A Geostatistical and Machine Learning Approach to Modelling the Effects of Water Stress, Climate Change, and Sanitation Infrastructure on the Spatial Distribution of Infectious Diseases

Jumoke Ifeoluwa Ogunremi [1], Philena Amie Sombo [1]

[1] *Western Illinois University*
Macomb, IL, 61455, USA

DOI: 10.22178/pos.129-30

LCC Subject Category: T1-995

Received 25.03.2026
Accepted 24.04.2026

Corresponding Author:
Jumoke Ifeoluwa Ogunremi

Abstract. Infectious disease is becoming an increasingly important environmental issue, particularly in rapidly urbanising LMICs that lack the infrastructure needed to manage these challenges. It is important to map their spatial and temporal interactions for prevention and adaptation. We discuss novel advances in geostatistics and machine learning (ML) designed to map the spatial structure of these factors that influence the spatial pattern of infectious diseases; this includes core geostatistical methods such as kriging, variogram modelling, spatial regression, and autocorrelation analysis, as well as ML models such as random forests, convolutional neural networks (CNNs), and long short-term memory (LSTM) networks. The advantage of multimethod geostatistical-ML approaches over monolithic ones is that they achieve greater accuracy, interpretability, and uncertainty management. Case studies from 2020–2025 demonstrate that remote sensing, hydrologic and infrastructure data can be used to augment cholera, malaria and dengue models. Some of the challenges are data quality assurance, data interpretability, scalability, and privacy issues related to health data. Future priorities should focus on explainable AI, federated learning, and climate-health digital twins, which will help create resilient, secure, and future-proof models applicable globally. Finally, the integration of geostatistics and machine learning is a promising interdisciplinary solution for disease prediction and for bolstering population resilience in a rapidly evolving world.

Keywords: Geostatistics; Machine Learning; Climate Change; Water Stress; Sanitation Infrastructure; Infectious Diseases; Spatial Epidemiology.

INTRODUCTION

Infectious diseases are among the most significant health and sustainable development issues worldwide. Despite all the strides made in health, sanitation, and technology, millions of people die annually of malaria, cholera, dengue, and diarrheal diseases, the majority of whom live in low and middle-income countries. The distribution of these diseases is very uneven and is influenced by environmental, climatic, and socio-economic factors. Recent studies have increasingly focused on the combined effects of water stress, climate change, and sanitation infrastructure on the spatial and temporal distribution of these diseases, as well as their interactions [1]. Limited water access, poor sanitation and hygiene practices, and warmer temperatures, altered rainfall patterns, and extreme weather events create conditions where pathogens can thrive, and vectors and reservoirs can flourish. The dynamics are important for predicting outbreaks and for population-level strategies. Water stress – the imbalance between water demand and supply is an important driver of health. The United Nations (2022) reported that more than 2 billion people are at risk of water stress, and almost half of the world's population will be at risk of water shortages within 30 years. The scarcity also means unsafe water is used, leading to the spread of cholera, typhoid, and hepatitis E, and poor water for hand washing and other personal hygiene needs, which contribute to the spread of pathogens via the faecal-oral route [2, 3]. A lack of regular rainfall and inadequate irrigation methods can also lead to stagnant water, which supports mosquito breeding and contributes to the transmission of malaria and dengue [4]. To predict risk, spatially explicit water models are essential.

