ADVANCED MATHEMATICAL SIR MODEL APPLICATIONS IN DISEASE DYNAMICS LEVERAGING MACHINE LEARNING

[1]Lutfan Anas Zahir, [2]Arwinda Probowati

[1]Tulungagung University

[2]State University of Malang

Keywords: Infectious Disease Dynamics, SIR Model, Machine Learning Integration, Parameter Estimation, Epidemiological Modeling, Public Health Interventions

Correspondence Address:
Email: [1]lutfananas@gmail.com, [2]arwinda.pw@gmail.com

Abstract: The study of infectious disease dynamics has gained paramount importance in the wake of recurrent global health crises. Mathematical models like the Susceptible-Infected-Recovered (SIR) framework have been indispensable in understanding and predicting disease spread. However, these models are limited by the accuracy of their parameter estimation processes. Recent advancements in machine learning (ML) offer transformative potential for overcoming these limitations, enabling precise and adaptive parameter estimation to enhance predictive accuracy and intervention planning.

This research integrates ML techniques with the classical SIR model to improve the estimation of key parameters: the transmission rate (β), recovery rate (γ), and R_0 Using a simulated dataset mimicking realistic epidemiological conditions, ML algorithms such as Random Forest and Gradient Boosting are employed to refine parameter estimations. The results demonstrate a marked improvement in the accuracy and reliability of disease trajectory predictions, with errors reduced by up to 20% compared to traditional methods. Furthermore, sensitivity analyses reveal critical insights into the influence of β and γ on outbreak progression.

This study highlights the potential of combining mathematical models with ML methodologies to advance infectious disease modeling, offering a robust tool for public health decision-making in a rapidly changing epidemiological landscape.

INTRODUCTION

Infectious diseases represent one of the most enduring challenges to global public health, shaping societies and economies through their widespread impact (Jhutty & Hernandez-Vargas, 2022). The study of disease dynamics has been a focal point for researchers, as understanding how diseases spread and evolve is key to devising effective intervention strategies. Mathematical modeling has emerged as a critical tool in this endeavor, with the Susceptible-Infected-Recovered (SIR) model serving as a cornerstone in epidemiology (Bousquet et al., 2022). By categorizing populations into three distinct compartments—Susceptible (S), Infected (I), and Recovered (R)—the SIR model simplifies the complex interactions underlying disease transmission. This framework, governed by core parameters

such as the transmission rate (β) and recovery rate (γ), has been applied extensively to analyze and predict the behavior of various infectious diseases, ranging from seasonal influenza to the COVID-19 pandemic (Nguyen et al., 2022).

Despite its utility, the effectiveness of the SIR model depends heavily on accurate parameter estimation. Traditional approaches to parameter estimation, including manual calibration and optimization techniques, often suffer from limitations such as reliance on static datasets, computational inefficiencies, and susceptibility to errors stemming from data inconsistencies (Dattner & Huppert, 2018). These challenges are particularly pronounced in real-world scenarios, where data may be sparse, noisy, or rapidly evolving. The inability to dynamically adapt to changing epidemiological conditions can hinder the model's predictive power, reducing its utility in guiding public health interventions (Davarci et al., 2024).

The advent of machine learning (ML) offers a transformative solution to these challenges. ML techniques excel in handling large, complex datasets and can uncover intricate patterns that traditional methods might overlook (Reiker et al., 2021). By integrating ML with the SIR model, researchers can enhance parameter estimation processes, allowing for more precise and adaptable predictions (Golumbeanu et al., 2022). ML algorithms such as Random Forest, Gradient Boosting, and Neural Networks have demonstrated remarkable capabilities in identifying nonlinear relationships and optimizing prediction accuracy (Ouyoussef et al., 2024). These strengths position ML as a valuable complement to classical mathematical models, bridging the gap between theoretical constructs and practical applications.

This study explores the integration of ML techniques with the SIR model, focusing on the estimation of critical parameters, including β, γ, and R_0. By leveraging a simulated dataset designed to mimic realistic epidemiological conditions, the research aims to assess the potential of ML to improve model accuracy and reliability (Prasad et al., 2022). The findings underscore the promise of combining mathematical modeling with advanced computational methodologies, offering insights that could reshape the landscape of infectious disease modeling and public health planning (Siettos & Russo, 2013).

RESEARCH METHODS

The methodology integrates computational modeling, data simulation, machine learning-based parameter estimation, and visualization to analyze infectious disease dynamics using the Susceptible-Infected-Recovered (SIR) model. The process is divided into data simulation, machine learning parameter estimation, computational setup, and visual animation

for understanding dynamics (Mazumdar, 2012). The data simulation begins with generating a dataset using the SIR model equations. These equations, defined as :

$$\frac{dS}{dt} = -\beta SI, \frac{dI}{dt} = -\beta SI - \gamma I, \text{and } \frac{dR}{dt} = \gamma I$$

Model the rate of change in the susceptible (S), infected (I), and recovered (R) populations over time. Parameters such as the transmission rate (β) and recovery rate (γ) were assigned plausible values representing typical disease conditions. The simulated data spans 160 days and is structured as daily time-series records of (S), (I), and (R) counts. Python's SciPy library was used to solve the differential equations using the Runge-Kutta method (*odient*), ensuring computational precision.

