Sarcouncil Journal of Internal Medicine and Public Health

ISSN(Online): 2945-3674

Volume- 05| Issue- 01| 2026

Research Article **Received:** 01-12-2025| **Accepted:** 25-12-2025 | **Published:** 10-01-2026

Integrating Spatial Modeling and Machine Learning for Infectious Disease Surveillance in U.S. Urban and Rural Settings

***Francis Ssenabulya Ssemujju*[1] *and Deborah Solomon*[2]**

[1] *Washington University in St. Louis, USA*

[2] *Kwame Nkrumah University of Science and Technology, Ghana*

Abstract: Background: The use of spatial models together with machine learning is changing how infectious diseases are monitored across the United States, making it easier to detect threats earlier, sort risks more accurately, and improve live updates on outbreak conditions. Conventional monitoring methods are usually slowed by late reports, scattered information, and uneven regional reaches have difficulty reflecting how infections spread differently from place to place or shift over time. Scope: By contrast, combining machine learning with geographic information system (GIS) allows diverse fusion of epidemiology, environmental, behavioral, and movement patterns to be merged into detailed predictions that pinpoint rising danger zones much more reliably. This narrative review synthesizes peer-reviewed literature published from 2020 to present, focusing on recent advancements in spatial-machine learning surveillance, examining methodological evolution, contrasting applications in urban versus rural contexts, and highlighting ethical and operational considerations. Findings: Studies suggest cities gain strong digital networks and abundant information flows that boost accuracy, whereas remote areas deal with scarce records, weak systems, or unequal access, restricting broader use. Issues like varying local patterns, mismatched sources, or unclear results emphasize combining approaches, using consistent analysis steps, along with open oversight. Implications: Future directions emphasize designs that protect user data; machine learning models suited for remote regions alongside broader use of One Health ideas to tackle diseases tied to animals or shifting climates. This review, taken together, shows how combining spatial data with machine learning could reshape disease tracking, offering more accurate forecasts, fairer outcomes, and stronger adaptability for U.S. public health efforts in varied regions.

Keywords: Machine Learning, Spatial Modeling, Infectious Disease Surveillance.

INTRODUCTION

Infectious disease surveillance is a foundation of global public health preparedness and offers an eventual response and mitigation strategy for epidemics (Shaikh *et al*., 2022). The World Health Organization advocates the concept of integrated disease surveillance as a strategy to build a strong disease surveillance program, a concept which was later changed to Integrated Disease Surveillance and Response (Lee *et al*., 2024). The past two decades has seen an increase in the degree of connectivity, the variability and change in the climate patterns, and rapid urbanization, leading to an increasing threat to the potential rapid spreading of diseases, thus demanding efficient, data-intensive disease surveillance systems (Yarmohammadtoosky & Xie, 2024). Manual case detection, laboratories, and passive accumulation of data surveillance activities have remained the backbone of epidemic intelligence-gathering tools over the past several decades (Khorshid *et al*., 2025). The issue with this stage is associated with delay, incompleteness, and lack of subtlety or variability between space and time, which are features associated with modern outbreaks (Khorshid *et al*., 2025). The COVID-19 outbreak highlighted the fragility of traditional disease surveillance pathways, which experienced delayed timelines to identify cases, as well as grain differences between space and time decisions, particularly between rural as well as urban zones (Rahman *et al*., 2022).

Spatial modeling (SM) emerged historically as an extension of epidemiologic mapping, initially developed as a response to cholera epidemics in the 19th century, and has since developed into a complex, and sophisticated discipline encompassing geographic information systems, remote sensing, and space-temporal statistics (Kiani *et al*., 2023). Its main purpose is to reveal where and when infectious diseases emerge, spread, and persist by associating environmental, demographic, and behavioral variables with disease incidence (Kiani *et al*., 2023). Simultaneously, machine learning (ML) has significantly changed the interpretation of spatial associations between various geographic, environmental, and demographic factors (Ekundayo, 2024). It makes use of algorithmic learning processes involving complex data integration, including health, environmental, and mobility data, to identify hidden or complex patterns to make predictive statements (Chakilam, 2022). Together, spatial modeling and machine learning provide a complementary analytical framework for spatial data interpretation, enabling

***Corresponding Author:** Deborah Solomon

DOI- https://doi.org/10.5281/zenodo.18205791

not only the identification of spatial patterns but also deeper insight into how and why disease dynamics vary across locations(Kopczewska, 2022).

A recent systematic literature review by Wang *et al*. (2025) highlighted the potential of integrating ML with spatial modeling to enhance disease forecasting accuracy, enable hotspot detection, and guide public-health resource allocation (Wang *et al*., 2025). For instance, Ledien *et al*. (2022), demonstrated that integrating linear models with ML algorithms improved estimates of the force-of-infection for Chagas disease across multiple geographies, revealing spatially varying risk patterns that traditional regression approaches failed to capture (Ledien *et al*., 2022). Similarly, Javaid *et al*. (2023), showed that WebGIS-enabled monitoring systems for vector-borne diseases can integrate real-time sensor data, spatial risk maps, and automated alerts, underscoring the value of combining geoinformatics with AI-driven analytics for early response (Javaid *et al*., 2023). These studies highlight the growing consensus that coupling ML with spatial modeling can substantially elevate the sensitivity and specificity of disease surveillance.

