UDC - 004.942:614.1

# COMPUTATIONAL MODELLING OF INFECTIOUS DISEASES AND FORECASTING THEIR SPREAD ON BIG CITIES

**Grigor V. Bezirganyan**
Technical University of Munich
Boltzmannstrasse 3, 85748 Garching, Germany
grigor.bezirganyan@tum.de
ORCID iD: 0000-0003-0428-1973
Federal Republic of Germany

**Henrik T. Sergoyan**
Yerevan State University
1 Alek Manukyan St, Yerevan 0025, RA
henriksergoyan@gmail.com
ORCID iD: 0000-0003-0163-7137
Republic of Armenia

doi: 10.56243/18294898-2022.2-56

**Abstract**

COVID-19 pandemic is a global crisis of an unprecedented global scale. Governments were faced with the challenge of designing and implementing policies with great uncertainties concerning their direct impacts and externalities. The effects of so called "policies" were often negative and in some cases further exacerbated the compounding healthcare and economic crises.

This paper presents how computer simulations can be used to test various policies in Yerevan, Armenia, before applying them in real life. Further, it is demonstrated how to identify the features that need to be collected to train well-performing and ethical AI models for healthcare management. Additionally, a machine learning model is introduced that helps to reduce the number of necessary PCR tests by around 98.7%.

***Key words***: covid-19, simulation, machine learning, disease modeling, strategic management, agent-based modeling, healthcare management.

**Introduction**

The COVID-19 pandemic has had a formative impact on the daily lives of people around the world since December 2019 at such a scale and velocity that is arguably unprecedented in modern history [1]. Numerous lockdowns were implemented by the governments to combat the spread of the virus since then [2]. Approaches to control the pandemic varied from country to country, and although some measures were successful in reducing the speed with which the virus spread, they also had a negative impact on economies, often resulting in recessions [3]. Hence there is a trade-off between the health risks posed by COVID-19 and the economic restrictions resulting from lockdowns and other limitations on economic activity. Therefore, it is important to determine the optimal policies that will minimize the death toll while maximizing economic activity.

Interest in agent-based simulations increased dramatically since the start of the pandemic, primarily because of the freedom and flexibility that they provide. In some cases, models for other infections were re-purposed for COVID-19, in other cases models were developed from scratch [4, 5]. In TU Berlin the MATSim transport mobility simulator was extended to also model the infection spread [6]. While MATSim uses spatial information from real cities for modeling, we opted for an approach that is less computationally complex, more scalable and more “tweakable”. The Covasim [7] simulator was developed from scratch to be fast and highly customizable. Covasim uses contact networks for virus transmission. While Covasim is much easier to use, more customizable and can use real-world demographic information, it doesn't take into account the spatial information of real cities. This means that while it is possible to generate simulations for abstract cities, it is not possible to adapt the simulations for specific real-world cities.

**Conflict setting**

In this paper, we introduce the Evid open-source and agent-based epidemic simulator written in Python. The simulator leverages spatial and demographic information of real cities while ensuring customizability and ease of use. The simulator allows us to model various types of government interventions aimed at reducing the likelihood of a healthcare system becoming overloaded beyond its capacity. We use the number of occupied ICU beds as a metric for healthcare system capacity and usage. The simulator outputs high-level statistics, as well as more granular information. The most granular outcome of the simulator is made similar to the contact tracing methods used by governments.

In the following sections, we demonstrate the structure and logic of the simulator and present an example of the data that it produces. We run the simulator in Armenia's capital city Yerevan with different policy scenarios and show how it can be used by governments for policy testing. We then demonstrate how Machine Learning methods can be used to predict the spread of a virus based on the contact tracing data generated by the simulator. We show that in the case of available data it is possible to reduce the number of necessary PCR tests by around 98.7% and still discover around 96% of the positive cases.

**Materials and Methods**

**Environment**

We want to create a virtual environment that will resemble the spatial and demographic structure of a real city. The environment defines the space where agents will live and interact with each other.

