RESEARCH INVENTION JOURNAL OF SCIENTIFIC AND EXPERIMENTAL SCIENCES 6(1):67-74, 2026

©RIJSES Publications ONLINE ISSN: 1115-618X

PRINT ISSN: 1597-2917

https://doi.org/10.59298/RIJSES/2026/616774

Integrating Spatial Omics with Environmental Exposure Data for Asthma Risk Prediction: Interpretability, Bias, Real-World Performance, Implementation, and Equity

Waiswa Arajab

Department of Pharmacy Kampala International University Uganda
Email: arajab.waiswa@studwc.kiu.ac.ug

ABSTRACT

Asthma is a complex, chronic respiratory disease influenced by both genetic and environmental factors, and it carries significant public health and economic burdens. Emerging spatial omics technologies, including spatial transcriptomics, proteomics, and metabolomics, enable high-resolution molecular characterization of tissues. At the same time, environmental exposure datasets capture temporally and spatially resolved risk factors such as air pollution, urban vegetation, and land-use patterns. Integrating these heterogeneous datasets can improve predictive models for asthma risk, enhance the interpretability of biological and environmental interactions, and inform precision public health interventions. Challenges remain in model interpretability, bias, equity, real-world validation, and implementation, particularly in ensuring fairness across diverse populations and maintaining data privacy. Approaches to data fusion, bias detection, and stakeholder engagement are critical to facilitate ethical and effective deployment. This review highlights current methodologies, practical considerations, and prospective deployment scenarios for integrating spatial omics with environmental exposure data to advance asthma risk prediction, with a focus on equity, reproducibility, and translational impact.

Keywords: Asthma, Spatial Omics, Environmental Exposure, Predictive Modeling, **and** Health Equity.

INTRODUCTION

Asthma is a chronic respiratory condition characterized by airway inflammation, airway hyperresponsiveness, and reversible airway obstruction [1]. The impact of asthma on public health is severe and costly. In the United States, the prevalence of asthma in adults and children is currently estimated to be 7.7% and 8.1%, respectively, and the total medical cost of asthma-related treatment was reported to be approximately $81 billion for 2013; given the increasing trend of asthma prevalence, asthma-related costs are projected to increase to $104 billion by 2025 [2]. Health disparities exist, and asthma is significantly more prevalent among certain demographic groups compared to others [1]. According to [2] (2018, the 5-year prevalence of asthma based on data from the National Health Interview Survey (NHIS) was higher among Puerto Ricans (12.7%), non-Hispanic Black (12.8%), and multiracial children (9.4%) in comparison to non-Hispanic White children (6.9%) across demographic groups; the 5-year prevalence of asthma based on the same data was also higher among people aged 18~34 (9.7%) and people aged 35~44 (9.1%) compared to people aged 65 and above (4.5%) across age groups [3]. Asthma has serious negative health consequences, with asthma exacerbations leading to emergency room visits, hospitalizations, lost school/work days, and even death. Environmental exposure, ranging from a person's general metabolism/biomarkers to long-term exposure to certain compounds, is an important risk factor influencing the development of asthma. While there are known environmental factors (pre-, peri-, and post-natal exposures) that affect personal exposures on a population level, quantifying and mapping environmental exposures such as pollution in a precise manner is widely claimed to be crucial for evaluating health impacts across various domains, including, but not limited to, asthma [1]. Spatial transcriptomics, spatial proteomics, and spatial metabolomics are three emerging spatial omics technologies that allow analyzing the distributions of RNA, protein, and metabolites simultaneously and in situ at tissue level, aiming to characterize tissues and organs better; thus, there is significant potential to assess disease risk mechanisms, including the risk of asthma, at different scales by integrating environmental exposure data into spatial omic analyses [2]. Moreover, despite the availability of substantial data across different areas, the integration of spatial omics with environmental exposure data remains uncommon due to the lack of methods to easily and effectively fuse these heterogeneous information sources [3]. The goal is thus to propose a methodology to facilitate the integration of information provided by different spatial omics technologies with environmental exposure data from varied sources. More specifically, the objectives are set to be the following: [1] to identify the modality of spatial omics data, the types of environmental exposure data, and the corresponding environmental exposure metrics that could serve as proxies for evaluating asthma risk in a nationwide study; [2] to review the current and emerging spatial omics technologies/modalities, datasets related to environmental exposure compliance, harmonization protocols and methods applied for different data types to summarize the information fitting the objectives – different environmental exposure data, chosen metrics, environmental exposure surveys, and suitable data integration methods[4, 5].

