JEPIBIO
Journal of Epidemiology and Infection Biology

Review Article

Computation and Comparative Modelling Approaches in Infectious Disease Prediction: A Review

Muhammad Shahzaib Hasnain*

School of Allied Health and Social Care, Faculty of Health, Medicine and Social Care, Anglia Ruskin University, East Rd, Cambridge Campus, United Kingdom

*Corresponding author:

Mr. M. Shahzaib Hasnain
(e-mail: shahzaibhasnain909@gmail.com)

Date of Receiving: 03/01/2025
Date of Acceptance 18/01/2025
Date of Publishing 10/06/2025

ABSTRACT Infectious diseases are one of the significant health concerns worldwide because of the high rate of disease transmission and multifactorial severity. The novel emerging and re-emerging pathogens necessitate the prediction tools that are precise and timely. Big data, algorithms, and real-time analytics have helped computational models to promote disease forecasting. Such models include mathematical models, machine learning methods and bioinformatics models. This review classifies the models of predicting infectious diseases into four groups, namely epidemiological, simulation, mathematical and bioinformatics-based. The assessment of each category is made on basis of structure, application, computation requirements as well as data dependency. Comparative analysis revealed that each model approach has its strengths and weaknesses as preferred. High scalability, low power consumption, and flexibility are the factors which are important to consider while choosing modelling approach. Correspondence analysis was also used in the review in order to put the connection between the types of pathogens and the organ systems. Respiratory and immune systems are the focus of the viral diseases. Diseases that are related to bacteria are linked with the respiratory and the nervous system. Parasites diseases exhibit close relation to blood and hepatic systems. The links of fungal infections are associated with the nervous system. This review will aid researchers and policymakers in selecting appropriate tools for infectious disease prediction and intervention planning.

KEYWORDS Algorithms, Computation, Infections, Prediction, Modelling

Introduction

The higher incidence of infectious diseases worldwide has remained a challenge to the health systems because of the speed at which they spread (Lloyd and May, 2001). It is contributed by its multifactorial aetiology and the fact that uncontrolled infectious disease spread can lead to massive epidemics such as COVID (Keshavamurthy *et al*, 2022). These diseases burden healthcare systems, disrupt economies, and challenge public health infrastructure, especially in resource-limited settings. The unpredictability and speed of transmission make timely detection and intervention essential (Wilke and Adami, 2002). The increased occurrence of novel pathogens and the recurrence of previously reported pathogens justifies the necessity of timely and optimal disease prediction to take course in intervention measures (Yang *et al*, 2020a). Addressing infectious diseases remains a top priority for ensuring population health and global stability.

Computational methods have transformed disease prediction methods and scalability. These methods use big data, complex algorithms and real time analytics to predict the outbreaks of the disease (Grover and Leskovec, 2016). These methods can also predict spreading patterns and burden to health. Such approaches include simple mathematical models to machine learning and bioinformatics-based approaches, and offer complementary insight into the disease dynamics (Dixon *et al*, 2022). The number and complexity of health-related data have grown, which has led to an augmented number of different types of computational models (Stojanović *et al*, 2019). Nevertheless, since there are numerous modelling strategies, there is an immediate necessity to reach comparisons with the usefulness, velocity, and information needs as well as correctness (Reveil and Chen, 2022). Such comparisons, in

To cite this article: Hasnain, M. S. (2025). *Computation and Comparative Modelling Approaches in Infectious Disease Prediction: A Review*. Journal of Epidemiology and Infection Biology 2(1):39-42.

addition to showing the relative merits and drawbacks of each approach, assist researchers and public health policymakers in choosing the most fitting tools regarding specific settings. Because of the interdisciplinary domain of infectious disease modelling, a blend of biology, medicine, statistics and computer science, there is an essential need to critically evaluate and compare these models.

