Computer Fraud and Security
ISSN (online): 1873-7056

PanDict to Enhance Pandemic Preparedness：A Multi-ModelFramework for Transmission Dynamics, Variant Evolution, and Hospitalization Resource Planning

Bohan Liu1

[1] Rice University, Houston, TX, 77005, USA

ABSTRACT

Global epidemics, like COVID-19, have substantial impacts on almost all countries in multiple aspects, such as economy, hospitalization, lifestyle, etc[1,2] . COVID-19 can spread to populations worldwide due, in part, to their high contagiousness, but more importantly, because of our inability to quickly address some of the most fundamental problems of a newly emerged virus: 1) How quickly will the virus spread? 2) Whether and under what conditions will new variants emerge? 3) How to arrange our resources accordingly? Since previous epidemic models were incapable of addressing these three most important questions, we developed the PanDict system, which can help address all three of the most essential problems discussed above. To elaborate, our model consists of three crucial parts, each tackling one of the three above-mentioned problems: 1) predicting the spread of the new virus in each local community and calculating its $R0$ value using our newly devised EPSEIRV model; 2) creating and using the $SI3R$ model to simulate variant competition; 3) forecasting hospitalization deficiencies in each state and producing visual representations of the projected demand using our IHOV model. In contrast to other vague and incorrect predictions/models, our EPSEIRV model accurately predicted the spread of the Omicron variant of Sars-CoV-2 in the United States and South Africa prior to their peaks. Moreover, in January 2022, we concluded that the $R0$ value of Omicron is around 18.8. The high infection speeds of these viruses allow them to widely circulate in the population before vaccines are fully developed. Thus, there will be inevitable surges in the number of patients, which can potentially overwhelm unprepared hospitals, hence making the IHOV model especially imperative. In a nutshell, when a novel disease emerges, the PanDict model can quickly and accurately predict how fast the disease spreads, whether the disease will successfully mutate, and how to arrange hospitalization resources to most efficiently mitigate suffering. These crucial functions can apprise our users of where the potential epidemic is heading and how to diminish its impact. Furthermore, the PanDict model will allow hospitalization systems to be much more prepared for upcoming surges of patients, which would significantly reduce excess deaths and hospitalization deficiencies. The system also supports related departments or corporation plans with the EPSEIRV model and $SI3R$ model during the contemporary epidemic.

Key words: Disease control, Infection prediction, Hospitalization resource arrangementIntroduction

Pandemics across the globe have long-lasting, profound impacts on all countries and on each individual. For instance, COVID-19 has crushed hospitalization systems worldwide and caused more than 6.3 million cumulative deaths while H1N1 directly resulted in about 575,400 deaths globally during its first year of circulation[2,3]. However, death and suffering are only the superficial consequences of a circulating epidemic as its harmful effects permeate almost all aspects of people's lives. The International Monetary Fund (IMF) estimates that from 2019 to 2020, the global median GDP fell by 3.9%, the worst recession since the Great Depression, demonstrating an imperative need for designing a system that mitigates the harm of

epidemics.[1]. The unprecedented impact of COVID-19 can, in part, be attributed to the inability of previous studies to obtain essential information about a newly emerged virus. Without the support of those basic information, policymakers failed to establish practical measures against the virus, and individuals were misinformed or simply not informed, causing severe public concerns. The problematic areas are as follows.

- Transmission forecast. Previous work has struggled to produce reliable transmission predictions because of some inherent flaws of the SEIRV model[4]. Even some authoritative organizations, including CDC, still cannot obtain the accurate value of the transmission coefficient ($R0$) of Omi- cron[5], let alone predict the growth of the epidemic. In addition, Dr. Fauci, one of the most trusted epidemiologists, has unexpectedly inaccurate and uninformative predictions in many well-known media, such as CNN[6], etc.

- Variant Prediction. Little to no work can be found about future variant prediction when confronted with a novel disease. For example, when the first wave of Omicron was close to an end, experts still had no clue what future variants will look like and under what conditions they can emerge[7]. Moreover, Dr. David Aronoff, an infectious disease expert and chair of the department of medicine at Indiana University School of Medicine, even said, "We really don't know if there's going to be another variant that may create a lot of havoc."[8]

- Hospitalization resource estimation. Highly contagious pandemics invariably threaten hospital- ization systems, as the number of patients can grow exponentially in merely a few days. Omicron, for instance, caused substantial staffing and resource shortage in one-fourth of all US hospitals, resulting in delays in elective surgeries[9]. While most hospitals in the US had excess resources and staff, due to the rapid, exponential increase in the demand for hospitalization in certain states, officials had no other choice than to deploy National Guard to "provide direct support to hospitals, care centers, and other medical facilities;" some even forced infected employees to keep working, putting patients at risk[10].

Realizing the urgent need to minimize the damage of global epidemics, we devised PanDict (Pandemic PreDiction, a computer system that uses the EPSEIRV model, the *SI*3*R* model,and the IHOV model to produce accurate projections of future epidemic situations and to address the three pressing issues above with each model tackling one of those problems. The system provides its users with a reliable transmission forecast, an informative variant prediction, and an projection of the demand for hospitalization resources. PanDict allows its users to input some most basic parameters of a novel disease and immediately receive the essential information needed to minimize its damage. Some detailed explanations of each part are as follows.

- First, to develop a reliable transmission forecast, we formulated the EPSEIRV model. The original SEIRV model contains a variable, α, that can't be determined experimentally, for it contains no physical meaning. Hence, it has a preset value(average of past diseases), which skews information from the data and exacerbates the accuracy of the model. Thus, we replaced those variables with quantifiable variables that introduce population density and time of exposure into the system, allowing the system to conduct simulations in each local community and obtain accurate results. We tested it with the real life data in South Africa and the United States, and our prediction of the infected population fits almost perfectly with the actual infected curve. We yielded an $R0$ value around 18.9 for Omicron, substantially higher than the conclusion of previous studies, "above 7.0"5. We published our findings on medium.com, around mid-January, attached.