Conceptual Framework

Asthma is a chronic respiratory disease characterized by variable and recurrent symptoms and airflow obstruction due to bronchial inflammation [7]. Individuals with asthma can experience a wide range of symptoms, including coughing, wheezing, shortness of breath, chest tightness, and sleep difficulties. According to the World Health Organization (WHO), asthma affects around 262 million individuals in the world and leads to about 461,000 deaths annually [8]. The representative and population-level daily asthma risk across the US is estimated using commonly available data sources. Another data integration analysis estimates asthma risk across the US using the collected datasets and generates a fine-resolution spatiotemporal surface to model and understand the association between asthma risk and environmental factors [3]. Asthma is used as a real application to illustrate the proposed methods. Model performance with and without spatiotemporal health data highlights the value of incorporating health-affecting signals into risk modeling [5].

Data Sources and Integration Methods

Spatial omics enables the observation of biological specimens within their spatial contexts and provides quantitative assessments of tissue composition [6]. These technologies have gained traction among biomedical researchers because they hold promise for revealing spatiotemporally dynamic molecular mechanisms underlying disease-associated environmental exposures [11]. Spatial omics data are stored as digital expression matrices of spatially resolved molecular profiles measured at varying tissue scales. For many spatial omics modalities, spatial entities such as whole tissues, tissue slices, and histological slides are positioned on single-entity platforms. However, spatial coordinates are not obtained from single-cell profiling techniques, resulting in a looser definition of spatial omics [10]. These spatial coordinates, determined by tissue-level experimental designs at facility sites, may be assigned with high accuracy after post-processing, supporting both site knowledge transfer and efficient resource use [9]. Data movement across different spatial scales is also facilitatory [8].

Spatial Omics Technologies

A wide array of spatial omics technologies now enables the molecular profiling of tissues at unprecedented spatial resolution [4]. These technologies fall broadly into three categories: tissue imaging, transcriptomics, and proteomics. Tissue imaging approaches, such as imaging mass spectrometry and multiplexed ion-beam imaging (MIBI), provide spatial distributions of metabolites, proteins, and lipids; imaging techniques yield spatial data in 2D coordinates and often do not require tissue slide pre-processing[5]. For many tissue types, spatial imaging is performed directly on fresh-frozen tissue without sample preparation. However, the indicators chosen to characterize tissue samples do not extend to explicit stress perturbations, while the provided visualization is non-map-like and can hinder exposure reconciliation [15]. Spatial transcriptomics (ST) technology enables the spatial localization of poly-A-tagged transcripts at single-cell resolution. A high-resolution tissue section is initially scanned using a bright-field microscope or a stereoscope, and an imaging map of the section is subsequently constructed [14]. A spatial transcriptomics slide with a defined pattern of oligonucleotide spots is then positioned above the tissue. These spots consist of oligonucleotides with a poly T capture sequence, a unique spatial bar code sequence, and a universal sequencing primer sequence [13]. The constructed imaging map of the scanned tissue is then registered to the spatial transcriptomics spot coordinates to yield a pose map. Such technology routinely captures hundreds of genes, yielding thousands of transcriptional states at the cellular level. Nevertheless, the available gene pool typically does not cover direct upstream-determinative stress exposures, nor does ST provide map-like representations or reconciliation with exposure models [12]. Spatially resolved proteomics technologies, such as MIBI, SWEF, imCortex, CODEX, and REAP-Seq, build on multicolor fluorescence imaging to provide subcellular information about protein expression. These methods usually involve incubation with fluorophore-tagged antibodies, which limits the number of candidate proteins available for multiplexing; those not amenable to cross-species architecturing remain unmeasured, and a single canonical protein for each candidate typically constrains interpretation of conditions for high-lockdown stability in standard biobank settings [13]. Malignant status or manual cropping, therefore, remains necessary for robust localization. While mass-spectrometry-based alternatives (e.g., imaging mass cytometry, Hyperion, and MIBI) accept a wider range of individuals and permit proxying through dominance [5], falsifying organic structure design remains a critical issue for atmospheric chemistry models [4].

