Advanced Multi-Mutation With Intervention Policies Pandemic Model

TEDDY LAZEBNIK1 AND GADDI BLUMROSEN2,3
1Department of Cancer Biology, Cancer Institute, University College London, London WC1E 6DD, U.K.
2Department of Digital Medical Technologies, Holon Institute of Technology, Holon 5810201, Israel
3Department of Computer Science, Holon Institute of Technology, Holon 5810201, Israel
Corresponding author: Teddy Lazebnik (t.lazebnik@ucl.ac.uk)

ABSTRACT A pandemic is a threat to humanity with potentially millions of deaths worldwide. Epidemiological models can be used to better understand pandemic dynamics and assist policymakers in optimizing their Intervention Policies (IPs). Most existing epidemiological models assume, sometimes incorrectly, that a pandemic is caused by a single pathogen, ignoring pathogen mutations over time that result in different pathogen strains with different characteristics. In addition, the existing models do not incorporate the effect of IPs like vaccinations and lockdowns during the fitting phase. In this work, we introduce a new model called Suspected-Infected-Vaccinated-Recovered-reInfected (SIVRI). This model extends the SIRS model with adaptation to incorporate available knowledge related to the different pathogen mutations together with multiple IPs. In order to find the model parameters we propose a new fitting procedure that supports the complex social, epidemiological, and clinical dynamics that occur during a pandemic. We examine the suggested SIVRI model in comparison to the SIRS and XGboost models on the COVID-19 pandemic in Israel that includes four COVID-19 mutations, and the vaccination and lockdown IPs. We show that the proposed model can fit accurately to the historical data and outperform the existing models in predictions of basic reproduction number, mortality rate, and severely infected individuals rate.

MSC: 34-04 | 65-05 | 68U20

INDEX TERMS Pandemic with mutations, agent-based model fitting, biological-epidemiological model.

I. INTRODUCTION Throughout history, pandemics had the potential to be a large-scale disaster to national and international populations, by causing significant mortality and economical instability [1]. A pandemic is in mostly a contagious disease that spread in the population by means of sexual activity [2], body fluids transfer [3], or airborne [4]. During a pandemic, the original pathogen can mutate as part of the reproduction process in the infectious hosts [5]. Most mutations produce changes that are not significant enough to change the biological-epidemiological behavior of the pathogen [6]. However, some pathogen mutations have modifications that enable the disease to be more contagious or deadly with more severe patients and higher mortality rate [7]. Available information regarding the pathogens’ structure and the mutations process has potential to improve medications, develop efficient vaccination, and change the pandemic spread course [8], [9], and can assist policymakers in adopting IPs with most effect [10], [11].

Mathematical models and computer simulation are powerful tools to represent and analyze biological and epidemiological processes [12]–[15]. In particular, the Susceptible-Infected-Recovered (SIR) model [16] has been widely adopted as the baseline mathematical model to represent epidemiological dynamics [4], [17]–[19]. The SIR model describes the spread of a single pathogen in a mixed population and its main focus lies on how the prevalence and dynamics of infection vary with the transmission capacity of the pathogen and the characteristics of the host immune response [20]. Multiple extensions to the SIR model have been proposed in order to include different biological [21], economic [22], [23], and pandemic management decision-support [24]–[27]. Particularly, the Susceptible-Exposed-Infected-Recovered (SEIR), Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Recovered-Susceptible SIRS models are commonly used as extensions to the SIR model [28]–[30]. A schematic view of the SIR, SIS, and SIRS models is shown in Fig. 1, where $\beta$ is the average infection rate, $\gamma$ is the average recovery duration, and $\lambda$ is the average duration of the acquired immunity after infection effect decline through time.
These models can fit and predict the course of a pandemic spread for a short period and specific population, assuming adequate historical data to find the model’s parameters [31], [32]. In the case where the infectious pathogen mutates, these models fail to capture the pandemic spread over time. Consequently, to enhance the model accuracy it is required to incorporate into the model the effect of the pathogen mutations from epidemiological documentation based on population testing.

Several models were adjudicated to handle more than one pathogen’s strains. Gubar et al. [33] used a SIR model with two strains. The authors assume that each strain has a unique infection and recovery rate and divide the infected group into two sub-groups: infected with no symptoms (asymptomatic) and symptomatic. This model does not take into consideration the similarity between the strains and is limited to only two strains. A SIRS model that allows each individual in the population to be infected by one strain and then reinfected by the other was proposed by Gordo et al. [34]. The authors represented the multi-strain dynamic in the population by using a meta-population of individuals. The model proposed by [34] has been validated on the influenza pandemic and is based on the collected influenza strain statistics that were collected between 1993 and 2006. The model provides a better estimation of the pandemic spread compared to other SIR-based models [34], but the model complexity, and the dependency on massive collected data, makes it cumbersome to adopt the model to other pandemics with limited collected data. A more recent SEIR (E-exposed) model with two strains for the COVID-19 pandemic proposed by Khayar and Allali [35]. The model assumes that infectious from one strain produces immunity to all other strains originated from the same pathogen. The model provides comparable results to the others, and the model was implied their model to examine the effect of the duration between exposed and infectious stages on the epidemiological model predictions. This model is not suitable for mutations that cause high functionality differences between the strain properties, in particular with the infectious property [36]. In a similar manner, Arruda et al. [37] proposed an SEIR model with arbitrary number of mutation and reinfection of the same strain dynamics. Fudolig et al. [38] proposed a multi-strain SIR based model with selective immunity by vaccination. The authors examined the influence of a new strain introduction is made to emerge in the population when a preexisting strain has reached equilibrium.

Non-model-based methods for the pandemic spread prediction are associated with a small number of assumptions, such as time serious neural networks. One can utilize such methods to deal with complex temporal dynamics with changes over time due to the appearance of mutations and IPs [39]. An ensemble of Recurrent Neural Networks (RNNs) was applied to capture the temporal epidemiological changes caused by the two first waves of the Coronavirus disease (COVID-19) [40]. To adopt the model to different countries while preserving the common nature of the disease they used transfer learning and incorporated to the model country-specific features representing the country-specific governmental constraints. Prediction of COVID-19 mortality based on blood tests with different machine learning models (neural networks, logistic regression, XGBoost, random forests, SVM, and decision trees) was examined in [41]. The XGBoost model, a scalable Tree Boosting System [42], achieved the highest prediction rate with an accuracy of 90 percent as early as 16 days before the outcome. The XGBoost model was also deployed successfully to estimate the number of COVID-19 infection cases overcoming 24 days in every province of South Korea [43]. These time series models achieved fair results but require a long training time and a large number of samples (records), and do not incorporate known medical priors like the one represented by transition probabilities between states based on immune efficiency, and the effect of the different characteristics of each strain.

