Bulletin of the *Transilvania* University of Braşov
Series III: Mathematics and Computer Science, Vol. 5(67), No. 1 - 2025, 241-258
https://doi.org/10.31926/but.mif.2025.5.67.1.18

EXPLOITING CELLULAR AUTOMATA AND LINEAR REGRESSION TO PREDICT DISEASE SPREAD

Alexandra BĂICOIANU [1], Rareş IVAN [2] and Iulian POPA [*,3]

Abstract

The Covid-19 epidemic has significantly impacted the world, highlighting the urgent need to understand and anticipate its mechanisms of spread. This essential knowledge is necessary to plan and conduct an immediate and adequate public health response. This study presents an approach to modeling the spread of Covid-19 using non-uniform cellular automata. The paper extends the application of a previously developed model that uses cellular automata and the SIRD epidemiological model for predicting the evolution of Covid-19. Originally developed for a different context, the model is now adapted to assess the progression of the pandemic in Germany and Italy, considering the potential impact of neighboring countries on the spread of the epidemic. Additionally, this approach expands the prediction model to countries lacking infection data (Switzerland) by using estimated parameters from neighboring countries and randomly initialized parameters for the target country. The study demonstrates the model's precision in tracking infection rates over time by employing reliable public data sources such as the World Health Organization and existing information from national health websites. The study not only furnishes valuable insights into the regional distribution of the epidemic's impact but also makes a significant contribution by extending the model's application beyond the borders of a single country. It introduces a strategy for extrapolating patterns of infection spread across borders, marking it as the first study of its kind with substantial importance in the field.

2020 *Mathematics Subject Classification:* 68Q80, 68T20, 92B20.

Key words: predictive modeling, public health, cellular automata, SIRD model, epidemic evolution, regional distribution, infection rate, Covid-19.

[1]Department of Mathematics and Computer Science, Faculty of Mathematics and Computer Science, *Transilvania* University of Braşov, Romania, e-mail: a.baicoianu@unitbv.ro

[2]Former student at Faculty of Mathematics and Computer Science, *Transilvania* University of Braşov, Romania, e-mail: ivanrarescosmin@gmail.com

[3*] *Corresponding author*, Ph.D. Student at Faculty of Mathematics and Computer Science, *Transilvania* University of Braşov, Romania, e-mail: iulian.popa@unitbv.ro

1 Introduction

The emergence of the unprecedented Covid-19 pandemic has highlighted the importance of accurate and reliable epidemiological models for predicting and controlling the spread of infectious diseases. These models serve as crucial tools for informing public health policies and actions. However, traditional epidemiological models, such as the Susceptible-Infectious-Recovered (SIR) model [1], [6] and its extensions, SEIR, SLIR, SIRD, SEIRU, SLIAR, have shown certain limitations in the face of the global Covid-19 crisis. For example, these models do not consider the spatial distribution of population and disease, which can significantly impact the spread of the disease [13].

Given the spatially heterogeneous nature of the Covid-19 pandemic, with uneven disease spread across regions and countries, models are urgently needed to capture this spatial heterogeneity and the complex interactions between regions. This needs to lead us to explore using cellular automata as a modeling tool to predict the spread of Covid-19.

Cellular automata, mathematical models first introduced by von Neumann and Ulam in the 1950s, are well suited for simulating systems with complex spatial interactions. In a cellular automaton, space is discretized into cells, with each cell following a deterministic rule that considers the states of neighbouring cells. Local interactions between cells can lead to complex global behaviour, making cellular automata an excellent tool for modeling complex systems such as infectious disease dynamics. However, conventional cellular automata models assume uniformity between cells, an assumption that is rarely met in real-world scenarios. Therefore, we chose non-uniform cellular automata, in which each cell (or region) is unique and behaves according to its characteristics. This non-uniformity allows for incorporating geographic variations and real-world population density, thus providing a more realistic model of Covid-19 spread.

Predictive modeling of health emergencies has progressed, with previous studies focusing on epidemic trends in China [8], [16]. The present research extends this work to Germany, introducing two key considerations: their distinct geographical and demographic characteristics and the influence of neighbouring countries on epidemic spread. Existing predictive models, including machine learning and regression models [1], [2], often overlook the geographical context. The analysis from [4] notes that the errors of models that do not consider spatial elements are between 13% and 225%, which is enormous. In addition, [5] asserts that we need space elements for a better-performing model. This study fills this gap by focusing on the geographical nuance of epidemic progression. Focusing on Covid-19 in Germany, a country known for its strong healthcare infrastructure and interregional mobility, we also examine how Germany, Italy, Austria and France might affect Switzerland. This approach improves our understanding of epidemics by discussing the impact of neighbouring regions.

In our solution, we merge cellular automata principles with an epidemiological model to simulate and predict the spread of Covid-19. The model uses a custom cellular automata map, formed by dividing a country's map into regions. Each

region forms one or more cells in our cellular automata model, with the daily number of active Covid-19 cases in each region serving as the primary data source for prediction. Prediction is performed using multiple linear regression, a statistical technique that predicts a dependent variable based on two or more independent variables [15]. In our model, the future number of Covid-19 cases forms the dependent variable, and the current state of a region and the states of neighbouring regions form the independent variables. This method allows each cell's prediction to be influenced not only by its own state but also by neighbouring cells' states, mimicking the disease transmission process in the real world.

