НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2026 Том 26 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2026 Vol. 26 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2026-26-1-177-184

# Reducing computational costs of agent-based modeling of respiratory infection spread using a machine learning-based surrogate model

**Ali Darwish[1]✉, Vasiliy N. Leonenko[2]**

[1,2] ITMO University, Saint Petersburg, 197101, Russian Federation
[2] Smorodintsev Research Instutute of Influenza, Saint Petersburg, 197022, Russian Federation

[1] adarwish@itmo.ru✉, darwishali320@gmail.com, https://orcid.org/0009-0007-3123-3789
[2] vnleonenko@itmo.ru, https://orcid.org/0000-0001-7070-6584

**Abstract**
Agent-Based Models (ABMs) have proven to be an effective tool for describing and predicting the dynamics of respiratory infections and forecasting future outbreaks and have helped health organizations control the disease by developing effective intervention strategy. The use of ABMs is accompanied by very high computational cost, which limits their use in real time. Replacing ABMs with machine learning-based models that can replicate the output or couple the two models together is a solution to the computational cost problem. This paper proposes a machine learning-based surrogate model to simulate an ABM simulating the spread of respiratory infection in Saint Petersburg to reduce simulation time and maintain equivalent accuracy in estimates. The research was based on evaluating the performance of a set of machine learning models under different approaches as surrogate models to use in place of ABM. Methods for generating ABM output chains were compared and evaluated through experiments using single-model approaches or ensemble approaches as a predictive model for each time step in the output (independent multi-output and regression chaining) or hybrid models between agent-based and machine learning. The results indicated that there are several models capable of replicating the simulation output sequence of the ABM with a slight superiority of eXtreme Gradient Boosting within the regression chaining approach. In the hybrid approach, the Long Short Term Memory model with the first values of the output sequence within the feature space outperformed the other models in obtaining more accurate results and achieved the lowest Mean Absolute Error and Root Mean Square Error.

**Keywords**
epidemiology, ABM simulations, surrogate modeling, machine learning, multi-output regression, regression chaining

**Acknowledgements**
The work was supported by the Russian Science Foundation in accordance with Agreement No. 24-25-20104 dated 12.04.2024 and the St. Petersburg State Autonomous Institution “Fund for the Support of Scientific, Scientific-Technical, Innovative Activities” on the provision of a regional grant for co-financing the project under agreement 24-25-20104 dated 22.05.2024 dated 03.06.2024 No. 169.

**For citation:** Darwish A., Leonenko V.N. Reducing computational costs of agent-based modeling of respiratory infection spread using a machine learning-based surrogate model. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2026, vol. 26, no. 1, pp. 177–184. doi: 10.17586/2226-1494-2026-26-1-177-184

УДК 004.85,614.4

# Снижение вычислительных затрат при агентном моделировании распространения респираторной инфекции с помощью суррогатной модели на основе машинного обучения

**Али Дарвиш[1]✉, Василий Николаевич Леоненко[2]**

[1,2] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
[2] НИИ гриппа им. А.А. Смородинцева, Санкт-Петербург, 197022, Российская Федерация

[1] adarwish@itmo.ru✉, darwishali320@gmail.com, https://orcid.org/0009-0007-3123-3789
[2] vnleonenko@itmo.ru, https://orcid.org/0000-0001-7070-6584

**Аннотация**

**Введение.** Агентные модели широко используются для моделирования распространения респираторных инфекций и разработки стратегий вмешательства. Однако высокая вычислительная сложность ограничивает их применение в задачах, требующих оперативного анализа. Решением данной проблемы является замена агентных моделей на модели-суррогаты, основанные на машинном обучении, либо комбинирование этих подходов для сокращения времени симуляции при сохранении точности. **Метод.** Выполнена оценка эффективности различных моделей машинного обучения, используемых в качестве суррогатов агентных моделей. Для обучения этих моделей была сгенерирована база данных, включающая входные и выходные данные, полученные в результате многократных запусков агентных моделей. Проведен анализ различных методов генерации выходных данных агентных моделей: использование одиночных моделей и ансамблей для предсказания значений на каждом временном шаге (подходы независимой множественной регрессии и цепной регрессии), а также гибридные модели, в которых часть выходных данных генерируется агентной моделью и используется в качестве входа для модели машинного обучения, завершающей построение выходной последовательности. **Основные результаты.** Эксперименты показали, что наилучшие результаты в аппроксимации выходной последовательности агентными моделями достигнуты при использовании модели eXtreme Gradient Boosting вместе с цепной регрессией. В гибридной архитектуре наилучшее качество предсказания достигнуто с использованием модели Long Short Term Memory, в которую включены начальные значения выходной последовательности. Данная модель обеспечила наименьшие значения Mean Absolute Error и Root Mean Square Error. **Обсуждение.** Полученные результаты показали потенциал применения моделей машинного обучения в качестве эффективной альтернативы агентным моделям для задач эпидемиологического моделирования, особенно в условиях ограниченного времени и ресурсов. Представленный подход может быть использован для оперативной оценки сценариев распространения инфекции и разработки профилактических стратегий.