In this research, we develop a biologically inspired epidemiological model named SIVRI (Suspected-Infected-Vaccinated-Recovered-reInfected). The suggested model supports an unlimited number of pathogen mutations (and as a private case, strains) and also includes the effect of
policy intervention during the fitting process. The pathogen modeling characterizes the relation between the mutations, that usually originate from the initial strain, by genetic inspired representation based on recently available genetic data, and a similarity matrix between the strains (based on their mutations). The quantified similarity between the strains affects the infection probability, recovery rate, and recovery probability. Then we build new epidemiological states that incorporate also the effect of different policy interventions like vaccination and lockdown. We further develop in this work an in silico tool, allowing researchers to investigate multiple scenarios in a relatively quick, affordable, and controlled manner. 

We show the feasibility of the suggested model to evaluate the mortality rate and mean basic reproduction number of the COVID-19 pandemic in Israel.

The proposed model allows a more accurate investigation of the epidemiological dynamics by taking into consideration mutation processes of the pandemic with detailed biological dynamics of the infection and re-infection processes. This work has a four-fold contribution: 1) providing a new model with high performance and short convergence time; 2) the model includes support to more than two strains by generalizing the SIRS model to an arbitrary number of mutations originated from an ancestor pathogen; 3) incorporating to the model the effect of IPs like vaccination and lockdown; and 4) providing an in silico tool, allowing clinical and epidemiological researchers as well as policymakers to investigate multiple scenarios in a relatively quick, cheap, and controlled manner.

This paper is organized as follows: In Section 2, we provide biological inspiration for the proposed model definition, followed by a formal introduction of the multi-mutation biological-epidemiological model. Afterward, we present a novel method for fitting the proposed model on a reach historical data using an agent-based simulation. In Section 3, we present the implementation of the model for the COVID-19 pandemic in Israel and evaluate the performance of the proposed model. In Section 4, we discuss the main advantages and limitations of the model. In Section 5, we conclude our contribution and propose future work.

II. SYSTEM MODEL

A. BIOMEDICAL INSPIRATION

A pathogen like a virus or bacteria can be transmitted by interaction with other infected individuals [21]. The infected individual initiates the immune response. First, there is the innate immune response initiate to try to extinguish the pathogen, which starts in a few minutes, and does not require recognition of the pathogen [44]. After the innate response, if needed, an adaptive immune response takes place, ranging from days to weeks, for a more continuous and pathogen specific immune reaction [45], [46]. During the attempt of the immune system to overpower the disease and recover, the reaction of the infected individual reflected by different level of symptoms, ranging from no symptoms at all (asymptomatic) up to severe symptoms. When the body can not fully combat the pathogen, it can lead to death, or long term effects after recovery [47].

The adaptive immune system tries to adapt the immune response to the specific pathogen by producing protein components called antibodies (humoral response) and cells like T-cells (cellular response), and then coordinating their reactions to eliminate the pathogen [48], [49]. The adaptive immune system efficiency decreases over time, which resulted by a decrease in pathogen specific antibodies, and T-cells concentration for instance [50]. In case of reinfection with the same or similar pathogen, the immune system response would have higher efficiency in comparison with the original infection with usually a shorter recovery period. Even a long time after the recovery, the body can sometimes recover by reactivation of the stored immune response from a long-term immune memory associated more with T-cells and B-cells (to reproduce specific antibodies) [51]. This immune memory mechanism can decrease the chances of the individual having severe symptoms in case of reinfection after recovery.

Moreover, in the case of several strain that originated in one pathogen, the similarity between the strains is associated with the mutations and their change to the proteins that build the pathogen’s surface. In particular, a pathogen has several proteins (on its membrane, or in its body) that are encoded by its DNA [52]. The proteins can be divided into segments called epitopes that are the recognizable sections by the adaptive immune system that enable antibodies and T-cells to bind to the pathogen and start a process to eliminate it [53]. A similarity between two pathogen mutations therefore needs to be evaluated in relation to the change it cause to the pathogen epitopes genes and binding properties.

B. MODEL FORMULATION

The proposed SIVRI model is represented by a spatial-temporal system. The clinical-epidemiological dynamics of individuals are represented by a temporal set of Ordinary Differential Equations (ODEs). The spatial locations of the individuals are described using a graph-based model. In particular, the model is implemented using an agent-based realization, where each individual in the population is represented by its current location (in the graph) and clinical-epidemiological state, as a timed finite state machine [26]. The model and its states are described in a stochastic manner per individual, where each individual has a separate set of equations [54].

1) EPIDEMIC DYNAMICS

The model considers a constant initial fixed number of individuals $N$. We assume a pandemic has $M := \{1, \ldots, m\}$ different strains, where $[2, \ldots m]$ are strains with mutation indices (a change in the pathogen DNA sequence) originated from “ancestor” single pathogen, $m = 1$. Each individual can be at any given time, at one epidemiological state $p \in \mathbb{R}^{8M+1}$. The states can be divided into five clinical matching categories: susceptible, infected, recovered,
reinfected, and deceased. In the first (susceptible) category, there is one \(i\)-susceptible state, which includes uninfected individuals from any pathogen strain or ones that are unvaccinated \(i\). The second group of infected consists of two states: \(i\)-infected and \(i\)-severely-infected, where \(i\)-infected are both symptomatic and asymptomatic individuals that infected by the \(i\) strain and \(i\)-severely-infected are only individuals with severe symptomatic that are likely to be hospitalized for life-saving medical care in intensive care units (ICU) for medical care. Additional epidemiological state in the infected group is a product of policy intervention of vaccination and is named the vaccinated state. The vaccinated state is part of the infected group since an individual that was to the pathogen is likely to develop protection against the pathogen like infected individuals. Though there might be side effects to the vaccine, we assume in this work that vaccinated individuals are asymptomatic-like. In the third (recovered) category, there are two states: \(i\)-short-recovered and \(i\)-long-recovered. The \(i\)-short-recovered state stands for the initial period of an individual’s recovery when the immune efficiency is high, usually as measured by antibody tests (serology testing). The \(i\)-long-recovered state where the immune efficiency declines over time, like the decrease in the antibody count after a few months from recovery or vaccination. The \(i\)-long-recovered state is only after the \(i\)-long-recovered, where the individual has a decline in his immune efficiency. However, has long memory immune system. We assume that only from \(i\)-long-recovered state the individual can be reinfected. The fourth category (reinfected) has two states: \(i\)-reinfected and \(i\)-severely-reinfected, where \(i\)-reinfected include symptomatic and asymptomatic individuals were reinfected by the \(i\) strain and \(i\)-severely-reinfected are symptomatic individuals that arrive at the ICU for intensive medical care treatment in the second or later infection.

