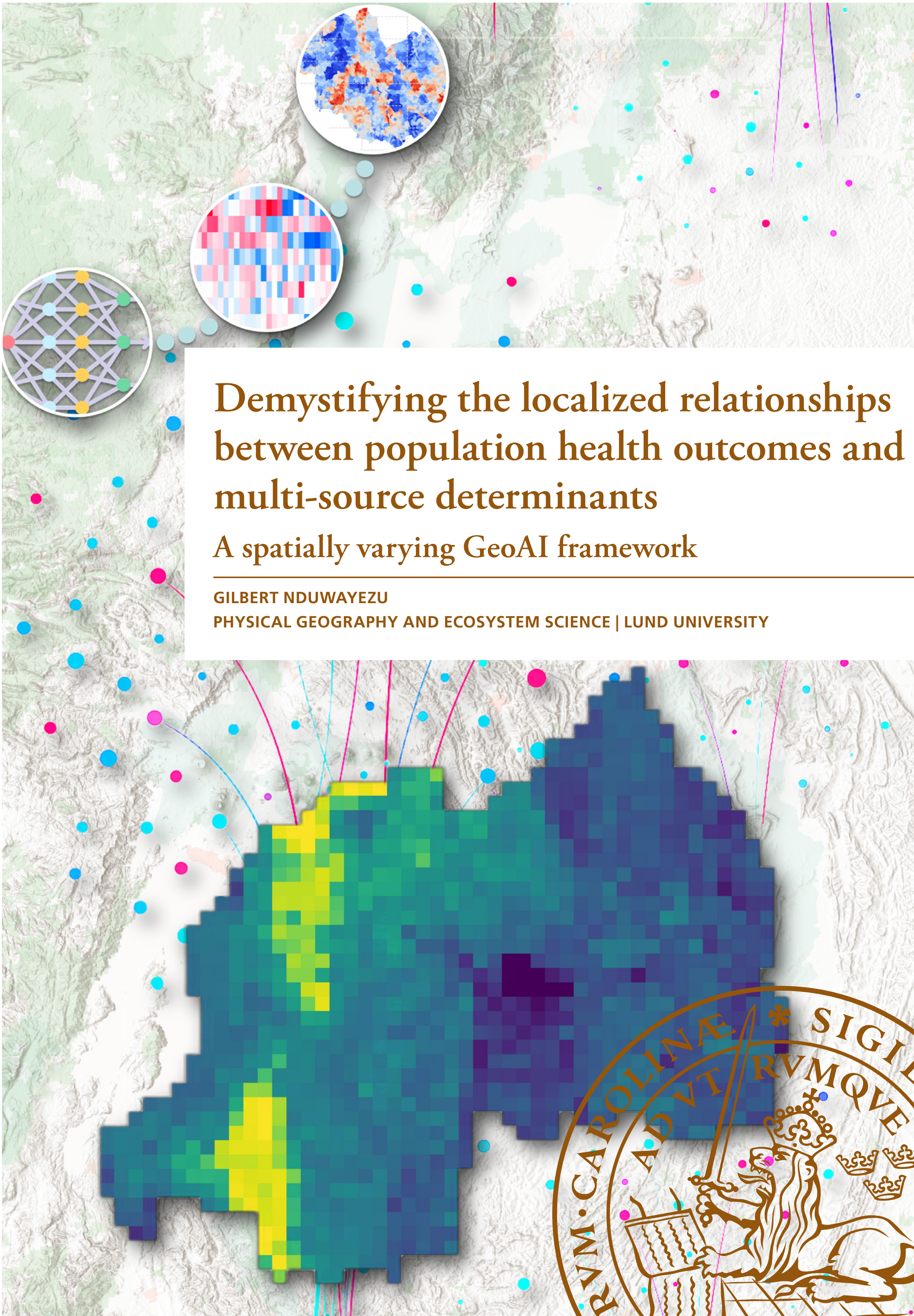
Demystifying the localized relationships between population health outcomes and multi-source determinants: A spatially varying GeoAI framework

Nduwayezu, Gilbert

2025



Demystifying the localized relationships between population health outcomes and multi-source determinants

A spatially varying GeoAI framework

GILBERT NDUWAYEZU

PHYSICAL GEOGRAPHY AND ECOSYSTEM SCIENCE | LUND UNIVERSITY

Demystifying the localized relationships between population health outcomes and multi-source determinants

Demystifying the localized relationships between population health outcomes and multi-source determinants:

A spatially varying GeoAI framework

Gilbert Nduwayezu

LUND UNIVERSITY

DOCTORAL DISSERTATION

Doctoral dissertation for the degree of Doctor of Philosophy (PhD) at the Faculty of Science at Lund University to be publicly defended on 11th December at 13.00 in Världen Geocentrum II, Sölvegatan 12, Lund

Faculty opponent
Professor Johannes Scholz

Organization: LUND UNIVERSITY

Document name: Doctoral dissertation

Date of issue: 2025-12-11

Author(s): Gilbert Nduwayezu

Sponsoring organization:

Title and subtitle: Demystifying the localized relationships between population health outcomes and multi-source determinants: A spatially varying GeoAI framework

Abstract:

Data-driven modeling frameworks have become essential tools for guiding surveillance strategies and informing public health policies across diverse population health challenges. Accurate, fine-scale disease estimates often lacking from direct surveys are critical for policy planning, given that spatial heterogeneity and nonlinear dynamics among determinants of health challenge classical models, limiting their utility for targeted public health interventions. Advances in geospatial artificial intelligence (GeoAI), and increased computational power, have enabled deeper insights into spatial non-stationarity, the shifting strength and direction of relationships across space. These advances enhance both the accuracy and contextual relevance of spatial modeling, supporting localized decision-making in population health outcomes. Drawing on diverse spatial frameworks, this thesis developed, tested, and applied localized spatially varying GeoAI methodologies, offering integrated modeling approaches to address stunting among children in the complexities of such public health concern. Paper I develops localized spatially varying approaches to reveal significant intra-area variation in stunting prevalence and nonlinear relationships using cross-sectional socioeconomic and fine-scale remotely sensed climatic and agroecological data to better characterise household microenvironments. The approach provided a more detailed understanding of how local environments shape nutrition outcomes and demonstrating the importance of considering both scale and nonlinearity in stunting research. Building on this, Paper II implements a hybrid spatial machine learning (ML) framework to detect fine-scale heterogeneity in stunting prevalence, while also quantify localized disparities that national-level surveys overlook. The framework captured spatially heterogeneous outcomes across most areas, with predictors exhibiting regions-specific effects that vary according to different thresholds of influence. Paper III advances the analysis by implementing a hybrid spatially varying deep learning (DL) approach, which captured convoluted nonlinear influence of fine socio-economic determinants on child stunting outcomes. The algorithm fairly captured variability in stunting outcomes, highlighting key child, maternal, and household determinants whose contributions varied across space, though limitations in training data size constrained broader generalizability. Paper IV further refines this perspective by introducing a predictive multilevel spatial ensemble learning (SEL) framework to produce small area estimates (SAEs) of stunting risk by combining geomasked household data with agroecological and remote sensing (RS) indicators. This approach demonstrated the capacity of predictive models to generalize beyond sampled survey clusters and produce continuous prevalence surfaces at scales as fine as 1 km². Overall, these papers highlight that trade-offs between interpretability, generalizability, and spatial scale in these analytical and predictive models remain challenging to navigate and must be evaluated case by case according to research priorities. The methodologies presented in this thesis aim to generate fine-scale, interpretable risk estimates that can support targeted nutrition interventions in data-scarce settings.

Key words: Population health, childhood stunting, multi-source data, spatially varying GeoAI, small area estimates, interpretability, localized interventions

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language English

Number of pages: 78

ISSN and key title:

ISBN: 978-91-89187-65-8 (print), 978-91-89187-66-5 (e-book)

Recipient’s notes

Price

Security classification

Signature

Date 2025-10-24

Demystifying the localized relationships between population health outcomes and multi-source determinants:

A spatially varying GeoAI framework

Gilbert Nduwayezu

LUND UNIVERSITY

Faculty of Science
Department of Physical Geography and Ecosystem Science

ISBN 978-91-89187-65-8 (print)

ISBN 978-91-89187-66-5 (electronic)

Printed in Sweden by Media-Tryck, Lund University,
Lund, 2025

NORDIC SWAN ECOLABEL
Printed matter
3041 0903

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

In loving memory of my mother

Table of Contents

Abstract

Data-driven modeling frameworks have become essential tools for guiding surveillance strategies and informing public health policies across diverse population health challenges. Accurate, fine-scale disease estimates often lacking from direct surveys are critical for policy planning, given that spatial heterogeneity and nonlinear dynamics among determinants of health challenge classical models, limiting their utility for targeted public health interventions. Advances in geospatial artificial intelligence (GeoAI), and increased computational power, have enabled deeper insights into spatial non-stationarity, the shifting strength and direction of relationships across space. These advances enhance both the accuracy and contextual relevance of spatial modeling, supporting localized decision-making in population health outcomes. Drawing on diverse spatial frameworks, this thesis developed, tested, and applied localized spatially varying GeoAI methodologies, offering integrated modeling approaches to address stunting among children in the complexities of such public health concern. Paper I develops localized spatially varying approaches to reveal significant intra-area variation in stunting prevalence and nonlinear relationships using cross-sectional socioeconomic and fine-scale remotely sensed climatic and agroecological data to better characterise household microenvironments. The approach provided a more detailed understanding of how local environments shape nutrition outcomes and demonstrating the importance of considering both scale and nonlinearity in stunting research. Building on this, Paper II implements a hybrid spatial machine learning (ML) framework to detect fine-scale heterogeneity in stunting prevalence, while also quantify localized disparities that national-level surveys overlook. The framework captures spatially heterogeneous outcomes across most areas, with predictors exhibiting regions-specific effects that vary according to different thresholds of influence. Paper III advances the analysis by implementing a hybrid spatially varying deep learning (DL) approach, which captured convoluted nonlinear influence of fine socio-economic determinants on child stunting outcomes. The algorithm fairly captured variability in stunting outcomes, highlighting key child, maternal, and household determinants whose contributions varied across space, though limitations in training data size constrained broader generalizability. Paper IV further refines this perspective by introducing a predictive multilevel spatial ensemble learning (SEL) framework to produce small area estimates (SAEs) of stunting risk by combining geomasked household data with agroecological and remote sensing (RS) indicators. This approach demonstrated the capacity of predictive models to generalize beyond sampled survey clusters and produce continuous prevalence surfaces at scales as fine as 1 km^2. Overall, these papers highlight that trade-offs between interpretability, generalizability, and spatial scale in these analytical and predictive models remain challenging to navigate and must be evaluated case by case according to research priorities. The methodologies presented in this thesis aim to generate fine-scale,

interpretable risk estimates that can support targeted nutrition interventions in data-scarce settings.

Popular summary

Population health outcomes emerge from complex spatial interactions among biological, environmental, and socioeconomic determinants, among others. Accurate understanding of these processes is critical for guiding targeted interventions, particularly in resource-limited settings where optimal allocation can substantially affect population well-being. Yet, many existing studies rely on coarse-scale analyses, producing global metrics that obscure local heterogeneity and limit both interpretability at fine scales and generalizability across contexts. This limitation is especially greater in low- and middle-income countries (LMICs), where health disparities are pronounced at local levels and interventions require precise targeting. In these data-scarce regions, remote sensing (RS) leverages satellite data combined with artificial intelligence (AI) to overcome limitations in data availability, providing a valuable alternative for assessing socioeconomic and agroecological factors critical to population health outcomes at a fine scale. Yet, classical analytical methods often fail to capture the complex, nonlinear relationships among health determinants, producing oversimplified models that inadequately reflect real-world complexity. Addressing these challenges requires methodologies that integrate localized interpretability, predictive generalizability, and spatial scale while preserving the natural pattern of spatial data. Geospatial artificial intelligence (GeoAI) has emerged as a powerful tool for population health analytics, particularly for capturing localized stunting risk patterns and generating high-resolution predictive models.

This study proposes an integrated approach allowing detailed local analysis across varied scenarios, bridging gaps of prior approaches and providing a consistent methodology applicable across varied epidemiological studies. Our framework integrates spatially varying coefficient (SVC) models with advanced artificial intelligence (AI) to capture nonlinear relationships, spatial dependencies, and local feature importance in childhood stunting and its complex determinants. To validate the proposed approach, we conducted analyses using two cross-sectional datasets on childhood stunting in Rwanda across three and five-year age groups, complemented by high-resolution climatic and agroecological data from multiple RS sources, to refine analysis of the microenvironment around the household and better capture the complex determinants of stunting among children.