**Ключевые слова**

эпидемиология, агентное моделирование, моделирование-суррогат, машинное обучение, множественная регрессия, цепная регрессия

**Благодарности**

Работа выполнена при финансовой поддержке Российского научного фонда в соответствии с Соглашением № 24-25-20104 от 12.04.2024 г. и Санкт-Петербургского государственного автономного учреждения «Фонд поддержки научной, научно-технической, инновационной деятельности» о предоставлении регионального гранта для софинансирования проекта в рамках соглашения № 24-25-20104 от 22.05.2024 г. от 03.06.2024 г. № 169.

**Ссылка для цитирования:** Дарвиш А., Леоненко В.Н. Снижение вычислительных затрат при агентном моделировании распространения респираторной инфекции с помощью суррогатной модели на основе машинного обучения // Научно-технический вестник информационных технологий, механики и оптики. 2026. Т. 26, № 1. С. 177–184 (на англ. яз.). doi: 10.17586/2226-1494-2026-26-1-177-184

## Introduction

Modeling the spread of acute respiratory infections increasingly relies on mechanistic mathematical models that describe the process of pathogen transmission through populations. These models have proven themselves as an effective tool for describing and predicting the spread dynamics of diseases such as influenza and COVID-19. The outbreak of respiratory infections poses an economic and social burden at the community and individual levels. Influenza infects about 5 million people annually and leads to the death of approximately 650,000 people [1]. The COVID-19 pandemic has infected millions of people worldwide and caused a very high mortality rate since 2019 [2]. In this regard, there is a great need for tools capable of modeling the spread of these infections and predicting future outbreaks in order to develop an effective intervention strategy to control and prevent the outbreaks.

There are many methods and models to predict, understand, and explain the dynamics of the epidemic. Statistical models, machine learning models, and deep learning models are particularly useful for prediction by analyzing historical disease data and associated data [3], but they are limited in analyzing, understanding, and explaining the dynamic process [4]. Compartmental models are used to understand, analyze, and predict disease dynamics. They have a simple structure and are computationally efficient in simulating large communities [5]. However, they are ineffective in providing long-term forecasts [4] and lack an accurate description of spread through assumptions about mixing patterns and are limited in capturing individual characteristics and spatial effects [6].

Agent-Based Models (ABMs) are capable of modeling many complex real-world scenarios and have been used effectively to model several behaviors in various domains [7, 8]. ABMs have become a powerful tool in modeling the spread of infectious diseases because of their ability to simulate interactions between individuals, track the spread of infectious diseases in heterogeneous populations, and simulate the effects of different interventions [9, 10]. ABMs can model interactions resulting from interactions between individuals and groups and can capture unexpected random patterns in epidemics [11]. They have the advantage of representing different characteristics of individuals that document variation within populations, allowing the simulation of very complex scenarios that are useful for epidemiologists in designing different experiments that are difficult to perform in the real world [12]. The main problem with using ABMs in epidemiological modeling is the computational complexity resulting from the interactions of individuals within the population, which leads to long running times to obtain results [13]. Computational complexity increases when matching to

real-world data, where running over a large parameter space is required [8, 14].

Surrogate models are one approach used to alleviate the computational limitations of ABMs. Surrogate models can be used to directly approximate the underlying function of an ABM or to explore the parameter space [15]. Recently, surrogate modeling applications have spread rapidly and their use has increased across a wide range of applications from various scientific disciplines [16]. Data-fitting models are a type of model used in surrogate model building [17] in which a computationally intensive model is replaced by a less computationally expensive model that approximates the original model. This is done by mapping the inputs and outputs of the original model using machine learning [18] and deep learning techniques [19].

In Russia, the use of ABMs has shown promising results in reproducing the dynamics of respiratory infections [9, 20]. Gaussian Process Regression (GPR) was used as a surrogate modeling method to reduce simulation time and conduct an uncertainty quantification of ABM that simulate the dynamics of COVID-19 spread in Saint Petersburg [21].

With this paper, we investigate a question raised in the article [21]: whether machine learning methods can be used successfully and efficiently as a method for surrogate modeling of ABM for simulating the dynamics of respiratory infection transmission in Saint Petersburg. The work [21] has suggested this possibility, but has only applied the GPR with changing kernels to the model inputs and outputs, and the modeling has been limited to data generated for COVID-19 infections with fraction of non-immune individuals in general population (*alpha* >= 0.4). In an attempt to stimulate further work on this topic, we compare several different potential methods and approaches, extending the research to include different types of respiratory infections with different values of the fraction of non-immune individuals.

## The ABMs used and the synthetic data

ABMs require a lot of detailed information representing a "synthetic population" in which individuals and the rules governing their behavior and interactions with each other and with the surrounding environment are defined. A synthetic population of the city Saint Petersburg was used [9, 20, 21]. The data consists of a set of tables representing hypothetical individuals and their housing, workplace and school information. A framework responsible for simulation is also required. The disease transmission dynamics in the used ABMs framework follow the model Susceptible-Exposed-Infectious-Recovered. The used ABMs framework is a pre-exist model implemented in Python[1]. The framework models the proportion of susceptible, non-infectious and infectious individuals within a synthetic population. Agents have age, sex, housing, school, and work information that govern their movement within the environment. All young people aged 7–17 go to school, and working-age adults (18–55 for males and 18–60 for females) can work [9, 20].

