An Agent-Based Digital Twin for Exploring Localized Non-pharmaceutical Interventions to Control COVID-19 Pandemic

Souvik Barat[1] · Ritu Parchure[2] · Shrinivas Darak[2] · Vinay Kulkarni[1] · Aditya Paranjape[1] · Monika Gajrani[1] · Abhishek Yadav[1] · Vinay Kulkarni[2]

Abstract

The COVID-19 epidemic created, at the time of writing the paper, highly unusual and uncertain socio-economic conditions. The world economy was severely impacted and business-as-usual activities severely disrupted. The situation presented the necessity to make a trade-off between individual health and safety on one hand and socio-economic progress on the other. Based on the current understanding of the epidemiological characteristics of COVID-19, a broad set of control measures has emerged along dimensions such as restricting people's movements, high-volume testing, contract tracing, use of face masks, and enforcement of social-distancing. However, these interventions have their own limitations and varying level of efficacy depending on factors such as the population density and the socio-economic characteristics of the area. To help tailor the intervention, we develop a configurable, fine-grained agent-based simulation model that serves as a virtual representation, i.e., a digital twin of a diverse and heterogeneous area such as a city. In this paper, to illustrate our techniques, we focus our attention on the Indian city of Pune in the western state of Maharashtra. We use the digital twin to simulate various what-if scenarios of interest to (1) predict the spread of the virus; (2) understand the effectiveness of candidate interventions; and (3) predict the consequences of introduction of interventions possibly leading to trade-offs between public health, citizen comfort, and economy. Our model is configured for the specific city of interest and used as an in-silico experimentation aid to predict the trajectory of active infections, mortality rate, load on hospital, and quarantine facility centers for the candidate interventions. The key contributions of this paper are: (1) a novel agent-based model that seamlessly captures people, place, and movement characteristics of the city, COVID-19 virus characteristics, and primitive set of candidate interventions, and (2) a simulation-driven approach to determine the exact intervention that needs to be applied under a given set of circumstances. Although the analysis presented in the paper is highly specific to COVID-19, our tools are generic enough to serve as a template for modeling the impact of future pandemics and formulating bespoke intervention strategies.

Keywords Covid19 pandemic · Digital twin of city · Agent based simulation · Simulation based control · What-if analysis

Introduction

In the midst of a pandemic like COVID-19, one of the key priorities of the public health administration is to understand the dynamics of the transmission of the pathogen (World Health Organization 2020; Organization et al. 2020) and use that knowledge to design effective control measures to keep its impact on public health within manageable and tolerable limits. In case of COVID-19, while the characteristics of the virus (i.e., mode of transmission and the typical trajectory of infection in an individual) are known to an extent from the existing research (World Health Organization 2020; Asadi et al. 2020; Cai et al. 2020; Wang et al. 2020; He et al. 2020a, b), the dynamics of its transmission and spread in a

✉ Vinay Kulkarni
    vinay.vkulkarni@tcs.com

    Souvik Barat
    souvik.barat@tcs.com

    Ritu Parchure
    ritu@prayaspune.org

    Aditya Paranjape
    aditya.paranjape@tcs.com

1   Tata Consultancy Services Research, Pune, India

2   Prayas Health Group, Pune, India

heterogenous population is not fully understood. It is known, though, that the spread of infection is related to people's movement, the nature of the area where people congregate (open-air versus closed), and number and frequency of proximal contacts. It is also known the demographic factors and comorbidity play a role in the spread of infection as well as its lethality. Therefore, the primary non-pharmaceutical intervention (NPI) of the public health authorities has been to restrict people's movement to varying degrees, i.e., through the so-called lockdowns. In addition to saving lives, lockdowns have been the primary instruments for managing the load on local healthcare systems.

The economic impact of the lockdowns imposed in 2020 has been recorded as being amongst the most adverse phenomena to impact the world economy (Fernandes 2020). Until the pandemic is brought under control through large-scale availability of medication or vaccines, the administrators need to decide whether or not lockdowns are needed, and their nature and duration. As such, there is no universal formula for answering these questions. This is because the dynamics of the spread of COVID-19 depend heavily on individual localities: their demographic profile, the prevalent social etiquette, the capacity of their healthcare systems, whether or not people comply with the administrative recommendations, etc. Therefore, devising effective tools and models (possibly on a continuously changing basis) to help administrators take decisions at a local level is an urgent requirement in the midst of the pandemic.

Use of statistical and mathematical models to understand the spread of a pathogen and to explore effective control measures is a well-established decision-making aid (Hethcote 1989; Marathe and Vullikanti 2013). A wide range of modeling, data visualization, and interpretation techniques have been developed to predict the spread of COVID-19 and to explore the efficacy of NPIs (Wynants et al. 2020). While some models have been found to be useful for exploring NPIs in a specific geography, others have been found wanting for their accuracy of prediction (Holmdahl and Buckee 2020). We believe that a universal model to predict the efficacy of NPIs for all geographies, countries, and cities across the world is a difficult proposition. Instead, a purpose-specific, locality-based, fine-grained model addressing a set of relevant aspects of interest can play a crucial role in decision-making for controlling the pandemic.

Contribution

In this paper, we develop a purpose-specific, configurable, extensible model of a city (referred to as a purposive digital twin of a city) which can be considered as a virtual environment to explore various hypotheses or interventions that policy-makers and public health practitioners might want
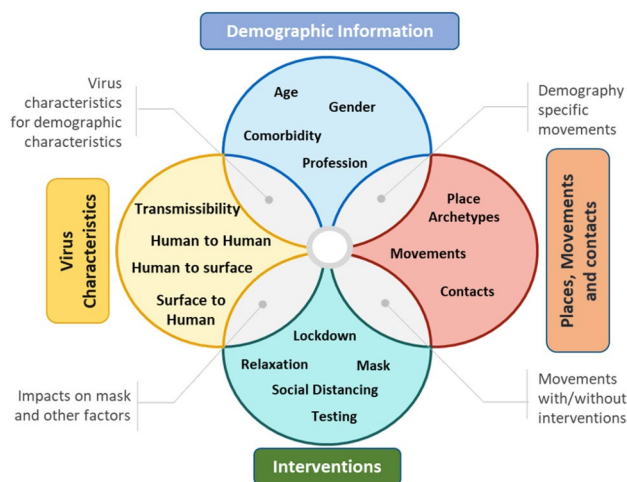
Fig. 1 Schematic of the digital twin of a city and the aspects of interest

to assess. Principally, we consider a city as an exceedingly complex system (Grieves and Vickers 2017), leverage the benefits of agent-based modeling paradigm (Macal and North 2009) [in compliance with epidemiological COVID-19 specific models (Silva et al. 2020; Rockett et al. rockett2020revealing; Agrawal et al. 2020; Kerr et al. 2020)], and rely on bottom–up simulation techniques to observe coarse-grained emergent properties of the system.

As shown in Fig. 1, the digital twin captures four aspects of interest:

1. *Epidemiological aspect:* Virus characteristics (Wang et al. 2020; He et al. 2020a, b).
2. *Demographic aspect of the city:* People archetypes, household structure, age, gender, and comorbidity of the population.
3. *Stochastic and context-specific socio-spatio-temporal movements of the population:* Movements specific to place, people, archetype, or profession during business-as-usual circumstances as well as under specific interventions.
4. *NPIs (referred to as interventions in this paper):* These include administrative interventions (such as closure of offices, shops, and schools or restrictions on business-as-usual movements of people), healthcare-specific interventions (such as testing, contact-tracing, and quarantine strategies) and interventions to promote appropriate social etiquettes (such as mask usage and social-distancing).

Unlike SIR/SEIR models (Li and Muldowney 1995; Agrawal et al. 2020), we do not explicitly model global behavioral patterns and infection spread dynamics using a set of differential or difference equations. Instead, we start by classifying

a city into a set of locality types (e.g., a residential area for well-to-do people with some shops and offices or a busy marketplace with low-income group residences and an office area). Thus, a locality type captures the people, place, and mobility characteristics succinctly. Each locality type is specified in terms of the necessary and sufficient agent types, each of which captures the state and the behavior of a real individual and their interaction with other individuals at a fine-grained level. Agent behavior is probabilistic, reflecting the choice of actions available with an individual to respond to an event of interest. Locality-level behavioral patterns emerge from the behavior of individual agents and their interaction with other agents. These behavioral patterns provide fine-grained information regarding who can come in proximal contact with whom in a given time frame. Combining this information with virus characteristics and with the health-related parameters and comorbidities of individual agents, we can determine how the infection spreads in a given locality using simulation. City-level infection spread can thus be arrived at inductively and can be used by the public health authorities to assess healthcare requirements and plan their intervention accordingly.

It is well known that different localities may require different kinds of intervention to control the spread of infection. The fine-grained nature of our model allows us to predict the efficacy of intervention techniques for individual localities, considering the differences between individual localities, and thus arrive at a bespoke optimal intervention for individual localities. For example, it is expected that a densely populated slum may require a strict lockdown and institutional quarantine centers where mildly infected patients can stay for 14 days, whereas a residential area for well-to-do people with some shops and offices can allow freer movement with mildly infected patients being quarantined inside their homes. By exploiting the simulatable nature of our city digital twin, we can validate of efficacy of such interventions in a quantitative manner.

Furthermore, while lockdowns or restrictions on business operations can mitigate the infection rates in some areas, these interventions hamper the growth or revival of an economy. A pragmatic solution, in all probability, will thus involve a set of locality-specific interventions that provide an acceptable trade-off between the spread of the infection and economic growth. The nature of this trade-off may vary from time to time, depending on the state of the pandemic, and socio-economic and political factors. The simulatable nature of city digital twin can provide a quantitative means of arriving at the desired trade-off.

We specify agent types using an actor-based language[1]. We populated the city digital twin as a set of interacting agents using the data available with city authorities and public health service agencies. We applied a robust construction, validation, and exploration methodology that has been extensively used for analyzing complex systems (Barat et al. 2017, 2019; Kulkarni et al. 2019). Authors from Prayas Health Group[2] validated all assumptions related to epidemiological and demographic aspects. Authors from Tata Consultancy Services Research[3] ensured correctness of city digital twin from a modeling perspective.

In this paper, we show how our fine-grained digital twin model can be used to predict and control the spread of COVID-19 virus in Pune City[4]. We specify (1) the epidemiological aspect using data published in peer-reviewed medical journals; (2) the demographic and movement aspects from official data available with Pune Municipality Corporation (PMC) and from the census data; and (3) interventions imposed in different localities/wards in the city from March 25, 2020 till date. The digital twin is first validated by correlating simulation results with the official data available with PMC and fine-tuned by interpreting various key parameters. The validated digital twin is then used to predict the spread of COVID-19 and estimate the load on the healthcare infrastructure for various candidate interventions. At the time of writing the paper, we note that our predictions continue to closely match the actual, recorded statistics as the epidemic unfolded in Pune[5].

While COVID-19 is part of the central theme of this paper, we recall how it has been widely advocated that the world needs to ensure a high level of preparedness for future pandemics, especially those caused by respiratory pathogens [see, for instance, Nuzzo et al. (2019)]. In this context, this paper must be seen not just as an exposition of our work on COVID-19 but as the presentation of a more general framework that is equally applicable to other pandemics and locales.

Organization

The remainder of this paper is organized as follows. Section 2 reviews different types of models, their capabilities, and limitations, justifying the need for a city digital twin. Section 3 presents technical details of our model. Section 4 describes how the model is configured and used for PMC. Section 5 evaluates the approach. Section 6 summarizes the paper and suggests directions for future work.

---

[1] http://www.esl-lang.org.

[2] https://www.prayaspune.org/health.

[3] https://www.tcs.com/creating-a-system-of-systems.

[4] https://en.wikipedia.org/wiki/Pune.

[5] https://indianexpress.com/article/cities/pune/pune-peak-load-on-critical-healthcare-to-be-sustained-until-oct-end-6562008/.

Exploration of the State-of-the-Art

Statistical and mathematical models (Marathe and Vullikanti 2013) often play a significant role in pandemic-related decision-making (Rhodes et al. 2020). For example, variants of susceptible-infected-recovered (SIR) models (Bailey et al. 1975) have been used for projecting infection transmission, understanding the load on medical infrastructure, and exploring control mechanisms for pandemics like HIV (Hallett et al. 2014), Hepatitis C (Martin et al. 2011), Severe Acute Respiratory Syndrome (SARS) (Lipsitch et al. 2003), H1N1 (Ferguson et al. 2006), and Ebola (Chretien et al. 2015).

Categorically, these models can be broadly divided into two types—(1) statistical models backed by experimental and/or historical data, and (2) computational models that faithfully represent the context under consideration. Computational models can be further divided into compartmental and agent-based (or microsimulation) models.

Statistical models (Simonsen et al. 2013; Samsuzzoha et al. 2013; COVID et al. 2020) predict the spread of a pathogen by estimating key epidemiological parameters, e.g., basic reproduction number (referred to as R_0) and the doubling time, from historical and real-time data collected from a specific area. These models are found to be useful only for the areas from where the data are collected and analyzed (Peng et al. 2020; Holmdahl and Buckee 2020) This is because, as explained in the previous section, the transmission of a pathogen depends significantly on local social factors. In the context of COVID-19, the construction of such statistical models for a specific city or country is a difficult proposition due to the unavailability of necessary data and lack of veracity of the available data: while data pertaining to hard-facts such as the number of deaths and the number of hospitalized people are known fairly accurately, the same cannot be said of the number of asymptomatic and mildly symptomatic patients in the geography under consideration.

SEIR Models

Compartmental models, chiefly the many variants of SIR and SEIR (Li and Muldowney 1995) models, represent different stages of infection (i.e., Susceptible, Exposed, Infected, Recovered, and Dead) as a set of compartments, and define the flow rules from one compartment to another in terms of differential or difference equations. In essence, compartmental models are constructed over an aggregated, homogenous population using a top–down modeling methodology. Among the wide variations and extensions of SIR model, the classic susceptible-exposed-infectious-recovered (SEIR) model has been widely adopted for COVID-19 (Prem et al. 2020; Agrawal et al. 2020; Keeling

et al. 2020; Teimouri 2020; Guan et al. 2020; Radulescu and Cavanagh 2020). It has been extended along two dimensions: (1) types of compartments, where stages such as asymptomatic, mild symptomatic, severe, hospitalized, quarantined, etc. augment the traditional compartments of SIR and SEIR models, and (2) stochasticity and temporal delays are added in the transition dynamics to model the uncertainty in aggregated movement from one compartment to other. For example, Giordano et al. (2020) use an extended SEIR model to capture eight distinct compartments: Susceptible, Infected, Diagnosed, Recognized, Ailing, Healed, Threatened, and Extinct stages of infection. The effects of various social-distancing interventions to control transmission and reduce the burden on healthcare system have been studied using an age-structured SEIR model (Prem et al. 2020) which is further extended to explore contact-tracing (Keeling et al. 2020; Agrawal et al. 2020).

From a modeling perspective, the key limitations of SEIR-based models are twofold: (1) inclusion of realistic (e.g., socio-economic) features requires a large number of compartments and parameters, increasing the difficulty of calibrating and validating the model parameters (Kerr et al. 2020), and (2) SEIR models are inherently unable to capture heterogeneities such as demographic and geographic characteristics (e.g., slum, well-to-do locality, housing societies etc.), household structure and dynamics (which is associated with almost 60–70% of total infection), and professional archetypes (Radulescu and Cavanagh 2020). The accuracy of SEIR-based models for COVID-19 is also questionable due to two key reasons. First, these models consider the estimated reproduction number [so-called R_t] to calculate the aggregate movement of population from the 'susceptible' compartment to the 'exposed' or 'infected' compartments (i.e., $S \rightarrow E$ or $S \rightarrow I$). The computation of R_t depends on historical or live data, and is specific to individual geographic areas. Finally, the movement from I (i.e., Infected) to R (i.e., recovered or dead) is grossly aggregated in SIR and SEIR models, usually ignoring the effect of demographic characteristics (e.g., age, comorbidity, and gender) of individuals.

Agent-Based Models

Agent-based modeling (ABM) can address some of the inherent limitations of SEIR/SIR models as it is capable of capturing the inherent heterogeneity of most populations. It allows individual (and heterogeneous) micro-elements within a given population or area to interact with each other (thus accurately reflecting the reality) to produce emergent, verifiable macro-behavior.

The ABM paradigm has been used as an aid to understand the spread of COVID-19 and the impact of interventions such lockdowns, contact-tracing, and social-distancing. For example, ABM has been used to understand the efficacy

of social-distancing and school closure in Australia (Chang et al. 2020). An agent-based simulator for an influenza epidemic has been repurposed to estimate the likelihood of human-to-human transmission of COVID-19 in a synthetic Singaporean population (Koo et al. 2020). Here, the authors showed how and why some interventions such as quarantine, school closure, and workplace distancing are more effective for the Singaporean population as compared to local containment and strict lockdowns. ABM has been used to assess public health measures or non-pharmaceutical interventions (NPIs) for reducing the contact rates (and thereby reducing the transmission of the virus) in the UK (Ferguson et al. 2020). A similar study has been conducted for the Indian cities of Mumbai and Bangalore in Agrawal et al. (2020).

