DOI: 10.32347/2412-9933.2026.65.88-96
UDC 004.891:519.86:616.9

Denys Nevinskyi
ORCID: https://orcid.org/0000-0002-0962-072X
Lviv Polytechnic National University, Lviv, Ukraine
PhD, Associate Professor of the Department of Electronic Means of Information and Computer Technologies
Yaroslav Vyklyuk
ORCID: https://orcid.org/0000-0003-4766-4659
Lviv Polytechnic National University, Lviv, Ukraine
DSc (Eng), Professor of the Department of Artificial Intelligence

Article history:
Received: 20.01.2026
Accepted: 03.02.2026
Published: 26.03.2026

MANAGERIAL ASPECTS OF REINFORCEMENT LEARNING FOR INFECTION CONTROL IN UKRAINE: A Q-TRAINING APPROACH

Abstract. *This paper addresses the critical challenge of optimizing managerial decisions for controlling the spread of viral infections, a problem that gained global significance following the COVID-19 pandemic. The primary objective of the research is to develop and implement a reinforcement learning-based system for the dynamic management of restrictive measures – such as self-isolation, mask-wearing, and the closure of institutions – to minimize the number of critical cases while reducing economic and social losses. The methodology integrates the SEIR epidemiological model with a complex multi-agent system. The simulation environment consists of functional blocks including buildings (homes, schools, offices, shops, and restaurants) and agents categorized by age and social roles. A Q-learning approach is applied to optimize restriction strategies. The reward function for the learning model is meticulously designed to balance three key objectives: minimizing the severity of restrictions, ensuring smooth transitions between quarantine levels, and reducing the total number of critically ill patients. The system was tested using geospatial and demographic data for the city of Kyiv, Ukraine, involving a simulated population of approximately 20,000 individuals. Three distinct scenarios were analyzed: a baseline with no restrictions, a scenario with pre-set static restrictions, and a dynamic management approach driven by reinforcement learning. The results demonstrate that the dynamic strategy significantly outperforms traditional approaches. Specifically, the proposed reinforcement learning model achieved a 14.43% reduction in the maximum number of critical cases in Kyiv compared to the static restriction scenario. Furthermore, the study provides a visualization of infection dynamics on a digital map, allowing for real-time observation of agent movements and state transitions (susceptible, infected, recovered, and critical). These findings suggest that the developed system serves as a powerful tool for public health authorities and government administrations to plan effective preventive measures against future epidemiological threats. Despite limited training time, the model proved capable of autonomous strategy generation, highlighting its potential for adaptive pandemic management.*

Keywords: infection prevention; reinforcement learning; Q-learning; mathematical modeling; infection control; optimization of medical resources

Introduction

Viral infections are diseases caused by the entry and influence of viruses on the body. Due to their structure, viruses can change frequently, which makes the task of developing drugs extremely difficult [1].

This problem was most acutely manifested during the COVID-19 pandemic, when the world was faced with a new coronavirus, and with it – with the fear of economic collapse and mass deaths. According to statistics, as of 08/13/2022, about 6.5 million people died in the world [2]. In terms of economic losses, while it is impossible to say exactly what the economic damage from the global COVID-19 pandemic will be, most major economies have lost at least 3.4 percent of their GDP in 2020. To put this figure into perspective, global GDP was estimated at approximately $84.54 trillion in 2020 – meaning that a 4.5 percent drop in economic growth results in a loss of economic output of almost $2.96 trillion [3].

Simulation has great potential in the fight against both current and future viral infections. Simulation can quickly facilitate the preparation of medical facilities and the training of large numbers of medical professionals and students [4]. It also helps to analyze safety and control systems for the spread of viruses in different environments, such as confined areas or buildings [5; 6].

Simulation systems can be used to monitor the behavior of a virus. Simulation provides an abstract representation of reality, conveying the details and characteristics of reality in a simple program. This allows us to predict possible variants of the spread of a viral infection and develop ways to prevent its further spread, minimizing further losses [7; 8].

Therefore, the aim of the work is to create a reinforcement learning system to optimize the strategy for preventing the spread of viral infections and to test this system for different configurations, analyze the impact and compare different types of strategies.

Literature Review

In the modern world, the problem of the spread of infectious diseases, in particular COVID-19, requires new approaches to modeling and analysis. Among the tools used to study epidemic processes, fuzzy multi-agent systems, extended SIR models [9] and deep learning methods [10] stand out. These approaches allow us to take into account the diversity of the population, individual characteristics of agents and complex interactions between pathogens.

So, let's analyze the articles on this topic in order to understand the need for research and implementation of such a system.

The study “Fuzzy Multi-Agent Simulation of COVID-19 Pandemic Spreading”[11] proposes an approach to modeling the spread of the COVID-19 pandemic based on fuzzy multi-agent systems. The agent parameters take into account the population distribution by age and the socioeconomic vulnerability index. Using real data on people from the West Indies (Guadeloupe, French West Indies). The results show that hospital capacity is exceeded and the number of deaths exceeds 2% of the infected population, which is close to reality.

The shortcomings of this study include limited results for other regions or populations, dependence on the accuracy of the input data and model parameters, and possible simplification of the simulation of the infectiousness of virus variants. An additional problem may be the lack of a detailed analysis of the impact of vaccination and other preventive measures on the spread of the disease within the model.