There are three input parameters to the model. $I_0$: initial number of infected individuals, λ: infection transmission coefficient, and α. The model starts by randomly selecting α from the population of susceptible individuals and then randomly selecting infected individuals according to $I_0$ with the number of days of infection equal to 3. The model then simulates the interaction of individuals on a daily basis. Contact between two agents occurs when the agents interact in the same residence, workplace, or school and the infection is transmitted based on the transmission coefficient and the individual infection rate that changes over time from the moment of infection. The number of days the agent is infected is tracked and the agent begins to transmit the infection on the third day until the eighth day of his infection. Once the agent passes the infectious period, the agent will move to the recovered state.

## Materials and Methods

### Simulation data set

To generate the data, each time a sample of synthetic data is taken, the ABM runs several times for different input values, and the results, namely, the number of infected people are saved. The model was executed on the following input values: sample size of synthetic data $n$ = 500,000, $I_0$ = 10, α within range [0.1, 0.99] are taken with a step of 0.05, λ within range [0.1, 0.99] are taken with a step of 0.1 and number of days of tracking infection cases *days* = 40.

The infection rate per 10,000 populations is then calculated by dividing the number of infected people by the sample size of the population and multiplying by 10,000. The data set was divided into two parts, a 70 % training set, and a 30 % test set.

### Approaches and models

We can describe the main problem as one-to-many prediction problem where a series of values [$y_1$ … $y_{40}$] is estimated from three known values α, λ, and $I_0$. Several models have been evaluated to compare their performance in predicting simulation outputs. These models represent a diverse set of approaches and methods, allowing for a comprehensive evaluation of surrogate modeling techniques under different assumptions and algorithmic complexities.

It is clear that the estimate of the number of infected people any day depends mainly on both α, λ, and prevalence on previous day as well as contact information between individuals and the individual infection rate. Contact information is not available and the prevalence information and individual infection rate are implicitly present in the number of new infections in each previous day.

The learning algorithms were applied in three different modeling approaches: standard, independent multi-output, and regression chaining. In the standard approach, models inherently support multiple-output regression where the model is given inputs and outputs and is trained in the original way to do the mapping between the inputs and outputs. The independent multiple-output modeling approach treated each target day in the output sequence independently with the assumption that there is no inherent correlation among the output sequences as

[1] Influenza_spatial: A spatial model for the spread of influenza Available at: https://github.com/vnleonenko/Influenza_spatial, free. (accessed: 05.02.2025).

shown in Fig. 1, *a*. This approach leverages the simplicity and modularity of training models for individual outputs while providing flexibility in tuning each model to the characteristics of its specific target variable. In contrast, the regression chaining approach explicitly modeled the sequential dependencies among the output variables. Starting with the input features, the first model predicted the initial value of the output sequence $y_1$. Subsequently, the predicted $y_1$ was combined with the input features to predict $y_1$, and this process continued iteratively until the final target value was predicted as illustrated by Fig. 1, *b*. This method takes advantage of potential inter-dependencies between consecutive outputs, thereby improving accuracy in scenarios where such relationships are significant.

The inclusion of approaches allowed for a detailed examination of the trade-offs between treating outputs as independent versus leveraging sequential dependencies. By incorporating a variety of algorithms and modeling strategies, the analysis aimed to identify the most effective surrogate modeling techniques for accurately capturing the dynamics of the original simulation. Linear Model (LM), Random Forest (RF), LightGBM (LGBM), eXtreme Gradient Boosting (XGB), GPR, and Long Short Term Memory (LSTM) models were trained and tested.