ABM is also used for several micro-level analyses: a three-layer agent network (that includes school, household, and a joint workplace community as layers) with a stochastic behavioral model was adopted to represent a synthetic population of Boston Metropolitan Area and to simulate the efficacy of social-distancing and contact-tracing (Aleta et al. 2020). A similar synthetic population of a metropolitan area in the United States was constructed by considering four place archetypes (i.e., household, school, work, and other) to understand the efficacy of NPIs, such as work-from-home, liberal leave, home isolation, self-isolation, and home isolation with household quarantine of ascertained cases (Chao et al. 2020). Kerr et al. (2020) have developed an open source agent-based simulator called Covasim to explore a wide set of interventions such as physical distancing, hygiene measures, and testing-related interventions that include symptomatic and asymptomatic testing, contact-tracing, and quarantine. Conceptually, it captures (1) age and population size-specific demographic information, (2) transmission networks of four social population archetypes (viz., households, schools, workplaces, and communities), and (3) age-specific disease severity (or age-specific epidemiological aspect, in the context of Fig. 1). ABMs have also been reportedly[6] used in collaboration with local health agencies and policy-makers to interventions such as reopening of schools in the United Kingdom, fever-screening in Nigeria, partial workplace and community reopening in Australia, and epidemic projections for Eswatini and in the American states of Oregon, Colorado, and Washington.

Overall, agent-based modeling demonstrates its ability to model the inherent heterogeneity of the population and household structures, age-specific variation of epidemiological characteristics, and microscale variations of intervention policies. It also helps to simulate how a situation can emerge from the interactions of multiple heterogeneous agents with relatively known temporal, spatial, and spatio-temporal

behavioral patterns i.e., movements of individuals, contact propensity, disease progression probability, and mortality. While it shows a clear benefit over statistical and compartmental model, the key limitation of this model is inherent computational complexity—it is computationally expensive and difficult to scale as all individual elements, such as citizens, places, transport infrastructure, and their individualistic behaviors, need to be modeled and simulated. Moreover, this calls for fairly detailed understanding of individual elements at least at archetype level. Therefore, agent-based model is well suited for analyzing localized contexts such as a locality or a city as opposed to entire country.

ABM in the Indian Context

In our view, existing agent-based models (Kerr et al. 2020; Chang et al. 2020; Koo et al. 2020) for pandemic control fall short of addressing the high level of heterogeneity associated with India, e.g., the Indian population includes a wide range of professional archetypes with a uniquely rich set of characteristics and local variations, a large variety of places where people come together, a large variety of reasons for different groups of people to come together, place-specific movement characteristics, and a large set of possible local interventions. Most of the agent-based models reported in literature consider three professional archetypes (namely: office-goers, school students, and others), age structure (classification at 5–10 year intervals), and two levels of interaction patterns, namely: interactions at home and interactions at public place, where public places include office, school, and other place.

Agent-based model proposed by TIFR and IISc (Agrawal et al. 2020; Harsha et al. 2020) have considered several heterogeneous demographic characteristics while analyzing the Indian cities of Mumbai and Bangalore. Their model considers house, school, college, office, factory, shop, commute medium (mainly train), and community space along with the other typical demographic aspects, such as age, gender, and comorbidity. The key limitations of their model are threefold: (1) they have considered the mean size of households in their final computation, which fails to bring out the key differences in the household structure of slums, residential areas for well-to-do people, households with senior members, etc.; (2) household infection, place-specific infection, and severity-specific counts (i.e., number of asymptomatic, symptomatic, and severely infected individuals) are computed using a set of aggregated equations (Agrawal et al. 2020) thus generalizing individualistic behaviors to an extent, and (3) a generalized and simple temporal model is used to specify the physical movement of individuals, thereby ignoring the daily and weekly temporal variations such as those in the crowding seen in public spaces, in the load on the public transport during various time of a day

---

[6] Optima Consortium for Decision Sciences http://optimamodel.com.

(e.g., office/school/factory hours), in public gathering at places of worship at specific times on a given day, or in the weekend crowding at malls.

Since our primary objective is to capture Indian cities , we consider fine-grained India-specific specializations along demographic aspect, movement aspects, and intervention aspects. We keep the epidemiological aspect unchanged from that reported in the literature (i.e., person-to-person, aerosol-based, and fomite-based transmissions).

1. *Demographic:* Wide range of household structures with varying areas, a wide range of family size from two to ten members, a wide range of professions (e.g., office-goers, workers, shop owners, drivers, housemaid, housewife, bank employee, school, and college students), a wide range of transport infrastructure, age range from 0 to 80 +, genders, and comorbidities such as hypertension, diabetes, and chronic pulmonary diseases (CPDs).
2. *Movement and contact:* People's movements and contact are complex, uncertain, and they exhibit significant heterogeneity along social, spatial, and temporal dimensions. For example, the movements and the contact propensity of office-goers can be different when they are in office, at home, in a shopping mall, or at a local shop. The footfall at all of these places is strongly dependent on the time of the day as well as the day of the week. Similarly, the contact rate of any given office-goer (or any other archetype) may differ significantly depending on their choice of commute (e.g., own car, shared cab, or public transport). We capture this heterogeneity by introducing a large (and extensible) list of places in our simulator, such as school, college, (large or small) office, factory, marketplace, shopping mall, local shop, wholesale market, place of worship, etc.
3. *Intervention:* The availability of such fine-grained elements in our digital twin model allows us to explore several micro-level temporal interventions (that are seen in the lockdown and the unlock process in India), such as time-dependent curfew, partial time-dependent closure of non-essential shops, closure of places of worship, limitations on public gathering, restrictions imposed on domestic help, etc.

In the next section, we present the details of our methodology and a fine-grained model for a city-specific digital twin.
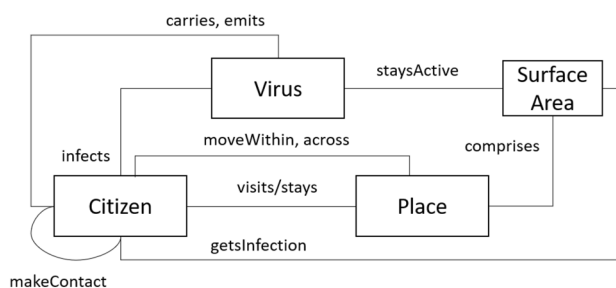

Fig. 2 Key concepts and relationships for the actor model

Methodology: Digital Twin of a City

General Principles

We visualize a city as a complex, dynamic system of systems (Boardman and Sauser 2006), whose macro-behavior (i.e., how a pandemic unfolds in the city with/without interventions) emerges from the behavior of its constituent micro-elements and their interactions, as shown in Fig. 2. Each of these constituent micro-elements has its own state, characteristics, and stochastic socio-spatio-temporal behavior.

From a system theoretic perspective, we characterize each micro-element as modular (i.e., encapsulates its own state, properties/characteristics, behavior and historical states/traces), reactive (i.e., interacts with other elements), composable (i.e., can be combined to form a larger element, e.g. place), autonomous (i.e., can act pro-actively without waiting for an external stimulus), and adaptive (i.e., can change its behavior over time). Moreover, these elements exhibit probabilistic spatio-temporal characteristics. It is cognizant of spatial relationship to exhibit place-specific behavior and to assess proximity with surrounding elements. Its behavior is cognizant of time and of progression of time.

We realize these micro-elements as extended forms of '*Agents*' (Barat 2018), where the canonical form of agent/actor abstraction (Agha et al. 1997; Macal and North 2009) is extended to capture the desired composability, uncertainty, spatio-temporal characteristics, and adaptation. An extended form of composition is considered for representing places like household, office, sub-area representing slum, and area for well-to-do population. We also consider relevant commuting means (e.g., bus, shared cab, and private cabs) of individual as a composite agent, since it has its own state (i.e., number of individuals and infected area), characteristics (i.e., area and capacity), and behavior (e.g., shared cab has one driver and passengers join and move out of vehicle).

Individual Citizens

A citizen has its own state about infection (e.g., susceptible, exposed, mildly infected, severely infected, and recovered) and location (e.g., staying at home or visiting workplace). They have individualistic characteristics, such as age, gender, comorbidity, and profession (e.g., office-goers, worker, students, and housewife). From behavioral perspective, each citizen may move from one place to other as the day progresses. These movements are uncertain and exhibit spatio-temporal characteristics—essentially a movement of an individual is a function of place, profession, and state of the individual. For example, a healthy office-goer may go to office during the weekdays and may have some propensity to shop for groceries or at a mall, and visit recreational places during weekends. In a given place, a citizen may move within that place and may come in contact with other citizens for a varying time and proximity. The same citizen is likely to exhibit significantly different movement pattern when unwell (the aforementioned adaptive property of agents). Here, we consider that all citizens are rational; emotive and psychological aspects of individual are not considered in this paper.

Epidemiological Aspect

From an epidemiological perspective, an infected citizen emits virus not only when sneezing and coughing but even while breathing normally. This leads to possible infection transmission either through person-to-person or through person-to-surface area-to-person (within a stipulated time). The transmission probability of person-to-person (through aerosol), person-to-surface-area, and surface-area-to-person (fomite infection) depends upon the virus characteristics which may change as virus mutates—we are not considering virus mutation in this paper.

Places

A place has a state (i.e., collective state of the individuals who are in the place), a set of properties or characteristics (such as area, operating hours and professions of permitted citizens), and behavior. The behavior of a place typically emerges from the behavior of its constituent micro-elements, i.e., citizens visiting/staying there. Infectiousness of these citizens, duration of their stay at the place, and plausible proximity of possible contacts play a role in virus spread dynamics of the place.

Digital Twin of a City

Holistically, a city is virtually represented as a digital twin by mimicking the state, the characteristics, and the behavior of all relevant constituent elements as interacting agents. The constructed digital twin is then used for understanding the spread of virus by simulating the behavior of the constituent agents and their interactions.

All agents of the constructed digital twin are first instantiated to the same states as reality and an estimated number of infected citizens are introduced in the digital twin. The behavior of all agents is simulated by triggering a sequence of discrete events that represent an hour (e.g., a simulation tick). We simulate the digital twin for pre-defined time epochs to understand how the pandemic unfolds in the city. This is compared with the real-life data for the same time epoch when the simulation is carried out post facto. The city digital twin thus validated is then used to check effectiveness of a candidate set of interventions through simulations.

The interventions (e.g., temporal restriction on movements, closure of specific places and commuting means, isolation of infected citizens through testing and contact-tracing, and mask adoption) are specified by introducing parametric changes in the characteristics and behavioral aspects of the agents. We adopt a systematic well-established three-step modeling and simulation methodology (Sargent 2013) to construct the digital twin of a city and use it for quantitative analysis. The steps are: (1) construction of a purposive digital twin of a city (a generic model for Indian cities), (2) contextualization and validation, for a specific city, and (3) experimentations and what-if scenario playing. These are explained in the following sections.

Constructing the City Digital Twin

We visualize a city digital twin as an extensible parameterized agent model that captures four interrelated aspects, namely, demographic, movements, epidemiological, and interventions, as highlighted in Fig. 1. Figure 3 depicts the meta-model of city digital twin.

Demographic Aspect

To support the wide spectrum of demographic heterogeneity of typical Indian cities in city digital twin, we consider variations of area specific population density, household structure and family size, occupational archetypes of local population, and three commonly considered demographic factors, namely: age, gender, and comorbidity of individuals.

As shown in Fig. 3, a city is a collection of administrative units or `Wards`, where each ward can be visualized as a set of `Localities` with unique characteristics in terms of `Citizens`, `Households`, and `Commercial Places`. The citizens have three properties namely `age`, `gender`, and `comorbidity`, where we consider three comorbidities, namely: `Hypertension`, `Diabetes`, and `Chronic Obstructive Pulmonary Disease` (COPD) that

Fig. 3 Meta-model of city digital twin
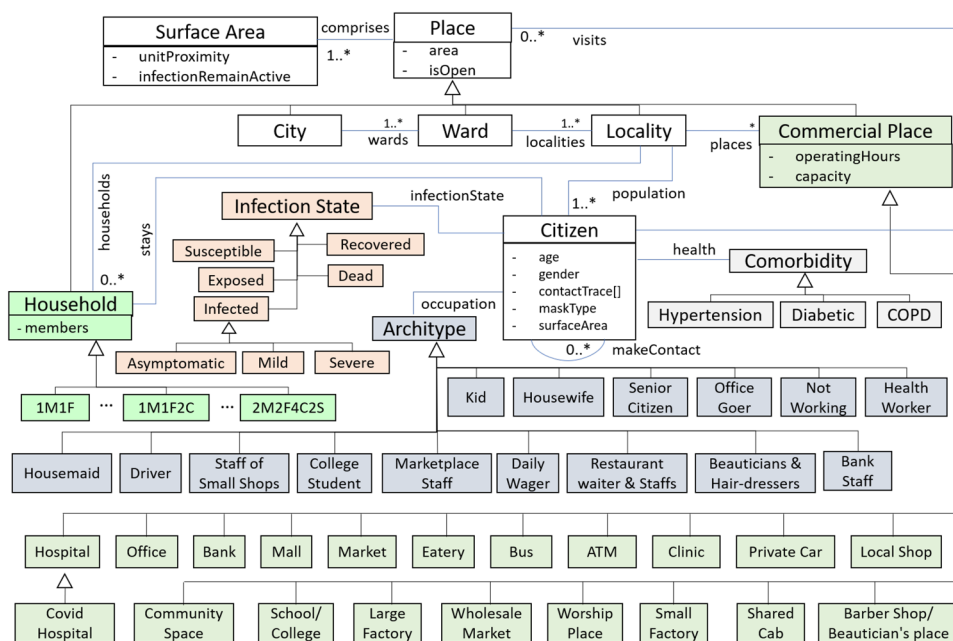
majorly influence severity of infection. A citizen belongs to a specific citizen `Archetype` as shown in the figure.

We have 21 types of household structures and 15 citizen archetypes. The household structures range from 2-member family (*1M1F*, i.e., 1 Male and 1 Female) to 12-member family (*3M3F4C2S*, i.e., 3 Male, 3 Female, 4 Children, and 2 Senior Citizens) that cohabit a house having specific area, which is a parameter for contextualization. Citizen archetypes cover a wide range that include commonly used archetypes, as seen in other agent-based models, such as *Kid*, *College Student*, *Senior Citizen*, and *Office-goers*. The list contains a set of unique and representative occupational archetypes. For example, a `Bank Staff` is different than `Office-goers` as they interact with other bank staffs and also with bank customers; a `Cab Driver` keeps interacting with varying number of passengers for different time span throughout the day; and a `Housemaid` visits multiple households and interacts with a fixed set of family members for close to 1 h or more in a day. Similarly, small shop keepers (e.g., `Staff of Small Shop`) interact with customers for short intervals, `Wagers` mostly work in congested places, and `Hairdressers` come in extended contact with their customers. Therefore, they have different propensity of getting and spreading infection while going about their daily routines.

Like household and citizen archetype, `Commercial Place` is also an extensible model element with 19 default commercial places, as shown in Fig. 3. Each commercial place has specific area (property of model element `Place`), and operating hours. These places have time-varying citizen visiting patterns and population densities over a day and over weekends. Therefore, they contribute differently in spreading infection (specific details are discussed in movement aspect).

Demographically, we visualize a city as a set of prototypical wards. Each ward is a combination of well-to-do and slum localities with representative set of households (with different structure and area), citizens (with different age, gender, comorbidities, and archetype), and commercial places. For example, a locality can be formed using two offices, three schools, hundreds of local shops, tens of barber shops, hundreds of clinics, and thousands of households with varying number of family members. Citizens from well-to-do localities may stay in relatively bigger houses with few family members as compared to slum area. Predominantly, the citizens from well-to-do area are office-goers, bank employees, health worker, and from other white-collar professions. On the other hand, slum areas are densely populated, and have smaller houses with bigger families. Typically, the citizens from slum area are engaged in professions (e.g., daily worker, drivers, housemaid, and restaurant staffs) that demand long working hours, may take them to crowded places, and do not have a fixed workplace.

In addition, we capture the distribution of commuting means (i.e., own car, bus, or shared cab) of the individuals in a locality. City administration and municipal corporation offices maintain record about ward count, number of localities in a ward, locality-specific population (including their age and gender), number of commercial places along with their types, number of registered public and private vehicles, and public transport infrastructure. For information related to comorbidity, guestimates available with healthcare organizations can be used.