The study found that sanitation deficits, topography, inadequate caregiving, poor education, limited antenatal care, low degree of urbanization, and climatic factors including rainfall and NDVI, proxies for several environmental processes may play important roles in driving stunting risk at fine spatial scales. Consistently across analyses, these influences were found to be convoluted and nonlinear, suggesting that stunting risk does not increase uniformly, instead varies by local context. This underscores the need for localized, easily adoptable methodologies to inform decision-making. In addition, we produced continuous surface maps covering unsampled areas, leveraging RS data to explain spatial variation in risk across vulnerable ecological areas. This twofold ability supporting both generalization and local decision-making addresses a major gap in current childhood stunting analytics, where models often struggle to balance interpretability and predictive accuracy performance. In practice, the approach can guide precise, location-specific interventions, optimize scarce health resources, and reduce spatial variation in outcomes.

This study integrates descriptive, exploration, and predictive analysis within a spatial framework, offering a blueprint for childhood stunting risk modeling, indicating how hybrid GeoAI approaches can process spatial epidemiological data into actionable insights for both research and policy. While spatial ensemble learning (SEL) generalized well to new dataset, its low explainability and the need to enhance causal analysis remain open challenges that future studies should address. Overall, the papers of this thesis highlight that trade-offs between interpretability, generalizability, and spatial scale in these analytical and developed predictive models remain challenging to navigate and must be evaluated case by case according to research priorities. The methodologies presented in this thesis aim to generate fine-scale, interpretable stunting risk estimates that can support targeted nutrition interventions under data-scarce conditions.

Populärvetenskaplig sammanfattning

Befolkningshälsa är resultatet av komplexa interaktioner mellan bland annat biologiska, miljömässiga och socioekonomiska faktorer. En korrekt förståelse av dessa processer är avgörande för att kunna styra riktade insatser, särskilt i resursbegränsade miljöer där en optimal fördelning av insatser kan ha stor inverkan på befolkningens välbefinnande. Många befintliga studier bygger dock på grova analyser, vilket döljer lokala variationer och begränsar både tolkningsbarheten på detaljerad nivå och möjligheten att generalisera mellan olika sammanhang. Denna begränsning är särskilt stor i låg- och medelinkomstländer (LMIC), där det finns stora skillnader i hälsa mellan olika områden och insatserna kräver precision. I dessa regioner med brist på data kan fjärranalys (RS) hjälpa till att mitigera begränsningarna i datatillgängligheten och är därmed ett användbart alternativ som använder satellitdata och artificiell intelligens (AI) för att bedöma socioekonomiska och agroekologiska faktorer som är avgörande för befolkningshälsa på detaljerad nivå. Klassiska analysmetoder misslyckas dock ofta med att fånga de komplexa, icke-linjära relationerna mellan hälsofaktorer, vilket resulterar i förenklade modeller som inte återspeglar verklighetens komplexitet. För att hantera dessa utmaningar krävs metoder som integrerar lokal tolkningsbarhet, prediktiv generaliserbarhet och rumslig skala samtidigt som det naturliga mönstret i rumsliga data bevaras. Geospatial artificiell intelligens (GeoAI) har visat sig vara ett kraftfullt verktyg för analys av befolkningshälsa, särskilt för att fånga upp lokala mönster av risk för tillväxthämning och för att skapa högupplösta prediktiva modeller.

Denna studie föreslår en integrerad metod som möjliggör detaljerad lokal analys i olika scenarier, överbrygger luckor i tidigare metoder och tillhandahåller en konsekvent metodik som kan tillämpas i olika epidemiologiska studier. Vårt ramverk integrerar modeller med rumsligt varierande koefficienter med avancerad artificiell intelligens (AI) för att fånga icke-linjära relationer, rumsliga beroenden och lokala särdrag som är viktiga för tillväxthämning hos barn och dess komplexa determinanter. För att validera den föreslagna metoden genomförde vi analyser med hjälp av två tvärsnittsdatauppsättningar om tillväxthämning hos barn i Rwanda i åldersgrupperna tre och fem år, kompletterade med högupplösta klimat- och agroekologiska data från flera RS-källor, för att förfina analysen av mikroklimatet kring hushållen och bättre fånga de komplexa determinanterna för tillväxthämning hos barn.

Studien visade att bristande sanitet, topografi, otillräcklig vård, dålig utbildning, begränsad mödravård, låg urbaniseringsgrad och klimatfaktorer som nederbörd och vegetation, som är indikatorer för flera miljöprocesser, kan spela en viktig roll för risken för tillväxthämning på fin rumslig skala. I alla analyser visade sig dessa influenser vara komplexa och icke-linjära, vilket tyder på att risken för tillväxthämning inte ökar enhetligt, utan varierar beroende på lokala förhållanden. Detta understryker behovet av lokaliserade, lättanvända metoder för att underlätta beslutsfattandet. Dessutom har vi tagit fram jämna prediktiva kartor som täcker områden som inte ingår i urvalet, med hjälp av fjärranalysdata för att förklara den rumsliga variationen i risk mellan olika utsatta ekologiska områden. Denna dubbla förmåga att stödja både generalisering och lokalt beslutsfattande fyller en viktig lucka i den aktuella analysen av tillväxthämning hos barn, där modellerna ofta har svårt att balansera tolkningsbarhet och prediktiv noggrannhet. I praktiken kan denna metod vägleda precisa, platsspecifika insatser, optimera knappa hälsoresurser och minska den rumsliga variationen i resultaten.

Denna studie integrerar beskrivande, utforskande och prediktiv analys inom ett rumsligt ramverk och erbjuder en mall för modellering av risker för tillväxthämning hos barn, vilket visar hur hybrid-GeoAI-metoder kan bearbeta rumsliga epidemiologiska data till användbara insikter för både forskning och politik. Medan rumsligt ensemble-lärande generaliserades väl till nya data, förblir dess låga förklarbarhet och behovet av att förbättra kausalanalysen öppna utmaningar som framtida studier bör ta itu med. Sammantaget framhåller artiklarna i denna avhandling att avvägningarna mellan tolkbarhet, generaliserbarhet och rumslig skala i dessa analytiska och utvecklade prediktiva modeller fortfarande är svåra att hantera och måste utvärderas från fall till fall utifrån forskningsprioriteringar. De metoder som presenteras i denna avhandling syftar till att generera finfördelade, tolkbara uppskattningar av risken för tillväxthämning som kan stödja riktade näringsinterventioner under förhållanden med brist på data.

List of papers

This thesis comprises the following papers, referenced in the text by Roman numerals and appended at the end.

I. **Nduwayezu, G.**, Kagoyire, C., Zhao, P., Eklund, L., Pilesjö, P., Bizimana, J. P., & Mansourian, A. (2024). Spatial machine learning for exploring the variability in low height-for-age from socioeconomic, agroecological, and climate features in the Northern Province of Rwanda. *GeoHealth*, *8*, e2024GH001027. https://doi.org/10.1029/2024GH001027

II. **Nduwayezu, G.**, Mansourian, A., Bizimana, J. P., & Pilesjö, P. (2025). Hybridizing spatial machine learning to explore the fine-scale heterogeneity between stunting prevalence and its associated risk determinants in Rwanda. *Geo-Spatial Information Science*, 1–21. https://doi.org/10.1080/10095020.2025.2459133

III. Kagoyire, C., **Nduwayezu, G.**, Oucheikh R., P., Bizimana, P., Pilesjö & Mansourian, A. From Kernel-Based to Neural Network Learned Spatial Weights: A Transferable GNNWR Workflow for Mapping Childhood Stunting Determinants. *(under review for publication in International Journal of Health Geographics)*.

IV. **Nduwayezu, G.**, Zhao, P., Pilesjö, P., Bizimana, J. P., & Mansourian, A. (2025). Multilevel small-area childhood stunting risk estimation: Insights from spatial ensemble learning, agro-ecological and environmentally remotely sensed indicators. *Environmental and Sustainability Indicators*, *27*, 100822. https://doi.org/10.1016/J.INDIC.2025.100822

Published papers during the doctoral program not included in this thesis

- Dusingizimana, T., **Nduwayezu, G.**, & Kjelqvist, T. (2024). Women's dietary diversity is associated with homestead production and market access: A cross-sectional study in rural Rwanda. *Maternal & Child Nutrition,* e13755. https://doi.org/10.1111/mcn.13755

- Ndagijimana, A., **Nduwayezu, G.**, Kagoyire, C., Elfving, K., Umubyeyi, A., Mansourian, A., & Lind, T. (2024). Childhood stunting is highly clustered in Northern Province of Rwanda: A spatial analysis of a population-based study. *Heliyon*, *10*(2), e24922. https://doi.org/10.1016/J.HELIYON.2024.E24922

- **Nduwayezu, G.**, Zhao, P., Kagoyire, C., Eklund, L., Bizimana, J. P., Pilesjö, P., & Mansourian, A. (2023). Understanding the spatial non-stationarity in the relationships between malaria incidence and environmental risk factors using Geographically Weighted Random Forest: A case study in Rwanda. *Geospatial Health*, *18*(1). https://doi.org/10.4081/gh.2023.1184

Author contributions

I. **NG** conceived the study and, together with the last author, led the development of the methodological framework. **NG** prepared the data and implemented the analyses. **NG** and co-authors jointly interpreted the results and wrote the manuscript.

II. **NG**, together with the second author, led the study design. **NG** prepared the data, implemented the methods, and performed the analysis. **NG** and co-authors interpreted the results, with **NG** leading the manuscript preparation.

III. **NG** participated in the study design and, together with the first author, led the methodology development and manuscript review.

IV. **NG**, together with the last author, conceived the study. **NG** independently prepared the data, led the methodology development, implementation, analysis, interpretation of results, and manuscript writing.

Abbreviations

AI	Artificial Intelligence
AICc	Corrected Akaike Information Criterion
ANN	Artificial Neural Networks
BMI	Body Mass Index
CatBoost	Category Boosting
DHS	Demographic and Health Surveys
DL	Deep Learning
EL	Ensemble Learning
ESDA	Exploratory Spatial Data Analysis
GBM	Gradient Boosting Machine
GAM	Generalized Additive Model
GLM	Generalized Linear Model
GeoAI	Geospatial Artificial Intelligence
GIS	Geographic Information System
GNNWR	Geographically Neural Network Weighted Regression
GW	Geographically Weighted
GWR	Geographically Weighted Regression
GWRF	Geographically Weighted Random Forest
GWSS	Geographically Weighted Summary Statistics
HAZ	Height-for-Age Z-score
HIV	Human Immunodeficiency Virus
IML	Interpretable Machine Learning
ITN	Insecticide-Treated Net
LMICs	Low- and Middle-Income Countries
LST	Land Surface Temperature
LightGBM	Light Gradient Boosting Machine
LR	Logistic Regression

MAUP	Modifiable Areal Unit Problem
MCC	Matthews Correlation Coefficient
MICS	Multiple Indicator Cluster Survey
ML	Machine Learning
NDVI	Normalized Difference Vegetation Index
OLS	Ordinary Least Squares
OOB	Out-of-Bag
PCA	Principal Component Analysis
PreLU	Parametric Rectified Linear Unit
R^2	Coefficient of determination
RF	Random Forest
RDHS	Rwanda Demographic and Health Survey
RS	Remote Sensing
RSS	Residual Sum of Squares
SAE	Small-Area Estimation
SAEs	Small-Area Estimates
SD	Standard Deviation
SGD	Stochastic Gradient Descent
SDGs	Sustainable Development Goals
SEM	Spatial Error Model
SEL	Spatial Ensemble Learning
SHAP	SHapley Additive ExPlanations
SLM	Spatial Lag Model
SRGCNNs	Spatial Regression Graph Convolutional Neural Networks
SVC	Spatially Varying Coefficient
SVM	Support Vector Machines
SWRF	Spatially Weighted Random Forest
SWNN	Spatially Weighted Neural Network
UK	Universal Kriging

WASH	Water, Sanitation and Health
WHO	World Health Organization
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

Acknowledgements

During these four years, my journey has not been a path filled with flowers, but rather one lined with thorns that often drained my energy and challenged my courage to continue. One of the most difficult moments was when I sustained a severe fracture in an accident that damaged my knee, leaving my leg weak and paralyzed. I spent four months hospitalized in a remote area of Rwanda, far from my family. To be honest, I nearly lost hope of continuing my PhD.