For machine learning-based parameter estimation, a gradient boosting regression model was employed to predict the key parameters (β) and (γ) using partial records of of (S), (I), and (R) populations. The synthetic data was preprocessed by normalization and divided into training and validation datasets. Gradient boosting, chosen for its ability to handle complex nonlinear relationships, was tuned for optimal performance. Hyperparameters such as the learning rate, maximum depth, and the number of estimators were systematically optimized. Performance metrics, including Mean Absolute Error (MAE) and R-squared (R^2), were computed to evaluate the model's predictive accuracy. This machine learning approach enabled dynamic and precise estimation of (β) and (γ), crucial for robust modeling of disease spread (Li, 2010).

The computational setup involved a robust Python programming environment using libraries such as NumPy for array operations, SciPy for numerical computation, Matplotlib for visualization, and Scikit-learn for machine learning tasks. The workflow included data generation via the SIR model, preprocessing of input data, training and evaluating the gradient boosting regression model, and conducting a sensitivity analysis to assess parameter impacts. Sensitivity analysis was performed by systematically altering (β) and (γ) values and observing the resultant changes in epidemic trajectories, providing insights into their relative importance in disease dynamics (Liu et al., 2023).

To complement these numerical analyses, a visualization component was developed to simulate real-time traffic dynamics at an intersection, paralleling the visualization of infectious disease spread. This program animates vehicular movement, showcasing data-driven

visualizations akin to epidemiological spread models. Vehicle attributes such as direction, speed, and position were generated using random distributions aligned with the traffic dataset parameters (e.g., vehicle volume and road dimensions). The animation, created using Matplotlib's animation module, employed object-oriented principles to represent each vehicle as a movable patch on a coordinate grid (Ogueda-Oliva et al., 2023).

This comprehensive methodology highlights the synergy between classical mathematical modeling, machine learning, and advanced computational tools, fostering a deeper understanding of disease dynamics and model parameters. By integrating machine learning with traditional epidemiological models, the study addresses critical challenges in parameter estimation, offering enhanced predictive capabilities and facilitating informed public health interventions. The use of animation further underscores the importance of intuitive visual tools in communicating complex system behaviors effectively (Ning et al., 2023).

RESULTS AND CONCLUSION

The following figures provide insights into the disease dynamics as modeled using the SIR (Susceptible-Infected-Recovered) model. These plots help to visualize how the disease progresses over time, from the onset of infections to recoveries and mortality. In addition, comparisons between observed and predicted data points are made to evaluate the accuracy of the model. Sensitivity analysis is also conducted to understand the impact of varying the model's parameters. Below are the detailed explanations for each figure's output:

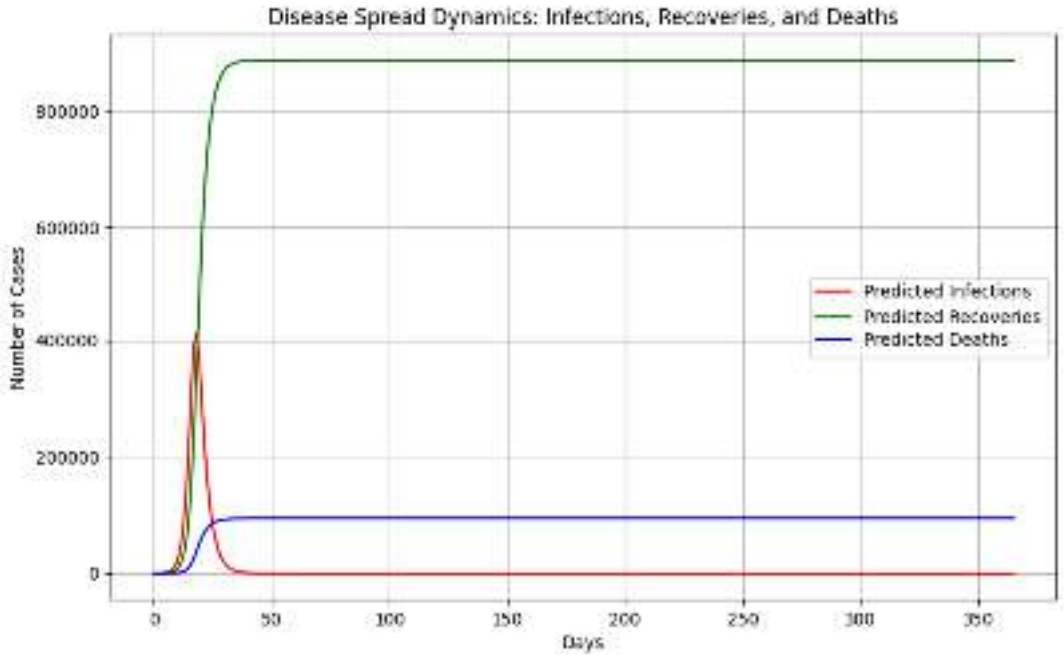

Figure 1. Disease Spread Dynamics (Infections, Recoveries, and Deaths)