- Secondly, to answer, "will new variants emerge and how?", we created the *SI*3*R* model, which simulates the competition between a mutant and the resident. After implementing it in the case of Omicron, we concluded that for a new variant to emerge, it has to overcome the public's immu- nity against Omicron. Otherwise, it would die out almost immediately. The theoretical foundation regarding super-infection and double-infection of the *SI*3*R* model references an article from *The American Naturalist*11. The

article stopped at discussing variant evolution on the conceptual level while the *SI3R* model allows for simulating competitions based on real life data.

- Finally, we devised the IHOV model to project the demand for hospitalization, extending an uncompleted Brown project[12]. The project initiated by Brown School of Public Health provides current hospitalization data, including number of available beds, occupation rate, etc. However, that project simply listed some possible cases based on guessed situations, so it serves no forewarning purposes to the public. In contrast, EPSEIRV models the infected population of each state as a function of time produced and inputs those projections to the IHOV model, allowing us to predict and show exactly where, when, and how much, for example, beds are needed.

As far as our knowledge is concerned, we were the first to construct an accurate simulation of the spread of Omicron. In a nutshell, our system can

1. Inform individuals, corporations, and the government of future trends of a newly emerged virus.

2. Reduce public concerns regarding the depressive future of pandemics, prepare health sectors for the possible emergence of new variants.

3. Help hospitals optimally arrange resources and minimize excess suffering and deaths.

We have only completed the IHOV model for the New England region, but future work may be easily implemented to expand the system to the entire US or even around the globe.

Previous Studies

Transmission Forecast

Almost all mathematics-based disease transmission models are variants of the SIR model. This section explains the SIR model and presents some popular variations of it[4].

The SIR model separates the population into three compartments, Susceptible, Infectious, and Removed. Everyone who has neither obtained immunity against the disease nor been infected with the disease is considered Susceptible, everyone who is infected and can infect others is regarded to be Infectious, and everyone who is immune to the disease is defined in the Removed category.

The SEIRV model implemented two additional compartments to the SIR model, Exposed and Vac- cinated. The original Infectious group in the SIR model is elaborated into two, taking patients' latent period into account; the Exposed group contains all who have been infected but are not yet infectious

Variant Prediction

Most researchers use mathematical or biological approaches to predict the emergence of future variants of a disease. This section explores the deficiencies of each approach and how our system resolves this problem. The biological approach to predicting future variants requires an extensive amount of experimentation and often takes a large amount of time before attaining adequate information. In the case of Omicron, the infectious disease expert, Dr. David Aronoff, who attempted to use biological methods to predict future variant emergence said after the surge of Omicron that they still "really don't know if there's going to be another variant that may create a lot of havoc."[8] This more mathematics-based article by Minus Van Baalen and Maurice Sabelis explored the competition between the resident strand and a mutant of a virus[11]. The study also explained the impact of that competition on the evolution of the virulence, as well as the infectiousness, of a disease. However, the study stopped on the theoretical level and did not provide a usable model for predicting variants. Thus, we devised the *SI3R* model to simulate the competition between the resident strand and a hypothetical mutant of a virus to explore the conditions under which new variants could emerge.

Demand For Hospitalization

Before our project, Brown school of Public Health was the only group that attempted to calculate the projected demand for hospitalization resources12. However, their study did not implement a disease transmission model that could estimate the spread of the disease; instead, they guessed the infected population and time then calculated future demand according to the guesses. Thus, it has no specificity to the disease and will not be instructive to policymakers who need predictions of exactly where and when hospitals will hit their capacity to prepare for the exponential growth of the number of patients.

Method

System Framework

The system aims to help minimize the damage of a newly emerged epidemic. It consists of three essential modules, each addressing one of previous studies' problems.

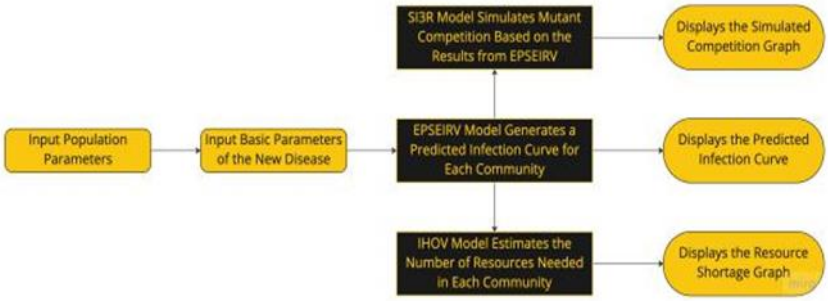

Figure 1. System Usage Flowchart

Shown in Figure 1, the process of the system is as follows. The user first inputs the population statistics of the target country and enters the date of the first case in each local community. The basic parameters of the newly emerged disease are also required. Those mainly include: incubation period, infectious period, and vaccine efficacy. Subsequently, the system produces and displays an accurate prediction of the spread of the new virus according to those parameters. With the projected infection curve, the second module will test the condition and probability of the emergence of a new variant using the *SI*3*R* model. Lastly, the system uses the projected infection curve to calculate the demand for health resources in each individual county of the target country and displays the projected data on a map(can download the actual numbers as well).

1. Transmission Forecast – the EPSEIRV Model

Review of SEIRV Model

As the augmented form of the SIR model, the SEIRV model divides the population into five key compartments: Susceptible, Exposed, Infectious, Removed, and Vaccinated. Below is the system of differential equations that describes the changes between each compartment:

$$\frac{dS}{dt} = -\beta I(\frac{S}{N})^{\alpha} - \psi\varepsilon S \tag{1}$$

$$\frac{dE}{dt} = \beta I(\frac{S}{N})^{\alpha} + \beta I(\frac{V}{N})^{\alpha} - \sigma E \tag{2}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{3}$$

$$\frac{dR}{dt} = \gamma I + \iota V \tag{4}$$

$$\frac{dV}{dt} = \psi\varepsilon S - \beta I(\frac{V}{N})^{\alpha} - \iota V \tag{5}$$

The SEIRV model is based on the following assumptions:

1. The total population remains constant(new births and deaths have trivial impacts and are not taken into account).

2. Measures against the disease remain unchanged.

3. People's living habits have no drastic changes.

4. The population is relative homogeneous.

5. An individual can't catch the disease twice.

The exponent α in Eqn. 1,2,5has no physical meaning and thus cannot be obtained in any way other than estimating with infection data. For most kinds of diseases, the value of α = 1.2 works the best, so it was set to be around 1.2 for novel diseases as well without any other supporting experiment or evidence. The model has only one degree of freedom (*beta*) with the exponent *alpha* predetermined; this lack of specificity significantly reduces its accuracy and robustness. In addition, this model does not account for changes in population density and social distancing. Thus, we made the following changes to the model.