To include the spatial information of real cities in our environment, we use data retrieved from OpenStreetMap.org [8]. It is also possible to use data from other sources, as long as the format meets the requirements: the data is in a tabular form, where each row contains the building's unique ID, district name, district id, coordinates and building type. Based on that data we build a virtual environment resembling the spatial structure of the city. An example of a virtual environment created on the data of the capital city of Armenia Yerevan is presented in Fig. 1.

Buildings can be public (restaurants, cafes, schools, etc) or residential. Residential buildings are divided into apartments. The number of apartments in each building is sampled from a predefined distribution. While it is possible to add more public buildings to the configuration, the simulator must always have public buildings of the following types: school, university, work and hospital.

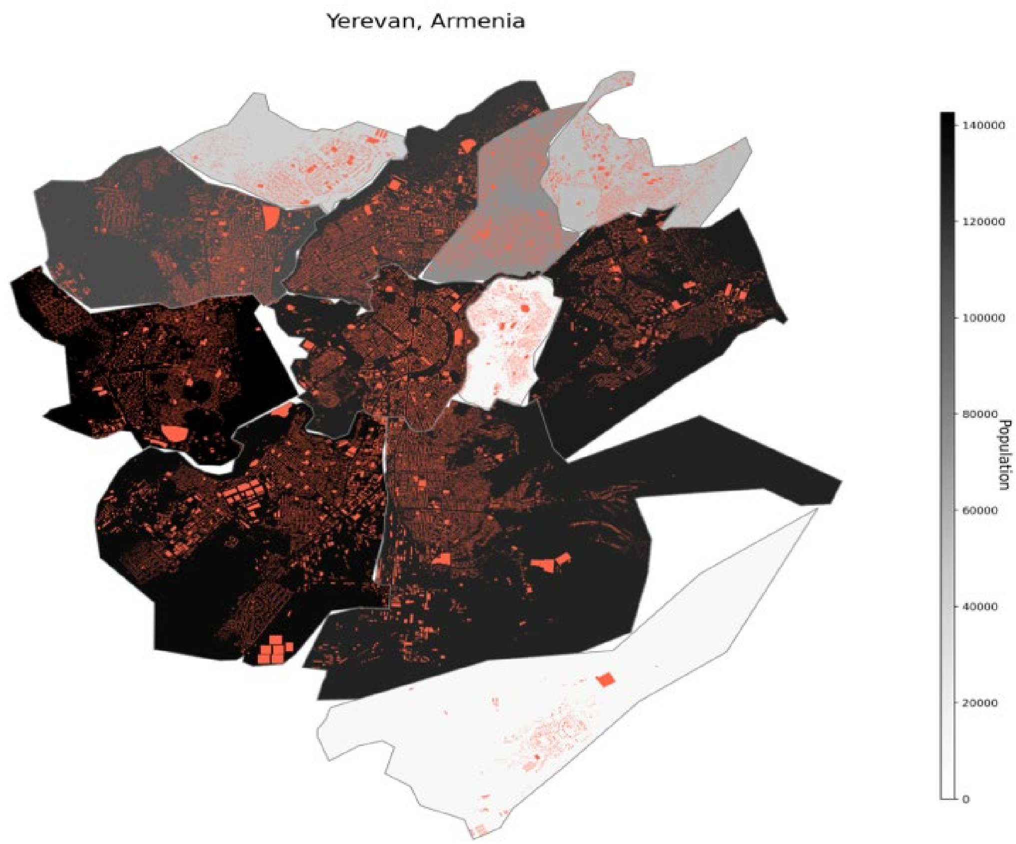


**Fig. 1 The visualization of the geospatial data of Yerevan, Armenia used for creating the virtual environment. The orange boxes are the buildings belonging to their administrative districts. The color of each district represents the population in the respective district.**

### Agents

After creating the buildings, agents are distributed by district, according to the predefined demographic statistics for each district. During the distribution, each agent is assigned an apartment, age group, and gender. Age groups also determine whether the agent will have a location for work or study.