The paper “Individual Variation Affects Outbreak Magnitude and Predictability in an Extended Multi-Pathogen SIR Model of Pigeons Visiting Dairy Farms”[12] examines the transmission of zoonotic diseases between animals and humans, which is an increasing risk, especially in agricultural contexts where individual heterogeneity plays an important role. In particular, interactions between pigeons and cows in indoor dairy farms can cause significant disease spread and economic losses to farmers, putting livestock, nearby human populations, and other wildlife at risk. The model extends the SEIRD framework to account for both intraspecific and interspecific pathogen spread, as well as the dynamics of pigeon movement, which play a critical role in the spread of infectious agents.

The limitations of this study may include limitations in generalizing the results to different systems and scenarios, the need for a more in-depth analysis of interactions between pathogens, and the impact of heterogeneity in host behavior on disease spread.

The study “Epidemic Control on a Large-Scale-Agent-Based Epidemiology Model using Deep Deterministic Policy Gradient”[13] analyzed pandemic mitigation measures such as lockdowns, rapid vaccination programs, school closures, and economic stimulus. These interventions can have positive or unintended negative consequences. The study attempts to model and determine the optimal intervention through automatic circular replication, but suffers from limitations related to the simulation objectives, the scale (several thousand individuals), the types of models that are not suitable for interventions, and the number of intervention strategies that can be explored (discrete versus continuous). To overcome these challenges, a policy optimization framework based on the Deep Deterministic Policy Gradient (DDPG) is used in a large-scale (100,000 individuals) epidemiological agent-based simulation with multi-objective optimization.

The limitations of this study include limitations in the ability to generalize the results to different scenarios, the need for a more detailed analysis of the types of models, and limitations in the number of intervention strategies that can be investigated.

The paper “Dynamic multi-strategy integrated differential evolution algorithm based on reinforcement learning for optimization problems”[14] considers the implementation of a multi-population structure into the differential evolution (DE) algorithm, which is proven to be an effective way to achieve algorithm adaptation and multi-strategy integration. However, existing studies have found that the choice of mutation strategy of each subpopulation during execution is fixed, which leads to poor self-adaptation of subpopulations. To solve this problem, this paper proposes a dynamic differential evolution algorithm with multi-strategy integration based on reinforcement learning. By applying reinforcement learning, each subpopulation can adaptively choose a mutation strategy depending on the current state of the environment (population diversity).

The paper “Synergistic Integration Between Machine Learning and Agent-Based Modeling: A Multidisciplinary Review”[15] explores the use of agent-based models and machine learning for decision-making. The authors point to the possibility of improving sequential decision-making using machine learning models that learn the behavioral patterns of agents. The study provides a comprehensive overview of the application of machine learning in agent-based models in four main scenarios: training a conditional microagent, interfering with the behavior of a microagent, emulation at the macro level, and sequential decision-making. The shortcomings of this study include the limited generalizability of the results to different application scenarios, the need for a more detailed analysis of additional components of the method, and the lack of a detailed analysis of the sensitivity of the method to changes in parameters and settings.

The paper "Continuous-discrete GeoSEIR(D) model for modelling and analysis of geo spread COVID-19"[16] investigates the use of a geospatial SEIR(D) model based on a multi-agent approach to model the spread of viral infections taking into account human activity and geodata. The model was validated on the example of COVID-19 in Lviv, analyzing different prevention strategies. A 50% reduction in the probability of infection due to masks postponed the peak by 150 days and reduced the number of patients by 25%, and a 75% reduction by 240 days with a 60% reduction. Restrictions on transport and public places led to a peak on day 165 with 2854 patients. Vaccination by 50%, 75% and 100% reduced the peak number of infected people by 34%, 57% and 94%, respectively. Weekend quarantines had little effect on the total number of infections. The combination of masks, travel restrictions, and vaccinations proved most effective, keeping the average number of cases at 8 with a peak of no more than 15 over four years. The study's limitations include limited applicability to other regions, insufficient accounting for social factors, and sensitivity to policy settings..

The paper "A Managerial Approach towards Modeling the Different Strains of the COVID-19 Virus Based on the Spatial GeoCity Model"[17] explores a modification of the GeoCity model to account for population age, activity schedules, and agent health status for realistic city modeling. The model was used to simulate the spread of three strains of COVID-19. Calculations showed that the virus is transmitted mainly through work and transport, and children are carriers, but not initiators of the epidemic. The dynamics of spread vary depending on weekdays and weekends.

Fast strains (omicron) spread more actively through transport, while slow strains (alpha) spread more actively through work and educational contacts. Immunocompetent individuals are the main carriers of fast strains, while immunocompromised individuals - slow ones. The shortcomings of the study include dependence on model assumptions and its limited generalizability.

