm{GA} = 55$ is mapped to 4; and with $\mathrm{CL} = 0$, since$0 \leq 0 < 1$, $\mathrm{CL} = 0$ is mapped to 0.
- T: $(-\infty, 0), [0, 10), [10, 20), [20, \infty)$.
    - Temperature is discretized into 4 intervals.
- H: $[0, 40), [40, 80), [80, \infty)$
    - Humidity are discretized into 3 intervals.

Table 5 shows the result of discretization from data shown in Table 4. Discretization is a key step in our process, as shown by our sensitivity analysis in Section IV, where results from two different sets of discretization boundaries are presented, the specific choices of interval boundaries do not affect the conclusions.

Table 5. Five data entries in Table 4 after discretization. NC = New Case, GA = Government Advocation, MU = Mask Use, SC = School Closure, CL = City Lockdown, MT = Mass Testing, ITB = International Travel Ban, CT = Contact Tracing, T = Temperature, and H = Humidity. For the first row of Table 4, with $R_t = 0.31, NC = 78, GA = 55, MU = -55, SC = 55, CL = 0, MT = 46, ITB = 44, CT = 55, T = 3.73, H = 48.47$, it is discretized as shown in the first row of this table, with $R_t = 0.31, NC = 1, GA = 4, MU = 4, SC = 4, CL = 0, MT = 4, ITB = 4, CT = 4, T = 1, H = 1$.

R_t	NC	GA	MU	SC	CL	MT	ITB	CT	T	H
0.31	1	4	4	4	0	4	4	4	1	1
0.72	1	4	3	4	4	0	0	4	2	1
1.34	0	4	3	0	0	3	3	4	3	2
1.91	1	4	4	4	0	4	4	4	2	0
2.14	2	4	0	4	3	3	3	0	1	0

IV. XAI Techniques

With data presented in the form shown in Table 5, we pose R_t prediction as an explainable classification task: given a data entry (row), (1) classify whether the R_t (for that row) is greater than some threshold θ, and (2) identify the features that are most influential to the classification.

A. *Methods for R_t Classification*

Three methods are used for classification: random forest, neural network and ECPI. Since the reliability of our explanation results depends on the quality of data and the discretization process, the classification step with performance evaluation not only makes a prediction about R_t, but also helps us to verify the correctness of discretization. Standard implementations for random forest and neural networks are used.

ECPI is an explainable classification algorithm based on probabilistic logic[27]. Roughly speaking, ECPI maps a dataset into a knowledge-base (KB) in probabilistic logic and performs classification with probabilistic logic inferences. Rules in an ECPI KB are Horn clauses with associated probabilities, such as

$$[p]\ Class_x \leftarrow FeatureValue_1, \dots, FeatureValue_n,$$

which is read as: with feature values 1 to n, the probability of class x is p. For instance,

$$[0.167]\ R_t \geq 1 \leftarrow \ \mathrm{MU} = 2,\ \mathrm{ITB} = 2,\ \mathrm{H} = 1.$$

is a rule in the KB and it is read as: with $\mathrm{MU} = 2, \mathrm{ITB} = 2,\ \mathrm{H} = 1$, the probability of $R_t \geq 1$ is $P(R_t \geq 1) = 0.167$. The probability p of each rule

$$h \leftarrow b_1, \dots, b_n$$

is $p = P(h|b_1, \dots, b_b)$. For a dataset with m features, its corresponding KB contains rules over a subset of k-combination $(k = 1, .., m)$ of features.

With the KB constructed, ECPI solves classification by computing probabilities of all classes with feature values in the query "appended" to the KB. The class with the highest probability is the resulting class. For instance, suppose we want to know whether $R_t \geq 1$ under the query $\mathrm{NC} = 2, \mathrm{GA} = 4, \mathrm{MU} = 4, \mathrm{SC} = 4, \mathrm{CL} = 4, \mathrm{MT} = 4, \mathrm{ITB} = 4, \mathrm{CT} = 4, \mathrm{T} = 1, \mathrm{H} = 1$, we append the lines in Table 6 to the KB, and then estimate $P(R_t \geq 1)$ and $P(R_t < 1)$, respectively. We return the class with the higher probability. We use the linear programming optimization approach introduced in [2] for the required probability estimation.

Table 6. A query example.

[1] NC = 2,	[1] GA = 4,	[1] MU = 4,	[1] SC = 4,	[1] CL = 4,
[1] MT = 4,	[1] ITB = 4,	[1] CT = 4,	[1] T = 1,	[1] H = 1.