Determining the Expression for Daily Infection

This section summarizes the derivation of the daily infection expression.

Assuming that the probability of one person meeting an infectious individual and capturing the disease is λ and that an average person spends μ fraction of their time interacting with others, then the probability of the person catching the disease at all equals to

$$\mu(1-(1-\lambda)I)$$

where I equals to the number of Infectious agents[13]. Because λ represents the probability of meeting one infectious individual and simultaneously receiving the disease, it should be almost infinitesimally small. Since $e = (1+\frac{1}{n})^n$ when n approaches infinity, we can approximate $e \sim (1-\lambda)^{-\frac{1}{\lambda}}$. Then, we may

rearrange the expression as follows:

$$\mu(1-(1-\lambda)I) = \mu(1-(1-\lambda)^{-\frac{1}{\lambda}\,-\lambda I}) = \mu(1-e^{-\lambda I})$$

This change to the expression connects daily infection number with population density and exposed time.

Exposed time is associated to the model by μ while this paragraph illustrates the model's connection to population density. Suppose r is the radius of a person's daily activity range, p is the probability of catching the disease upon encounter, N is the population of the target community, and A is the area. Then, we can express

λ as $p\frac{\pi r^2}{A}$. Substituting A with $\frac{N}{D}$, D representing the population density. Finally we yield $\lambda = p\frac{\pi r^2 D}{N}$, showing that the infection coefficient λ is directly proportional to the population density.

Equations of the EPSEIRV Model

Thus, the system of differential equations in the EPSEIRV model are as follows:

$$\frac{dS}{dt} = -\mu(1-e^{-p\frac{\pi r^2 D}{N}I})S - \psi\varepsilon S \tag{6}$$

$$\frac{dE}{dt} = \mu(1-e^{-p\frac{\pi r^2 D}{N}I})(S+V) - \sigma E \tag{7}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{8}$$

$$\frac{dR}{dt} = \gamma I + \iota V \tag{9}$$

$$\frac{dV}{dt} = \psi\varepsilon S - \mu(1-e^{-p\frac{\pi r^2 D}{N}I})V - \iota V \tag{10}$$

S represents the number of people who are susceptible to the disease, *E* represents the number of people who have caught the disease but are not yet infectious,*I* represents the number of people who are infectious, *R* represents the number of people who have gained immunity against the disease, *V* represents the number of people who are under vaccination but are not fully vaccinated, *t* represents time in days, ψ represents the the proportion of newly vaccinated individuals to the susceptible population, ε represents vaccine efficacy, σ represents the inverse of disease incubation period in days, γ represents the inverse of disease infectious period in days, ι represents the inverse of the time needed to gain immunity fro vaccines in days, *p* represents the probability of one person capturing the disease upon encounter with an infectious agent, *D* represents population density, *N* represents the number of people in population,*r* represents the radius of an average person's daily activity range, and finally, μ represents the fraction of time an average person spends interacting with others. μ and *p* are social parameters in this model while the rest are basic parameters.

The following chart elucidates the changing relationship between each compartment:

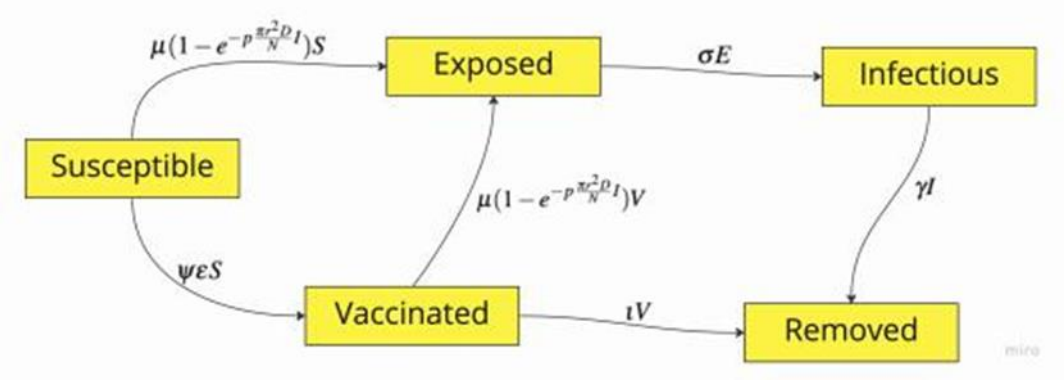

Figure 2. Change between each compartment of the EPSEIRV model summarized in a flowchart.

On day 1, E is set to 1, R is set to the number of people who have received a booster shot before day −14(14 days prior to the first case), V is set to the number of people who received a booster shot between day −14 and day 0,I is set to 0, and S is set to $N-E-V-R$.

SEIRV Equations EPSEIRV Equations

$$\frac{dS}{dt} = -\beta I(\frac{S}{N})^{\alpha} - \psi\varepsilon S \qquad \frac{dS}{dt} = -\mu(1-e^{-p\frac{\pi r^2 D}{N}I})S - \psi\varepsilon S$$
$$\frac{dE}{dt} = \beta I(\frac{S}{N})^{\alpha} + \beta I(\frac{V}{N})^{\alpha} - \sigma E \qquad \frac{dE}{dt} = \mu(1-e^{-p\frac{\pi r^2 D}{N}I})(S+V) - \sigma E$$
$$\frac{dI}{dt} = \sigma E - \gamma I \qquad \frac{dI}{dt} = \sigma E - \gamma I$$
$$\frac{dR}{dt} = \gamma I + \iota V \qquad \frac{dR}{dt} = \gamma I + \iota V$$
$$\frac{dV}{dt} = \psi\varepsilon S - \beta I(\frac{V}{N})^{\alpha} - \iota V \qquad \frac{dV}{dt} = \psi\varepsilon S - \mu(1-e^{-p\frac{\pi r^2 D}{N}I})V - \iota V$$

Figure 3. Comparison between the SEIRV model equations and the EPSEIRV model equations.