During my recovery, I lost my beloved mother just three months ago. She was the one who showed me the meaning of life, who supported me unconditionally, and who fulfilled both parental roles with strength and grace. She nurtured in me a passion for education, and both spiritually and materially, gave me everything she could. After her passing, I found myself almost lacking the courage to face the world without her. Every day, I still hear her voice reminding me to stay resilient, no matter what life brings. I cherish the memories we made during her four and a half years battle with illness. I feel grateful that I was by her side until her final breath. Though I began this PhD journey after she started losing her memory, I carried her words in my heart: *"Work hard, no matter what."* I never had the chance to tell her that I kept going, even after the accident that took my knee, but I did. I kept fighting to become the person she always believed I could be. A son she would be proud of. *Ma, ndagukunda.*

But after all, I was fortunate. I had people by my side who took my hand and encouraged me to fight, to hold on, and to keep going. First and foremost, I express my deepest gratitude to my principal supervisor, Prof. Ali Mansourian, for his exceptional guidance, patience, and constant support. His unwavering belief in me shaped not only this research but also the researcher I have become. He never imposed his academic standards but gave me the freedom to explore, while always being there to guide me with clarity and technical insight. His mentorship laid the foundation for my future academic career, and for that, I am truly grateful.

To my co-supervisors Prof. Petter Pilesjö, Dr. Lina Eklund, and Associate Prof. Jean Pierre Bizimana thank you for your invaluable contributions and support throughout my research. Working with you has been an academically enriching and personally meaningful experience, and I feel fortunate to have had your support.

I also extend heartfelt thanks to Prof. Anna Maria Johnsson, who served as my institutional representative and chair of my dissertation. Your unwavering support and mentorship were a source of motivation and strength.

This work was supported by the Swedish International Development Cooperation Agency (SIDA) through its collaboration with the University of Rwanda (grant no. 11277), with support from the UR-Sweden Programme Coordination Offices in

Rwanda and Sweden. Thank you, Prof. Etienne Ruvebana, Prof. Aline Umubyeyi, Peter Musoni, Darius Murangwa, Edward Kasumba, Oreste Mwiseneza, and Claudine Mukalinguyeneza, for managing all the necessary logistics so efficiently. I'm also grateful to the team leaders of the undernutrition sub-programme: Prof. Theoneste Ntakirutimana, and Dr. Eric Matsiko from the UR side, and Prof. Gunilla Krantz along with all supervisors from the Swedish universities. Your coordination and guidance have been key to my successful academic journey.

To Dr. Pengxiang Zhao and the researchers at the GIS Centre at Lund University, thank you for the stimulating discussions and technical support. I especially appreciate Pengxiang's many evenings spent helping me debug scripts I learned so much through your selfless dedication.

I am also grateful to my fellow PhD colleagues in the undernutrition sub-programme, especially Dr. Theogene Dusingizimana, Clarisse Kagoyire and Albert Ndagijimana, for your collaboration, discussions, and support. Our shared experiences enriched my understanding of data analysis and research methodology. I wish you both all the best in your future careers.

Throughout my PhD, I also received support from several researchers who generously responded to my queries, shared data, and helped clarify aspects of my research. Special thanks to Vestine Uwilingiyimana, Maurice Mugabowindekwe, and many others whose names are not all listed here thank you for helping shape my research into what it is today.

To my friends and colleagues in Lund, Maxime, Leonard, Joseph, Olive, and Emmanuel thank you for making Lund feel like a second home. Your presence made the long days and the distance from my family more bearable.

And lastly because they come first in my heart, I give my deepest thanks to my family. To my wife and children, Alvah, Ava, and Avi: you are my strength, my joy, my everything. Your love and support kept me going when I thought I couldn't continue. Without you, this PhD would not have been completed. There is so much I could say enough to fill hundreds of pages but to keep it simple: I love you.

I am deeply grateful to my family, mentors, and friends for walking with me through the thorns of this journey. May God bless you abundantly.

Rationale and thesis structure

Population health outcomes arise from the interplay of diverse biological, environmental, and socioeconomic factors, among others, all varying across space. In low-resource settings, where more precision and targeted intervention are most needed, classical analytical approaches are often inadequate, due to their reliance on coarse spatial resolution data and linear assumptions that obscure critical local variation, thereby limiting both actionable insight and transferability across regions. Specifically, in the Northern of Rwanda, child stunting, a core component of the undernutrition burden, has been observed for more than two decades, despite sustained clinical and nutritional interventions (Habtu et al., 2022; McLean et al., 2018). With most households dependent on rain-fed agriculture, food security and income remain highly sensitive to climatic and ecological factors often overlooked in child stunting analyses. Local, timely, and spatially explicit insights that incorporate these dimensions are essential for more effective, targeted responses. Despite the growing use of GeoAI in geospatial research, previous studies remain largely confined to traditional ML and statistical methods, often overlooking spatially varying relationships and failing to integrate interpretability with prediction, both of which are critical for understanding complex population health determinants.

This thesis addresses these limitations by proposing an integrative GeoAI framework that integrates these dimensions, enabling scalable, locally informed health analysis adaptable to varied contexts, aiming to describe, explain, and predict the spatial heterogeneity of child stunting in Rwanda. This dissertation is organized as a compilation of research papers, framed around the goal of using GeoAI to deepen spatial epidemiological understanding of child stunting. **Paper I** developed spatially varying models to explore nonlinear, location-specific relationships between childhood stunting and socioeconomic, climate, and agroecological factors among children under three. **Paper II** applied a hybrid ML framework to identify global and local determinants of stunting risk among children below the age of five. **Paper III** introduced a hybrid spatially varying DL model to capture complex spatial dependencies and key factors of stunting in children under three. **Paper IV** implemented a multilevel SEL approach to generate predictive SAEs maps of stunting risk using household, agroecological, and RS data. Through these efforts, the dissertation seeks to bridge the research and knowledge gaps by providing a more accurate and nuanced description of population health data, thereby facilitating more effective spatial analysis and decision-making processes.

The thesis synthesizes cross-paper findings, discusses implications for public health policy, and highlights future research directions. This thesis begins by outlining the persistent challenge of childhood stunting and urgent need for locally targeted interventions informed by a nuanced understanding of population health determinants. It then presents the aims and methods focused on characterizing spatial epidemiology of childhood stunting through localized analytical and predictive approaches. The results and discussion sections evaluate the application of GeoAI techniques, exploring both their analytical and predictive capabilities in capturing complex population health outcomes' patterns. Finally, future work is proposed to enhance the developed models by advancing causal inference, forecasting, and interpretability in spatial-temporal health data.

Background, related studies and motivation

In this section, we first review the epidemiology of childhood stunting, its multifaceted determinants, emphasizing the need for localized decision-making interventions. Second, we characterize geospatial population health data using spatially explicit methods, highlighting the importance of localized analysis. Third, we explore SVC and GeoAI approaches to enhance analytical and predictions of stunting risk. Finally, we discuss the interpretability and the limitations of current methods for geospatial health.

Childhood stunting: epidemiology, multifaceted determinants, and public health implications

Stunting in children (low height-for-age) is a major global public health concern, reflecting chronic undernutrition during the first 1000 days of a child growth (UNICEF et al., 2025; WHO, 2006). Undernutrition, defined as insufficient intake of energy and nutrients, manifests as wasting, underweight, or stunting (UNICEF et al., 2025). Stunting, assessed using the HAZ, is a key indicator of long-term nutritional deprivation, with children scoring below -2 SD classified as stunted, and -3 SD severely stunted (WHO, 2006), reflecting both biological and environmental growth constraints (Osgood-Zimmerman et al., 2018). Globally, around 151 million children under five are affected, with the highest burden in Sub-Saharan Africa and South Asia as depicted in Figure 1 (UNICEF et al., 2025). Within countries, stunting disproportionately affects children from disadvantaged households and marginalized communities (Mertens et al., 2023).

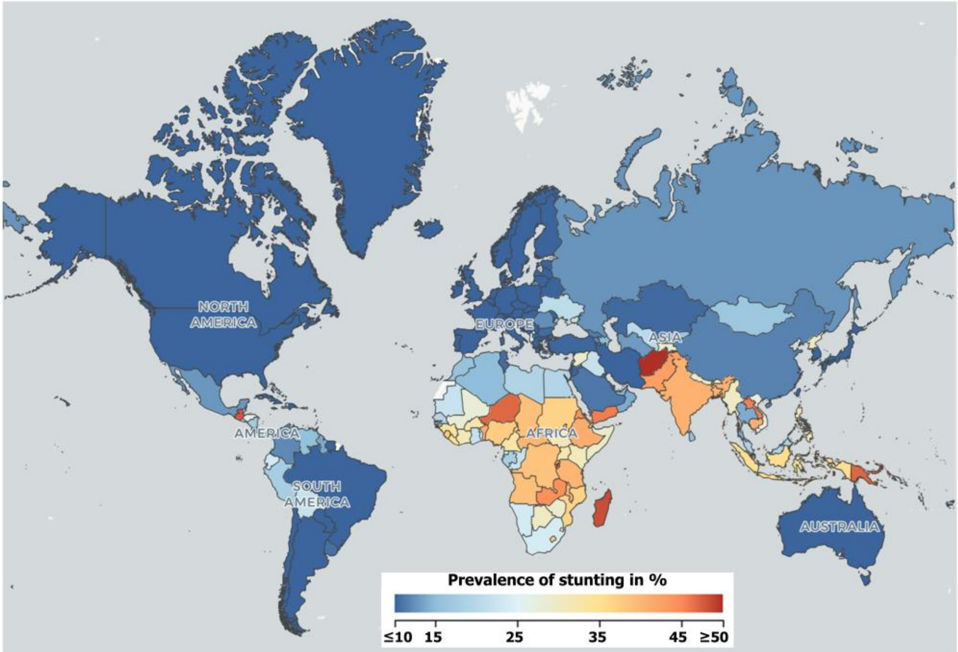


Figure 1: Global geographical distribution of stunting among children under five, presented as percentage prevalence by countries, with the highest levels in Sub-Saharan Africa, and Southern Asia. Colours indicate severity: very low (<10%) to very high (≥50%). Data reflect the most recent modeled estimates as of 2019 (IHME, 2020).

Stunting, a core component of the undernutrition burden, is shaped by diverse determinants: socio-economic inequality, poor education, limited access to healthcare, inadequate sanitation, unsafe water, recurrent infections, and food insecurity, among others (Baffour et al., 2023). In low- and middle-income countries (LMICs), inadequate nutrition is the primary driver of child stunting, yet evidence shows that climate change, agroecological conditions, and socio-economic factors further exacerbate it through multiple pathways (Aheto & Dagne, 2021; Baffour et al., 2023; Khaki et al., 2024; Nduwayezu et al., 2024). For example, extreme rainfall and droughts disrupt crop production, while sparse vegetation and poor soil fertility may reduce food availability, directly impacting dietary security (Christian & Dake, 2022; Khaki et al., 2024). Additionally, poor sanitation and waterborne diseases further impair nutrition and growth (Aheto & Dagne, 2021). Livestock ownership and access to Animal-sourced Foods (ASFs) are crucial for mitigating stunting, as these foods provide essential nutrients for development (Christian & Dake, 2022; Haileselassie et al., 2020). Studies show that children from livestock-owning households or those receiving protein-rich supplements exhibit better growth, emphasizing the need for integrated, multi-faceted solutions to tackle stunting (Dusingizimana et al., 2024). These determinants interact to perpetuate malnutrition across generations (UNICEF et al., 2025). The consequences extend beyond growth deficits to impaired cognitive and motor development, reduced school performance, higher disease susceptibility, and lower adult economic productivity (Osgood-Zimmerman et al., 2018), reinforcing cycles of poverty and health disparities (Baffour et al., 2023). To address childhood stunting, this study aligns with the SDGs, particularly SDG 2 (zero hunger), SDG 3 (good health and well-being), and SDG 10 (reduced inequality) (UN-HABITAT, 2018), to design effective interventions of such public health issue (Annan, 2018), which build on integrated approaches across the wider range of determinants, supported by different descriptive, analytical and geospatial predictive methodologies (Blanford, 2025).