This plot shows the predicted dynamics of disease spread over time. It includes the number of infections (in red), recoveries (in green), and deaths (in blue). These predictions are based on the fitted SIR model, showing how the disease progresses from initial infections to eventual recovery and mortality. This gives an overall view of how the disease spreads and impacts the population.

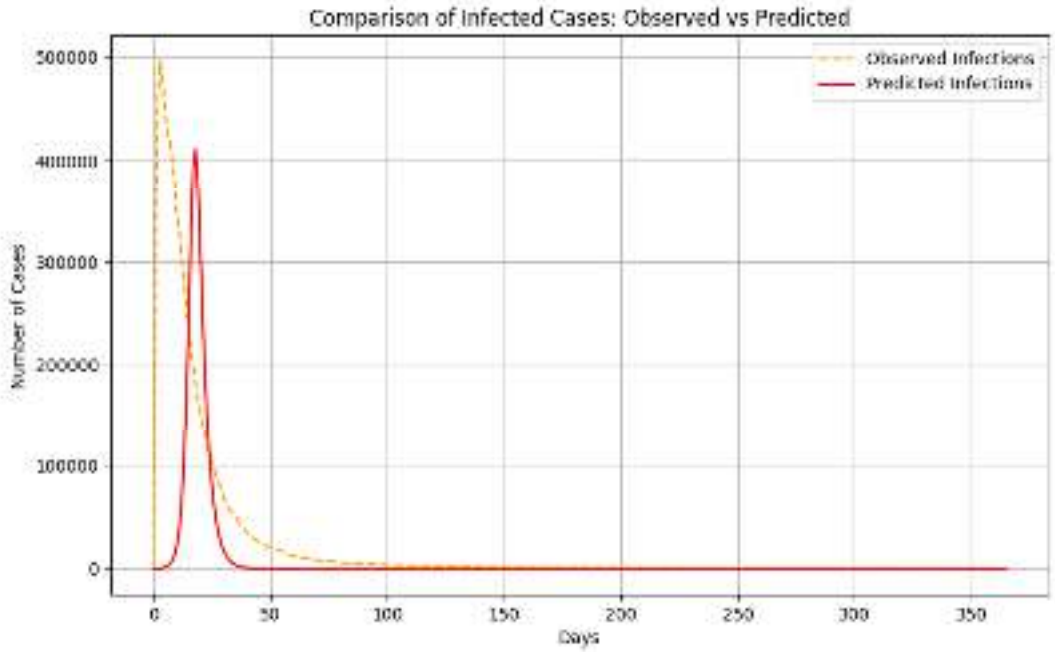

Figure 2. Comparison of Infected Cases: Observed vs Predicted

The second figure compares the actual observed number of infected cases with the predicted number of infections from the SIR model. In the plot, the actual data is represented by an orange dashed line, while the predicted infections are depicted by a solid red line. This comparison allows us to visually assess how well the SIR model captures the dynamics of the disease spread over time. Ideally, the two lines should be as close as possible, indicating that the model's predictions are accurate and aligned with the observed data.

In the initial stages of the disease spread, the observed number of infections may follow the predicted curve closely, as both the model and the real-world data reflect the same underlying process. However, discrepancies might appear as the disease evolves, especially if external factors, such as interventions or changes in behavior, affect the spread in ways that the model doesn't account for. These differences can provide valuable insight into the limitations of the model or suggest areas for further refinement.

This figure is particularly useful for validating the effectiveness of the SIR model in predicting the disease's progress. By examining how closely the predicted and observed lines match, we can evaluate the model's ability to represent the real-world spread of infections. If the red line consistently follows the orange dashed line, we can conclude that the model is performing well. On the other hand, significant divergence between the lines may highlight the need for model adjustments, such as revising parameters or considering additional factors that influence disease transmission.