Determining Parameters

We will use the parameter determination process of the Omicron variant of Sars-CoV-2 as an example to demonstrate how each model parameter can be obtained.

Basic Parameters of Omicron

According to Centers of Disease Control and Prevention (CDC) investigations, the incubation period of Omicron is approximately 3 days, and the mean infectious period of Omicron is about 11 days[14]. CDC reports also reveal that the effectiveness of unboosted vaccines is insufficiently low, and British Broadcasting Corporation (BBC) claims that the efficiency of booster shots is roughly 80%[15,16]; therefore, we decided to onlyinclude booster shots as an effective type of vaccination. In the United States, the boosted population has an average daily increase of 600,000 (0.18% of the entire population) while in South Africa, no one has received a booster shot yet, so we decided to abandon the Vaccinated category when modeling the spread in South Africa[3].

Determining Social Parameters and Modeling

As shown by the statistics of United States Department of Labor, below is a list of major daily activities and time spent by an average American(in hours)[17]:

Major Activity	Average Time	Potential Fraction	Resulting Time
Personal Care Activities(Sleeping, Grooming…)	10.74	0	0
Household Activities	2.01	0	0
Purchasing Goods and Services	0.38	0.2	0.076

Caring for and Helping Household Members	0.43	0	0
Caring for and Helping Non-household Members	0.14	1	0.14
Working and Work-Related Activities	3.02	0.58	1.7516
Attending Class(Education)	0.18	0.54	0.0972
Homework and Research	0.19	0	0
Organizational, Civic, and Religious Activities	0.18	1	0.18
Socializing and Communicating	0.54	0.6	0.324
Watching Television	3.05	0	0
Participating in Sports, Exercise, and Recreation	0.37	1	0.37
Telephone Calls, Email, and Mail	0.22	0	0
Travel	0.79	0.05	0.0395

The activities marked red are the ones that potentially involve in person interactions with others. However, for each activity, only a portion of the times listed above will imply in person interactions. Supporting facts are as follows:

1. Roughly 80 percent of the population shops online[18].

2. 42 percent of the workforce works from home[17].

3. About 40 percent of socializing time is online[19].

4. Private cars dominate the American Commute. Only 5% of US commuters use public transportation[20].

5. 46 percent of students receive only online instruction[21].

Therefore, the number of hours that involve potential interactions with others can be calculated as: $0.38 \times 0.2 + 0.14 + 3.02 \times 0.58 + 0.18 \times 0.54 + 0.18 + 0.54 \times 0.6 + 0.37 + 0.79 \times 0.05 \approx 2.97$ hours, which means $\mu = \frac{2.97}{24}$. With all the parameters specified, we graphed this system of differential equations by compiling it into a Python

program. Then, we utilized the least rooted mean squared error(RMSE) to determine the best representative p value and the most accurate model for the spread of Omicron in the US.

The fraction of time people spend interacting with others in South Africa is not much different from that of the United States because it is a comparison between two groups with large populations and similar pandemic-control regulations.

2. Variant Prediction – the *SI*3*R* Model

Setting up the *SI*3*R* model

Upon the emergence of a novel virus, not only are we concerned about the spread of the resident strand of the virus, the public is also perturbed by the possibility of the appearance of new variants, particularly when we can't obtain sufficient information about when and under what conditions new variants may emerge. Since little to no work has done to investigate the competition between a mutant and a resident, we devised the *SI*3*R* model to simulate the competition and thus determine the conditions under which a new variant might emerge.

The model consists of five compartments, Susceptible, Infectious$_1$, Infectious$_2$, Infectious$a$, and Removed. The model's assumptions include those of the improved SEIRV model and that the mutated strand cannot evade the public's immunity against the resident variant. Therefore, the Susceptible compartment represents the portion of the population that has caught neither of the strands and is thus prone to receive the virus. Infectious$_1$ contains those who have been infected with the resident strand and are capable of spreading it. Individuals in the Infectious$_2$ compartment can spread the mutant while those in the Infectious$a$ compartment can transmit both strands. Since both strands, the resident and the mutant, according to the prediction, cannot escape the immunity against the other strand, individuals removed from any of the three infectious compartments would be defined in Removed. This system of equations is not extremely detailed but is sufficiently functional to generate general results. More future work could be implemented to enhance the accuracy, for example, adding the Exposed categories.

Generating Equations for the *SI*3*R* model

As established above, the expression for daily infection number of one variant can be modeled by $\mu(1-e^{-\alpha}I)S$. Thus, combining the infected number of the five possible transmission routes results in $\mu(5-e^{-a1}I1\ -e^{-\alpha1}Ia\ -e^{-a2}I2\ -e^{-\alpha2}Ia\ -e^{-\alpha a}Ia)S$. Then, dividing up the infected population into one of the three Infectious groups yields the correct model.

Equations of the *SI*3*R* model

Below is the system of differential equations that describe the change of each compartment.

$$\frac{dS}{dt} = -\mu(5-e^{-a1}I1\ -e^{-\alpha1}Ia\ -e^{-a2}I2\ -e^{-\alpha2}Ia\ -e^{-\alpha a}Ia)S \tag{11}$$

$$\frac{dI_1}{dt} = \mu(2-e^{-a1}I1\ -e^{-\alpha1}Ia)S-\mu(1-e^{-a2}I2\ )I_1\ -\theta1I_1 \tag{12}$$

$$\frac{dI_2}{dt} = \mu(2-e^{-a2}I2\ -e^{-\alpha2}Ia)S-\mu(1-e^{-a1}I1\ )I_2\ -\theta2I_2 \tag{13}$$

$$\frac{dI_a}{dt} = \mu(1-e^{-\alpha a}Ia)s+\mu(1-e^{-a2}I2\ )I_1\ +\mu(1-e^{-a1}I1\ )I_2\ -\theta aIa \tag{14}$$

$$\frac{dR}{dt} = \theta1I_1\ +\theta2I_2\ +\theta aIa \tag{15}$$

S represents the number of people who are susceptible to both the resident and the mutant, I_1 represents the number of people who are infected by the resident strand, I_2 represents the number of people who are infected by the mutated strand,Ia represents the number of people who are infected by both strands, R represents the number of people who have gained immunity, $\theta1$ represents the inverse of the infectious period of the resident strand, $\theta2$ represents the inverse of the infectious period of the mutated strand, θa represents the inverse of the infectious period of individuals carrying both strands, $\alpha1$ represents the resident strand's transmission rate from a person in

the Infectious$a$ compartment, α2 represents the mutated strand's transmission rate from a person in the Infectious$a$ compartment, α_a represents the probability of an individual who carries both strands transmits all two to a susceptible individual, a_1 represents the resident strand's transmission rate from a person in the Infectious$_1$ compartment. a_2 represents the mutated strand's transmission rate from a person in the Infectious$_2$ compartment, and finally, μ represents the fraction of time an average person spends interacting with others.

