Abstract: Nearly every aspect of healthcare is quantifiable. These aspects extend from calculating the extensiveness of a certain disease for risk management, to evaluating effectiveness of treatment to different patients, solving problems of resource allocation. This movement of enumerate healthcare seems to be reasonable, since lots of aspects of healthcare are moving to a more standardized, agreeable measure as form of communication and the needs for collaboration of ideas. The amount of data generated by health care increase in minutes. Such large amount of data require us better data-driven approach. Any system of healthcare will encounter the problem with filtering the core insights that will truly bring values to both the patients and the system. Therefore, modeling has been a great tool for healthcare to narrow down and extracting these core insights.

Introduction: Mathematical modeling of healthcare is most noticeable in studying the spread of infectious disease. Researchers have used it for evaluating and planning means of interventions. The terrible global events of COVID 19, amongst the advance of the means of disease surveillance and rapid diagnostics test have made the application for mathematical modeling more demanding. Mathematical models have been used to formulate and test basic hypothesises, dividing strategies and policies for disease control, and extract the key parameters for studying infectious disease's behaviour. In fact, the field mathematical epidemiology is the combination of rigorous statistical testing with mathematical models to derive these key insights to understand these diseases.

In this topic, we will define what models in healthcare is and how we understanding transmit diseases. We will discuss the SIR model and the occasional SEIR model in healthcare. Then we finalize with our recommendations for modeling in healthcare using more extensive methods such as machine learning (deep learning) and finally, our (failed) example and (failed) attempts of implementing deep learning to predict COVID-19 rate.

We will define two key terms: modeling in healthcare and infectious disease. **Healthcare modeling** refers to the use of mathematical and statistical models to analyze and understand healthcare data and make predictions, evaluation or recommendations about healthcare outcomes. This can include predicting the likelihood of certain medical events or conditions, such as the likelihood of a patient developing a particular disease or the likelihood of a patient responding to a particular treatment. For COVID 19 and infectious disease, it's predicting the likelihood of a patients get infected and transmit a disease, or in another word, estimating and predicting the reproduction rate of a disease (more on reproduction rate later.)

Healthcare modeling can be used in a variety of applications, including population health management, disease prediction and prevention, cost analysis and resource planning, and personalized medicine. It can involve the use of machine learning algorithms, statistical modeling techniques, and other data-driven approaches to analyze healthcare data and make informed decisions about healthcare outcomes. Modeling appear as a means of transport, also median for these intersections. Nowadays we can see how healthcare systems are moving towards a more data-driven approach: we're creating facilities opportunities for cross communications of knowledge and resources - where third parties companies like 23andMe provided partnerships with a number of healthcare organizations and research institutions. For example, the company has a partnership with the Broad Institute of MIT and Harvard, a research organization that focuses on genomics and biomedical research. Through this partnership, 23andMe has contributed genetic data to the Broad Institute's database, which is used for research purposes.

Some examples of healthcare modeling include predicting the likelihood of hospital readmissions, identifying risk factors for certain diseases, and predicting the effectiveness of different treatments for different patients. "Healthcare modeling can be used to improve the quality and efficiency of healthcare delivery" (Opioid Epidemic & Health IT - Health IT Playbook.

https://www.healthit.gov/playbook/opioid-epidemic-and-health-it/) and to optimize the allocation of healthcare resources.

Transmitting of disease is the spread of the disease from one person to another, which we can equivalently view as ways a disease multiplies. Epidemics can be observed through this reproduction of disease through direct contact, environmental factors, or through the air for respiratory infections.

Modeling transmit disease refers to the use of mathematical models to study and predict the spread of a disease within a population. These models can take into account various factors that can influence the transmission of a disease, such as the rate at which the disease is transmitted from person to person, the proportion of the population that is susceptible to the disease, and the effectiveness of interventions such as vaccination or quarantine measures. Models are usually simplified, understood as direct contact. A person must directly interacts with another in the model for the calculation of the rate of transmission can kick in. But we can argue that these constant chances of transmission don't exist in the real world, where interaction with higher immunity is not consistent with interaction with people with low immunity. Similar to modeling in different fields like economics or physics, healthcare modeling can run into a risk of oversimplification. Though the discussed factor can be canceled out due in the real world, infection rates emerge from other means as well such as contaminated objects or through airs, more often than not the real-world rates are unpredictable. Therefore models tend to be limited to few variables and factors. They are restricted to the basic hypothesis and models, and inform only the immediate response after a certain policy.