The effects of climate change also include changes in ecology and epidemiology due to shifts in temperature, humidity, and precipitation, which can affect vector and pathogen biology. Geographical range shifts of malaria and dengue vectors, outbreaks of diarrheal diseases due to disruptions in water supply from floods, droughts, and hurricanes [5, 6]. Traditional deterministic epidemiological models do not adequately capture these non-linear and delayed effects, which have driven the development of geostatistical models and machine learning approaches. Sanitation is a modulator of the environmental effects on transmission. WHO (2022) estimated that over 1.7 billion people have no access to basic sanitation and that over 430,000 people die annually from diarrheal disease. The lack of sanitation is a powerful multiplier of climate variability: rainfall leads to drainage failures and the spread of pathogens; water scarcity brings them together. The vulnerabilities are especially marked in slums in fast-growing urbanising areas [7]. Therefore, knowledge of the spatial variability of sanitation, climate and hydrology is important in mapping risk. Geostatistics and machine learning techniques are well-suited for this task. Geostatistics - kriging, variogram modelling, and spatial autocorrelation - measure spatial correlation and fill in the gaps in risk [8], but make assumptions of stationarity and linearity, and are limited in complex systems. Researchers use machine learning techniques such as random forests, gradient-boosted trees, and neural networks to model high-dimensional relationships among environmental, climatic, and socio-economic variables [9]. Integrated models combine the best of both worlds: spatial structure/uncertainty from geostatistics, prediction from ML. For example, authors [10] added a random forest with a kriging model for the purposes of improving the malaria risk prediction by 20% in the Ethiopian highlands; On the other hand, authors [11] applied a random forest geospatial model with sanitation, water stress and precipitation to predict cholera in West Africa.

This review compiles the latest studies (2020-2025) on geostatistical and ML prediction of infectious disease transmission cycles, using water stress, climate change, and sanitation. It:

1) gives a theoretical background of each approach,

2) reviews recent implementations and hybrid combinations, and

3) points out relevant challenges for data fusion, scalability of models, and translating policies.

The aim is to illustrate the use of a hybrid spatial modelling approach to improve disease surveillance and adaptive health interventions, in accordance with SDG 3 (Good Health and Well-Being) and SDG 6 (Clean Water and Sanitation).

Background and Related Work. Since the 1920s, the goal of spatial epidemiology has been to account for the spatial distribution of disease using geographical, climatic, and infrastructural information, and this has become increasingly important in the changing world today. Today, fine-scale geospatial information, along with the complexity of computing and ML, enables us to model the heterogeneous, non-linear, and dynamic distribution of disease [9]. Spatial epidemiological models originated in the mining and environmental sciences, where geostatistics was needed. Spatial dependence and risk prediction at unsampled sites are measured and estimated using variogram analysis, kriging, and spatial autocorrelation [8]. Early models used sparse data to interpolate malaria prevalence in Africa and related rainfall, temperature, and altitude to risk using spatial and geographically weighted regression (GWR) [12]. Researchers used spatial regression to predict cholera hotspots in coastal Bangladesh based on groundwater salinity, and they applied variogram-based geostatistics to study leishmaniasis and schistosomiasis in Latin America.

However, in most geostatistical models, it is assumed that the system is stationary and linear, assumptions that are generally not satisfied in real epidemiological systems (which are typically non-stationary, heterogeneous, and coupled). Consequently, researchers have turned to ML methods capable of handling high-dimensional data. Various methods have been used to combine environmental factors (such as temperature, vegetation indices, and soil moisture) with socio-economic factors (population density and sanitation coverage) to generate disease risk maps, including support vector machines (SVMs), decision trees, random forests, and neural networks