B. Methods for Explanation

We use two different XAI methods for identifying decisive features in this study: ECPI and Shapley additive explanations. Both methods give "local'" explanations in that for a given query, they identify the decisive features for that query. Thus, with both methods, we first perform classification, then identify the features that are most influential to the classification. We do this analysis to all individual entries in our dataset, and then aggregate the results to form global interpretations.

Generally speaking, ECPI computes explanations by identifying the subset of features that gives the same classification result as the full set. For a dataset with p features, suppose that a query (with p feature values) gives class c, to find the top k features explaining the classification, we compute $P(c)$ with $\binom{p}{k}$ "partial queries'". For instance, for the query shown in Table 6, we compute $P(R_t < 1) = 0.66$ and $P(R_t \geq 1) = 0.34$ thus conclude $R_t < 1$. To find a 1-feature explanation for this classification, we compute $P(R_t < 1)$ with 10 individual features (these are the 1-combinations from the set of 10 features) shown in Table 6. The result is shown in Figure 4. We can see that

$P(R_t < 1) > 0.5$ for $\mathrm{CT} = 4, \mathrm{MT} = 4, \mathrm{CL} = 4$ with its highest value obtained with $CL = 4$. Thus $CL = 4$ is the one-feature explanation for this query.

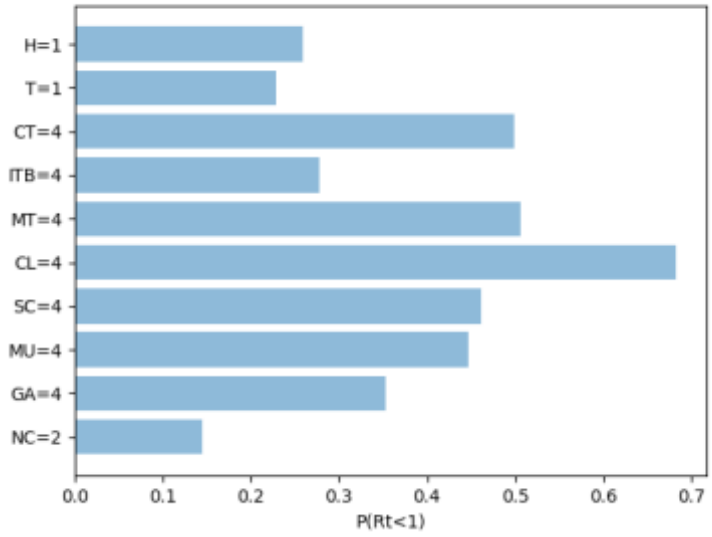


Figure 4. $P(R_t < 1)$ with each of the 10 individual features in the set of features values in Table 6.

SHAP is based on Shapley value [28], a game theory concept that assigns a unique distribution of a total surplus generated by the coalition of all players in a cooperative game. In our context, each feature with its value, e.g., NC = 0 or MU = 3, are viewed as "players" in the game where the outcome is in one of the two classes (R_t being either greater than 1 or not). Shapley value φ for each feature-value describes its "contribution" to the outcome classification. Formally,

$$\varphi_j(v) = \sum_{S \subseteq \{x_1,\ldots,x_p\} \{x_j\}} \frac{|S|!\,(p - |S| - 1))!}{p!} \left(v\left(S \cup \{x_j\}\right) - v(S)\right)$$

where *S* is set of the features used in the model, *x* is the vector of feature values of the instance to be explained and p is the number of features. $v(S)$ is the prediction for feature values in set *S* that are marginalized over features that are not included in set *S*. We use the TreeSHAP method presented in [29] to calculate the Shapley values and features with the highest Shapley values are deemed as the most influential ones. For instance, given the query shown in Table 6, the calculated Shapley

values are shown in Figure 5. Here, we see that $\mathrm{CL} = 4$ is the most decisive feature, followed by $CT = 4$.

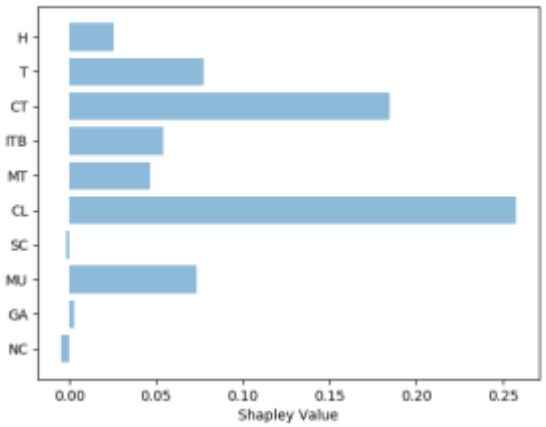


Figure 5. Shapley values for the 10 features values in Table 6.