Beyond global applications, the United States has witnessed a dramatic expansion of ML-supported spatial surveillance, particularly during the COVID-19 pandemic. Studies led by U.S. research groups have leveraged county-level mobility data, socioeconomic indicators, and spatially explicit ML models to forecast COVID-19 incidence and identify emerging hotspots (Lucas *et al*., 2022; Ramchandani *et al*., 2020; R. Wang *et al*., 2021). Geospatial surveillance of vector-borne diseases within the United States has also incorporated the power of ML and geospatial information. This is depicted in the Los Angeles West Nile virus instance, where researchers applied geospatial and temporal temperature data with the use of ML algorithms to delineate risk and detect vulnerable transmission areas within the Los Angeles region (Skaff *et al*., 2020). Complementing this, Boligarla *et al*. (2023) demonstrated that analyzing geolocated Twitter messages using ML models can effectively predict Lyme disease incidence across U.S. counties, revealing how digital behavioral patterns can supplement spatial surveillance in areas with limited clinical reporting (Boligarla *et al*., 2023). `

Yet, much work remains to be done. One of the biggest challenges is the imbalance of data infrastructure between the countryside and cities. Cities tend to have well-organized electronic health records (HER) infrastructure, enhanced capacity for tests, enhanced internet connectivity, and available auxiliary data (e.g., mobility data, data from social networks, and environmental data) (Adeleke *et al*., 2025). The countryside, however, might lag due to data inadequacies, a lower capacity of diagnostic facilities, lagging reporting time, and the unavailability of data infrastructures for digital health (Kumar *et al*., 2025). This issue becomes even more important since Zhang *et al*. (2024), argue that monitoring emerging zoonotic diseases using data-standardized surveillance and IT-enabled data systems cannot be achieved within the countryside because of the lacking fundamental resources that could aid the development of such technologies (Zhang *et al*., 2024).

Furthermore, a considerable number of surveillance systems implemented within the U.S. currently involve a relatively low level of coupling between the workflows of both SM and ML. The partitioning of data that crosses political boundaries, the regional variability of the usage of any common GIS identifier, and the absence of a standard analytic data pipeline are the factors that impede the large-scale adoption of the approach (Guralnik, 2024). Moreover, even if those cutting-edge models are successfully developed through collaborations of the academic and federal communities, it is often not easily integrated into the day-to-day operations of the local departments of public health that often experience staff and computational constraints (Iyamu *et al*., 2022). Lastly, even if it is true that considerable progress has been achieved, the coupling of GIS and machine learning within the U.S. surveillance systems remains patchy and extremely variable across the various states and counties (Svynarenko *et al*., 2025).

The purpose of this narrative review is to compile current literature findings related to the integration of spatial modeling and machine learning to discuss applications, challenges, and future opportunities related to the development of fair and data-informed surveillance systems that could ultimately strengthen national response and readiness capabilities.

Publisher: SARC Publisher

LITERATURE REVIEW

Evolution of Spatial Epidemiology and Machine Learning Integration

The field of spatial epidemiology has also been deeply transformed over the last two decades, moving forward from traditional geographical mappings of distributions and toward much more complex and data-driven models that combine geostatistics, machine learning algorithms, and multimodal digital data (Morrison *et al*., 2024; Resch *et al*., 2025). Conventional models of spatial epidemiology relied on mappings of diseases to locate geographic clusters and estimate the relationship between various environmental and/or demographic factors and the prevalence of diseases through regression models and other measures of geographical autocorrelation (Morrison *et al*., 2024). Nevertheless, the advent of modern digital data and the rapid increase in computational modeling and artificial intelligence algorithms has brought further a paradigm shift toward a predictive model that captures the complex and large dimensional behavior of diseases (Resch *et al*., 2025), representing the cornerstone of a modern surveillance system of infectious diseases globally and ever more within the USA (Kianfar *et al*., 2025).

Perhaps the strongest methodological breakthrough concerns the combination of the use of ML within models of spatial prediction. This includes the fact that a model of the transmission of Chagas disease, developed by Ledien *et al*. (2022), shows that the use of algorithms of the ML type led to a considerable improvement over linear models of classical estimation of the force of infection (Ledien *et al*., 2022). At the same time, the authors also noted the importance of models of the ensemble type, such as random forests and gradient boosting machines, that are used to extract the complex nonlinear dependencies between the vulnerability of transmission and the ecological predictors of the transmission risk during accurate spatiotemporal predictions (Ledien *et al*., 2022). This combination also clearly represented a breakthrough and a shift within the framework of the methodology of the spatial studies of epidemiology of the past, which typically used fixed-effect regression models that are ineffective for the preimage of complex dependencies both in time and space (Ledien *et al*., 2022). Likewise, Piovezan *et al*. (2022), combining the two approaches of the spatial model and the ML model, successfully predicted the distribution of eggs of the *Aedes aegypti* mosquito in Brazil and showed the positive effect of the additional precision within the vector surveillance through the combination of the metrics of the environment, the climatic conditions, and the previously used places of oviposition of the vector (Piovezan *et al*., 2022).

Looking beyond vector surveillance, the field of digital epidemiology has enabled the use of a broader range of data to detect the geographical pattern of diseases using novel data streams and the applications of ML algorithms. Zhang *et al*. (2021) undertook a systematic review, and the authors concluded that the inclusion of internet search information and climatic information increases the predictability of diseases that are sensitive to climatic factors and has a higher geographical resolution than the traditional data (Zhang *et al*., 2021).

Spatial modeling has also been influenced by the advancements of deep learning. Wang *et al*. (2021) introduced a deep model involving a multi-granularity spatiotemporal neural network that greatly enhanced the accuracy of influenza surveillance models through the concurrent understanding of local characteristics, regional transmission, and national transitions throughout the United States (Wang *et al*., 2021). This model directly answered a long-existing concern of spatial epidemiology related to the incorporation of the complexity of spatiotemporal dependencies characterized by administrative, demographic, and environmental factors into models. The direct incorporation of geographic characteristics into a neural network model opened the way to the development of novel models of the future.

Further within the field of public health, the importance of the adoption of AI technologies to improve surveillance has been highlighted to involve the use of frameworks and technologies such as ML and Natural Language Processing (NLP), which can identify and categorize important information related to health within online platforms and databases (Olawade *et al*., 2023). The authors highlighted the need to utilize AI-enabled surveillance to address the complexity of modern data and improve the techniques of both spatial epidemiology and predictive analytics.