[13]. Authors [14] used an ensemble ML to map malaria transmission over the African continent using satellite and survey data. Authors [15] showed that random forests and gradient boosting models could forecast dengue in Southeast Asia, where rainfall variability and drought prevalence are considered important variables. While predictions are a strength of ML, they lack the spatial interpretability that geostatistics offers. The autocorrelation measures, such as Moran's I and the Getis-Ord Gi, are explicit measures of dispersion, whereas the ML models are "black boxes"; this has led to the development of new hybrid geostatistical-ML approaches that aim to combine the strengths of spatial statistics and AI for interpretation and prediction. Authors [10] applied random forests and kriging to create high-resolution malaria risk maps in Ethiopia, achieving higher spatial accuracy than previously reported. Authors [10] employed random forests and kriging to generate malaria risk maps in Ethiopia with higher spatial accuracy than previously reported. In addition, remote sensing has been used to improve the accuracy of the cholera prediction in the Ganges Delta through the application of ML: authors [16] leveraged vegetation and water reflectance indices derived from Sentinel data to enhance the prediction of cholera in the Ganges Delta region by applying ML; authors [11] showed that the soil moisture index derived from satellite data is strongly associated with sanitation and cholera in West Africa; this is an important development, driven by sanitation and socio-economic factors. Although income and education were important, poor sanitation quality was a significant predictor of diarrheal disease in sub-Saharan African cities, according to the authors [7]. Researchers have effectively reproduced lagged impacts of environmental factors using spatiotemporal machine learning methods such as RNNs and LSTMs. Authors [17] built an LSTM model for Dengue in Indonesia that outperforms static models, and hybrid space-LSTM models were applied to determine the lagged effects of rainfall anomalies on cholera in Bangladesh. However, there are still many issues to be addressed. Researchers often develop disease- or location-specific models that cannot be easily transferred; they also collect data with highly variable quality and resolution, especially in resource-poor environments; apply methods inconsistently; and struggle to interpret models and quantify uncertainty. The next steps should be on data standardisation, explainable AI, and connecting up multi-source data such as climate, water, sanitation, and population. The message from the facts is clear: geostatistics and ML complement each other, and their combination is the way forward for infectious disease modelling in the face of environmental and infrastructural challenges.

RESULTS AND DISCUSSION

Geostatistical Modelling Techniques. Geostatistics are an important tool for quantifying the spatial variability and dependence of infectious diseases in relation to environmental and built features. Geostatistics is a quantitative approach to conceptualise infectious disease distribution as a continuous surface from a finite, irregularly spaced set of observations of disease density [8]. It can be used to model the non-random distribution of infectious diseases driven by climate, hydrology, and socio-economic factors. The subsequent subsections briefly describe the main techniques of infectious disease mapping.

Spatial Interpolation is a technique for estimating risk at unsampled locations using nearby samples. The most commonly used method is kriging, which exploits spatial autocorrelation to produce predictions with low estimation error and is better than inverse distance weighting (IDW) methods by incorporating both similarity and distance. Ordinary kriging assumes stationarity, but regression and universal kriging can include variables (such as temperature, rainfall, and sanitation coverage). Co-kriging uses correlated information (e.g., rainfall and vegetation indices) [18]. Other studies have also employed the Kriging technique to create maps of malaria prevalence from national surveys and to map climatic and topographic factors [10].

Variogram modelling is a fundamental part of kriging analysis and describes spatial correlation as a function of distance. The nugget (measurement or micro-scale variance), the sill (total variance) and the range (distance at which the observations are independent) are used to quantify spatial variance [19]. In Bangladesh, for instance, a variogram analysis indicated a correlation scale of ∼40 km for cholera, suggesting regional-scale environmental risk factors, such as groundwater salinity [18]. Anisotropic Variograms are used for models that allow for different variances in different directions.

Spatial autocorrelation is a test for whether the distribution of disease at one location is associated with that at another. Researchers used Moran's I and Geary's C to determine global clustering in Southeast Asia. They linked the observed clustering to non-compliance with hygiene standards and irregular rainfall, which could be addressed by targeted interventions such as water quality management and vaccination [15]. In West Africa, Getis-Ord Gi* was applied to detect hot and cold spots, which were attributed to poor hygiene and irregular rainfall and prioritised for intervention, such as improvements in water quality and vaccination [11].

Spatial regression takes the description a step further, inferring the effects of environmental and socio-economic predictors on disease while accounting for spatial effects. Spatial Lag Models (SLM) model the effects of nearby observations, and Spatial Error Models (SEM) account for spatial dependence in model residuals due to unmeasured forces. Authors [14] showed that there are spatial lags in temperature and precipitation associated with malaria in East Africa. Geographically Weighted Regression (GWR) and Multiscale GWR (MGWR) allow the impact of a variable to vary spatially [20]. Authors [7] used GWR to show that sanitation has varying effects in African cities, depending on population density and infrastructure.

In addition, geostatistics offers uncertainty quantification. Kriging variance maps, resampling and BH models provide a measure of prediction uncertainty. Maps of predicted risk and uncertainty help decision-makers focus on the high-risk, highly uncertain areas [21]. The development of Bayesian geostatistical modelling has greatly shaped how researchers use survey-based and satellite-derived environmental covariates to model malaria risk in sub-Saharan Africa [8].

