The Past Helps The Future: Coupling Differential Equations with Machine Learning Methods to Model Epidemic Outbreaks

Yulia Abramova[1] and Vasiliy Leonenko[2[0000-0001-7070-6584]]

[1] ITMO University, 49 Kronverksky Pr., St. Petersburg, Russia, 197101
[2] ITMO University, 49 Kronverksky Pr., St. Petersburg, Russia, 197101
vnleonenko@itmo.com

Abstract. The aim of the research is to assess the applicability of methods of artificial intelligence to the analysis and prediction of infectious disease dynamics, with an aim to increase the speed of obtaining predictions along with enhancing quality of the results. To ensure the compliance of the forecasts with the natural laws governing the epidemic transmission, we employ Physics-Informed Neural Networks (PINN) as our main tool for the forecasting experiments. With the help of numerical experiments, we show the applicability of the approach to infectious disease modeling based on coupling classic approaches, namely, SIR models, and the cutting-edge research related to machine learning techniques. We compare the accuracy of different implementations of PINN along with the statistical models in the task of forecasting COVID incidence in Saint Petersburg, thus choosing the best modeling approach for this challenge. The results of the research could be incorporated into surveillance systems monitoring the advance of COVID and influenza incidence in Russian cities.

Keywords: Mathematical Epidemiology, SIRD Models, Physics-Informed Neural Networks, Forecasting, Python.

1 Introduction

Simple mathematical models, such as SIR models, based on differential equations, are known from the beginning of XX century and are still in wide use for epidemic outbreak prediction, having a great ability to generate easily interpretable results. In fact, a large part of the decision-making frameworks developed around the world to analyze and combat COVID-19 were based on SIR models (e.g., [1], [2]). However, they are often unable to capture peculiarities of disease dynamics related to stochastic effects, as well as to consider the uncertainty in input data. The differential model output is 100% determined by the values of the input parameters and is represented by a smooth incidence curve, although the real incidence usually has fluctuations around the trend. Also, despite the fact that SIR models belong to the simplest explanatory models available for simulating disease dynamics, they still require implementation of the calibration algorithm and might suffer from the local optima problem, giving incoherent results. In this aspect SIR models are more complicated to handle than the statistical models, - especially to the domain specialists not directly connected with modeling and differential

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_23

equations, - so they are far from being “out-of-the-box” solution for decision making in epidemiology. At the same time, the statistical approaches, although suitable for long-term forecasting of seasonal illnesses [3], are hardly applicable for predicting peaks and outbreak longevity because of the complicated physical laws governing outbreak incidence dynamics.

To provide a more effective prediction tool which is also easy to use, machine learning approaches can be employed [4]. While more sophisticated and often more accurate than simple statistical models, they also have their disadvantages. Particularly, LSTM networks which are commonly used for forecasting purposes in mathematical epidemiology [5], [6], require massive datasets for training, which are often not available, especially in cases of infections caused by novel or mutated viruses.

One of the ways to overcome the problem of the ML-based predictors, which do not “know” the laws of incidence dynamics due to scarce amount of training data, is essentially to feed them those laws directly, thus obtaining a hybrid approach incorporating ML techniques and SIR-type models. Thus, there exists an opportunity of enhancing the quality of cutting-edge methods with the help of age-old [7] differential equations. The corresponding approach is called PINN (physics-informed neural networks) [8]. Although PINNs have already received appreciation in the domains like physics, their application to epidemiological problems is still not widespread. In most of the corresponding papers known to the authors, PINNs are used to predict cumulative dynamics of epidemics in big territories, of the scale of separate countries, thus the regarded incidence time series are rather smooth and could be easily handled by neural networks. The efficiency of PINN applied to noisy city-level incidence data is yet to be assessed.

In this research, we use PINN for the aim of predicting COVID-19 incidence in Saint Petersburg and comparing its efficiency depending on modifications of PINN. The research question addressed is whether PINN can be considered an efficient “out-of-the-box” solution for the problem of disease forecasting. The answer on that question will clarify whether PINN can become a core component of decision-making frameworks used by domain specialists and policy makers to plan control measures on the city level.