Environmental Exposure Datasets

Studies provide a wide variety of datasets on environmental exposure through adsorbents and land-use regression (LUR) models [6]. A variety of pollutants with distinct exposure metrics can be represented, both spatially and temporally. They all suffer from measurement errors and cannot be used directly beyond the original study area [7]. Strategies are needed to harmonize different exposure datasets accordingly [8]. Local estimated personas (LEPs) represent multi-way exposures developed from a diverse range of materials [4]. LEPs comprise an environmental analogue of comprehensive genomic profiles, integrating air pollutants, noise, meteorological conditions, and other environmental extremes from both local and broader sources, using machine-learning modelling methods to assist exposure risk mapping across dedicated populations within the framework of exposome health cohorts [5]. Both the modelling framework and LEPs from 60 cohorts, longitudinal or cross-sectional, urban, rural, or national, covering temporal, regional, or occasionally national global scaling, are made available within an online environmental exposome framework [7]. Another effort has focused on optimizing the joint use of spatially biased exposure datasets with a single-personal monitor providing a unique multi-scale estimate of indoor and outdoor air quality from short-term colocations on urban scales [4]. Environmental Similarity Learning enables the framework to derive multiple local estimated personal exposures on datasets beyond the original city and monitor. Instruction files are also provided to accompany the Python codes shareable on demand [2].

Data Fusion Techniques

Data fusion techniques integrate multiple datasets to enhance analysis quality and support better decision-making [15]. Typical applications encompass indoor environmental control practices aimed at asthmatic management, sensor monitoring at both indoor and outdoor hotspots, and predictive analytics targeting a decline in lung function among asthma patients [16]. Within the healthcare sector, transfer learning and neural network models have gained traction for prognostication pertaining to medical outcomes such as melanoma screening and bio-signal analysis [11]. The distinctively multidimensional nature of collected datasets in these domains presents specific challenges during integration, including the imbalance of informative yet incomplete observations, uncertainty regarding a suitable model, and the practical exchange of expert knowledge [10]. Addressing these hurdles can substantially elevate the overall efficacy of machine learning endeavours within healthcare and climate monitoring frameworks [6]. In the context of the exposome-asthma pipeline, various techniques are available to harmonise environmental exposure (time-series) data with biomolecular (high-dimensional) data [3]. Different strategies permit the accommodation of diverse forms of temporal exposure data as well as the fusion of other large-scale data typologies such as imaging data or genetic sequencing information [4].

Model Development and Interpretability

Predictive modeling combines exposure data with spatial transcriptomics to identify risk biomarkers for asthma—characteristics associated with susceptibility to the disease [3]. An environmental dataset characterizes exposure across locations and time frames, with mode-specific methods accommodating multimodal data. Individually, spatial transcriptomics and exposure matrices contribute linear Gaussian latent variable models that admit joint modeling of continuous and count responses [5]. The models can be fitted with Hamiltonian Monte Carlo methods that produce Markov chain Monte Carlo samples of the parameter posterior distribution. Interpretability is critical for characterizing and communicating the dependencies encoded in exposure–characteristic–risk associations [6]. To address this need, the design opts for interpretable models and interpretable data representations. Exposure characteristics are thus computed in a data-driven fashion from the environmental dataset, and several global post hoc explanation methods are available to highlight important features for sharing insights [7]. Secondary exposure–transcriptomic–dimensionality reduction models extend the interpretability framework further [14]. Bias and fairness considerations guide practices to avoid reinforcing existing inequalities through exposure data. Bias manifests in two ways: risk predictions shift across geographic populations, and population composition changes through data partitions [7]. Five datasets permit assessment of bias transferability and monitoring of features' potential explication of bias-related shifts. To detect unwanted shifts that arise, several established bias-detection techniques identify features whose marginal distribution differs between populations [13]. Mitigation strategies adapt procedure designs to minimize the inclusion of biased features in models, suggested through an equity lens. Transparency practices promote reproducibility and targeted distribution of methods and findings [12].