Classical GIS approaches for spatial modeling of population health

Population health is fundamentally shaped by geography (Moore & Carpenter, 1999). Geospatial data, as a unique form of data, recording the spatial health attributes and their interactions of different locations (Moraga, 2019), allows accurate comprehensive descriptive, analytical and predictive, pivotal tasks for the success of spatial analysis in population health (Kirby et al., 2017). As such, the success in representing and analyzing geospatial information is key to ensuring the methodological rigor in public health (Wilson & Wakefield, 2020), calling for a careful and detailed approach to describing geospatial data (Lin & Wen, 2022). Understanding geospatial data and providing accurate descriptions are crucial for designing effective spatial models across diverse analysis tasks (Clark et al., 2024). In that context, spatial predictive models heavily relying on accurate descriptive and analytical tasks (Blanford, 2025), involve constructing appropriate models based on identified spatial patterns and interpreted relationships (Egaña et al., 2025).

These population health data, heterogeneous across space, are modeled in either continuous or discrete perspective (Anselin, 2010). The former assumes that spatial relationships vary continuously across space, while the latter, by contrast, assumes that coefficients vary across discrete subregions of the data (Guo et al., 2025), and are often captured by various spatially stratified modeling frameworks (Wang et al., 2024b; Wang et al., 2010). From a modeling perspective, population health outcomes such as child growth failure are rarely uniform across regions; instead, their prevalence shifts with socioeconomic, agroecological, and environmental contexts (Moraga, 2019). The discrete approach, however, captures broad contrasts between areas but risk oversimplifying health disparities by imposing hard boundaries. Continuous models, by contrast, acknowledge that population health determinants from nutritional to environmental change continuously with geographic locations, revealing spatial risk gradients hidden by stratified models (Konstantinoudis et al., 2020). The population health data are often collected at the household level, providing detailed spatial information on individual outcomes (Tuson et al., 2020). Yet, it is rarely feasible to design policies tailored to every individual location (Wang et al., 2010), meaning that even continuous models of health heterogeneity are typically applied to aggregated areal data that align with how public health interventions are implemented. To upscale sampled point data to discrete areal data, aggregation- based geomasking is applied with the dual goals of protecting privacy and providing a degree of utility for spatial analysis (Seidl, 2025). This process prevents identification of specific locations, households, or individuals represented in the dataset (Wang, 2024a). As shown in Figure 2, this dissertation focused on exploratory, analytical, and predictive modeling within a continuous spatial framework, applied to both point-referenced and aggregated areal data (Wang, 2024a), which better reflect the gradual transitions in health risks shaped by environment, behavior, and socioeconomics (Lin & Wen, 2022), and thus provide a more precise basis for explanatory and predictive intervention.

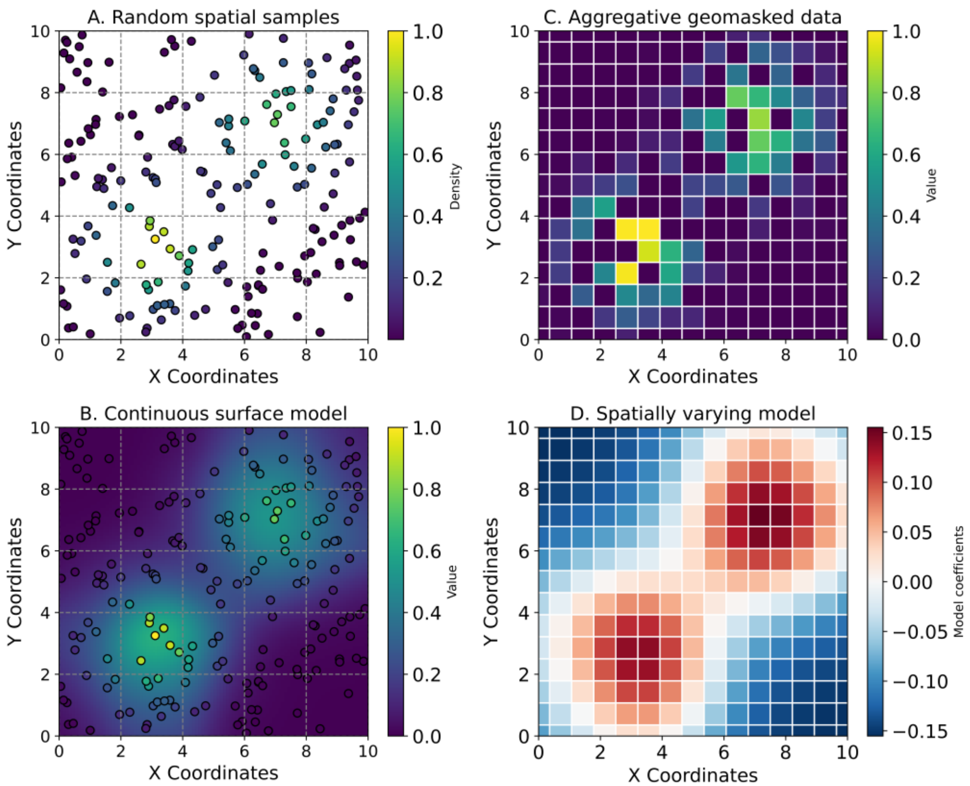


Figure 2: Workflow for exploring and modeling continuous spatial structure from random samples. Random spatial samples (A) were used to reveal underlying spatial variation, followed by the generation of a continuous surface estimates using geostatistical methods (B). Sample values were then interpolated onto a continuous spatial grid to estimate localized means (C), which were further smoothed and analyzed to reveal spatially varying trends centered around the mean (D).

The main sources of these geospatial health data include GPS-tagged household mostly from cross-sectional surveys (e.g., DHS and MICS) (Blanford, 2025), census data, and climatic and agroecological indicators such as rainfall and vegetation indices derived from RS imagery (Eberth et al., 2021; Kirby et al., 2017). Due to data unavailability or temporal gaps in cross-sectional surveys, these remotely sensed indicators continuously and consistently available, complement household-level data, enabling more robust and dynamic analyses (Clark et al., 2024). Therefore, spatial epidemiological studies integrate multi-scale datasets, to fuse individual or household information with broader community and environmental contexts, capturing the complex drivers of population health outcomes (Pradhan et al., 2025). However, challenges remain, including inconsistent spatial resolution of data, missing or incomplete data, and privacy concerns when linking sensitive individual-level information to geographic locations (Seidler et al., 2025). In the same vein, that privacy of surveyed household health data presents a significant challenge, especially when working with multisource, misaligned datasets that require aggregative geomasking (Alahmadi & Moraga, 2025; Hampton et al., 2010). This process can introduce spatial issues, particularly the spatial change of support (Lee et al., 2009), in the forms of the modifiable areal unit problem (MAUP), and ecological fallacy (Openshaw & Taylor, 1979), where conclusions drawn from aggregated spatial scales may fail to reflect individual-level variation. MAUP is a fundamental issue in spatial analysis, where results are inherently dependent on the choice of spatial units, while ecological fallacy refers to the incorrect application of group-level inferences to individuals (Comber & Harris, 2022).

To address these issues, methods such as Moran's I or variogram modeling can be utilized to assess spatial autocorrelation and variability, offering a more accurate understanding of scale effects (Comber et al., 2022). This thesis adopted spatial datasets with multiple geomasked resolutions, to support the ethical and methodological requirements of childhood stunting researches to ensure respondent confidentiality (Burgert-Brucker et al., 2018; Janocha et al., 2021), preserving spatial structure (Bharti et al., 2019), and aligning with the scale of relevant socio-demographic, environmental and agroecological covariates (Alahmadi & Moraga, 2025). The spatially balanced sampling design of these surveyed household data also emerges as a key consideration to ensure representative spatial coverage (Diggle et al., 2010), critical in population health studies such as childhood stunting. This approach as opposed to classical sampling theories, ensure that sampling locations are well spread out across the survey area (Koldasbayeva et al., 2024), avoiding clusters of nearby samples and better capturing the spatial variation in disease risk (Olatunji et al., 2021), to ensuring spatial dependency, and heterogeneity in disease prevalence (Kamgno et al., 2025), critical for validity and reliability of spatial varying model outcomes (Fratesi et al., 2025).

Consequently, population health outcomes, such as childhood stunting, which are rarely distributed uniformly, exhibit two main spatial effects, namely, spatial heterogeneity and spatial dependency (Anselin, 2010), reflecting variations in socio-economic status, access to healthcare, nutrition, sanitation, and education. Spatial heterogeneity, often addressed with spatially varying models, captures non-stationarity across space, while dependence reflects spatial autocorrelation among nearby observations (Anselin, 1988a; Brunsdon et al., 1996). Recognizing these patterns enables the identification of vulnerable populations and the targeting of interventions where they are most needed (Moraga, 2019). National or regional averages often mask localized "hotspots" where children are most vulnerable, highlighting the need for fine-grained spatial analyses (Osgood-Zimmerman et al., 2018). Epidemiologists can assess spatial dependency in health outcomes using global spatial autocorrelation measures such as Global Moran's I (Getis & Ord, 1992), which provide a single test statistic that summarize the spatial patterns for the entire study area (Getis, 2010), and Lagrange Multiplier diagnostics (Anselin, 1988b), which provide targeted test statistics to identify spatial dependence in regression models. While useful as an initial test, such global tests do not detect localized clusters of high or low disease burden (Eberth et al., 2021). Methods to identify localized patterns with consistently high or low disease burden statistics include Anselin's Local Moran's I (LISA) (Anselin, 1995), or Getis-Ord Gi* (Hot Spot Analysis) (Ord & Getis, 1995), pivotal in guiding targeted interventions and resource allocation (Eberth et al., 2021). In the same vein, spatial econometric models account for spatial autocorrelation by explicitly modeling outcomes at a given location as dependent on those of neighboring locations (Anselin, 2010). This is achieved by including spatially lagged terms in the regression model applied to the dependent variable (SLM), explanatory variables (SLX), the error term (SEM), or combinations thereof (Anselin, 1989). However, these models can account for global autocorrelation in the spatial data, and neither method allows the relationships to vary over space (Eberth et al., 2021). Hence, to further describe local variability of these spatial data, a GWR in the form of Geographically Weighted Summary Statistics (GWSS) model, a non-stationary spatial statistics (Brunsdon et al., 2002), provides the spatial local mean, local Pearsons's correlation, under the premise of spatial autocorrelation to bridge gaps of these global descriptive statistical techniques (Gollini et al., 2015). Most importantly, these spatial models provide a descriptive picture of spatial data, significantly enhance the performance of analytical and spatial prediction tasks in population health outcomes (Zhang et al., 2024a). Overall, the inherent spatial structure and dependencies in such health data violate assumptions of randomness (Eberth et al., 2021), necessitating specialized spatially varying modeling approaches to ensure the validity of conclusions and avoid misleading policy implications (Lin & Wen, 2022).

GeoAI in population health: exploring analytical and predictive boundaries

"Essentially, all models are wrong, but some are useful" - George E. P. Box

As noted in the section above, geospatial health data pose significant challenges owing to inherent spatial dependence and heterogeneity, which are central to GeoAI and demand specialized modeling techniques (Goodchild & Li, 2021). Fundamentally, spatial heterogeneity arises in two distinct forms, namely form-based, which reflects differences in observed values across locations, and process-based, refers to the spatial variation in the relationships between variables across different locations (Goodchild, 2004). This latter, process-based heterogeneity, is often manifested as spatial non-stationarity and is captured by various SVC GeoAI frameworks (Anselin & Amaral, 2024). In extreme cases of population health, local spatial regression models are vital for uncovering such spatial heterogeneity that global models can obscure, thereby limiting the effectiveness of spatial analysis (Lin & Wen, 2022). While a global analysis may indicate a consistent positive association between socio-economic deprivation and population health across a region (Kanevski, 2025), local models can identify areas where this relationship reverses due to differing environmental or agroecological conditions, such as variations in food availability or water quality (Deng et al., 2025). Such spatial variation underscores the need for local modeling approaches to accurately capture the complex and heterogeneous drivers of population health outcomes (Maitra et al., 2025). The integration of spatial statistical modeling and GeoAI methods have proven effective (Li, 2020), in addressing spatial epidemiological issues exhibiting spatial heterogeneity due to the complementary strengths each approach offers (Oulaid et al., 2025).