On day 1, I_1 is set to the number of patients in reality, I_2 is set to 1, R is set to the number of people who have

been infected and removed in reality, Ia is set to 0, and S is set to $Totalpopulation - I_1 - I_2 - Ia - R$.

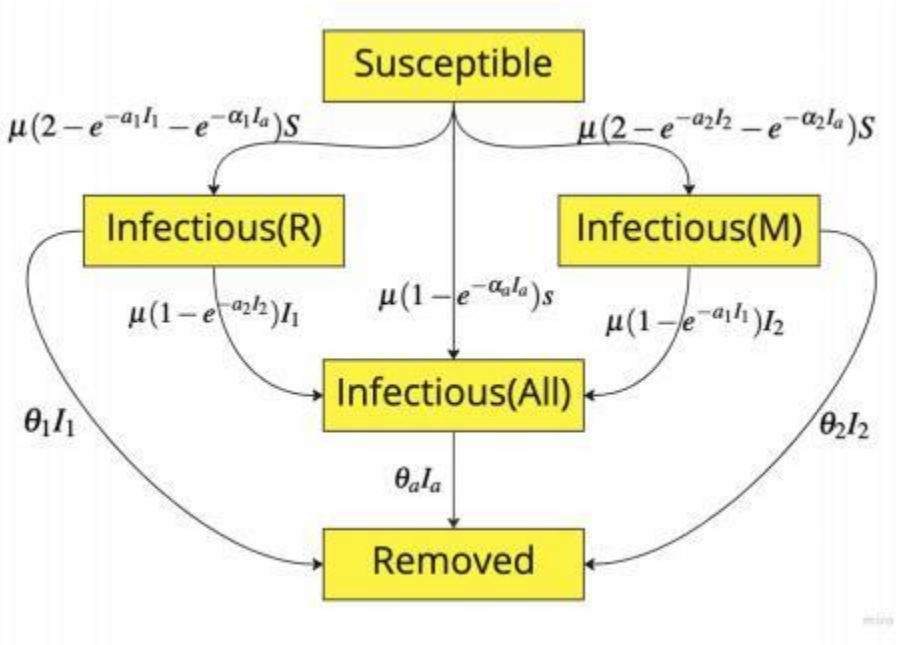

Figure 4. Change between each compartment of the $SI3R$ model summarized in a flowchart.

3. Estimating the Demand for Hospitalization – the IHOV Model

Using the EPSEIRV model, we can thus simulate the growth of the new variant in each local community as a function

of time. Then, obtain the number of hospitalization resources needed (n) in each hospital using Equation16:

$$n = I \times h - (O \times V + T) \tag{16}$$

in which I represents the infected population of that target community as a function of time yielded by the EPSEIRV model, h represents the hospitalization rate, O represents the number of already occupied beds, V represents the emergency vacating rate, and T represents the total number of beds available. Established in the EPSEIRV section as part of our improvement, the infectiousness and transmission curve of a virus closely relate to the population density and exposed period. Those factors tie closely to the individuality of each local community (e.g. Brooklyn, New York would have a much denser population distribution than Exeter, New Hampshire). Hence, the $I(t)$ function yielded by the EPSEIRV model allows us to accurately model the transmission data of each community. Due to our time limitation, we were only able to implement the IHOV model on the state level in the New England region. However, with simple data gathering and inputting, future work can be done to implement the IHOV model to each county and even globally.

We have implemented all New England states' parameters to the code, including population, population density,

number of beds available, and bed occupancy rate, etc. Assuming that 40 percent(can be adjusted) of all occupied beds will be emptied for the steep increase in the number of patients, the system then calculates the number of beds available to patients of the novel disease. Moreover, since exposed time is one of our parameters, our user can adjust the exposed time of communities to test the effect of quarantine policies or social distancing policies.

For our code, we used the Basemap package to produce a visual display of the resulting data. It allows us to generate different types of maps and borders. The Odeint package was used to simulate system of differential equations.

Results

The results section has three parts, each covering a part of the system. Since the system aims to generate useful guiding information when new diseases emerge, this section will discuss the results yielded from running the system on Omicron data in the United States.

Infection Forecast

The best-fitting p for the data obtained in the United States equals to 3.57×10^{-8}. Please note that we used positive test percentage instead of new daily cases because the percentage better represents the status of the entire population, for only a small portion of the population is tested each day. In addition, we subtracted the percentage of Delta variant positive cases(used the model on Delta as well) from the total positive percentage. Also, we assumed day 0 to be November 22, 2021 since the first detected case in the US is a traveller who returned from South Africa on November 2222.