Predictive Modeling Approaches

Longitudinal analysis of electronic health record data provides patients' histories over time but is limited when behavioral changes or clinical services vary across jurisdictions [12]. Global model training at the population scale sacrifices specificity to local exposure. Environmental datasets, especially for measuring environmental pollutants over different time windows, are often poorly aligned and exhibit noise and drift, complicating systematic integration [13]. Existing healthcare analytics research tends to ignore privacy concerns related to sensitive patient information or focus solely on the privacy of the data itself rather than the organizational or institutional data use under different regulatory frameworks [14]. Deploying nationwide predictive models is inherently accompanied by considerable uncertainties regarding generalizability and external validity across diverse populations and geographic regions [15].

Interpretability Methods

While spatial omics and environmental exposure datasets capture different biological and environmental processes, they may describe overlapping areas of exposure, and thus fusion methods can still be beneficial [5]. A data fusion framework permits flexible specification of temperature exposures and allows the inclusion of heterogeneous spatial-omics features such as genotype, transcriptome, and epigenome data at cell type or population levels [4]. Relevant geospatial layers are updated per project and curated to minimize measurement error [11]. Machine learning approaches for air pollution data were leveraged, yet the conceptual clarity of statistical methods also seemed relevant [3]. Estimation with either technique generates a distribution of predictions for each target pixel given the observed data. Statistical models are well acknowledged in the literature and remain viable candidates for integration of any structured, time series observations [6]. A framework called spatial multi-species indicator kriging for the joint mapping of several variables accommodates these requirements: spatio-temporal mappings can be proposed where a limited number of sampling sites are available for each variable of interest without detrimental assumptions on the correlation structure [16].

Bias and Fairness Considerations

Risk-prediction models, widely employed across disciplines, must address bias and fairness concerns prior to deployment: definition depends on the objectives of the target group, potentially involving their characteristics and intersectional aspects [4]. For spatial-omics–environmental-exposure–asthma-risk prediction, bias originates from algorithm, evolutionary dynamics, and methodological choices; in detail, inference may depend on spatial or temporal biological data also correlated with asthma risk, macroenvironmental exposure affecting microenvironments linked to asthmatic variables, and nonrepresentational datasets notably, uneven population density with high-risk zones with limited data, leading to overconfidence in low-damage zones[7]. At the genome scale, experimental design, data participation, and specific biological information, such as ancestry and temporal knowledge, may also entrench biological dependence [5]. Multi–environmental–exposure data permit the study of compounds' simultaneous relations, yet past compounds and contagions exert long–lag influences [8]. Individual models assess model-usage bias, namely, whether the model is better in a subset of populations, ages, or spaces. Approaches include: categorical features only; Global-Drift and Disparate-Impact; Feature-M≥α Skilled; Logit-Poly. Global-Drift probes when characteristics exhibit drift, mutual dependencies, or multiple individuals for influence determination [9].

Real-World Performance and Validation

Asthma is a complex, heterogeneous chronic respiratory condition that contributes to considerable morbidity and mortality globally [13]. Projections indicate that the prevalence of asthma will increase substantially in the coming years, with an associated increase in healthcare costs largely attributed to the development of more expensive therapies and a greater frequency of hospitalizations due to exacerbations [14]. Numerous studies have established connections between asthma and various environmental exposures, such as exposure to particulate matter with a diameter of 2.5 or less (PM2.5), volatile organic compounds (VOCs), nitrogen dioxide (NO_2), and formaldehyde [17]. However, many of these studies either investigate exposure-response relationships in the general population or focus on groups with a diminished number of environmental indicators, often to the exclusion of spatially dependent data or data from even broader domains that are nevertheless relevant [13]. The dearth of knowledge surrounding the environmental determinants of health is further compounded by the concentration of individuals living and working in urban environments on a global scale [10, 6].

External Validation and Generalizability

External validation and generalizability remain pivotal throughout the data-fusion enterprise. Such rigorous tests model performance across distinct datasets and various populations [11]. In the present case, multiple exposure datasets were used in the risk model training to predict an independent population [10]. Two datasets were provided for internal validation, one from the same year, one older, yet they relied on single spatial omics modalities. Complementary analyses on coarser, publicly available, yet relevant molecular feature sets further assessed the model's robustness to diverse measured biology [9]. Spatial omics with targeted biological assessment garnered attention in environmental-health investigation of exposures influencing human health [16]. The majority of models privy to multiple datasets still prescribe spatial omics alone, neglecting previous coarser molecular exploration [2].