In the fifth (deceased) category there is only the state, which indicates individuals who died due to the disease.

Due to the similarity between the different pathogen strains (each with its unique difference in its mutations), the infection and recovery probabilities (both short and long term), and infection rate differ. Therefore, for a strain \(i\), with a baseline infection probability of \(b_i\), and individual recovery history within \(j \in \mathbb{P}(M)\) (the power group of the group of strains), the personalized infection probability \(B_i\) is as follows:

\[
B_i := b_i - \sum_{i < j \leq m} a_{ij} \beta_j \mathbb{I}(a_j > \epsilon_j)
\]

where \(a_{ij}\) is the similarity coefficient between the strain \(i\) and \(j\) based on the number of significant mutations (mutations that change the binding properties of the pathogen’s strains epitopes), \(a_j\) is the \(j\)th strain antibodies (and/or T-cell) concentration in the body, \(\epsilon_j\) is the \(j\)th antibodies (and/or T-cell) threshold that we assume satisfies high immunity efficiency, and \(\mathbb{I}(x)\) is the indication function that return 1 if the condition \(x\) is fulfilled and 0 otherwise. The condition states that the \(j\)th strain pathogen from the historical record is dominant enough, having adequate concentration of antibodies (and/or T-cells), to contribute to the immunity efficiency and to decrease the infection probability. In case the condition is not fulfilled, then the immune efficiency declines, and the individual move from short-term to long-term immunity state. Since the overall set of pathogen mutations are assumed to have a pairwise disjoint sets of biological marker, then the infection probability satisfy the probability constraints and \(0 \leq B_i\).

Similarly, an individual recover from the \(i\)-infected with a rate \(\Gamma_i\) where the baseline recovery rate if the \(i\) strain is \(\gamma_i\), such that

\[
\Gamma_i := \gamma_i + \sum_{i < j \leq m} a_{ij} \beta_j \mathbb{I}(a_j > \epsilon_j)
\]

\(\Gamma_i\) representing a rate, and non feasible negative values of rate, \(\Gamma_i < 0\), are set to zero, \(\Gamma_i = 0\).

In the same sense, an individual recover from \(i\)-reinfection with strain \(i\) at a rate \(\Lambda_i\) where the baseline reinfection recovery rate of the \(i\) strain is \(\lambda_i\), such that

\[
\Lambda_i := \lambda_i + \sum_{i < j \leq m} a_{ij} \beta_j \mathbb{I}(a_j > \epsilon_j).
\]

Likewise, \(\Lambda_i\) is defined for \(0 \leq \Lambda_i\).

To formally define the duration of the transformation between short-term and long-term recovery, we use the exponential decay in the antibodies (and/or T-cell) with a threshold value of \(\epsilon_i\) that differ for each strain, as follows:

\[
f_j(a_i) = e^{-\xi a_i} \wedge f_i(a_i) = \begin{cases} 1, & (a_i \leq \epsilon_i \land a_i^{-1} > \epsilon_i \land a_i^{-1} = 0) \\ \vee(a_i^{-1} = 1) \\ 0, & otherwise. \end{cases}
\]

The threshold value \(\epsilon_i\) and antibodies (and/or T-cell) decrease value \(\xi\) can be set based on clinical medical priors, or experimentally found by fitting the model on historical data.

We define a system characterized by nine model equations related to the \(8M + 1\) possible epidemiological states. In Eq. (5), \(dS_i(t)\) is the dynamical amount of individuals that are susceptible to the strain \(i \in M\) over time. It is affected by the amount of individuals that currently infected with strain \(i\) in either stage of infections in a probability \(B_i\).

\[
dS_i(t) = -S_i(t)(I_i^0(t) + I_i^s(t) + I_i^I(t) + I_i^1(t))
\]

In Eq. (6), \(dI_i^0(t)\) is the dynamical amount of non-severely infected by strain \(i \in M\) individuals over time. It is affected by the following four terms. 1) Susceptible individuals that are infected by any \(i\)-infected individual is becoming infected in probability \(B_i\). 2) Individuals recover at a rate \(\Gamma_i\). 3) Individuals become severely infected at a rate \(\Phi_i\). 4) Individuals that are severely infected recover at a rate \(\Phi_i^1\), becoming non-severely infected again.

\[
dI_i^0(t) = B_i S_i(t)(I_i^0(t) + I_i^s(t) + I_i^I(t) + I_i^1(t)) - \Gamma_i I_i^0(t) - \Phi_i I_i^0(t) + \Phi_i^1 I_i^1(t)
\]

In Eq. (7), \(dI_i^s(t)\) is the dynamical amount of severely infected by strain \(i \in M\) individuals over time, affected by the
terms: 1) individuals become severely infected at a rate $\Phi_i^d$ from non-severely infected; 2) individuals that are severely infected recover at a rate $\Phi_i^{r'}$, become non-severely infected; and 3) individuals that are severely infected die at a rate $\Phi_i^d$.

$$\frac{dI_s^d(t)}{dt} = \Phi_i^d I_s^d(t) - (\Phi_i^{r'} + \Phi_i^d) I_s^d(t)$$ (7)

In Eq. (8), $dR_i^s(t)$ is the dynamical amount of short-term recovered individuals from strain $i \in M$ over time, affected by the terms: 1) non-severely infected individuals recover at a rate $\Gamma_i$, move to short-term recovered state; 2) non-severely reinfected individuals recover at a rate $\Lambda_i$, move to short-term recovered again; and 3) short-term recovered individuals become long-term recovered individuals at a rate $k_i$

$$\frac{dR_i^s(t)}{dt} = \Gamma_i I_s^d(t) - k_i R_i^s(t) + \lambda_i I_i^d(t)$$ (8)

where $k_i$ is the time when the immunity move from short term to long term. Formally, it is defined as

$$k_i := \min (f_i(a_l^i)) = 1.$$  

In Eq. (9), $dR_i^u(t)$ is the dynamical amount of long-term individuals recovered from strain $i \in M$ over time, and affected by the terms: 1) short-term recovered individuals become long-term recovered individuals at a rate $k_i$; and 2) long-term recovered individuals are reinfect in strain $i$ with a probability $B_i$ by any individual that infected by strain $i$.