To simulate the mobility of the agents, we need them to move between districts and buildings and interact with agents. For this, we use the Mesa agent-based modeling framework [9]. Mesa uses discrete steps for simulation. During one simulation step, every agent performs a predefined set of instructions. In our case, each step corresponds to one hour, hence, we need to do 24 simulation steps to simulate one day.

In each step each of the agents does two operations: move and infect. Agents move to a building based on a probability distribution conditioned on the age group of the agent, day of the week, and the time of the day. By adjusting the probabilities it is also possible to enforce a lockdown by setting the probabilities of respective facilities to 0.

### Disease transmission

After an agent finishes the moving step, a portion of agents which are in the same building or apartment is sampled as contacted agents (the contact probability is defined for each building type). If the agent is infected, we calculate the probability of its contacts getting infected according to [10]. Some of the parameters that contribute to the calculation of the probability are room size, speaking frequency, loudness, whether people wear masks, etc. By default, the parameters are adapted for COVID-19 according to [10], but they can be adjusted for other diseases.

After infection, agents go through an incubation period, during which they don't have any symptoms and cannot infect other people. The incubation period is a random number of steps, sampled from a normal distribution with a default mean of 48 and a standard deviation of 7. After the incubation period is over, agents get severity status, which can be asymptomatic, mild, or severe, based on a predefined probability distribution. If the status is asymptomatic, agents continue to move as usual but can infect other agents. In case of mild severity, agents are quarantined in their apartments, and if the status is severe, they are taken to a hospital, where the possibility of dying is much lower. Hospitals have a finite maximum capacity. If there are no ICU beds available, agents are quarantined in their homes, until space becomes available and they are moved to a hospital.

If the infected agent does not die within a predefined number of days (about 10), they become more healthy, and cannot infect others or get infected. In that case, they are removed from the simulator, since they cannot affect the outcome of the simulator anymore.

**Research results**

We configured the simulator for the capital of Armenia, Yerevan. The city is divided into 12 districts, with a total population of 1080311 people [11]. We also configured information such as the age and sex distribution [12], and the number of residential buildings and apartments in each district [13] for the simulator. The number of cafes, universities, schools, offices, shops and other buildings is retrieved from the OpenStreetMap [8].

For the parameters that we couldn't find in open source databases (such as areas or volumes of the facilities, probability of having a contact in the facility, number of the total ICU beds, etc.), we used our judgment to assign approximate values, which later can be tuned to get the more accurate output.

We ran the simulator for 30 simulation days (i.e. 1440 steps) and logged only contacts of infected people. This generated around 3.4 GB of data.

Within those 60 days, we had 75333 infections 1631 of which died. We can also observe a higher death rate when there are no ICU beds available (Fig. 2).

To demonstrate how the simulator can be used for testing policies, we run the simulator for 31 days with different parameters and policies and with 100 initially infected agents. We summarize the results of the experiments below.