Prospective Deployment Scenarios

Integrating spatial omics data with environmental exposure data will help to advance the understanding of the relationship between the environment and asthma [8]. Decisions on deploying hybrid frameworks in real settings should consider the scientific, technical, and regulatory context [6]. Six prospective deployment efforts are proposed: measuring the effects of urban vegetation, mass transit points, and site disturbance on airborne dust with and without preemptive interventions probed through dust concentration in one cohort study; measuring the effects of urban green cover on asthma onset in a different cohort; evaluating the effect of varnish usage on room air quality before and after applying protective finishes in the same study; assessing the effect of preemptive cleaning of porous materials after fire incidents within the same cohort; and appraising the effect of cleaning different surfaces on the same cohort[14]. For each effort, real-world success hinges on the availability of hygienic and biogenic data, model performance, and community support [1, 10].

Implementation Considerations

Spatial machine learning models such as support vector regression, random forests, and multi-target regression can utilize various environmental input datasets to predict air pollutant concentrations across a region. Informative spatial predictor variables are identified through stepwise selection or multi-index strategies [9]. More than 20 previous environmental samples enable the transferability of input measures across space [10]. Engaging 10, 20, and 30 additional monitors on health exposure and disease toward three-distribution classification can further test sparsity and evacuation [11]. For airborne particulate matter detection, specific sites on the major road network are prioritized [12]. Interpretation of spatial-prediction models plays a key role in ensuring their usability for environmental and health studies [13]. Variable importance measures and uncertainty quantification methods complement the models to support relatively reliable inference [4]. Data sharing, as well as privacy and confidentiality, have received considerable attention throughout the public health domain. Multiple sharing platforms and frameworks, such as the National Institute of Health Data Sharing Policy and the Data Sharing and Integration Framework, have been established to promote data sharing and encourage sharing-dependent activities [16]. However, due to the sensitive nature of health records and the continued accumulation of environmental hygiene data nationwide, sharing health or environ-indexed data still harbors significant barriers. To facilitate the generation of pooling samples while protecting data privacy and owner rights, relevant governmental laws and regulations need to be strictly adhered to [17]. Development of de-identification methodologies enables the removal of direct identifiers and the transformation of quasi-direct identifiers to diminish re-identification risks. Data-governance policies are fundamental to defining data access levels, use constraints, individual obligations, and derived-data ownership for protected health information and environmental data [12].

Computational Infrastructure

Fulfilling asthma risk of harm assessments involves predictive modeling using massive spatial multicellular omics data, environmental exposures, and other covariates [15]. Individual-level spatial omics profiles reflect the composition of tissues accessible through biopsies, reflecting exposures on time scales slower than tissue turnover. Time-series environmental exposure data are cataloged by public authorities and available at varying spatial resolutions; higher-resolution data can require extensive harmonization [16]. At the local scale, exposures are temporally variable, compatible with time-series databases such as publicly available air pollution, terrestrial moisture, businesses, and traffic [15]. Many modalities cannot inject temporal data. Feasible scaling begins with high-capacity but low-correlation multimodal marginal embeddings from HD spatial omics extinction or outlier detection; up to $\approx 90\%$ loss across all biophysical spatial morphometric and covariance molecules and tags means no low-dimensional feature distribution is expected [12]. To mitigate potential exposure ascertainment bias and excessive learning, fuse extracted high-maxwellian-meso-mimetic-space-metamode embeddings along the temporal dimensional fixture with the high-capacity, low-correlation volum/max average unoccupied-textural-dynamics-w-carb-margext/data-free delivery from readout HD spatial-mf odel [2]. Curation of ambient exposure life-risk and non-visit tracing of many-day low-variability exposure full-death-blank attributes reduces ascertainment bias while maintaining the requisite exposure-size fill and considerable low-fidelity full-blank lower-information sufficiency[16].

Workflow Transparency and Reproducibility

Workflow transparency and reproducibility represent essential principles for facilitating science, applied research, and observational studies [10]. Many countries, scientific communities, funding agencies, and journals promote the adoption of open science and open research practices through principles and checklists, policies, recommendations, and guidelines [9]. To build transparent and reproducible models and systems, the Data, Computer Code, and Workflow (DCW) collaboration framework proposes jointly documenting the arrival, processing, transformation, analysis, storage, and dissemination of data, computer code, and workflow, a division aligned with the spatial omics and environmental exposome pipeline [12]. DCW rising requirements extend to provenance and workflow metadata specifications, model specification and storage, calibration records, evaluation reports, and methodological and contextual information [4]. The original data sources, processing methods, and parameter settings follow best practices from other kinds of pipeline data (https://www.pipeline-partners.com/) and granularity recommended for applied studies: include supplementary files, readmes for datasets and processing tools, relevant software locations, references, and ground-truth datasets only [8]. The exposition offers the minimal descriptive information often required for compliance [5].