Machine Learning and AI-Based Approaches. Artificial intelligence (AI) and machine learning (ML) are becoming increasingly crucial in spatial epidemiology, particularly for capturing non-linear relationships among environmental risk factors, sanitation services, and disease. Moreover, ML methods do not impose any prior assumptions about the distribution, and can help discover hidden patterns in a variety of large-scale data. With the increasing accessibility of open health data, climate archives and remote-sensing imagery, the use of ML and real-time predictions, and fine-scale risk mapping are now possible [9]. The most popular supervised algorithms for disease-risk mapping are random forests, support vector machines (SVMs), and gradient boosting. Random forests can handle multicollinearity and missing data, and they can handle non-linear interactions. Using random forests as a mapping tool, authors [11] found that the most important variables for predicting cholera risk in West Africa were sanitation and rainfall variability, and that performance was better than that of traditional regression. Authors [15] used gradient boosting to link drought event frequency and dengue in Southeast Asia, demonstrating the importance of water scarcity to vector ecology. In Southeast Asia, SVMs correctly predict malaria risk from remote-sensing-based temperatures, vegetation, and rainfall anomalies with over 85% accuracy [22]. Still, they are sensitive to kernel functions and are not well scalable unless optimised.

Deep learning has introduced a revolution in spatiotemporal modelling by extracting the hierarchical features of data. The same type of CNN used for images is now being adopted for gridded maps of the environment and diseases. Using a CNN trained on satellite and climate data, authors [23] were able to predict dengue outbreaks in Thailand, accounting for small-scale variability that was not detectable via traditional spatial regression. Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) networks, are ideal for modelling time lags from the environment to disease. Authors [17] reported an 18% improvement in dengue forecasting across Indonesia using lagged rainfall and temperature in an LSTM. Authors [6] state that temporal modelling is crucial for diseases sensitive to climate, such as cholera, malaria, and leptospirosis.

Some other useful techniques include unsupervised learning, such as clustering, self-organising maps (SOMs), and principal component analysis (PCA). In poor surveillance districts, authors [24] used SOMs to group districts in Bangladesh based on sanitation, rainfall, and water quality, which were strongly related to cholera incidence, suggesting that the clusters could be useful in those districts. PCA can be used to identify the smallest number of correlated parameters that describe the data; these parameters represent meaningful components of the data.

Explainable AI (XAI) is on the rise. While deep models have high accuracy, they lack transparency for policy. SHAP and LIME provide explanations of predictions at the global and local levels. Authors [11] used SHAP values in random forests to determine that precipitation anomalies and access to sanitation services are the most influential factors in cholera risk, enabling targeted interventions. However, there are issues. The effectiveness of models depends on the quality and representativeness of the data, and data quality is often poorly documented in resource-poor settings, leading to biases that can result in inequitable health policies. Deep learning models require large, labelled datasets and substantial computational resources, but researchers are now mitigating these challenges through new paradigms such as cloud computing, transfer learning, and federated learning. To conclude, ML and AI have revolutionised spatial epidemiology, and hybrid geostatistical-ML systems will remain a key player in the future, amid the increased threat of climate and environmental hazards.

Case Studies and Applications. Integrated geostatistical ML techniques have been used throughout Africa, Asia and Latin America to estimate environmental variables and predict outbreaks of infectious disease. The Malaria Atlas Project from authors [14] was an early large-scale demonstration that used satellite remote sensing, random forests, and kriging to produce high-resolution malaria risk maps across sub-Saharan Africa. The prevalence was strongly related to rainfall, temperature and vegetation, and the hybrid model performed better than the standard spatial regression by over 15% in data-sparse areas.