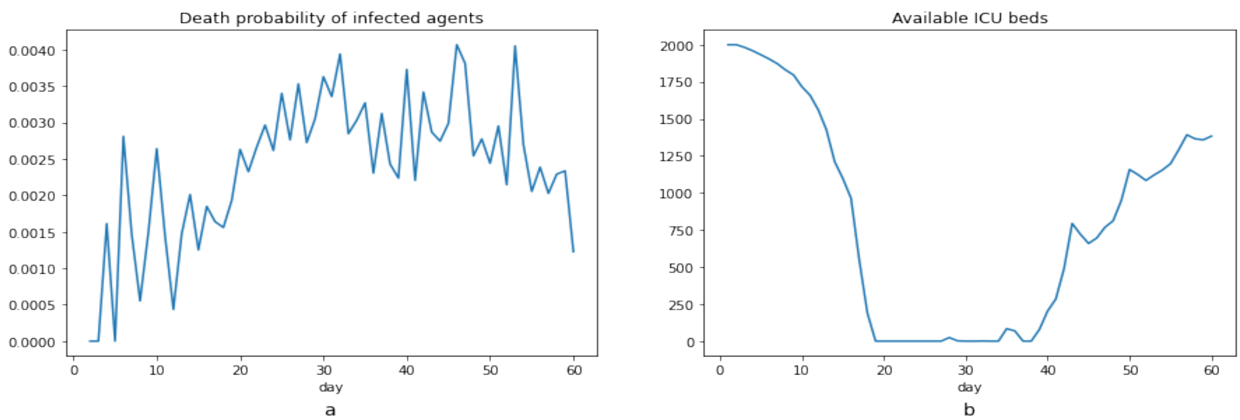


**Fig. 2 There exists a moderate negative correlation of 0.625 between the probability of dying of infected agents (a) and the number of available ICU beds (b). When there are no beds available the probability of dying is higher.**

- **Experiment 1** (Fig. 3 (a)): We run the simulator as usual without any restrictions. In 31 days around 7500 agents were infected and without any measures to stop the virus, the number of infections kept increasing.
- **Experiment 2** (Fig. 3 (b)): We introduced the mandatory mask policy from the first day till the last. In this case, the total number of infections was under 1000, and the infection wave died in around 20 days.
- **Experiment 3** (Fig. 3 (c)): the simulation started as usual and in day 15 we introduced the mandatory mask policy. We can see how the trend of infection changes from increasing to decreasing right after the mask mandate.
- **Experiment 4** (Fig. 3 (d)): the simulation started as usual, from day 15 cafes and gyms were closed. In this scenario no one wears masks, and we can see that while the trend changes after the policy change, the cases are still much higher compared to the mandatory mask mandate in experiment 3.
- **Experiment 5** (Fig. 3 (e)): There is a mandatory mask mandate from the first day, and the gyms and cafes are closed from day 15. In this scenario, closing gyms and cafes didn't have much impact, as on day 15 the infection wave was dying already.
- **Experiment 6** (Fig. 3 (f)): In this scenario, a full lockdown on all public facilities was introduced from day 15. As expected a full lockdown is effective and reduces the number of active cases to almost 0 within 15 days.
- **Experiment 7** (Fig. 3 (g)): In this scenario, there is a full lockdown and mask mandate from day 15. The mask mandate is not having an impact because of full lockdown agents don't have contact with other agents other outside their household.
- **Experiment 8** (Fig. 3 (h)): In this experiment, we enforce a mask mandate from the first day and a full lockdown from day 15. The full lockdown doesn't affect the curve for the same reason mentioned in experiment 5.

Next, we use the synthetic data generated by the simulator to train Machine Learning models for predicting the risks of people catching the virus, given that they had contact with an infected person.

The real-life scenario that we try to replicate with synthetic data is the following: a government is using a contact tracing application to identify contacts with infected people. The people who have contact with an infected person are tested for COVID-19 using a PCR test. For the sake of simplicity and without loss of generality we assume that everyone in the city uses the contact tracing application.

The government would like to reduce the number of necessary tests by using machine learning algorithms on the contact tracing data. The data has to be granular enough to give good enough predictions but at the same time needs to use features that are within ethical and legal bounds. For example, it cannot contain any information that can be used for personal identification.

For this, we clean and preprocess the 60-day data described in the previous subsection. After the preprocessing the data for training has the structure shown in Table 1. As we can see the data doesn't hold any personal features and cannot be used for personal identification.