Results

With SHAP and ECPI, we study the two classification cases for $0 \leq R_t < 1$ and $0 \leq R_t < 2$. For each case, we compute the top feature that is the most influential. There are 228 and 435 entries with $0 \leq R_t < 1$ and $0 \leq R_t < 2$, respectively. The results are shown in Figure 6.

From Figures 6, several qualitative interpretations can be obtained. We can see that for $R_t \leq 1$, *City Lockdown (CL)* and *Contact Tracing (CT)* are most influential for $k = 1, 2$. Both measures are effective when they take value 4, meaning they are implemented for more than 15 days. For $R_t \leq 2$, *City Lockdown (CL), Contact Tracing (CT), Mask Use (MU), Mass Testing (MT), School Closure (SC)* are most influential. Humidity plays no role completely whereas warm temperature could be helpful. The bars shown on *New Cases (NC)* might be interpreted as: when *NC* is sufficiently small (<10), it is likely to stay in that way.

Figure 7 shows influential factors for daily new cases in three ranges: 0-10, 10-100, and >100 for $R_t \leq 1$ and $R_t \leq 2$, respectively. For $R_t \leq 1$ we see that *CL* is the most effective single measure,

when there are more than 10 cases per day. For $R_t \leq 2$, the top effective measures are *CL*, *CT*, *MT* and *MU*.

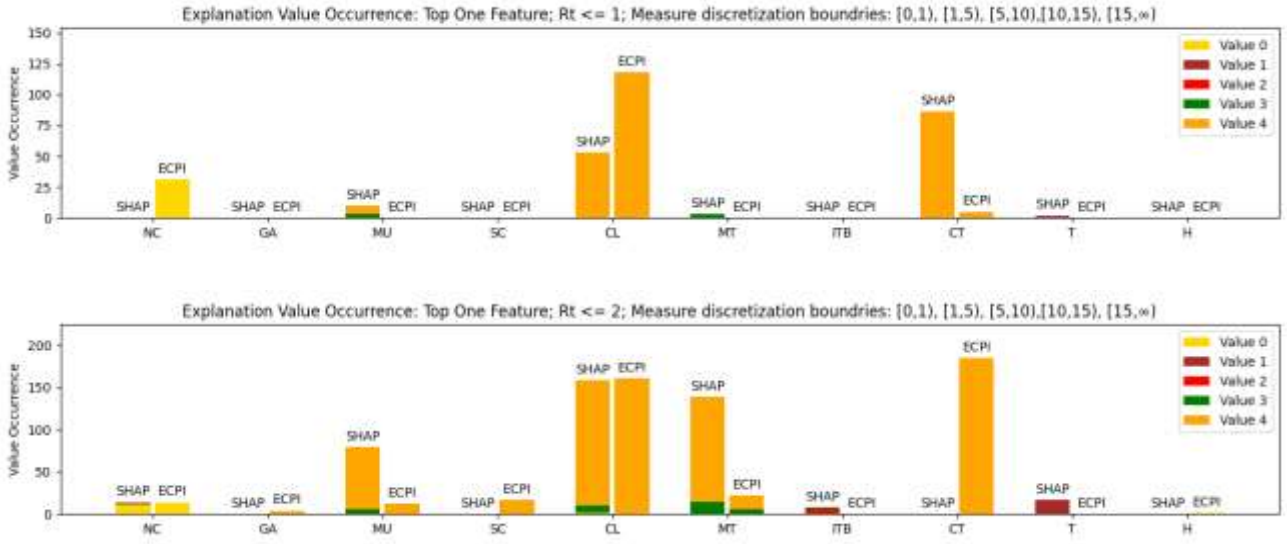


Figure 6. Top influential factors for $R_t \leq 1$ and $R_t \leq 2$. The x-axis is the control measures together with temperature and humidity, and the y-axis is the number of entries in the dataset identified by SHAP or ECPI as the explanations. A measure with more entry occurrences is considered to be more effective. The top figure shows the most effective measure to achieve $R_t \leq 1$. SHAP considers $CT = 4$ (implementing contact tracing for 15 days or more) as the most effective for achieving $R_t \leq 1$; and ECPI identifies $CL = 4$ (implementing city lockdown for 15 days or more) as the most effective. SHAP considers $CL = 4$ influential as well, but not to the same level as *CT*. The bottom figure shows the most effective measure to achieve $R_t \leq 2$. SHAP considers $CL = 4$ (implementing city lockdown for 15 days or more) as the most effective measure for achieving $R_t \leq 2$, and $MT = 4$ and $MU = 4$ (implementing massive testing and mask use for 15 days or more) influential as well, but not to the level of *CL*. ECPI considers $CT = 4$ (implementing contact tracing for 15 days or more) as the most effective measure; and $CL = 4$ being almost as effective as *CT*.