Authors [11] employed the random forest (RF) machine learning model, in combination with Moran's I autocorrelation, to model cholera risk in West Africa using satellite rainfall, water stress indices, and sanitation coverage. The framework identified areas of critical sanitation conditions (hot zones) in conjunction with precipitation variability, and geostatistical diagnostics added interpretability and consistency, thereby facilitating targeted public-health action. In coastal Bangladesh, authors [18] integrated support vector regression with variogram analysis. They identified groundwater salinity and post-monsoon rainfall as the major factors, leading to a cholera risk map validated by a field survey.

Authors [10] demonstrated that a random forest ordinary kriging hybrid technique generated continuous malaria risk surfaces in the Ethiopian highlands with an increase in accuracy of 20% compared with malaria-only techniques, and also produced probabilistic uncertainty maps for allocation of resources. In Thailand, authors [23] used CNNs on urban heat, rainfall, and vegetation imagery to identify previously unseen urban clusters associated with poor drainage with over 90% accuracy. Authors [17] modelled dengue in Indonesia using LSTM networks with 94% accuracy, leveraging lagged temperature and humidity data, demonstrating the benefit of temporal memory for vector-borne disease forecasting.

Authors [25] applied GWR to a random forests model to explore how sanitation and diarrheal disease are shaped by spatial variation in semi-arid Brazil, finding that drought areas were at higher risk of poor sanitation, and that climate infrastructure interactions are synergistic. Authors [16] used Sentinel-2 data with gradient boosting for cholera prediction in the Ganges Delta, achieving 82% accuracy, and chlorophyll was used.

Table 1 – Summary of Key Case Studies Integrating Geostatistics and Machine Learning for Disease Modelling (2020–2025)

Study	Region/ Country	Disease Focus	Model / Methods	Key Predictors	Major Findings
[14]	Sub-Saharan Africa	Malaria	Random Forest + kriging	Rainfall, temperature, vegetation index	15% improvement in predictive accuracy; enhanced risk mapping for low-surveillance areas
[18]	Bangladesh (Coastal)	Cholera	Support Vector Regression + Variogram Analysis	Groundwater salinity, rainfall, groundwater depth	Identified high-risk zones; groundwater salinity is a major determinant of outbreaks.
[11]	West Africa	Cholera	Random Forest +	Rainfall variability,	Detected cholera hot

Study	Region/ Country	Disease Focus	Model / Methods	Key Predictors	Major Findings
			Moran's I	sanitation coverage, water stress	zones; improved interpretability via spatial diagnostics
[10]	Ethiopia	Malaria	Random Forest + Ordinary Kriging	Rainfall, temperature, and land cover	20% accuracy improvement; probabilistic risk mapping and uncertainty quantification
[23]	Thailand	Dengue Fever	Convolutional Neural Network (CNN)	Urban heat index, vegetation, rainfall	90% accuracy; identified new urban dengue clusters linked to poor drainage
[17]	Indonesia	Dengue Fever	Long Short-Term Memory (LSTM)	Lagged rainfall, humidity, temperature	94% accuracy; effective spatiotemporal forecasting
[25]	Brazil	Diarrheal Disease	GWR + Random Forest	Sanitation index, drought frequency, and income level	Revealed spatial heterogeneity; sanitation quality is a key local determinant
[16]	Ganges Delta	Cholera	Gradient Boosting + Remote Sensing	Water quality, turbidity, and vegetation health	82% accuracy; linked water pollution directly to disease outbreaks

Challenges and Limitations. Even with significant methodological advances, there are still obstacles to the practical application of geostatistical ML frameworks. These range from data quality through computational cost, interpretability, generalisation, ethics and policy translation.

The issues of data quality and spatial heterogeneity remain constant challenges. In low- and middle-income areas, disease surveillance is inadequate, atmospheric interference or missing data in satellite observations occur, and data collection methods vary from region to region [14]. There is also a lack of consistency in the sanitation and water-stress indicators, such as the definitions of "access to clean water", "latrine coverage", and "drainage quality" [7]. The harmonisation of multi-source datasets from climate, hydrology, infrastructure, and demographics also needs to be carefully handled across spatial and temporal scales, as errors can lead to biased results and erroneous conclusions.