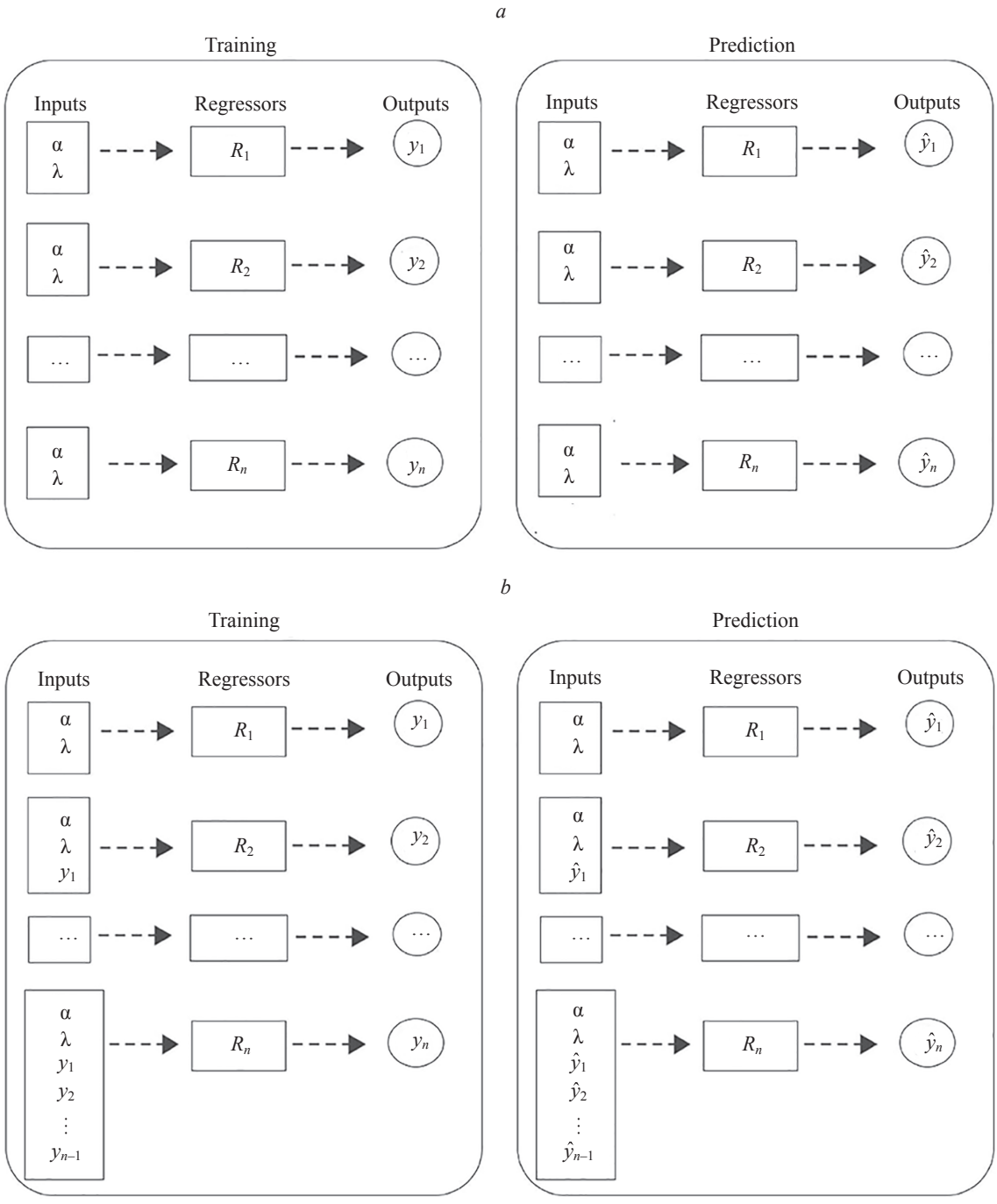


*Fig. 1*. The training and testing process in multi-output and regression chaining approaches: independent multi-output approach (*a*); regression chaining approach (*b*).

*R* denotes regression model, *n* is the index in the output sequence, and $\hat{y}$ represents the model-predicted value

**Metrics**

During the training process the Mean Squared Error (MSE) was minimized while during testing we evaluated the performance of different models using the MAE and the Root Mean Square Error (RMSE) as key performance indicators.

## Results and Discussion

In total, we made 20 different experiments to replicate the behavior of the ABM. All models were implemented with minimal hyper-parameter optimization, and all computations were performed on a personal computer. The metrics of all models are presented in Tables 1 and 2.

**Standard approach**

Three methods (GPR, RF, and XGB) were applied in an original manner. GPR was initially used using additive kernel [21], and the error on the test set was (*MAE* = 30.66, *RMSE* = 78.79). We considered it as the main comparison, because GPR is frequently found in the process-based simulation literature [21, 22]. The results shows that XGB model outperformed the main comparison model (*MAE* = 29.89, *RMSE* = 78.63).

**Independent multi-output approach**

We applied four methods (LM, RF, LGBM and XGB). Table 1 shows that none of the models achieved better results than the main comparison model.

**Regression chaining approach**

The same four methods (LM, RF, LGBM and XGB) were also applied. The results indicate that the XGB model outperforms all other models in replicating the behavior of the ABM (*MAE* = 28.6, *RMSE* = 77.62). Actual valuespredicted incidence and prevalence for the XGB model are shown in Fig. 2, *a*. The figure shows the quality of the XGB model in repeating behavior for daily infections and also for the prevalence calculated from the total daily infections over the previous several days according to ABM used.

**Models across different approaches**

In the LM, both the independent outputs and regression chain approaches yielded the same MAE (30.8) and RMSE (80.31). This uniform performance suggests that the linear model lacks the flexibility to distinguish between sequential dependencies and independent predictions, limiting its effectiveness in approximating ABM behavior. LGBM performance varied significantly between independent multi-output and the regression chaining approach. The increasing error on the regression chaining approach can most likely be explained by the propagation of successive prediction errors. RF performed consistently across standard and independent multi-output approaches. The independent multi-output approach slightly outperformed others (*MAE* = 32.41). However, the independent multi-output approach showed an increase in RMSE metric (*RMSE* = 87.41), likely because the model was optimized to make fewer errors but with larger values.

**Results analysis**

We note from the results shown in Table 1 that the error values are close to each other despite the difference in model and approach. In addition, the linear model within the second approach was able to achieve good results compared to other models. This is what made us explain the difference in results between models and approaches by their ability to benefit from sequence information either implicitly during the first approach or explicitly in the third approach. The XGB model within the third approach was able to achieve a balance by reducing the number and value of errors in predicting the output series and gave the best result among all models and approaches.