Fig. 4 Epidemiological dynamics
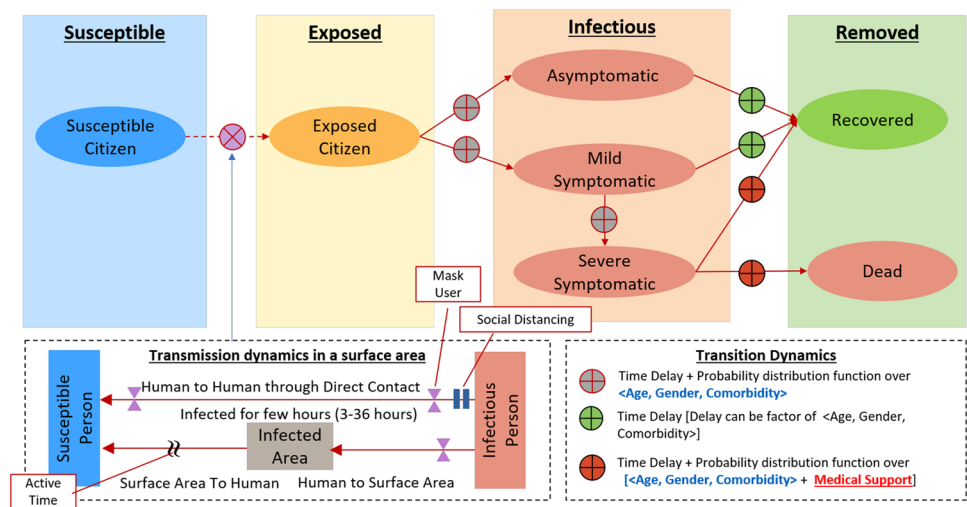
Movements

Business-as-usual movements of citizens can be broadly classified into movements within a `Place` (e.g., within office, school, and mall) and movements between places (e.g., home to car, car to office, office to shop, and shop to home). These movements depend on the demographic factors as well as the state of the citizens and are stochastic in nature. The key factors are: citizen archetype (who is the person), place archetypes (where the person is), and current time (when).

The movement within a place is inherently random; however, we adopt a rationalistic view to define movement patterns and movement frequency of an individual at a specific place. Essentially, they are functions over the tuple <citizen archetype, place archetype>. For example, a doctor stays in a specific location during the clinic hours; and visiting citizens (patients) wait in a waiting area for specific time (a range of time interval) and go to doctor's place before moving out of the place. Similar pattern is seen in hospital, bank, and barber shops. However, they may have different area and population density at a specific time; therefore, number and frequency of proximal contacts may differ significantly for different places. The movements within place for all <citizen archetype and place archetype> tuple are defined in our digital twin. While most of the combinations follow well-defined patterns, some movements are specified as random movements.

Movement between places can be further divided into three sub-categories, namely: (a) deterministic movements conforming to fixed pattern and time, (b) stochastic movements conforming to a fixed pattern, and (c) random movements. Deterministic movements conforming to fixed pattern and time primarily conform to three factors time of the day and week, state of the individual (mainly health- and infection-related state), and operating hours of the (commercial)

place. The examples of such types of movements are: office-goers going to office, students going to school, shop keepers going to shop, housemaid going to houses, and barbers going to barber shop. We use event–condition–action (ECA) paradigm to specify movement for professional place for each archetype. We further consider distribution of commuting means (defined in demographic aspect) to mimic commuting dynamics. The movements for professional places and utilization of commuting means conform to well-defined temporal patterns, which local administration has realistic guesstimate.

Stochastic movements conforming to a fixed pattern, such as movements to grocery shops, clinics, bank and ATM, eateries, barber shops, and wholesale market, are probabilistic in nature and conform to a repetitive pattern (daily, weekly, monthly, or other time interval) for each citizen archetype. We augment probabilistic actions in event–condition–action paradigm to specify this type of movements. Here, local administration is usually unawre of individual level movements, but they have reasonable understanding about the average footfall of each of the places (i.e., average footfall in bank, ATM, and clinic) in a day. We consider this information to validate assumption about probabilistic event–condition–action specification.

Random movements, such as visit to worship place, mall, community space, other localities, or other households (as guest), exhibit significant uncertainty. We use probabilistic event–condition–action where triggering event and action both are probabilistic in nature. We use heuristics (validated by local administration and/or public health organization, such as Prayas) to specify the involved probabilities and pattern for each archetype.

Due to all types of movements at different places, citizens make contacts with other citizens; we define a contact as a colocation of two or more citizens in a `Surface Area` of a place. Each citizen maintains (remembers) all its contact

using `contactTrace` attribute of citizen (we use this attribute to realize contact-tracing, which is described as Intervention aspect).

Epidemiological Aspect

We define infection dynamics by comprehending established facts and peer-reviewed literatures on COVID-19 virus. Here, we focus on two types of dynamics, namely: infection transition dynamics over infection stages, and virus transmission dynamics. Existing literature (He et al. 2020a, b; Kucharski et al. 2020) considers four prominent stages of a COVID-19-infected citizen, namely: susceptible, exposed, infected, and removed, as shown in Fig. 4. All citizens (irrespective of their age, gender, and other demographic characteristics) who are not yet infected are considered as susceptible. As shown in the figure, a susceptible citizen can be exposed to the virus either through aerosol or fomite. An exposed citizen becomes infectious after a time delay—typically after 3 days (this time delay is a parameter in our digital twin). In this stage, a citizen can remain asymptomatic or may develop mild symptoms. Subsequently, some mild symptomatic citizens may develop severe symptoms necessitating hospitalization. Age, gender, and comorbidities of individual are the primary factors for progression of infection in the individual (Guan et al. 2020). All asymptomatic and mildly symptomatic citizens recover after a time delay (a parameter in the digital twin). In addition, the medical treatment available also determines progression of severely symptomatic individuals or citizens. We assume that all severely symptomatic citizens undergo testing and get admitted to a special Hospital for COVID-19 patients, as shown in Fig. 3 (`Covid Hospital`). Recovered patients resume business-as-usual activities or are removed from the digital twin in case of death.

The transition time delays and transition propensity for citizens with various age, gender, and comorbidities are parameters in our digital twin. Table 1 depicts the transition propensities and Table 2 presents transition propensities, delays, and ranges (a synthesis of infection transmission dynamics (Chen et al. 2020; Guan et al. 2020; Chen and Li 2020).

We capture two types of virus transmission dynamics—(1) transmission at `Household` and transmission at `Commercial Places` (of the classification presented in Fig. 3). The household transmission to a susceptible citizen is computed based on three factors: (1) hourly household transmission rate (a model parameter), (2) citizens, i.e., family members, visitors, and housemaid, in the household at specific hour of the day, and (3) infection-related states of individuals in the household.

Transmission from an infected citizen to a susceptible citizen in a commercial place can happen when the susceptible citizen is within the proximity (where proximity distance is a model parameter) of an infected citizen for a specific time span (a model parameter). Essentially, a susceptible citizen is in same `Surface Area` (defined by `unitProximity` attribute with 2 m^2 as default value) of a `Place` with an infected citizen (see Fig. 3 for model element description). A transmission at commercial place is a function of: (1) hourly transmission rate, (2) individuals in proximal contact, and (3) infection-related states of all individuals in proximal contact.

A fomite infection is a two-step process, as shown in Fig. 4. An infected person infects surface area with a probability as she sneezes/coughs/breaths. This surface area stays infected for a span of time (a parameter named `infectionRemainActive` with value range 3–36 h) during which a susceptible citizen coming in proximal contact may get exposed to the virus (a parameter in digital twin).

We consider attack rates (i.e., infection probability) during incubation and infectious phase are 5% and 27%, respectively, as default values. Similarly, the rate of transmission through fomite are 5% and 15%, respectively, for incubation and infectious phases (Li et al. 2020).

Interventions

In India, a nationwide stringent lockdown was imposed on March 25, 2020. Different forms of lockdowns continued till May 2020 and then a gradual unlock started in a phased manner. Initially, all commercial places except few essential services were closed, which restricted more than 95% of the business-as-usual activities including offices, schools, colleges, non-essential shops, and factories. Household visits for housemaids, workers, and relatives were also restricted during the initial phases. We visualize these interventions, termed as `Intervention Strategy`, as a collection of changes in the form of business-as-usual activities, testing uptake, and behavioral characteristics of the citizens, e.g., wearing masks. We define a coherent set of interventions as intervention strategy indicating necessary changes on model elements of city digital twin (shown in Fig. 3) for a span of (simulation) time, as shown in Fig. 5. We consider interventions along three dimensions: administrative intervention, health care-related intervention, and social intervention. Administrative interventions are related to citizen movements and (partial) closure of places. Examples are (1) complete or partial closure of commercial places, such as offices, schools, and malls, (2) total or time-bound restriction of movements in and out of a locality (or a ward or a city, etc), (3) restriction of non-essential movement in a locality (or a ward or a city, etc), and (4) total ban or part-capacity operation of public transport and shared cabs. Introducing these interventions amounts to changing values

Table 1 Age, gender, and comorbidity properties

Comorbidity	Gender	Infection stage	80 + years	70–79 years	60–69 years	50–59 years	40–49 years	30–39 years	20–29 years	10–19 years	0–9 years
No history	Male	Dead	9.69	5.32	2.4	0.73	0.2	0.1	0.04	0.01	0.002
		Severely infected	22.86	12.54	5.66	1.73	0.47	0.23	0.09	0.02	0
		Asymptomatic	15.43	17.49	18.87	19.65	19.91	19.95	19.98	20	20
		Mild symptomatic	61.71	69.96	75.47	78.62	79.62	79.81	79.93	79.98	80
	Female	Dead	5.91	3.24	1.46	0.45	0.12	0.06	0.02	0.01	0.001
		Severely infected	13.94	7.65	3.45	1.05	0.29	0.14	0.05	0.01	0.003
		Asymptomatic	17.21	18.47	19.31	19.79	19.94	19.97	19.99	20	20
		Mild symptomatic	68.85	73.88	77.24	79.16	79.77	79.89	79.96	79.99	80
Diabetic	Male	Dead	13.649	7.49	3.377	1.006	0.28	0.144	0.057	0.014	0.003
		Severely infected	32.2	17.67	7.97	2.37	0.66	0.34	0.13	0.03	0.01
		Asymptomatic	13.56	16.47	18.41	19.53	19.87	19.93	19.97	19.99	20
		Mildly symptomatic	54.24	65.87	73.63	78.1	79.47	79.73	79.89	79.97	79.99
	Female	Dead	8.41	4.615	2.081	0.636	0.179	0.093	0.035	0.008	0.002
		Severely infected	19.84	10.89	4.91	1.5	0.42	0.22	0.08	0.02	0
		Asymptomatic	16.03	17.82	19.02	19.7	19.92	19.96	19.98	20	20
		Mildly asymptomatic	64.13	71.29	76.07	78.8	79.66	79.83	79.93	79.98	80
Hypertensive	Male	Dead	12.424	6.818	3.074	0.962	0.267	0.137	0.053	0.013	0.003
		Severely infected	29.31	16.08	7.25	2.27	0.63	0.32	0.13	0.03	0.01
		Asymptomatic	14.14	16.78	18.55	19.55	19.87	19.94	19.97	19.99	20
		Mildly symptomatic	56.55	67.13	74.2	78.18	79.5	79.74	79.9	79.97	79.99
	Female	Dead	7.485	4.107	1.852	0.59	0.167	0.087	0.034	0.008	0.002
		Severely infected	17.66	9.69	4.37	1.39	0.39	0.2	0.08	0.02	0
		Asymptomatic	16.47	18.06	19.13	19.72	19.92	19.96	19.98	20	20
		Mildly symptomatic	65.87	72.25	76.5	78.89	79.69	79.84	79.94	79.98	80
COPD	Male and female	Dead	19.077	10.468	4.72	1.443	0.391	0.214	0.08	0.018	0.004
		Severely infected	45	24.69	11.14	3.4	0.92	0.51	0.19	0.04	0.01
		Asymptomatic	11	15.06	17.77	19.32	19.82	19.9	19.96	19.99	20
		Mildly symptomatic	44	60.25	71.09	77.28	79.26	79.6	79.85	79.97	79.99

Table 2 Transition temporality

| Transition	Duration
Exposed to infectious	48–72 h
Asymptomatic to recovered	14 days
Mildly to severely symptomatic	6 days
Mildly symptomatic to recovered	21 days
Severely symptomatic to recovered	23–28 days
Severely symptomatic to dead	17–29 days

of appropriate model parameters of city digital twin model elements (shown in Fig. 3) suitably. For example, `isOpen` attribute of `Place` is used for opening and closing a place; `operatingHours` attribute of `Commercial Place` is used for partial time-dependent closure of places. Similarly, `isOpen` attribute of `Household` is considered to prohibit guest and housemaids in the households, `isOpen` attribute of `Locality` is used for defining containment zones, and `isOpen` attribute of `Ward` is used for closing boarders. The allowed passengers of `Cab` and `Bus` can be changed by changing capacity (of `Commercial Place`) attribute to an allowed number. We also have a provision to introduce sanitization of a place where a percentage of contaminated `Surface Area` can be considered as clean surface area as result of sanitization.

Subsequently, it reduces the change of fomite-related infection (see infection transmission logic of epidemiological aspect). From health care perspective, testing of severely infected citizens and admitting them to Covid hospital are mandatory responsibilities for city administrators as discussed in epidemiologic aspect. Interventions from health care standpoint include testing of mildly infected citizens (in addition to severely infected citizen), contact-tracing, and isolation of detected mildly infected citizens. The change in testing uptake is realized by (randomly) testing a certain percentage (a parameter) of mildly infected citizens. This mimics the scenario of encouraging citizens to undergo Covid testing who are experiencing Covid-like symptoms. Essentially, we control testing uptake of a city/ward/locality by tweaking the percentage of mildly symptomatic citizens in our model.

Contact-tracing is realized by testing a percentage of contacts of citizens, who are tested positive. We first compute who all have shared same surface area from infectious stage to testing time of an infected citizen (through `contactTrace` attribute of `Citizen` model element), and then, we select a specific percentage of citizens (contact-tracing percentage—a parameter) for contact-tracing. One can change uptake of contact-tracing by changing contact-tracing percentage in the model.

The isolation of citizens is achieved through home quarantine and institutional quarantine. In our model, the home quarantine is realized by (temporarily) muting all daily movements of the citizen and confining all household members at home using `isOpen` parameter of `Household`. Institutional quarantine is like hospitalization, where citizens are sent to institutional quarantine center till citizen recovers from the infection. Who can avail home quarantine and who should go to quarantine facility is a locality-specific administrative-decision—typically, citizens from well-to-do localities may avail home quarantine and citizens from slum area avail quarantine centers we use an attribute of locality model element named as
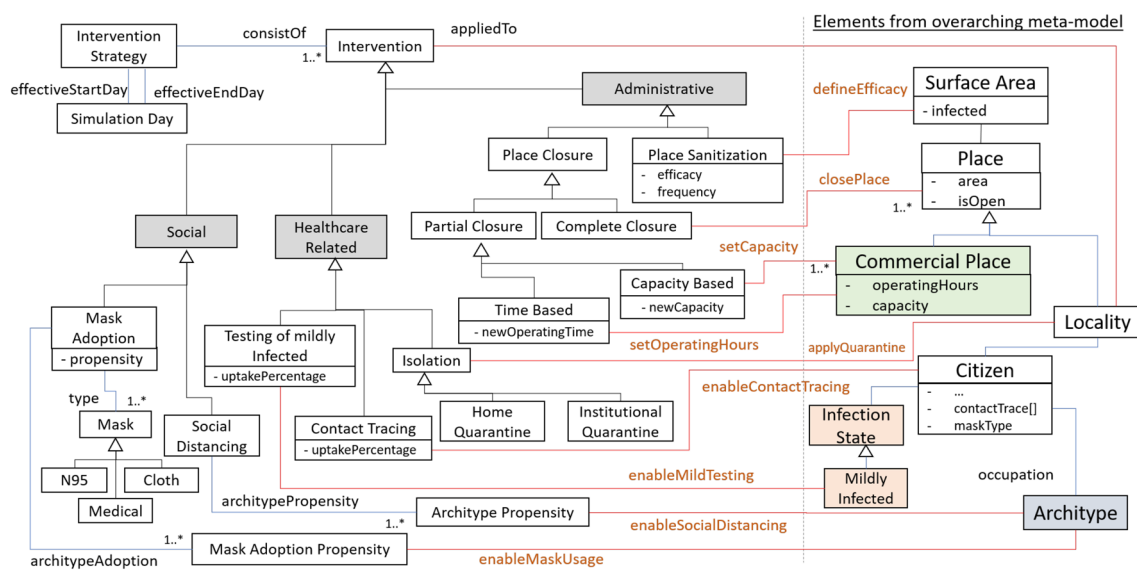
Fig. 5 Intervention meta-model and relationship with core model
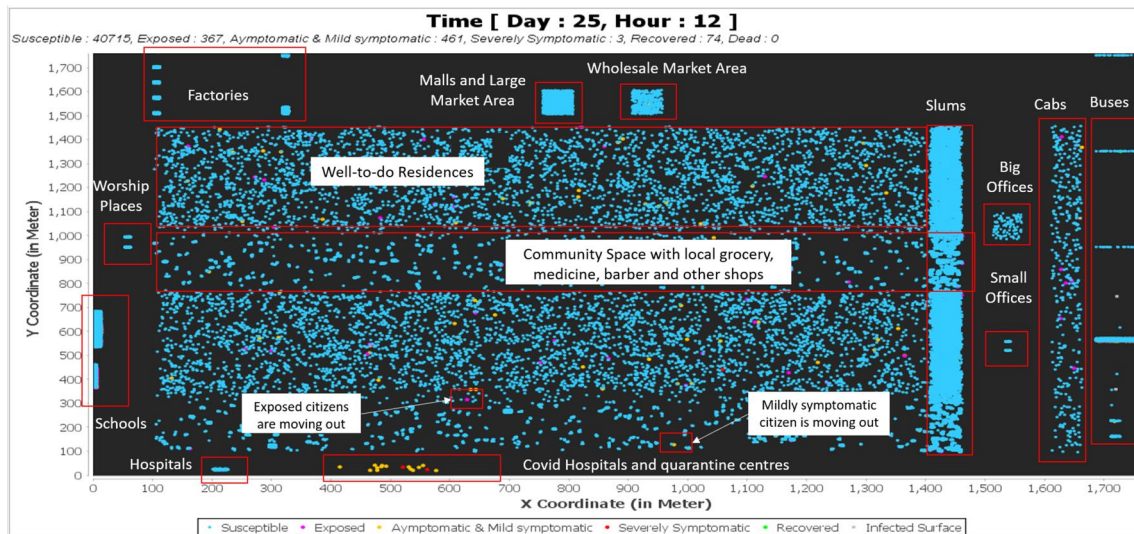
Fig. 6 A screenshot of the simulator, showing the ward-level agents and localities in the early phase of a simulation

`isolationOption` to indicate (a) if isolation should be enforced the citizens of a locality, and (b) what kind of isolation (i.e., home or institutional quarantine) should be enforced.