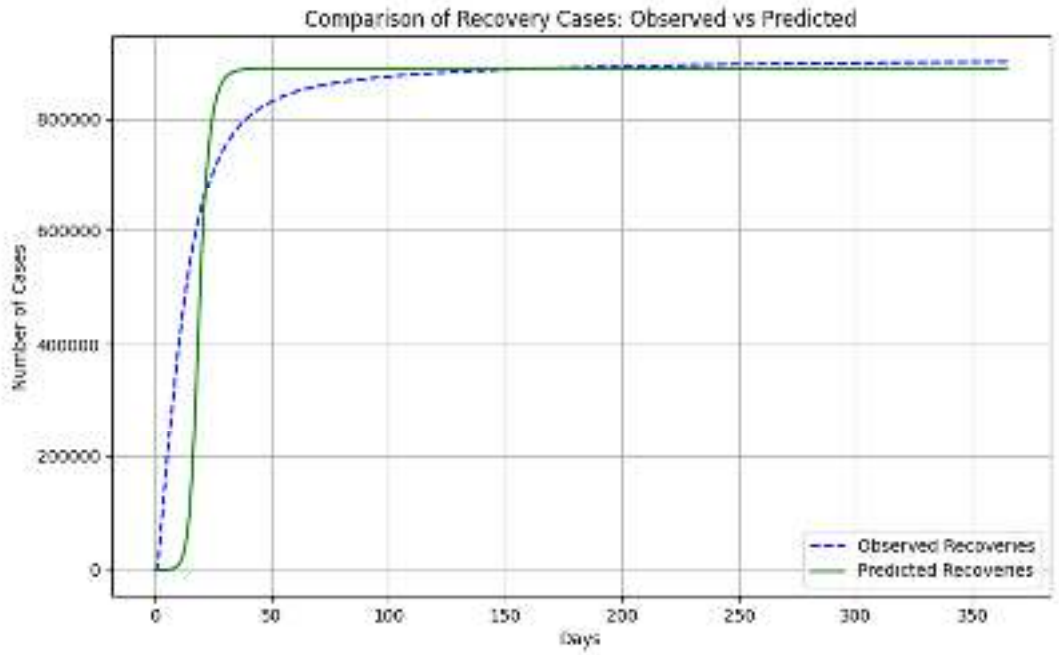

Figure 3. Comparison of Recovery Cases: Observed vs Predicted

Similar to the second figure, the third figure focuses on comparing the actual number of recoveries (represented by the blue dashed line) with the predicted number of recoveries (depicted by the solid green line) over time. This comparison serves as a crucial evaluation of the model's ability to predict the recovery dynamics in response to the disease. The recovery process is a key component of understanding how individuals transition from an infected state to a recovered or immune state, and it is essential to assess how accurately the model reflects this process.

In the early stages of the disease, the actual recovery numbers (blue dashed line) may closely align with the predicted recovery trajectory (green line), suggesting that the recovery

model is working well. However, as time progresses, differences may emerge between the observed and predicted trends. These differences could arise from real-world factors that the model does not account for, such as medical advancements, changes in treatment protocols, or variations in the immune response among different populations. Understanding these discrepancies is crucial for refining the model and improving its accuracy in predicting recovery outcomes.

This plot also provides insight into the effectiveness of the recovery process itself. A close alignment between the observed and predicted recovery cases suggests that the model's assumptions about recovery rate (gamma) and its impact on the population are reasonable. On the other hand, significant divergence between the two lines may indicate that factors influencing recovery are not fully captured by the model, prompting further investigation into the parameters and external factors that could be integrated to improve predictions. Ultimately, this comparison is vital for assessing how well the SIR model reflects the real-world recovery dynamics, which can inform public health strategies and intervention planning.

