Application of PanDict System Based on EPSEIRV and SI3R Models in Epidemic Forecasting and Healthcare Resource Planning

Bohan Liu

Phillips Exeter Academy
Exeter, USA 03833
E-mail: lbh833@126.com

Abstract—Global epidemics, such as COVID-19, have had a significant impact on almost all countries in terms of economy, hospitalization, lifestyle, and other aspects. Part of the reason is their high infectivity, but more importantly, due to the speed of virus transmission, the probability of new varieties appearing, and the conditions under which they appear, we cannot predict, making it a major challenge for us to arrange resources reasonably when the virus appears. Due to the inability of previous epidemic models to solve these three most important problems, we have developed the PanDict system, which can help solve all three basic problems discussed above. For a detailed explanation, our model consists of three key components that address the above issues: predicting the spread of new viruses in each local community and using our newly designed EPSEIRV model to calculate its R0 value; Creating and using the SI3R model to simulate variant competition; Predict the insufficient hospitalization in each state and use our IHOV model to generate a visual representation of the predicted demand. Compared to other vague and incorrect predictions/models, our EPSEIRV model accurately predicted the transmission of the severe acute respiratory syndrome coronavirus type 2 Omicron variant in the United States and South Africa before reaching its peak. In addition, the high infection rate of viruses allows them to spread widely among the population before vaccines are fully developed. As a result, the number of patients will inevitably surge, which will make hospitals overwhelmed, making the IHOV model particularly necessary. The PanDict model can quickly and accurately predict the speed of disease transmission, whether the disease will successfully mutate, and how to arrange hospitalization resources to most effectively alleviate pain. In addition, the PanDict model enables the hospitalization system to be more prepared for the upcoming surge in patients, which will greatly reduce excess deaths and insufficient hospitalization.

Keywords-Disease Control; Infection Prediction; Hospitalization Resource Arrangement

I. INTRODUCTION

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 575400 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 that 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.

For the Transmission forecast model, previous work has struggled to produce reliable transmission predictions because of some inherent flaws of the SEIRV model [4]. Even some authoritative organizations, including the CDC,

still cannot obtain the accurate value of the transmission coefficient (R0) of Omicron [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].

For the Variant Prediction model, 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 would look like and under what conditions they could 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].

For hospitalization resources, highly contagious pandemics invariably threaten hospitalization systems, as the number of patients can grow exponentially in merely a few days. Omicron, for instance, caused substantial staffing and resource shortages 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 the 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, this study devised PanDict (Pandemic PreDiction, a computer system that uses the EPSEIRV model, the SI3R 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 a projection of the demand for hospitalization resources. PanDict allows its users to input some of the 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.

Realizing the urgent need to minimize the damage of global epidemics, this study devised PanDict (Pandemic PreDiction, a computer system that uses the EPSEIRV model, the SI3R 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 a projection of the demand for hospitalization resources. PanDict allows its users to input some of the 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, the author 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 model's accuracy. Thus, the author 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. This study tested it with 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. This study yielded an R0 value around 18.9 for Omicron, substantially higher than the conclusion of previous studies, "above 7.05". The author published our findings on medium.com around mid-January, attached here. Secondly, to answer, "Will new variants emerge and how?" the author created the SI3R model, which simulates the competition between a mutant and a resident. After implementing it in the case of Omicron, the author concluded that for a new variant to emerge, it has to overcome the public's immunity against Omicron. Otherwise, it would die out almost immediately. The theoretical foundation regarding

super-infection and double-infection of the SI3R 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, the author 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 the 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 goes, the author was the first to develop an accurate simulation of the spread of Omicron. In a nutshell, our system serves two purposes. Firstly, it informs individuals, corporations, and the government about the future trends of newly emerged viruses. This helps reduce public concerns about the potential impact of pandemics and allows the health sector to better prepare for the emergence of new variants. Secondly, it assists hospitals in efficiently allocating resources, thereby minimizing unnecessary suffering and deaths. Currently, the author has only completed the IHOV model for the New England region. However, future work can easily expand the system's coverage to encompass the entire United States or even the entire globe.

II. PREVIOUS STUDIES

A. *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 vacant. 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. At the same time, the Infectious group includes those who were infected and have become infectious. The Vaccinated compartment comprises of people who have been vaccinated but have not yet gained immunity (those who have gained immunity are considered Removed). A more complete review of the SEIRV model is provided in the Method section.