We consider two types of social interventions in our model: mask usage and social-distancing. The efficacy of the mask is realized through a set of model configurations and parameterizations.

1. Our model recognizes an extensible list of mask types with precise efficacy of spreading and protecting virus. Three mask types are added as default model elements, namely: N95 with efficacy 95%, medical mask with efficacy 70%, and cloth mask with efficacy 25% (Tian et al. 2020).

2. We use parametric model elements to capture archetype-specific distributions of mask type adoption. For example, one can specify all healthcare staff use N95; adoptions of N95, medical mask, and cloth mask among the senior citizens are—20%, 40%, and 40%. These adoption distributions help to set the value of `maskType` attribution of individual `Citizen` in our model.

3. The transmission dynamics from an infected citizen to susceptible citizen is cognizant of the efficacy of the mask (as shown in Fig. 4). Essentially, the probability of infection spread is adjusted based on (a) mask usage of susceptible citizen, (b) mask usage of collocated infected citizen, and (c) mask types of the both citizens.

Likewise, the social-distancing is realized using a parametric model element to capture archetype-specific adoption of social-distancing norm. We realize this norm by ensuring a citizen, who is following social-distancing norm, will try to avoid collocating in a `Surface Area` with others if there is any free surface area in a `Place`.

Contextualization and Validation

A digital twin of a city is realized by implementing meta-model shown in Fig. 3 as a set of parametric agent types, where agent types: (1) structurally conform to the meta-model relationships; (2) encapsulate necessary variables to capture states, characteristics, and historical traces; and (3) mimic business-as-usual behaviors using probabilistic event-condition-action rules. To use this digital twin for a specific city, all agents need to be contextualized and validated with respect to a specific city.

Contextualization

Contextualization of city digital twin is a process of instantiating the meta-model of Fig. 3 in terms of the requisite number of agents leading to a purposive simulatable model of the city. The model comprises demographic configuration, movement configuration, and instantiation of all agents to their representative states. Demographic configuration defines number of wards of a city, localities within each ward, locality-specific households, and commercial establishments (e.g., offices, factories, schools, shops, etc.), number of citizens in each locality, and their demographic characteristics. Movement configuration specifies movement characteristics embellished with appropriate probabilities and parameters. An agent has multiple state variables; however, infection state is of primary importance. Infection can be introduced in a locality by changing infection state of (a

(a) After 12 hours from Figure 6　　(b) After 2 months from Figure 6　　(c) After 12 hours from (b)
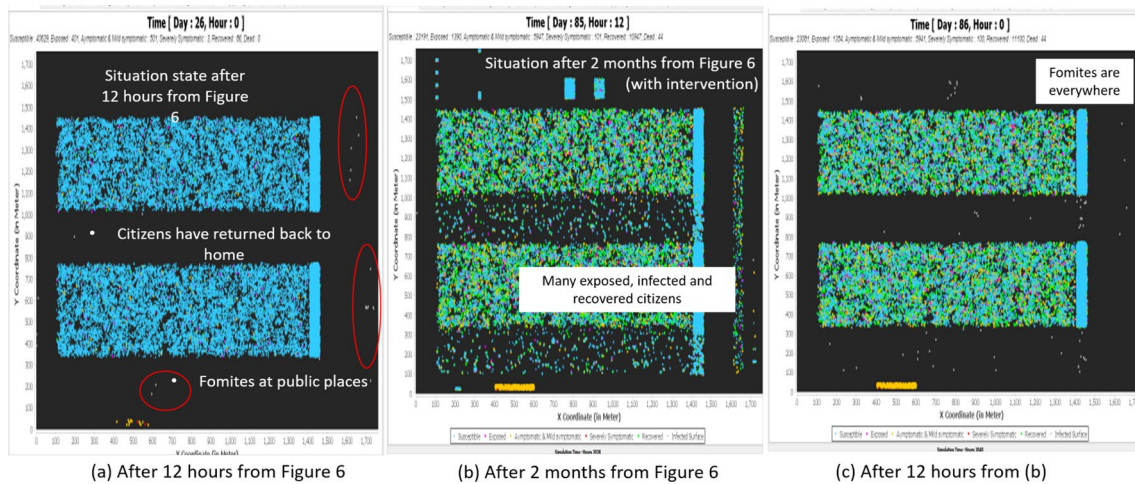
Fig. 7 Simulation dashboard for illustration

limited number of) its citizens from the default 'Susceptible' state to 'Exposed' state. We capture all configuration and instantiation parameters in spreadsheets for ease of use.

Simulation

Contextualized digital twin can be simulated. We have chosen the simulation time 'tick' to be an hour. Figure 6 shows a screenshot of the visualization environment accompanying our simulator. A ward is represented here as 1.5×1.5 km^2 block. Blue color dots denote susceptible citizens. As business-as-usual activities take them to various places thus bringing in contact with other citizen agents and infected surfaces, they get exposed (denoted by magenta colored dots) and may get infected (denoted by orange colored dots for mildly symptomatic and red colored dots for severely symptomatic) based on duration and frequency of proximal contacts, and age, gender, and comorbidities of the agents. Thus, macro-behavior of entire city in terms of infection transmission, as illustrated in Fig. 7, emerges over time from the micro-behavior of various agents.

Simulation continues for the desired time epoch several months in our case by repeating simulation tick specific number of times. Thus, simulation gives an estimate of the likely spread of infection over the time epoch along with fine-grained information such as age, gender, comorbidities, hotspots, etc. The candidate set of interventions for controlling the pandemic are experimented with by setting up suitable what-if scenarios and simulating for them using real data. Table 3 depicts the decide rata used to check effectiveness of these interventions. Simulation run generates the necessary data for computing the desired KPIs, as shown in Fig. 8.

Validation

The confidence on simulation-based estimations depends on: (1) how faithfully does the digital twin represent the city ward, (2) methodological rigor, (3) accuracy of the technology used, and (4) how close the simulation results are with respect to the reality. We use well-established techniques of simulating business- and mission-critical systems to establish faithfulness of the constructed digital twin with respect to the reality (Sargent 2013). They are:

1. Conceptual model validation that determines the theories and assumptions is reasonable for the intended purpose.
2. Computerized model validation that ensures all conceptual elements is appropriately represented using a robust implementation technology.
3. Operational validation that determines the simulation outputs is sufficiently close to reality.
4. Data validity that ensures the reliability and adequacy of inputs data.

Conceptual model validation is ensured by validating all aspect specific assumptions. Epidemiological assumptions and infection-related probabilities are introduced as per the data and results available in peer-reviewed literature. The correct interpretation of these properties is ensured by epidemiologists within our group. Demographic aspect and movement-related assumptions are ascertained using data available with city administration authority. We also took help of demographic expert within our group to validate citizen movement-related assumptions. The candidate set of interventions can be obtained (as in the case of Pune) from the city's administration authorities.

Table 3 Key performance indicators (KPIs)

KPI	Description	Illustration
SEIR graphs for ward and localities	Progression of active susceptible, exposed, active infected, cumulative recovered and cumulative death counts for ward and localities	Figure 10a
New cases of infection, recovery and death	Day wise new cases of mildly infected and asymptomatic cases, severely symptomatic, recovered and death	Figure 10d
Cumulative cases	Cumulative cases of mild, severe and death cases	Figure 10e
Load on hospitals	Active cases in hospital and ventilators	Figure 10b
Load on institutional and home quarantine	Number of citizens who are under home quarantine and intuitional quarantine.	Figure 10c
Load on testing	(1) Number of severely infected citizen tested, (2) number of mildly infected citizens tested	Figure 10f
contact-tracing	Number of citizens are traced and tested through contact-tracing	Figure 10f
Testing efficacy	% of positive cases for overall testing, household testing and contact-tracing	Not shown
Demographic distribution of infected citizens	Gender-specific distribution, age-specific distribution, medical history-specific distribution and occupational archetype-specific distribution of all infected citizens (based on cumulative numbers)	Not shown (similar to hospitalization)
Demographic distribution of hospitalized citizens	Gender-specific distribution, age-specific distribution, medical history-specific distribution and occupational archetype-specific distribution of hospitalized citizens (based on active numbers)	Figure 10j, k
Demographic distribution of death	Gender-specific distribution, age-specific distribution, medical history- specific distribution and occupational archetype-specific distribution of death persons (based on commutative numbers)	Not shown (similar to hospitalization)
Source of infections	(1) Counts of transmission place (i.e., the place from where the virus is transmitted), (2) counts of archetypes who spread the virus	Figure 10h, i
Way of transmission	How virus is transmitted, i.e., either through aerosol at household, aerosol in the commercial places, or through fomite	Figure g
Infection fatality rate	Infection fatality rate, cumulative death upon cumulative infected	Not shown
Average infected family members at households	Average % of infected family members at household for slum and well-to-do localities	Not shown
% of impacted households	% of households where at least one family member is infected in slum and well-to-do localities	Not shown

Validity of computerized model is ensured bottom–up. A city is modeled as a set of interacting agent types to specify people, place, movement, virus characteristics, and candidate set of interventions. We used a robust agent/actor technology named ESL (Clark et al. 2017) that has been used to model other complex system of systems such as a telecom company (for minimizing customer churn (Barat et al. 2020)) and a retail chain (for optimizing shop stock replenishment (Barat et al. 2019a, b)). Individual agent-type specification was first validated with domain experts for completeness, correctness, and consistency, and then tested on real data as part of the city digital twin. Thus, suitable modeling abstraction, rigorous model construction, and robust technology ensure validity of the computerized model.

We contextualize digital twin for a city, introduce an amount of infection, and compare simulation results with the real observations to ensure operational validity. We chose death count as the metric to establish operational validity as it was reliably recorded on a daily basis.

To ensure data validity, we considered data from city administration authority as input. However, we acknowledge that limited data are available on the distribution of comorbidity and disease prognostic in Indian population. High fidelity data inputs on several socio-demographic can further improve simulation prediction. We consider it as a threat to internal validity (Onwuegbuzie 2000) of our approach.

What-If Analysis

We followed an iterative human-in-the-loop simulation-based approach, as shown in Fig. 9. We simulated the contextualized digital twin for the desired time epoch. The simulation results were interpreted to arrive at key metrics such as new cases (infections), deaths, load on hospital infrastructure (number of new admissions and nature of critical care required), load on quarantine centers (number of new mildly infected patients staying in small houses or
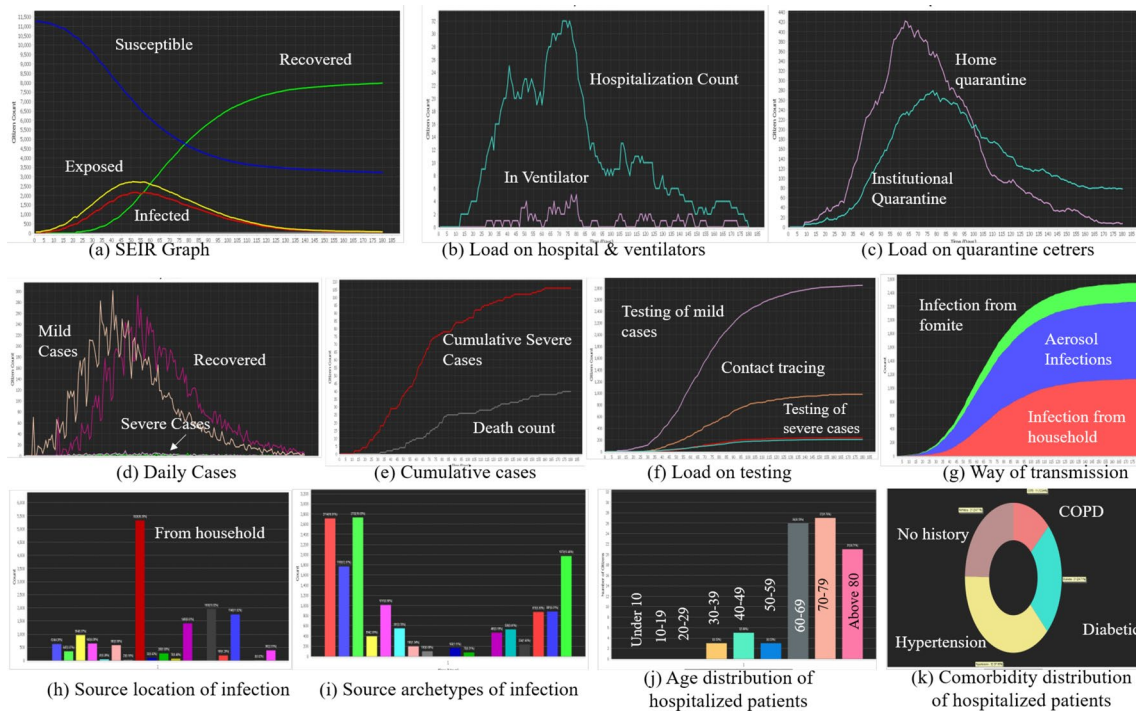
Fig. 8 Snapshots from a simulation wherein no intervention was considered
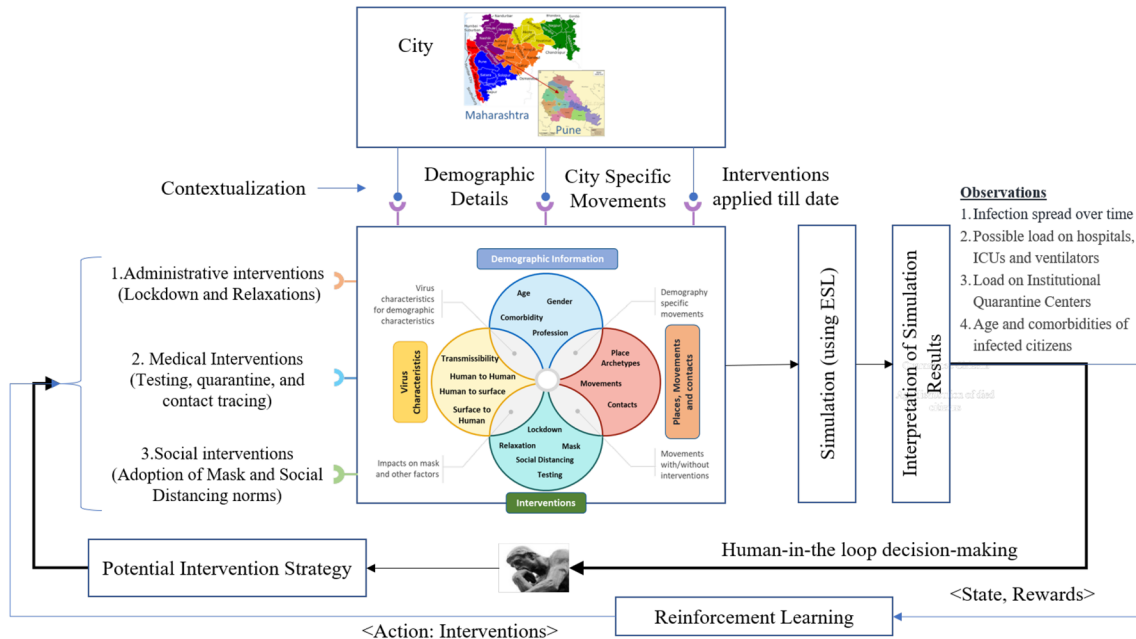

Fig. 9 Human-in-the-loop and reinforcement learning-based what-if explorations

slums), and load on testing infrastructure (whom to test and when). Akin to design of experiment, we set up appropriate what-if scenarios to ascertain efficacy of the candidate set of interventions to improve the pandemic situation in

terms of the key metrics. We tried them out one-by-one and later in combinations. The fine-grained nature of digital twin representing a city ward helped ascertain identification of ward-specific interventions. This helped us to come up with

Table 4 Area and population of few wards in Pune

Prototypical area	Ward	Total area (km²)	Well-to-do area (m²)	Slum area (m²)	Well-to-do population	Slum population	Total population	% population in slum
Residential area	Sahakar nagar	9.2	8,830,000	370,000	126,912	78,529	205,441	38.2246
	Kothrud	16.26	15,424,257	835,743	103,524	141,742	245,266	57.79113
Market Area	Bhavani Peth	2.9	2,380,802	519,198	157,936	106,851	264,787	40.35357
Office area	Nagar road	29.1	28,586,850	513,150	186,489	76,408	262,897	29.06385
	Aundh	40.75	40,262,827	487,173	268,804	72,540	341,344	21.25129

ward-specific recommendations that were backed by data. The ward-level numbers were projected onto city level to obtain the big picture.