Data Governance and Privacy

To mitigate the risks of disclosing sensitive geospatial information in asthma research, it is essential to establish robust governance structures for data access and sharing [5]. Data governance encompasses both technical and administrative processes aimed at safeguarding the integrity of datasets and the security of personally identifiable information (PII) [9]. The establishment of clear governance policies and procedures enhances accountability, compliance, and trust [7]. Efforts should focus on ensuring that all asthma-related data collection, usage, and dissemination practices adhere to the policies of the Grand Challenges in Environmental Exposure and Health Data Science community. Houston et al. recommend that the implementation of the Spatial Omic Exposure (SOMEx) data-collection system be contingent upon the existence and formalization of data-governance policies [13]. Asthma represents a complex interplay of environmental, genomic, and other data inputs in determining risk [11]. Integrating spatial omics datasets with asthma risk geospatial and temporal environmental exposures constitutes a further extension of this complexity [12]. In addition to these data assets, specific datasets such as participant covariates and follow-up information on asthma diagnosis and treatments must also be collected in order to properly parameterize models of interest [14].

Equity and Policy Implications

Asthma and other diseases associated with environmental exposures disproportionately affect already vulnerable communities [7]. Therefore, assisting public health authorities, environmental organizations, and community networks in identifying high-risk populations and implementing interventions involves not only addressing technical and interpretability challenges but also prioritizing the equitable assessment of exposure data and risk predictions [3]. Specific considerations include: access and benefit-sharing mechanisms, stakeholder engagement, regulatory frameworks, and ethical guidelines that influence the effective application of the developed models, thus pursuing a translational approach beyond technical advancements [15, 16].

Access and Benefit Sharing

Equitable access to data, tools, and methods while promoting fair deliberation of subsequent uses can be supported through various mechanisms [8]. Partnerships with institutions serving underrepresented populations can facilitate data transfers while involving stakeholders that enhance study relevance [17]. In addition, complying with applicable data use agreements and accreditation requirements can help foster scientific collaborations and data sharing, respectively [2].

Stakeholder Engagement

Involving stakeholders helps ensure that the proposed models advance the research and needs of the community. Candidates for engagement include the community residing in the area of interest and ecologists and public health practitioners focused on the region's trees and vegetation [3]. Modes of engagement range from public meetings that introduce the research objective and gather opinions on community needs to presentations of drafted manuscripts that elicit feedback [1]. Materials include an informative postcard outlining the research objective, a letter inviting participation in public meetings to better align with community needs, and a drafted manuscript for expert review [2, 7].

Regulatory and Ethical Frameworks

Local and national regulations, guidelines, and frameworks govern public health research in spatial omics, environmental exposure, and asthma risk [6]. Regulations include the Family Educational Rights and Privacy Act, the Federal Information Security Modernization Act, the Federal Information Processing Standards, the Health Insurance Portability and Accountability Act, the Privacy Act, the Safe and Accurate Framing of Environmental Research Act, and the Environmental Protection Agency Adequate Sensory and Health Monitoring Organization [7]. Ethical Considerations, including the International Federation of Networks of Emerging Infectious Disease, the International Conference on Harmonization Good Clinical Practice, and the National Institutes of Health, Guide to the Public Health Research Data Sharing Policy, provide additional relevant guidance [1].