2 Methods

2.1 SIRD Model

The SIRD model is one of the primary models for studying and modeling the spread of epidemics. It is based on a simpler SIR model [7] but includes an additional category, thus dividing the population into four categories: Susceptible (S), Infected (I), Recovered (R), and Dead individuals (D). In this system, each class of individuals is assigned to a specific compartment, and transitions between compartments represent the movement of individuals between different states. When “Susceptibles” (S) come into contact with infected individuals, they can also become infected and transition to the “Infected” category (I). “Infected” individuals (I) either recover and are moved to the “Recovered” category (R) or, they may die and move to the Dead category (D). All the rates, i.e. of infection transmission α, of recovery β, and of mortality γ, are considered constant. We assume that individuals who recover from the virus gain complete

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_23

immunity against future infection, so our model is limited to the simulation of a single COVID-19 wave. The total population N = S+I+R+D is considered to be constant. The corresponding differential equations have the following form:

$$\frac{dS}{dt} = -\beta SI,\ \frac{dI}{dt} = \beta SI - \gamma I - \delta I, \frac{dR}{dt} = \gamma I, \frac{dD}{dt} = \delta I.$$

2.2 Physics-Informed Neural Network

Physics-informed neural networks (PINNs) are specialized neural networks designed for addressing supervised learning tasks while incorporating the principles of physics, especially those associated with intricate nonlinear partial differential equations and ordinary-differential-equations (ODEs) [8]. Models based on PINNs adhere to physical laws by incorporating a loss function that includes the residuals from physics equations and boundary conditions. These models utilize automatic differentiation to compute the derivatives of the neural network output concerning its inputs (spatial and temporal coordinates, and model parameters) [9]. By reducing the loss function, the network can accurately approximate solutions [10], [11]. By leveraging these differential equations, PINNs enhance the learning process, enabling the algorithm to converge toward the correct solution, even when the available training data is limited.

In a specific application, PINNs were utilized for data approximation and the identification of unknown paramters of the SIRD model. Following an in-depth analysis of the system, the model can calculate the coefficients within the system of differential equations within a predetermined range, facilitating interpretation of the results [12].

Typically, a PINN architecture consists of multiple fully connected layers with numerous neurons and incorporates non-linear activation functions between them. The input for a PINN comprises batches of time steps, while the output represents tensors that convey the network's estimations of the compartments within the SIRD model at each time step [13]. These estimations are constrained by the conditions derived from the SIRD system:

$$f_1 = \frac{dS}{dt} - (-\beta SI), f_2 = \frac{dI}{dt} - (\beta SI - \gamma I - \delta I), f_3 = \frac{dR}{dt} - \gamma I, f_4 = \frac{dD}{dt} - \delta I.$$

Per training iteration, we compute the usual data error defined as

$$loss_U = mean(S - S')^2 + mean(I - I')^2 + mean(R - R')^2 + mean(D - D')^2$$

Here, X is the actual data that the model was provided, X' is the prediction the model computed. We obtain the physics-informed part of the loss function, the residual error, per training iteration as $loss_F = mean(f_1)^2 + mean(f_2)^2 + mean(f_3)^2 + mean(f_4)^2$. The neural network's parameters can be acquired through the process of minimizing the mean squared error loss. This loss function incorporates a weighting factor denoted as μ, which falls within the range of μ ϵ [0,1] and helps balance the importance of accurately reproducing the data and conforming to the differential equations: $loss = \mu loss_U + (1 - \mu)loss_F$.

2.3 Data

In our research, we work with daily incidence cases and cumulative incidence data covering a period of 359 days from July 5, 2020, to June 28, 2021 [14]. To enhance the computational speed, we reduced the dataset using every tenth record of current and

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_23

cumulative incidence. Although our experiments concentrated exclusively on one particular epidemic season, the PINN described here can be easily adapted for the diverse epidemic seasons.

3 Results

3.1 Forecasting of post-peak incidence

The prediction quality of the models was evaluated based on the root mean square error (RMSE), which measures the average squared difference between the predicted and actual values. We examined various architectures of our fully connected neural network, which forms the basis of PINNs calibrated to two types of input data: daily registered cases and cumulative incidence data (Fig 1). After the first layer, the activation function was Rectified Linear Unit (ReLU) in both cases. We established that for cumulative data, it is more advantageous to consider ReLU as the activation function, whereas for daily registered cases the hyperbolic tangent (tanh) is more suitable. We also compared our results with ARIMA predictions, with parameters (2,2,3).