Nevertheless, modeling transmit disease can inform public health policy and decision-making by providing estimates of the potential impact of different interventions on the spread of a disease. It can also help to identify high-risk groups or areas and inform the allocation of resources to control the spread of disease. Models complexity can rangefrom mathematical models calculating the spread and infectiousness of three human pathogen

(Mathematical models of infectious disease transmission - Nicholas C.Grassly and Christophe Fraser) but can be pretty simple as a simulation, such as ones from *Washington Post* by Harry Steven (https://www.washingtonpost.com/graphics/2020/world/corona-simulator/). Fascinatingly, they can help us answer whether specific policies are more resource efficient to implement than others. For example, we learn by looking at the simulation from Washington Post, Harry successfully answered quarantined people with infections by early testing and diagnose (by seperate the red dots indicating infected individuals into separate "box") can be equally effective as quarantining the whole population (by getting the dots to stay in a position instead of moving around). We learn from the experiment two key points: the policy to quarantine infected individual can become more cost effective option, but the second keypoint is that implementing both policies will help drastically bring the disease to extinction.

SIR models stands for Susceptible, Infected and Recovered person from a disease. **SEIR** add another E for exposed, which means people get the disease, but incubate the disease without showing any symptoms. SIR is the model that demonstrates above. Then model getting a bit more complicated by adding the factor of "exposed" turning SIR to SEIR, which is especially fitting with diseases with no symptoms during incubation period such as COVID-19 or SARS. This time, we investigating whether we can containe the disease in time, with different rate of infection and different rate of incubation, which demonstrated in *Outbreak* by Kevin Simler (https://meltingasphalt.com/interactive/outbreak/). From the experiment, we learn other keypoints: early detection and diagnoses, especially more effective detection tools help drastically mitigate the spread of the disease and effective disease containment.

But even with SEIR, that's not how people would behave. People usually commute in real life to a specific location like a school or a workplace, not drifting around randomly. They are likely to abide by the policies and certainly not touching each other when they commute. What about other policies in action, such as social distancing, or different communities interaction, how we corporate that to our model? Each of these policies and social practice is another

parameters we have to consider, adds another "flavor" to our models collectively. Dr. Grand Anderson has done an excellent job incorporating these policies, creating each model simulating each, or a combination, of the policies in action (*Simulating a pandemic* - Grand Anderson, https://www.youtube.com/watch?v=gxAaO2rsdIs). But even with his new add-ons flavors, we can't just create one simulation models for one policy, evaluate it, then move on to the next. Such action not only waste time, but the environment it create also low in validity, since reality also contains unpredictability. Therefore, we need a better, more robust method of evaluating data, can apply real-world factors but also incorporates multiple variables. If we succeed, our model can predict future disease rates outside of just reflecting current disease trends. One way we can improve our model is by implementing machine learning (deep learning) in predicting and estimating transmitted disease productive rate. Using machine learning to investigate a disease is nothing new, but surprisingly, using machine learning to optimize control measures for infectious diseases is only researched recently. This can involve developing algorithms to identify the most effective interventions for reducing the spread of an infectious disease, such as vaccination campaigns or quarantine measures.

We have experimented with using artificial neural network such as Long short-term memory (LSTM) using Keras Python library and Recurrent neural network (RNN) using Darts to predict areas with higher rate of COVID. As a demonstration on how we can implement this prediction to certain areas, we will try to predict the total number of confirmed case in different countries. Alongside with visualization alone, for more complicated method using RNN, we will use the following measure to evaluate the model's performance: The Mean Absolute Percentage Error (MAPE) which calculated the sum of the individual absolute errors divided by each period separately, the root mean square percentage error (RMSPE), the root mean square error (RMSE) and finally, the R-squared. Surprisingly, the simple methods using Keras produce more tangible result than more complicated RNN model, as the result is closer to the true values. The result of our models are horrible, but RNN models with the reduced loss value

demonstrated there is potential for improvement judging our model is feasible. We believe the well-resarched methodology is possible, and can be used to propose robust models. Here we propose using multiple parameters such as the policies/ intervention from healthcare systems and governments, using rates of people taking care of themselves/ following the policies; how a disease spreads within a population, and how it is affected by various factors such as vaccination rates, population density, and public health interventions.