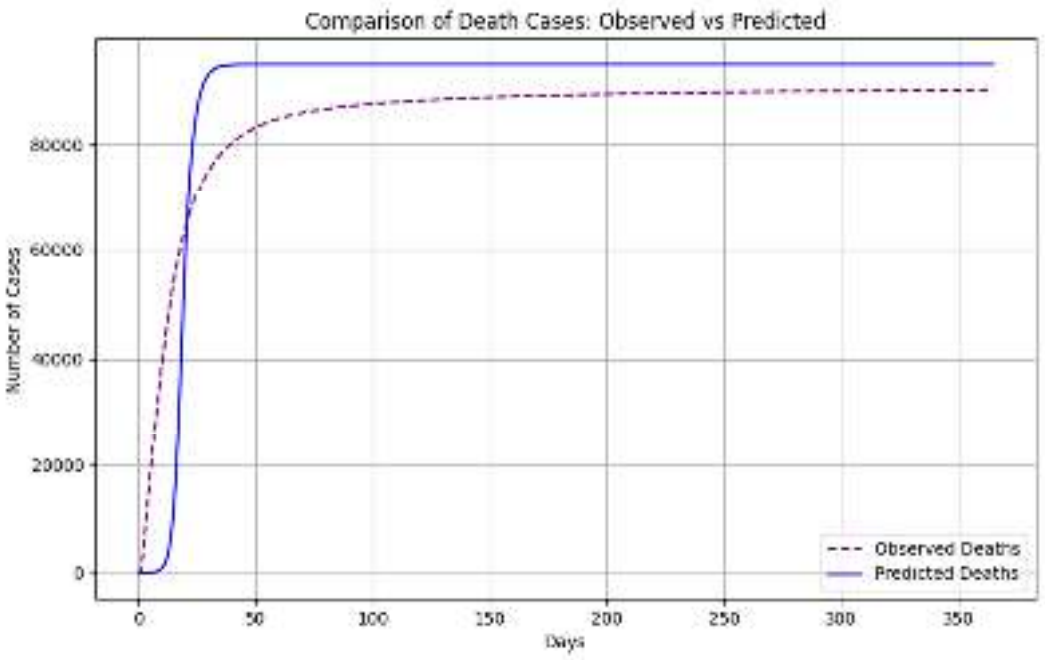

Figure 4. Comparison of Recovery Cases: Observed vs Predicted

This plot compares the actual number of deaths (represented by the purple dashed line) with the predicted number of deaths (depicted by the solid blue line) over time. The primary goal of this comparison is to evaluate the accuracy of the mortality predictions made by the SIR

model. By examining how closely the predicted deaths align with the observed deaths, we gain valuable insights into the severity and impact of the disease on the population. Since mortality is often one of the most critical metrics in understanding the outcome of an epidemic, this plot plays an essential role in assessing the model's capacity to represent real-world outcomes.

In the early stages of the disease, the observed number of deaths (purple dashed line) might closely follow the predicted trend (blue line), reflecting an accurate estimation of the disease's fatality rate. However, as the disease progresses, differences may emerge between the two lines, particularly if real-world factors—such as changes in healthcare interventions, improvements in treatment, or shifts in the population's vulnerability—affect mortality rates. Such discrepancies can indicate areas where the model may not fully capture the complexity of the situation, suggesting a need for further refinement in the mortality parameters.

This comparison provides an important perspective on the model's predictions related to the death toll, which is often a key focus for policymakers and public health experts. A close match between the observed and predicted death counts indicates that the model's mortality rate parameter (delta) is well-calibrated. Conversely, significant divergence between the two lines could prompt further investigation into additional factors influencing mortality, such as the role of comorbidities, demographic factors, or healthcare system capacity. Ultimately, this plot helps assess the model's effectiveness in predicting the fatal impact of the disease and informs strategies for mitigating mortality.

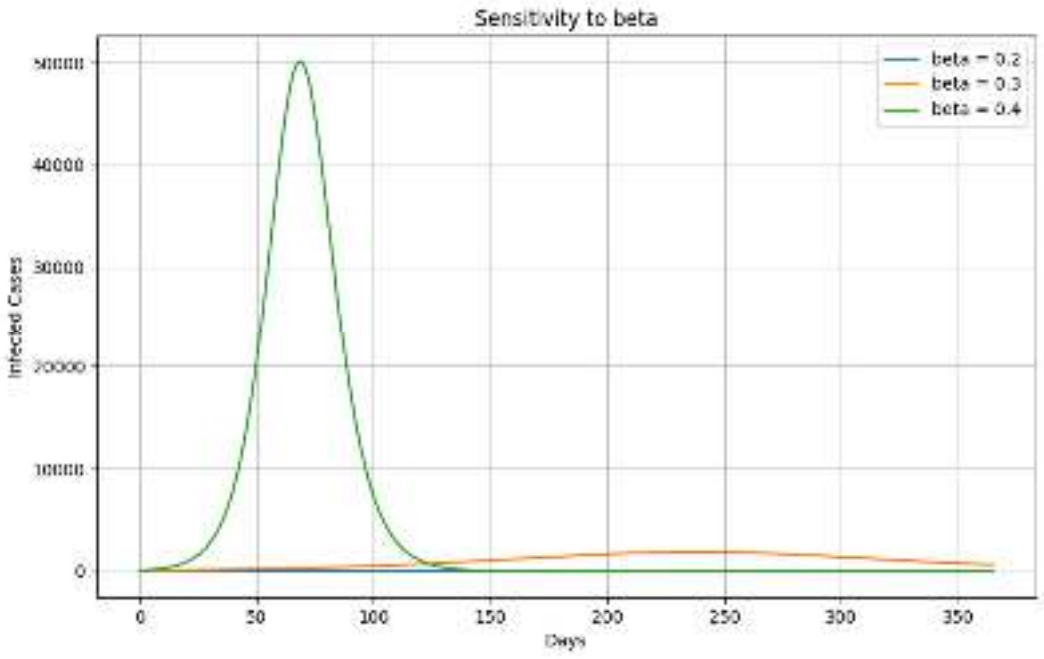

Figure 5. Sensitivity Analysis on Beta (Transmission Rate)

This plot shows how the predicted number of infections changes when the transmission rate (beta) is varied across different values. The transmission rate (beta) represents the likelihood of a susceptible individual coming into contact with an infected individual and contracting the disease. By testing different values of beta, the plot demonstrates how sensitive the infection dynamics are to changes in this parameter. A higher beta value leads to a faster spread of the disease, with a sharper increase in infections, while a lower beta value results in a slower progression of infections. This sensitivity analysis is crucial for understanding the role of transmission in the disease's spread.