During the COVID-19 pandemic, the trend of enhancing the integration of ML and spatial modeling accelerated across the world. Budd *et al*. (2020), revealed that the application of digital technologies such as spatial mobility data, contact

Publisher: SARC Publisher

tracing tools, and geolocation-enabled notifications acted as key factors for COVID-19 management strategies and greatly enhanced real-time surveillance (Budd *et al*., 2020). At the same time, Asadzadeh *et al*. (2020) also found that the combination of digital healthcare technologies and spatial analytic tools enabled the optimization of emergency management operations during the COVID-19 pandemic (Asadzadeh *et al*., 2020).

Wastewater-based epidemiology (WBE) has also emerged as a significant addition to the field of spatial-ML integration. Taking a recent review of the literature into account, the involvement of enhanced models of wastewater using ML for COVID-19 tracking has been identified to improve accuracy of hotspots and forecasting using geospatial mapping and various algorithms of ML (Torabi *et al*., 2023). Further contributing to the related field, the involvement of WBE networks integrated with GIS has been acknowledged to enable a more accurate tracking of SARS-CoV-2 trends (Cuadros *et al*., 2024).

As the development of both ML and spatial models progressed around the world, U.S. scientists sought to combine these technologies into a national surveillance program. This could be observed in the Los Angeles West Nile Virus (WNV) data project's creation of a spatiotemporal ML model using temperature-dependent intensity levels that correlated well with a wide range of climatically trans critical zones and locations of the WNV transmission clusters across the broader Los Angeles area (Skaff *et al*., 2020). This could be further observed using geolocation and geosocial data mining and Natural Language Generation (NLG) algorithms leading to the development of a distributed geolocation model of Lyme's U.S. incidence using over a million geolocated Twitter posts, further highlighting the capacity of geolocation-ML algorithms and NLG models to be used within national surveillance frameworks (Boligarla *et al*., 2023). Early models of COVID-19 forecasting using U.S. county-level data and related ML algorithms also found that the coupling of spatial lag variables contributed to enhanced accuracy and reduced volatility of quantitative models of U.S. COVID-19 epidemiologic trends (Vahedi *et al*., 2021).

Lastly, the ever-growing risk of zoonotic diseases has also led to the adoption of innovative approaches for the integration of GIS and ML. Zhang *et al*. (2024) recently demonstrated that the use of GIS, AI, and Remote Sensing (RS) technologies enhances zoonotic surveillance systems through early detection, determination of geographical risk areas, and predictive modeling of transmission paths between species (Zhang *et al*., 2024).

Taken together, these international and U.S. developments demonstrate a clear trend: the field of spatial epidemiology is moving from a geographic, descriptive approach towards a predictive model using AI and powered by a variety of data streams and scalable algorithms. This represents a starting point for a methodology and framework that could understand the implications of the combination of these new surveillance technologies and the needs of both the urban and rural contexts.

Applications of Machine Learning and Spatial Modeling Integration in Urban vs. Rural Infectious Disease Surveillance

Machine learning (ML) and Spatial Modeling methods have become increasingly indispensable for infectious disease surveillance (Villanueva-Miranda *et al*., 2025), yet their performance, implementation, and impact differ widely between urban and rural settings. The United States presents a critical case study of this divide. Urban centers benefit from dense healthcare systems, higher reporting frequency, richer electronic data sources, and stronger digital infrastructure, enabling more rigorous Machine Learning-Geographic Information System (ML-GIS) integration (Ghildayal *et al*., 2024). Conversely, rural regions often face sparse case counts, limited digital connectivity, and underdeveloped surveillance capacity, making the adoption and accuracy of advanced analytic frameworks more challenging (Brown & Davis, 2025). Understanding these divergent contexts is essential for designing equitable surveillance systems capable of early detection, precision mapping, and real-time response.

Urban Settings: High-Resolution Data and Strong Model Performance

Urban areas provide ideal conditions for ML-enabled spatial surveillance due to their extensive Electronic Health Record (EHR) systems, detailed geographic identifiers, high population densities, and diverse digital data streams (Ghildayal *et al*., 2024; Mollalo *et al*., 2024). Urban settings also generate large amounts of behavioral data in digital forms, for example, tweets, mobility patterns, and search activities, which can be used in ML-related surveillance systems (Boligarla *et*

Publisher: SARC Publisher

al., 2023). Spatial analysis of more than 1.3 million geolocated tweets revealed strong alignment between digital Lyme disease discussions and county-level incidence patterns, highlighting the predictive value of social media data in densely populated environments where digital participation is high (Boligarla *et al*., 2023). The high level of variation in the mentioned datasets significantly improves the forecasting accuracy, often complementing and sometimes outperforming conventional surveillance metrics.

Digital health infrastructure enhances the integration of ML and Spatial Modeling in an urban setting. During the COVID-19 crisis, cities possessing extensive testing infrastructure, data interoperability, and high-resolution reporting, like NYC, were able to provide enough detail on cases to enable accurate real-time forecasting through machine learning, as evidenced by predictive modeling in NYC by (Zhang *et al*., 2025). Additionally, they possess more developed WebGIS and Internet of Things (IoT) surveillance systems. The monitoring of vector-borne diseases in real-time by WebGIS requires an extensive network of environmental sensors sensitive to micro-variability, enabling the continuous flow of geospatial data, and automatic alert generation, which is impractical without considerable resource investment, as is the case in urban settings only (Javaid *et al*., 2023).