In the study "Modeling and analysis of different scenarios for the spread of COVID-19 by using the modified multi-agent systems – Evidence from the selected countries"[18]: the SIR model based on a multi-agent system using mobile cellular automata for modeling the spread of COVID-19 was improved. Improved rules for agent interaction and methods for correlating model parameters with real geographical, social and medical indicators were proposed. This allows modeling the spatial distribution of the virus both in a separate region and at the country level, taking into account transport, shops, educational institutions, parks, etc. The model also assesses the impact of quarantine, transport restrictions, mask regime and social distancing. Experiments were conducted to assess the impact of individual and combined measures to combat the pandemic. A method for comparing the dynamics of the model with real data was proposed, which made it possible to assess the effectiveness of government measures in the Chernivtsi region (Ukraine). The spread of the virus was also simulated in Slovakia, Turkey, and Serbia, where the model's forecast showed high accuracy. The study's shortcomings include dependence on the model's assumptions and possible limitations in its application to other countries and scenarios.

As we can see from the above works, simulations are carried out on one environment and the results are tied and use data from one region. Therefore, it is necessary to build a system for modeling viral infections in which it would be possible to set the parameters of the environment. Also, a reinforcement learning model that would dynamically, depending on the state of the model, change the parameters of the environment, that is, impose restrictions. This will make the simulation more realistic and close to real events, as well as more adaptive for various types of data. When using multi-agent systems, it is necessary to divide agents into different risk zones, which will change the probability of their transition from one state to another.

Materials and Methods

Epidemiological model

In this paper, we will use the SEIR model with data description described in our previous work.[19]

Multi-agent system

In this work, the multi-agent system [20] is divided into blocks, each of which has its own content. The functional blocks are as follows:

- Buildings – an object with parameters in which people can be located and interact.
- Model – the above described model of infection spread.
- Persons (Agents) – categories of persons, each with their own role, buildings they visit, and methods of interaction (Fig. 1).

Assuming that there are areas ("Home", "School", "Office", "University" "Shop", "Restaurant"), an agent can be inside one of the described areas at any given time. It is assumed that transmission of infection occurs with a certain probability when a susceptible agent is in the same cell as an infected agent.

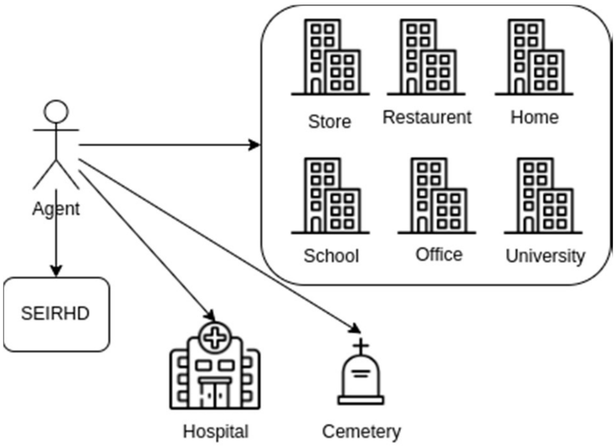

Figure 1 – Model of a multi-agent system

The simulator models a day as 24 separate hours, with each person potentially changing location every hour. At the end of the day, each person's infection status is updated. If contact tracing is active, each person's contacts from previous days are updated. The updated predicted infection status and other state variables are returned to the government as observations. The process is repeated as long as the infection remains active.

Reinforcement learning model

The environment will be the above-described multi-agent system to which constraints will be applied

The actions that the model can perform will be the introduction of constraints to reduce the level of morbidity, as well as the removal of constraints so that agents can perform their actions as fully as possible.

The reward function consists of three parts: the fewer constraints the better, the smoother the constraints the better, and the fewer critically ill patients the better.

The reward function, which denotes the fewer constraints the better, is designed to reward agents for actions that correspond to lower levels of constraints (lower stages).

Mathematical representation:

Normalized rewards per stage are calculated using the formula:

$$r_{stage(i)} = \frac{i^t}{\max_j\left(j^t\right)},\ i \in \left[0, num\_stages - 1\right], \quad (1)$$

where i – is a stage (a discrete value from 0 to num_stages−1); t – is a degree to ensure nonlinear growth of the reward. This parameter can be changed depending on the model parameters or optimized empirically. In this work, it was empirically established that the best agreement with the real situation occurs at a value of 1.5.

Reward for action a corresponding to the stage choice:

$$r = -r_{stage}\left(a\right), \quad (2)$$

where $r_{stage}\left(a\right)$ – normalized reward for the selected stage a.

The smoother the better reward function is designed to minimize abrupt changes between constraint stages to ensure a smooth transition between them.

Mathematical representation:

The difference between the stage in the previous step ($stage_{i-1}$) and the stage at the current step ($stage_{i-1}$) :

$$\Delta_{stage} = |stage_i - stage_{i-1}| \quad (3)$$

The reward is calculated as the average of the differences between all stages at each step:

$$r = -\Delta^{mean}{}_{stage} = = -\frac{1}{N}\sum_{i=1}^{N}\left|stage_i\left(i\right) - stage_{i-1}\left(i\right)\right|, \quad (4)$$

where N— number of measured stages.

The reward of fewer critical cases is the better it compares the number of infected, critical, or dead people between the current and previous step to limit the increase in the number of cases.

$$r = -\frac{\sum_{i=1}^{N} clip\left(\frac{summary - prev\_summary}{prev\_summary}, 0, \infty\right)}{N}, \quad (5)$$

where N – number of cases under consideration; $clip(x, 0, \infty)$ constrains the value from below to 0 and from above to infinity to avoid negative increments.

The total reward is calculated as the weighted average of the above functions.