V. Discussion and Conclusion

As a data-driven modeling approach, our work is limited by a number of factors. Firstly, all results are based on data collected from the selected 18 countries and regions during the period of 23/01/2020 to 03/04/2020. We select these countries and regions to cover wide ranges of implemented non-pharmaceutical control measures as well as temperature and humidity conditions. For instance, South Korea, Hong Kong, Macau, and Taiwan are included and they have not implemented city lockdown during the studied period; Singapore and Australia are included and they both have higher temperatures than other places. Although results obtained might not be generalizable, they are about

these regions during the said period. Thus, when applying these results to other regions and other time, they should be viewed as indicative.

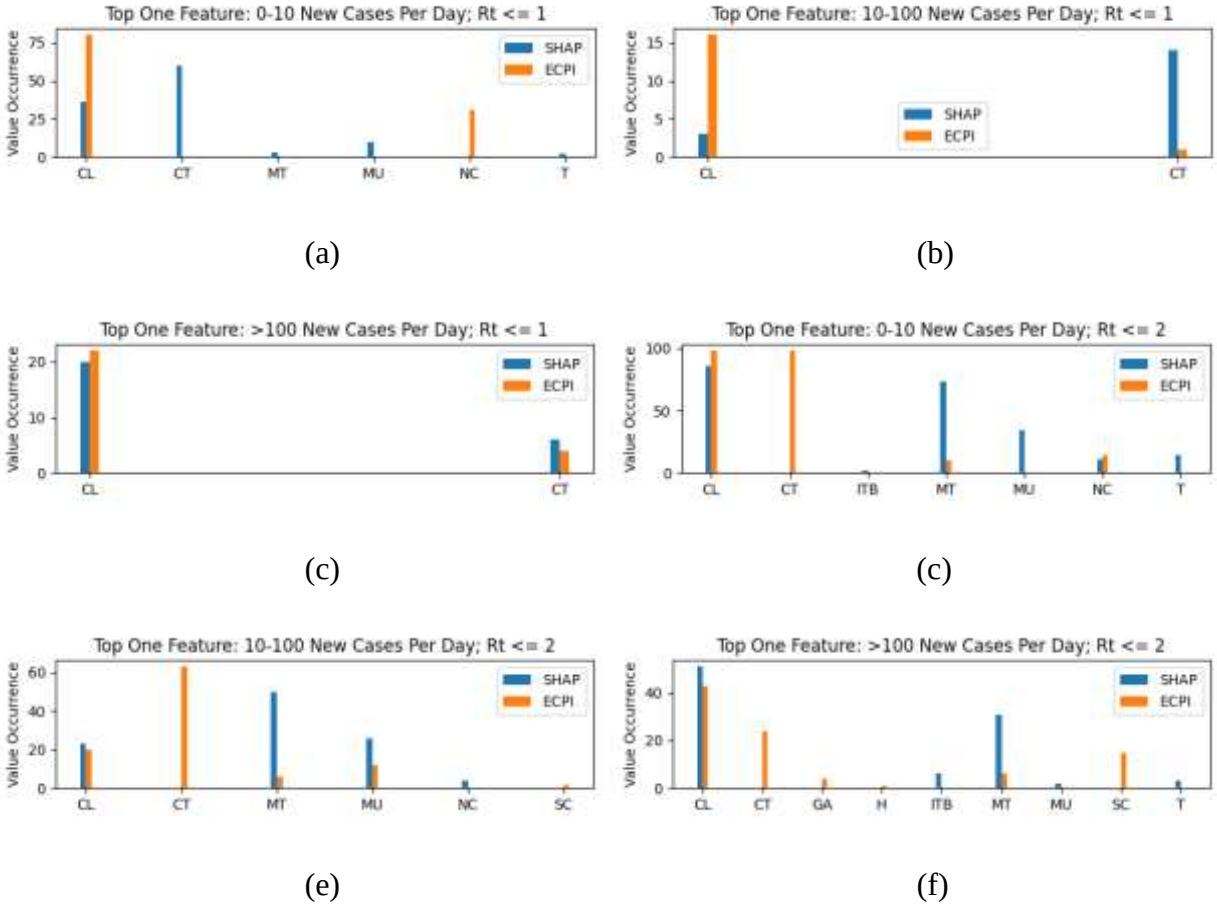


Figure 7. Influential factors for daily new cases in different ranges. The x-axis is the control measures; and the y-axis is the number of entries in the dataset identified by SHAP or ECPI as the explanations. A measure with more entry occurrences is considered to be more effective. (a) shows the most influential measure to achieve $R_t \leq 1$when the number of daily new cases is in the range of 0-10. SHAP and ECPI consider *CT (contact tracing)* and *CL (city lockdown)* as the most influential measure, respectively; (b) shows the most influential measure when the number of daily new cases is in the range of 10-100. SHAP and ECPI consider *CT* and *CL* as the most influential measures. (c) shows the most influential measure to achieve $R_t \leq 1$ when the number of daily new cases is greater than 100. Again, SHAP and ECPI consider *CT* and *CL* as the most influential measure. (d-f) show results under the same conditions for $R_t \leq 2$.

Secondly, the data used is inherently ambiguous, e.g., “contact tracing"” and “mass testing"” have been implemented at different countries, but it is unlikely the same standard has been applied. The exact definition of “city lockdown” is ambiguous as well. We have loosely taken it as various measures aimed at people staying at home to various degrees, with closure of leisure and recreation facilities. Consequentially, although our methods are quantitative, due to the qualitative nature of the data, one should read our results qualitatively.

Thirdly, we rely on the calculated R_t to label our data into different classes (e.g., $R_t \leq 1$ vs. $R_t > 1$), which are then used to construct our models. The calculation method is reported in [25] with parameters found in [26]. These are the core assumptions in this work. Namely, we assume the underlying serial interval distribution can be modelled with a Gamma function and this function is the same for all countries