The capability of the model extends beyond the borders of a single country. It can simulate the spread of Covid-19 from two countries into a common neighbouring country. This particular feature of cross-border modeling allows a broader regional perspective on the pandemic, capturing the dynamics of the spread of the disease across national borders. This is particularly important for global public health planning, where understanding the cross-border transmission of the disease is essential for coordinating international responses and mitigation strategies.

2 Materials and methods

There are several essential components that support the conduct of this current study. The fundamental aspect is to understand the predictive modeling systems, which serve as the basis for designing research, building models, validating, and assessing their performance [1],[3],[10]. Central to this study is the Susceptible-Infectious-Recovered-Deceased (SIRD) model, which serves as cornerstone of the experiments. This epidemiological model categorizes the population into distinct categories, namely susceptible, infected, recovered, and deceased. These categories offer crucial parameters necessary for the functioning of cellular automata for the prediction of disease spread [6], [11].

The cellular automata [18], [19] are network-like structures used to encapsulate distinct stages of the epidemic. Each cell represents a population subset characterized by health statuses, encompassing susceptibility, infection, recovery, or mortality. All this data, available at [20], on the evolution of the epidemic was collected from reliable sources such as the World Health Organization. They provide a nuanced picture of disease spread's geographical and temporal dynamics. In addition, another essential part of the model is represented by the linear regression for predictive analysis. By delineating correlations and estimating outcomes based on input data, linear regression help to determine the influence of neighbouring regions on the propagation dynamics of the epidemic [10], [15]. Overall, a comprehensive understanding of these concepts will serve as a guide and provide a solid foundation for comprehending the fundamentals of the following model.

2.1 Specificity of cellular automata

Creating a cellular automaton for a country

Creating the cellular automata space involves several computational and spatial analyses that ensure an accurate and practical representation of the geographical area in question. The key is to transpose a real-world map into a spatial space that symbolizes, through each cell, a specific part of the territory while accurately preserving the structural and spatial relationships between the different regions. The selection of cell size is an essential aspect of this process. This is determined based on the size of the smallest geographical region that can fit into a single cell. It should be noted that this is only sometimes the physically smallest region. The underlying reason is the potential for significant disparities in different regions' sizes. If some regions are extremely small, they are combined with one of the adjacent regions. This approach is not strictly standardized and is usually made ad-hoc, ensuring each region is properly represented in the cellular automata matrix.

Considering a map of Germany, see Figure 1 a) - The map of Germany, divided into regions, using the Mercator projection, this is transformed into a cellular space, with small regions combined with neighbouring ones. A Python script then overlays a grid of equally spaced vertical and horizontal lines on the chosen map, dividing it into cells. Selecting the dimensions of the cells is a crucial aspect of this process. This is determined based on the size of the smallest geographical region that can fit within a single cell. It is important to note that this is not always the smallest region in a physical sense. The rationale behind this is the potential for significant disparities in the size of different regions. In cases where some regions are extremely small, they are combined with one of the adjacent regions. This approach is not strictly standardized and is usually performed ad hoc, ensuring that each region is appropriately represented in the cellular automaton grid. By using this method, we ensure a balance between the resolution of the cellular grid and the practical representation of geographical regions. The result is shown in Figure 1 b) and this matrix, with values from 0 (outside the country) to the total number of regions, is $19X14$, with other unused cells outside the borders. The process requires precision, reproducing geographic space into cell space.

The corresponding details for each region, including cell ID, $region$, and $CellCount$ occupied by them, are kept in a CSV file. This file helps for easy reference and efficient data management, see Table 1. Creating a cellular automaton space is a complicated process that requires meticulous attention to detail and precise rendering of the geographic space within the cellular frame boundaries. As seen in Germany, the process can require manual intervention and ad hoc decisions, especially when dealing with smaller regions.

Note that the cellular automata space's accuracy depends on the original map's quality and the careful division and assignment of cells to regions. Therefore, periodic quality reviews and checks are necessary to maintain the integrity and accuracy of the cellular automata space. As the scale and complexity of the area

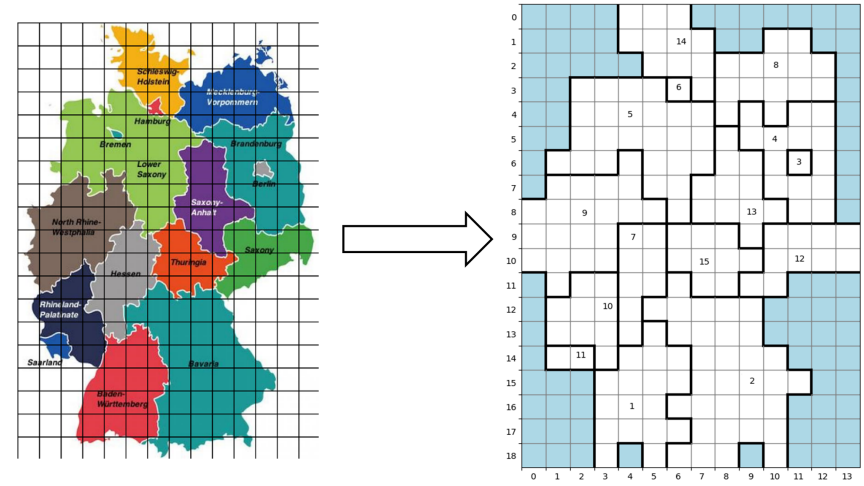

Figure 1: Map of Germany (a) transformed into the cellular space (b)