Case Study: Pune City

We applied our digital-twin-based approach to predict and control the spread of COVID-19 pandemic for Pune city[7], a western state of Maharashtra in India. The urban area of Pune has a population around 4 million and a geographic area of approximately 330 km². The city is divided into 41 administrative zones which are referred to as wards. We classified residential localities into two: (1) slum a densely populated (typically, upwards of 60,000 population per km²) area with small houses packed together each inhabiting a large number of people, and (2) well-to-do area consisting of apartments and houses with a reduced population density (mid-1000s per km²).

Based on the composition of residential areas and commercial places, we classified Pune city wards into three prototypical wards: (1) residential ward: primarily well-to-do residential area with a small population living in slums and few busy small-size market areas (e.g., Sahakar Nagar and Kothrud); (2) market ward: primarily a densely populated busy market place with residential area comprising of well-to-do and slum localities (e.g., Bhavani Peth); and (3) office ward: primarily a business district comprising large modern office complexes, some small office buildings, shopping malls, and large modern residential complexes (e.g., Nagar Road and Aundh). Table 4 presents details of the five wards which we used for analysis.

Contextualization

Description of the three prototypical wards in terms of the four aspects constituting the digital twin follows:

1. *Epidemiological aspects:* The epidemiological characteristics, i.e., transmission and transition dynamics along with infection probabilities, of COVID-19 are same across Pune, and they are same as rest of the India/world. Hence, we consider default characteristics as discussed in Sect. 3.

2. *Demographic aspects and movements:* We consider the age, gender, and comorbidity-related heterogeneities are uniform across all three prototypical wards, as shown in Table 5. The heterogeneities of household structures and commercial places of three prototypical wards are highlighted in Tables 6 and 7.

3. *Interventions:* Pune city complies with all nationwide lockdown and unlock guidelines[8]. A stringent lockdown was imposed on March 25, 2020 across all wards in Pune along with rest of the India. All commercial places except few essential services were closed, which restricted more than 95% of the business-as-usual activities including office, school, colleges, non-essential shops, and factories. Housemaid, workers and relatives to visit any household were also restricted during that time. Different forms of lockdowns continued till May 2020, and then, relaxation/unlock started with varying movements and place closures, as illustrated in Table 8. All lockdown and unlock- based interventions in Pune are applied to all three prototypical wards to comprehend the spread of virus in different wards in Pune.

We initialized ward-specific digital twins with approximated infected citizens. We used heuristics based on reported infection counts from various wards in Pune during the second week of March to estimate initial infection counts of these prototypical wards.

We considered March 10 as the starting point of our analyses and started with 100–120 infected citizens per 100K citizens in all three prototypical wards. However, we found some of the residential wards (e.g., Kothrud) and office wards (e.g., Aundh) had a smaller number of reported cases in the second week of March as compared to other wards. We mimicked this laggard infection behavior by introducing a sub-category characterized by fewer infections (20–25 infected citizens per 100K citizens) on March 10.

---

7 https://en.wikipedia.org/wiki/Pune.

8 https://en.wikipedia.org/wiki/COVID-19_pandemic_lockdown_in_India.

Table 5 Household characteristics

Prototypical ward type	Locality type	Typical household structures	Typical household size	Visitor patterns
Residential wards	Well-to-do	Nuclear families → 1M1F (2): 10%, 1M1F1C (3): 25%, 1M1F2C (4): 25% Joint families → 1M1F1C2S (5): 10%, 1M1F2C2S (6): 10%, 2M1F2C2S (7): 10% Room Sharing → 4M (4): 5%, 4F (4): 5%	(1) 40% household size 400–800 ft^2 (2) 60% household size > 800 ft^2	Housemaid 2 times a day for an hour
	Slum	Nuclear families → 1M1F (2): 5%, 1M1F1C (3): 10%, 1M1F2C (4): 10% Joint families→ 1M1F2C1S (5): 10%, 1M1F2C2S (6): 10%, 1M1F3C2S (7): 10%, 4M2F3C1S(10): 10%, 3M2F2C1S (8): 20% Room sharing →4M (4): 10%, 4F (4): 5%	(1) 60% household size < 500 ft^2 (2) 40% household size > 500 ft^2	Relatives (once a week)—from same SES with similar distribution of profession
Market wards	Well-to-do	Nuclear families → 1M1F (2): 10%, 1M1F1C (3): 20%, 1M1F2C (4): 15% Joint families → 1M1F2C2S (6): 15%, 2M2F4C2S (8): 10%, 3M3F3C1S (10): 10%, 3M3F4C1S (11): 10% Room sharing → 4M (4): 5%, 4F (4): 5%	(1) 40% household size 400–800 ft^2 (2) 60% household size > 800 ft^2	Housemaid 2 times a day for an hour
	Slum	Nuclear families → 1M1F (2): 5%, 1M1F1C (3): 10%, 1M1F2C (4): 10%, 1M1F3C: (5) 3%, 2M1F2C (5): 7% Joint families → 2M2F4C1S (9): 15%, 3M2F1C2S (8): 10%, 3M3F4C1S (11): 20%, 3M2F2C1S (8): 10% Room sharing → 4M (4): 5%, 4F (4): 5%	(1) 60% household size < 500 ft^2 (2) 40% household size > 500 ft^2	Relatives (once a week)—from same SES with similar distribution of profession
Office wards	Well-to-do	Nuclear families→ 1M1F (2): 10%, 1M1F1C (3): 25%, 1M1F2C (4): 20% Joint families → 1M1F1C2S (5): 10%, 1M1F2C2S (6): 15%, 2M1F2C2S (7): 15% Room sharing → 4M (4): 5%, 4F (4): 5%	(1) 40% household size 400–800 ft^2 (2) 60% household size > 800 ft^2	Housemaid 2 times a day for an hour
	Slum	Nuclear families→ 1M1F (2): 5%, 1M1F1C (3): 5%, 1M1F2C(5): 10% Joint families → 1M1F2C2S (6): 5%, 1M1F3C1S (6): 5%, 2M2F4C1S (9): 10%, 2M1F2C2S (7): 10%, 3M3F4C1S (11): 10%, 3M2F2C1S (8): 10% Room sharing→ 4M (4): 5%, 4F (4): 5%	(1) 70% household size < 500 ft^2 (2) 30% household size > 500 ft^2	Relatives (once a week)—from same SES with similar distribution of profession

Table 6 Characteristics of commercial places and place-related movements

Place name	Area (ft²)	Opening times	Visiting patterns (for validation)	Prototypical ward type: residential ward	Prototypical ward type: market ward	Prototypical ward type: office ward
Government School	300–320	6 a.m. to 3 p.m. (weekdays)	Staffs: 20–25, students: 800–1200	4	4	4
Private School	1000–1200	6 a.m. to 3 p.m. (weekdays)	Staffs: 60, students 2500–3000	4	4	4
Small Shop	20–30	6 a.m. to 10 p.m. (weekdays)	Staffs: 1–2, visitors: 50–70/day	1830 (vegetable vendors: 227, grocery shop: 227, other hawkers: 328, other shops: 980, pharmacy: 68)	1560 (vegetable vendors: 194, grocery shop: 194, other hawkers: 832, other shops: 328, pharmacy: 68)	1830 (vegetable vendors: 227, grocery shop: 227, other hawkers: 328, other shops: 980, pharmacy: 68)
Market place shops	30–50	7 a.m. to 8 p.m. (all days)	Staffs: 3–5 people; visitors: 100–150/day	40	1600	71
Malls	15,000–18,000	11 a.m. to 11 p.m. all days	Staffs: 80–100; visitors-weekdays-200–300; weekend-1000–1200	1	1	1
Banks	90–110	9 a.m. to 5 p.m. weekdays	Staffs: 12–15; visitors: approx 100–120/day	40	48	7
ATM	2 to 3	24 × 7	Visitors: 50–60/day	112	84	84
Small offices	40–50	10 a.m. to 7 p.m. weekdays	Staff: 20–100	325	488	976
Big offices	150–300	10 to 7 weekdays	Staffs > 100	4	8	24
Warship Places	100–150	7 a.m. to 8 p.m. all days	Visitors: 50–200/day (mostly senior citizens)	12	12	12
Small Eatery	20–30	8 a.m. to 8 p.m. 6 days	Staff: 4–5, visitors: 100–200/day	960	960	1040
Hospital	150–170	24 × 7	Staff: 20–50, visitors: 100–300/day	8	8	12
Clinic/lab	20–25	10 a.m. to 2 p.m. and 6 p.m. to 9 p.m. 6 days	Staff: 2; visitor: 50–70/day	96	48	48
Factories	80–300	9 a.m. to 5 p.m.	Staffs > 100	Small: 30, big: 1	Small: 15, big: 1	Small: 8, big: 2
Bus	30	6 a.m. to 10 p.m. (all days)	Capacity: 40	32	32	40
Shared cab	5	24 × 7	Capacity: 4	1600	1600	2000
Barber shop/beautician	30–50	7 a.m. to 8 p.m. all days	Staff: 1–2, visitors: 30–50/day	120	120	120

Table 7 Demographic distribution

Citizen archetypes	Age range (years)	Gender distribution
Beautician	20–60	Male: 70%, female: 30%
College student	16–25	Male: 60%, female: 40%
Daily wage worker	18–55	Male: 80%, female: 20%
Driver	18–60	Male: 100%
Health Worker	18–60	Male: 40%, female: 60%
House maid	18–60	Female 100%
House wife	20–60	Female 100%
Market place staff	18–60	Male: 70%, female: 30%
Office-goer	18–60	Male: 70%, female: 30%
Restaurant staff	18–60	Male: 70%, female: 30%
School kid	5–16	Male: 50%, female: 50%
Senior citizen	60–90	Male: 50%, female: 50%
Small shop Staff	20–60	Male: 80%, female: 20%
Average	30–31	Male: 50%, female: 50%
Medical history distributions		
School and college kid medical	No medical history: 100%, hypertensive: 0%, diabetic: 0%, COPD: 0%	
Senior citizen medical	No medical history: 44%, hypertensive: 40%, diabetic: 20%, COPD: 6%	
Other archetype	No medical history: 62%, hypertensive: 25%, diabetic: 10%, COPD: 3%	

Validation

The city administrative organization, i.e., Pune Municipal Corporation (PMC), made available pandemic-related data for each ward on a daily basis. We used these data, augmented suitably in consultation with demography and epidemiology experts, to initialize ward-specific digital twin and to validate the simulation results. In particular, the conceptual validity of the assumptions related to the number of places and citizen archetypes are ensured by corroborating ward-specific data available with PMC. The closeness of archetype movements with/without interventions is validated by local demographic experts. We provided a graphical visualization aid of archetype movements, as shown in Fig. 10, to ensure the validity of movement-related assumptions. In cognizant with the data validity step, we considered official and reliable data sources to contextualize Pune specific digital twin. For example, data shown in Table 4 are taken from PMC site, and data shown in Tables 5, 6 and 7 are derived from PMC data. For example, the population size and number of households at slum and well-to-do localities in respective wards are derived from PMC city sanitation plan, 2012, and Table 8 is constructed based on the actual interventions that are applied to Pune. As part of operational validity, we compared death count of Pune, as shown in Fig. 11. In addition to city-level death count, we also compared death counts of all 5 wards that we considered in our analysis.

Early Explorations

A nationwide lockdown was imposed from March 25, 2020 in India, and thereafter, the government of India has issued a series of advisories on possible interventions for states and cities to follow. However, it was observed that a common set of interventions have resulted into different impact on different part of the state and city as the spread of virus, fatality rate, and load on hospitals depend on the local situation of the epidemic. The local administration had to carefully weigh the epidemic situation on the ground and the socio-economic advantages of opening up against the burden on hospital load. Our digital twin contextualized for Pune city allowed us to analyze different intervention strategies such as restricted/staggered timings of offices/shops/malls, intermittent lockdown for a shorter period, offices/shops/restaurants with limited occupancy, etc., supplemented with necessary testing uptake. During the lockdown phase in Pune (i.e., March 25 to May 31), the main objective was to limit the spread of the virus so as to minimize the load on hospital infrastructure and to arrive at accurate enough estimate of number of ICU beds and ventilators. The key focus was to identify and isolate mildly infected and asymptomatic citizens that can be safely isolated either at home or quarantine centers.

In the light of limited capacity for testing (around 700–1000 per day) and inadequate number of quarantine centers in the early phase of pandemic, we focused on identifying locality-specific time-bound interventions so as to utilize the available public health infrastructure judiciously.

Table 8 Definition of intervention strategies and timelines

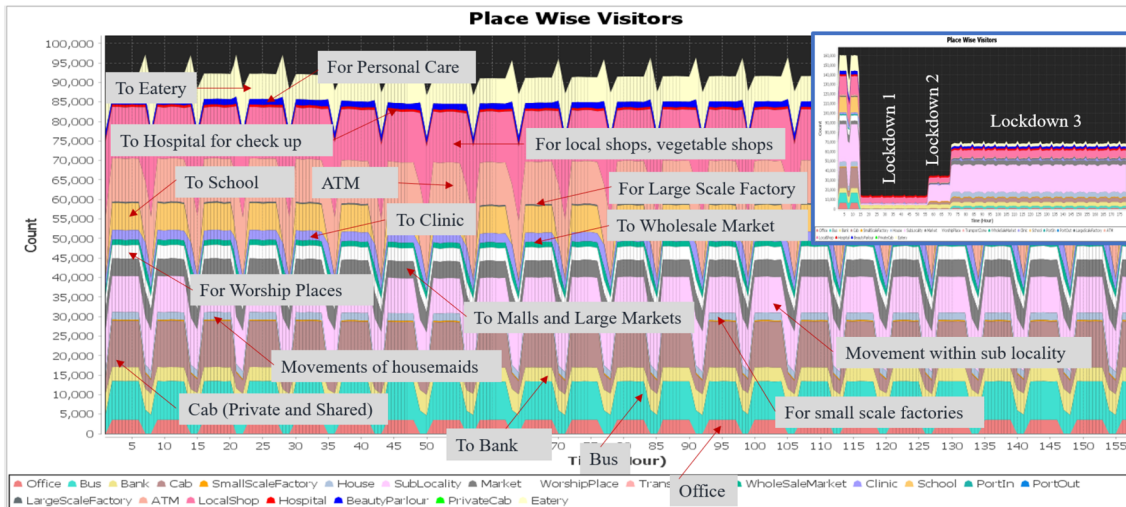| Intervention strategy	Lockdown 1	Lockdown 2	Lockdown 3	Unlock 1	10-day lockdown	Unlock 2	Unlock 3	Unlock 4
Effective date	March 25 to May 3, 2020 (Phase 1 and 2)	May 4 to May 17, 2020 (Phase 3)	May 18 to May 31, 2020 (Phase 4)	June 1 to July 12, 2020	July 13 to July 23, 2020	July 24 to August 2, 2020	August 3 to September 5, 2020	September 6, 2020 present
Administrative interventions								
Public movements	10 a.m. to 2 p.m. for essential work	10 a.m. to 7 p.m. for essential work	7 a.m. to 7 p.m. for essential work	7 a.m. to 7 p.m. for essential work	10 a.m. to 2 p.m. for essential work	7 a.m. to 7 p.m. for essential work	No time restriction	
Essential shops (e.g., grocery)	1	1	1	1	1	1	1	1
Bank and ATM	1	1	1	1	1	1	1	1
Clinics	0.3	0.3	0.5	0.7	0.3	0.7	1	1
Barber and beautician places	Closed	Closed	0.2	0.2	Closed	0.5	0.5	0.5
Local non-essential shops	0	0.05	0.3	0.5	0	0.85	0.85	1
Non-essential shops located in market places	0	0	0.3	0.5	0	0.85	0.85	1
Shops at mall	Closed							
Housemaid	Not allowed	Not allowed	0.5	0.5	Not allowed	1	1	1
Factories	Closed	0.05	0.5	0.5	Closed	1	1	1
Small offices	10% offices with 50% staffs	15% offices with 50% staffs	25% offices with 50% staffs	25% offices with 80% staffs	10% offices with 50% staffs	50% offices with 80% staffs	85% offices with 80% staffs	85% offices with 80% staffs
Big offices	Closed	Closed	25% with 10% staffs	25% with 20% staffs	Closed	50% with 10–20% staffs	60% with 10–20% staffs	60% with 10–20% staffs
School and Colleges	Closed							
Public cabs	Not allowed	10% with driver + one person	100% with Driver + one person	100% with driver + one person	Not allowed	1	1	1
Bus	Not allowed	Not allowed	20% with 50% occupancy	40% with 50% occupancy	Not allowed	60% with 50% occupancy	75% with 50% occupancy	75% with 50% occupancy
Eateries	Closed						20% with 50% occupancy	20% with 50% occupancy
Testing related interventions								
Severely infected	1	1	1	1	1	1	1	1
Mildly infected	0	0.05	0.05	0.05	0.05	0.05	0.05	0.05
contact-tracing	0.05	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Household contacts	0	1	1	1	1	1	1	1