Challenges and Limitations

Asthma remains a pressing public health issue. Despite the increasing availability of large, high-dimensional, and heterogeneous datasets related to its risk factors, progress on preventive interventions is still limited[17]. The rise of spatial omics technology has revolutionized biospecimen analysis at an unprecedented spatial resolution. Such spatial datasets containing genomic and transcriptomic information from human biospecimens, captured together with geographic coordinates, present rich opportunities for research [16]. Various spatial information, such as neighbourhoods, land use, vegetation coverage, etc., have been shown to have close relationships with environmental exposures such as pollution and may serve as proxies to support the geospatial characterisation of the environment surrounding these biospecimens[15]. Geospatial identification of such environmental exposures around human biospecimens can further enhance asthma-related risk analyses. The tightly correlated exposome and spatial omic information, however, can hardly be analysed simultaneously, even by state-of-the-art machine learning methods, particularly with the diverse forms of the two informative pieces. Their co-analysis remains a challenge to be solved for asthma-risk analysis [4]. The integration of spatial omics and time-varying environmental exposure datasets for predicting individual asthma risks has not been conducted yet. Concerns on the interpretability of the model emerge, considering that the constructed environmental dataset widely carries the influence of social-economic factors, including policies, regulations, socio-economics, and land-use. Furthermore, how to ensure the real-world deployment efficacy is uncertain [13]. In summary, the development of a method to predict individual asthma risk by jointly modelling the co-variation of molecular omics, exposome, and high-dimensional spatial variables is demonstrated. Various strategies to address interpretation, bias, and real-life implementation are further given [18-24].

CONCLUSION

Integrating spatial omics with environmental exposure data presents a promising approach to advance asthma risk prediction and inform precision public health strategies. By leveraging high-resolution molecular profiles alongside spatiotemporal environmental datasets, predictive models can better identify at-risk individuals and elucidate underlying disease mechanisms. Real-world implementation requires careful attention to interpretability, bias mitigation, data governance, and equitable access, ensuring that vulnerable populations benefit from advances in predictive analytics. Future research should focus on refining data-fusion methodologies, validating models across diverse populations, and developing ethical and policy frameworks that enable robust, reproducible, and inclusive asthma risk prediction at scale.

REFERENCES

1. Soyiri IN, Sheikh A, Reis S, Kavanagh K, Vieno M, Clemens T, Carnell EJ, Pan J, King A, Beck RC, Ward HJ. Improving predictive asthma algorithms with modelled environment data for Scotland: an observational cohort study protocol. BMJ Open. 2018 May 1;8(5):e023289.
2. Fecho K, Ahalt SC, Appold S, Arunachalam S, Pfaff E, Stillwell L, Valencia A, Xu H, Peden DB. Development and application of an open tool for sharing and analyzing integrated clinical and environmental exposures data: asthma use case. JMIR Formative Research. 2022 Apr 1;6(4):e32357.
3. Ugwu CN, Ugwu OP, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Ejemot-Nwadiaro RI, Okon MB, Egba SI, Uti DE. Sustainable development goals (SDGs) and resilient healthcare systems: Addressing medicine and public health challenges in conflict zones. Medicine. 2025 Feb 14;104(7):e41535.
4. Bae WD, Alkobaisi S, Horak M, Park CS, Kim S, Davidson J. Predicting health risks of adult asthmatics susceptible to indoor air quality using improved logistic and quantile regression models. Life. 2022 Oct 18;12(10):1631.
5. Maitre L, Guimbaud JB, Warembourg C, Güil-Oumrait N, Petrone PM, Chadeau-Hyam M, Vrijheid M, Basagaña X, Gonzalez JR, Exposome Data Challenge Participant Consortium. State-of-the-art methods for exposure-health studies: results from the exposome data challenge event. Environment International. 2022 Oct 1;168:107422.
6. Gautam Y, Johansson E, Mersha TB. Multi-omics profiling approach to asthma: an evolving paradigm. Journal of Personalized Medicine. 2022 Jan 7;12(1):66.
7. Ugwu OP, Alum EU, Ugwu JN, Eze VH, Ugwu CN, Ogenyi FC, Okon MB. Harnessing technology for infectious disease response in conflict zones: Challenges, innovations, and policy implications. Medicine. 2024 Jul 12;103(28):e38834.
8. Bae WD, Kim S, Park CS, Alkobaisi S, Lee J, Seo W, Park JS, Park S, Lee S, Lee JW. Performance improvement of machine learning techniques predicting the association of exacerbation of peak expiratory flow ratio with short-term exposure level to indoor air quality using adult asthmatics' clustered data. PLoS One. 2021 Jan 7;16(1):e0244233.
9. Thessen AE, Grondin CJ, Kulkarni RD, Brander S, Truong L, Vasilevsky NA, Callahan TJ, Chan LE, Westra B, Willis M, Rothenberg SE. Community approaches for integrating environmental exposures into human models of disease. Environmental health perspectives. 2020 Dec 28;128(12):125002.
10. Ongesa TN, Ugwu OP, Ugwu CN, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Okon MB, Ejemot-Nwadiaro RI. Optimizing emergency response systems in urban health crises: A project management approach to public health preparedness and response. Medicine. 2025 Jan 17;104(3):e41279.
11. Flynn C, Guha A, Majumdar S, Srivastava D, Zhou Z. Towards algorithmic fairness in space-time: filling in black holes. arXiv preprint arXiv:2211.04568. 2022 Nov 8.
12. Machado AF, Hu F, Ratz P, Gallic E, Charpentier A. Geospatial disparities: A case study on real estate prices in Paris. arXiv preprint arXiv:2401.16197. 2024 Jan 29.