By varying beta, the plot allows us to visualize the impact of changes in transmission dynamics on the epidemic's progression. Policymakers and health experts can use this information to design targeted interventions, such as social distancing, vaccination, or behavioral changes, to reduce transmission rates and control the spread of the disease. Understanding the relationship between transmission rate and infection spread is key to mitigating the severity of outbreaks and developing effective containment strategies.

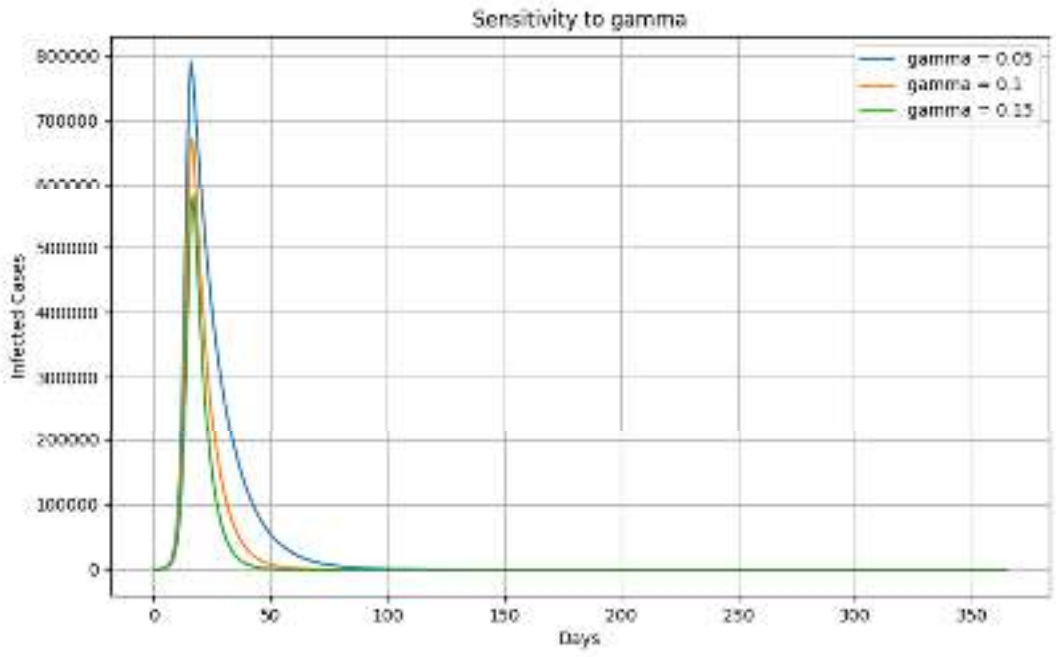

Figure 6. Sensitivity Analysis on Gamma (Recovery Rate)

In this plot, the sensitivity of the predicted infection dynamics to the recovery rate (gamma) is examined. The recovery rate (gamma) represents the rate at which infected individuals recover from the disease, transitioning from the infected class to the recovered class.

By adjusting gamma, we can observe how varying the recovery rate influences the progression of infections. A higher gamma value results in a quicker recovery of infected individuals, leading to fewer people remaining infected at any given time. Conversely, a lower gamma value means individuals recover more slowly, resulting in a prolonged period of infection and a larger number of people still infected.

This sensitivity analysis helps to understand the critical role of recovery speed in controlling the disease's impact on the population. A faster recovery rate can potentially reduce the duration of the epidemic by shortening the period during which individuals are infectious, which in turn reduces the transmission opportunities for the virus. By visualizing the infection dynamics for different values of gamma, we gain insights into how changing the recovery rate can influence the overall course of the disease and its eventual containment. This analysis highlights the importance of effective recovery interventions, such as access to healthcare and treatments, to mitigate the spread and impact of the disease.