Wastewater-based epidemiology (WBE), one of the most impactful surveillance innovations of the COVID-19 era, has expanded more effectively in metropolitan counties, with citywide severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA levels strongly tracking clinical and syndromic surveillance indicators across multiple waves (Hopkins *et al*., 2023). Spatiotemporal ML models applied to urban wastewater SARS-CoV-2 signals have achieved high forecast precision due to frequent sampling, stable flow dynamics, and consistent population inputs across urban treatment facilities (Torabi *et al*., 2023). These characteristics provide a level of modeling stability that is difficult to achieve in smaller or decentralized wastewater systems. Collectively, evidence from these domains indicates that ML-based spatial modeling approaches achieve their highest performance where data are dense, structured, and continuously updated, conditions that characterize urban environments.

Rural Settings: Sparse Data, Uncertain Predictions, and Limited Infrastructure

While urban areas benefit from data abundance, rural areas are hampered by systemic barriers issues hindering the efficiency of the implementation process of ML-GIS tools (Sosin & Carpenter-Song, 2024). In many non-urban counties in the US, there is less access to healthcare facilities and less adoption of EHR systems, hence less reporting, and less spatial detail, all of which hamper the predictive capability of machine learning models (Ekren *et al*., 2025). Inequities in early outbreak detection and response are contributed to by issues of limited infrastructure, especially evident in evidence suggesting the lack of capability in some non-urban areas to provide efficient AI-driven surveillance systems for zoonotic outbreaks (Zhang *et al*., 2024).

Wastewater-based epidemiology (WBE) faces similar challenges, as many rural communities depend on septic systems rather than centralized treatment facilities. Sparse and inconsistent sampling makes spatiotemporal ML modeling difficult to implement, creating sharp contrasts with metropolitan counties that maintain stable, high-frequency wastewater surveillance grids (Torabi *et al*., 2023).

Digital behavioral data, which are a critical enabling resource for ML monitoring, also show large urban-rural disparities. The urban areas show significantly more social behavior on social media platforms and geocoded digital traces, whereas there may be very fewer instances of signals required for successful Natural Language Processing (NLP) inference or modeling studies, as evident from disparities in Lyme disease indicators extracted from tweets (Boligarla *et al*., 2023).

Finally, even if rural data were available, geographic heterogeneity may still negatively affect the accuracy of machine learning models (Ledien *et al*., 2022). The presence of non-stationarity in geographic spaces, that is, variations in patterns and relationships under different environments, may hamper the applicability of ML-SM models developed for urban settings, and considerable variation may exist in the force of infection estimated in various non-urban environments (Ledien *et al*., 2022; Yin *et al*., 2024).

Publisher: SARC Publisher

Methodological and Ethical Considerations in Integrating Spatial Modeling and Machine Learning

The integration of spatial modeling and machine learning (ML) into infectious disease surveillance creates powerful new capabilities but also introduces a layered set of methodological, ethical, and regulatory challenges. These challenges matter not only for model accuracy and operational performance, but also for health equity, public trust, and the ethical governance of AI-enabled public health systems in the United States and beyond.

Methodological Integration of Spatial Modeling and Machine Learning

Recent empirical studies show that many ML or regression-based models for infectious disease fail to fully leverage spatial dependence, limiting their ability to capture localized transmission dynamics and cross-regional interactions (Bao *et al*., 2025; Jiao *et al*., 2021; B. Lucas *et al*., 2023). Similarly, Santangelo *et al*. (2023) emphasized that, despite the superior accuracy of certain types of ML models compared to regression models, spatial non-stationarity and local environments are often neglected by most models (Santangelo *et al*., 2023).

There is now an emerging set of research and applications combining spatial visualization, geostatistical interpolation, and classification or regression using machine learning approaches. Recent progress indicates that through integrating spatial clustering, kernel density maps, and machine learning approaches for risk modeling, it is now possible to accurately forecast patterns of outbreaks under fluctuating demographics and environments (Cheng *et al*., 2024). Findings from early warning systems based on AI are also indicative of the model's superiority by integrating spatiotemporal variables, including mobility, exposure, and incidence, as against models based on temporal patterns (Villanueva-Miranda *et al*., 2025).

These methodological considerations are especially salient for urban–rural applications. ML-based risk prediction frameworks developed using dense, high-resolution urban data often fail to generalize when healthcare access, reporting intensity, and spatial covariates differ across rural settings, underscoring the need for spatially aware domain adaptation and locally calibrated models (Liu *et al*., 2025). Comparable challenges are reported in regions where spatial resolution is coarse and surveillance systems remain uneven or fragmented, making stable prediction difficult to achieve (Qasrawi *et al*., 2025).

Similar observations are found for other safety-critical domains, where applications of AI are emerging. In mining applications, only those hazard identification systems, combined with AI and machine learning, where spatial and temporal relationships with mining equipment locations, tunnel shape, and other mining-environment-related variables are strongly entwined, prove to be reliable, thereby discrediting models that remain spatially ignorant (Arthur *et al*., 2025). Similar observations from other computer vision applications clearly establish the use of spatial map inputs and other spatial variables serving as essential components, leading to significant improvements in automatic hazard detection within spatial environments (Oluwagbade, 2025).

Data Quality, Heterogeneity, and Uncertainty

Large-scale ML-based spatial modeling approaches are highly dependent on the quality, completeness, and representativeness of the data they use. Noise in reporting, differences in case definitions, as well as variability in testing rates, may generate systematic errors in the prediction outputs of the models, which can worsen in rural as well as resource-constrained settings (Al Meslamani *et al*., 2024). The same issues also exist in surveillance systems having patchy geographic representation, where there could be delays in the time of reporting, as well as a lack of digitalization, where even the most advanced algorithms fail to provide reliable outputs in terms of the inherent weaknesses in the data structure (Adewumi, 2025).