The weighted average for these rewards can be calculated using the following general formula:

$$r_{weighted} = \frac{\sum_{i=1}^{N} w_i \cdot r_i}{\sum_{i=1}^{N} w_i}, \quad (6)$$

where $r_{weighted}$ – is the weighted average of rewards; r_i – individual reward for the i-th component; w_i – weight corresponding to the i-th component (the more important the component, the greater its weight); N – total number of components.

Implementing restrictions

Self-isolation

To implement this restriction, a parameter was set for an individual agent that if he was sick, the infected person stayed at home, that is, he was attached to an object of the house class and did not leave it until he entered the Recovered state or the Critical state, then he was moved to the hospital.

Wearing masks

To implement this restriction, a coefficient was used that multiplied by the probability of infection spread. For wearing masks, the coefficient was chosen as 0.6. If this restriction was selected as active, this coefficient was multiplied by the coefficient of the agent's transition to the Infected state before recalculating the agent's states, thereby reducing the probability of transition.

Building closures

To implement the restriction as a closure of buildings in an object of the corresponding class, the is_locked parameter was set to positive and because of this, the number of possible visitors and employees for a separate category of buildings was set to 0. This did not allow agents to be in a building of a certain category.

Results and Discussion

So, now having implemented the model and created a software implementation of the work, we will conduct a number of simulations with different parameters to analyze the work of the model and its results.

In this work, we will conduct simulations with data for city Kyiv (Ukraine). After the simulations, we will analyze the results of the work of the two models (Fig. 2, Fig. 3).

Three scenarios will be used during the experiments:

1. Running a simulation without restrictions.
2. Conducting a simulation with pre-set time constraints.
3. Conducting simulation with dynamic constraints.

By pre-set constraints, we mean that before running the model, a scenario will be implemented for which day which constraints will be implemented. The static constraint scenario will be based on the results of previous experiments and analysis of the simulation without constraints.

Dynamic constraints refer to the use of reinforcement learning that will introduce constraints during the simulation, i.e. at the end of the day, it will analyze the results of the simulation day and introduce the necessary constraint.

Restrictions will be divided into several levels:

0) No restrictions.
1) Self-isolation.
2) Self-isolation and social distancing.
3) Self-isolation, mask-wearing and social distancing office closures.
4) Self-isolation, mask-wearing and social distancing, closure of establishments.

The charts will display the level of restriction and its duration. This will be a red arrow above the quantitative data and the level number. Therefore, from the charts you will be able to see which strategy was chosen and which restrictions were imposed in which periods (Fig. 2).

To better understand how the spread simulation proceeds, we will also perform a map representation of the simulation with dynamic constraints. The environment describes the model, system agents, training environment, and system locations. The figure shows a visualization of the spread of infection on a map. The map shows the city of Kyiv. Areas of a certain color are buildings between which agents can move.

Buildings are painted in different colors depending on their category. The category of buildings according to the color is shown in Table 1.

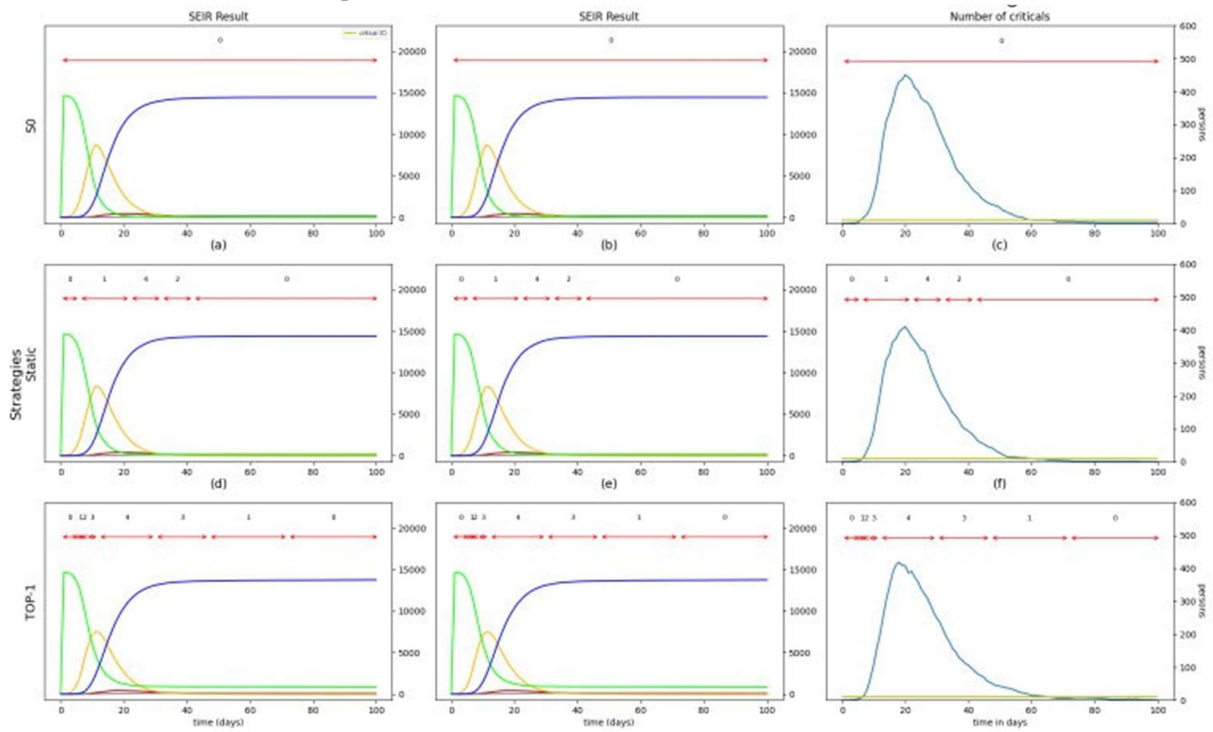

Figure 2 – Comparison of simulations of the spread of infection in the absence of restrictions (S0), static restrictions (Static) and dynamic restrictions (TOP-1)