In short, the SEIRV model has an exponent on the infection expression, α, which has no physical meaning, so it can't be determined with precision at all. It is pre-set to be around 1.2, a value that works best for most kinds of diseases. Thus, with α predetermined, the model only contains one degree of freedom: the contagiousness measure, β, which significantly undermines its accuracy. In contrast, the author improved the model by discarding the guessed variables and introducing two additional parameters derived from manipulating probability equations of encounters and the fraction of time people spend interacting with others. This improvement allowed us to accurately model the spread of the Omicron in January.

B. *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, the author 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.

C. Demand for Hospitalization

Before our project, the Brown School of Public Health was the only group that attempted to calculate the projected demand for hospitalization resources [12]. 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 and 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.

III. METHOD

A. 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 the previous studies' problems.
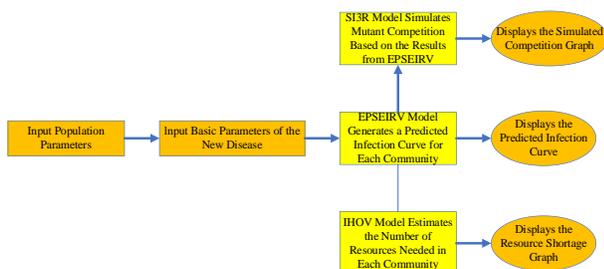
Figure 1.   System Usage Flowchart

As 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 the 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 SI3R 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 (you can download the actual numbers as well).

IV. TRANSMISSION FORECAST – THE EPSEIRV MODEL

A. Review of the SEIRV Model

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

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

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

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

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

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 relatively homogeneous.

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

The exponent α in (1), (2), (5) has 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, the author made the following changes to the model.

B. *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:

$$\mu(1-(1-\lambda)^I) \qquad (6)$$

where I equal 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, author can approximate $e = (1-\lambda)^{-\frac{1}{\lambda}}$. Then, author may rearrange the expression as follows:

$$\mu(1-(1-\lambda)^I) = \mu(1-(1-\lambda)^{\frac{1}{\lambda}^{-\lambda I}}) = \mu(1-e^{-\lambda I}) \qquad (7)$$

This change in the expression connects daily infection numbers with population density and exposed time. Exposed time is associated with 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, the author can express λ as $p\frac{\pi r^2}{A}$, Substituting A with $\frac{N}{D}$. D represents the population density. Finally, author yield $\lambda = p\frac{\pi r^2 D}{N}$. Showing that the infection coefficient λ is directly proportional to the population density.

C. *Equations of the EPSEIRV Model*

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

$$\frac{dS}{dt} = -\mu\left(1-e^{-p\frac{\pi r^2 D}{N}I}\right)S - \psi\varepsilon S \qquad (8)$$

$$\frac{dE}{dt} = \mu\left(1-e^{-p\frac{\pi r^2 D}{N}I}\right)(S+V) - \sigma E \qquad (9)$$

$$\frac{dI}{dt} = \sigma E - \gamma I \qquad (10)$$

$$\frac{dR}{dt} = \gamma I + \iota V \qquad (11)$$

$$\frac{dV}{dt} = \psi\varepsilon S - \mu\left(1-e^{-p\frac{\pi r^2 D}{N}I}\right)V - \iota V \qquad (12)$$

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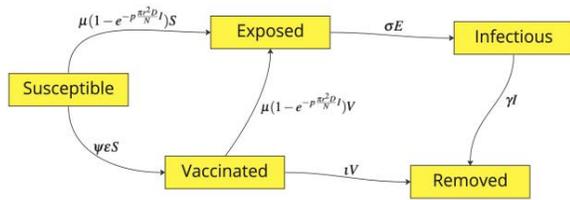
Figure 2.   Change between each compartment of the EPSEIRV model summarized in a flowchart.

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 $N - E - V - R$.

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

V.   DETERMINING PARAMETERS

The author 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.

A. Basic Parameters of Omicron

According to Centers for Disease Control and Prevention (CDC) investigations, the incubation period of Omicron is approximately three 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 the British Broadcasting Corporation (BBC) claims that the efficiency of booster shots is roughly 80% [15][16]; therefore, the author decided only to include 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 the author decided to abandon the Vaccinated category when modeling the spread in South Africa [3].