Table 1: The arrangement of the 15 regions in the cellular space

ID	Region	CellCount	ID	Region	CellCount
1	Baden-Wurttemberg	17	**9**	Nordrhein-Westfalen	20
2	Bayern	32	**10**	Rheinland-Pfalz	9
3	Berlin	1	**11**	Saarland	2
4	Brandenburg	14	**12**	Sachsen	10
5	Bremen, Niedersachsen	25	**13**	Sachsen-Anhalt	11
6	Hamburg	1	**14**	Schleswig-Holstein	11
7	Hessen	8	**15**	Thuringen	9
8	Mecklenburg-Vorpommern	14			

under consideration increases, more sophisticated and automated methods may be required to manage the generation and manipulation of the cellular automata space. However, the basic principles and steps described here would remain the same, ensuring a consistent approach to creating cellular automaton spaces for different countries or regions.

Creating the cellular automaton for multiple countries

To create the cell space for several countries, individual matrices are combined into one, considering different conditions, such as the lack of regional data for a particular country.

After generating each country's individual cellular automata matrices, the next essential step is to amalgamate them into a single composite matrix. In order to proceed with this step, it is essential to ensure uniform cell dimensions across all countries. The initial step in this process involves identifying the smallest region,

which will determine the standard cell size. Due to the irregularity of regional borders, some cells may contains multiple regions. In such cases, the cell will be assigned based on the region occupying the largest area.

Integrating these matrices requires carefully analyzing these countries' geographical alignment and adjacency. The matrices are merged in an ad-hoc manner, ensuring that geographical coherence is maintained. This involves aligning each country's individual cellular automatic matrices so that their relative geographic positions and boundaries accurately intersect in the combined matrix.

Germany, Italy, Austria, France and Switzerland are combined into a single matrix to study cross-border phenomena, see Figure 2. The merging of matrices

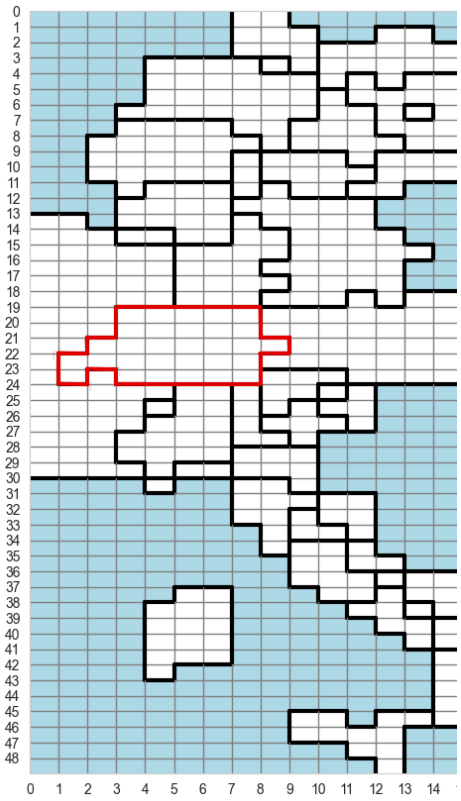

Figure 2: The resulting cellular automaton space derived from combining matrices

is done ad-hoc, ensuring that geographical consistency is maintained. This involves aligning each country's individual cellular automatic matrices so that their relative geographic positions and boundaries intersect precisely in the combined matrix. In the final combined matrix, Figure 2, each cell continues to represent a specific region in Germany and Italy or a single region in Switzerland, Austria and France. This allows us to study and simulate data spread across these countries and to gain insights into how variables or phenomena might cross national borders. This process, while requiring careful planning and precise execution, pro-

vides a robust and flexible framework for studying larger geographical areas that cross national borders. Despite its complexity, creating such a combined space for cellular automata provides a valuable cross-border modeling and analysis tool.

SIRD model

The system proposed in [8] focuses on cells in a fixed two-dimensional array, each representing the number of infected individuals at a given time step. This is based on a discrete version of the SIRD model, which estimates changes in the number of infections. The general equation of the SIRD model is modified to include interactions between neighbouring cells in the grid. Each cell has four adjacent cells, which are assigned unique interaction coefficients. These coefficients thus play a significant role in determining current and future infection levels within the cell and in neighbouring cells, see Figure 3. The significance

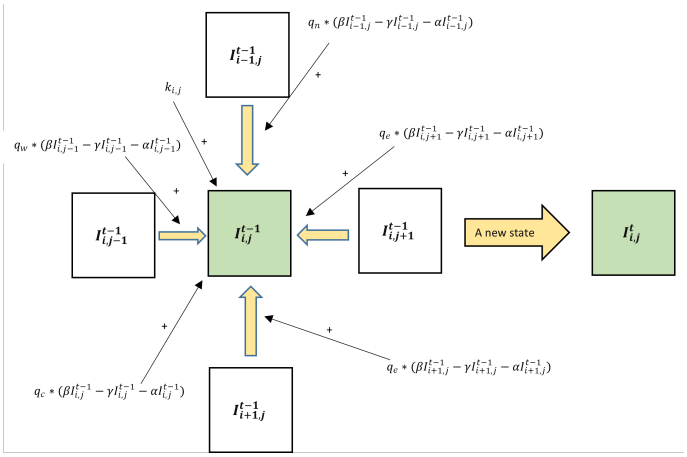

Figure 3: Cellular automaton transition diagram model

of each variable in the equations given in the Figure 3 is justified and explained in section 3.2 of the source mentioned. We have time series data, which shows the number of people infected over a period of time. Each cell has a sequence of values, each representing the number of infected people in that cell at each point in time. Using this data and referring to Figure 3, which explains how a new state is formed, a system of linear equations was constructed for each cell, stated in matrix form [8]:

$$\Delta I_{i,j} = P_{i,j} q_{i,j}$$

where:

$$\Delta I_{i,j} = [\Delta I^1_{i,j}, \Delta I^2_{i,j}, \ldots, \Delta I^N_{i,j}]^T$$

is a N-dimensional vector with each element $\Delta I^t_{i,j} = I^t_{i,j} - I^{t-1}_{i,j}, t = 1, \ldots, N$ $P_{i,j}$ is a $NX6$ dimensional matrix where each row t is

$$[1, P^{t-1}_{c,i,j}, P^{t-1}_{n,i,j}, P^{t-1}_{e,i,j}, P^{t-1}_{s,i,j}, P^{t-1}_{w,i,j}]^T$$

$q_{i,j}$ is a 6-dimensional vector with

$$[k_{i,j}, q_{c,i,j}, q_{n,i,j}, q_{e,i,j}, q_{s,i,j}, q_{w,i,j}]^T.$$

In these equations, $\Delta I_{i,j}$ is a vector demonstrating the change in the number of infections, $P_{i,j}$ is a matrix with information about the current state of the cell and its neighbors, and $q_{i,j}$ is a vector containing the coefficients to be optimized.

The goal is to find the values of $q_{i,j}$ for all cells (i, j) such that the estimated changes in the number of infected individuals at time t, taking into account interactions with neighbouring cells, represented by $P_{i,j}$, accurately approximate the actual changes represented by $\Delta I_{i,j}$. This is treated as a multiple linear regression problem, aiming to optimize $q_{i,j}$.

This method facilitates a reliable prediction of disease spread. Note that the applied equations are taken from [8], and interpreting a new state, as explained in Figure 3, is crucial for understanding the proposed model.

2.2 Prediction model

The assumed methodology makes regional data adjustments to the model based on a cellular automaton for infection modeling. Infected individuals are estimated for each cell, and prediction models are modified to predict the spread of infection using a set of equations ([19]-[21]) derived from the previous study [8]. Figure 4 represents the steps taken for a prediction.

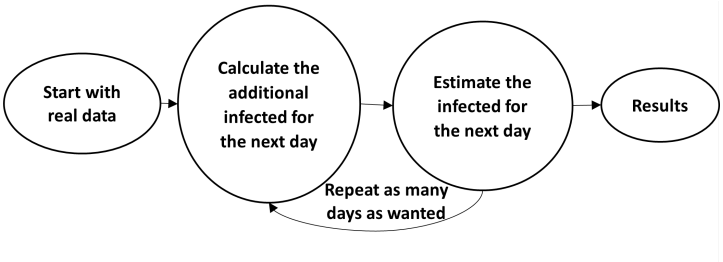

Figure 4: The steps from prediction model

In this study, the approach employed is multiple linear regression, a statistical technique used to model the relationship between a dependent variable and multiple independent variables. Unlike simple linear regression, which involves only one independent variable, multiple linear regression can handle two or more predictors, making it a powerful tool for understanding complex relationships in data. The general form of the multiple linear regression equation is:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \ldots + \beta_n * X_n + e$$

Here, Y represents the dependent variable, β_0 is the y-intercept, $\beta_1, \ldots, \beta_n$ are the coefficients for the independent variables $X_1, \ldots, X_n$ and e is the error term. Multiple linear regression is chosen for its ability to handle complex datasets and

deliver meaningful insights in various fields such as economics, finance, social sciences, and natural sciences.

Building on the discussion of our multiple linear regression and cellular automata-based prediction models, we consider a situation where no data exist for a given country, but data are available for its neighbours. In this scenario, we use Switzerland as the country of interest and Italy, Germany, Austria and France as neighbouring countries with data. As a preliminary setup, it is assumed that the q-values, i.e., the parameters for the multiple regression model, are known for neighbouring countries and are estimated from the available infection data using the multiple linear regression model. However, no data are available for Switzerland to estimate these q-values.

To address this issue, we initialize q values for Switzerland randomly. This is a viable strategy as it is considered that infections may spread from Italy, Germany, Austria and France, which share borders with Switzerland. Transmission of infections from these neighbours to Switzerland would account for a proportion of all infections in Switzerland. Since the q values for Switzerland are random, this means that although the model will generate predictions for the spread of infections in this country, these predictions will be inherently less reliable than those for neighbouring countries, for which we have actual infection data and estimated q parameters . Concisely, this approach extends the prediction model to countries with missing infection data, using estimated parameters from neighbouring countries and randomly initialized parameters for the target country. This method provides a strategy for extrapolating patterns of infection spread across borders, although the accuracy of such predictions could be better due to the lack of data-driven parameter estimation.

Adapting the prediction model to another epidemic situation is relatively straightforward, given the flexible nature of the tools and techniques used. The basis of the model, cellular automata, is highly adaptable and capable of representing a variety of scenarios. Although the process of transforming a country map into a cellular automaton requires some flexibility and creativity due to its ad-hoc nature, it can be done regardless of the country or region of interest.