We train our data from the beginning of 2020 up until the beginning of May 2021. We try to predict the US's COVID total confirmed case. Here's our prediction from the overall 3 methods with RNN model for the US's case: LSTM, RNN vanilla method, and Gated recurrent unit (GRU):
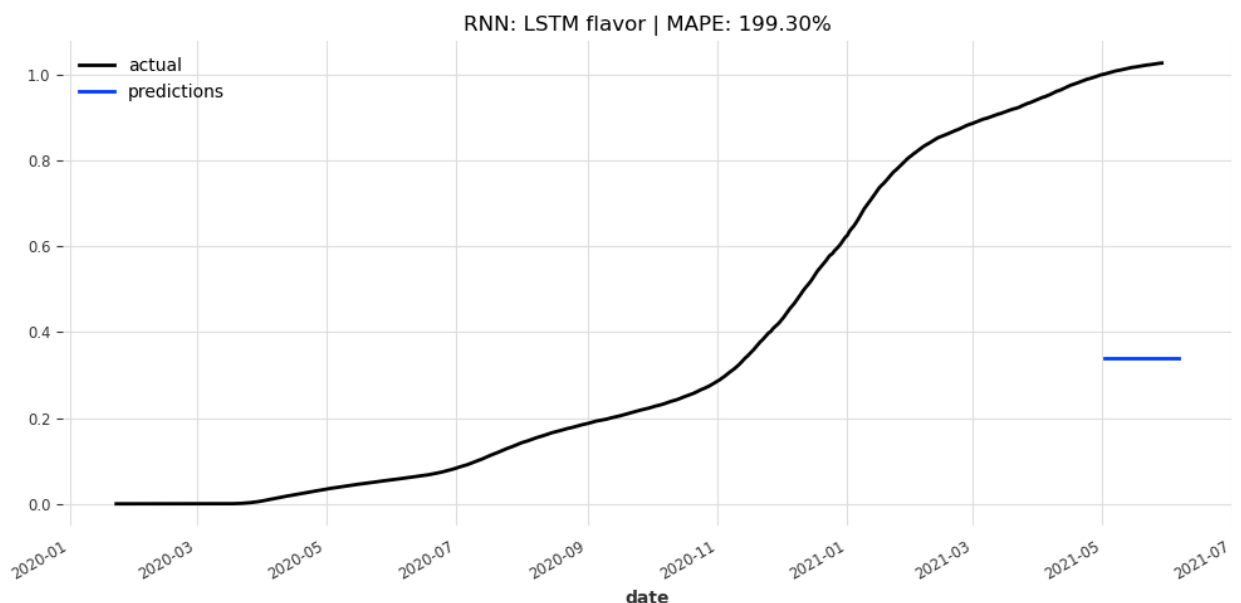

The graph depicted LSTM method prediction. All three methods produced the same predictions, gave horrible values. But we learn through observing the loss calculated during training:

```
beginning the training of the LSTM RNN:
Epoch 99: 100%|██████████| 29/29 [00:00<00:00, 46.62it/s, loss=0.115, v_num=logs, train_loss=0.0933,
val_loss=0.466]
training of the LSTM RNN completed: 65.71 sec
Predicting DataLoader 0: 100%|██████████| 1/1 [00:00<00:00, 55.56it/s]
LSTM :
MAPE  :  199.3029
RMSPE :  0.6659
RMSE  :  0.6763
-R squared  :  15366746993394475597824.0000
se  :  0.0871
```

That our model can be improved! Training loss is constantly reduced through each iteration of our model (epoch we set as 100) meaning our model can be improved. We will look into further calibration of the parameters by increasing our epochs from 100 to 300, then maybe our model can perform better! (*For further detail about models performance, we provide the performance results of the three models at the appendix below!*)