$$\frac{dR_i^u(t)}{dt} = k_i R_i^s(t) - B_i R_i^s(t)(I_s^d(t) + I_i^d(t) + I_j^d(t))$$ (9)

In Eq. (10), $dI_i^u(t)$ is the dynamical amount of non-severely reinfected by strain $i \in M$ individuals over time. It is affected by the terms: 1) long-term recovered individuals who are infected by any of the $i$-infected individuals become non-severely reinfected in probability $B_i$; 2) individuals recover at a rate $\Lambda_i$; 3) individuals become severely reinfeceated at a rate $\Psi_i^d$; and 4) severely reinfected individuals recover at a rate $\Psi_i^d$ and become non-severely reinfected again.

$$\frac{dI_i^u(t)}{dt} = B_i R_i^s(t)(I_s^d(t) + I_i^d(t) + I_j^d(t))$$ (10)

In Eq. (11), $dI_i^u(t)$ is the dynamical amount of severely reinfected by strain $i \in M$ individuals over time, and is affected by the terms: 1) non-severely reinfeceated individuals become severely reinfeceated at a rate $\Psi_i^d$; 2) individuals that are severely reinfected recover at a rate $\Psi_i^d$, becoming non-severely reinfeceated again; and 3) severely reinfeceated individuals die at a rate $\Psi_i^d$.

$$\frac{dI_i^u(t)}{dt} = \Psi_i^d I_i^u(t) - (\Psi_i^d + \Psi_i^d) I_i^u(t)$$ (11)

In Eq. (12), $\frac{dD_i(t)}{dt}$ is the dynamical amount of deceased individuals over time due to strain $i \in M$, and is affected by

$$\frac{dD_i(t)}{dt} = \Phi_i^d I_s^d(t) + \Psi_i^d I_i^u(t)$$ (12)

The dynamics of Eqs. (5-12) can be summarized in Eq. (13) for each strain $i \in M = [1, \ldots, m]$ as follows:

$$\frac{dS_i(t)}{dt} = -B_i S_i(t)(I_s^d(t) + I_i^d(t) + I_j^d(t))$$

$$\frac{dI_i^s(t)}{dt} = B_i S_i(t)(I_s^d(t) + I_i^d(t) + I_j^d(t))$$

$$\frac{dI_i^u(t)}{dt} = \Psi_i^d I_i^u(t) - (\Psi_i^d + \Psi_i^d) I_i^u(t)$$

$$\frac{dD_i(t)}{dt} = \Phi_i^d I_s^d(t) + \Psi_i^d I_i^u(t).$$ (13)

The initial conditions of Eq. (13) defined for the beginning of a pandemic as follows:

$$S_i(0) = N - 1, \quad I_i^s(0) = I_i^u(0) = R_i^0(0)$$

$$R_i^0(0) = I_i^u(0) = I_i^u(0) = D_i(0) = 0, \quad I_s^d(0) = 1$$ (14)

A schematic transition between epidemic states of an individual is shown in Fig. 2.

2) INTERACTION DYNAMICS

The social model is a graph-based model $G := (V, L \subset V \times V)$ where $V$ is a set of nodes, which represents locations and $L$ is the set of edges, which represents the connections between these locations. Each individual in the population is assigned to one node $v_j \in V$ in an undirected, connected graph $G$. Each agent has a separate clock (since individuals are represented by timed finite state machine based agents) where at each point in time they either stay in the same node (location) at a probability $\alpha$, or move to one of the neighbor nodes (e.g., $v_k \in V$ | $(v_j, v_k) \in L$) with probability $1 - \alpha$. We assume the transition between the nodes is immediate. At a fixed rate, the epidemiological dynamics are executed simultaneously over all graph’s nodes. Individuals that are located in the same node $v_j \in V$ can interact in a pair-wise manner and therefore can infect each other.

3) AGENT-BASED REPRESENTATION

The proposed model can be numerically solved using the agent-based technique [54]. An individual is defined by a tuple $\rho = (\mu, l, \tau)$ where $\mu$ is the epidemiological state of the individual, $l$ is the current individual’s location (corresponding to the node’s index), and $\tau$ is an inner clock counting the
FIGURE 2. A schematic view of an individual’s transformation between epidemiological states for a single pathogen (i’th) variant. Each individual is represented by a separate agent. The population dynamics in a pandemic, with multiple pathogen’s variants, can be predicted by aggregation of all agents’ behavior (multi-agent model), each with his immune efficiency to previous infection (cross immunity) or vaccination.

time pass from the last change in the epidemiological state. The epidemiological individual state changes based on the model equations (see Eq. (13)). To extend the model from a single agent representing an individual to a population, the parameters, and the probabilities can represent the mean value of the population. Then, for each time step, based on a synchronized clock, the population moves on the graph according to some policy.

C. MODEL FITTING
The proposed model epidemiological prediction quality depends on the model parameters fitting procedure and on the size, diversity, and accuracy of the historical data used to train the model.

1) INACCURACIES IN THE SAMPLED DATA
The sampling of the population statistic through epidemiological testing and medical data evaluation is biased, which can induce errors in the model. One example is the definition for severe cases that can differ between populations (states) and in some times its medical criterion can slightly change over time [55], [56]. Another example is the determination of the death cause, in particular when the deceased individual had other background diseases besides being positive for the pathogen [57], [58]. In addition to the non accurate medical data, the epidemiological data is commonly under-sampled in comparison with the real data [59], [60]. For example, less active sampling during the weekend compared to the rest of the week. In addition to the medical accuracy of the samples and the inaccuracies in the sampling procedure, there is the effect on data of the epidemiological actions to decrease the pandemic spread taken by the policymakers like the government [61], [62]. For example, a government can force a lockdown for some period [61] or vaccinate a portion of its population [63]–[65]. These interventions are complex nonlinear processes that affect and alter the course of the natural pandemic spread and make the fitting more complex.