B. Determining Social Parameters and Modeling

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

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:

As shown in TABLE I. , Roughly 80 percent of the population shops online [18]. Forty-two percent of the workforce works from home [17]. About 40 percent of socializing time is online [19]. Private cars dominate the American Commute. Only 5% of US commuters use public transportation [20]. Forty-six 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 $\mu = \dfrac{2.97}{24}$ With all the parameters specified, author graphed this system of differential equations by compiling it into a Python program. Then, the author 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 amount of time people spends interacting with others in South Africa is not much different from that of the United States because it compares two groups with large populations and similar pandemic-control regulations.

TABLE I.   THE STATISTICS OF THE UNITED STATES DEPARTMENT OF LABOR

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

VI.   VARIANT PREDICTION – THE SI3R MODEL

A. Setting up the SI3R model

Upon the emergence of a novel virus, not only are authors concerned about the spread of the resident strand of the virus, but the public is also perturbed by the possibility of the appearance of new variants, particularly when the author can't obtain sufficient information about when and under what conditions new variants may emerge. Since little to no work has been done to investigate the competition between a mutant and a resident, the author devised the SI3R model to simulate the competition and thus determine the conditions under which a new variant might emerge.

The model consists of five compartments: Susceptible, Infectious1, Infectious2, Infectious, 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 contains those who have been infected with the resident strand and are capable of spreading it. Individuals in the Infectious2 compartment can spread the mutant, while those in the Infectious 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, by adding the Exposed categories.

B. Generating Equations for the SI3R model

As established above, the expression for daily infection number of one variant can be modeled by $\mu\left(1-e^{-\alpha I}\right)S$. Thus, combining the infected number of the five possible transmission routes results in $\mu\left(5-e^{-a_1 I_1}-e^{-\alpha_1 I_a}-e^{-a_2 I_2}-e^{-\alpha_2 I_a}-e^{-\alpha_a I_a}\right)S$. Then, dividing up the infected population into one of the three Infectious groups yields the correct model.

C. Equations of the SI3R model

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

$$\frac{dS}{dt}=-\mu\left(5-e^{-a_1 I_1}-e^{-\alpha_1 I_a}-e^{-a_2 I_2}-e^{-\alpha_2 I_a}-e^{-\alpha_a I_a}\right)S \quad (13)$$

$$\frac{dI_1}{dt}=\mu\left(2-e^{-a_1 I_1}-e^{-\alpha_1 I_a}\right)S-\mu\left(1-e^{-a_2 I_2}\right)I_1-\theta_1 I_1 \quad (14)$$

$$\frac{dI_2}{dt}=\mu\left(2-e^{-a_2 I_2}-e^{-\alpha_2 I_a}\right)S-\mu\left(1-e^{-a_1 I_1}\right)I_2-\theta_2 I_2 \quad (15)$$

$$\frac{dI_a}{dt}=\mu\left(1-e^{-\alpha_a I_a}\right)s+\mu\left(1-e^{-a_2 I_2}\right)I_1+\mu\left(1-e^{-a_1 I_1}\right)I_2-\theta_a I_a \quad (16)$$

$$\frac{dR}{dt}=\theta_1 I_1+\theta_2 I_2+\theta_a I_a \quad (17)$$

S represents the number of people who are susceptible to both the resident and the mutant, I1 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, I_a represents the number of people who are infected by both strands, R represents the number of people who have gained immunity, θ_I represents the inverse of the infectious period of the resident strand, θ_2 represents the inverse of the infectious period of the mutated strand, θ_a represents the inverse of the infectious period of individuals carrying both strands, α_1 represents the resident strand's transmission rate from a person in the Infectiousa compartment, α_2 represents the mutated strand's transmission rate from a person in the Infectiousa compartment, α_a represents the probability of an individual who carries both strands transmits all two to a susceptible individual, a1 represents the resident strand's transmission rate from a person in the Infectious1 compartment. a2 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, I_a is set to 0, and S is set to $(Population) - I_1 - I_2 - I_a - R$.
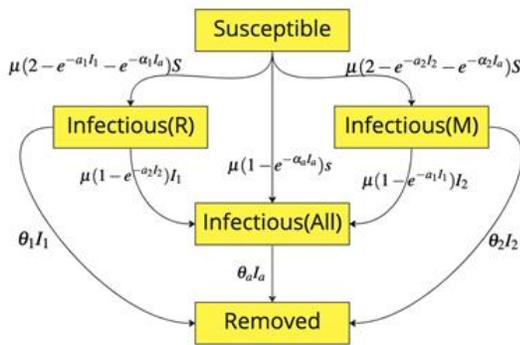
Figure 4.   Change between each compartment of the SI3R model summarized in a flowchart.