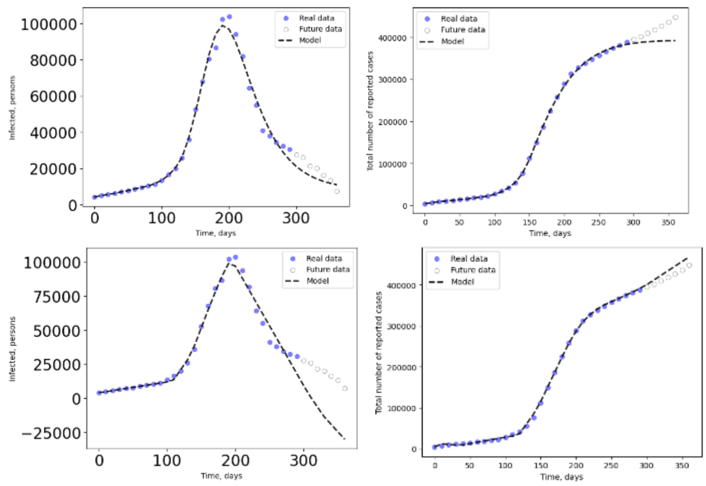

Fig. 1. Upper row: PINN's (tanh) prediction of incidence (left) and cumulative infected (right). Lower row: PINN's (ReLU) prediction of incidence (left) and cumulative infected (right).

It is worth noting that the resulting incidence forecast made by ARIMA outperforms both versions of PINN: namely, RMSE for the prediction by ARIMA is 2541.22, whereas RMSE for PINN (ReLU) is 15786.79 and RMSE for PINN (tanh) is 6516.33. Also, one can notice that the incomplete data for the experiments were taken in such a way that the peak incidence was included in the dataset, i.e., we predicted the second half-wave of declining incidence. It is obvious that the ARIMA forecasting made on incomplete data before the peak will be unable to replicate the falling of incidence and will show a growth to infinity instead. At the same time, our experiments showed that

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_23

in such conditions PINNs behave in similar fashion, despite the laws of disease incidence incorporated into the loss function. In fact, the form of the prediction curve in PINN rather replicates the activation function than the typical incidence curve. Since predicting the peak incidence and the day of maximal number of the infected is of crucial importance for healthcare specialists, the next series of experiments were dedicated to the modification of PINN algorithm to make it useful for peak forecasting.

3.2 Peak prediction

Let us consider the case when we impose an additional condition at the boundary by explicitly assuming a particular duration of the outbreak t*, i.e. the value of modeled incidence which corresponds to t* should be close to zero. Since, we don't have sufficient historical data to derive the typical epidemic duration (in case of COVID-19), or, alternatively, the variation in this value is rather big (in case of seasonal influenza), we iterated the value of the desired epidemic duration over a certain range and assessed the forecasting error in each of the iterations. The modified formula for the loss function is $loss = \mu loss_U + (1-\mu) loss_F + loss_T,$ where $loss_T$ is the additional boundary condition. Using the synthetic dataset, we compared the performance of PINNs depending on presence or absence of the boundary condition in the loss function. The inclusion of the additional boundary condition greatly enhanced the prediction quality. As a result, we use this modified error function for addressing prediction tasks on actual data, which we do below, assuming that 160 initial incidence points are used for training. We considered a range of values from 260 to 380 to search for peak predictions. We selected several durations for consideration (Fig. 2, left), for which we executed 10 iterations of the PINN with different initial seeds to account for the output variation due to the stochasticity of the algorithm. This allowed us to generate a total of 10 peak predictions for each hypothetical length of the epidemic waves. The training process was being halted once the error function fell below a predetermined threshold of 0.0001. The accuracy of fitting to the known incidence points (calibration error) and the difference between the expected and the actual incidence trajectory (prediction error) are given in Table 1. We show the quality of the peak prediction made according to these trajectories in Fig. 2, right. The maximal incidence reached during an outbreak and the day of its occurrence is a crucial epidemic indicator for the healthcare specialists, since it largely defines the maximal workload of the healthcare system in terms of the resources spent and hospital beds allocated. The points on the graph represent the prediction bias for the peak day (dt) and the ratio between the modeled and actual outbreak peak heights (dh). The ideal peak prediction scenario corresponds to the intersection of dt = 0 (vertical dashed line) and dh = 1 (horizontal dashed line).