The presence of wastewater, mobility, and environmental data introduces another level of heterogeneity. In wastewater prediction models, spatiotemporal variation needs to address the variability in catchment boundaries, rates of flow, as well as the size of the population, so as not to mistake the presence of engineered systems as actual signals in the epidemiological process (Torabi *et al*., 2023). Considering uncertainties that can pervade the entire ML process, the use of probabilistic forecasts has gained traction as a way to reduce overconfidence in outbreak prediction tasks (Santangelo *et al*., 2023).

In the wake of heightened methodology standards, literature has shifted toward standardized analytical pipelines. New syntheses describe

Publisher: SARC Publisher

structured research pipelines that cover topics like data harmonization, spatial feature engineering, bias audits, as well as other notions like external model validations over varied geographies to enhance generalizability (Wang *et al*., 2025).

Interpretability and Explainability in Spatial–ML Models

Increasing complexity of ML models, particularly deep learning models where spatial structure is implicitly represented, raises important interpretability challenges (Luo *et al*., 2023). From the perspective of medical AI in general, opaque models were demonstrated to mask sources of bias, analyze error, and make accountability challenges, giving rise to a significant demand for the systematic use of model explanation techniques and clear documentation of the process, as described in the work of (Hanna *et al*., 2025). In the context of spatial disease surveillance, the issue essentially corresponds to the demand for risk maps and feature attribution algorithms that are interpretable.

Model interpretability techniques like SHAP values, partial dependency plots, and spatial saliency maps can also shed light on the role of geography variables in outbreak prediction, thus improving practitioner confidence in the automatic alerts produced by machine learning models (Cheah *et al*., 2025). The role of interpretability in computational pathology also addresses the importance of transparency in developing unbiased AI models, as domain experts must understand the workings of the AI models to properly audit the outputs produced (Montezuma *et al*., 2025).

Privacy and Governance of Geolocated Health Data

The presence of fine geolocation information, which forms the basis of spatial modeling, also represents a challenge in terms of privacy as well as governance. Ethical reviews concerning AI in the healthcare sector argue that the use of geolocation information in relation to patient data, patient movement information, as well as genomic information could potentially make it easier to work towards the re-identification of individuals, particularly when information is shared across institutional settings, as proposed by (Tilala *et al*., 2024). To mitigate these risks, public health systems require strong safeguards, data minimization, rigorous de-identification, and governance mechanisms designed to prevent exacerbation of inequities or erosion of public trust (Dankwa-Mullan, 2024).

Gaps in regulations add to the complexity of the scenario. AI research has identified challenges in AI compliance frameworks in the U.S. healthcare industry in relation to accountability, security, as well as regulation of AI systems using large amounts of health information in continuously updated databases (Parker *et al*., 2024). General models for AI ethics state the importance of considering transparency, fairness, as well as privacy in AI systems in deciding the use of spatial data to direct activities in the healthcare sector (Radanliev, 2025).

Future Directions and Research Gaps

Despite the tremendous advancements made in the process of integrating machine learning algorithms in spatial surveillance of infectious diseases, there are still several research as well as systems-level gaps that remain in the way of the widespread adoption of this technology in the United States of America. This research as well as systems-level gaps cover various levels, ranging from technology to methodology, along with ethics.

Technical Gaps: Toward Standardized Spatial-ML Pipelines

One of the key technical challenges associated with integrating machine learning with spatial modeling and GIS is the lack of standardized spatial-ML pipelines. This means that the current models are based on ad-hoc workflows designed by research groups, government entities, or other partners. This has resulted in a lack of replication, as well as variability in model accuracy for geographically diverse areas. Moreover, the surveillance modeling in terms of spatial feature engineering, geocoding, uncertainty estimation, and model verification has been diverse, ranging from state to state, which has made it quite challenging to adapt the models at the national level (Cheah *et al*., 2025; Santangelo *et al*., 2023).

Further technical work will be necessary to create integrated toolkits for spatiotemporal preprocessing, machine learning feature extraction, as well as region-to-region comparison. Recently, the idea has been put forward that spatiotemporal deep learning models might form the basis for rigorously repeatable national models, but this could only be achieved if integrated geospatial as well as temporal metadata are applied (Wang *et al*., 2021).

Moreover, the intake of high frequency data, like the levels of viruses in wastewater, satellite-measured climate variables, entomological trap

Publisher: SARC Publisher

captures, or logged population mobility, for the purpose of integrating it into the model also has limitations in most U.S. departments of health (Dórea & Revie, 2021).

Data Heterogeneity and Underrepresentation of Rural Communities

This disparity creates challenges for ML systems, which rely heavily on rich, continuous data streams. When trained predominantly on urban data, models often fail to generalize exacerbating prediction errors in rural regions and amplifying inequities in early detection. Li *et al*. (2021) found that COVID-19 forecasting models trained on urban-dominant datasets systematically misestimated rural trajectories due to lower case densities and inconsistent reporting. (Li *et al*., 2021)

Addressing these gaps requires research into rural-adaptive ML approaches. These may include spatial transfer learning, synthetic minority oversampling for sparse geographies, small-area Bayesian modeling, and hybrid ML-geostatistical frameworks that compensate for patchy reporting.

Opportunities in Hybrid Modeling and Multimodal Integration

Future innovation lies in hybrid modeling efforts that merge classical epidemiological frameworks, spatial statistics, deep learning, and real-time data feeds. Future research should explore hybrid pipelines that incorporate geostatistical smoothing with ML-based risk classification, graph neural networks for county-level transmission mapping, physics-informed ML models that encode epidemiological constraints, integration of genomic epidemiology with spatial ML, and continuous wastewater signal prediction with deep spatiotemporal models

These multimodal systems will help close the gap between the interpretability of mechanistic models and the prediction capabilities offered by machine learning algorithms.