SVC one of modeling backbone in subsymbolic GeoAI methodologies (Finley, 2011; Sahana et al., 2023; Wang et al., 2025), are an extension of traditional regression models that capture how relationships between variables vary across different locations (Comber et al., 2024), producing locally varying regression coefficient estimates, suggesting that the importance of an explanatory variable may vary over space (Brunsdon et al., 1996). The most common SVC formulated under frequentist framework includes GWR (Brunsdon et al., 1996; Fotheringham, 2009), extended in MGWR (Fotheringham & Sachdeva, 2022; Oshan et al., 2019), and other variants to incorporate temporally explicit data (Fotheringham et al., 2015). Under GWR modeling framework, the relationship between predictors and the response varies continuously across geographic space using a local weighting matrix (Comber et al., 2023). This local regression method captures spatial variation by assigning greater weight to nearby observations using a kernel function, such as Gaussian, exponential, or bi-square (Fotheringham et al., 2022), each differing slightly in how weights decline with distance (Fotheringham et al., 2023). This kernel function needs a specified bandwidth that defines the neighborhood of influence around each point (Lu et al., 2017). Depicted in Figures 3 and 4, the bandwidth can be either fixed in terms of a geographic distance, or adaptive by adjusting to a set number of neighboring observations (Fotheringham et al., 2023). Larger bandwidths suggest broader regional trends by incorporating more distant observations, while smaller bandwidths focus on nearby data to emphasize local details (Brunsdon et al., 2002). The selection of the bandwidth is typically estimated by cross-validation (Binbin et al., 2020), by minimizing a corrected AICc, balancing a bias-variance trade-off to capture spatial trends and avoid overfitting (Fotheringham et al., 2023). The general formulation of a GWR model and other SCV and ML hybrids are summarized in Table 1. Typically, GWR explicitly account for spatial nonstationarity resulting in a surface of local goodness-of-fit metrics, and associated measures of uncertainty, enabling detailed understanding of how predictors influence disease risk (Fotheringham et al., 2023). However, the application of GWR may yield biased estimates due to repeated use of observations to estimate local regression parameters, too few observations for each regression that leads to overfitting (Binbin et al., 2020), and multicollinearity among local coefficients (e.g. for instances of local collinearity overlooked by global estimates) (Comber et al., 2024). MGWR addresses these issues by extending the flexibility of the spatial scale, allowing relationships between exposures and outcomes to vary across different spatial scales (Oshan et al., 2019). Within this framework, applied to spatial epidemiology (Eberth et al., 2021), broad-scale predictors (e.g. climate) operate over larger bandwidths, while localised factors (e.g. socio-economic) vary over smaller bandwidths, reflecting spatial scale differences in influence (Wolf et al., 2018). As displayed in Figure 5, MGWR adaptively selecting bandwidths for each variable (Fotheringham et al., 2017), enhances both flexibility and interpretability, making it particularly effective for modeling multiscale population health phenomena (Eberth et al., 2021).

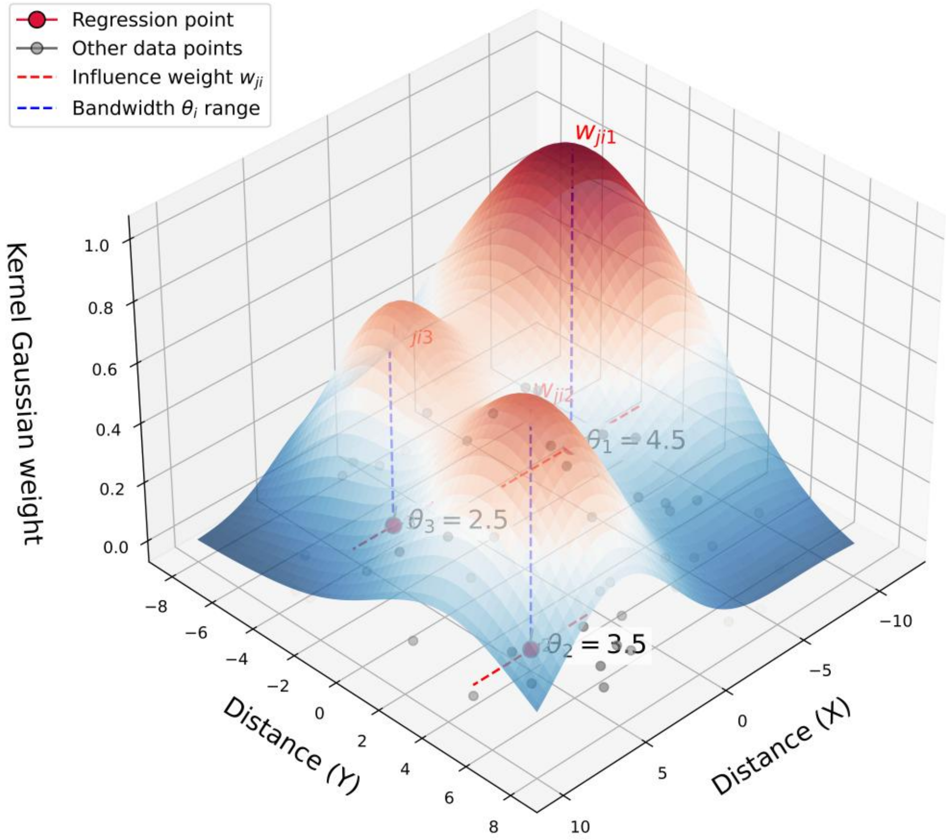


Figure 3: Gaussian kernel functions with varying bandwidth parameters (θ) (Modified from Fang et al., (2021)). Each curve represents the kernel weight as a function of distance, showing how increasing θ broadens the kernel's spread.

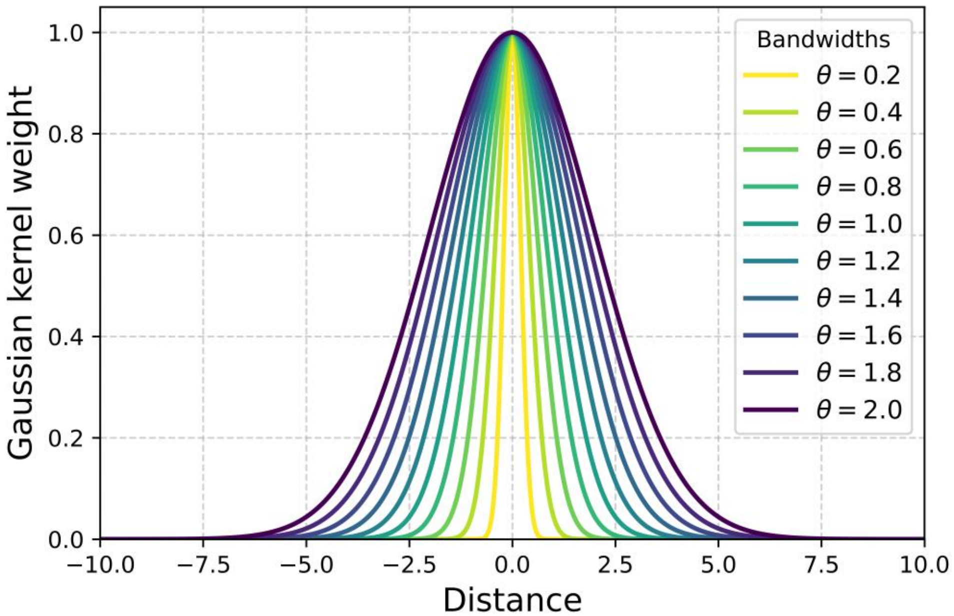


Figure 4: Adaptive Gaussian kernel surfaces illustrating spatial weighting in MGWR (Modified from Fang et al., (2021)). Regression centers vary in bandwidth (θ) and influence (w), defining localized kernel contributions. The surface depicts cumulative influence, with point opacity reflecting proximity-weighted intensity.

Yet, both GWR and MGWR, due to their rudimentary kernel-based parameterisation fail to handle very large, complex, and high-dimensional spatial datasets with complex interactions and non-linearities (Fotheringham et al., 2017). Therefore, MGWR is often viewed as better suited for exploratory rather than inferential analysis (Comber et al., 2024). Conversely, ML methods like decision trees (Breiman, 2001), SVM (Vapnik, 1995), and ANN (Hornik et al., 1989), excel at detecting complex non-linear patterns (Kmoch et al., 2025). These approaches also provide scalability and automation, enabling large-scale, high-resolution mapping that can support timely, data-driven decision-making in public health interventions (Ahmed et al., 2021). Although these ML capture such convoluted non-linear relationships, they often ignore local spatial dependencies to provide interpretable, location-specific coefficients like GWR does (Jiao & Tao, 2025). A hybrid GeoAI approach synergizes GWR with ML techniques to accounts for spatial heterogeneity by producing location-specific coefficients, and captures non-linear relationships and interactions within health data (Credit, 2022). As shown in Figure 6, GWRF blends GWR's spatially varying coefficients with the non-linear predictive power of RF to capture spatial context through localized training on neighborhood data (Georganos et al., 2021; Santos et al., 2019). In doing so, GWRF introduces two new hyperparameters bandwidth and local.w (Georganos & Kalogirou, 2022), in addition to the standard RF hyperparameters number of trees (ntree or n_estimators) and features per split (mtry or max_features) (Breiman, 2001), thereby enhancing model accuracy in spatially heterogeneous contexts. Similarly, GNNWR employs neural attention mechanisms to estimate nonstationary weights (Du et al., 2020), accommodating complex spatial dependencies and interactions of spatial epidemiology data (Kianfar et al., 2025). Typically, GNNWR extends standard GWR by employing a SWNN (Du et al., 2020), where globally estimated OLS coefficients are modulated by neural-network-derived (Wu et al., 2021), non-stationary spatial weights to yield spatially varying regression coefficients and capture both local spatial heterogeneity and non-linear associations (Yin et al., 2024). Another GeoAI approach, SRGCNNs (Zhu et al., 2022b), utilizes graph convolutional networks to model spatial relationships in a way similar to GWR, offering a DL paradigm for spatial regression health data analysis. Motivated by the need for model transparency, these approaches generate localized performance metrics, such as spatially explicit R^2 and residuals (Lin & Wen, 2022), crucial for identifying geographic disparities and context-specific drivers of population health.

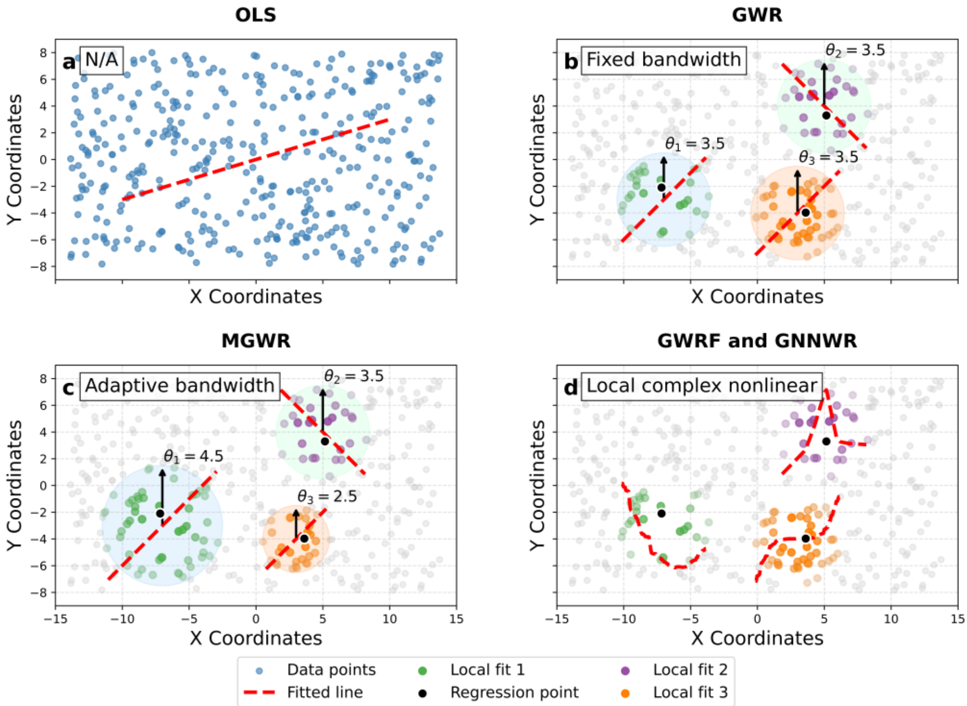


Figure 5: Comparison of global and local regression models for spatially varying relationships. (a) OLS provides a single global fit across all samples. (b) GWR applies fixed-bandwidth local linear models within local neighborhoods. (c) MGWR uses adaptive bandwidths to account for spatial heterogeneity. (d) GWRF and GNNWR fit local nonlinear models using RF and DL. Local bandwidths and regression points illustrate how model structure adapts to spatial variation.