This review will seek to make a thorough comparison between the major computational models being used in the prediction of an infectious disease (Egli *et al*, 2020; Santangelo *et al*, 2023). It can be categorized into four major groups, i.e. epidemiological models, simulation models, mathematical models, and bioinformatics-based. The approaches are going to be detailed against how they are structured, where they are applicable, their computing performance, and the dependency on data. Finally, this review will be useful in comprehending the way these models could be integrated or chosen depending on types of disease, timeline of predictions, and available data.

Epidemiological Modelling Approaches

Epidemiological models serve as foundational tools for disease surveillance and outbreak prediction(Horrocks and Bauch, 2020). These models utilize structured and unstructured data from surveillance systems, environmental records, and digital platforms to monitor and forecast disease trends (Lloyd and May, 2001). Classical epidemiological methodologies include outbreak investigations and disease surveillance systems, which have significantly evolved with the integration of computational algorithms and whole-genome sequencing (Dixon *et al*, 2022; Tao *et al*, 2023). Event-based surveillance platforms such as Google Flu Trends, ProMED-Mail, and HealthMap gather data from news sources, social media, and governmental reports to identify emerging disease threats.

Recent advances in computational epidemiology have led to the development of algorithmic prediction models for vector-borne and waterborne diseases, often using meteorological data and population demographics (Meng *et al*, 2019; Kim and Ahn, 2021). Techniques like Bayesian regression, support vector regression, and neural networks enhance predictive accuracy, particularly in spatial epidemiology (Krantz *et al*, 2020; Yang *et al*, 2020a; Yang *et al*, 2020b). While epidemiological models are valuable for real-time tracking and public health response planning, they are heavily reliant on the quality, timeliness, and granularity of input data, which can limit their accuracy and applicability.

Simulation Modelling Approaches

The food industry requires One Health approach as an essential method to tackle SCVs. The One Health approach unifies Simulation models are used to create hypothetical models of real-life systems and to determine how well-studied diseases will behave, under a variety of testing conditions. These are system dynamics, agent-based models (ABMs), discrete event models (DE) and Monte Carlo simulations (Stojanović *et al*, 2019; Waleed *et al*, 2020; Latkowski and Dunin-Kęplicz, 2021). Both approaches are based on their different ways of representing time and processes, and uncertainty. As an example, ABMs model interactions of individual agents, and have an emergent behavior at the population level, where system dynamics models concentrate on aggregated feedback loops and flows over time (Waleed *et al*, 2020; Sansone *et al*, 2022). Raw data available has been useful together with simulation frameworks such as, CoPE, SIM -D, PanSim and CityCOVID which have been effective in modelling complex interventions like, social distancing, vaccinations and mobility restrictions (Latkowski and Dunin-Kęplicz, 2021). Moreover, the incorporation of such instruments as the harmonic decomposition or wavelet analysis enabled the simulation models to detect the seasonal and environmental impacts on the spread of diseases (Krantz *et al*, 2020; Zhang *et al*, 2023). Although flexible and realistic, simulation models are often computationally taxing and necessarily parameter intensive, thus not as suitable in circumstances requiring very quick deployment when insufficient data and expertise is at hand.

Mathematical Modelling Approaches

Mathematical models provide an explanation of how a disease is transmitted and progresses by use of equations (Yang *et al*, 2020b; Tedeschi, 2023). These models are classified in detail as deterministic and stochastic frameworks. Discrete epidemic models relatively common are the classical compartmental models, i.e. Susceptible-Infected-Recovered (SIR), and Susceptible-Exposed-Infectious-Recovered (SEIR), used to gain insight into disease thresholds, epidemic curves, and effects of interventions (da Silva *et al*, 2020; Yang *et al*, 2020b; Liu *et al*, 2021). These models have been used to set policymaking in departments during epidemics such as the coronavirus pandemic (COVID-19) and measles.

Further developed procedures like the Sparse Identification of Nonlinear Dynamics (SINDy) give dynamic system insights based on the case notification data (Horrocks and Bauch, 2020). Stochastic models, of which Ebola models are also examples, bring elements of probability, which explains how these models take into consideration random transmission events (Asher, 2018). Although mathematical models are easy to introduce and apply to limited calculation resources, they tend to rely on assumptions of population homogeneity and invariance of parameters that are less applicable in real-life conditions, where heterogeneity and variability is likely to appear.