VII. ESTIMATING THE DEMAND FOR HOSPITALIZATION − THE IHOV MODEL

Using the EPSEIRV model, the author 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 (18):

$$\frac{dR}{dt} = \theta_1 I_1 + \theta_2 I_2 + \theta_a I_a \qquad (18)$$

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 related 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 model the transmission data of each community accurately.

Due to our time limitation, the author was 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 in each county and even globally. The author has implemented all New England states' parameters to the code, including population, population density, number of beds available, 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.

VIII. 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.

A. Infection Forecast

The best-fitting p for the data obtained in the United States equals to 3.57×10^{-8}. Please note that the author used positive test percentages

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, the author subtracted the percentage of Delta variant positive cases (used the model on Delta as well) from the total positive percentage. Also, the author assumed day 0 to be November 22, 2021, since the first detected case in the US is a traveler who returned from South Africa on November 22 [22].
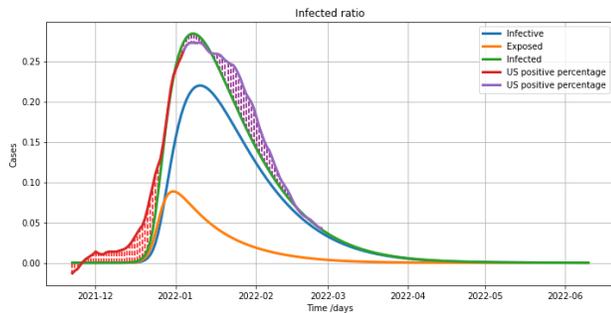
Figure 5.   Projected Positive Percentage as a Function of Time in the US Produced by the EPSEIRV Model

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 the author 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 based on the data before that day. Our model accurately predicted the peak of the 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.
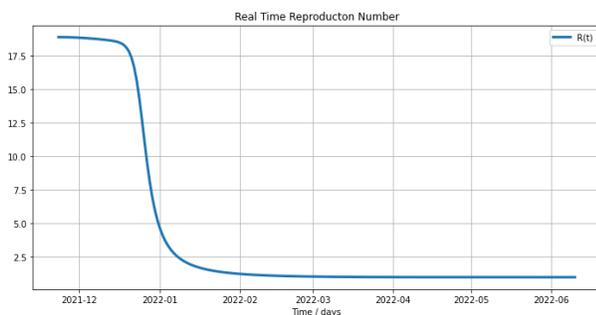
Figure 6.   Real-Time Reproduction Number as a Function of Time in the US Produced by the EPSEIRV Model

Figure 6 is the yielded real time reproduction number of Omicron in the US as a function of

time by the EPSEIRV model. This concludes that the R0 value of Omicron in the US is approximately 18.8, and its real-time 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 the author published our data on Medium.com, the CDC announced that the R0 value of Omicron was about 75. Now, with more projects examining the R0 number of Omicron in the US, an R0 number above 13 for Omicron has become somewhat of 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 error). Apparently, Omicron's absurdly high contagiousness is not well considered in the 1.2 value of alpha in the original model.

B. Variant Prediction
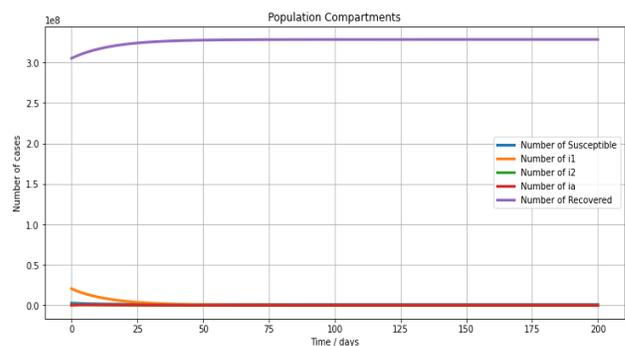
Figure 7.   Simulation of the Competition Between Omicron and a Hypothetical Mutant

Figure 6 models the increase of accumulative positive Omicron cases in the United States; it is found that the increase of accumulative positive Omicron cases will be slow until late December 2021. However, it will increase rapidly by March 2022, which more directly shows that most of the population will be infected with Omicron before the end of March 2022. After that, the cumulative number of cases will not increase and level off.