Table 1: Overview of classical spatially varying and GeoAI models in population health

Model	Mathematical formula	Explanation	Role in population health modeling
Global Moran's I	$I = \frac{\sum_i \sum_j w_{ij} z_i . z_j / S_0}{\sum_i z_i^2 / n}$	where z_i is the deviation from the mean, w_{ij} is the spatial weight between locations i and j, S_0 is the sum of all weights, and is n the number of observations.	In the form of incremental spatial autocorrelation, measures global spatial autocorrelation to evaluate scale-dependent clustering at increasing distances using autocorrelation. Reveals optimal spatial scale for modeling spatial effects (Eberth et al., 2021).
Local Moran's I (LISA)	$I_i = \frac{(x_i - \bar{x}) \sum_j w_{ij} (x_j - \bar{x})}{S^2}$	Local Moran's I_i measures the degree of spatial clustering around location i, where $x_i - \bar{x}$ s the deviation from the mean, w_{ij} is the spatial weight between i and j, and S^2 is the global variance of the variable.	Maps local stunting clusters; detects local "hotspots" and "coldspots". Pinpoints target areas for focused intervention (Koldasbayeva et al., 2024).
GWSS	1. GW mean: $m(z_i) = \frac{\Sigma_{j=1} w_{ij} z_j}{\Sigma_{j=1} w_{ij}}$ 2. GW Pearson's correlation: $\rho(z_i, y_i) = \frac{c(z_i, y_i)}{s(z_i) s(y_i)}$ with GW covariance:	1. where $m(z_i)$ is the local mean of variable z at location i, z_j are nearby values, and w_{ij} are spatial weights based on proximity. 2. GW Pearson's correlation z_i, y_i at location i is the ratio of the GW covariance $c(z_i, y_i)$, calculated using spatial weights w_{ij} and local means $m(z_i)$, $m(y_i)$, to the product of the GW standard deviations $s(z_i)$ and $s(y_i)$.	Locally weighted mean, Pearson's using spatial weights (Brunsdon et al., 2002). Visualizes spatially varying data relationships; explores spatial non-stationarity (Gollini et al., 2015).

	$c(z_i, y_i) = \frac{\Sigma_{j=1}^{n} w_{ij}[(z_j - m(z_i))(y_j - m(y_i))]}{\Sigma_{j=1}^{n} w_{ij}}$		
GWR	$y_i = \beta_0(s_i) + \sum_{k=1}^{p} \beta_k(s_i) X_{ik} + \epsilon_i, \quad i = 1, \ldots, n$	where y_i is the response variable at location s_i, modeled as a function of p predictor variables X_{ik} and their spatially varying coefficients $\beta_k(s_i)$. $\beta_0(s_i)$ is the intercept, and ϵ_i is the error term, typically assumed to follow a normal distribution $N(0, \sigma^2)$.	Regression coefficients vary locally to reflect geographic variation (Brunsdon et al., 1996). Models spatially varying associations; maps local drivers of child stunting.
MGWR	$y_i = \beta_0(s_i) + \sum_{k=1}^{p} \beta_k(s_i, h_k) X_{ik} + \epsilon_i, \quad i = 1, \ldots, n$	where h_k is the variable-specific bandwidth for the k-th predictor.	Variable-specific bandwidths for each coefficient, handling multi-scale processes (Fotheringham et al., 2017). Disentangles spatial effects at different scales for each predictor.
GAM	$y = \alpha + \sum_{k=1}^{p} f_k(x_{ik}) + \epsilon_i$	where y is the response variable, α represents the intercept, ϵ_i denotes the residual error, and $f_k(x_{ik})$ refers to the smooth functions, such as splines.	GAM captures complex nonlinear relationships between predictors (x_{ik}) and response (Hastie, 2017; Hastie & Tibshirani, 1995).
GWRF	$y_i = a(u_i, v_i)x_i + \varepsilon$	where $(u_i, v_i)x_i$ is the prediction of the RF model with the spatially weighted random forest (SWRF) calibrated on location i, and u_i, v_i are the coordinates of the centroid of the spatial unit i.	RF predictions made locally, leveraging spatially weighted samples (Georganos et al., 2021; Georganos & Kalogirou, 2022). Captures nonlinear, spatially varying risk associations in childhood stunting.
GNNWR	$y_i = w_0(u_i v_i).\beta_0^{OLS} + \sum_{j=1}^{k} w_j(u_i v_i).\beta_j^{OLS}.X_{ij} + \varepsilon_i$ $i = 1,2, \cdots, n$ with a spatially weighted neural network (SWNN) given by:	where $w_j(u_i v_i)$ represents the non-stationary weight for the model parameter β_j^{OLS}. where $[d^S{}_{i1}, d^S{}_{i1}, \cdots, d^S{}_{in}]^T$ represents the distances from point i to other samples.	ANN regression with spatially adaptive weights for each location (Du et al., 2020). Adapts to complex nonlinear and spatial patterns in determinants (Wu et al., 2021).

	$w(u_i v_i) = SWNN\ ([d^S{}_{i1},\ d^S{}_{i1}, \cdots, d^S{}_{in}]^T)$		
SEL	Ensemble learning (EL): $$\hat{\mu}_{EL} = \sum_{k=1}^{K} \omega_k\ \hat{\mu}_k$$ with the, Universal kriging with external drift (UK): $$Z_{KED}(s_0) = \sum_{i=1}^{n} \omega_i^{KED}\, Z(s_i)$$	where $\hat{\mu}_{EL}$ is the EL, $\hat{\mu}_1$,......., $\hat{\mu}_K$ define potential candidate estimators, ω_k are weights assigned to each estimator. where $Z(s_i)$ are the predicted risk values at sampled locations, ω_i^{KED} are the kriging weights and n is the number of sampled locations.	This hybrid framework combines ensemble ML models (e.g., RF, XGB) with UK to address spatial heterogeneity, whereby providing spatially explicit predictions and risk maps for intervention planning (Hu & Tang, 2023).

While both GWRF and GNNWR models capture the nonlinear and complex interactions between population health and socio-ecological conditions, their applicability remains limited when extrapolating beyond observed data to unsampled contexts. To this end, EL with geostatistics models have attracted many attentions in spatial disease modeling due to their ability to capture complex interaction among features (Ahmed et al., 2021), and for their limited expert knowledge requirements (Davies & Van Der Laan, 2016), allowing for spatial nonlinearity across the regions (Jemeļjanova et al., 2024), for improving disease risk predictions (Moraga, 2019). Built on that framework, recent advancements explored the capabilities of both supervised learning (Zhang et al., 2024b), and geostatistics (Wang et al., 2024c), within the hybrid geostatistical regression frameworks (Fouedjio & Arya, 2024), to respectively predict, downscale, and to explicitly capture local variation in surface earth properties. Similarly, (Hengl et al. (2018) implemented a spatial RF and applied buffer distances on observations in the form of a distance matrix, a method also used by Milà et al. (2024) to deal with model bias and prediction suboptimality, as those from the aspatial RF. On the same basis, Ahn et al. (2020) inspired by Hengl et al. (2018), applied the PCA to the distance vector using only the geographic coordinates for spatial estimation without other covariates. Inspired by Lundberg & Lee (2017), Li (2024) introduced GeoShapley combined spatial data with coordinates to measure spatial effects in ML models for advancing the model interpretability. Although these models offer predictive power, coordinate-based frameworks, alone or with covariates, provide limited interpretability for disease-risk decision-making (Jemeļjanova et al., 2024), since the influence of location cannot be directly assessed (Liu, 2024). Being said, they are still issues related to the interpretability and transferability of these methods which are conditioned by their rigorous parameterization and experiments made on well generalized real cases with enough large structured datasets (Zhao et al., 2024). This study advances the population health risk prediction by envisioning a hybrid ML and geostastical classification approach as an alternative to the existing similar studies to refine the explainability critical for localized public health decision making. Overall, this robustness rooted in these models to capture complex, non-linear interactions and handle high-dimensional data through learned relationships instead of predefined kernels, makes these hybrid GeoAI methodologies ideal for our study (Luo et al., 2021), providing a powerful tool for targeted interventions and evidence-based public health planning. With this study, we provide critical insights into selecting appropriate spatially varying GeoAI methodologies for spatial modelling tasks, ultimately improving the reliability of localized descriptive, analytical and predictive models of childhood stunting.

Explainability and interpretability in population health outcomes

Understanding why a model produces specific predictions is essential for designing effective public health interventions (Loh et al., 2022). For instance, a variety of classical ML tools, such as GLM (Nelder & Wedderburn, 1972), GAM (Hastie & Tibshirani, 1995), and decision tree variable importance and partial dependence plots (Ahmed et al., 2021), considered as early incarnation of IML, has been widely used in scientific studies (Ahmed et al., 2025). Traditional approaches output single regression coefficients, and global variable importance metrics, which provide insight into predictor effects but have notable limitations (Molnar & Freiesleben, 2024). Coefficients assume linearity and global stationarity, while standard variable importance measures indicate which predictors matter most without clarifying directionality, interactions, or context-specific effects, often obscuring nuanced relationships in complex, non-linear datasets (Jiang et al., 2024). To this end, analytical spatially varying GeoAI models such GWRF offer local variable importance, and GNNWR produces location-specific regression coefficients, providing nuanced insights into how predictors vary across space (Santos et al., 2019). These models visualize these local coefficients surfaces, and adhering to a single-color bar ensures consistency across the maps, allowing for direct comparisons of patterns and magnitude between variables (Stofer, 2016). This broader visualization strategy enables users and analysts to assess both trends and reliability across multiple variables. Complimenting these, XAI techniques, such as SHAP, overcome these gaps by quantifying both global and local contributions of predictors, capturing non-linear interactions and spatially varying effects (Lundberg et al., 2020; Lundberg & Lee, 2017). In population health, SHAP can reveal how socio-economic, environmental, and spatial factors drive outcomes like childhood stunting, and, when combined with spatial visualization, highlights high-impact drivers across regions (Nduwayezu et al., 2025b). Inspired by Lundberg & Lee (2017), Li (2024) introduced GeoShapley to measure spatial effects in ML models for advancing the interpretability of these spatial models. These recent advances in spatial SHAP further allow visualization of geographic gradients in feature effects, highlighting spatially varying drivers of outcomes such as childhood stunting. This approach bridges the gap from traditional linear or tree-based models to advanced, interpretable predictive frameworks capable of guiding targeted public health interventions. Overall, this thesis generates fine-scale, interpretable risk surfaces that can support targeted nutrition interventions in data-constrained settings, and enable a more demystified modeling, providing interpretable insights into the drivers of child undernutrition, enabling actionable intelligence for decision-makers.

Aim and objectives

The overall aim of this thesis is to develop, test, and apply methodologies for the detection, explanation, and prediction of the spatial heterogeneity in population health datasets. The thesis provides information on the potentials and challenges associated with integrating multi-source geospatial, household survey, and remotely sensed data to understand local determinants of stunting among children, map small-area prevalence in Rwanda, and support geographically targeted public health policy interventions to the areas of the most needs. The following three specific objectives were pursued to accomplish this aim:

1. To quantify and statistically analyze the spatial patterns of child stunting prevalence using spatially varying and clustering approaches across multiple scales **(Paper I).**
2. Identify and interpret localized nonlinear determinants of childhood stunting by applying hybrid GeoAI approaches that integrate multisource determinants **(Papers II & III).**
3. Develop and evaluate predictive GeoAI models that generalize to unsampled locations using spatial ensemble methods to generate SAEs of childhood stunting risk **(Paper IV).**

To achieve these objectives, this study addresses the following questions:

- How can spatial patterns in childhood stunting be statistically and visually quantified?
- How can nonlinear relationships between localized stunting determinants and spatial contexts inform targeted health policy models; and particularly, how effective are climatic and agroecological indicators in analyzing childhood stunting?
- How predictive inference models generalize to produce probabilistic risk estimates of childhood stunting, enabling localized interventions?

Figure 6 illustrates the overall structure of the thesis and how the contributing papers are integrated into the final synthesis.