Bioinformatics-Based Modelling Approaches

The blossom of genomic and transcriptomic data has led to bioinformatics becoming an effective method of predicting infectious diseases in a molecular form. These models are based on machine learning techniques consisting of random forests (RF), support vector machines (SVM) and k-nearest neighbors (kNN), to study genetic sequences and determine disease associations, drug targets, and resistant markers (Wilke and Adami, 2002; Brett *et al*, 2017; Wolf *et al*, 2018; Zhang *et al*, 2020). Investigation of miRNA-disease or lncRNA-disease associations is based on tools such as Path-Based MiRNA-Disease Association (PBMDA) and Ensemble Kernel Ridge

Regression-based models (EKRRMDA) (You *et al*, 2017; Wolf *et al*, 2018; Horrocks and Bauch, 2020). These in silico techniques are especially applicable when it comes to learning biology of pathogens, regulation of genes, and molecular interaction. Bioinformatics models have the advantage of being quicker and cheaper than experimental ones, and thus they are useful in achieving rapid drug screening and characterization of pathogens. They are however very reliant on having the availability and quality of, large scale, robust genomic data. In addition, some ML models might not be interpretable due to the complexity and opacity aspect which is important to clinical validation.

Comparative Analysis of Modelling Approaches

Every model has its own range of benefits and limitations (Table 1). Epidemiological models are quick and are effective in short-term projections and real-time response though timely and accurate surveillance are very important. The simulation modelling is informative to behavior and policy situations although labor-intensive and very demanding with regards to data. Mathematical models are mathematically transparent, and computationally cheap, although they can distort complicated dynamics. On the contrary, the bioinformatics-based modelling is better in scaling to the molecular level and prediction; however, they need large amounts of quality omics-scale data and powerful computing capabilities. Regarding the speed, mathematical and bioinformatics models tend to take less time in case of the availability of data, whereas simulation and epidemiological models might be time-consuming as they utilize the use of real-time and big-scale data. Data type required is also different-epidemiological and simulation models require population level and behavior data, the mathematical models require parameterized equations and bioinformatics models need genomic and transcriptomic sequences.

Table 1: Comparison of Infectious Disease Modelling Approaches

Modelling Approach	Primary Use	Data Requirements	Strengths	Limitations
Epidemiological Models	Real-time tracking, outbreak detection	Surveillance, environmental, social media	Fast, practical, useful for public health planning	Data quality dependent, may lack granularity
Simulation Models	Policy scenario analysis, behavioural modelling	Demographic, spatial, intervention data	Captures complex dynamics and interactions, flexible	High computational cost, requires detailed assumptions
Mathematical Models	Theoretical analysis, short- to mid-term forecasts	Transmission rates, population data	Simple to implement, analytically tractable, computationally efficient	Oversimplifies, assumes population homogeneity
Bioinformatics Models	Molecular prediction, drug/target discovery	Genomic, transcriptomic, molecular datasets	High-resolution, scalable, useful in personalized medicine and pathogen analysis	Requires omics data, less interpretable, not suited for population-level modelling

Modelling approaches applied on disease systems

The correspondence analysis (CA) of majority of studies conducted during 2010 to 2024 revealed the relationship between various types of pathogens (virus, bacteria, parasitic and fungi) and the body systems affecting (Fig. 1). Majority of the viral diseases were strongly linked with respiratory, immune and reproductive systems, which pointed to a high rate of viral infections in these fields-COVID-19, the flu, HIV are a few of such viral infections (Raghavan *et al*, 2017). Bacterial pathogens specifically relate to certain diseases of the nervous, respiratory, and reproductive systems which include bacterial meningitis, pneumonia and sexually transmitted diseases such as chlamydia or syphilis (Schreuder *et al*, 2021; Yang and Li, 2021; Goyal and Singh, 2023). Diseases related to parasites like malaria and leishmaniasis reveal great relations with the hematologic, hepatic, integumentary and visceral systems indicating that they are systemic and blood related (Semenza and Suk, 2018). There are fungal pathogens and they are linked to the nervous system which is in line with fungal meningitis (Choi *et al*, 2016; Egli *et al*, 2020).