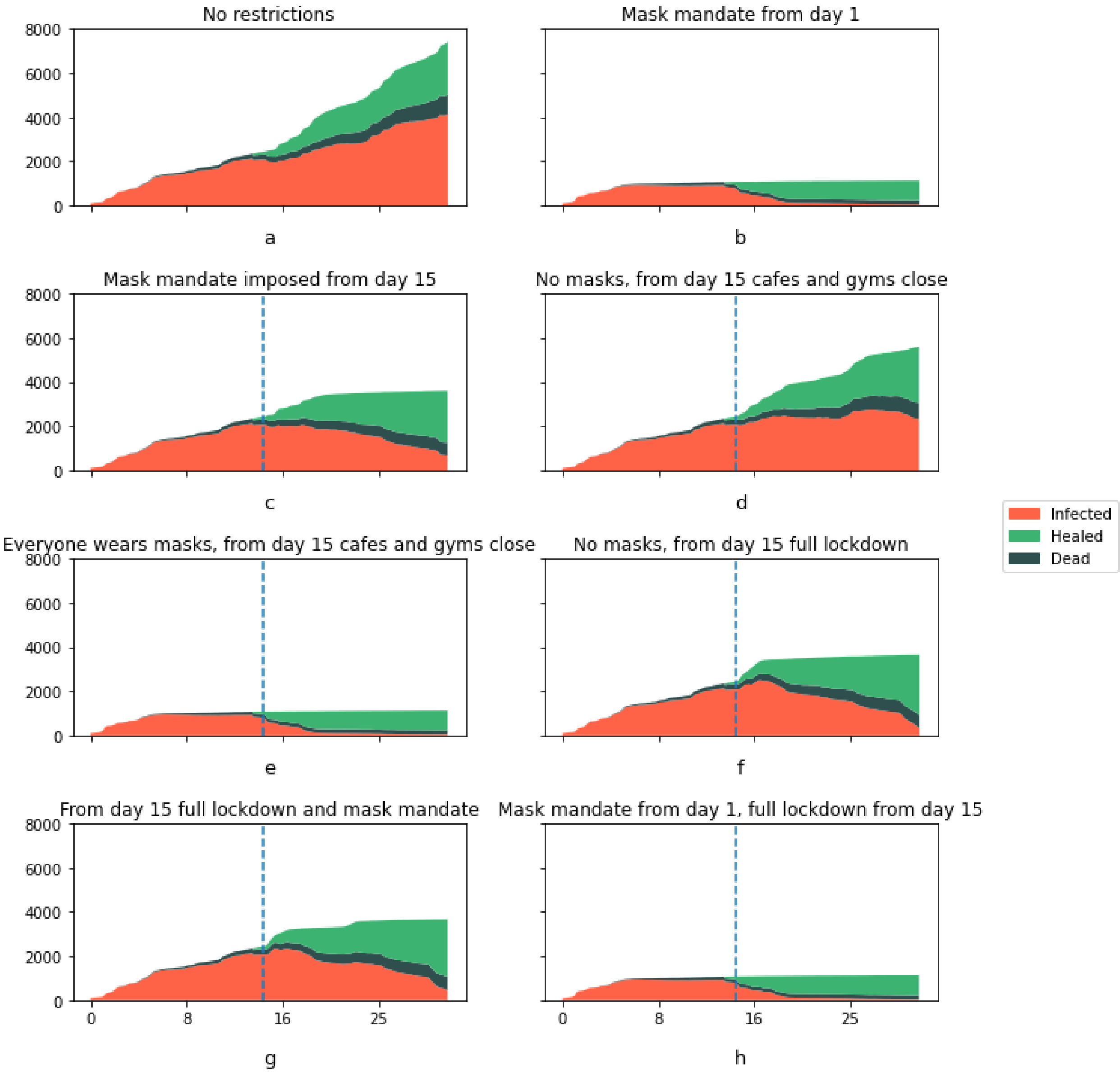


**Fig. 3 Experiment results of different policies:**
**(a) No policies, (b) Mask mandate from day 1, (c)**
**Mask mandate from day 15, (d) Cafes and gyms are closed from day 15, (e) Mask mandate from day 1, and cafes and gyms are closed from day 15, (f) Full lockdown from day 15, (g) Mask mandate and full lockdown from day 15, (h) Mask mandate from day 1 and full lockdown from day 15**

**Table 1**

**Confusion matrix on the test set**

| | 0 | 1 |
|---|---|---|
| 0 | 7521327 | 950469 |
| 1 | 12624 | 12624 |

We use an XGBoost [14] model for training, with a max-depth of 4 and use positive class scaling to deal with class imbalance. We choose to use XGBoost as even after several years of its introduction, it is still one of the state-of-the-art methods for tabular data [15].