The error metric values shown in Table 1 are calculated as the average of the error values for all output sequences. When the error metric values are reviewed for each output sequence in the test set, it is found that 5–10 examples out of 40 examples in the output set have very high error values that cause the average error to increase. These high error producing examples are common across models, but they do not have a distinct pattern, which means that the models

*Table 1.* The MAEs and RMSEs of the testing set for all models applied in three modeling approaches

| Model | MAE | RMSE |
|---|---|---|
| Standard | | |
| GPR | 30.66 | 78.79 |
| RF | 32.44 | 87.29 |
| XGB | **29.89** | **78.63** |
| Independent multi-output | | |
| LM | **30.80** | **80.31** |
| RF | 32.41 | 87.41 |
| LGBM | 32.15 | 86.32 |
| XGB | 32.60 | 87.58 |
| Regression chaining | | |
| LM | 30.80 | 80.31 |
| RF | 32.77 | 87.88 |
| LGBM | 35.70 | 99.55 |
| XGB | **28.60** | **77.62** |

*Table 2.* The MAEs and RMSEs of the testing set for all models applied in extended modeling approach

| Model | MAE | RMSE |
|---|---|---|
| Expand by the first three time steps | | |
| GPR | 11.44 | 29.12 |
| XGB | 12.55 | 34.38 |
| LSTM | **11.37** | **28.67** |
| Expand by the first four time steps | | |
| GPR | 10.43 | 26.25 |
| XGB | 11.77 | 32.84 |
| LSTM | **9.39** | **23.60** |
| Expand by the first five time steps | | |
| GPR | 8.90 | 21.65 |
| XGB | 10.71 | 29.22 |
| LSTM | **7.65** | **19.93** |

*Note.* Bold values indicate the lowest MAE and RMSE across models.

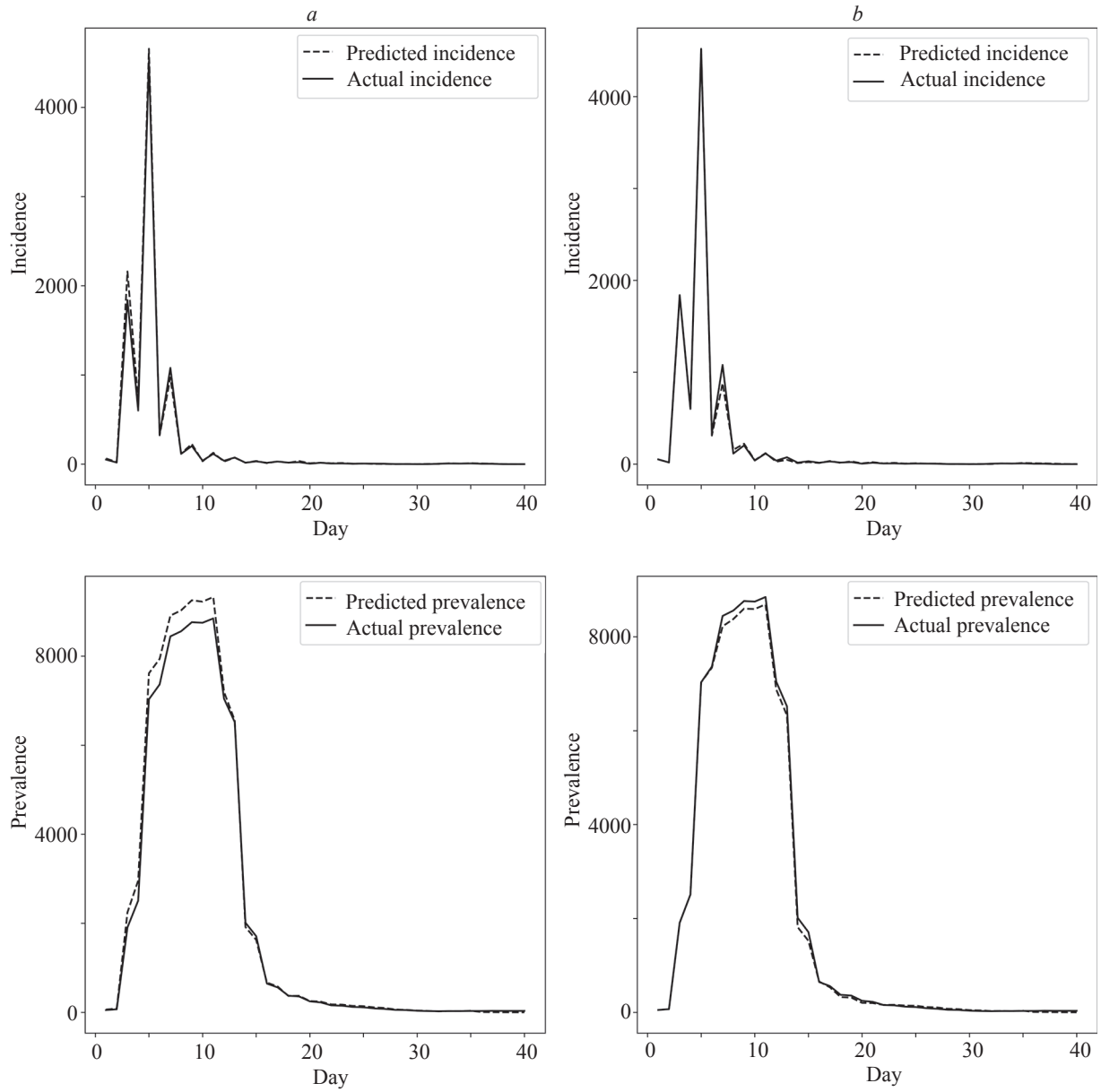


*Fig. 2*. Actual incidence and prevalence for one example generated by ABM (α = 0.95, λ = 0.6) values predicted incidence and prevalence: prediction using XGB regression chaining approach (*a*), prediction using extended models with the first five time steps (*b*)

do not fail to learn the critical dynamics for a given value of the input parameters. Furthermore, we observed that the error values in each sequence were caused by the prediction error at the beginning of the sequence between the third and tenth day, which is the period representing the highest infection values, and therefore the period that requires a high calculation time.