2) MODEL FITTING ASSUMPTIONS
In this work, the model-fitting has been performed using the official published historical data by the WHO and neglects the effect of under-sampling [66]. This is because the accuracy improvement of the measured medical data is not in the scope of this work. Still, We assume that the goodness of fit of the suggested model will be the same with more accurate data. In addition, due to the important effect of the policy intervention, we do incorporate the intervention policies data in addition to the epidemiological data to the suggested fitting model, the policy interventions. Consequently, to fit the proposed model we suggest the following types of data:

- Epidemiological data: the daily number of infected, recovered, and deceased individuals. In addition, new mutation introduction dates.
- Clinical data: the daily number of new ICU cases, representing the severely (re)infected individuals.
- Intervention policies data: binary values indicates if a lockdown is taking place and the daily number of fully (two doses and up to six months from the second dose) vaccinated individuals.
- Social data: distribution of the populations into settlements and the average traffic between them.

3) MODEL FITTING PROCEDURE
Based on this data, one can use the fitting method proposed by [22]. Namely, set the model’s parameter values at random at the beginning. At each iteration of the algorithm, compute the fitness of the simulation using a given metric $d$. Afterward, for each parameter $p$ run the simulation for $p \pm \lambda$ where $\lambda > 0 \in \mathbb{R}$ is a local environment coefficient. After computing for all the parameters, one obtains a $x$-dimensional sphere, where $x$ is the number of parameters in the model, compute the numerical gradient in the parameter space and perform a learning step using the gradient descent method [67]. Repeat this process until the norm of the gradient of the algorithm is smaller than some predefined threshold $\epsilon > 0 \in \mathbb{R}$. Unlike in [22], the fitness
function used in the fitting process is defined as follows to take into consideration more historical data, (15), as shown at the bottom of the page, where \( s_t \) is the model’s prediction to the \( t_{th} \) day and \( h_t \) is the historical record for the same day. In addition \( h_t[I], h_t[D], h_t[R], \) and \( h_t[ICU] \) are the daily number of new infection, deceased, recovered, and hospitalized individuals, respectively. For a period \([t_0, t_f]\) where the comparison between the two dynamics takes place, the fitness function takes the form

\[
fitness(s, h)^2 := \Sigma_{t\in[t_0, t_f]} (W(t) d(s_t, h_t)^2), \tag{16}
\]

where \( W : \mathbb{N} \to \mathbb{R}^+ \) is a weight function such that \( \Sigma_{t\in[t_0, t_f]} W(t) = 1, \forall t \in [t_0, t_f] : W(t) \geq 0 \) and if \( t_i > t_j \) than \( w(t_i) > w(t_j) \).

The proposed fitness metric (Eq. (15)) can be improved if more detailed historical data is available. One option, in particular, is assuming the infection data is a classified permutation. In such a case, one would be able to compare the number of infected individuals for each mutation produced in the simulation compared to the historical data.

### III. MODEL EVALUATION

#### A. EVALUATION MODEL DATA

To evaluate the proposed model accuracy we used the COVID-19 virus pandemic in Israel in a period between March 1, 2020, and July 1, 2021. The data include epidemiological, policy interventions, and medical clinical data and is publicly available. The data was diverse, included in the examined period four major strains (all based on mutations from the original pathogen), and policy intervention and the data collection was according to the international committee standards [68], [69].

The epidemiological data were retrieved from the world health organization (WHO), the social data from google maps, the clinical data from the Israeli ministry of health, and the intervention policies data was manually collected from the Israeli government’s official website’s news. From WHO, we retrieved the daily number of infected, recovered, and dead individuals out of the assumed population of five million individuals. From google maps, we built our spatial locations model, where each settlement with more than 10,000 individuals (picked manually) are the vertices, and the average daily traffic between two settlements of over 10,000 transformations are the edges. The probability that an individual stays in the same node (\( \alpha \)) is computed by the portion of transformations originated in the node \( v_j \in V \) divided by the overall population size in the same node. From the Israeli ministry of health, we collected the daily number of new hospitalized individuals in the ICU and the daily number of individuals that are fully vaccinated (obtained both injections of the vaccine). In addition, we obtain the total number of individuals registered in the health system which represents the initial size of the susceptible population. Finally, from the Israeli government’s official website we collected information related to the three lockdowns: the first lockdown took place between March 25, 2020, and April 18, 2021; second lockdown between September 25, 2020, and October 17, 2020; and the third lockdown was between December 27, 2020, and January 5, 2021. In addition, a one-day lockdown took place on April 29, 2020.

Formally, the data is than organized as a table such that each record representing the amount of individuals in the population at each epidemiological state on a daily basis, divided by the infectious mutation if such information is available. In addition, a binary feature indicating if their is a lockdown or not is included.

#### B. MODEL FITTING QUALITY

To evaluate the fitting quality we examined three fundamental epidemiological parameters: basic reproduction number, mortality rate, and severe (hospitalized) cases rate. These parameters are fundamental parameters used in epidemiology and are assumed to capture the epidemiological dynamics, including the effect of policy interventions. The basic reproduction number, defined as \( R_0(t) = (I(t) - I(t - 1))/ (R(t) - R(t - 1)), \) provides an estimation to the average pandemic spread, where \( R_0 > 1 \) indicates on continuous outbreak while \( R_0 < 1 \) indicates on a decay in the pandemic decay. Policy interventions like lockdowns or vaccinations aim to decrease \( R_0(t) \). The mortality rate and the severe cases rate, are both markers of the severity of the pandemic, but also indicate the medical treatment quality and success, which can decrease the mortality and severe cases rate.

The fitting results are shown in Fig. 3. The fitting seems to track well over the real data and capture the dynamics of the main parameters over time, where it seems that the fitting has an averaging the data, and excludes intermediate artifacts, as expected. The fitting has a low mean square error (MSE) of \( 0.024 \pm 0.07, 0.0008 \pm 0.0004, \) and \( 0.0004 \pm 0.0003, \) for the \( R_0 \), mortality rate, and severe cases rate, respectively.

#### C. MODEL PREDICTIONS

A prediction of the epidemic spread is essential for policymakers to design their IP (intervention policy) and control the epidemic spread. To evaluate the prediction power of the model and its generalization properties for periods in a range of weeks and months we predict the three epidemiological
parameters values that were used for fitting (the basic reproduction number, mortality rate, and severe cases rate) based on fitting on historical data. For this, we trained the model from the start of the pandemic and predicted the values of the parameters for two months between May 2, 2021, and July 1, 2021. The fitting results are shown in Fig. 4c.