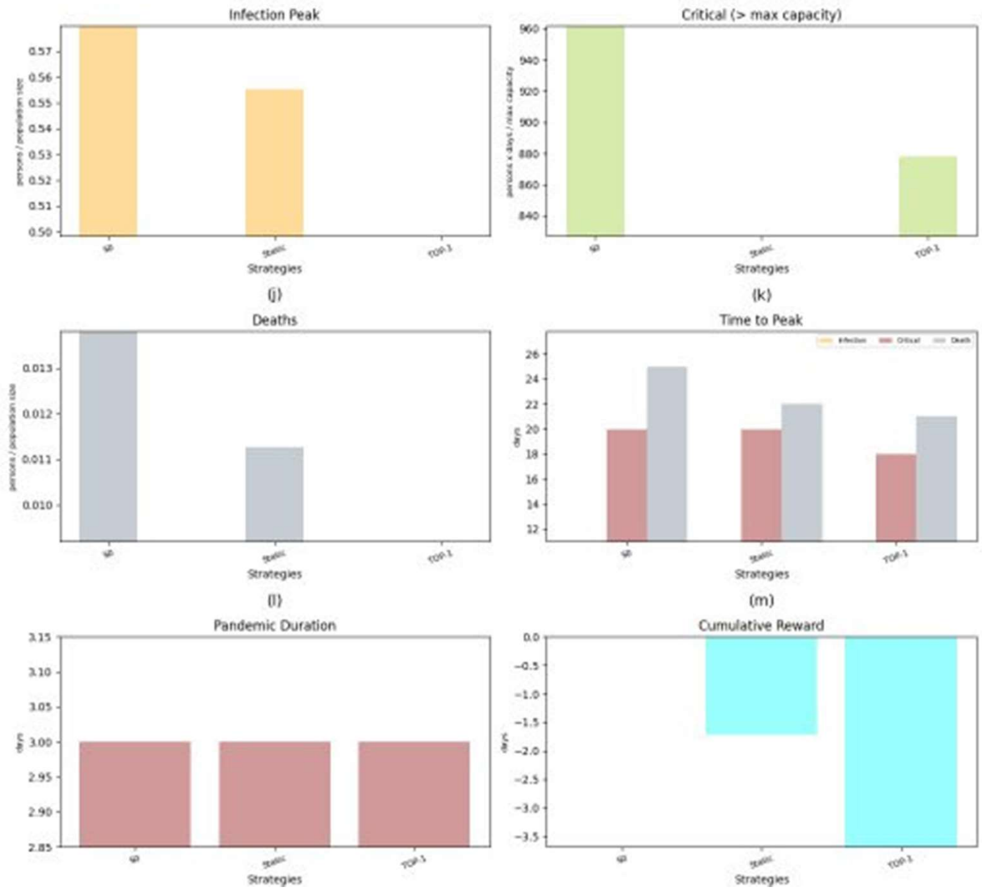

Figure 3 – Comparison of metrics (peak infection, disease duration, number of deaths, and disease growth rate) for simulations with no constraints (S0), static constraints (Static), and dynamic constraints (TOP-1)

Table 1 – ***Number of buildings used in the simulation***

Building	Color
University	Red
Home	Gray
Store	Purple
Hospital	Yellow
School	Green
Restaurent	Blue

Description of results with data for the city of Kyiv (Ukraine)

First, let's analyze the distribution of agents by building. The number of people is calculated based on the number of buildings within a radius of 3000m from the center of Kyiv, that is, it will be about 20,000 people (Table 2). This number was chosen to optimize training and simulation time, since the number of buildings of different categories and the number of people is quite large.

The initial population was distributed across residential, educational, and commercial buildings according to demographic statistics for Kyiv.

Now let's run the simulations. The line on top shows the period of the restriction.

The results show that the implementation of static regulation significantly reduced the number of fatalities, critical states and the maximum level of critical cases, which indicates the effectiveness of the restrictions introduced (Table 3). It is also noticeable that reinforcement learning was able to achieve similar results in the simulation as the scenario with static restrictions.