**Extended model**

Analysis of the results led us to test the output sequence prediction by using the first part of the sequence as input to the model. Three experiments were performed, the first by adding the first three values of the sequence, then the first four values, then the first five values. For each experiment, the GPR model from the first approach, the XGB model from the third approach, and the LSTM model with teacher forcing were used. The metrics of all models are presented in Table 2. The choice of LSTM is justified by its ability to effectively model temporal dynamics and long-term dependencies, features that are particularly relevant given the sequential nature of the problem. The teacher forcing technique was integrated to stabilize the training process by mitigating the accumulation of prediction errors during sequence generation.

As expected, the results in Table 2 show a significant improvement in modeling in all tested models with close error values, with the LSTM model outperforming the other models in all cases. Actual values predicted incidence and prevalence for extended models with the first five time steps are shown in Fig. 2, *b*. The figure shows an improvement in the accuracy of repeating the behavior for daily incidence and also for prevalence.

**Recommended model**

By monitoring the performance of the ABM during execution, it is observed that the values of the first days of the of the simulated output sequence are generated in less time than the rest of the sequence, as the number of

infected people is small and thus the number of simulations is smaller. Therefore, we propose to combine the ABM with the extended LSTM model with $n$-time steps ($n = 3, 4, 5$) so that we get a trade-off between the accuracy of the results and the required time.

## Conclusion

In this proof-of-concept work, we compare several machine learning-based models and approaches for Agent-Based Model (ABM) surrogate modeling simulating the spread of respiratory infection in Saint Petersburg in order to identify the most suitable methods and approaches as surrogates for modeling the complex behavior of ABMs. We conclude that machine learning-based surrogate models can replicate the behavior of the ABM model under study, suggesting that they can be used in a practical and straightforward manner. The proposed model is a case study, and we do not claim that these results will apply to the ABM model when used on population-wide data. Therefore, our plan in our future work is to evaluate the performance on simulation output sequences resulting from the application to the population as a whole.

## References


1. Wei S., Lin S., Zhao W., Song S., Yang Y., He Y., Shu Z., Zhong L., Ti L. The prediction of influenza-like illness using national influenza surveillance data and Baidu query data. *BMC Public Health*, 2024, vol. 24, no. 1, pp. 513. https://doi.org/10.1186/s12889-024-17978-0
2. Wang H., Paulson K.R., Pease S.A., Watson S., Comfort H., Zheng P., et al. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *The Lancet*, 2022, vol. 399, no. 10334, pp. 1513–1536. https://doi.org/10.1016/S0140-6736(21)02796-3
3. Wang P., Zheng X., Li J., Zhu B. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals*, 2020, vol. 139, pp. 110058. https://doi.org/10.1016/j.chaos.2020.110058
4. Rakhshan S.A., Nejad M.S., Zaj M., Ghane F.H. Global analysis and prediction scenario of infectious outbreaks by recurrent dynamic model and machine learning models: A case study on COVID-19. *Computers in Biology and Medicine*, 2023, vol. 158, pp. 106817. https://doi.org/10.1016/j.compbiomed.2023.106817
5. Augustin C.M., Gsell M.A.F., Karabelas E., Willemen E., Prinzen F.W., Lumens J., Vigmond E.J., Plank G. A computationally efficient physiologically comprehensive 3D–0D closed-loop model of the heart and circulation. *Computer Methods in Applied Mechanics and Engineering*, 2021, vol. 386, pp. 114092. https://doi.org/10.1016/j.cma.2021.114092
6. Sturniolo S., Waites W., Colbourn T., Manheim D., Panovska-Griffiths J. Testing, tracing and isolation in compartmental models. *PLoS Computational Biology*, 2021, vol. 17, no. 3, pp. e1008633. https://doi.org/10.1371/journal.pcbi.1008633
7. Li J., Rombaut E., Vanhaverbeke L. A systematic review of agent-based models for autonomous vehicles in urban mobility and logistics: Possibilities for integrated simulation models. *Computers, Environment and Urban Systems*, 2021, vol. 89, pp. 101686. https://doi.org/10.1016/j.compenvurbsys.2021.101686
8. Dehkordi M.A.E., Lechner J., Ghorbani A., Nikolic I., Chappin E., Herder P. Using machine learning for agent specifications in agent-based models and simulations: A critical review and guidelines. *Journal of Artificial Societies and Social Simulation*, 2023, vol. 26, no. 1, pp. 9. https://doi.org/10.18564/jasss.5016
9. Leonenko V., Arzamastsev S., Bobashev G. Contact patterns and influenza outbreaks in Russian cities: A proof-of-concept study via agent-based modeling. *Journal of Computational Science*, 2020, vol. 44, pp. 101156. https://doi.org/10.1016/j.jocs.2020.101156
10. Pereira F.H., Schimit P.H.T., Bezerra F.E. A deep learning based surrogate model for the parameter identification problem in probabilistic cellular automaton epidemic models. *Computer Methods and Programs in Biomedicine*, 2021, vol. 205, pp. 106078. https://doi.org/10.1016/j.cmpb.2021.106078
11. Hunter E., Mac Namee B., Kelleher J.D. A taxonomy for agent-based models in human infectious disease epidemiology. *Journal of Artificial Societies and Social Simulation*, 2017, V. 20. N 3. P. 2. https://doi.org/10.18564/jasss.3414
12. Perumal R., van Zyl T.L. Surrogate-assisted strategies: the parameterisation of an infectious disease agent-based model. *Neural Computing and Applications*, 2025, vol. 37, no. 2, pp. 627–638. https://doi.org/10.1007/s00521-022-07476-y
13. Zhang Y., Li Z., Zhang Y. Validation and calibration of an agent-based model: A surrogate approach. *Discrete Dynamics in Nature and*