Table 1. Median RMSE depending on the assumed outbreak duration

t*, days	260	280	310	330	360	380
Calibration	565.9	392.9	527.2	**349.2**	457.0	437.9
Prediction	26186	20073	13651	**12153**	20123	18961

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_23

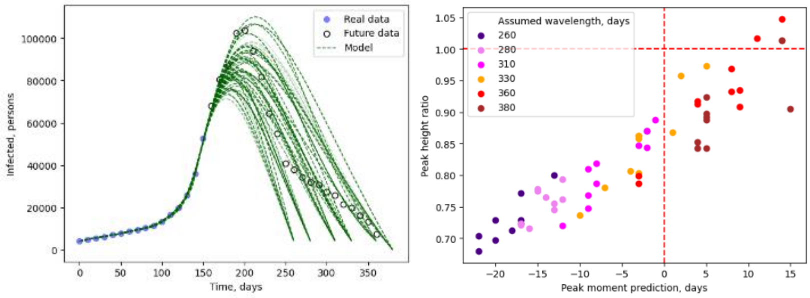

Fig. 2. Left: PINN's prediction represented by a color-coded gradient illustrating possible scenarios for each sample value across ten iterations. Right: Biases of the peak prediction for the daily incidence data depending on the assumed wavelength

In the ideal case, we would expect from calibration error (Table 1) to be convex as a function of assumed disease duration, and, moreover, to have the local minimum at the same value of outbreak duration t* as the function of predicted error, with both giving us the duration of the real outbreak (approx. 360 days). This result could be indeed reached on synthetic data generated by SIRD model, with the smooth and concave incidence curve, however, it is not the case for the real outbreak. As we can see from the tables, values of median RMSE fluctuate with changing t*. The optimal values of RMSE are reached for t*=330 and this value also delivers us the optimal forecast quality, but it is not equal to 360. As it is demonstrated in Fig. 2 (right), even for the assumed wavelength corresponding to the lowest median RMSE (t*=330, marked with orange dots) the bias both in peak height and peak day might be dramatic, depending on the simulation run. Also, it is interesting that almost all of the predictions underestimated the peak height. We consider it a peculiarity related to the regarded incidence dataset which cannot be attributed to the forecasting method itself.

4 Discussion

In this research, we described the modeling techniques for the retrospective analysis and forecasting of COVID-19 incidence in Saint Petersburg based on physics-informed neural networks. We showed that PINN has several advantages compared with SIR models and classical statistical approaches. Apart from SIR models, PINN can reproduce variation in incidence data contrary to the smooth forecasting curves provided by solving differential equations (which is however not very visible in our graphs due to data sampling and big wavelength in general). When used as a part of the forecasting framework with data assimilation, PINN calculates consecutive forecasts much faster than SIR. In general, PINN does not require complex calibration procedures, like SIR models, which is especially beneficial when the data reanalysis is performed on weekly basis.

Compared to standard ML models, PINN can produce an adequate forecast without requiring too much data for training. This is essential because the field of epidemiology generally does not provide big datasets for training.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_23

Apart from simple statistical approaches, like ARIMA, PINN can foresee and thus to reproduce a typical form of the disease incidence curve thanks to the physical laws incorporated into its loss function. This gives PINN an advantage of accurately predicting the incidence before the peak of the outbreak, which cannot be done using ARIMA and similar methods. However, for that purpose PINN requires a modification with adding a boundary condition, which was described in detail in the previous section. Also, when predicting the disease incidence after reaching the peak, PINN loses its benefits and can be easily replaced by ARIMA.

Adding assumed length of the outbreak into the loss function helped obtain more reasonable peak predictions and limit the calibration time by discarding options with unreasonable outbreak lengths. At the same time, in small training datasets different forecasts with dramatically different assumed wavelengths can demonstrate similar RMSE. However, it cannot be considered a problem of the PINN, because it is a known issue of forecasting on small incidence data when a plausible forecast could be only obtained when approaching the incidence peak [16].