and regions studied. These are the assumptions also made in [25] and [26]. Moreover, we are aware that R_t is an estimate that can be approximated with more than one method, some authors such as [9] gives a much smaller estimate of R_t for Beijing in January (they estimate R_t being close to 0.5 whereas our calculation shows it is greater than 2; although ours drops to below 0.5 after February 10, same as theirs), different results might be obtained if R_t is estimated differently.

Lastly, we have ignored the consideration where as COVID-19 progresses, case reporting mechanisms have been evolved. Thus, there might be some discrepancy when comparing numbers of newly reported cases at different times. Although we acknowledge such discrepancy may exist, we believe it does not invalidate our results due to our discretization and classification approach, where values are mapped to classes with small discrepancies "masked" in this process. Alternatively, we can pose this analysis as a regression/XAI problem - instead of constructing classification models, we can build regression models that predict R_t and then identify key factors in such predictions. This approach potentially requires more and higher quality data and asks for different machine learning models. We plan to explore this direction in future work.

In conclusion, we applied two explainable AI methods in studying the importance of factors affecting the spread of COVID-19. We find city lockdown and contact tracing being the two most effective control measures, surpassing mass testing, school closure, international travel ban and mask use. As countries are considering lifting city lockdown, to prevent resurgent disease, effort should be put to developing privacy preserving, practical and effective contact tracing techniques.

References

[1] S. M. Lundberg, S. Lee, A unified approach to interpreting model predictions, in: Proc. of NIPS, 2017.

[2] X. Fan, S. Liu, T. C. Henderson, Explainable ai for classification using probabilistic logic inference, arXiv (2020).

[3] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, AI Journal (2019).

[4] O. Biran, C. V. Cotton, Explanation and justification in machine learning: A survey, in: Proc. of IJCAI-17 Workshop on Explainable AI, 2017.

[5] D. Doran, S. Schulz, T. R. Besold, What does explainable AI really mean? A new conceptualization of perspectives, in: Proc. of AI*IA, 2017.

[6] E. Elenberg, A. Dimakis, M. Feldman, A. Karbasi, Streaming weak sub-modularity: Interpreting neural networks on the fly, in: Proc. of NIPS, 2017.

[7] M. T. Ribeiro, S. Singh, C. Guestrin, Anchors: High-precision model-agnostic explanations, in: Proc of AAAI-18, 2018.

[8] S. Wachter, B. Mittelstadt, L. Floridi, Transparent, explainable, and accountable AI for robotics, Science Robotics (2017).