Since the existence of a mutant is hypothetical, the author 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, the author obtained Figure 7. The purple line represents the Removed category, the orange line is Infectious1, the minuscule green tip is Infectious2, and the almost straight red line is Infectiousa. This study 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 a 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.

C. Estimation for Hospitalization Resources
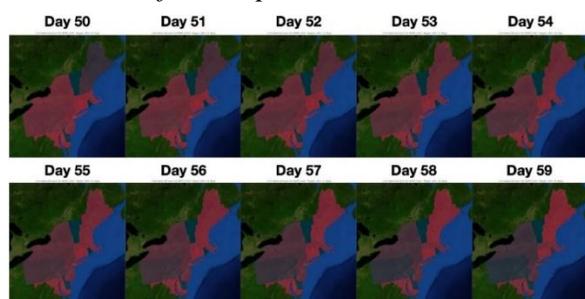
Figure 8.   Visual Display of the IHOV Model Results

For demonstration and testing purposes, the author ran the system based on Omicron data prior to January 6, 2022. Figure 8 shows 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, the author 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.

IX.   CONCLUSION

The COVID-19 pandemic substantially impacts almost all countries in multiple aspects and can spread to populations worldwide due to its high contagiousness. Therefore, it is a serious challenge for us to quickly address some of the most basic problems with emerging viruses.

In this study, the author focused on the many issues of the current disease transmission department that were exposed in the uncontrollable COVID-19 pandemic. The author built the EPSEIRV model, designed the SI3R model to expand our understanding of mutant competition, programmed the IHOV model to project the demand for inpatient resources, and created the overall PanDict system, which informs the public about infection prediction and variant emergence, and local resource shortages when new viruses emerge. The author modified the original SEIRV model to significantly improve its accuracy by eliminating the inaccurate index, α, and introducing population density and exposure time into the system of equations.

In addition, the EPSEIRV model can be further improved by moving people from the Removed category back to the Susceptible compartment to address reinfection. Secondly, the SI3R model supports mutant competition simulation, which can be used to better understand and predict the emergence of new variants. Besides, the IHOV model uses the generated predictions to estimate the local demand for inpatient resources. Based on this estimation, the proper arrangements of limited health resources can significantly reduce the staffing and resource shortages that hospitals experienced during the Omicron outbreak.

As a result, our system can minimize inpatient resource shortages and reduce unnecessary

economic losses and human casualties when new viruses appear. This will be good for the public and the health sector to better prepare for the emergence of new variants.

X. CODE AVAILABILITY

The relevant code is publicly available on GitHub at: https://github.com/123Bohan/LBH.git.

REFERENCES

[1] Stephanie Oum, Jennifer Kates, and Adam Wexler, "Economic Impact of COVID-19 on PEPFAR Countries," Global Health Policy, Feb 07, 2022.

[2] Kristine M. Smith,Catherine C. Machalaba,Richard Seifman... & William B. Karesh.(2019). Infectious disease and economics: The case for considering multi-sectoral impacts. One Health.

[3] Edouard Mathieu, Hannah Ritchie, Lucas Rodés-Guirao, Cameron Appel, Daniel Gavrilov, Charlie Giattino, et al. Coronavirus pandemic (covid-19). Our World Data (2020).

[4] R. M. Nayani Umesha Rajapaksha, Millawage Supun Dilara Wijesinghe, K. A. Sujith Prasanna Jayasooriya, W. M. Prasad Chathuranga Weerasinghe, B. M. Indika Gunawardana, An Extended Susceptible-Exposed-Infected-Recovered (SEIR) Model with Vaccination for Predicting the COVID-19 Pandemic in Sri Lanka. (Preprints)

[5] NCIRD & DVD. Omicron variant: What you need to know. N, CDC, 2018.

[6] Kevin, B. Fauci predicts omicron covid wave will peak in U.S. by the end of January. N, CNBC, Dec. 29, 2021.