The parameters of the SIRD epidemiological model - mortality rate, infection rate, and incubation period - would require adjustments based on the specific characteristics of the new epidemic. These parameters are often derived from empirical studies and are usually readily available in the early stages of an epidemic. In addition, changing a country's regional data (deceased, cured, and infected persons) is simple. Public health organizations and ministries usually provide this information and should be available in standardized formats such as CSV files. The template is designed to handle this type of data, which means it can be reconfigured with minimal effort. However, the effectiveness and accuracy of the model would depend on the quality of the data and the specific characteristics of the epidemic, including factors such as public health interventions, social behaviour, and virus characteristics. Therefore, although the technical process of model fitting is simple, a comprehensive and accurate prediction may require a nuanced understanding of the specific epidemic context.

The model is also flexible to adapt to situations where infection data are not available for a particular country but are available for neighbouring countries, bringing innovative points to this strategy. In such scenarios, estimated parameters from neighbouring countries with data are used and randomly initialized for the country of interest. This allows the model to generate predictions of the spread of infection for that country, albeit with a lower degree of accuracy due to the random nature of the initialization.

The adaptive nature of the model extends to situations involving different epidemics. Modifying regional data (dead, recovered, infected) is straightforward, as health organizations usually provide this information in standardized formats. The parameters of the SIRD epidemiological model (mortality rate, infection rate, and incubation period) are also easy to modify. However, the effectiveness and accuracy of the model would depend on the quality of the data and the specific characteristics of the epidemic, including public health interventions, social behaviour, and virus traits.

3 Results and interpretation

The dataset used for this research is crucial as it serves as a basis for modeling and predicting the spread of infection. A dataset available at [20], [7] was used, which includes information collected from reliable sources such as the World Health Organization and the official websites of the respective countries. The data, accessed on 27.05.2023, are stored in CSV format, which is widely used and easy to manage due to its simplicity and readability. The dataset includes several key attributes, namely the number of people infected, those who have been cured, and those who have died from the infection. For this paper, three main parameters were identified and used to understand the evolution of the disease:

1. mortality rate (0.21)
2. infection rate 1.4-2.5 [17], in the experiments value 1.4 is used
3. incubation period (5.1 days) [14].

For this example, we will consider the case of Germany. The cellular automaton for Germany was created as detailed in Section 2.1. Germany's Covid-19 data as of 14.05.2020 was analyzed and split into training and test data. The model was constructed using multiple intervals over the entire duration of the pandemic, yielding similar outcomes throughout. This paper illustrates the model's performance using the first fifty-seven weeks of data in the training phase, while the next two weeks were reserved for testing. The fifty-eighth and fifty-ninth weeks of the pandemic period, utilized for testing, demonstrate a declining trend in the overall number of cases, see Figure 5 a). In Berlin, the number of cases exceeds 1000, represented by red, during the initial three days of this period, after which it declines below this threshold, transitioning to orange. Similarly, the Bayern region exhibits a reduction in the number of cases, indicated by the diminishing intensity

of the orange color. During this timeframe, certain regions, such as Brandenburg, maintain a consistent infection rate, of approximately 20 cases, and the color alternates between yellow and light orange. This approach ensures robust training of the prediction model and facilitates an unbiased evaluation of its performance. We can compare the prediction with the actual data by applying multiple linear regression to the cellular automaton, see Figure 5. The results of the cellular automaton can be visually represented by using colors for each cell according to the severity of Covid-19 infection in that region: yellow signifies regions with a case count below 20 per cell, orange intensity denotes a range between 20 and 100 cases and red coloration is reserved for cells with more than 1000 cases. The different shades of colors represent different levels of severity, from areas where no cases have been reported to regions experiencing high numbers of cases.

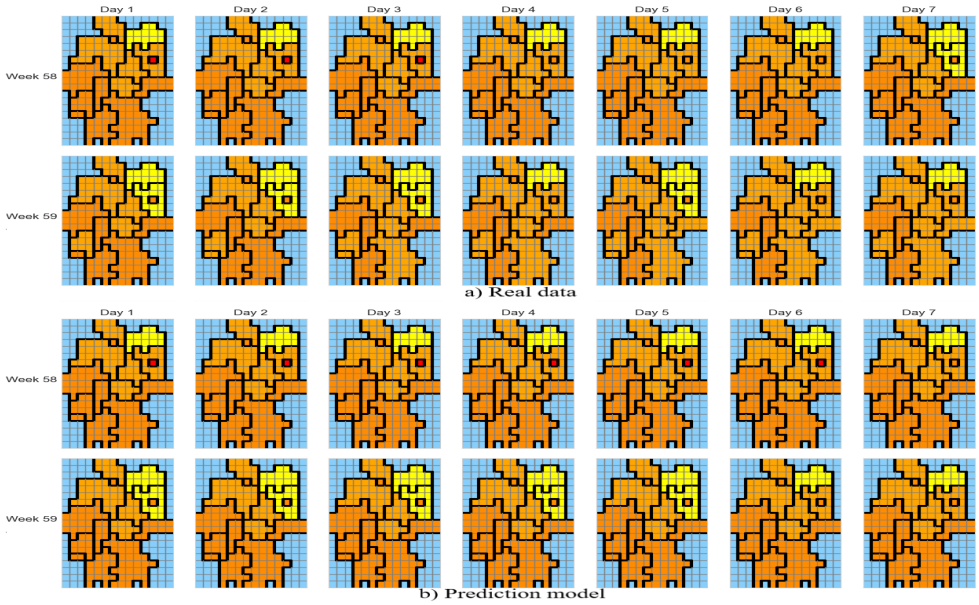

Figure 5: Actual cases (a) and prediction result for the next two weeks (b)