The prediction model seems to follow the real-data trends with Pearson correlation of 0.918, 0.763, and 0.802, and mean square error (MSE) of 0.043, 0.051, and 0.033, for the $R_0$, mortality rate, and severe cases rate, respectively. For all parameters, the prediction accuracy at the start of the prediction period is higher. This can be explained by the training data capturing more accurately the recent epidemiological statistics at the start of the prediction period. The predictions are smoother in the suggested model compared to the real data, as expected from a high-quality statistical fitting process. There is a higher difference and lower correlation of the mortality rate and for the severe cases rate between the predicted and the real data. It can be explained for the mortality rate by instantaneous jumps in death rate, which can be also linked to the limited sample process with a limited number, medical care treatment, and data collection bias. The higher estimation of severe cases and the mortality rate mainly towards the end of the prediction can be explained by better medical care in the months of the testing or the effect of vaccination that was not captured well in the training data and is a topic for further investigation.

D. MODEL CONVERGENCE

A major property to evaluate stochastic model properties is its two main convergence properties: model convergence time of its parameters; and model parameters final residual error.

Figure 5 describes the mean value of the three fitting parameters residuals over 40 iterations. The error residuals were compared to the asymptotic value after 500 iterations. To ensure the asymptotic value we cross-verified it with a two-tailed paired T-test and show the mean of the distribution has a P-value of less than 0.01. The iteration number that ensures that the residual value is less than 0.1 percent from the maximal (0.001, compared to maximum parameters values of 1) was found to be $n = 37$ iterations. The results of the mean value of the parameters are shown in Fig. 5. From these
results, we can see that the mortality rate and the $R_0$, even converge close to their final value only after 4 iterations, where the severe cases rate, which is a more sensitive parameter that is related to complex model iterations require longer convergence time.

### E. COMPARISON WITH OTHER MODELS

To evaluate the performance of the proposed model we compare the model prediction accuracy to the common analytical SIRS model [70], and to the efficient machine learning model, XGBoost [42]. The SIRS model was chosen since our proposed model extends it by introducing multi-mutation dynamics and clinical-epidemiological priors. The XGBoost model was chosen since epidemiological properties prediction over time is a time series task, and XGboost is state of the art machine learning technique for clinical time-series tasks in general [39], [71], and in epidemiology in particular [43], [72]. The XGboost model is obtained the same data as the proposed model (see Section III-A) up to a desired date as the training cohort. The model is trained on the data with the k-fold ($k = 5$) cross-validation approach [73] such that at each day $t$ we predicted the next day $t + 1$ similarly to the proposed model. Afterward, during the testing phase, we query the trained model on the testing cohort, using the latest day of the training cohort as the input of the sequential queries process.

We compare the model’s mean square error for the three epidemiological parameters over one month (30 days), given data of three months (90 days) before. The fitting phase used the WHO [66] data about the daily number of susceptible, infected, recovered, and dead individuals. The proposed model was fitted as described in Section II-C3. Then the prediction phase continued the simulation for additional 30 days with a time step of 15 minutes (resulting in 2880 steps in time total). For fare comparison, we choose a test period that was without significant PIs (policy intervention).

The SIRS model’s model contains three parameters: the average infection rate ($\beta$), the average recovery rate ($\gamma$), the average rate which recovered individuals return to the susceptible statue due to loss of immunity ($\lambda$). During the fitting phase, these values were estimated using the least mean square [75] method, and during the prediction phase, the historical values of the beginning of the month are set
as the initial condition and the model is solved using the fourth-order Runge-Kutta algorithm [76]. The XGBoost is trained on a four-dimensional feature dataset with the epidemiological data from WHO [66]. The training and prediction phases have been performed using the Python XGBoost package.\(^4\)

The results summary for the prediction of the proposed model in comparison to the other models is shown in Table 1. The proposed model outperforms both the model-based SIRS model which it extends and the more data-based machine-learning model of XGBoost on average. The SIRS model did not outperform our model, which approves our model is a modified improved extension for the SIRS model, and for two out of the five periods, the XGBoost model slightly outperforms the proposed model.

Since the SIRS model does not contain severe infected and diseased states, we compare the other two parameters with the XGBoost only. The results are given in Tables 2 and 3. As for the \((R_0)\), our suggested model also outperforms on average the XGBoost model. Still, like for the \((R_0)\), we do see that in some periods the predictions based on the XGBoost model are slightly better. This indicates that the pure data approach (with some medical priors), with data-driven features, and higher computations has some advantages over the model-based approach.

### IV. DISCUSSION

The proposed SIVRI (Suspected-Infected-Vaccinated-Recovered-ReInfected) model developed in this study allows us to better represent and predict pandemic spread dynamics for a long period including changes to the pandemic’s pathogen throughout the mutation process and intervention policies (such as lockdowns and vaccination). The proposed model extends the traditional SIRS model by introducing a deceased group, splitting the infected state into non-severely and severely infected individuals, introducing transformation between short-term and long-term immunity memory states rather than a single recovery state, divide between reinfection and infection for a second time or more via long-term immunity memory, and by introducing spatial dynamics. On top of that, the model takes into consideration an arbitrary number of strains \(m\) and the similarity between them in the case of the strain mutations being mutated from an ancestor pathogen, as presented in Fig. 2.

Based on the proposed model, we introduced a novel fitting method taking into consideration four types of data: epidemiological, clinical, intervention policies, and social. The fitting process integrated the different information stream signals

---

\(^4\)The original code can be found at https://github.com/dmlc/xgboost
TABLE 2. The daily prediction error for morality rate for the SIVRI model in comparison to the XGboost model for 5 different time periods. Bold font indicates the most accurate model for each period.

| Fitting Period | Mar 2020 - Jun 2020 | Jun 2020 - Oct 2020 | Oct 2020 - Dec 2020 | Dec 2020 - Feb 2021 | Feb 2021 - May 2021 |
|----------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| Testing period | Jun 2020            | Oct 2020            | Dec 2020            | Feb 2021            | May 2021            |
| Proposed model | 0.28                | 0.35                | 0.19                | 0.07                | 0.11                |
| XGboost model  | 0.39                | 0.32                | 0.26                | 0.14                | 0.06                |

into one fitting process like in the case of vaccination it reduced the likelihood to be infected, or in case of a lockdown, it freeze the spatial state of an individual represented by an agent. This flexibility of the model enables more flexibility to the model to correspond to temporal changes like pathogen mutations or vaccinations that can alter the pandemic course as shown in Section II-C3.