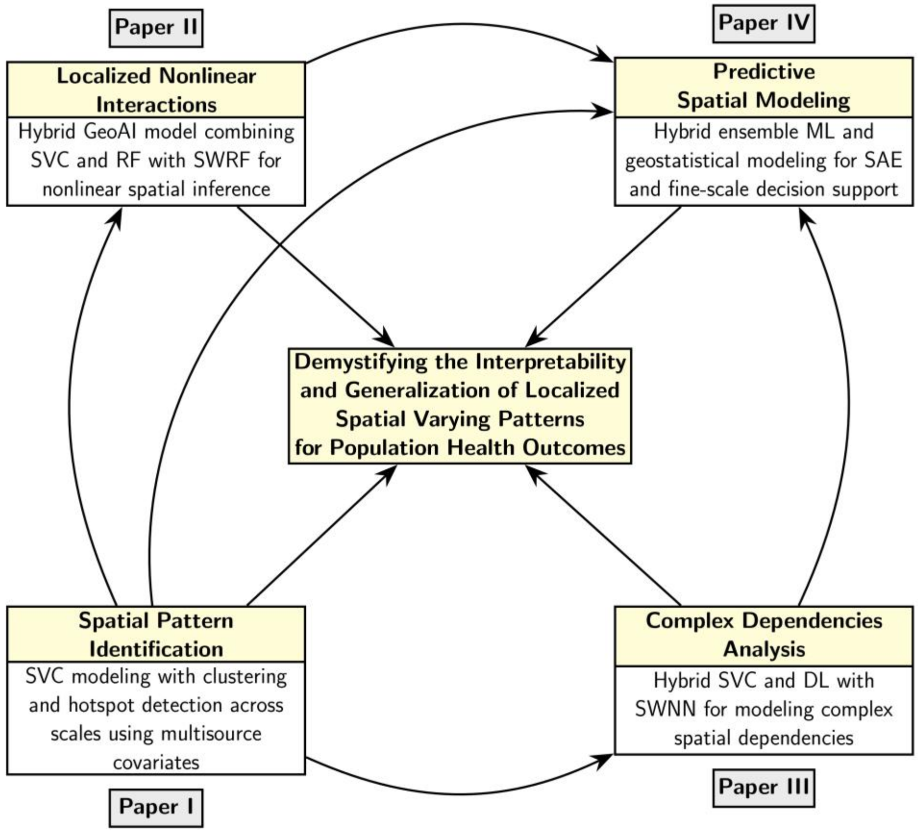


Figure 6: Dissertation structure showing linkages among four papers on spatial modelling (**Paper I**), local determinants (**Papers II** & **III**), and predictive estimation of population health outcomes (**Paper IV**). Arrows denote conceptual and methodological contributions across papers.

Materials and methods

Overview and study area

Rwanda's Northern Province historically exhibiting higher rates of child stunting relative to other regions (NISR et al., 2021), served as the primary study area for this research (**Paper I**, **III**, and **IV**). This region is typified by rugged terrain, high population density, and predominantly smallholder agricultural systems, which are acutely susceptible to both agroecological and socioeconomic stressors. These characteristics make it a critical landscape for investigating how local environmental and social determinants interact to shape child nutritional outcomes. To complement this regional focus, national-scale patterns were assessed using data from the 2019-2020 RDHS (**Paper II**). The RDHS provided standardized anthropometric data for children under five, alongside detailed maternal, household, and demographic variables. To safeguard respondent confidentiality during data curation, survey cluster coordinates were randomly displaced by up to 2 km in urban areas and 5 km in rural areas (Burgert-Brucker et al., 2018; Janocha et al., 2021). To address the resulting spatial uncertainty, data were aggregated at sector jurisdictional and gridded resolutions (1 km^2 and 5 km^2), minimizing positional error while maintaining spatial reliability for subsequent analyses.

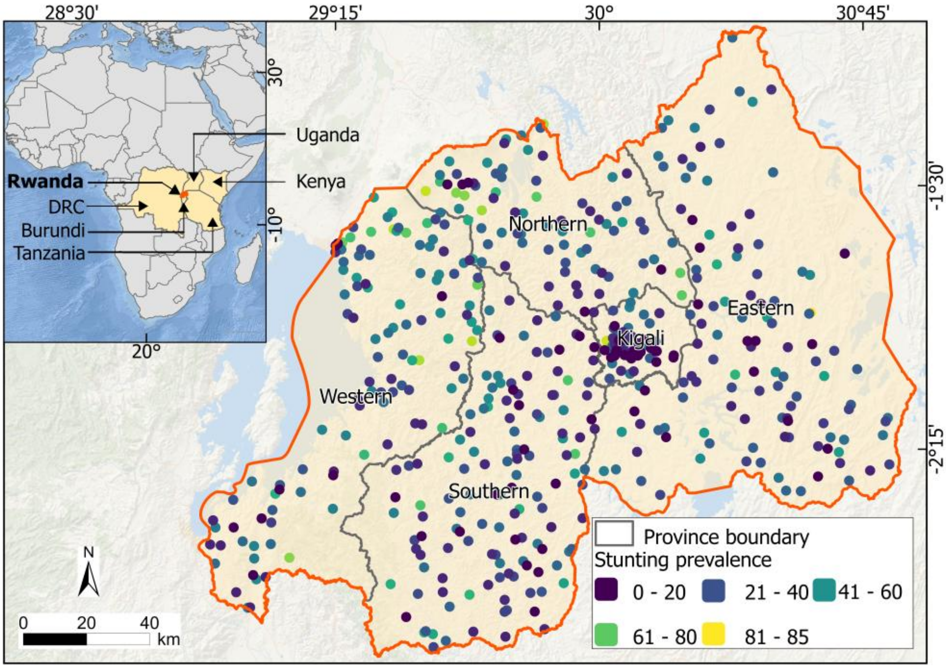


Figure 7: Geographical location of Rwanda (the left upper figure), and stunting prevalence at household cluster level in Rwanda (the main figure). These geo-located data formed the basis for the modeled interpolated surface, and average stunting prevalence rate at the sector level, a jurisdictional zone in Rwanda used in **Paper II**.

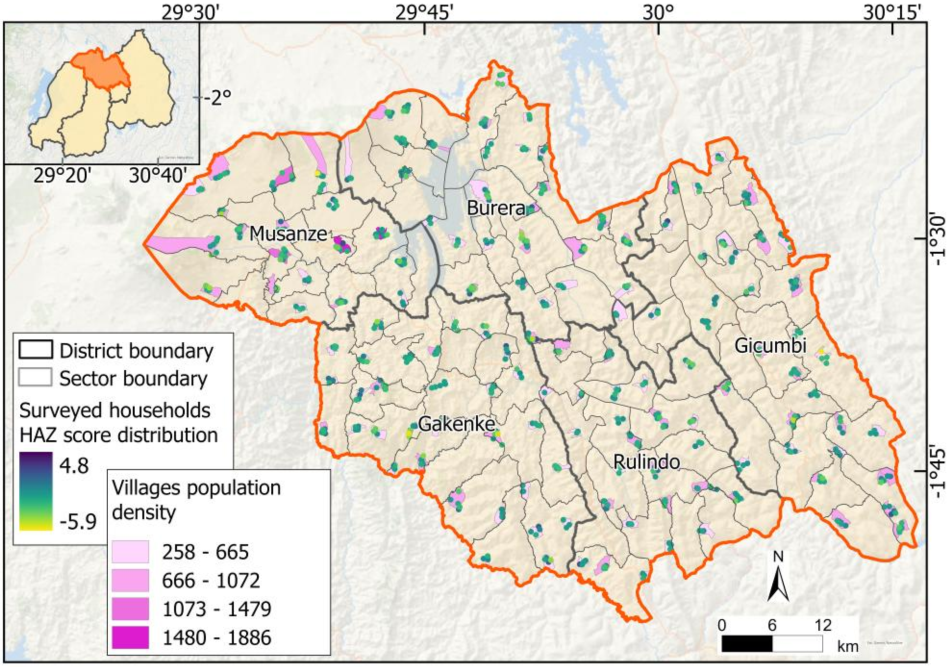


Figure 8: The location map of the Northern Province in Rwanda (the right upper figure), and the spatial distribution of the sampled households in the study area (the main figure). These childhood HAZ prevalence was initially collected at the household level (**Paper III**), then aggregation-based geomasking to generate continuous areal HAZ prevalence at the sector level (**Paper I**), further gridded at 2.2 × 2.2 km resolution, and finally dichotomized into stunted and non-stunted categories for use in **Paper IV**.

Paper I

The aim of this study was to explore the spatial heterogeneity of stunting (low HAZ) among children aged between one month and three years in Rwanda's Northern Province, and to analyze how socioeconomic, agroecological, and climatic factors contribute to its variability. We utilized cross-sectional data from 601 households, geomasked through aggregation at the local administrative level, and applied SVC models and interpretable ML techniques. Built on balanced spatial sampling (Diggle et al., 2010; Olatunji et al., 2021), and WHO anthropometric guidelines (WHO, 2006), we captured different child growth failure metrics and related socioeconomic attributes in the northern province. To do this, two-stage cluster sampling was employed (Katz, 1995), randomly selecting 137 from 2,744 villages using a spatial grid to ensure geographic representativeness, followed by systematic household sampling proportional to village population density. The sample size was calculated using the standard formula for prevalence studies (Carlin & Hocking, 1999), assuming a 40.5% stunting prevalence (NISR et al., 2021), 95% confidence, 5%

margin of error, and a design effect of 1.5 (Katz, 1995), yielding 553 households. Adjusting for 10% non-response (Jensen et al., 2022; Prince, 2012), the final sample size was 615 households. Anthropometric measurements of children and mothers were collected following standard protocols. After excluding 14 children with missing anthropometric data, 601 samples were included in the analysis. Stunting was the primary outcome, with predictors spanning seven categories: socio-demographic factors, health, nutrition, childcare, violence, livestock, and healthcare access. Motivated by existing health literature, the household survey data were integrated with additional agroecological (soil fertility, slope, elevation), and climate factors (rainfall, LST, NDVI), given their significant epidemiological pathways that influence multiple dimensions of food security and contribute to chronic undernutrition (Aheto & Dagne, 2021; Baffour et al., 2023; Khaki et al., 2024; Nduwayezu et al., 2024). Grid-based aggregation was used to convert household-level point data into spatially continuous aggregate data Gribov & Krivoruchko (2020), while protecting respondent confidentiality. Spatial incremental autocorrelation was assessed using Moran's I and Getis-Ord Gi* to identify clusters of stunting hotspots and coldspots (ESRI, 2024). We fitted multiple SVC models, complemented by GAM as IML (Hastie & Tibshirani, 1995). First, we fitted OLS regression, which served as a global benchmark but was limited by its assumption of spatial stationarity. To overcome this, we further fitted a GWR model (Brunsdon et al., 1996), which estimates local coefficients by weighting observations according to geographic proximity. To reflect local variations in stunting risk factors, we used an adaptive spatial kernel, and the optimal bandwidth was selected through minimization of the corrected AICc, to ensure the model robustness as described in Table 1. Building on this, we calibrated a MGWR model (Fotheringham et al., 2023), by first selecting an adaptive spatial kernel to capture local variations in the relationships between predictors and childhood stunting prevalence. This process ensures that the model captures localized effects without overfitting. We assessed the model performance with localized R^2 to measure spatially explicit explanatory power and Local Moran's I to detect spatial autocorrelation in residuals, ensuring robust and interpretable spatial predictions. Moreover, the interpretation focused on model coefficients the nonlinear effects of predictors. Finally, to capture nonlinearities in predictor outcome relationships, critical for localized public heath decision making, the analysis integrated GAM (Hastie, 2017). These models were trained with thin plate regression splines, and were used to reveal nonlinear effects such as U-shaped or bell-shaped responses to slope, NDVI, and soil fertility. This paper integrated these SVC models to provide a robust framework capable of uncovering both the spatial heterogeneity and the nonlinear dynamics of stunting determinants in the Northern Province of Rwanda.