Interpretability and transparency are major concerns in deep and ensemble models. Although these models have good predictive power, their internal logic can be difficult to understand due to their "black-box" nature [9], making it challenging to incorporate them into public health policymaking, where causal attribution is required for decision-making. Some approaches, such as Explainable AI (XAI) methods SHAP and LIME [11], are starting to tackle this issue but are underutilised in the epidemiological context.

Scalability and computational cost are also constraints for real-world deployment. Both deep learning and classical geostatistical approaches (kriging, the Bayesian hierarchical model) require extensive computational resources to achieve high-resolution disease mapping. Researchers address these challenges using cloud computing and GPU acceleration, but costs, reproducibility issues, and data transfer problems remain, particularly in resource-limited settings [17]. They are exploring promising directions such as approximate Bayesian inference and spatial subsampling.

There is limited model generalisation and transferability due to the context-sensitive nature of the relationship between the environment and disease. The impact of rainfall on malaria, for instance, depends on local vector ecology, whereas the impact of sanitation differs between urban and rural settings [25]. This is because the model may become overly complex and fit the specific patterns of the local data, leading to poor performance on test data. Transfer learning and domain adaptation are potential solutions which are not often used in spatial epidemiology.

Researchers face methodological tensions when they combine geostatistics and machine learning. Real data often violate stationarity and isotropy assumptions, so researchers must reconcile probabilistic geostatistical outputs with deterministic machine learning predictions. To avoid over-optimistic error estimates and inflated uncertainty [10], they rely on spatial cross-validation.

But there are other limitations, such as data ethics and privacy. Fine-grained spatial health information can be susceptible to re-identification threats, especially for disadvantaged groups [13], and mobile, social, and sensor information exacerbates privacy concerns. A way forward is federated learning and other privacy-preserving frameworks that enable collaborative model training without sharing raw data.

Uncertainty quantification is still not uniform among the methods. The uncertainty (kriging variance, Bayesian posteriors) is a natural part of geostatistics, but many ML models only give point estimates. Bayesian neural networks and ensemble-based uncertainty estimation are emerging ideas, but not yet commonly used in operational surveillance systems [21].

Future Research Directions. Innovation is needed across technical, ethical, and institutional aspects to advance future geostatistical ML disease modelling. Several directions seem to be particularly promising.

The convergence of AI and GIS will likely serve as the foundation for future decision-support systems. Environmental change monitoring using automated platforms that combine geostatistics, ML, and GIS, supported by remote sensing, IoT sensors, and epidemiological surveillance, can be implemented to detect emerging hotspots in near real time [13]. Interoperability of systems and institutions is a prerequisite.

Digital twins are a revolutionary field. A climate health digital twin would continually integrate environmental and epidemiological data to simulate the dynamics of diseases under different interventions and climate scenarios, making static risk maps adaptable and enabling real-time decision-making tools [9]. The spatial structure and uncertainty would come from geostatistics, and the predictive engine would come from ML.

Data scarcity and privacy issues are addressed by federated and transfer learning. Federated learning is a paradigm that enables multiple institutions to train shared models without sharing raw data, which is useful when sharing is legally or politically restricted. Transfer learning: Using models trained in one context (e.g., dengue in Southeast Asia) in another (Latin America), with a need for fewer extensive local datasets [17]. Incorporating these into geostatistical frameworks would allow for global connections and local calibration of disease prediction.

CONCLUSIONS

Environmental changes, water scarcity, and poor sanitation are major global health problems. With the rise of new and emerging infectious diseases driven by climate change, urbanisation, and infrastructural inequity, combining geostatistics with machine learning is a powerful tool for capturing the coupled spatial and temporal dynamics of these transformations. Geostatistics provides interpretability and formal uncertainty quantification, while ML offers high-dimensional pattern recognition and flexibility. They work together to identify emerging hotspots, attribute environmental drivers and anticipate future risks.

The results of this review have revealed that the hybrid frameworks, random forest kriging, GWR with ML, and deep learning geostatistical ensembles have enhanced predictive accuracy for environmentally mediated diseases. Examples from Africa, Asia, and Latin America show that adding water stress, sanitation, and climate variables improves model performance and policy relevance, and helps create fine-scale maps of water-related risks that inform early warning and resource allocation in vulnerable areas.