## Литература


1. Wei S., Lin S., Zhao W., Song S., Yang Y., He Y., Shu Z., Zhong L., Ti L. The prediction of influenza-like illness using national influenza surveillance data and Baidu query data // BMC Public Health. 2024. V. 24. N 1. P. 513. https://doi.org/10.1186/s12889-024-17978-0
2. Wang H., Paulson K.R., Pease S.A., Watson S., Comfort H., Zheng P., et al. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21 // The Lancet. 2022. V. 399. N 10334. P. 1513–1536. https://doi.org/10.1016/S0140-6736(21)02796-3
3. Wang P., Zheng X., Li J., Zhu B. Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics // Chaos, Solitons & Fractals. 2020. V. 139. P. 110058. https://doi.org/10.1016/j.chaos.2020.110058
4. Rakhshan S.A., Nejad M.S., Zaj M., Ghane F.H. Global analysis and prediction scenario of infectious outbreaks by recurrent dynamic model and machine learning models: A case study on COVID-19 // Computers in Biology and Medicine. 2023. V. 158. P. 106817. https://doi.org/10.1016/j.compbiomed.2023.106817
5. Augustin C.M., Gsell M.A.F., Karabelas E., Willemen E., Prinzen F.W., Lumens J., Vigmond E.J., Plank G. A computationally efficient physiologically comprehensive 3D–0D closed-loop model of the heart and circulation // Computer Methods in Applied Mechanics and Engineering. 2021. V. 386. P. 114092. https://doi.org/10.1016/j.cma.2021.114092
6. Sturniolo S., Waites W., Colbourn T., Manheim D., Panovska-Griffiths J. Testing, tracing and isolation in compartmental models // PLoS Computational Biology. 2021. V. 17. N 3. P. e1008633. https://doi.org/10.1371/journal.pcbi.1008633
7. Li J., Rombaut E., Vanhaverbeke L. A systematic review of agent-based models for autonomous vehicles in urban mobility and logistics: Possibilities for integrated simulation models // Computers, Environment and Urban Systems. 2021. V. 89. P. 101686. https://doi.org/10.1016/j.compenvurbsys.2021.101686
8. Dehkordi M.A.E., Lechner J., Ghorbani A., Nikolic I., Chappin E., Herder P. Using machine learning for agent specifications in agent-based models and simulations: A critical review and guidelines // Journal of Artificial Societies and Social Simulation. 2023. V. 26. N 1. P. 9. https://doi.org/10.18564/jasss.5016
9. Leonenko V., Arzamastsev S., Bobashev G. Contact patterns and influenza outbreaks in Russian cities: A proof-of-concept study via agent-based modeling // Journal of Computational Science. 2020. V. 44. P. 101156. https://doi.org/10.1016/j.jocs.2020.101156
10. Pereira F.H., Schimit P.H.T., Bezerra F.E. A deep learning based surrogate model for the parameter identification problem in probabilistic cellular automaton epidemic models // Computer Methods and Programs in Biomedicine. 2021. V. 205. P. 106078. https://doi.org/10.1016/j.cmpb.2021.106078
11. Hunter E., Mac Namee B., Kelleher J.D. A taxonomy for agent-based models in human infectious disease epidemiology // Journal of Artificial Societies and Social Simulation. 2017. V. 20. N 3. P. 2. https://doi.org/10.18564/jasss.3414
12. Perumal R., van Zyl T.L. Surrogate-assisted strategies: the parameterisation of an infectious disease agent-based model // Neural Computing and Applications. 2025. V. 37. N 2. P. 627–638. https://doi.org/10.1007/s00521-022-07476-y
13. Zhang Y., Li Z., Zhang Y. Validation and calibration of an agent-based model: A surrogate approach // Discrete Dynamics in Nature and