Fig. 10 Archetype-specific business-as-usual movement, in the absence of interventions

Fig. 11 Simulated and actual death count for validation
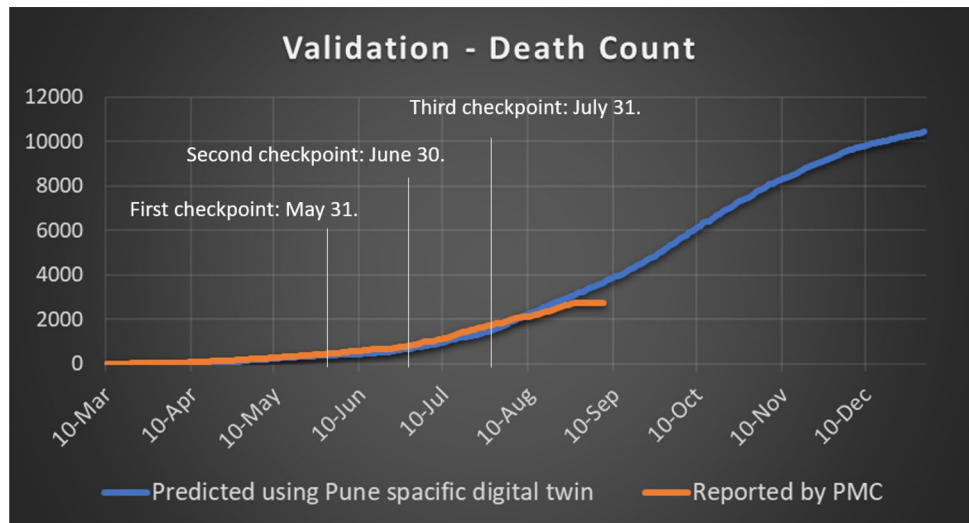
Some of the interventions explored were: (1) isolating localities having higher number of infections, (2) increasing contact-tracing and household testing in localities having higher number of infections, and (3) enforcing intuitional quarantine for mildly infected patients staying in slum areas. We discuss some of the early what-if scenarios played out using the city digital twin.

Efficacy of Initial Lockdown

First, we wanted to ascertain whether the strict lockdown imposed on 25th March was indeed necessary or not. Figure 12 illustrates possible virus spread with/without interventions using SEIR graphs. Figure 12a shows the virus spread had there been no strict lockdown. Figure 12b shows effect of the first lockdown that was imposed on March 25,

and Fig. 12c depicts the consequence of partial relaxations applied on May 4.

Efficacy of Increased Testing

As shown in Fig. 12, the predicted impact for allowing essential movements as part of lockdown 2 (see Table 8) was threefold increment of the infection situation at that point of time. Therefore, we analyzed how limited testing capacity can be effectively utilized to control such surge of infection count. The impact of increased testing for three prototypical wards is shown in Fig. 13. We also evaluated, active infection, and need for hospitalization of well-to-do and slum areas of all three prototypical wards under three different testing uptakes, as shown in Fig. 14.
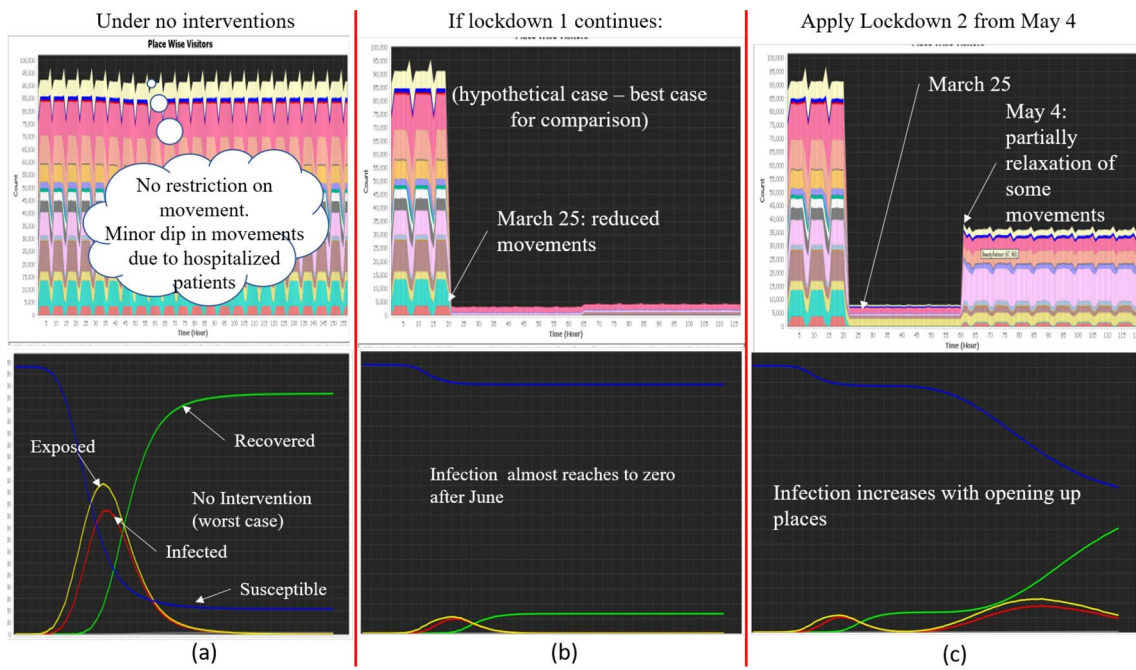
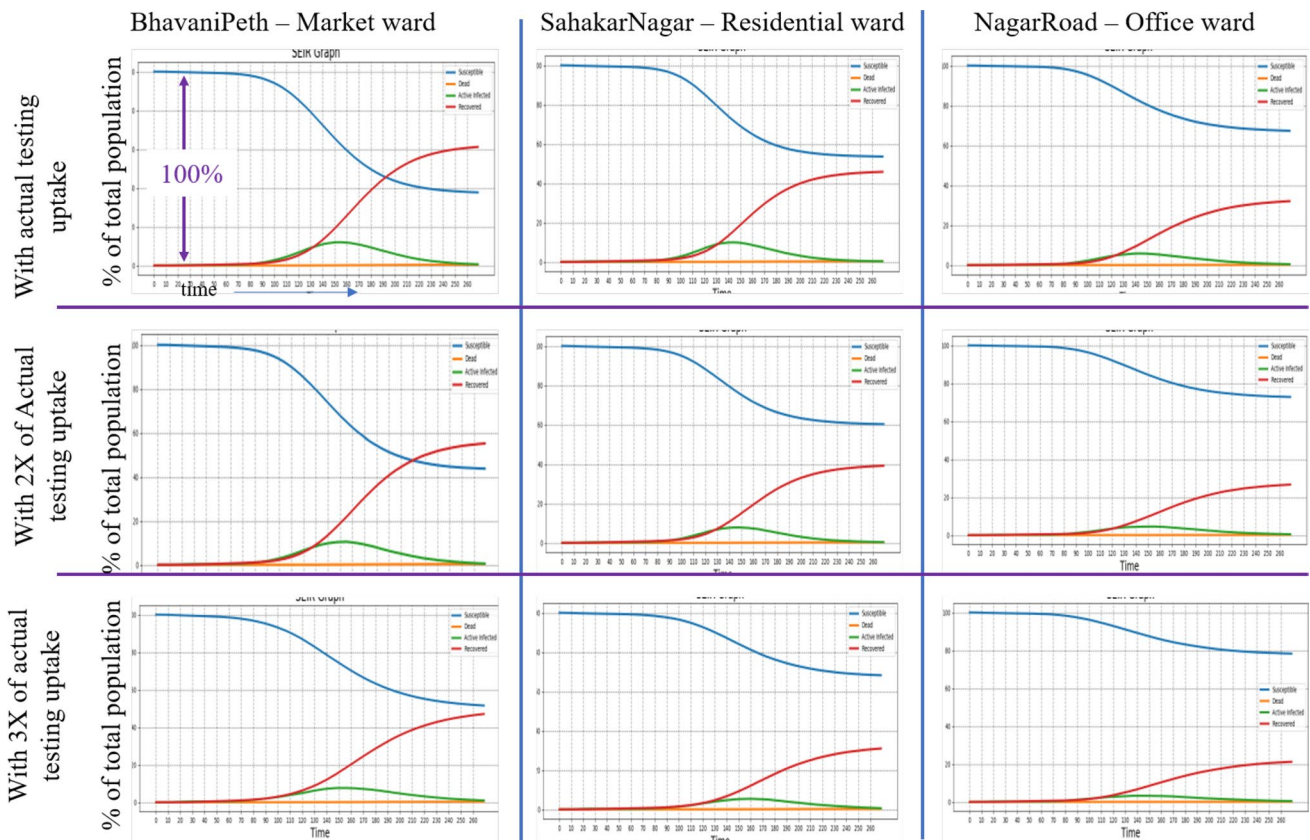**Fig. 12** Simulation under initial lockdown conditions

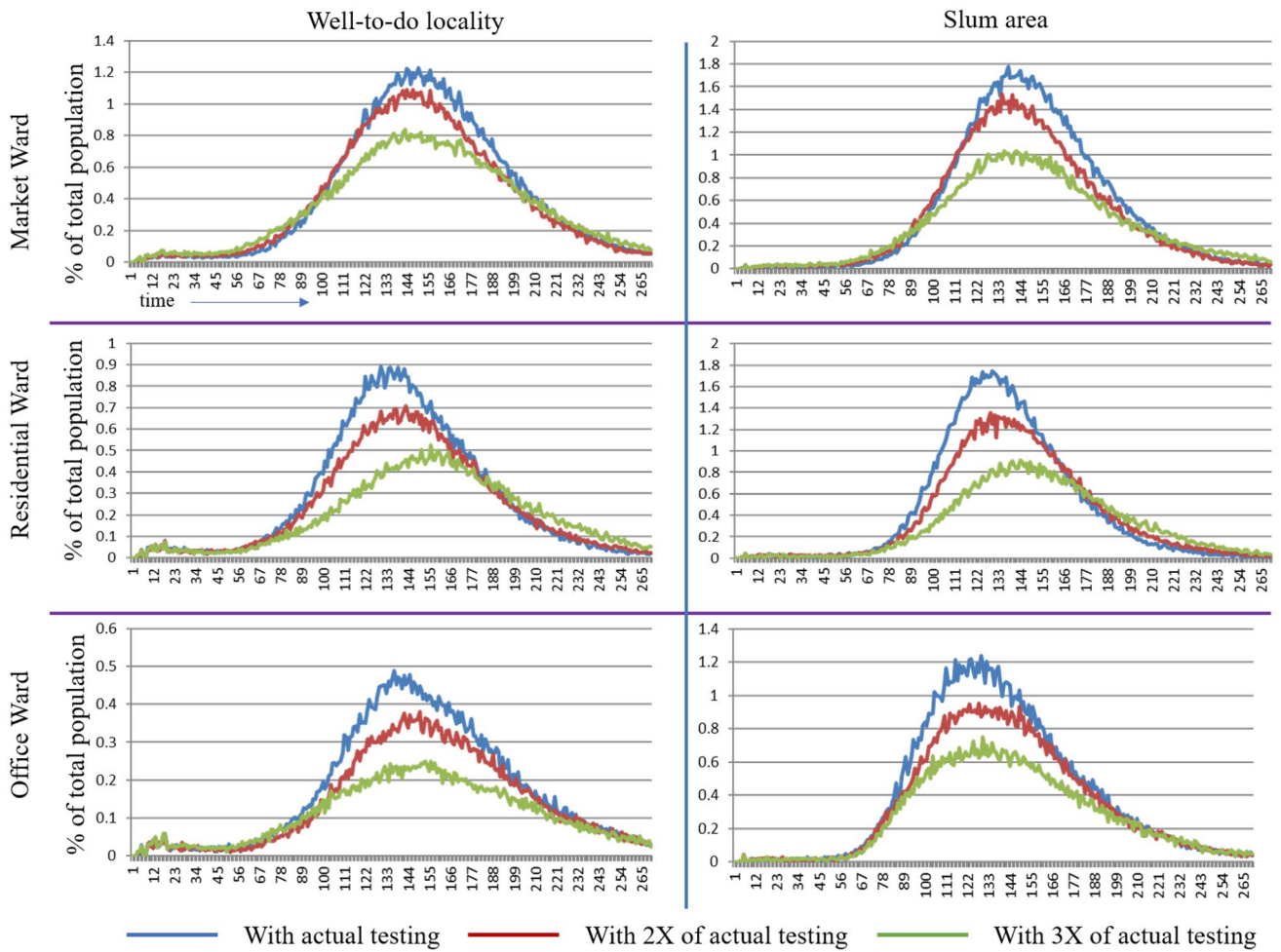**Fig. 13** Ward-specific SEIR graphs illustrating the impact of increased testing

Fig. 14 Impact of increased testing uptake on well-to-do and slum areas of three prototypical wards
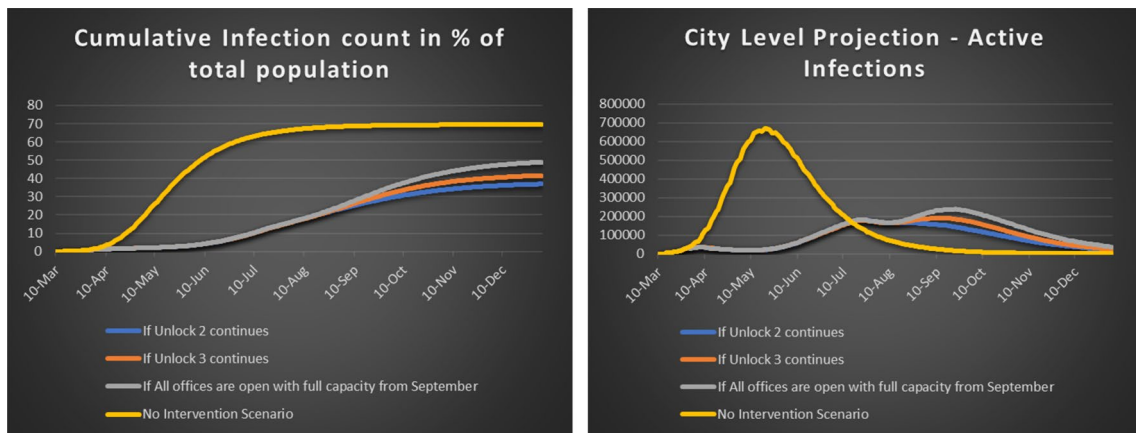