REFERENCES

1. WHO. (2023). World Health Statistics 2023. Retrieved from https://www.who.int/publications/b/69040

2. Bain, W., Yang, H., Shah, F. A., Suber, T., Drohan, C., Al-Yousif, N., DeSensi, R. S., Bensen, N., Schaefer, C., Rosborough, B. R., Somasundaram, A., Workman, C. J., Lampenfeld, C., Cillo, A. R., Cardello, C., Shan, F., Bruno, T. C., Vignali, D. a. A., Ray, P., & Kitsios, G. D. (2021). COVID-19 versus Non–COVID-19

Acute Respiratory Distress Syndrome: Comparison of Demographics, Physiologic Parameters, Inflammatory Biomarkers, and Clinical Outcomes. *Annals of the American Thoracic Society, 18*(7), 1202–1210. doi: 10.1513/annalsats.202008-1026oc

3. Hutton, G., & Chase, C. (2017). Water Supply, Sanitation, and Hygiene. In *The World Bank eBooks* (pp. 171–198). doi: 10.1596/978-1-4648-0522-6_ch9

4. Ahmed, S. H., Shaikh, T. G., Waseem, S., Zahid, M., Ahmed, K. A. H. M., Ullah, I., & Hasibuzzaman, M. A. (2024). Water-related diseases following flooding in South Asian countries – a healthcare crisis. *European Journal of Clinical and Experimental Medicine, 22*(1), 232–242. doi: 10.15584/ejcem.2024.1.29

5. Semenza, J. C., Rocklöv, J., & Ebi, K. L. (2022). Climate Change and Cascading Risks from Infectious Disease. *Infectious Diseases and Therapy, 11*(4), 1371–1390. doi: 10.1007/s40121-022-00647-3

6. Ebi, K. L., & Hess, J. J. (2020). Health risks due to climate change: Inequity in causes and consequences. *Health Affairs, 39*(12), 2056–2062. doi: 10.1377/hlthaff.2020.01125

7. Chaudhry, D. (2023). Climate Change and Health of the Urban Poor: The role of environmental justice. *The Journal of Climate Change and Health, 15*, 100277. doi: 10.1016/j.joclim.2023.100277

8. Diggle, P. J. & Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health Methods and Applications.* CRC Press.

9. Rezaei, M. (2025). Artificial intelligence in knowledge management: Identifying and addressing the key implementation challenges. *Technological Forecasting and Social Change, 217*, 124183. doi: 10.1016/j.techfore.2025.124183

10. Hussain, S. S. A., Bedi, S., Yadav, C. P., Mohanty, A. K., Mahatme, K., Tyagi, S., Krishnan, N. M. A., Kota, S. H., & Sharma, A. (2025). Hybrid models combining trend and seasonality components with machine learning algorithms provide accurate forecasting of malaria incidence. *PLOS Global Public Health, 5*(10), e0004500. doi: 10.1371/journal.pgph.0004500

11. Campbell, A. M., Racault, M., Goult, S., & Laurenson, A. (2020). Cholera Risk: A machine learning approach applied to essential climate variables. *International Journal of Environmental Research and Public Health, 17*(24), 9378. doi: 10.3390/ijerph17249378

12. Tewara, M. A., Yunxia, L., Lin, W., Barong, B. H., Mbah-Fongkimeh, P. N., Zhaolei, Z., Xinhui, L., Miao, Z., Liu, X., & Xue, F. (2020). Geographically weighted regression modelling of the spatial association between malaria cases and environmental factors in Cameroon. *Research Square*. doi: 10.21203/rs.2.13021/v2

13. Zhou, G., He, X., Yang, K., Li, L., Guo, H., Wang, G., Li, J., Chen, Y., & Yang, Y. (2023). Effects of temperature and relative humidity on behaviour and physiological indices in goats. *Small Ruminant Research, 229*, 107126. doi: 10.1016/j.smallrumres.2023.107126