[7] Laura, U. Expect more worrisome variants after omicron, scientists say. N, Associated Press, Jan. 15, 2022.

[8] Zappa Martina,Verdecchia Paolo & Angeli Fabio. Is the competition between variants the end of the severe acute respiratory syndrome coronavirus 2 pandemic? A journey from Wuhan to XBB.1.16. European journal of internal medicine.

[9] Topf Katherine G, Sheppard Michael, Marx Grace E, Wiegand Ryan E, LinkGelles Ruth, Binder Alison M, and Barbour Kamil E. Impact of the COVID-19 Vaccination Program on case incidence, emergency department visits, and hospital admissions among children aged 5-17 Years during the Delta and Omicron Periods-United States, December 2020 to April 2022. PloS one(12).

[10] Ao Li. Analysis on the development status and challenges of new employment patterns under the epidemic. Proceedings of 3rd International Conference on the Frontiers of Innovative Economics and Management (FIEM 2022)(pp.393-404).BCP Business & Management.

[11] Sandhu Simran K, Morozov Andrew Yu, and Farkas József Z. Modelling evolution of virulence in populations with a distributed parasite load. Journal of mathematical biology(1-2).

[12] Zhang Frank W, Meghoo Colin A, Staats Katherine L, Hayes Elizabeth Perkins, Metzner Mitch, Sobel Julia,Hultquist Eric, et al. A Spatiotemporal Tool to Project Hospital Critical Care Capacity and Mortality From COVID-19 in US Counties. American journal of public health,2021,111(6).

[13] Theodore K,David I. Law of mass action and saturation in SIR model with application to Coronavirus modelling. Infectious Disease Modelling,2021,6.

[14] Jansen Lauren, Tegomoh Bryan, Lange Kate, Showalter Kimberly, Figliomeni Jon, Abdalhamid Baha, et al. Investigation of a SARS-CoV-2 B.1.1.529 (Omicron) Variant Cluster - Nebraska, November-December 2021. MMWR. Morbidity and mortality weekly report,2021,70(5152).

[15] Klaassen Fayette, Chitwood Melanie H, Cohen Ted, Pitzer Virginia E, Russi Marcus, Swartwood Nicole A, et al. Changes in population immunity against infection and severe disease from SARS-CoV-2 Omicron variants in the United States between December 2021 and November 2022. medRxiv : the preprint server for health sciences,2022.

[16] Tian Z. COVID Mutation and Effect on Vaccine[C]//Eliwise Academy.Proceedings of the 5th International Conference on Economic Management and Green Development (ICEMGD V).Springer,2021:53-65.

[17] Anne R,Pablo G. Child Care Time, Parents' Well-Being, and Gender: Evidence from the American Time Use Survey. Journal of child and family studies,2016,25.

[18] Luwei Z,Sheng J,Shaoyao H. Spatial and Behavioral Studies of Daily Shopping Activities of Urban Residents in the Context of COVID-19[C]. A New Idea for Starting Point of the Silk Road: Urban and Rural Design for Human: Proceedings of the 14th International Conference on Environment-Behavior Studies (EBRA 2020).

[19] Krings Victoria C.,Steeden Ben,Abrams Dominic,Hogg Michael A., and Blanchard Anita L. The effects of COVID-19 on virtual working within online groups[J]. Group Processes & Intergroup Relations,2021,24(2).

[20] Ziye Lin. Analysis of the Used Car Market in the United States. Proceedings of 3rd International Conference on the Frontiers of Innovative Economics and Management (FIEM 2022)(pp.100-106).BCP Business and Management.

[21] Xiaoling Chen. Education Informationization in the Context of the Epidemic: Problems and Strategic Selection. Proceedings of International Conference on China and the World in the Context of the Globalization of COVID-19 (CWCGC)(pp.146-158).

[22] PérezSantosa Lissette, KourCardellá Vivian, TejeroSuárez Yahisel, MacíasRoig Lisandra M, PintosSaavedra Yanet, MederoDíz Dailyn, et al. Epidemiological Characterization of Patients in the First Eight Weeks Following Detection of SARS-CoV-2 B.1.1.529 (omicron) Variant in Cuba.[J]. MEDICC review,2022,24(3-4).

[23] Breban, Romulus & Vardavas, Raffaele & Blower, Sally. (2007). Theory versus Data: How to Calculate R0? PloS one. 2(3):e282.