Fig. 15 Impact of infections under different interventions

Simulation results indicate that increased testing is advantageous but is not linear to infection spread. Increased testing in densely populated wards (e.g., BhavaniPeth) during

the early phase of epidemic showed far greater benefit compared to low population density wards such as NagarRoad. Moreover, increased testing beyond certain degree is shown
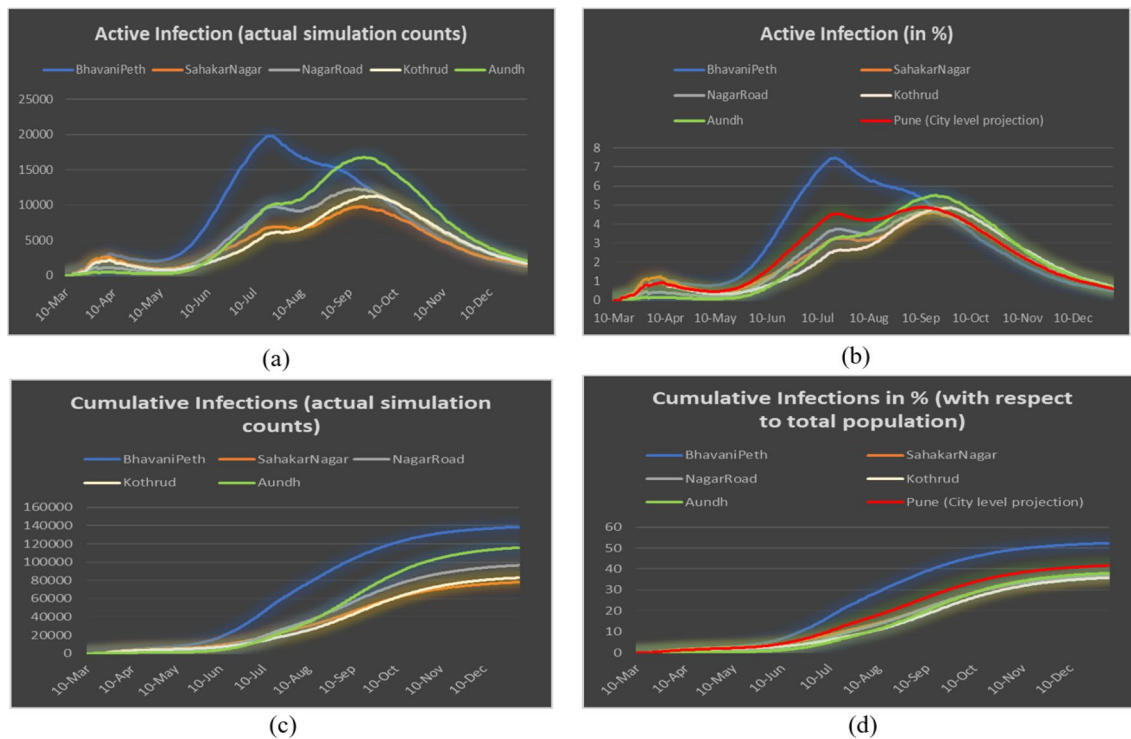
Fig. 16 Infection situation in different prototypical wards under Unlock 3.0 intervention

diminishing advantage. Therefore, the recommendation was to increase testing uptake in densely populated slum area significantly as compared to less density area.

Explorations for Unlock

In India, the unlock process, Unlock 1.0, was started from June 1. The second phase of unlock, Unlock 2.0, lasted 1–31 July, and Unlock 3.0, Unlock 4.0, and Unlock 5.0 were put into effect in August, September, and October, respectively. However, Pune has seen restrictive/muted unlock as it was one of the top three cities in India with respect to number of infection. As shown in Table 8, Unlock 1.0 with limited opening of commercial places continued till July 12, a 10-day strict lockdown was imposed between July 13 to July 23, and then, phased unlocks were imposed with adequate testing and contract tracing. Prior to implementing these unlock strategies, several candidate interventions were evaluated and their efficacies with respect to death counts, peak load on critical health care (oxygen beds, ICU, ventilators, etc.) and load in institutional quarantine centers were predicted through what-if simulations on Pune city digital twin. As an illustration, the cumulative count and active count of infection under four different interventions are shown in Fig. 15.

Our analysis shows that different wards reach peak at different stages of epidemic. As shown in Fig. 16, wards with higher population density, like BhavaniPeth, reach peak as early as mid-July, whereas other wards see the peak in September and October. We also see from Fig. 16b that two residential wards, namely SahakarNagar and Kothrud, with more or less identical characteristics exhibit similar active infection pattern, but reach peak at different times as they were in different state in terms of initial infection when first lockdown started in March 25. City-level estimation of active infection, shown using red color line in Fig. 16b, exhibits quite a different pattern from the constituent wards. This justifies our claim that city-level situation is an aggregation of ward-specific situations, and hence, analysis of city-level situation may lead to interventions that may not be effective at ward level. The trends on cumulative infections, shown in Fig. 16c, d, show that around 35–40% of crowded dense market ward population is infected by mid-August. In comparison, only 10–15% of population in residential wards (e.g., Kothrud and Aundh) is infected by mid-August. The model predicted that with the sequence of interventions, the peak load on critical health care (oxygen beds, ICU, ventilators, etc.) remain sustained until end-October[9], as shown in Fig. 17a.

---

[9] https://indianexpress.com/article/cities/pune/pune-peak-load-on-critical-healthcare-to-be-sustained-until-oct-end-6562008.
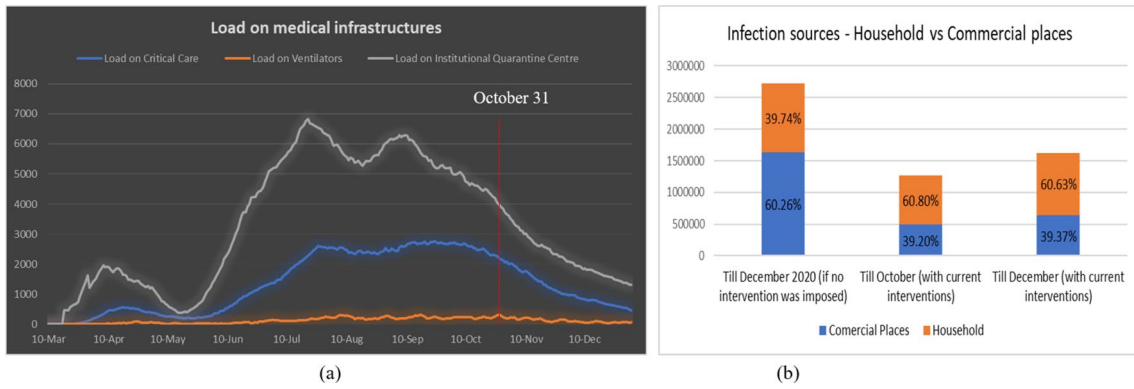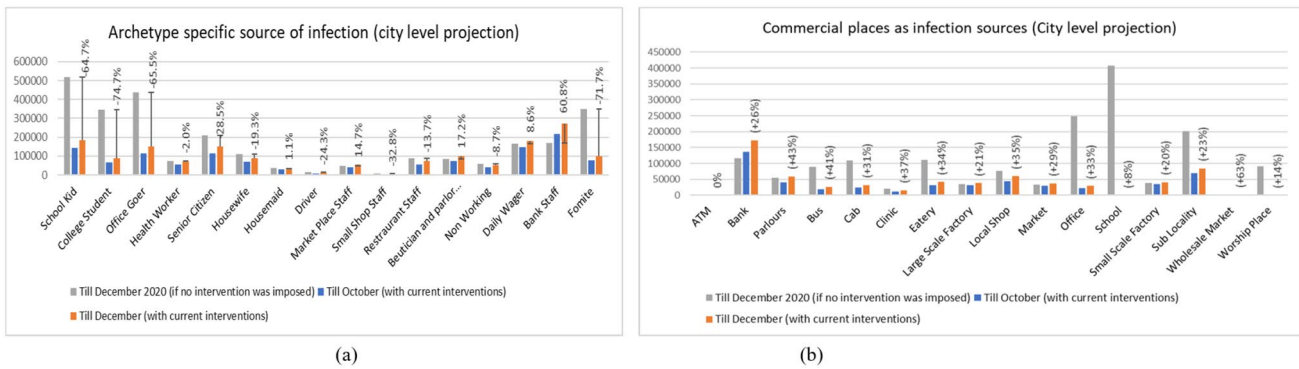
Fig. 17 Analysis outcome

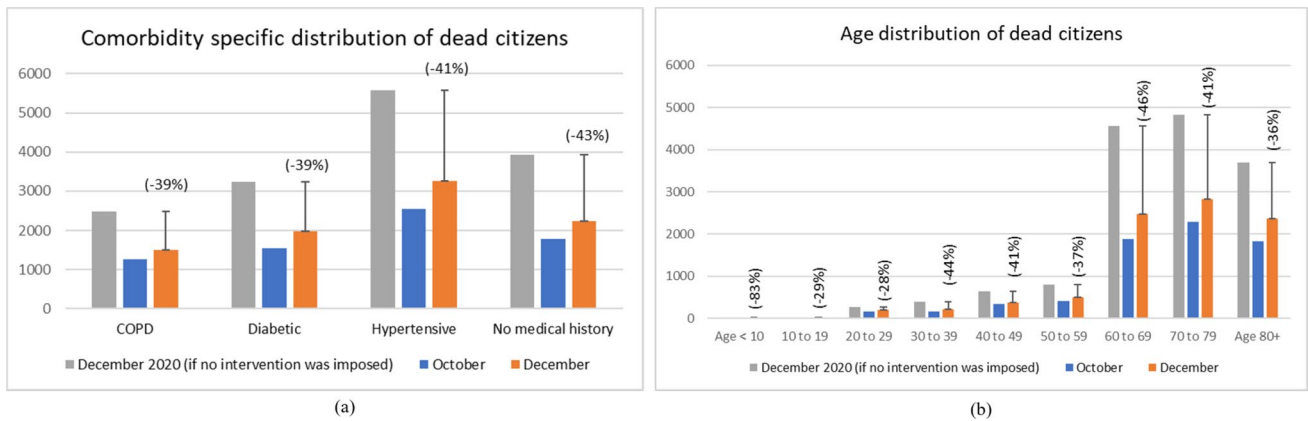**Fig. 18** Citizen archetype and place-specific analysis

Fig. 19 Analysis of comorbidity and age

Further analysis shows that majority of the infections, as high as 60%, are being seen to occur through household contacts. A comparison of virus spread in household and commercial places under no intervention and sequence of interventions till Unlock 3.0, shown in Fig. 17b, indicates that the virus spread at outside household is significantly controlled through the series of interventions.

Analysis of citizen archetypes and commercial places as sources of infection spread, shown in Fig. 18a, b, respectively, indicates that long-duration interactions in enclosed and crowded places are likely to increase transmission risk. Correlation of Fig. 18a, b indicates that complete closure of schools and colleges and offices operating with reduced staffs helped to reduce the spread of virus by
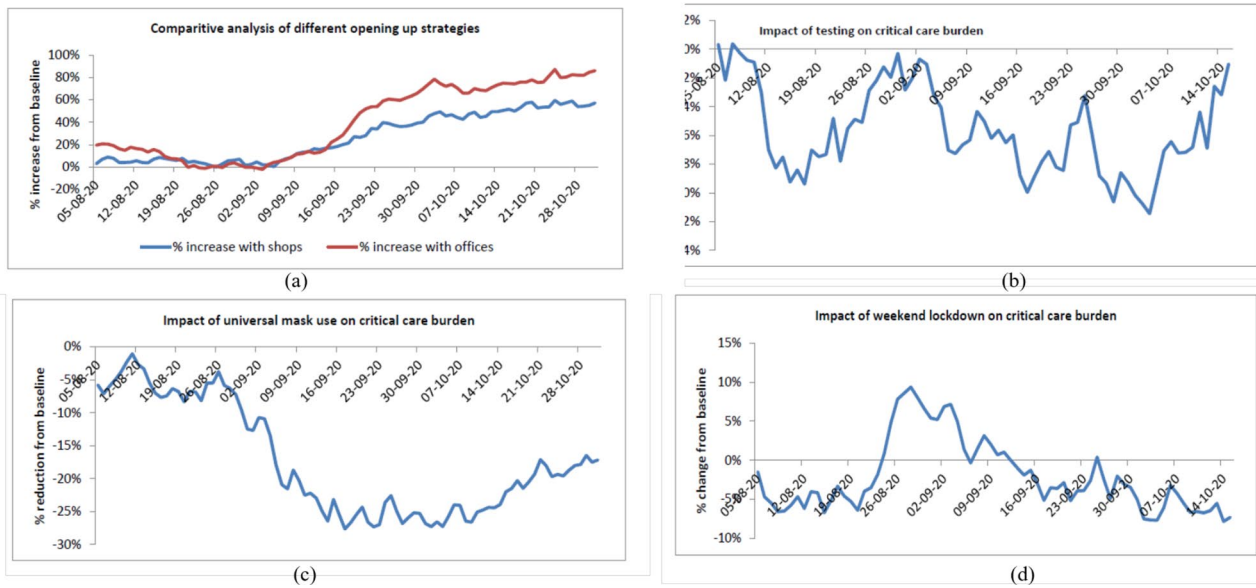
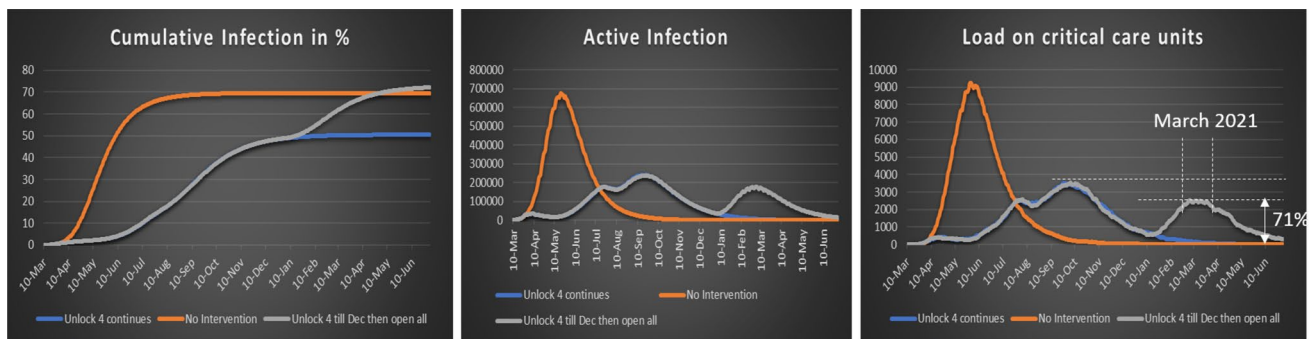**Fig. 20** Summary of experimental outcomes

Fig. 21 A hypothetical experimental result of opening up places from January 2021

respective archetypes (shown using error bars in Fig. 18a). Those closures also helped to reduce fomite-related infections significantly. The orange bar of Fig. 18b indicates place-specific propensity of spreading virus. It indicates bank, parlour, bus, office, and crowded market which has greater propensity of virus spread (i.e., additional precautions need to be taken while visiting those places). School can be also categorized in similar characteristics when they are open. An analysis on comorbidity and age distribution on death, shown in Fig. 19, indicate the impact of lockdowns and unlocks on different age groups and individuals with comorbidities. A representative sample of specific interventions explored through what-if scenario playing are listed below:

1. *Opening up:* We undertook comparative analysis to assess the impact of opening up different places with load on hospitals for critical care admissions as the primary metric. The analysis showed that opening of offices is more detrimental than opening of shops, as shown in Fig. 20a.

2. *Weekend lockdown:* Weekend lockdown (allowing only essential services and imposing restriction on all other movements on weekend) has limited efficacy in reducing the burden, as shown in Fig. 20b. This efficacy will be further reduced if there is overcrowding during the weekdays.

3. *Universal use of masks:* Behavioral intervention as in use of mask was found significantly effective as compared to no use of mask. Our simulations showed that load on hospital for critical care admissions reduced up

to 25% when if 80% of population wore masks in public places, refer Fig. 20c.

4. *Testing:* Testing helped in slowing down the epidemic. The model results show that doubling the current testing rate is likely to reduce the hospital load up to 10% during peak months. Screening of symptomatic cases was found to be a better strategy compared to increasing the contact-tracing efforts among non-household contacts. However, increased testing also increased the burden on institutional quarantining facilities and home isolation by 35–40% (refer Fig. 20d).

5. *A hypothetical experimentation:* Simulation results of a hypothetical experimentation of continuing Unlock 4.0 till December 2020 and then opening up everything are shown in Fig. 21. The number of infections shoots up in mid-Jan climbing rapidly till mid-Apr before tapering somewhat. There is commensurate surge in critical care admissions to hospitals that peaks from mid-Mar to mid-Apr. This indicates that a phased and carefully planned approach to opening up of schools, offices, and other public places is necessary even when the present infection level is negligible (see active infection level in December 2020, as shown in Fig. 21b).

Discussion

The present model provides information that would be useful to plan public health measures as well as prepare health care systems for impending case load. It provides a ward-level analysis of the unfolding situation in the city. As observed for Pune city, different wards (and sub-localities within wards) achieved peaks at different time points, owing to differences in demographic profile, movement dynamics, and infection levels during early months of epidemic. The ward-level picture can help in designing context-specific, flexible, and responsive public health strategies.