Real-Time Predictive Systems and Streaming Data Analytics

A key frontier in spatial-ML surveillance is the deployment of real-time predictive systems capable of processing streaming data. The importance of live prediction models, live prediction, and live hotspot detection was well demonstrated in relation to the COVID-19 pandemic, in which the pandemic has brought attention to the use of live prediction models, but currently, most mathematical models in ML-GIS systems remain updated in a batching manner, as opposed to live feeds. WebGIS systems in surveillance systems currently provide the basis for live prediction, but this can only be made effective using models that are updated dynamically as more data becomes available, as has been noted in the most recent surveillance systems' evaluation (Javaid *et al*., 2023). Future research must also address concept drift, as the relationship between diseases and the environment can change over time relationships over time and that may degrade model performance unless automatically detected and corrected.

CONCLUSION

The blending of spatial analytics, machine learning, and other innovative technologies is significantly altering the outbreak surveillance landscape in the United States. This review shows that these technologies promote improvements in outbreak prediction, shed greater insight into the concept of spatial heterogeneity, as well as the analytic potential offered by these advances in the United States. Advancing towards a more balanced and effective surveillance infrastructure will require standardized data, greater support in non-urban areas, and efforts in hybrid, explainable modeling. Incorporating ethical, privacy-enhancing, and fairness constraints will also be necessary, as systems leverage more sensitive geolocation and behavior data. Through the synergistic use of technological innovation, effective governance, and resource allocation, the spatial-ML surveillance approach can help realize the vision of improved detection, greater precision, and stronger national health security.

REFERENCES

1. Adeleke, O. J., Jovanovich, K., Ogunbunmi, S., Samuel, O., & Kehinde, T. O. "Comprehensive Exploration of Smart Cities: A Systematic Review of Benefits, Challenges, and Future Directions in Telecommunications and Urban Development." IEEE Sensors Reviews (2025).
2. Adewumi, I. P. "Critical analysis of infectious disease surveillance and response system in Nigeria." Discover Public Health 22.1 (2025): 272.
3. Al Meslamani, A. Z., Sobrino, I., & de la Fuente, J. "Machine learning in infectious diseases: potential applications and limitations." Annals of Medicine 56.1 (2024): 2362869.

Publisher: SARC Publisher

4. Arthur, A. A., Annankra, J. A., & Yakin, Z. "Examining the role of AI and machine learning in improving hazard detection and predictive analytics for accident prevention in mining operations." World Journal of Advanced Engineering Technology and Sciences 15.3 (2025): 640-646.
5. Asadzadeh, A., Pakkhoo, S., Saeidabad, M. M., Khezri, H., & Ferdousi, R. "Information technology in emergency management of COVID-19 outbreak." Informatics in medicine unlocked 21 (2020): 100475.
6. Bao, Y., Huang, I., Li, Q., Zhang, Z., Xing, Y., Hou, D., & Ye, J. "A framework for modeling county-level COVID-19 transmission." Frontiers in Public Health 13 (2025): 1608360.
7. Boligarla, S., Laison, E. K. E., Li, J., Mahadevan, R., Ng, A., Lin, Y., ... & Nasri, B. "Leveraging machine learning approaches for predicting potential Lyme disease cases and incidence rates in the United States using Twitter." BMC medical informatics and decision making 23.1 (2023): 217.
8. Brown, K. E., & Davis, S. E. "Gaps in artificial intelligence research for rural health in the United States: a scoping review." Journal of the American Medical Informatics Association (2025): ocaf206.
9. Budd, J., Miller, B. S., Manning, E. M., Lampos, V., Zhuang, M., Edelstein, M., ... & McKendry, R. A. "Digital technologies in the public-health response to COVID-19." Nature medicine 26.8 (2020): 1183-1192.
10. Chakilam, C. "Integrating Machine Learning and Big Data Analytics to Transform Patient Outcomes in Chronic Disease Management." Journal of Survey in Fisheries Sciences 9.3 (2022): 118-130.
11. Cheah, B. C., Vicente, C. R., & Chan, K. R. (2025). Machine Learning and Artificial Intelligence for Infectious Disease Surveillance, Diagnosis, and Prognosis. *Viruses*, *17*(7), 882.
12. Cheng, Y., Bai, Y., Yang, J., Tan, X., Xu, T., & Cheng, R. "Analysis and prediction of infectious diseases based on spatial visualization and machine learning." Scientific Reports 14.1 (2024): 28659.
13. Cuadros, D. F., Chen, X., Li, J., Omori, R., & Musuka, G. "Advancing public health surveillance: Integrating modeling and GIS in the wastewater-based epidemiology of viruses, a narrative review." Pathogens 13.8 (2024): 685.
14. Dankwa-Mullan, I. "Health equity and ethical considerations in using artificial intelligence in public health and medicine." Preventing chronic disease 21 (2024): E64.
15. Dórea, F. C., & Revie, C. W. (2021). Data-driven surveillance: Effective collection, integration, and interpretation of data to support decision making. *Frontiers in Veterinary Science*, *8*, 633977.
16. Ekren, E., Maleki, S., Curran, C., Watkins, C., & Villagran, M. M. "Health differences between rural and non-rural Texas counties based on 2023 County Health Rankings." BMC health services research 25.1 (2025): 2.
17. Ekundayo, F. "Using machine learning to predict disease outbreaks and enhance public health surveillance." World J Adv Res Rev 24.3 (2024): 794-811.
18. Ghildayal, N., Nagavedu, K., Wiltz, J. L., Back, S., Boehmer, T. K., Draper, C., ... & Block, J. P. "Public health surveillance in electronic health records: lessons from PCORnet." Preventing Chronic Disease 21 (2024): E51.
19. Guralnik, E. "US public health surveillance, reimagined." *Learning Health Systems*, 8.4 (2024): e10445.
20. Hanna, M. G., Pantanowitz, L., Jackson, B., Palmer, O., Visweswaran, S., Pantanowitz, J., ... & Rashidi, H. H. "Ethical and bias considerations in artificial intelligence/machine learning." Modern Pathology 38.3 (2025): 100686.
21. Hopkins, L., Persse, D., Caton, K., Ensor, K., Schneider, R., McCall, C., & Stadler, L. B. "Citywide wastewater SARS-CoV-2 levels strongly correlated with multiple disease surveillance indicators and outcomes over three COVID-19 waves." Science of The Total Environment 855 (2023): 158967.
22. Iyamu, I., Gómez-Ramírez, O., Xu, A. X., Chang, H. J., Watt, S., Mckee, G., & Gilbert, M. "Challenges in the development of digital public health interventions and mapped solutions: findings from a scoping review." Digital health 8 (2022): 20552076221102255.
23. Javaid, M., Sarfraz, M. S., Aftab, M. U., Zaman, Q. U., Rauf, H. T., & Alnowibet, K. A. "WebGIS-based real-time surveillance and response system for vector-borne infectious diseases." International Journal of Environmental Research and Public Health 20.4 (2023): 3740.
24. Jiao, J., Chen, Y., & Azimian, A. "Exploring temporal varying demographic and economic