14. Bhatt, S., Weiss, D. J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U., Battle, K. E., Moyes, C. L., Henry, A., Eckhoff, P. A., Wenger, E. A., Briët, O., Penny, M. A., Smith, T. A., Bennett, A., Yukich, J., Eisele, T. P., Griffin, J. T., Fergus, C. A., & Gething, P. W. (2015). The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. *Nature, 526*(7572), 207–211. doi: 10.1038/nature15535

15. Bisanzio, D., Bosa, H. K., Bakamutumaho, B., Nasimiyu, C., Atwine, D., Kyabayinze, D., Olaro, C., Breiman, R. F., Njenga, M. K., Mwebesa, H., Aceng, J. R., & Reithinger, R. (2025). Modelling case burden and duration of the Sudan Ebola virus disease outbreak in Uganda, 2022. *Emerging Infectious Diseases, 31*(9), 1829–1832. doi: 10.3201/eid3109.241545

16. Adewumi, I. O. (2025). AI for Cholera Outbreak Prediction, Real-Time Tracking, and Low-Resource Diagnostics using Federated and Privacy-Preserving Machine Learning. *Research Square*. doi: 10.21203/rs.3.rs-7441133/v1

17. Liu, Q., Chi, S., Dmytruk, K., Dmytruk, O., & Tan, S. (2022). Coronaviral infection and interferon response: The Virus-Host Arms Race and COVID-19. *Viruses, 14*(7), 1349. doi: 10.3390/v14071349

18. Manaf, M., Ali, Z., & Scholz, M. (2026). Integrating random forest-based regression kriging for analysing the spatial variability of rainfall in arid and semi-arid regions. *Scientific Reports, 16*(1), 5298. doi: 10.1038/s41598-026-36074-4

19. Cressie, N. (2021). A few statistical principles for data science. Australian & New Zealand *Journal of Statistics, 63*(1), 182–200. doi: 10.1111/anzs.12324

20. Brunsdon, C., Fotheringham, A., & Charlton, M. (2002). Geographically weighted summary statistics — a framework for localised exploratory data analysis. *Computers Environment and Urban Systems, 26*(6), 501–524. doi: 10.1016/s0198-9715(01)00009-6

21. Giorgi, E., Diggle, P. J., Snow, R. W., & Noor, A. M. (2018). Geostatistical Methods for Disease Mapping and Visualisation Using Data from Spatio-temporally Referenced Prevalence Surveys. *International Statistical Review, 86*(3), 571–597. doi: 10.1111/insr.12268

22. Chen, Z., Chong, K. C., Wong, M. C., Boon, S. S., Huang, J., Wang, M. H., Ng, R. W., Lai, C. K., & Chan, P. K. (2021). A global analysis of replacement of genetic variants of SARS-CoV-2 in association with containment capacity and changes in disease severity. *Clinical Microbiology and Infection, 27*(5), 750–757. doi: 10.1016/j.cmi.2021.01.018

23. Chutia, D., Borah, K., Singh, L., Sarmah, D., & Singha, L. (2026). Updates on dengue virus infection: Epidemiology, molecular pathogenesis, and clinical strategies. *Vacunas, 27*(3), 500652. doi: 10.1016/j.vacun.2026.500652

24. Agboka, K. M., Abdel-Rahman, E. M., Salifu, D., Kanji, B., Ndjomatchoua, F. T., Guimapi, R. A., Ekesi, S., & Tobias, L. (2025). Towards combining self-organising maps (SOM) and convolutional neural networks (CNN) to improve model accuracy: Application to phenotypic resistance in malaria vectors. *MethodsX, 14*, 103198. doi: 10.1016/j.mex.2025.103198

25. De Oliveira, A. F., Da Costa Leite, I., & Valente, J. G. (2015). Global burden of diarrheal disease attributable to the water supply and sanitation system in the State of Minas Gerais, Brazil: 2005. *Ciência & Saúde Coletiva, 20*(4), 1027–1036. doi: 10.1590/1413-81232015204.00372014