The model provides information on total number of active cases (both mild/moderate and severe cases) over a period of time. The insight becomes useful for planning to cope with additional health care demands—such as additional need for isolation beds, oxygen beds, Intensive Care Units, and ventilators. When juxtaposed to the number of detected cases, it gives a fair idea of how many actual infections are being missed. It can help set realistic targets about how much testing needs to be expanded. The model can also provide outputs on positivity rates in household and non-household contacts across different wards/localities, and places where majority transmissions occur. Such granular information can help in better targeting (in terms of who to test, where to test) and optimal utilization of available resources.

The model shows that mask use and increased testing substantially reduce the burden on health care. These findings are consistent with other prediction studies from India

and elsewhere. The efficacy of these interventions in the present analysis assumed a high level of compliance. However, in real life, such high levels may be difficult to achieve. The model has the flexibility to not only consider different levels of testing uptake and mask use, but further complexities seen in real life. For example, one can set rules about who wears masks, how many people wear masks, and at what places people wear/not wear masks. Similarly, different scenarios such as early vs late testing, different levels of testing uptake, and compliance levels of isolation of positive cases can be simulated in the model. Such detailed analysis can give a more realistic picture of possible benefits of NPI. It can help in advocating for the need for awareness and behavioral change interventions, which is possibly the most neglected aspect of COVID-19 responses in India.

Evaluation

Our city digital twin for COVID-19 is best seen as an aid to explore the efficacy and impact of candidate set of interventions for controlling the pandemic. The simulatable nature of city digital twin makes it amenable to what-if and if-what scenario playing thus leading to data-driven justification-backed decision-making. City administration found our predictions and analyses leading to control interventions useful. Our ward-wise (city-wide) prediction of COVID-19 deaths, which is computed by corelating prototypical wards with actual wards and aggregating the death counts observed from prototypical wards, matches closely with the reality. The state of epidemic in Pune in terms of deaths, infections, and load on hospital in the months of September and October is unfolding the way which we predicted in August[10]. Our analyses of household infection and infections at other places correlate with Indian evidence on transmission dynamics of COVID-19 (Laxminarayan et al. 2020). Such close resemblance with reality on multiple dimensions of COVID-19 epidemic establishes the efficacy of our fine-grained model to predict and control COVID-19 pandemic in Pune city. Given the parameterized nature of the model, we think that it can be easily configured to other comparable cities.

Useful Design Choices

Given the limited data and uncertain dynamics of the overall system, we had to make a set of well-founded choices to ensure that city digital twin is an accurate representation of ground reality. The key choices are discussed below:

---

[10] https://indianexpress.com/article/cities/pune/pune-peak-load-on-critical-healthcare-to-be-sustained-until-oct-end-6562008/.

1. *System theoretic view:* System theoretic view to conceptualize a city as a complex system of systems distinguishes our approach from popular compartmental model-based approaches. Our focus on capturing the existing heterogeneity in people, places, transportation, and health infrastructure, etc., so as to enable specification of individualistic micro-behavior led us to choose agent-based modeling paradigm. We extended the canonical agent-based modeling paradigm to help specify probabilistic and spatio-temporal behavior of individual agents. As a result, the emergent behavior of the complex system of system faithfully reflected the reality. We believe that these extensions stand out from prevalent agent-based approaches.

2. *Aspects of interest:* Our assumption to capture a city around four interrelated aspects, namely: epidemiological aspect, demographic aspect, movement aspect, and intervention aspect, turned out to be necessary and sufficient for prediction and control of virus spread. We believe that the separation of concerns and making dependencies therein explicit help manage the complexity of virus spread dynamics. The extended agent-based modeling technology is rich to specify these aspects and their relationships at the desired level of granularity, thus preventing oversimplification of the inherent complexity.

3. *Citizen and place archetypes:* The other key factor that helps us to differentiate from other agent-based approaches is the conceptualization of an extensive list of agent types to capture necessary heterogenous characteristics of citizens and places. In our digital twin, an agent type encodes specific characterization in terms of demography (e.g., age and gender), profession (e.g., office-goers, Students, senior Citizen, housewife, healthcare professional, driver, housemaid, market staff and shop owners), place (e.g., apartment, slum, shop, office, clinic, hospital, and grocery store), transportation (e.g., bus, car and cab), health condition (e.g., diabetes, hypertension, and chronic pulmonary disease), and so on. In particular, an agent type specifies business-as-usual behavior of a well-demarcated and well-defined subset of the heterogeneous population. Therefore, a collection of key representative agent types together can characterize an entire population of a city.

4. *Proven methodology and technology:* Finally, the use of proven model construction and validation methodology augmented with the ESL technology that has been found effective for analyzing complex business systems have together led to a robust and pragmatic approach.

Limitations

Every model has its limitations and this model is no exception. Like all other models, this model too relies on a set of assumptions; inaccurate assumptions may reduce the accuracy of the model prediction significantly. Some model parameters have significant uncertainty. Adverse impact of the uncertainty is contained to the extent possible by taking due precaution in analysis. Still, the input data should not be grossly incorrect. From technology perspective, the scalability of agent-based is an inherent limitation. At present, ESL technology infrastructure is capable of supporting a few million agents. Augmented with use of stratified sampling, the technology has catered to predict and control COVID-19 epidemic in large cities. However, scaling it to be able to cater to state and country level is not possible with standard computing power.

Future Directions

As our understanding of COVID-19 improves, we plan to incorporate the learnings from other ongoing COVID-19 initiatives. It is relatively easy to introduce new archetypes of citizens and places. This extension should also help to configure the digital twin for other comparable cities. Digital twin can be extended to cater to reinfection, mutated virus, and vaccine. How to strike a balance between control of epidemic and revival of economy is an important next step which we are considering.

Concluding Remarks

We presented a novel agent-based digital twin of a city to support simulation-based approach to predict and control COVID-19 epidemic. The defining characteristic of the city digital twin is a set of suitable agent types necessary to capture heterogeneity in terms of people, places, transport infrastructure, health care infrastructure, etc. As a result, we are able to construct a fine-grained model of the city that is amenable to what-if and if-what scenario playing. We populated the city digital twin using data from the city administration, together with suitable augmentation. The fine-grained nature of digital twin enabled us to address the critical concerns such as the rate and the extent of the spread of the epidemic, demographic, and comorbidity characteristics of the infected people, load on the healthcare infrastructure in terms of specific needs such as number of admissions requiring critical care (supplementary oxygen, ventilator support, intensive care, etc.), load on institutional quarantine centers, and so on. We set up appropriate what-if scenarios to identify the most effective intervention from the candidate set to control

epidemic as well as bring back normalcy. We vetted the simulation results against epidemic-related data released by city administration daily. Though the approach has been validated in the specific context of Pune city, we believe that the parameterized nature of our digital twin makes the approach applicable for other comparable cities with minor effort. We expect that the framework detailed in this paper can serve as a template for developing simulators as aids to decision-making should such epidemics occur in the future.

Compliance with Ethical Standards

Conflict of interest The authors declare that they do not have any conflict of interest.

References

Agha GA, Mason IA, Smith SF, Talcott CL (1997) A foundation for actor computation. J Funct Programm 7(1):1–72

Agrawal S, Bhandari S, Bhattacharjee A, Deo A, Dixit NM, Harsha P et al (2020) City-scale agent-based simulators for the study of nonpharmaceutical interventions in the context of the Covid-19 epidemic. J Indian Inst Sci 1–39

Aleta A, Martin-Corral D, y Piontti AP, Ajelli M, Litvinova M, Chinazzi M et al (2020) Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the Covid-19 epidemic. medRxiv

Asadi S, Bouvier N, Wexler AS, Ristenpart WD (2020) The coronavirus pandemic and aerosols: does Covid-19 transmit via expiratory particles?. Taylor & Francis, Routledge

Bailey NT et al (1975) The mathematical theory of infectious diseases and its applications. Charles Griffin & Company Ltd, High Wycombe

Barat S (2018) Actor based behavioural simulation as an aid for organisational decision making (Unpublished doctoral dissertation). Middlesex University

Barat S, Khadilkar H, Meisheri H, Kulkarni V, Baniwal V, Kumar P, Gajrani M (2019) Actor based simulation for closed loop control of supply chain using reinforcement learning. In: Proceedings of the 18th international conference on autonomous agents and multiagent systems, pp 1802–1804

Barat S, Kulkarni V, Clark T, Barn B (2017) An actor-model based bottomup simulationan experiment on Indian demonetisation initiative. In: 2017 winter simulation conference (WSC), pp 860–871

Barat S, Kulkarni V, Clark T, Barn B (2019) An actor based simulation driven digital twin for analyzing complex business systems. In: 2019 winter simulation conference (WSC), pp 157–168

Barat S, Kulkarni V, Kumar P, Bhattacharya K, Natarajan S, Viswanathan S (2020) Towards effective design and adaptation of CSP using modelling and simulation based digital twin approach. In: Proceedings of the 2020 summer simulation conference, pp 1–12

Boardman J, Sauser B (2006) System of systems-the meaning of. In: 2006 IEEE/SMC international conference on system of systems engineering, p 6

Cai J, Sun W, Huang J, Gamber M, Wu J, He G (2020) Indirect virus transmission in cluster of Covid-19 cases, Wenzhou, China, 2020. Emerg Infect Dis 26(6):1343–1345

Chang SL, Harding N, Zachreson C, Cliff O M, Prokopenko M (2020) Modelling transmission and control of the Covid-19 pandemic in Australia. arXiv:2003.10218

Chao DL, Oron AP, Srikrishna D, Famulare M (2020) Modeling layered non-pharmaceutical interventions against sars-cov-2 in the United States with corvid. medRxiv

Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395(10223):507–513

Chen Y, Li L (2020) Sars-cov-2: virus dynamics and host response. Lancet Infect Dis 20(5):515–516

Chretien J-P, Riley S, George DB (2015) Mathematical modeling of the west Africa Ebola epidemic. Elife 4:e09186

Clark T, Kulkarni V, Barat S, Barn B (2017) ESL: an actor-based platform for developing emergent behaviour organisation simulations. In: International conference on practical applications of agents and multi-agent systems, pp 311–315

COVID I, Murray CJ, et al (2020) Forecasting the impact of the first wave of the covid-19 pandemic on hospital demand and deaths for the USA and European economic area countries. medRxiv

Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M et al (2020) Report 9: impact of non-pharmaceutical interventions (NPIS) to reduce covid19 mortality and healthcare demand

Ferguson NM, Cummings DA, Fraser C, Cajka JC, Cooley PC, Burke DS (2006) Strategies for mitigating an influenza pandemic. Nature 442(7101):448–452

Fernandes N (2020) Economic effects of coronavirus outbreak (Covid-19) on the world economy. Available at SSRN 3557504

Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, Colaneri M (2020) Modelling the Covid-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 1–6

Grieves M, Vickers J (2017) Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. In: Transdisciplinary perspectives on complex systems. Springer, pp 85–113

Guan W-J, Liang W-H, Zhao Y, Liang H-R, Chen Z-S, Li Y-M et al (2020) Comorbidity and its impact on 1590 patients with Covid-19 in China: a nationwide analysis. Eur Respir J 55(5)

Hallett TB, Menzies NA, Revill P, Keebler D, Borquez A, McRobie E, Eaton JW (2014) Using modeling to inform international guidelines for antiretroviral treatment. Aids 28:S1–S4

Harsha P, Juneja S, Patil P, Rathod N, Saptharishi R, Sarath A, Vaidhiyan NK (2020) Covid-19 epidemic study II: phased emergence from the lockdown in Mumbai. arXiv:2006.03375

He S, Peng Y, Sun K (2020) Seir modeling of the Covid-19 and its dynamics. Nonlinear Dyn 101(3):1667–1680

He X, Lau EH, Wu P, Deng X, Wang J, Hao X (2020) Temporal dynamics in viral shedding and transmissibility of Covid-19. Nat Med 26(5):672–675

Hethcote H (1989) Three basic epidemiological models. In: Gross L, Hallam TG, Levin SA (eds) Applied mathematical ecology. Springer, Berlin

Holmdahl I, Buckee C (2020) Wrong but usefulwhat Covid-19 epidemiologic models can and cannot tell us. N Engl J Med

Keeling MJ, Hollingsworth TD, Read JM (2020) The efficacy of contact tracing for the containment of the 2019 novel coronavirus (covid-19). medRxiv

Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Hart G, Rosenfeld K et al (2020) Covasim: an agent-based model of Covid-19 dynamics and interventions. medRxiv

Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, Dickens BL (2020) Interventions to mitigate early spread of sars-cov-2 in Singapore: a modelling study. Lancet Infect Dis

Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S et al (2020) Early dynamics of transmission and control of Covid-19: a mathematical modelling study. Lancet Infect Dis

Kulkarni V, Barat S, Clark T (2019) Towards adaptive enterprises using digital twins. In: 2019 winter simulation conference (WSC), pp 60–74

Laxminarayan R, Wahl B, Dudala SR, Gopal K, Neelima S, Reddy KJ (2020) Epidemiology and transmission dynamics of Covid-19 in two Indian States. Science 370(6517):691–697

Li MY, Muldowney JS (1995) Global stability for the seir model in epidemiology. Math Biosci 125(2):155–164

Li W, Zhang B, Lu J, Liu S, Chang Z, Cao P et al (2020) The characteristics of household transmission of Covid-19. Clin Infect Dis

Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L (2003) Transmission dynamics and control of severe acute respiratory syndrome. Science 300(5627):1966–1970

Macal CM, North MJ (2009) Agent-based modeling and simulation. In: Proceedings of the 2009 winter simulation conference (WSC), pp 86–98

Marathe M, Vullikanti AKS (2013) Computational epidemiology. Commun ACM 56(7):88–96

Martin NK, Vickerman P, Hickman M (2011) Mathematical modelling of hepatitis c treatment for injecting drug users. J Theor Biol 274(1):58–66

Nuzzo JB, Mullen L, Snyder M, Cicero A, Inglesby TV (2019) Preparedness for a high-impact respiratory pathogen pandemic (Technical Report). Johns Hopkins Center for Health Security

Onwuegbuzie AJ (2000) Expanding the framework of internal and external validity in quantitative research

Organization WH et al (2020) Coronavirus disease 2019 (covid-19): situation report, 82. World Health Organization

Peng L, Yang W, Zhang D, Zhuge C, Hong L (2020) Epidemic analysis of covid-19 in china by dynamical modeling. arXiv preprint arXiv :2002.06563

Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N et al (2020) The effect of control strategies to reduce social mixing on outcomes of the Covid-19 epidemic in Wuhan, China: a modelling study. Lancet Public Health

Radulescu A, Cavanagh K (2020) Management strategies in a seir model of Covid 19 community spread. arXiv:2003.11150

Rhodes T, Lancaster K, Lees S, Parker M (2020) Modelling the pandemic: attuning models to their contexts. BMJ Glob Health 5(6):e002914

Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray K-A (2020) Revealing Covid-19 transmission in Australia by sars-cov-2 genome sequencing and agent-based modeling. Nat Med 26(9):1398–1404

Samsuzzoha M, Singh M, Lucy D (2013) Uncertainty and sensitivity analysis of the basic reproduction number of a vaccinated epidemic model of influenza. Appl Math Model 37(3):903–915

Sargent RG (2013) Verification and validation of simulation models. J Simul 7(1):12–24

Silva PC, Batista PV, Lima HS, Alves MA, Guimarães FG, Silva RC (2020) Covid-abs: an agent-based model of Covid-19 epidemic to simulate health and economic effects of social distancing interventions. Chaos Solitons Fractals 139:110088

Simonsen L, Spreeuwenberg P, Lustig R, Taylor RJ, Fleming DM, Kroneman M (2013) Global mortality estimates for the 2009 influenza pandemic from the glamor project: a modeling study. PLoS Med 10(11):e1001558

Teimouri A (2020) An seir model with contact tracing and age-structured social mixing for Covid-19 outbreak. medRxiv

Tian L, Li X, Qi F, Tang Q-Y, Tang V, Liu J, Tang L (2020) Calibrated intervention and containment of the Covid-19 pandemic. arXiv :2003.07353

Wang Y, Wang Y, Chen Y, Qin Q (2020) Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (Covid-19) implicate special control measures. J Med Virol 92(6):568–576

World Health Organization et al (2020) Modes of transmission of virus causing Covid-19: implications for IPC precaution recommendations: scientific brief, 27 march 2020 (Technical Report). World Health Organization

World Health Organization et al. (2020) Corona-virus disease (Covid-19) outbreak, vol 14. https://www.who.int. Accessed Apr

Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E et al. (2020) Prediction models for diagnosis and prognosis of Covid-19: systematic review and critical appraisal. BMJ 369

Springer