Publisher: SARC Publisher

disparities in COVID-19 infections in four US areas: Based on OLS, GWR, and random forest models." Computational urban science 1.1 (2021): 27.
25. Khorshid, M. A., Aly, G. E., Elabd, M. A., Kotb, M. A. K., El-Nokrashy, A. F., & Saad, N. G. E. D. "Real-Time Violence Detection Across Domains: Evaluating Model Robustness in Varied Surveillance Contexts." 2025 Intelligent Methods, Systems, and Applications(IMSA). IEEE, (2025).
26. Kianfar, N., Sartorius, B., Lau, C. L., Bergquist, R., & Kiani, B. "The future of spatial epidemiology in the AI era: enhancing machine learning approaches with explicit spatial structure." Geospatial Health 20.1386 (2025): 1386.
27. Kiani, B., Lau, C., & Bergquist, R. "From Snow's map of cholera transmission to dynamic catchment boundary delineation: current front lines in spatial analysis." Geospatial Health 18.2 (2023).
28. Kopczewska, K. "Spatial machine learning: new opportunities for regional science." The Annals of Regional Science 68.3 (2022): 713-755.
29. Kumar, P., Lionis, C., Andoko, D., Rahman, Z., Anastasaki, M., & Awankem, B. "Evaluation of Diagnostic Services in Rural and Remote Areas: Bottlenecks, Success Stories, and Solutions." Journal of Surgical Specialties and Rural Practice 6.1 (2025): 32-37.
30. Ledien, J., Cucunubá, Z. M., Parra-Henao, G., Rodríguez-Monguí, E., Dobson, A. P., Adamo, S. B., ... & Nouvellet, P. "Linear and Machine Learning modelling for spatiotemporal disease predictions: Force-of-Infection of Chagas disease." PLoS neglected tropical diseases 16.7 (2022): e0010594.
31. Lee, A. C., Iversen, B. G., Lynes, S., Desenclos, J. C., Bezuidenhoudt, J. E., Flodgren, G. M., & Pyone, T. "The state of integrated disease surveillance globally: synthesis report of a mixed methods study." Public Health 228 (2024): 85-91.
32. Liu, M., Liu, Y., & Liu, J. "Machine learning for infectious disease risk prediction: a survey." ACM Computing Surveys 57.8 (2025): 1-39.
33. Lucas, B., Vahedi, B., & Karimzadeh, M. "A spatiotemporal machine learning approach to forecasting COVID-19 incidence at the county level in the USA." International Journal of Data Science and Analytics 15.3 (2023): 247-266.
34. Lucas, T. C., Nandi, A. K., Keddie, S. H., Chestnutt, E. G., Howes, R. E., Rumisha, S. F., ... & Weiss, D. J. "Improving disaggregation models of malaria incidence by ensembling non-linear models of prevalence." Spatial and spatio-temporal epidemiology 41 (2022): 100357.
35. Luo, L., Li, B., Wang, X., Cui, L., & Liu, G. "Interpretable spatial identity neural network-based epidemic prediction." Scientific Reports 13.1 (2023): 18159.
36. Mollalo, A., Hamidi, B., Lenert, L. A., & Alekseyenko, A. V. "Application of Spatial analysis on electronic health records to characterize patient phenotypes: systematic review." JMIR Medical Informatics 12.1 (2024): e56343.
37. Montezuma, D., Porz, R., Ameisen, D., L'Imperio, V., Serbanescu, M. S., Temprana-Salvador, J., ... & of Digital, E. S. "Unbiased AI: Addressing Bias in Computational Pathology." Mayo Clinic Proceedings: Digital Health (2025): 100302.
38. Morrison, C. N., Mair, C. F., Bates, L., Duncan, D. T., Branas, C. C., Bushover, B. R., ... & Rundle, A. G. "Defining spatial epidemiology: a systematic review and re-orientation." Epidemiology 35.4 (2024): 542-555.
39. Olawade, D. B., Wada, O. J., David-Olawade, A. C., Kunonga, E., Abaire, O., & Ling, J. "Using artificial intelligence to improve public health: a narrative review." Frontiers in Public Health 11 (2023): 1196397.
40. Oluwagbade, E. "Computer Vision Techniques for Automated Detection of Hazardous Events in Open-Pit and Underground Mines." (2025).
41. Parker, V. J., Economou, N. J., & Silcox, C. "AI governance in health systems: aligning innovation, accountability, and trust." Preprint at https://healthpolicy. duke. edu/sites/default/files/2024-10/AI% 20Governance% 20in% 20Health% 20Systems. pdf (2024).
42. Piovezan, R., de Azevedo, T. S., Faria, E., Veroneze, R., Von Zuben, C. J., Von Zuben, F. J., & Sallum, M. A. M. "Assessing the effect of Aedes (Stegomyia) aegypti (Linnaeus, 1762) control based on machine learning for predicting the spatiotemporal distribution of eggs in ovitraps." Dialogues in health 1 (2022): 100003.