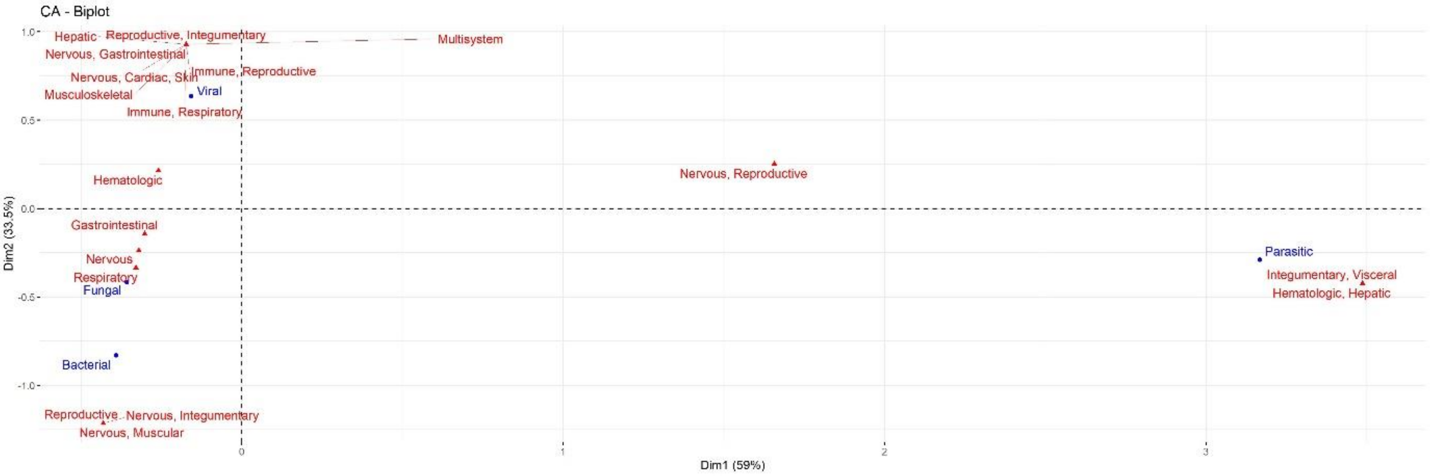

Fig. 1: Correspondence analysis of all the diseases studied with modelling approaches.

In conclusion, there is no particular model that would be better than the other universally; it only matters which one to choose depending on the use-case scenario, data availability, resolution that is necessary, and compute potential. Hybrids or integrated modelling solutions may provide more complete infectious disease prediction systems, which use elements of more than one model.

Declaration of Competing Interest

The authors declare that they have no competing or conflict of interests.

Author Contributions

MSH: Conceptualization, formal analysis, Writing—review and editing. All authors have read and agreed to the published version of the manuscript.

References

Asher, J. (2018). *Forecasting Ebola with a regression transmission model*. Epidemics 22:50-55. doi: https://doi.org/10.1016/j.epidem.2017.02.009

Brett, T. S., J. M. Drake, and P. Rohani. (2017). *Anticipating the emergence of infectious diseases*. J. R. Soc. Interface. 14(132):20170115. doi: doi:10.1098/rsif.2017.0115

Choi, J., Y. Cho, E. Shim, and H. Woo. (2016). *Web-based infectious disease surveillance systems and public health perspectives: a systematic review*. BMC Public Health 16(1):1238. doi: 10.1186/s12889-016-3893-0

da Silva, R. G., M. Ribeiro, V. C. Mariani, and L. D. S. Coelho. (2020). *Forecasting Brazilian and American COVID-19 cases based on artificial intelligence coupled with climatic exogenous variables*. Chaos Solitons Fractals 139:110027. doi: 10.1016/j.chaos.2020.110027