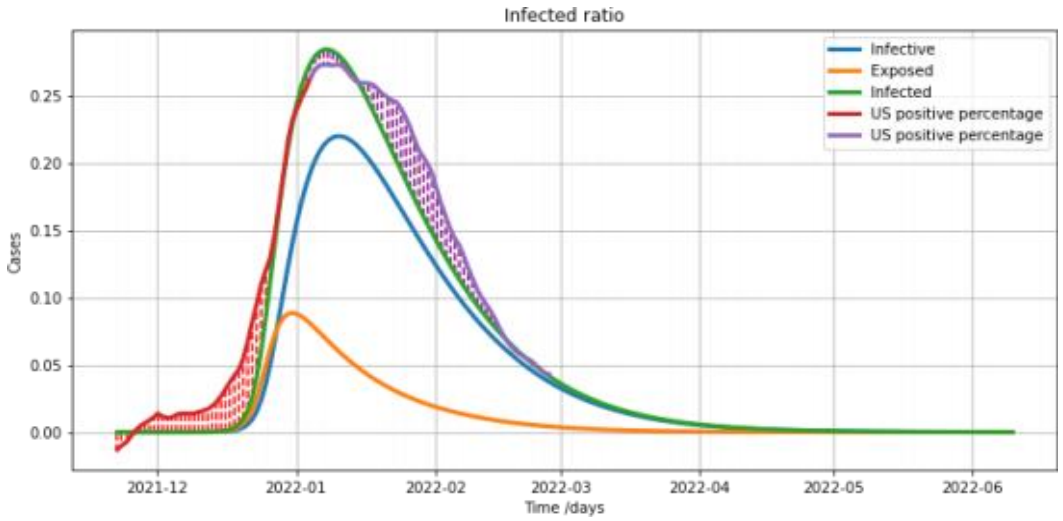

Figure 5. Projected Positive Percentage as a Function of Time in the US Produced by the EPSEIRV Model

As shown in Figure 5, in which the vertical and horizontal axes respectively represent infected population percentage and time, the red line is the data we used to generate our prediction(green line) while the purple line represents the real life data of the ensuing days. Our prediction was generated on January 6 and was based on the data prior to that day. Our model accurately predicted the peak of infected percentage and produced a projected infection curve that is almost identical to that in real life. The dotted lines mark the error of our predicted trajectory, which, shown in the picture, is negligibly small.

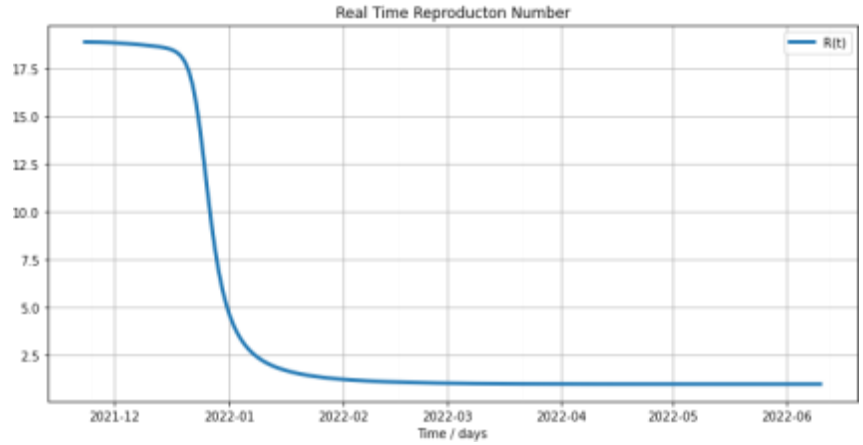

Figure 6. Real Time Reproduction Number as a Function of Time in the US Produced by the EPSEIRV Model

Figure 6is the yielded realtime reproduction number of Omicron in the US as a function a time by the EPSEIRV model. This concludes that the $R0$ value of Omicron in the US is approximately 18.8, and its realtime reproduction number ($R(t)$) decreases as a larger portion of the population becomes infected or removed. This is a reliable $R0$ value because the infection curve of this $R0$ value perfectly models the actual infection data. (Note: $R0$ is a concept in epidemiology that estimates an infectious agent's propensity for epidemic transmission. Simply explained, the $R0$ value of a disease means how many secondary infections were caused by the very first case in a fully susceptible population while $R(t)$ is defined as the average secondary infections caused by one patient relative to time[23].) An $R0$ value of 18.8 is extremely high for a disease, which is consistent with its absurdly quick circulation speed in the US. Before we have published our data on Medium.com, CDC announced that the $R0$ value of Omicron was about 7[5]. Now, with more projects examining the $R0$ number of Omicron in the US, an $R0$ number above 13 for Omicron had become somewhat a consensus, reaffirming our produced result.

In contrast, Figure 7 shows the performance of the original SEIRV model when α, according to its published essay, is set to 1.2. All other parameters were set the same way as the EPSEIRV model. While also using the same amount of data, the SEIRV model performs significantly worse than our proposed EPSEIRV model (the shaded region also represents its prediction's error). Apparently, Omicron's absurdly high contagiousness is not well considered in the 1.2 value of alpha in the original model.

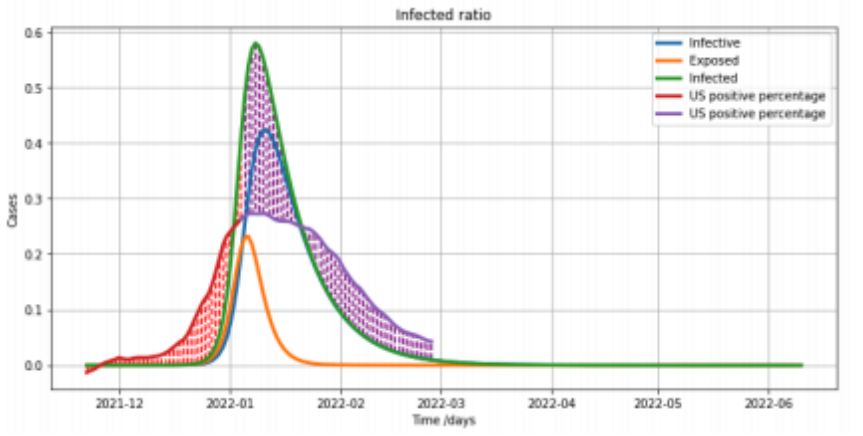

Figure 7. Projected Positive Percentage as a Function of Time in the US Produced by the SEIRV Model

Figure 8represents the change of each of the five compartments in the EPSEIRV model except for the vaccinated

group. Shown in the graph, almost the entirety of the population will obtain immunity against Omicron by the end of March, either from vaccines or from infection. This prediction was made at the beginning of January 2022 and has now been confirmed by other studies.