Comparison of predictions with real data showed the model's ability to learn and generalize models. Root mean square error (RMSE) was used to assess model performance. The RMSE for regions such as Mecklenburg-Vorpommern and Sachsen-Anhalt was small, suggesting high prediction accuracy, while for Berlin, it was significantly higher, Figure 6. The results suggest the model's effectiveness in predicting disease severity in different regions of Germany and its potential application in other regions and contexts, see Figure 7. This figure comprehensively analyzes the daily error per cell for the other regions except *Berlin* over the 14 days of testing. This analysis allows a comparative view of the model performance in different regions of Germany.

Analyzing the individual regions, it can be seen that the model maintains a relatively low error rate. This is particularly encouraging as it demonstrates the ability of the model to provide reasonable predictions across regions. Some

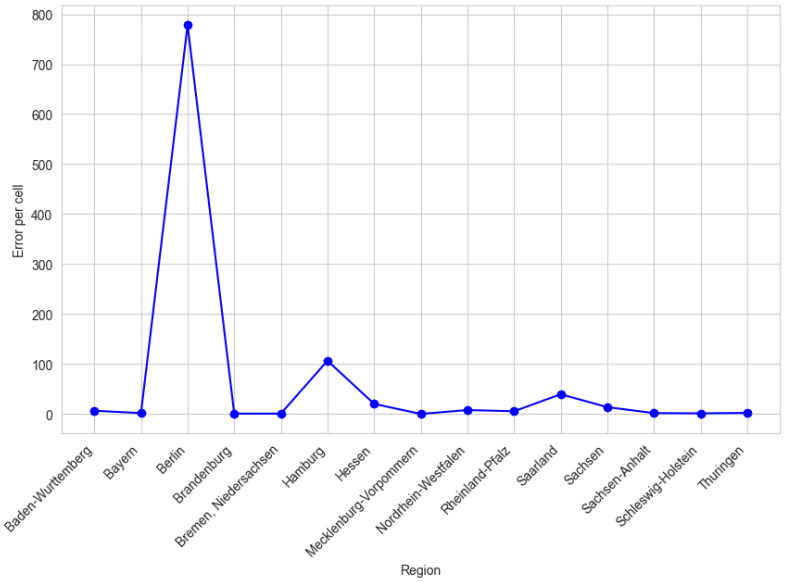

Figure 6: RMSE Results

regions show a steady increase in error per cell over the 14 days, but overall the error remains low. For example, *Brandenburg* maintains a very low error rate, increasing to only about 0.9 on day 14. In *Mecklenburg-Vorpommern*, the error remains particularly low, below 0.23, indicating excellent prediction accuracy in this region. In contrast, *Saarland* shows a more significant value, reaching an error of 39.62 on day 14. However, the errors in regions like this are still below an acceptable limit, indicating a consistent performance of the model. The regions *Berlin* and *Hamburg* have been omitted from this analysis because of their considerably high error rate (778.39 for Berlin and 106.73 for Hamburg), which differs strongly from the error rates of the other regions.

The spread of the disease among people can be unpredictable depending on several factors, being influenced by adjacency in terms of road network and capacity, the implementation of government-imposed restrictive measures and isolation protocols, and the vaccination rate, where applicable. Analyzing the obtained results and population dynamics (based on [21]) across different regions of Germany shows a direct association between the RMSE value and density. Berlin has the highest population density in Germany, with 4086 inhabitants per square kilometer, followed by Hamburg, which registers 2439 inhabitants per square kilometer.

The population size plays an important role in the prediction process. Given that Berlin is geographically represented by a single cell (see Figure 1), we expanded the map by subdividing this cell into four and sixteen smaller units, but the results were quite similar, check Table 2 for details.

The following analysis focuses on cross-border transmission of the disease from Germany, Italy, France and Austria to Switzerland. This involves assessing the spread of the disease between countries, an essential part of understanding and

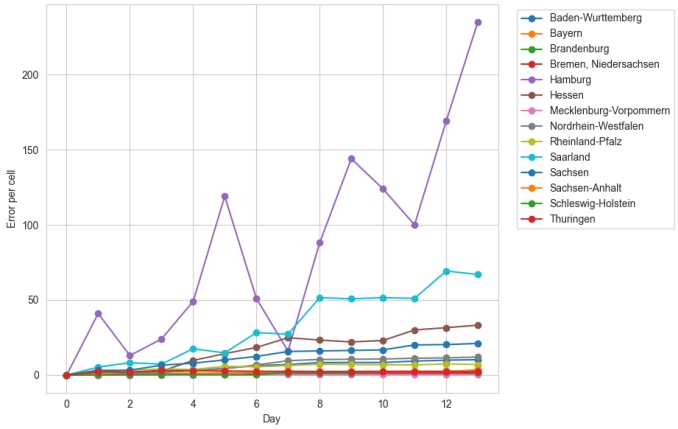

Figure 7: Daily error per cell for the other regions(without Berlin)

combating pandemics, given the interconnectedness of the contemporary world. The cellular automaton for these five countries is illustrated in Figure 2, providing a visual representation of these countries. The graphical representation provides a quantitative perspective on this issue and shows the total number of active cases in Switzerland concerning the potential cases coming from the neighbouring countries. It serves as a tool to visualize the transmission of the disease between countries, thus allowing us to observe the potential influence of neighbouring countries on the epidemic evolution in Switzerland.