Publisher: SARC Publisher

43. Qasrawi, R., Issa, G., Thwib, S., AbuGhoush, R., Amro, M., Ayyad, R., ... & Kong, J. D. "The Role of Machine Learning in Infectious Disease Early Detection and Prediction in the MENA Region: A Systematic Review." Informatics in Medicine Unlocked (2025): 101651.
44. Radanliev, P. "AI ethics: Integrating transparency, fairness, and privacy in AI development." Applied Artificial Intelligence 39.1 (2025): 2463722.
45. Ramchandani, A., Fan, C., & Mostafavi, A. "Deepcovidnet: An interpretable deep learning model for predictive surveillance of covid-19 using heterogeneous features and their interactions." Ieee Access 8 (2020): 159915-159930.
46. Resch, B., Kolokoussis, P., Hanny, D., Brovelli, M. A., & Kamel Boulos, M. N. "The generative revolution: AI foundation models in geospatial health—applications, challenges and future research." International Journal of Health Geographics 24.1 (2025): 6.
47. Santangelo, O. E., Gentile, V., Pizzo, S., Giordano, D., & Cedrone, F. "Machine learning and prediction of infectious diseases: a systematic review." Machine Learning and Knowledge Extraction 5.1 (2023): 175-198.
48. Shaikh, T. G., Waseem, S., Ahmed, S. H., Swed, S., & Hasan, M. M. "Infectious disease surveillance system in Pakistan: challenges and way forward." Tropical Medicine and Health 50.1 (2022): 46.
49. Skaff, N. K., Cheng, Q., Clemesha, R. E., Collender, P. A., Gershunov, A., Head, J. R., ... & Remais, J. V. "Thermal thresholds heighten sensitivity of West Nile virus transmission to changing temperatures in coastal California." Proceedings of the Royal Society B 287.1932 (2020): 20201065.
50. Sosin, A. N., & Carpenter-Song, E. A. "Reimagining Rural Health Equity: Understanding Disparities And Orienting Policy, Practice, And Research In Rural America: Article examines rural health equity." Health Affairs 43.6 (2024): 791-797.
51. Svynarenko, R., Kim, H., Stansberry, T., Oh, C., Sarkar, A., & Lindley, L. C. (2025). *Analysis of Geospatial Variations in Healthcare Across Rural Communities in the US Using Machine Learning*. *13*(13), 1504.
52. Tilala, M. H., Chenchala, P. K., Choppadandi, A., Kaur, J., Naguri, S., Saoji, R., ... & Tilala, M. "Ethical considerations in the use of artificial intelligence and machine learning in health care: a comprehensive review." Cureus 16.6 (2024).
53. Torabi, F., Li, G., Mole, C., Nicholson, G., Rowlingson, B., Smith, C. R., ... & Blangiardo, M. "Wastewater-based surveillance models for COVID-19: A focused review on spatio-temporal models." Heliyon 9.11 (2023).
54. Vahedi, B., Karimzadeh, M., & Zoraghein, H. "Spatiotemporal prediction of COVID-19 cases using inter-and intra-county proxies of human interactions." Nature communications 12.1 (2021): 6440.
55. Villanueva-Miranda, I., Xiao, G., & Xie, Y. "Artificial Intelligence in Early Warning Systems for Infectious Disease Surveillance: A Systematic Review." Frontiers in Public Health 13 (2025): 1609615.
56. Villanueva-Miranda, I., Xiao, G., & Xie, Y. "Artificial Intelligence in Early Warning Systems for Infectious Disease Surveillance: A Systematic Review." Frontiers in Public Health 13 (2025): 1609615.
57. Wang, R., Wu, H., Wu, Y., Zheng, J., & Li, Y. "Improving influenza surveillance based on multi-granularity deep spatiotemporal neural network." Computers in Biology and Medicine 134 (2021): 104482.
58. Wang, Y., Zhong, J., & Kumar, R. "A systematic review of machine learning applications in infectious disease prediction, diagnosis, and outbreak forecasting." (2025).
59. Yarmohammadtoosky, S., & Xie, Y. "Data-Driven Analysis of Climate Change Impacts on Public Health: A Systematic Review and Future Directions." 2024 IEEE International Conference on Big Data (BigData). IEEE, (2024).
60. Yin, Z., Ding, J., Liu, Y., Wang, R., Wang, Y., Chen, Y., ... & Du, Z. "GNNWR: an open-source package of spatiotemporal intelligent regression methods for modeling spatial and temporal non-stationarity." Geoscientific Model Development Discussions 2024 (2024): 1-22.
61. Zhang, L., Guo, W., & Lv, C. "Modern technologies and solutions to enhance surveillance and response systems for emerging zoonotic diseases." Science in One Health 3 (2024): 100061.
62. Zhang, R., Tai, J., Yao, Q., Yang, W., Ruggeri, K., Shaman, J., & Pei, S. "Behavior-driven forecasts of neighborhood-level COVID-19 spread in New York City." PLOS

Publisher: SARC Publisher

Computational Biology 21.4 (2025): e1012979.

63. Zhang, Y., Bambrick, H., Mengersen, K., Tong, S., & Hu, W. "Using internet-based query and climate data to predict climate-sensitive infectious disease risks: a systematic review of epidemiological evidence." International Journal of Biometeorology 65.12 (2021): 2203-2214.

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:
Ssemujju, F. S. & Solomon, D. “Integrating Spatial Modeling and Machine Learning for Infectious Disease Surveillance in U.S. Urban and Rural Settings." *Sarcouncil Journal of Internal Medicine and Public Health* 5.1 (2025): pp 1-12.

Publisher: SARC Publisher