13. Cheng S, Blanco MN, Sheppard L, Shojaie A, Szpiro A. Variable importance measure for spatial machine learning models with application to air pollution exposure prediction. arXiv preprint arXiv:2406.01982. 2024 Jun 4.
14. Ugwu CN, Ugwu OP, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Ejemot-Nwadiaro RI, Okon MB, Egba SI, Uti DE. Medical preparedness for bioterrorism and chemical warfare: A public health integration review. Medicine. 2025 May 2;104(18):e42289.
15. Vlaanderen J, Portengen L, Chadeau-Hyam M, Szpiro A, Gehring U, Brunekreef B, Hoek G, Vermeulen R. Error in air pollution exposure model determinants and bias in health estimates. Journal of Exposure Science & Environmental Epidemiology. 2019 Mar;29(2):258-66.
16. Xu H, Cox S, Stillwell L, Pfaff E, Champion J, Ahalt SC, Fecho K. FHIR PIT: an open software application for spatiotemporal integration of clinical data and environmental exposure data. BMC medical informatics and decision making. 2020 Mar 11;20(1):53.
17. Comess S, Akbay A, Vasiliou M, Hines RN, Joppa L, Vasiliou V, Kleinstreuer N. Bringing big data to bear in environmental public health: challenges and recommendations. Frontiers in artificial intelligence. 2020 May 15;3:31.
18. Paul-Chima UO, Ugwu CN, Alum EU. Integrated approaches in nutraceutical delivery systems: optimizing ADME dynamics for enhanced therapeutic potency and clinical impact. RPS Pharmacy and Pharmacology Reports. 2024 Oct;3(4):rqae024.
19. Boyd A. Enhancing Environmental Data Resources in Cohort Studies: ALSPAC exemplar (ERICA). International Journal of Population Data Science. 2018 Sep 11;3(4).
20. Kerr GH, Van Donkelaar A, Martin RV, Brauer M, Bukart K, Wozniak S, Goldberg DL, Anenberg SC. Increasing racial and ethnic disparities in ambient air pollution-attributable morbidity and mortality in the United States. Environmental health perspectives. 2024 Mar 6;132(3):037002.
21. Paul-Chima UO, Nneoma UC, Bulhan S. Metabolic immunobridge: Could adipose-derived extracellular vesicles be the missing link between obesity, autoimmunity, and drug-induced hepatotoxicity?. Medical Hypotheses. 2025 Sep 28:111776.
22. Ho YH, Liu Z, Lee CC, Mostafavi A. ML4EJ: Decoding the role of urban features in shaping environmental injustice using interpretable machine learning. Cities. 2026 Feb 1;169:106546.
23. Richey MM, Bang J, Sivaraman V. Targeting disparate spaces: new technology and old tools. Frontiers in Public Health. 2024 Apr 23;12:1366179.
24. Trupin L, Balmes JR, Chen H, Eisner MD, Hammond SK, Katz PP, Lurmann F, Quinlan PJ, Thorne PS, Yelin EH, Blanc PD. An integrated model of environmental factors in adult asthma lung function and disease severity: a cross-sectional study. Environmental Health. 2010 May 20;9(1):24.

CITE AS: Waiswa Arajab (2026). Integrating Spatial Omics with Environmental Exposure Data for Asthma Risk Prediction: Interpretability, Bias, Real-World Performance, Implementation, and Equity. RESEARCH INVENTION JOURNAL OF SCIENTIFIC AND EXPERIMENTAL SCIENCES 6(1):67-74. https://doi.org/10.59298/RIJSES/2026/616774