*Society*, 2020, vol. 2020, pp. 6946370. https://doi.org/10.1155/2020/6946370
14. Miksch F., Jahn B., Espinosa K.J., Chhatwal J., Siebert U., Popper N. Why should we apply ABM for decision analysis for infectious diseases?—An example for dengue interventions. *PLoS One*, 2019, vol. 14, no. 8, pp. e0221564. https://doi.org/10.1371/journal.pone.0221564
15. Lamperti F., Roventini A., Sani A. Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, 2018, vol. 90, pp. 366–389. https://doi.org/10.1016/j.jedc.2018.03.011
16. Silionis N.E., Liangou T., Anyfantis K.N. Deep learning-based surrogate models for spatial field solution reconstruction and uncertainty quantification in Structural Health Monitoring applications. *Computers & Structures*, 2024, vol. 301, pp. 107462. https://doi.org/10.1016/j.compstruc.2024.107462
17. Peherstorfer B., Willcox K., Gunzburger M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, 2018, vol. 60, no. 3, pp. 550–591. https://doi.org/10.1137/16M1082469
18. Hamrani A., Akbarzadeh A., Madramootoo C.A., Bouarab F.Z. Machine learning surrogate modeling for meshless methods: Leveraging universal approximation. *International Journal of Computational Methods*, 2023, vol. 20, no. 08, pp. 2141022. https://doi.org/10.1142/S021987622141022X
19. Jeong Y., Lee S.I., Lee J., Choi W. Data-efficient surrogate modeling using meta-learning and physics-informed deep learning approaches. *Expert Systems with Applications*, 2024, vol. 250, pp. 123758. https://doi.org/10.1016/j.eswa.2024.123758
20. Leonenko V., Lobachev A., Bobashev G. Spatial modeling of influenza outbreaks in Saint Petersburg using synthetic populations. *Lecture Notes in Computer Science*, 2019, vol. 11536, pp. 492–505. https://doi.org/10.1007/978-3-030-22734-0_36
21. Matveeva A., Leonenko V. Application of Gaussian process regression as a surrogate modeling method to assess the dynamics of COVID-19 propagation. *Procedia Computer Science*, 2022, vol. 212, pp. 340–347. https://doi.org/10.1016/j.procs.2022.11.018
22. Angione C., Silverman E., Yaneske E. Using machine learning as a surrogate model for agent-based simulations. *PLoS One*, 2022, vol. 17, no. 2, pp. e0263150. https://doi.org/10.1371/journal.pone.0263150

Society. 2020. V. 2020. P. 6946370. https://doi.org/10.1155/2020/6946370
14. Miksch F., Jahn B., Espinosa K.J., Chhatwal J., Siebert U., Popper N. Why should we apply ABM for decision analysis for infectious diseases?—An example for dengue interventions // PLoS One. 2019. V. 14. N 8. P. e0221564. https://doi.org/10.1371/journal.pone.0221564
15. Lamperti F., Roventini A., Sani A. Agent-based model calibration using machine learning surrogates // Journal of Economic Dynamics and Control. 2018. V. 90. P. 366–389. https://doi.org/10.1016/j.jedc.2018.03.011
16. Silionis N.E., Liangou T., Anyfantis K.N. Deep learning-based surrogate models for spatial field solution reconstruction and uncertainty quantification in Structural Health Monitoring applications // Computers & Structures. 2024. V. 301. P. 107462. https://doi.org/10.1016/j.compstruc.2024.107462
17. Peherstorfer B., Willcox K., Gunzburger M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization // SIAM Review. 2018. V. 60. N 3. P. 550–591. https://doi.org/10.1137/16M1082469
18. Hamrani A., Akbarzadeh A., Madramootoo C.A., Bouarab F.Z. Machine learning surrogate modeling for meshless methods: Leveraging universal approximation // International Journal of Computational Methods. 2023. V. 20. N 08. P. 2141022. https://doi.org/10.1142/S021987622141022X
19. Jeong Y., Lee S.I., Lee J., Choi W. Data-efficient surrogate modeling using meta-learning and physics-informed deep learning approaches // Expert Systems with Applications. 2024. V. 250. P. 123758. https://doi.org/10.1016/j.eswa.2024.123758
20. Leonenko V., Lobachev A., Bobashev G. Spatial modeling of influenza outbreaks in Saint Petersburg using synthetic populations // Lecture Notes in Computer Science. 2019. V. 11536. P. 492–505. https://doi.org/10.1007/978-3-030-22734-0_36
21. Matveeva A., Leonenko V. Application of Gaussian process regression as a surrogate modeling method to assess the dynamics of COVID-19 propagation // Procedia Computer Science. 2022. V. 212. P. 340–347. https://doi.org/10.1016/j.procs.2022.11.018
22. Angione C., Silverman E., Yaneske E. Using machine learning as a surrogate model for agent-based simulations // PLoS One. 2022. V. 17. N 2. P. e0263150. https://doi.org/10.1371/journal.pone.0263150

## Authors

**Ali Darwish** — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0007-3123-3789, adarwish@itmo.ru, darwishali320@gmail.com
**Vasiliy N. Leonenko** — PhD (Physics & Mathematics), Researcher, Smorodintsev Research Institute of Influenza, Saint Petersburg, 197022, Russian Federation; Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 31767480200, https://orcid.org/0000-0001-7070-6584, vnleonenko@itmo.ru

## Авторы

**Дарвиш Али** — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0007-3123-3789, adarwish@itmo.ru, darwishali320@gmail.com
**Леоненко Василий Николаевич** — кандидат физико-математических наук, научный сотрудник, НИИ гриппа им. А.А. Смородинцева, Санкт-Петербург, 197022, Российская Федерация; доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 31767480200, https://orcid.org/0000-0001-7070-6584, vnleonenko@itmo.ru

*Received 02.05.2025*
*Approved after reviewing 22.11.2025*
*Accepted 20.01.2026*

*Статья поступила в редакцию 02.05.2025*
*Одобрена после рецензирования 22.11.2025*
*Принята к печати 20.01.2026*