We also make the same predictions using LSTM but with a much simpler method from keras. This time we can view how training loss and validation loss decrease over each epoch, and how our model predict visually.
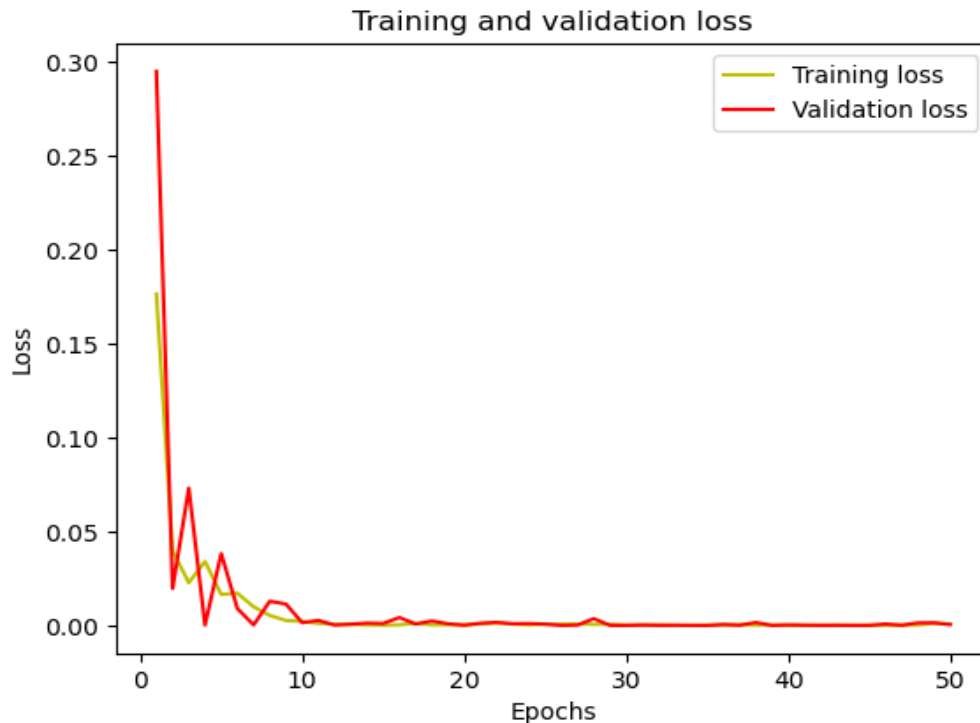
The training loss and validation loss decreased over each iteration, meaning that our model actually learning and adapting to our data. Let's look at the model performance when predicting US cases.
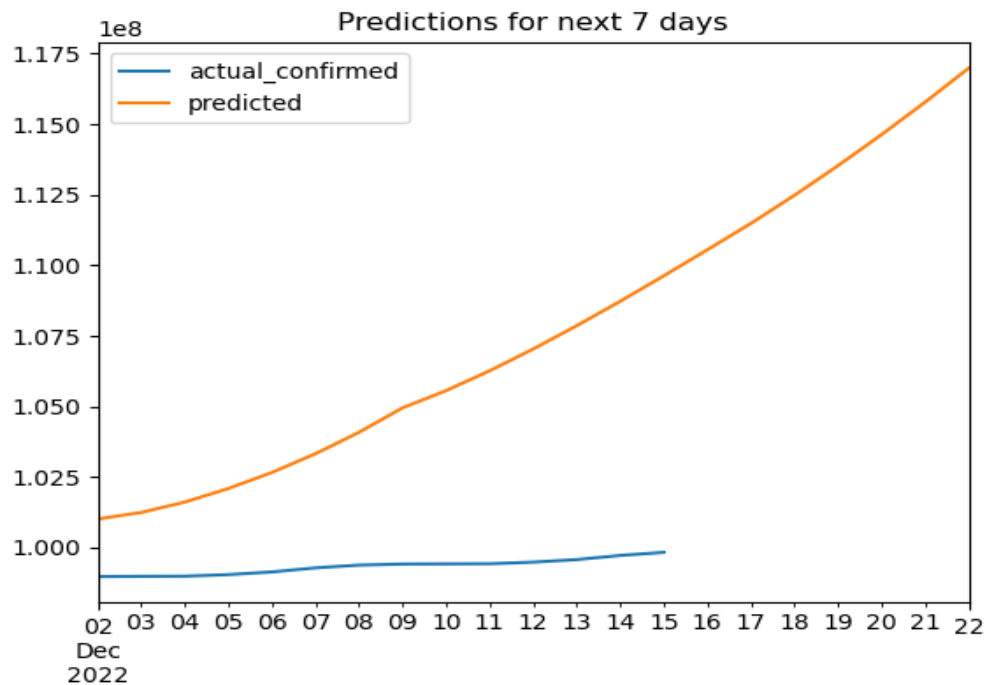
The prediction model is far from accurate to the true value of the US. We have re-train our model multiple times, and this is the closest we got. We believe there are methods more robust than ours, and there are factors that we have overlooked and didn't include in our model. Hopefully our future projects and researches, we can incorporate these factors into our models and create a more consistent prediction model.