Dixon, S., R. Keshavamurthy, D. H. Farber, A. Stevens, K. T. Pazdernik, and L. E. Charles. (2022). *A Comparison of Infectious Disease Forecasting Methods across Locations, Diseases, and Time*. Pathogens 11(2)doi: 10.3390/pathogens11020185

Egli, A., J. Schrenzel, and G. Greub. (2020). *Digital microbiology*. Clinical Microbiology and Infection 26(10):1324-1331. doi: https://doi.org/10.1016/j.cmi.2020.06.023

Goyal, S., and R. Singh. (2023). *Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques*. Journal of Ambient Intelligence and Humanized Computing 14(4):3239-3259. doi: 10.1007/s12652-021-03464-7

Grover, A., and J. Leskovec. (2016). *node2vec: Scalable Feature Learning for Networks*. In: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 855-864 p.

Horrocks, J., and C. T. Bauch. (2020). *Algorithmic discovery of dynamic models from infectious disease data*. Scientific Reports 10(1):7061. doi: 10.1038/s41598-020-63877-w

Keshavamurthy, R., S. Dixon, K. T. Pazdernik, and L. E. Charles. (2022). *Predicting infectious disease for biopreparedness and response: A systematic review of machine learning and deep learning approaches*. One Health 15:100439. doi: https://doi.org/10.1016/j.onehlt.2022.100439

Kim, J., and I. Ahn. (2021). *Infectious disease outbreak prediction using media articles with machine* *learning models*. Scientific Reports 11(1):4413. doi: 10.1038/s41598-021-83926-2

Krantz, S. G., P. Polyakov, and A. Rao. (2020). *True epidemic growth construction through harmonic analysis*. J Theor Biol 494:110243. doi: 10.1016/j.jtbi.2020.110243

Latkowski, R., and B. Dunin-Kęplicz. (2021). *An Agent-Based Covid-19 Simulator: Extending Covasim to the Polish Context*. Procedia Computer Science 192:3607-3616. doi: https://doi.org/10.1016/j.procs.2021.09.134

Liu, X.-X., S. J. Fong, N. Dey, R. G. Crespo, and E. Herrera-Viedma. (2021). *A new SEAIRD pandemic prediction model with clinical and epidemiological data analysis on COVID-19 outbreak*. Applied Intelligence 51(7):4162-4198. doi: 10.1007/s10489-020-01938-3

Lloyd, A. L., and R. M. May. (2001). *Epidemiology. How viruses spread among computers and people*. Science 292(5520):1316-1317. doi: 10.1126/science.1061076

Meng, Q., X. Liu, J. Xie, D. Xiao, Y. Wang, and D. Deng. (2019). *Epidemiological characteristics of bacillary dysentery from 2009 to 2016 and its incidence prediction model based on meteorological factors*. Environ Health Prev Med 24(1):82. doi: 10.1186/s12199-019-0829-1

Raghavan, A., D. E. Rimmelin, K. V. Fitch, and M. V. Zanni. (2017). *Sex Differences in Select Non-communicable HIV-Associated Comorbidities: Exploring the Role of Systemic Immune Activation/Inflammation*. Curr HIV/AIDS Rep 14(6):220-228. doi: 10.1007/s11904-017-0366-8

Reveil, M., and Y.-H. Chen. (2022). *Predicting and preventing COVID-19 outbreaks in indoor environments: an agent-based modeling study*. Scientific Reports 12(1):16076. doi: 10.1038/s41598-022-18284-8

Sansone, M., P. Holmstrom, S. Hallberg, R. Nordén, L.-M. Andersson, and J. Westin. (2022). *System dynamic modelling of healthcare associated influenza -a tool for infection control*. BMC Health Services Research 22(1):709. doi: 10.1186/s12913-022-07959-7

Santangelo, O. E., V. Gentile, S. Pizzo, D. Giordano, and F. Cedrone. (2023). *Machine Learning and Prediction of Infectious Diseases: A Systematic Review*. Mach. Learn. Knowl. Extr. 5(1):175-198.