[9] K. Leung, J. T. Wu, D. Liu, G. M. Leung, First-wave covid-19 transmissibility and severity in china outside hubei after control measures, and second-wave scenario planning: a modelling impact assessment, The Lancet (2020).

[10] H. Tian, Y. Liu, Y. Li, C.-H. Wu, B. Chen, M. U. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang, et al., An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in china, Science (2020).

[11] M. U. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, et al., The effect of human mobility and control measures on the covid-19 epidemic in China, Science (2020).

[12] H. Lau, V. Khosrawipour, P. Kocbach, A. Mikolajczyk, J. Schubert, J. Bania, T. Khosrawipour, The positive impact of lockdown in wuhan on containing the covid-19 outbreak in china, Journal of travel medicine 27 (3) (2020).

[13] A. Pan, L. Liu, C. Wang, H. Guo, X. Hao, Q. Wang, J. Huang, N. He, H. Yu, X. Lin, et al., Association of public health interventions with the epidemiology of the covid-19 outbreak in wuhan, China, Jama 323 (19) (2020) 1915–1923.

[14] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday, et al., The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, China: a modelling study, The Lancet Public Health (2020).

[15] P. Klepac, L. W. Pomeroy, O. N. Bjørnstad, T. Kuiken, A. D. Osterhaus, J. M. Rijks, Stage-structured transmission of phocine distemper virus in the dutch 2002 outbreak, Proceedings of the Royal Society B: Biological Sciences 276 (1666) (2009) 2469–2476.

[16] P. Klepac, H. Caswell, The stage-structured epidemic: linking disease and demography with a multi-state matrix approach model, Theoretical Ecology 4 (3) (2011) 301–319.

[17] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, et al., Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts, The Lancet Global Health (2020).

[18] Y. Ng, Z. Li, Y. X. Chua, W. L. Chaw, Z. Zhao, B. Er, R. Pung, C. J. Chiew, D. C. Lye, D. Heng, et al., Evaluation of the effectiveness of surveillance and containment measures for the first 100 patients with covid-19 in Singapore–january 2–february 29, 2020 (2020).

[19] V. C. Cheng, S.-C. Wong, V. W. Chuang, S. Y. So, J. H. Chen, S. Sridhar, K. K. To, J. F. Chan, I. F. Hung, P.-L. Ho, et al., The role of community-wide wearing of face mask for

control of coronavirus disease 2019 (covid-19) epidemic due to sars-cov-2, Journal of Infection (2020)

[20] C. I. Jarvis, K. Van Zandvoort, A. Gimma, K. Prem, P. Klepac, G. J. Rubin, W. J. Edmunds, Quantifying the impact of physical distance measures on the transmission of covid-19 in the UK, BMC medicine 18 (2020) 1–10.

[21] B. Nussbaumer-Streit, V. Mayr, A. I. Dobrescu, A. Chapman, E. Persad, I. Klerings, G. Wagner, U. Siebert, C. Christof, C. Zachariah, et al., Quarantine alone or in combination with other public health measures to control covid-19: a rapid review, Cochrane Database of Systematic Reviews (4) (2020).

[22] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, A. Rinaldo, Spread and dynamics of the covid-19 epidemic in Italy: Effects of emergency containment measures, Proceedings of the National Academy of Sciences 117 (19) (2020) 10484–10491.

[23] N. Ferguson, D. Laydon, G. Nedjati Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, G. Cuomo-Dannenburg, et al., Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand (2020).

[24] T. Greenhalgh, M. B. Schmid, T. Czypionka, D. Bassler, L. Gruer, Face masks for the public during the covid-19 crisis, Bmj 369 (2020).

[25] S. Flaxman, S. Mishra, A. Gandy, H. Unwin, H. Coupland, T. Mellan, H. Zhu, T. Berah, J. Eaton, P. Perez Guzman, et al., Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on covid-19 in 11 European countries, Tech. rep., Imperial College London (2020).

[26] J. T. Wu, K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, G. M. Leung, Estimating clinical severity of covid-19 from the transmission dynamics in wuhan, China, Nature Medicine (2020) 1–5.

[27] N. Nilsson, Logic and artificial intelligence, Artificial Intelligence 47 (1-3) (1991) 31–56.

[28] S. Hart, Shapley Value, Palgrave Macmillan UK, London, 1989, pp. 210–216. URL https://doi.org/10.1007/978-1-349-20181-5_25.

[29] S. M. Lundberg, G. G. Erion, S. Lee, Consistent individualized feature attribution for tree ensembles, CoRR abs/1802.03888 (2018).