Table 2 – ***Quantitative simulation indicators***

Type	S			E			I			R		
	Max	Mean	STD	Max	Mean	STD	Max	Mean	STD	Max	Mean	STD
No restrictions	634	109	123	362	147	78	9221	963	2182	14802	1310	3287
Static constraint	384	87	115	294	110	59	7392	888	1337	14198	2429	3821
Dynamic costraint	509	98	121	239	137	73	8382	953	2197	15003	1409	3826

*Table 3 – **Comparison of simulation indicators***

Тип	**Criticals**			**Reduction in the number of critically ill patients, %**
	Max	**Mean**	**STD**	
No restrictions	492	94.32	122.61	0
Static constraint	449	89.38	104.96	8.74
Dynamic constraint	421	82.94	115.19	14.43

The result obtained indicates the need to improve the initial data, conduct more experiments and increase the training time of the model. However, even with limited training time, the model was able to develop a strategy that is potentially effective. For a more detailed analysis, it is worth comparing the results by superimposing them on each other.

So, for a better understanding, let's present the results and the process of passing the simulation on a map. The presentation configurations are described above.

As we can see in the visualization, almost all agents are green, meaning they may be infected. After the simulation starts, the agents will change color depending on their state.

Buildings of different colors will reflect buildings and their categories, for example, buildings with a gray circle are residential buildings, meaning just houses, buildings with a red circle are hospitals, etc.

On the twentieth iteration, as we can see, quite a few agents are yellow, that is, they are in the "Infected" state, and their placement has also changed, which indicates the movement of agents between buildings during the simulation. We also already see a small number of agents in blue, which indicates that they have acquired immunity. Also, if you zoom in, you will see a few agents in a critical state, that is, in red.

As we can see from the figure, most of the agents are already blue, which indicates that they are in the "Recovered" state and have acquired immunity to infection. We also see more individuals in red, which in turn shows that the number of individuals in critical condition has increased (Fig. 4).

Conclusions

The work developed a complex multi-agent system for modeling the spread of viral infections, which includes the SEIR epidemiological model, agents of different categories (age, social) and a dynamic environment (buildings, locations) with the possibility of implementing restrictive measures. An approach based on reinforcement learning was implemented for dynamic management of restrictive measures (self-isolation, wearing masks, closing institutions, etc.) in order to minimize the number of critical cases and overall losses.

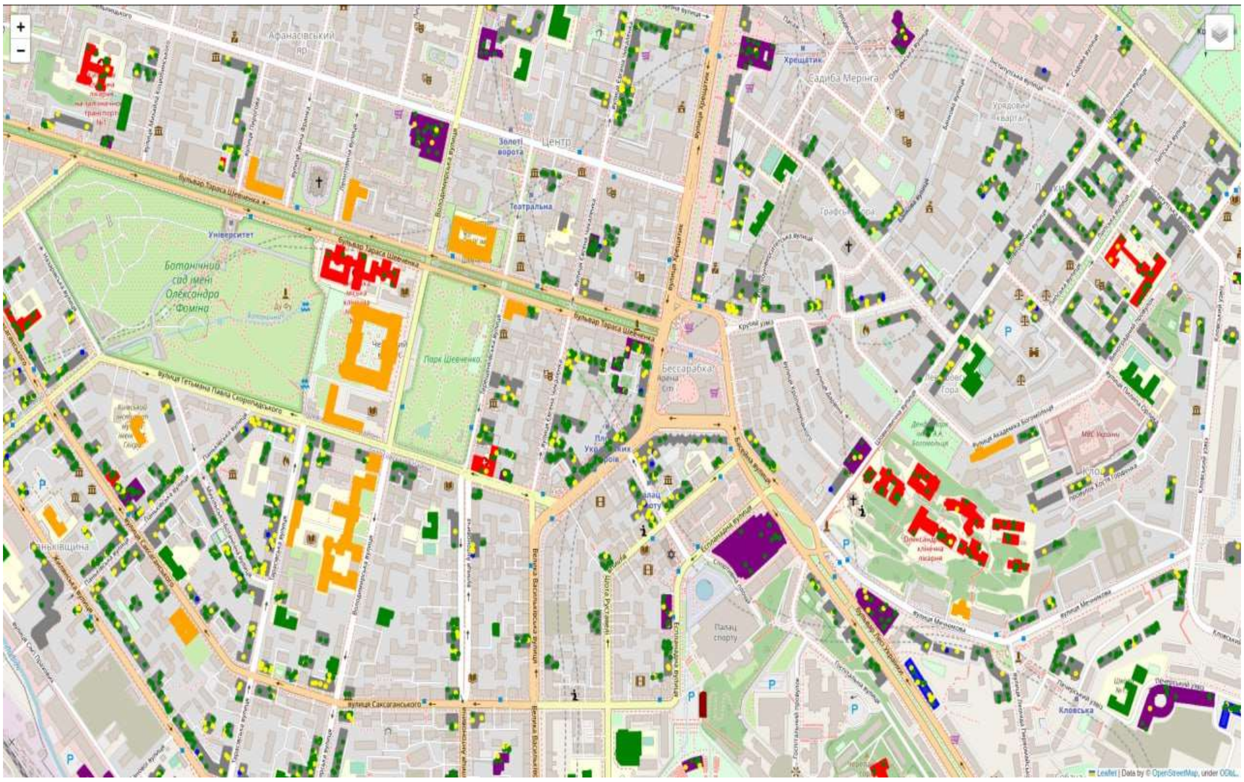

Figure 4 – Visualization of the simulation over Ukraine on the map with dynamic constraints. Thirtieth iteration (thirtieth day)

A comparative analysis of simulation results for citiy Kyiv (Ukraine) – was conducted under three scenarios: without restrictions, with static restrictions. and with dynamic restrictions based on reinforcement learning. The results showed that the dynamic constraint management strategy allows achieving better indicators, in particular, reducing the number of critical cases by 14.43% in Kyiv, compared to static restrictions. Thus, in Kyiv the maximum number of critical cases decreased from 492 to 42.