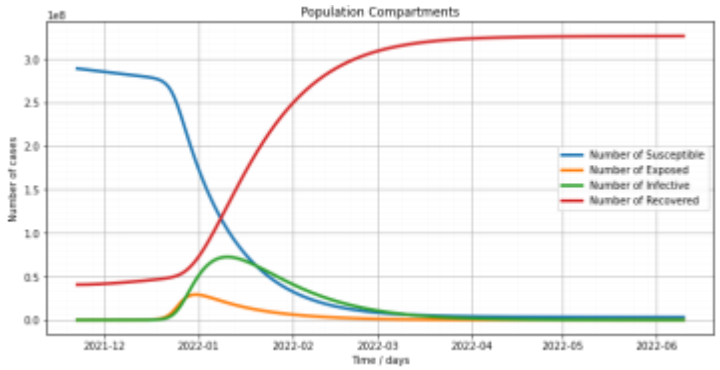

Figure 8. Change of Each Compartment Relative to Time in the US Produced by the EPSEIRV Model

Figure 9models the increase of accumulative positive Omicron cases in the United States, which more directly shows that most of the population will catch Omicron before the end of March 2022.

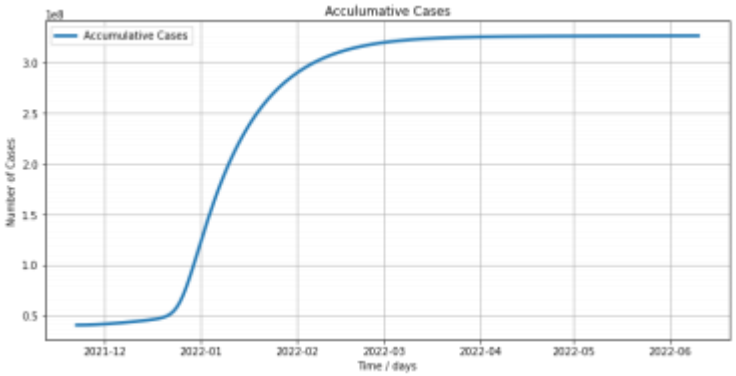

Figure 9. Increase of Accumulative Cases Relative to Time in the US

Variant Prediction

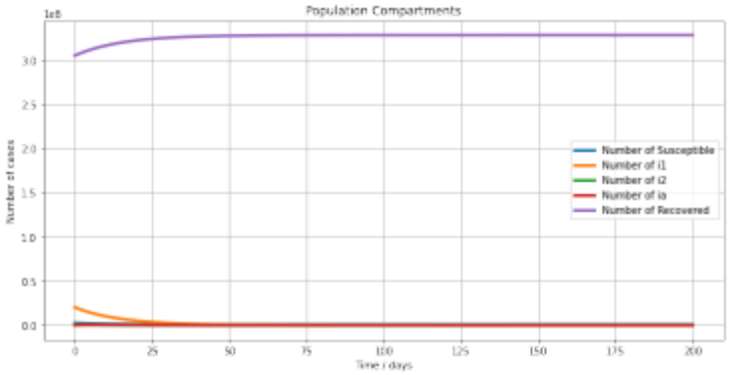

Figure 10. Simulation of the Competition Between Omicron and a Hypothetical Mutant

Since the existence of a mutant is hypothetical, we can't determine the specific parameters of it. However, the model still produced illustrative results. Assuming that on March 5, 2022, a new mutant of Sars-CoV-2 appears, thus setting day 1 to March 5, 2022, we obtained Figure 10. The purple line represents the Removed category, the orange line is Infectious$_1$, the minuscule green tipis Infectious$_2$, and the almost straight red line is Infectious$_a$. We tested with a range of input parameters, from absurdly high transmission rate to low transmission rate and from long infectious periods to short infectious periods. Nonetheless, it had negligible impact on the growth of the number of patients in each Infectious compartment for both strands died out very quickly in all scenarios due to the large number of removed individuals. Thus, under the assumptions of the model, it is not likely that a new mutant of Omicron can

prevail. In conclusion,assuming that no drastic changes happen to the US population, for a new variant to prevail,

it has to overcome the immunity against Omicron.

Estimation for Hospitalization Resources

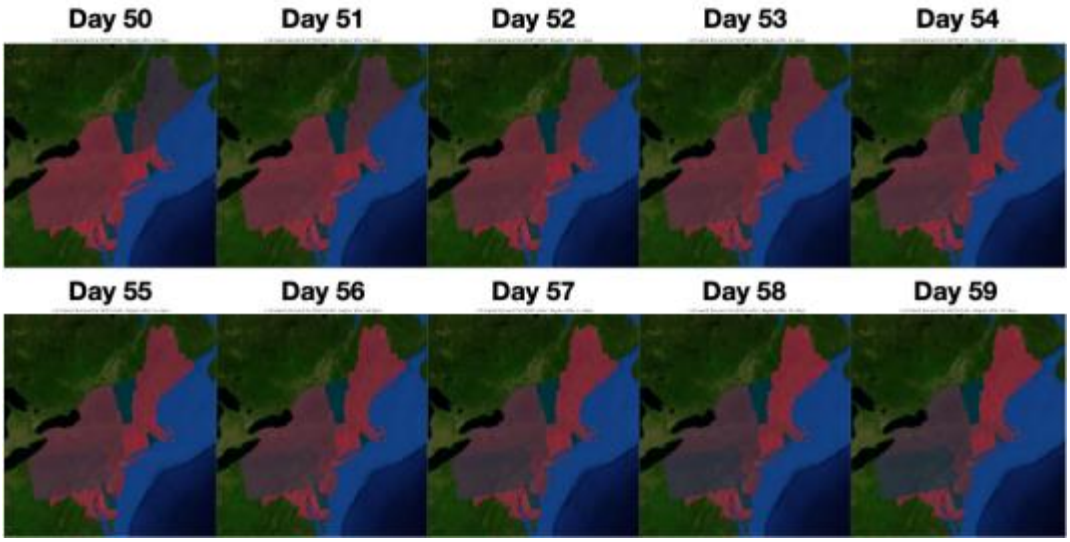

Figure 11. Visual Display of the IHOV Model Results

For demonstration and testing purposes, we ran the system based on Omicron data prior to January 6, 2022. Figure 11is the need for hospitalization resources in each state between 50-60 days since the first case. Users of the system can input different day numbers to access the estimation for that day with a slider. Again, we only finished the Northeast region of the United States for it is time consuming to enter the parameters mentioned in Methods section. Future work may expand the system to the US or even around the world.