Figure 8 displays the prediction results for Switzerland. For this, the initial fifty-eight weeks of data were utilized for the training phase, the next two weeks were reserved for testing. These results are considered satisfactory, as they produce an error rate of 793 cases on day 14, in relation to the demographic magnitude of Switzerland, exceeding 8 million residents.

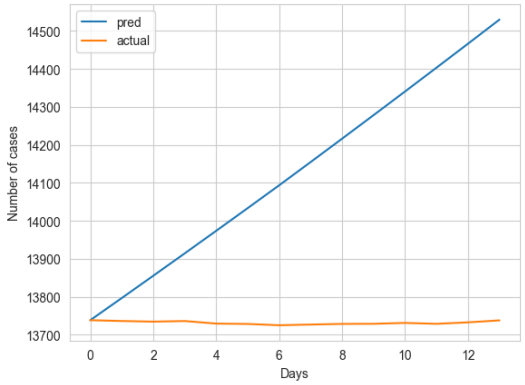

Figure 8: Predicted number of cases from neighbouring countries vs. number of cases

Figure 9 gives a clearer picture of how the spread of the disease in Germany,

Region	19x14 cells	38x28 cells	76x56 cells
Baden-Wurttemberg	6.66	6.83	7.04
Bayern	1.84	1.84	1.88
Berlin	778.39	774.7	709.62
Brandenburg	0.9	0.97	1.00
Bremen, Niedersachsen	0.75	0.88	0.9
Hamburg	106.73	144.27	189.04
Hessen	20.4	22.96	24.22
Mecklenburg-Vorpommern	0.23	0.22	0.21
Nordrhein-Westfalen	8.02	8.7	9.17
Rheinland-Pfalz	5.26	6.0	6.21
Saarland	39.62	39.47	39.68
Sachsen	13.87	13.54	13.3
Sachsen-Anhalt	1.96	2.29	2.56
Schleswig-Holstein	1.38	1.42	1.44
Thuringen	2.35	2.64	2.83

Table 2: RMSE values for multi cell representation

Italy, Austria or France could impact the epidemiological situation in Switzerland. It highlights the critical role of international cooperation and coordination in managing such infectious diseases.

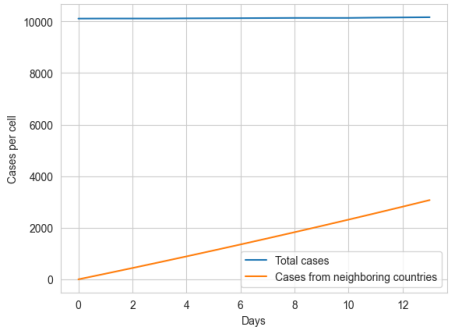

Figure 9: Cases from neighbouring countries vs. total number of cases

4 Discussions

4.1 Aim and main findings

Our approach demonstrated that implementing a cellular automaton and an epidemiological SIRD model can provide a robust tool for simulating the spread of infectious diseases. Several papers deal with the specificity of the Covid-19 pandemic in Germany [9], [12]. Nevertheless, this approach provides a more nuanced

understanding of disease dynamics at the granular level, considering regional particularities and interactions between adjacent countries.

With data from Germany serving as the central case study, the experiments have shown how the proposed model can be trained successfully utilizing real-world data. Minor differences between model predictions and real data indicate the model's validity. However, some variations were identified, particularly in regions with higher disease severity, highlighting areas where further improvements are needed.

The ability of the model to explore the disease's spread over international boundaries, a crucial component of pandemic management in our society, was demonstrated through an examination of international transmission from a neighboring country to Switzerland.

4.2 Strengths and limitations

We want to highlight some strengths and limitations of this study. To the best of our knowledge, this is the first study quantifying the model's ability to extend beyond the borders of a single country. The spread of Covid-19 from two countries into a shared neighboring country can be modeled using this approach. This unique aspect of cross-border modeling enables a more comprehensive regional view of the pandemic by reflecting the dynamics of the disease's transnational spread. In addition, the results obtained in this paper highlight the value of using visual representations, such as cellular automata, to convey complex epidemiological data in an accessible manner. By translating raw data into color-coded severity levels, this approach can make the science of epidemiology more accessible to a wider, non-specialist audience.

Other essential strengths of the current study include the fact that the model can easily be replicated on another country/epidemiological model as long as we have viable data and the fact that the model can predict the spread of disease across borders and can help identify trends in the spread of Covid-19 virus.

By all means, there are also some limitations, one of which is the need for more data - quite a lot of data is needed (total cases, cured persons, and dead persons), but only sometimes all three are reported. Another limitation may be that using the SIRD model and cellular automata may oversimplify the reality because not all people are the same, and the model does not consider the same protective measures against the spread of the virus. Experiments have shown that there can be significant errors in regions represented by a single cell. In the examples provided in section 3, Berlin and Hamburg had the most significant errors (although Hamburg was much smaller than Berlin). Furthermore, the linear regression model does not consider a large number of variables, and the general model does not consider possible mutations.

5 Conclusions

This study demonstrates the utilization of cellular automata for modeling disease spread, particularly at the intersection of computing and public health.