From the model's confusion matrix (Table 1) on the test set, we can see that we get around 96% of True Positive Rate (TPR) and 89% of True Negative Rate. This means that if we perform tests only on the agents predicted as positive by the model, we will need to do only 963093 tests instead of 8484957 and will miss only 537 positive cases. This means that we can reduce the number of required tests by around 98.7% and still detect 96% of the positive cases.

**Conclusions**

In this paper we introduced the novel Evid open-source, agent-based pandemic simulator that allows us to simulate pandemics on virtual replicas of real cities and tests policies before applying them in real life.

We also demonstrated with an example that the synthetic dataset can be used to develop ethical AI models for managing the pandemic. We developed an example model that helps us to reduce the number of necessary tests by around 98.7%. It is important to note that while this holds on a synthetic dataset, the number will be different in a real-world scenario because of less data availability and more noisy datasets.

## References


1. World Health Organization Novel Corona Virus (2019-nCoV) Situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
2. Hannah Ritchie, Edouard Mathieu, L.R.G.C.A.C.G.E.O.O.J.H.B.M.D.B., Roser M. Coronavirus Pandemic (COVID-19) (2020) //Our World in Data 2020, https://ourworldindata.org/coronavirus.
3. Zumbrun J. Corona Virus, Slump is worst since great depression, Will It Be as Painful? https://www.wsj.com/articles/coronavirus-slump-is-worst-since-great-depression-will-it-be-as-painful-11589115601, 2021.
4. Silva P.C., Batista P.V., Lima H.S., Alves M.A., Guimarães F.G., Silva R.C. COVID-ABS, An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions (2020) //Chaos, Solitons & Fractals 2020, 139, 110088. doi:10.1016/j.chaos.2020.110088
5. Hussain T., Ozair M., Ali F. S., Assiri T.A., Mahmoud E.E. Sensitivity analysis and optimal control of COVID-19 dynamics based on SEIQR model (2021) //Results in Physics 2021, 22, 103956, doi:10.1016/j.rinp.2021.103956.
6. Müller S. A., Balmer M., Charlton B., Ewert R., Neumann A., Rakow C., Schlenther T., Nagel K. Using Mobile Phone Data for Epidemiological Simulations of Lockdowns, Government Interventions, Behavioral Changes and Resulting Changes of Reinfections. p. 2020.07.22.20160093. doi:10.1101/2020.07.22.20160093.
7. Kerr C.C., Stuart R.M., Mistry D., Abeysuriya R.G., Rosenfeld K., Hart G.R., Núñez R.C., Cohen J.A., Selvaraj P., Hagedorn B. and others, Covasim: An Agent-Based Model of COVID-19 Dynamics and Interventions, 17, e1009149. doi:10.1371/journal.pcbi.1009149.
8. OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org. https://www.openstreetmap.org, 2017.
9. Kazil J., Masad D., Crooks A. Utilizing Python for Agent-Based Modeling, The Mesa Framework, In Proceedings of the Social, Cultural, and Behavioral Modeling, Thomson R., Bisgin H., Dancy C., Hyder A., Hussain M., Springer International Publishing, Cham, 2020.- p. 308-317.

10. Lelieveld J., Helleis F., Borrmann S., Cheng Y., Drewnick F., Haug G., Klimach T., Sciare J., Su H., Pöschl U. Model Calculations of Aerosol Transmission and Infection Risk of COVID-19 in Indoor Environments, 17, 8114, doi:10.3390/ijerph17218114.
11. Grigor Bezirganyan and Henrik Sergoyan, Covid-19, Synthetic contact-tracing dataset, V2, 2022. doi:10.17632/7WWRYN57TR.2.
12. Statistical Committee of the Republic of Armenia, Population report. https://armstat.am/file/article/demog_2019_2.pdf, 2019.
13. Statistical Committee of the Republic of Armenia, RA, Housing Fund and Utility in 2019, https://armstat.am/file/article/bn_fond_2020_1.pdf, 2019.
14. Chen T., Guestrin C., XGBoost: A Scalable Tree Boosting System. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; ACM, New York, NY, USA, 2016, KDD ’16, p. 785–794, doi:10.1145/2939672.2939785.
15. Shwartz-Ziv R., Armon A., Tabular Data: Deep Learning Is Not All You Need. Information Fusion 2022, 81, 84–90. doi:10.1016/j.inffus.2021.11.011.