Conclusion

In this study we focused on the current disease transmission department's inabilities that were exposed by the uncontrollable COVID-19 pandemic; we produced the EPSEIRV model, devised the *SI*3*R* model to expand our knowledge on mutant competition, programmed the IHOV model to project the demand for hospitalization resources, and created the overall PanDict system, which informs the public of infection prediction, variant emergence, and local resource shortages when a new virus emerges. Our modifications to the original SEIRV model drastically enhance its accuracy by eliminating the inaccurate exponent, α, and introduce population density, as well as exposure time, to the system of equations. In January 2022, the EPSEIRV model accurately predicted the growth of Omicron in the US and calculated the $R0$ value to be around 18.8. In May, the number Omicron infection cases, again, started to increase. This is due to people's fading

immunity against Omicron, which challenges the assumption of the EPSEIRV model. Further improvements could be made to the EPSEIRV model to address repeated infections. This

could be done by moving people from the Removed category back to the Susceptible compartment at a speed with which people lose their immunity. Secondly, the *SI*3*R* model enables for mutant competition simulation, which can be used to better understand and predict the emergence of new variants. Moreover, the IHOV model uses the produced predictions to estimate the need for hospitalization resources on a local scale. It will inform its users where, when, and how many, for example beds, are needed. With that estimation, proper arrangements of the limited health resources can substantially reduce the staffing and resource shortages hospitals experienced during the Omicron outbreak. Therefore, when new viruses appear, our system can lessen excess deaths, minimize significant hospitalization resource shortages, reduce public concern, and prune unnecessary economic impacts.

Code Availability

Code for the US model is linked here.

References

1. Stephanie, O. Economic impact of covid-19 on pepfar countries. https://www.kff.org/global-health-policy/issu e-brief/economic-impact-of-covid-19-on-pepfar-countries/#:~:text=The%20toll%20the%20COVID%2D19, downturn%20since%20the%20Great%20Depression. (2022).
2. Kristine, S. Infectious disease and economics: The case for considering multi-sectoral impacts. https: //www.ncbi.nlm.nih.gov/pmc/articles/PMC6330263/ (2019).
3. Hannah, R. *et al.* Coronavirus pandemic (covid-19). *Our World Data* (2020). https://ourworldindata.org/coron avirus.
4. Rajapaksha, R. M. N. U., Wijesinghe, M. S. D., Jayasooriya, K. A. S. P., Weerasinghe, W. M. P. C. & Gunawardana, B. M. I. An extended susceptible-exposed-infected-recovered (seir) model with vaccination for predicting the covid-19 pandemic in sri lanka. https://www.medrxiv.org/content/10.1101/2021.06.17. 21258837v2.full-text (2021).
5. NCIRD & DVD. Omicron variant: What you need to know. https://www.cdc.gov/coronavirus/2019-ncov/vari ants/omicron-variant.html?s_cid=11734:omnicron20variant:sem.ga:p:RG:GM:gen:PTN:FY22 (2022).
6. Kevin, B. Fauci predicts omicron covid wave will peak in u.s. by end of january. https://www.cnbc.com/2021/ 12/29/fauci-predicts-omicron-covid-wave-will-peak-in-us-by-end-of-january-.html (2022).
7. Laura, U. Expect more worrisome variants after omicron, scientists say. https://www.usnews.com/news/world/ articles/2022-01-15/expect-more-worrisome-variants-after-omicron-scientists-say (2022).
8. Deborah, B. Will the coronavirus pandemic ever end? https://www.voanews.com/a/will-the-coronavirus-pan demic-ever-end-/6393534.html (2022).
9. Bianca, C. Hospitals feel crush of omicron surge. https://www.nbcdfw.com/news/health/hospitals-feel-crush-o f-omicron-surge/2858350/ (2022).
10. Zach, S. National guard helps medical facilities with covid-19 peak. https://www.nationalguard.mil/News/Art icle/2900042/national-guard-helps-medical-facilities-with-covid-19-peak/ (2022).
11. Minus, B. The dynamics of multiple infection and the evolution of virulence. https://www.jstor.org/stable /2463102?seq=1 (1995).
12. Ashish, J. & Thomas, T. Us hospital capacity. https://globalepidemics.org/hospital-capacity/ (2021).
13. Theodore, K. & David, I. Law of mass action and saturation in sir model with application to coronavirus modelling. https://www.sciencedirect.com/science/article/pii/S2468042720300725?via%3Dihub/ (2020).

14. Lauren, J. *et al.* Investigation of a sars-cov-2 b.1.1.529 (omicron) variant cluster — nebraska, novem- ber–december 2021. http://dx.doi.org/10.15585/mmwr.mm705152e3externalicon (2021).
15. Michelle,R. Booster at least 80% effective against severe omicron. https://www.bbc.com/news/health-59696499 (2021).
16. NCIRD & DVD. Understanding how covid-19 vaccines work. https://www.cdc.gov/coronavirus/2019-ncov/va ccines/different-vaccines/how-they-work.html (2022).
17. BLS. American time use survey summary. https://www.bls.gov/news.release/atus.nr0.htm (2021).
18. Agatha, A. 41 online shopping statistics retailers should know in 2022. https://fitsmallbusiness.com/online-sho pping-statistics/ (2021).
19. Joseph, J. Effect of covid-19 on virtual socializing in the u.s. 2020, by gender. https://www.statista.com/statist ics/1108550/virtual-socializing-with-friends-and-family-during-coronavirus-usa-gender/ (2021).
20. Bureau,U. C. New census report shows public transportation commuters concentrated in large metro areas of the united states. https://www.census.gov/newsroom/press-releases/2021/public-transportation-commuters.html (2021).
21. Tea, I., Turkalj. 28 online education statistics and facts to learn in 2021. https://comfyliving.net/online-educati on-statistics/ (2021).
22. HHS. First confirmed case of omicron variant detected in the united states. https://www.cdc.gov/media/releas es/2021/s1201-omicron-variant.html (2021).
23. Breban, R., Vardavas, R. & Blower, S. Theory versus data: How to calculate r_0 ? https://journals.plos.org/plos one/article?id=10.1371/journal.pone.0000282 (2007).