Despite being based on ODEs, PINN does not guarantee that the conservation laws governing the epidemic will be complied in a model. As an example, in many cases PINN shows a negative number of infective persons at the second part of the outbreak and thus the trajectory should be artificially cut at zero. This should be taken into account when using PINN to derive disease indicators different from the predicted incidence (for instance, the number of the recovered individuals).

PINN cannot deliver plausible parameter values for the ODE on which it is based. While this could be true on smooth simulated data, in general case it does not work, as our experiments showed. Thus, apart from ODEs, PINN cannot be used to assess the disease parameter values. If there is a need to solve this task using ML techniques, there are better solutions for that, for instance, NeuroODE [17], since it does not add a stochastic component to the obtained solution.

The conclusions made during the experiments might be somewhat limited because only one wave of COVID-19 was used for testing. In future studies, we plan to generalize the results further by justifying them on additional incidence data.

To sum up, based on our research, we can conclude that the technique of PINN is indeed capable of addressing the challenges which are more often tackled by SIR models, but it also has its drawbacks, and cannot be named a universal out-of-the-box approach. The selection of the proper tool for epidemic surveillance largely depends on the modeling aim and it is up to the researcher to make an ultimate choice of the most suitable technique.

Acknowledgments. This research was supported by The Russian Science Foundation, Agreement #22-71-10067.

Disclosure of Interests. The authors declare no competing interests.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_23

References

1. Read, J., et al.: Novel coronavirus 2019-nCoV (COVID-19): early estimation of epidemiological parameters and epidemic size estimates. Philos. Trans. R. Soc. B. 376, 20200265 (2021)
2. Maier, B., Brockmann, D.: Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science 368, 742–746 (2020)
3. Kondratyev, M. and Tsybalova, L., 2021. Long-term forecasting of influenza-like illnesses in Russia. International Journal of Pure and Applied Mathematics, 2013, 89(4), 619–641.
4. Shahid, F., Zameer, A. and Muneeb, M., 2020. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos, Solitons & Fractals, 140, p.110212.
5. Z. Yang et al., "Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions," J.Thoracic Dis., vol. 12, no. 3, pp. 165–174, Mar. 2020
6. Leonenko, V. N., Bochenina, K. O., & Kesarev, S. A. (2017). Influenza peaks forecasting in Russia: Assessing the applicability of statistical methods. Procedia Computer Science, 108, 2363-2367.
7. William Ogilvy Kermack and Anderson G McKendrick. "A contribution to the mathematical theory of epidemics". In: Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character 115.772 (1927), pp. 700–721.
8. M. Raissi, P. Perdikaris, and G.E. Karniadakis. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". In: Journal of Computational Physics 378 (2019), pp. 686–707. issn: 0021-9991.
9. A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind. Automatic differentiation in machine learning: a survey. arXiv:1502.05767, 2015.
10. C. Rao, S. Hao, and L. Yang. Physics-informed deep learning for incompressible laminar flows. Theoretical and Applied Mechanics Letters, 10:207–212, 2020.
11. Salah A Faroughi, Pingki Datta, Seyed Kourosh Mahjour, and Shirko Faroughi. Physics-informed neural networks with periodic activation functions for solute transport in heterogeneous porous media. arXiv preprint arXiv:2212.08965, 2022.
12. Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next. Journal of Scientific Computing, 92(88).
13. S. Shaier, M. Raissi, and P. Seshaiyer. Data-driven approaches for predicting spread of infectious diseases through dinns: Disease informed neural networks. Letters in Biomathematics, 9(1):71–105, 2022
14. Kouprianov, A.: Monitoring COVID-19 epidemic in St. Petersburg, Russia: Data and scripts (2021). https://github.com/alexei-kouprianov/COVID-19.SPb.monitoring
15. Huaman, I., & Leonenko, V. Does Complex Mean Accurate: Comparing COVID-19 Propagation Models with Different Structural Complexity. Computational Science. In ICCS 2023: Computational Science – ICCS 2023 (pp. 270–277)
16. Leonenko, V. N., & Ivanov, S. V. (2018). Prediction of influenza peaks in Russian cities: Comparing the accuracy of two SEIR models. Math. Biosci. Eng, 15(1), 209-232.
17. Chen RT, Rubanova Y, Bettencourt J, Duvenaud DK. Neural ordinary differential equations. Advances in neural information processing systems. 2018;31.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63772-8_23