## ՎԱՐԱԿԻՉ ՀԻՎԱՆԴՈՒԹՅՈՒՆՆԵՐԻ ՀԱՇՎՈՂԱԿԱՆ ՄՈԴԵԼԱՎՈՐՈՒՄ ԵՎ ԴՐԱՆՑ ՏԱՐԱԾՄԱՆ ԿԱՆԽԱՏԵՍՈՒՄ ՄԵԾ ՔԱՂԱՔՆԵՐՈՒՄ


**Բեզիրգանյան Գ.Վ.[1], Սերգոյան Հ.Տ.[2]**
*[1]Մյունխենի տեխնիկական համալսարան*
*[2]Երևանի պետական համալսարան*



ԿՈՎԻԴ-19 համաճարակը աննախադեպ մասշտաբի համաշխարհային ճգնաժամ է: Կառավարությունները բախվեցին օրենքների պլանավորման և իրականացման մարտահրավերին՝ կապված դրանցում առկա անմիջական և արտաքին անորոշություններով: Նշված քաղաքականության հետևանքները հաճախ բացասական են եղել և որոշ դեպքերում էլ ավելի են սրել առողջապահական և տնտեսական ճգնաժամերը: Սույն հոդվածում ներկայացված է, թե ինչպես կարող է համակարգչային սիմուլյացիան օգտագործվել Երևանում տարբեր քաղաքականություններ փորձարկելու համար, նախքան դրանք իրական կյանքում կիրառելը: Ավելին, աշխատանքում նկարագրված է թե ինչպես կարելի է գտնել այն պարամետրերը, որոնք պահանջվում են առողջապահական կառավարման համար արդյունավետ և էթիկական արհեստական բանականության մոդելներ պատրաստելու համար: Վերջում ներկայացված է մեքենայական ուսուցման մոդել, որն օգնում է նվազեցնել անհրաժեշտ ՊՇՌ թեստերի քանակը մոտ 98,7%-ով:



***Բանալի բառեր.*** ԿՈՎԻԴ-19, սիմուլյացիա, մեքենայական ուսուցում, հիվանդությունների մոդելավորում, ստրատեգիաների կառավարում, ագենտային մոդելավորում, առողջապահության կառավարում:

# ВЫЧИСЛИТЕЛЬНОЕ МОДЕЛИРОВАНИЕ ИНФЕКЦИОННЫХ ЗАБОЛЕВАНИЙ И ПРОГНОЗИРОВАНИЕ ИХ РАСПРОСТРАНЕНИЯ В КРУПНЫХ ГОРОДАХ

**Безирганян Г.В.[1], Сергоян Г.Т.[2]**
[1]*Мюнхенский технический университет*
[2]*Ереванский государственный университет*

Пандемия COVID-19 - глобальный кризис беспрецедентного масштаба. Правительства столкнулись с проблемой разработки и реализации политики с большой неопределенностью в отношении ее прямого воздействия и внешних факторов. Последствия указанной политики часто были негативными, а в некоторых случаях еще больше усугубляли кризисы в сфере здравоохранения и экономики. В этой статье показано, как можно использовать компьютерное моделирование для проверки различных политик в Ереване, перед их применением в реальной жизни. Далее демонстрируется, как определить параметры, которые нужны для разработки эффективных и этичных моделей искусственного интеллекта для управления здравоохранением. Кроме того, представлена модель машинного обучения, которая помогает сократить количество необходимых ПЦР-тестов примерно на 98,7%.

***Ключевые слова:*** COVID-19, моделирование, машинное обучение, моделирование заболеваний, стратегический менеджмент, агентное моделирование, управление здравоохранением.

Submitted on 26.04.2022.
Sent for review on 27.04.2022.
Guaranteed for printing on 14.06.2022.