The proposed model predication was tested on the data from the COVID-19 outbreak in Israel, using the historical data from March 1, 2020, to May 1, 2021 (14 months), as shown in Fig. 3. The fitting phase used the WHO [66] data about the daily number of susceptible, infected, recovered, and dead individuals. In addition, the data about the number of vaccinated and lockdowns in Israel from the Israeli ministry of health are taken into consideration as well. The proposed model was fitted as described in Section II-C3, obtaining a MSE of 0.024 ± 0.07, 0.0008 ± 0.0004, and 0.0004 ± 0.0003 for the basic reproduction number ($R_0$), mortality rate, and severe cases rate, respectively.

Furthermore, we showed that the model prediction for the fundamental epidemiological parameters of the basic reproduction number ($R_0$), mortality rate, and severe cases between May 2, 2021, and July 1, 2021, obtained a daily MSE of 0.043, 0.051, 0.033, respectively, as shown in Figs. (4a-4c). Moreover, we presented that the model parameters converge to their steady-state value with an error of fewer than 0.1 percent for all the three epidemiological parameters in less than 37 iterations of the model, making it relatively stable and applicable in the manner of computation time.

V. CONCLUSION AND FUTURE WORK

The proposed SIVRI model is a theoretical platform that has a potential to assist policymakers in the decision-making process by providing predictions on the effect of different policy interventions on the pandemic spread.

We plan the following future work: 1) include in the model and in the immune system efficacy more accurate information related to the mutation structure similar to the one used in [34]; 2) incorporate into the model more policy interventions to make the model reflect better the pandemic dynamics and have more accurate model’s parameters (like a building-level interaction model as proposed by [80]); 3) add an expose and not infected states; 4) use correction model to compensate for testing inaccuracies; 5) take into account the differences of the immune response between the antibodies and T-cell; and 6) include the effect of re-vaccination; These future directions can represent better the real pandemic dynamics and further improve the SICRI model prediction quality.

DECLARATIONS

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
CONFLICTS OF INTEREST/COMPETING INTERESTS
The authors have no relevant financial or non-financial interests to disclose.

DATA AVAILABILITY
All the data used in this research is available online, the sources are cited in the text.

CODE AVAILABILITY
The code is available via written request from the authors.

AUTHOR CONTRIBUTIONS
Conceptualization, data gathering, formal analysis and investigation, coding, manuscript preparation, were performed by Teddy Lazebnik; Conceptualization, Formal analysis, investigation and manuscript preparation were performed by Gaddi Blumrosen.

ACKNOWLEDGMENT
The authors would like to thank Svetlana Bunimovich-Mendrazitsky for the thoughtful discussions on the mathematical modeling. (Teddy Lazebnik and Gaddi Blumrosen contributed equally to this work.)

REFERENCES
[1] A. A. Conti, “Historical and methodological highlights of quarantine measures: From ancient plague epidemics to current coronavirus disease (COVID-19) pandemic,” Acta Bio-Med. Atenei Parmensis, vol. 91, no. 2, pp. 226–229, 2022.
[2] H.-F. Huo, Q. Yang, and H. Xiang, “Dynamics of an edge-based SEIR model for sexually transmitted diseases,” Math. Biosci. Eng., vol. 17, no. 1, pp. 669–699, 2020.
[3] X. Wang, Z. Wang, and H. Shen, “Dynamical analysis of a discrete-time SIS epidemic model on complex networks,” Appl. Math. Lett., vol. 94, pp. 292–298, Aug. 2019.
[4] I. Cooper, A. Mondal, and C. G. Antonopoulos, “A SIR model assumption for the spread of COVID-19 in different communities,” Chaos, Solitons Fractals, vol. 139, Oct. 2020, Art. no. 110057.
[5] C. Pan, B. Cheung, S. Tan, C. Li, L. Li, S. Liu, and S. Jiang, “Genomic signature and mutation trend analysis of pandemic (H1N1) 2009 influenza virus,” PLoS ONE, vol. 5, no. 3, e9549, Mar. 2010.
[6] D. Wilson, D. Falush, and G. McVean, “Genetic diversity in the SIR model of pathogen evolution,” Proc. Roy. Soc. London A, vol. 2006, no. 12, pp. 302–315, Dec. 2005.
[7] M. G. M. Gomes, L. J. White, and G. F. Medley, “Infection, reinfection, and vaccination under suboptimal immune protection: Epidemiological perspectives,” J. Theor. Biol., vol. 228, no. 4, pp. 539–549, Jun. 2004.
[8] M. Agarwal and A. S. Bhadauria, “Modeling spread of polio with the role of drug treatment,” J. Theor. Biol., vol. 237, no. 3, pp. 302–315, Dec. 2005.
[9] A. Yusuf, S. Qureshi, M. Inc, A. I. Aliyu, D. Baleanu, and A. A. Shaikh, “Two-strain epidemic model involving fractional derivative with Mittag–Leffler kernel,” Chaos, Interdiscipl. J. Nonlinear Sci., vol. 28, no. 12, Dec. 2018, Art. no. 123121.
[10] I. Nesteruk, “Statistics-based predictions of coronavirus epidemic spreading in mainland China,” Innox. Biosyst. Bioeng., vol. 4, no. 1, pp. 13–18, Feb. 2020.
[11] T. Lazebnik, N. Aaroni, and S. Bunimovich-Mendrazitsky, “PDE based geometry model for BCG immunotherapy of bladder cancer,” Biosystems, vol. 100, Feb. 2021, Art. no. 104319.
[12] S. Towers, K. Vogt Geisse, Y. Zheng, and Z. Feng, “Antiviral treatment for pandemic influenza: Assessing potential repercussions using a seasonally forced SIR model,” J. Theor. Biol., vol. 289, pp. 259–268, Nov. 2011.
[13] S. He, Y. Peng, and K. Sun, “SEIR modeling of the COVID-19 and its dynamics,” Nonlinear Dyn., vol. 101, no. 3, pp. 1667–1680, Aug. 2020.
[14] D. Bognanni, D. Hanley, D. Kolliner, and K. Mitman, “Economics and epidemics: Evidence from an estimated spatial econ-sir model,” Int. Labor Econ., Bonn, Germany, Tech. Rep. IZA DP 13797, 2020. [Online]. Available: https://www.iza.org/publications/dp/13797/economics-and-epidemics-evidence-from-an-estimated-spatial-econ-sir-model
[15] S. Mwalili, M. Kimathi, V. Ojiambo, D. Gathungu, and R. Mbogo, “SEIR model for COVID-19 dynamics incorporating the environment and social distancing,” BMC Res. Notes, vol. 13, no. 1, p. 352, Dec. 2020.
[16] S. Meoin, N. Nickaen, A. Roointan, N. Borhani, Z. Heidary, S. H. Javannard, J. Ghasiari, and Y. Ghesiari, “Inefficiency of SIR models in forecasting COVID-19 epidemic: A case study of Isfahan,” Sci. Rep., vol. 11, no. 1, p. 4725, Dec. 2021.
[17] G. Chowell, L. Sattenspiel, S. Bansal, and C. Viboud, “Mathematical models to characterize early epidemic growth: A review,” Math. Biosci., vol. 18, pp. 66–97, Sep. 2016.
[18] E. Gubair, V. Tayniksky, and Q. Zhu, “Optimal control of heterogeneous mutating viruses,” Games, vol. 9, no. 4, p. 103, Dec. 2018.
[19] I. Gordo, M. G. M. Gomes, D. G. Reis, and P. R. A. Campos, “Genetic diversity in the SIR model of pathogen evolution,” PLoS ONE, vol. 4, no. 3, pp. e84876, Mar. 2009.
[20] O. Khayr and K. Allali, “Global dynamics of a multi-strain SEIR epidemic model with general incidence rates: Application to COVID-19 pandemic,” Nonlinear Dyn., vol. 102, no. 1, pp. 489–509, Sep. 2020.
[21] T. Lazebnik and S. Bunimovich-Mendrazitsky, “The signature features of COVID-19 pandemic in a hybrid mathematical model—Implications for optimal work–school lockdown policy,” Adv. Theory Simul., vol. 4, no. 5, 2021, Art. no. 2000298.
[22] T. L. Drake, Z. Chalabi, and R. Coker, “Cost-effectiveness analysis of pandemic influenza preparedness: What’s missing?” Bull. World Health Org., vol. 90, no. 12, pp. 940–941, Dec. 2012.
A. Casadevall, “Passive antibody administration (immediate immunity) and herd immunity against SARS-CoV-2 virus: Implications from current and past knowledge,” Pathogens Disease, vol. 78, no. 3, Apr. 2020, art. no. fta025.