Schreuder, A., M. Prokop, E. T. Scholten, O. M. Mets, K. Chung, F. A. A. Mohamed Hoesein, C. Jacobs, and C. M. Schaefer-Prokop. (2021). *CT-Detected Subsolid Nodules: A Predictor of Lung Cancer Development at Another Location?* Cancers (Basel) 13(11)doi: 10.3390/cancers13112812

Semenza, J. C., and J. E. Suk. (2018). *Vector-borne diseases and climate change: a European perspective*. FEMS Microbiol Lett 365(2)doi: 10.1093/femsle/fnx244

Stojanović, O., J. Leugering, G. Pipa, S. Ghozzi, and A. Ullrich. (2019). *A Bayesian Monte Carlo approach for predicting the spread of infectious diseases*. PLOS ONE 14(12):e0225838. doi: 10.1371/journal.pone.0225838

Tao, D., D. Zhang, R. Hu, E. Rundensteiner, and H. Feng. (2023). *Epidemiological Data Mining for Assisting with Foodborne Outbreak Investigation*. Foods 12(20):3825.

Tedeschi, L. O. (2023). *Review: The prevailing mathematical modeling classifications and paradigms to support the advancement of sustainable animal production*. animal 17:100813. doi: https://doi.org/10.1016/j.animal.2023.100813

Waleed, M., T.-W. Um, T. Kamal, A. Khan, and Z. U. Zahid. (2020). *SIM-D: An Agent-Based Simulator for Modeling Contagion in Population*. Applied Sciences 10(21):7745.

Wilke, C. O., and C. Adami. (2002). *The biology of digital organisms*. Trends in Ecology & Evolution 17(11):528-532. doi: https://doi.org/10.1016/S0169-5347(02)02612-5

Wolf, F. A., P. Angerer, and F. J. Theis. (2018). *SCANPY: large-scale single-cell gene expression data analysis*. Genome Biology 19(1):15. doi: 10.1186/s13059-017-1382-0

Yang, W., J. Zhang, and R. Ma. (2020a). *The Prediction of Infectious Diseases: A Bibliometric Analysis*. Int J Environ Res Public Health 17(17)doi: 10.3390/ijerph17176218

Yang, Y., and D. Li. (2021). *Medical Data Feature Learning Based on Probability and Depth Learning Mining: Model Development and Validation*. JMIR Med Inform 9(4):e19055. doi: 10.2196/19055

Yang, Z., Z. Zeng, K. Wang, S. S. Wong, W. Liang, M. Zanin, P. Liu, X. Cao, Z. Gao, Z. Mai, J. Liang, X. Liu, S. Li, Y. Li, F. Ye, W. Guan, Y. Yang, F. Li, S. Luo, Y. Xie, B. Liu, Z. Wang, S. Zhang, Y. Wang, N. Zhong, and J. He. (2020b). *Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions*. J Thorac Dis 12(3):165-174. doi: 10.21037/jtd.2020.02.64

You, Z.-H., Z.-A. Huang, Z. Zhu, G.-Y. Yan, Z.-W. Li, Z. Wen, and X. Chen. (2017). *PBMDA: A novel and effective path-based computational model for miRNA-disease association prediction*. PLOS Computational Biology 13(3):e1005455. doi: 10.1371/journal.pcbi.1005455

Zhang, W., S. Liu, N. Osgood, H. Zhu, Y. Qian, and P. Jia. (2023). *Using simulation modelling and systems science to help contain COVID-19: A systematic review*. Systems Research and Behavioral Science 40(1):207-234. doi: https://doi.org/10.1002/sres.2897

Zhang, Y., X. Zhang, A. N. Lane, T. W. Fan, and J. Liu. (2020). *Inferring Gene Regulatory Networks of Metabolic Enzymes Using Gradient Boosted Trees*. IEEE J Biomed Health Inform 24(5):1528-1536. doi: 10.1109/jbhi.2019.2931997