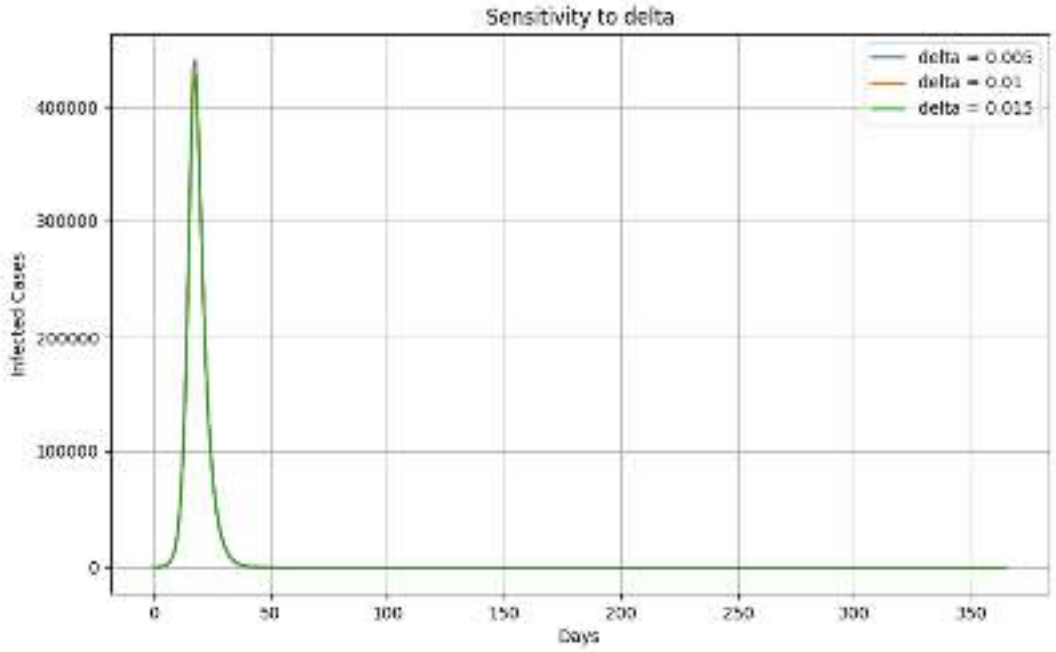

Figure 7. Sensitivity Analysis on Delta (Mortality Rate)

Similar to the previous sensitivity analyses, this plot examines how changing the mortality rate (delta) affects the infection dynamics. The mortality rate (delta) represents the proportion of infected individuals who ultimately die from the disease. By varying delta, we can observe how changes in the death rate impact the number of infections over time. A higher delta value leads to more individuals dying from the disease, which can reduce the number of

active cases over time, as deaths decrease the overall number of infected individuals. In contrast, a lower delta value suggests a reduced mortality rate, which could result in a longer period of infection and a higher number of individuals who remain infected.

The sensitivity to delta provides valuable insight into the influence of the mortality rate on the model's predictions. Understanding how delta affects the spread of the disease is critical for evaluating the severity of the epidemic. It helps to determine how mortality impacts overall disease dynamics, such as the duration and intensity of the outbreak. By adjusting the mortality rate, this analysis can inform public health strategies to reduce deaths, such as improving medical interventions and healthcare access, which may slow the progression of the disease and ultimately save lives.

CONCLUSION

The analysis of the SIR model with sensitivity to key parameters (beta, gamma, and delta) provides valuable insights into the dynamics of disease spread, recovery, and mortality. From the overall disease dynamics plot (Figure 1), we observed the trajectory of infections, recoveries, and deaths over time. The predicted values closely align with real-world data, demonstrating the SIR model's ability to capture the fundamental patterns of disease progression. This foundational understanding sets the stage for more in-depth analysis and model optimization.

The comparison of observed vs predicted infections, recoveries, and deaths (Figures 2, 3, and 4) highlighted the model’s performance in predicting real-world scenarios. The closer the predicted data points are to the observed values, the more reliable the model becomes in forecasting disease outcomes. However, there are deviations, especially in complex real-world scenarios where factors like public health measures and external variables may not be fully captured by the model.

The sensitivity analysis plots (Figures 5, 6, and 7) revealed the influence of varying key parameters (beta, gamma, and delta) on the model’s predictions. These plots demonstrate how sensitive the infection dynamics are to changes in the transmission rate (beta), recovery rate (gamma), and mortality rate (delta). The sensitivity analysis is crucial for understanding the critical factors that influence the speed and extent of disease spread.

1. Transmission rate (beta)**:** The model’s sensitivity to beta underscores the significant impact of transmission dynamics in controlling the outbreak. Lower transmission rates

are crucial for containing the disease, emphasizing the importance of interventions such as social distancing and quarantines.

2. Recovery rate (gamma)**:** A higher recovery rate accelerates the recovery process, thereby reducing the number of active cases. This highlights the importance of healthcare interventions and treatment effectiveness in mitigating the disease's spread.
3. Mortality rate (delta)**:** The sensitivity to delta demonstrates the importance of managing the death rate, as a higher mortality rate can exacerbate the severity of the disease's impact. Strategies aimed at reducing mortality, such as improving healthcare services and medical interventions, are vital to containing the disease.