Visualization of the spread of infection on a map of cities allows you to clearly observe the dynamics of changes in the number of agents in different states (susceptible, infected, recovered, critical) and assess the effectiveness of the applied restrictive measures. The proposed system for modeling and optimizing restrictive measures can be useful for health care and public administration authorities in planning and making decisions to combat the spread of viral infections, especially in pandemic conditions.

Conflict of Interest. The authors confirm that there are no financial, personal, or other interests that could be considered a potential conflict of interest regarding the publication of this article.

Funding. The research was conducted without financial support.

Data Availability. All data are available in digital or graphical form within the main text of the manuscript.

Use of Artificial Intelligence. The authors confirm that no artificial intelligence tools were used in the creation of this work.

References

1. Viral disease. *Wikipedia*. Sep. 16, 2022. Accessed: Oct. 31, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Viral_disease&oldid=1110529921
2. COVID-19 Coronavirus Infection. *Public Health Center of Ukraine*. Accessed: Oct. 31, 2024. [Online]. Available: https://phc.org.ua/kontrol-zakhvoryuvan/inshi-infekciyni-zakhvoryuvannya/koronavirusna-infekciya-covid-19
3. Impact of the coronavirus pandemic on the global economy - Statistics & Facts. *Statista*. Accessed: Oct. 31, 2024. [Online]. Available: https://www.statista.com/topics/6139/covid-19-impact-on-the-global-economy/#editorsPicks
4. SmythOS. Examples of Multi-Agent Systems in Action: Key Use Cases Across Industries. Accessed: Nov. 08, 2024. [Online]. Available: https://smythos.com/artificial-intelligence/multi-agent-systems/examples-of-multi-agent-systems/
5. Gressman, P. T., & Peck, J. R. (2020). Simulating COVID-19 in a university environment. *Mathematical Biosciences*, 328, 108436. https://doi.org/10.1016/j.mbs.2020.108436
6. Batista, B., Dickenson, D., Gurski, K., Kebe, M., & Rankin, N. (2020). Minimizing disease spread on a quarantined cruise ship: A model of COVID-19 with asymptomatic infections. *Mathematical Biosciences*, 329, 108442. https://doi.org/10.1016/j.mbs.2020.108442
7. Mettler, S. K., et al. (2021). The importance of timely contact tracing – A simulation study. *International Journal of Infectious Diseases*, 108, 309–319. https://doi.org/10.1016/j.ijid.2021.04.029
8. Xu, Z., Wu, B., & Topcu, U. (2021). Control strategies for COVID-19 epidemic with vaccination, shield immunity and quarantine: A metric temporal logic approach. *PLoS One*, 16 (3), e0247660. https://doi.org/10.1371/journal.pone.0247660
9. Engbert, R., Rabe, M. M., Kliegl, R., & Reich, S. (2021). Sequential Data Assimilation of the Stochastic SEIR Epidemic Model for Regional COVID-19 Dynamics. *Bulletin of Mathematical Biology*, 83 (1), 1. https://doi.org/10.1007/s11538-020-00834-8
10. Reinforcement learning. *GeeksforGeeks*. Accessed: Apr. 22, 2024. [Online]. Available: https://www.geeksforgeeks.org/what-is-reinforcement-learning/
11. Baz, D. E., & Doncescu, A. (2023). Fuzzy Multi-Agent Simulation of COVID-19 Pandemic Spreading. *arXiv*. https://doi.org/10.48550/ARXIV.2310.17986
12. Lazebnik, T., & Spiegel, O. (2023). Individual Variation Affects Outbreak Magnitude and Predictability in an Extended Multi-Pathogen SIR Model of Pigeons Vising Dairy Farms. *arXiv*. https://doi.org/10.48550/ARXIV.2310.08613
13. Deshkar, G., Kshirsagar, J., Hayatnagarkar, H., & Venugopalan, J. (2023). Epidemic Control on a Large-Scale-Agent-Based Epidemiology Model using Deep Deterministic Policy Gradient. *arXiv*. https://doi.org/10.48550/ARXIV.2304.04475
14. Yang, Q., Chu, S. C., Pan, J. S., Chou, J. H., & Watada, J. (2024). Dynamic multi-strategy integrated differential evolution algorithm based on reinforcement learning for optimization problems. *Complex and Intelligent Systems*, 10 (2), 1845–1877. https://doi.org/10.1007/s40747-023-01243-9
15. Zhang, W., Valencia, A., & Chan, N. B. (2023). Synergistic Integration Between Machine Learning and Agent-Based Modeling: A Multidisciplinary Review. *IEEE Transactions on Neural Networks and Learning Systems*, 34 (5), 2170–2190. https://doi.org/10.1109/TNNLS.2021.3106777
16. Vyklyuk, Y., Nevinskyi, D., & Hazdiuk, K. (2024). Continuous-discrete GeoSEIR(D) model for modelling and analysis of geo spread COVID-19. *Intelligence-Based Medicine*, 10, 100155. https://doi.org/10.1016/j.ibmed.2024.100155
17. Vyklyuk, Y., Nevinskyi, D., Chopyak, V., Škoda, M., Golubovska, O., & Hazdiuk, K. (2023). A Managerial Approach towards Modeling the Different Strains of the COVID-19 Virus Based on the Spatial GeoCity Model. *Viruses*, 15 (12), 2299. https://doi.org/10.3390/v15122299