Conclusion: In *Mathematical models of Infectious Disease Tranmission* by Nicholas Grassly and Christophe Fraser, the dynamic between the infectious disease and the host is much more complicated than a model can demonstrated. Until there's a robust and consistent method of implementing and quantifying the factors that impact how a disease can progress, healthcare mathematical models are the best method of simplifying and summarizing the real world. It's the challenge of an infectious disease epimiologist to disentangle the behaviour of both the disease and the host, while preserving the rules for these behaviours to be predicted and identifiable.

Reference:

https://meltingasphalt.com/interactive/outbreak/ by Simler, K. (2020, March 16). *Outbreak*. Melting Asphalt. Retrieved December 5, 2022, from https://meltingasphalt.com/interactive/outbreak/

Ayer, T., Chhatwal, J., & Linas, B. P. (2020). *COVID-19 Simulator*. www.covid19sim.org. Retrieved December 6, 2022, from https://www.covid19sim.org/

Kruijshaar, M. E., Barendregt, J. J., & Hoeymans, N. (2002). *The use of models in the estimation of disease epidemiology* [Doctoral Dissertation]. https://www.scielosp.org/pdf/bwho/v80n8/v80n8a04.pdf

Grassly, N., Fraser, C. Mathematical models of infectious disease transmission. *Nat Rev Microbiol* 6, 477–487 (2008). https://doi.org/10.1038/nrmicro1845

Woolhouse, M. (2011). How to make predictions about future infectious disease risks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1573), 2045-2054. https://doi.org/10.1098/rstb.2010.0387
or
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3130384/

Vanagas, G., Krilavičius, T., & Man, K. L. (2018). Mathematical Modeling and Models for Optimal Decision-Making in Health Care. *Computational and Mathematical Methods in Medicine*, *2019*. https://doi.org/10.1155/2019/2945021

Bjørnstad, O.N., Shea, K., Krzywinski, M. et al. The SEIRS model for infectious disease dynamics. Nat Methods 17, 557–558 (2020). https://doi.org/10.1038/s41592-020-0856-2

Opioid Epidemic & Health IT - Health IT Playbook. https://www.healthit.gov/playbook/opioid-epidemic-and-health-it/

Appendix:

The result of three RNN models training using darts:

beginning the training of the LSTM RNN:
Epoch 99: 100%|███████████| 29/29 [00:00<00:00, 46.62it/s, loss=0.115, v_num=logs, train_loss=0.0933, val_loss=0.466]
training of the LSTM RNN completed: 65.71 sec
Predicting DataLoader 0: 100%|███████████| 1/1 [00:00<00:00, 55.56it/s]
LSTM :
MAPE : 199.3029
RMSPE : 0.6659
RMSE : 0.6763
-R squared : 15366746993394475597824.0000
se : 0.0871

beginning the training of the GRU RNN:
Epoch 99: 100%|███████████| 29/29 [00:00<00:00, 48.48it/s, loss=0.11, v_num=logs, train_loss=0.0911, val_loss=0.474]
training of the GRU RNN completed: 62.80 sec
Predicting DataLoader 0: 100%|███████████| 1/1 [00:00<00:00, 45.46it/s]
GRU :
MAPE : 209.6602
RMSPE : 0.6771
RMSE : 0.6877
-R squared : 467975998695823638528.0000
se : 0.0871

beginning the training of the Vanilla RNN:
Epoch 99: 100%|███████████| 29/29 [00:00<00:00, 61.31it/s, loss=0.114, v_num=logs, train_loss=0.109, val_loss=0.461]
training of the Vanilla RNN completed: 52.79 sec
Predicting DataLoader 0: 100%|███████████| 1/1 [00:00<00:00, 83.32it/s]
Vanilla :
MAPE : 200.5710
RMSPE : 0.6673
RMSE : 0.6778
-R squared : 1159422527176881015593923583 67232.0000
se : 0.0871

The three plot accordingly:

RNN: LSTM flavor | MAPE: 199.30%

RNN: GRU flavor | MAPE: 209.66%

RNN: Vanilla flavor | MAPE: 200.57%