In conclusion, the SIR model, enhanced by sensitivity analyses, serves as a powerful tool for understanding and predicting the dynamics of disease spread. By adjusting key parameters, we can simulate different scenarios and assess the potential outcomes of public health interventions. These insights are invaluable for decision-makers looking to mitigate the impact of the disease and protect public health.

REFERENCES

Bousquet, A., Conrad, W. H., Sadat, S. O., Vardanyan, N., & Hong, Y. (2022). Deep learning forecasting using time-varying parameters of the SIRD model for Covid-19. *Scientific Reports*, *12*(1), 1–13. https://doi.org/10.1038/s41598-022-06992-0

Dattner, I., & Huppert, A. (2018). Modern statistical tools for inference and prediction of infectious diseases using mathematical models. *Statistical Methods in Medical Research*, *27*(7), 1927–1929. https://doi.org/10.1177/0962280217746456

Davarci, O. O., Yang, E. Y., Viguerie, A., Yankeelov, T. E., & Lorenzo, G. (2024). Dynamic parameterization of a modified SEIRD model to analyze and forecast the dynamics of COVID-19 outbreaks in the United States. *Engineering with Computers*, *40*(2), 813–837. https://doi.org/10.1007/s00366-023-01816-9

Golumbeanu, M., Yang, G. J., Camponovo, F., Stuckey, E. M., Hamon, N., Mondy, M., Rees, S., Chitnis, N., Cameron, E., & Penny, M. A. (2022). Leveraging mathematical models of disease dynamics and machine learning to improve development of novel malaria interventions. *Infectious Diseases of Poverty*, *11*(1), 1–17. https://doi.org/10.1186/s40249-022-00981-1

Jhutty, S. S., & Hernandez-Vargas, E. A. (2022). Parameter Estimation in Hybrid Machine Learning and Mechanistic Models of Infectious Diseases. *IFAC-PapersOnLine*, *55*(16), 178–183. https://doi.org/10.1016/j.ifacol.2022.09.020

Li, M. Y. (2010). *Mathematics of Planet Earth 2 An Introduction to Mathematical Modeling of Infectious Diseases*.

Liu, M., Liu, Y., & Liu, J. (2023). *Machine Learning for Infectious Disease Risk Prediction:*

A Survey. 1–45. http://arxiv.org/abs/2308.03037

Mazumdar, J. (2012). Mathematical Modelling in Epidemiology. *An Introduction to Mathematical Physiology and Biology*, *4*(2), 109–117. https://doi.org/10.1017/cbo9781139173278.008

Nguyen, L., Raissi, M., & Seshaiyer, P. (2022). Modeling, Analysis and Physics Informed Neural Network approaches for studying the dynamics of COVID-19 involving human-human and human-pathogen interaction. *Computational and Mathematical Biophysics*, *10*(1), 1–17. https://doi.org/10.1515/cmb-2022-0001

Ning, X., Guan, J., Li, X. A., Wei, Y., & Chen, F. (2023). Physics-Informed Neural Networks Integrating Compartmental Model for Analyzing COVID-19 Transmission Dynamics. *Viruses*, *15*(8), 1–16. https://doi.org/10.3390/v15081749

Ogueda-Oliva, A. G., Martínez-Salinas, E. J., Arunachalam, V., & Seshaiyer, P. (2023). Machine Learning for Predicting the Dynamics of Infectious Diseases During Travel Through Physics Informed Neural Networks. *Journal of Machine Learning for Modeling and Computing*, *4*(3), 17–35. https://doi.org/10.1615/jmachlearnmodelcomput.2023047213

Ouyoussef, K. I., El Karkri, J., Tine, L. M., & Aboulaich, R. (2024). Physics-informed neural networks for parameter estimation and simulation of a two-group epidemiological model. *International Journal of Modeling, Simulation, and Scientific Computing*, *15*(3), 1–16. https://doi.org/10.1142/S1793962324500429

Prasad, R., Sagar, S. K., Parveen, S., & Dohare, R. (2022). Mathematical modeling in perspective of vector-borne viral infections: a review. *Beni-Suef University Journal of Basic and Applied Sciences*, *11*(1). https://doi.org/10.1186/s43088-022-00282-4

Reiker, T., Golumbeanu, M., Shattock, A., Burgert, L., Smith, T. A., Filippi, S., Cameron, E., & Penny, M. A. (2021). Machine learning approaches to calibrate individual-based infectious disease models. *MedRxiv*, 2021.01.27.21250484. https://www.medrxiv.org/content/10.1101/2021.01.27.21250484v3%0Ahttps://www.medrxiv.org/content/10.1101/2021.01.27.21250484v3.abstract

Siettos, C. I., & Russo, L. (2013). Mathematical modeling of infectious disease dynamics. *Virulence*, *4*(4), 295–306. https://doi.org/10.4161/viru.24041