18. Vyklyuk, Y., Manylich, M., Škoda, M., Radovanović, M. M., & Petrović, M. D. (2021). Modeling and analysis of different scenarios for the spread of COVID-19 by using the modified multi-agent systems – Evidence from the selected countries. *Results in Physics*, 20, 103662. https://doi.org/10.1016/j.rinp.2020.103662

19. Vyklyuk, Y., Nevinskyi, D., & Boyko, N. (2023). GEOCITY – A NEW DYNAMIC-SPATIAL MODEL OF URBAN ECOSYSTEM. *Journal of the Geographical Institute Jovan Cvijic SASA*, 73 (2), 187–203. https://doi.org/10.2298/IJGI2302187V.

20. Allioui, H., et al. (2022). A Multi-Agent Deep Reinforcement Learning Approach for Enhancement of COVID-19 CT Image Segmentation. *Journal of Personalized Medicine*, 12 (2), 309. https://doi.org/10.3390/jpm12020309

Невінський Денис Володимирович
ORCID: https://orcid.org/0000-0002-0962-072X
Національний університет «Львівська політехніка», Львів, Україна
Кандидат технічних наук, доцент кафедри електронних засобів інформаційно-комп'ютерних технологій
Виклюк Ярослав Ігорович
ORCID: https://orcid.org/0000-0003-4766-4659
Національний університет «Львівська політехніка», Львів, Україна
Доктор технічних наук, професор кафедри штучного інтелекту

УПРАВЛІНСЬКІ АСПЕКТИ НАВЧАННЯ З ПІДКРІПЛЕННЯМ ДЛЯ КОНТРОЛЮ ІНФЕКЦІЙ В УКРАЇНІ: ПІДХІД НА ОСНОВІ Q-НАВЧАННЯ

Анотація. *Розглянуто критично важливе завдання оптимізації управлінських рішень для контролю поширення вірусних інфекцій – проблема, що набула глобального значення після пандемії COVID-19. Основною метою дослідження є розробка та впровадження системи на основі навчання з підкріпленням для динамічного управління обмежувальними заходами (такими як самоізоляція, масковий режим та закриття установ) задля мінімізації кількості критичних випадків за одночасного зменшення економічних і соціальних втрат. Методологія поєднує епідеміологічну модель SEIR із комплексною мультиагентною системою. Середовище моделювання складається з функціональних блоків, включаючи будівлі (будинки, школи, офіси, магазини та ресторани), та агентів, класифікованих за віком і соціальними ролями. Для оптимізації стратегій обмежень застосовано підхід Q-навчання. Функція винагороди для моделі навчання ретельно розроблена для збалансування трьох ключових цілей: мінімізації суворості обмежень, забезпечення плавних переходів між рівнями карантину та скорочення загальної кількості критично хворих пацієнтів. Система була протестована з використанням геопросторових і демографічних даних міста Києва (Україна) для симуляції популяції чисельністю приблизно 20 000 осіб. Було проаналізовано три сценарії: базовий без обмежень, сценарій із заздалегідь встановленими статичними обмеженнями та підхід динамічного управління на основі навчання з підкріпленням. Результати демонструють, що динамічна стратегія значно перевершує традиційні підходи. Зокрема, запропонована модель навчання з підкріпленням забезпечила зниження максимальної кількості критичних випадків у Києві на 14,43% порівняно зі сценарієм статичних обмежень. Крім того, дослідження надає візуалізацію динаміки інфекції на цифровій карті, що дозволяє спостерігати за переміщенням агентів та зміною їхніх станів (сприйнятливі, інфіковані, одужалі та критичні) у режимі реального часу. Отримані результати свідчать про те, що розроблена система є потужним інструментом для органів охорони здоров'я та державного управління під час планування ефективних превентивних заходів проти майбутніх епідеміологічних загроз. Попри обмежений час навчання, модель виявилася здатною до автономної генерації стратегій, що підкреслює її потенціал для адаптивного управління пандеміями.*

Ключові слова: профілактика інфекцій; навчання з підкріпленням; Q-навчання; математичне моделювання; інфекційний контроль; оптимізація медичних ресурсів

Link to publication

APA Nevinskyi, D., & Vyklyuk, Y. (2026). Managerial Aspects of Reinforcement Learning for Infection Control in Ukraine: A Q-Training Approach. *Management of Development of Complex Systems*, 65, 88–96, dx.doi.org\10.32347/2412-9933.2026.65.88-96.

ДСТУ Невінський Д. В., Виклюк Я. І. (2026). Управлінські аспекти навчання з підкріпленням для контролю інфекцій в Україні: підхід на основі Q-навчання. *Управління розвитком складних систем.* Київ, 2026. № 65. С. 88 – 96, dx.doi.org\10.32347/2412-9933.2026.65.88-96.