The research highlights the potential of integrating cellular automata with traditional epidemiological models, such as SIRD, to simulate the spread of infectious diseases such as Covid-19. This model has been effectively trained using real-world data, mainly from Germany. Despite the good fit of the model to accurate data, it was fascinating to observe variations, especially in Berlin, indicating that there is room for refinement.

The evaluation of disease transmission from Germany, Italy, Austria and France to Switzerland further strengthens the applicability of the proposed model. This additional insight demonstrates the broader potential of the model for international pandemic management, going beyond the scope of the original study. Indeed, this cross-border analysis underlines the model's adaptability for large-scale health strategies.

Also, translating raw data into color-classified severity grades using cellular automata makes complex epidemiological data more tractable. The application of this model has highlighted its practical implications and can contribute to public health decision-making. Integrating traditional epidemiological models with computational methods such as cellular automata can lead to a richer understanding of disease dynamics.

There are many opportunities for improvement, among which we mention that the proposed model can have increased complexity, incorporating various factors of disease spread; parameters can be adjusted over time; alternative predictive models with improved accuracy, such as machine learning algorithms, can be used; geographical scope can be extended by applying the model globally. Through these modifications, the model could be refined to provide more accurate and valuable information for mitigating the impact of infectious diseases.

References

[1] Alanazi, S., Kamruzzaman, M.M., Alruwaili, M., Alshammari, N., Alqahtani, S. and Karime, A., *Measuring and preventing COVID-19 using the SIR model and machine learning in smart health care*, Journal of Healthcare Engineering **2020** (2020).

[2] Anastassopoulou, C., Russo, L., Tsakris, A., and Siettos, C., *Data-based analysis, modelling and forecasting of the COVID-19 outbreak*, PLOS ONE. **15** (2020).

[3] Anderson, R.M. and May, R.M., *Infectious diseases of humans: dynamics and control*, OUP Oxford, 1991. Available at: `https://books.google.ro/books?id=HT0--xXBguQC`.

[4] Anirudh, A., *Mathematical modeling and the transmission dynamics in predicting the Covid-19 - What next in combating the pandemic*, Infectious Disease Modelling **5** (2020), 366-374.

[5] Bohner, M., Streipert, S. and Torres, D.F.M., *Exact solution to a dynamic SIR model*, Nonlinear Analysis: Hybrid Systems **32** (2019), 228-238.

[6] Brauer, F., *Compartmental models in epidemiology*, ed. by F. Brauer, P. van den Driessche, and J. Wu, Springer, New York, 2008.

[7] Dong, E., Du, H., and Gardner, L., *An interactive web-based dashboard to track COVID-19 in real time*,The Lancet Infectious Diseases **20** (2020), 533-534.

[8] Eosina, P., Arymurthy, A.M. and Krisnadhi, A. A., *A non-uniform continuous cellular automata for analyzing and predicting the spreading patterns of COVID-19*, Big Data and Cognitive Computing **6** (2022), no. 2, Article ID 46. Available at: `https://www.mdpi.com/2504-2289/6/2/46`.

[9] Hajek, A., De Bock, F., Wieler, L. H., Sprengholz, P., Kretzler, B. and König, H.-H., *Perceptions of health care use in Germany during the COVID-19 pandemic*, International Journal of Environmental Research and Public Health **17** (2020), no. 24, Article ID 9351.

[10] Hastie, T., Tibshirani, R. and Friedman, J., *The elements of statistical learning: data mining, inference and prediction*, 2009, Springer. Available at: `http://www-stat.stanford.edu/~tibs/ElemStatLearn/`.

[11] Hethcote, H.W., *The mathematics of infectious diseases*, SIAM Review **42** (2000), no. 4, 599-653.

[12] Küchenhoff, H., Günther, F., Höhle, M. and Bender, A., *Analysis of the early COVID-19 epidemic curve in Germany by regression models with change points*, Epidemiology & Infection **149** (2021), e68.

[13] Keeling, M.J., and Rohani, P., *Modeling infectious diseases in humans and animals*, Princeton University Press, 2008. Available at: `http://www.jstor.org/stable/j.ctvcm4gk0`.

[14] Li, Q., Guan, X., Wu P., et al., *Early transmission dynamics in Wuhan, China, of novel Coronavirus-infected pneumonia*, New England Journal Mededicine **382** (2020), no. 13, 1199-1207.

[15] Montgomery, D.C., Peck, E.A. and Vining, G.G., *Introduction to linear regression analysis (4th ed.)*, Wiley & Sons, 2006.

[16] Sun, K., Chen, J., and Viboud, C., *Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study*, The Lancet Digital Health. **2** (2020).

[17] Viceconte, G. and Petrosillo, N., *COVID-19 R0: Magic number or conundrum?*, Infectious Disease Reports **12** (2020), no. 1.

[18] Weisstein, E.W., *Cellular automaton (MathWorld - A Wolfram Web Resource)*, available at: `http://mathworld.wolfram.com/CellularAutomaton.html`.

[19] White, S., Rey, Á. and Sánchez, G., *Using cellular automata to simulate epidemic diseases*, Applied Mathematical Sciences **3** (2009), 959-968.

[20] *COVID-19 Data repository by the center for systems science and engineering (CSSE) at Johns Hopkins University.* (2020), available at: `https://github.com/CSSEGISandData/COVID-19`.

[21] *States of Germany*, available at: `https://en.wikipedia.org/wiki/States_of_Germany`.