M. Kobeliszcze, M. T. Traugott, I. Medits, F. S. Spitzer, A. Zoufaly, L. Weselsliindtner, C. Simonitsch, T. Seitz, W. Hoepler, E. Puchhammer-Stöckl, S. W. Aberle, M. Födinger, A. Bergthaler, M. Kundi, F. X. Heinz, K. Stiasny, and J. H. Aberle, “Dynamics of CD4 T cell and antibody responses in COVID-19 patients with different disease severity,” Frontiers Med., vol. 7, p. 711, Nov. 2020.

R. J. Cox and K. A. Brodkast, “Not just antibodies: B cells and T cells mediate immunity to COVID-19,” Nature Rev. Immunol., vol. 20, no. 10, pp. 581–582, Oct. 2020.

D. A. Winer, S. Winder, L. Shen, P. P. Wadia, J. Yantha, G. Palbers, H. Tou, P. Wu, F. Davidovin, M. N. H. X. Leong, A. Glassford, M. Caimol, J. A. Kenkel, T. F. Tedder, T. McLauglin, D. B. Miklos, H.-M. Dosch, and E. G. Englemann, “B-cells promote insulin resistance through modulation of T cells and production of pathogenic IgG antibodies,” Nature Med., vol. 17, pp. 610–617, Apr. 2011.

M. Hellerstein, “What are the roles of antibodies versus a durable, high quality ‘Fc’ cell response in protective immunity against SARS-CoV-2?” Vaccine X, vol. 6, Dec. 2020, art. no. 100076.

M. Grifoni, D. Weiskopf, S. I. Ramirez, J. Mateus, J. M. Dan, C. R. Modrbochcher, S. A. Rawlings, A. Sutherland, L. Premkumar, R. S. Jadi, D. Marrama, A. M. de Silva, A. Frazer, F. A. Carlin, J. A. Greenbaum, B. Peters, F. Krammer, D. M. Smith, S. Crotty, and A. Seta, “Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals,” Cell, vol. 181, no. 7, pp. 1489–1501, Jun. 2020.

M. Shen, Y. Zhou, J. Ye, A. A. A. Al-Maskri, Y. Kang, S. Zeng, and S. Cai, “Recent advances and perspectives of nucleic acid detection for coronavirus,” J. Pharmaceutical Anal., vol. 10, no. 2, pp. 97–101, Apr. 2020.

S. Mukherjee, D. Torrowski, R. Detroja, S. B. Mukherjee, and M. Frenkel-Morgenstern, “Immunoinformatics and structural analysis for identification of immunodominant epitopes in SARS-CoV-2 as potential vaccine targets,” Vaccines, vol. 8, no. 2, p. 290, Jun. 2020.

C. M. Macal, “To agent-based simulation from system dynamics,” in Proc. Winter Simul. Conf., Dec. 2010, pp. 371–382.

M. Ou, J. Zhu, P. Ji, H. Li, Z. Zhong, B. Li, J. Pang, J. Zhang, and X. Zheng, “Risk factors of severe cases with COVID-19: A meta-analysis,” Nature Med., vol. 148, p. e175, Aug. 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32782055/?doi=10.1016/j.nmed.2020.06.001

P. Weiss and D. R. Murdoch, “Clinical course and mortality risk of severe COVID-19,” Lancet, vol. 395, no. 10229, pp. 1014–1015, Mar. 2020.

J. W. Zylke and H. Bauchner, “Mortality and morbidity: The measure of a pandemic,” J. Amer. Med. Assoc., vol. 324, no. 5, pp. 458–459, 2020.

A. Sette, “Targets of T cell responses to SARS-CoV-2 coronavirus,” Cell, vol. 181, no. 7, pp. 1489–1501, Jun. 2020.

S. Mukherjee, D. Tworowski, R. Detroja, S. B. Mukherjee, and K. M. Jagodnik, F. Ray, F. M. Giorgi, and A. Lachmann, “Correcting for under-reported COVID-19 case numbers: Estimating the true scale of the pandemic,” medRxiv, Apr. 2020.

T. Lazebnik and A. Alexi, “Comparison of pandemic intervention policies in several building types using heterogeneous population model,” Commun. Nonlinear Sci. Numer. Simul., vol. 107, 2021.