Paper II

The aim of this study was to investigate how different socioeconomic, maternal health, and environmental factors affect the risk of childhood stunting in Rwanda, and to examine how these relationships vary across space. We utilized Bayesian-modeled surface prevalence data from the 2019-2020 RDHS (NISR et al., 2021), in combination with geospatial covariates and hybrid RF models to quantify the importance of localized determinants associated with childhood stunting. Stunting prevalence was derived from modeled surface rasters (5×5 km resolution), which interpolated RDHS cluster-level measurements using Bayesian geostatistical methods (Burgert-Brucker et al., 2018). These surfaces provided a continuous outcome variable, expressed as the percentage of children under five who were stunted. Exposure variables were selected through a literature review and included socioeconomic/demographic, maternal health, and environmental indicators known to influence childhood nutrition, such as parental literacy, women's anemia status, use of improved water sources, open defecation, insecticide-treated net use, antenatal care visits, and delivery care quality (Boah et al., 2022; Kismul et al., 2017; Pattnaik et al., 2021; Tangena et al., 2023; Uwiringiyimana et al., 2019; Vollmer et al., 2017). All predictors were scaled to 0-1 and aggregated to the sector jurisdictional level to account for DHS cluster displacement. The core modeling framework in this study contrasted the global RF with a GWRF to assess fine-scale heterogeneity in stunting prevalence. We first trained the global RF as a benchmark model (Breiman, 2001). However, because the global RF does not account for spatial dependence as described in methods section, the study advanced to GWRF (Georganos & Kalogirou, 2022), which integrates spatial kernels into the RF algorithm by assigning higher weights to geographically proximate observations. This hybrid GeoAI builds separate sub-models for each spatial unit, incorporating only its neighboring units, which were defined either by a distance threshold (fixed kernel) or by the number of nearest neighbors (adaptive kernel). Rooted in the randomization inherent to training a RF (Breiman, 2001), this hybrid approach allowed the model to simultaneously capture nonlinear predictor effects and local spatial variation while mitigating overfitting risks. To prevent overfitting prior to fitting the RF and GWRF models, we tested multiple hyperparameter combinations using 10-fold cross-validation (Santos et al., 2019), selecting the optimal set based on performance with the adaptive kernel. For comparison, we applied the same parameters consistently to both the RF and GWRF models. Specifically, we used GWRF for further inferences due to its local implications and interpretations, such as spatially varying relationships, local feature importance, and targeted health decision-making. Finally, to ensure interpretability, the study moved from GAM in **Paper I**, to employ advanced IML tools such as global and interaction partial dependence plots (Molnar, 2022), to identify thresholds in predictors and revealing local interaction effects between childhood stunting with its factors. The methodological frameworks in **Paper II**, demonstrated how hybrid GeoAI approaches improve predictive performance while preserving policy-relevant interpretability.

Paper III

This study introduced a spatially varying DL model to assess the spatial heterogeneity of childhood stunting determinants among children under three in Northern Rwanda. We cross-sectionally analyzed point household survey data from 601 households in Northern Rwanda to capture the fine-scale determinants of stunting risk. The dependent variable was the HAZ, and predictor features were child health and care (age, sex, birthweight, breastfeeding practices, childcare), maternal factors (BMI, social support), and household conditions (sanitation, electricity, kitchen gardens, farmland, milk consumption). Missing data were addressed using ML-based imputation (Azur et al., 2011), and feature selection combined RF importance, ridge regression, and forward selection. The methodological framework of this paper extended GWR and MGWR to incorporate neural network architectures through the GNNWR model. We first fitted GWR and MGWR as baselines, each calibrated with adaptive spatial kernels and bandwidths optimized through AICc minimization. As described in Table 1, while these SVC models capture spatial nonstationarity (Fotheringham et al., 2017), their reliance on predefined kernel functions limits their ability to estimate complex nonlinearities and high-dimensional interactions among variables. To address this, we implemented GNNWR by embedding an ANN within the GWR framework (Du et al., 2020), using a SWNN to compute nonstationary weights, thereby enabling the estimation of local regression coefficients while simultaneously capturing nonlinear dependencies among predictors (Yin et al., 2024). We trained the GNNWR by iteratively optimizing neural network hyperparameters including four hidden layers, the PReLU activation function, the SGD algorithm, a dropout rate of 0.2, and a learning rate of 0.01. The GNNWR achieved higher explanatory power ($R^2 \approx 0.66$ on training data) compared to GWR and MGWR, and uncovered localized, nonlinear relationships between child health, maternal care, and household conditions, revealing both broad spatial trends and fine-scale heterogeneity in stunting risk determinants. In this paper, the integration of neural networks with SVC demonstrated the potential of novel subsymbolic GeoAI methodologies to reveal complex, spatially varying determinants to provide actionable insights for localized interventions against childhood stunting.

Paper IV

The aim of this study was to examine how agroecological, environmental, and socioeconomic conditions influence small-area childhood stunting risk in Rwanda's Northern Province, and how these relationships vary across spatial scales. We utilized cross-sectional household survey data in **Paper I** and **III**, combined with satellite-derived and agroecological indicators, to implement a multilevel SAE framework using SEL and XAI. The study targeted children aged between one month and three years, with childhood stunting assessed using HAZ. Response data were spatialized to multiple resolutions, including 1 km² and 5 km² grid cells, village, and sector levels. Preprocessing accounted for spatial dependence and scale effects using Moran's I and Getis-Ord Gi statistics (Tiwari et al., 2023), while incremental spatial autocorrelation was applied to detect the distance at which spatial clustering peaked, which occurred at 8 km (ESRI, 2024). Stunting prevalence values were dichotomized into stunted and non-stunted categories, yielding an imbalanced dataset (Koldasbayeva et al., 2024). Explanatory variables included elevation, slope, soil fertility, rainfall, NDVI, livestock density, urbanicity, toilet type, and water access. These were derived from RS datasets aggregated using zonal statistics. These RS covariates were selected based on their hypothesized importance for estimating socioeconomic conditions and their availability as shown to correlate with childhood stunting in previous work (**Paper I**, **II &III**). We introduced a so called, SEL, a novel two-stage modeling framework that integrates stack ensemble and geostatistical methods for spatial prediction of childhood stunting risk. In the first stage, an EL model was trained to estimate stunting prevalence at sampled locations by capturing convoluted, non-linear relationships with environmental, agroecological, and health-related covariates using SHAP. By considering the predicted 5 km^2 grid at sampled points as realizations of an underlying continuous spatial process literature (Burgert-Brucker et al., 2018; Janocha et al., 2021), these predictions are then incorporated as external drift in a UK model (Wiedemann et al., 2023), which accounts for spatial autocorrelation to produce probabilistic maps of stunting risk. To address potential scale effects, maps were generated at multiple spatial resolutions, enabling a more robust and flexible assessment of spatial variation in stunting prevalence across the study region. Through a rigorous evaluation of different stages of the population modeling pipeline data input, model selection, and outcome assessment our findings underscore the efficacy of the model, providing robust estimates and quantifying uncertainty. RF consistently outperformed individual learners, with SEL providing enhanced discriminability between stunting and non-stunting probabilities. SHAP values were used for model interpretation (Lundberg & Lee, 2017). The multilevel SEL framework in this paper demonstrated the value of capturing scale effects and spatial non-stationarity in stunting risk estimation.

Results and discussion

Spatial patterns of stunting among children: statistical and visual interpretability

All four papers consistently revealed that child stunting in Rwanda is characterized by fine-scale spatial clustering. Across all four studies, spatial analysis consistently revealed that stunting is not uniformly distributed, but follows distinct spatial patterns (i.e., strong clustering and local heterogeneity) across provinces, districts, and even sectors. At the national level, exploratory spatial statistics (e.g., GWSS, Moran's I, Getis-Ord Gi*) identified significant hotspots in Rwanda's Northern and Western provinces. **Paper II** identified strong geographic heterogeneity, with RDHS cluster data showing stunting hotspots concentrated in the Northern and Western provinces, consistent with the broader pattern in **Paper I**, but further revealing small, localized clusters that the RDHS data alone obscures, underscoring the persistence of childhood stunting across ages. For that, Using Moran's I and Getis-Ord Gi analyses, **Paper I** identified marked clustering of stunting prevalence, with persistent hotspots and coldspots concentrated in Musanze, Gakenke, and Gicumbi districts, highlighting the need for spatially targeted, scale-sensitive interventions. Similarly, in **Paper IV** using incremental spatial autocorrelation, across all spatial supports, statistically significant positive spatial autocorrelation was detected, with the strength of clustering peaking at finer scales (household and 1 km^2 grid) and attenuating at coarser aggregations (5 km^2 grid). Notably, the detection of clusters persisted across multiple spatial scales, emphasizing the robustness of local patterns despite varying administrative resolutions (Figure 9).

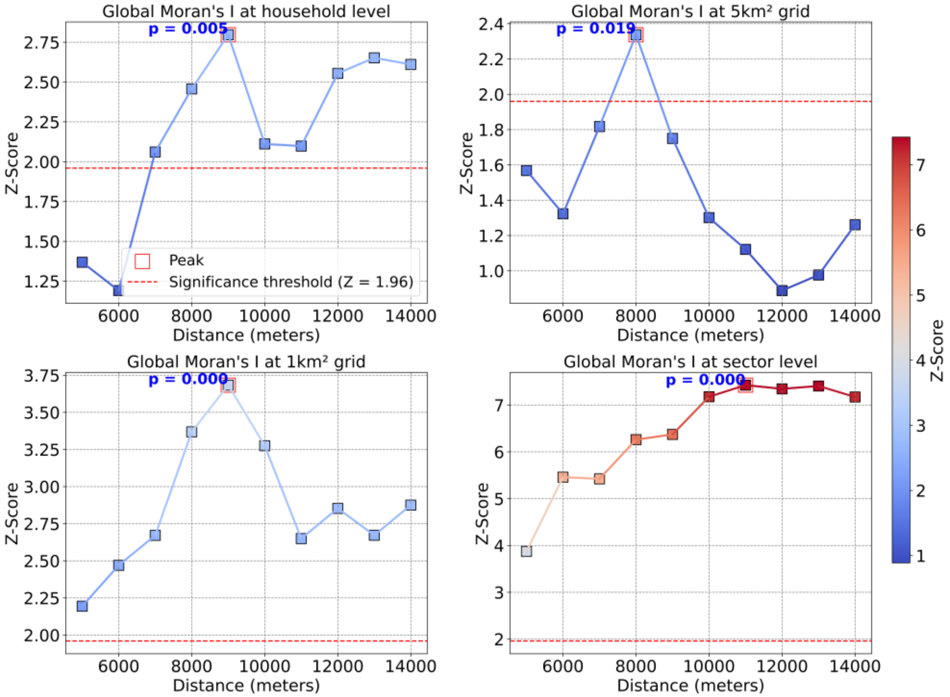


Figure 9: Global Moran's I Z-scores at household, 1km² grid, 5km² grid, and sector jurisdictional levels. Each line segment is colored based on the Z-score intensity. Square markers represent distance intervals. The corresponding p-value at which the peak occurs is also displayed in bold red. The dashed red line indicates the statistical significance threshold (Z = 1.96).

Paper I also computed a local Moran's I to examine the residuals of SVC models; even after accounting for predictors, Moran's I values indicated that spatial dependence persisted within OLS model, reflecting unobserved localized processes (Figure 10).

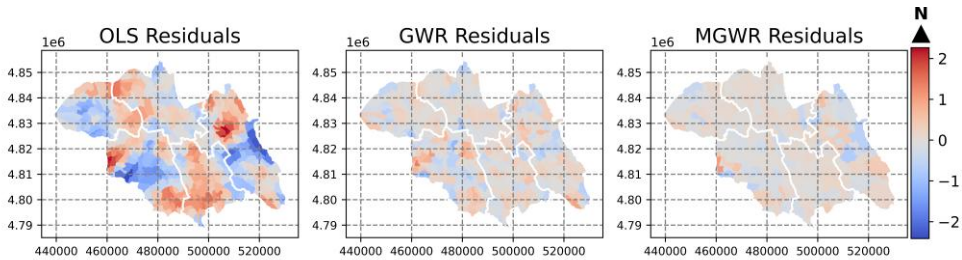


Figure 10: Spatial distribution of the standardized models' residuals; OLS, GWR, and MGWR. A smaller standardized error indicates higher model performance.

The visualization of predictive surfaces in **Paper IV**, added another dimension. Predicted prevalence surfaces derived from SEL provided fine-resolution maps at 1-5 km², capturing stunting gradients across Northern Province (Figure 15).

Together, these results suggest that statistical tests and visual mapping mutually reinforce each other: clustering can be formally confirmed, while spatial visualization makes the results more interpretable for policy makers and stakeholders (Stofer, 2024), offering a powerful entry point for geographically targeted interventions (Stofer, 2016), while also providing a benchmark for evaluating model robustness.

Localized determinants of child stunting and their policy implications

The importance of localized non-linear and complex determinants of child stunting was most evident in **Papers I**, **II**, and **III**. **Papers I**, **II**, and **III** converge on the conclusion that child age, maternal, and household factors dominate as determinants of stunting, though their strength and direction vary geographically, consistent with existing literature. **Paper I** confirmed that stunting risk increased with elevation, rainfall, but their impact was magnified in districts with poor child care practices (number of days the child was left alone) (Yoneshiro et al., 2025). NDVI and slope showed U-shaped effects (Figure 11), meaning that both very low and very high values were detrimental, indicating nonlinear ecological influences (Nduwayezu et al., 2024). This further reflects the dual burden of low agricultural productivity in degraded landscapes and difficult farming conditions in steep, high-rainfall areas (Nduwayezu et al., 2025b).

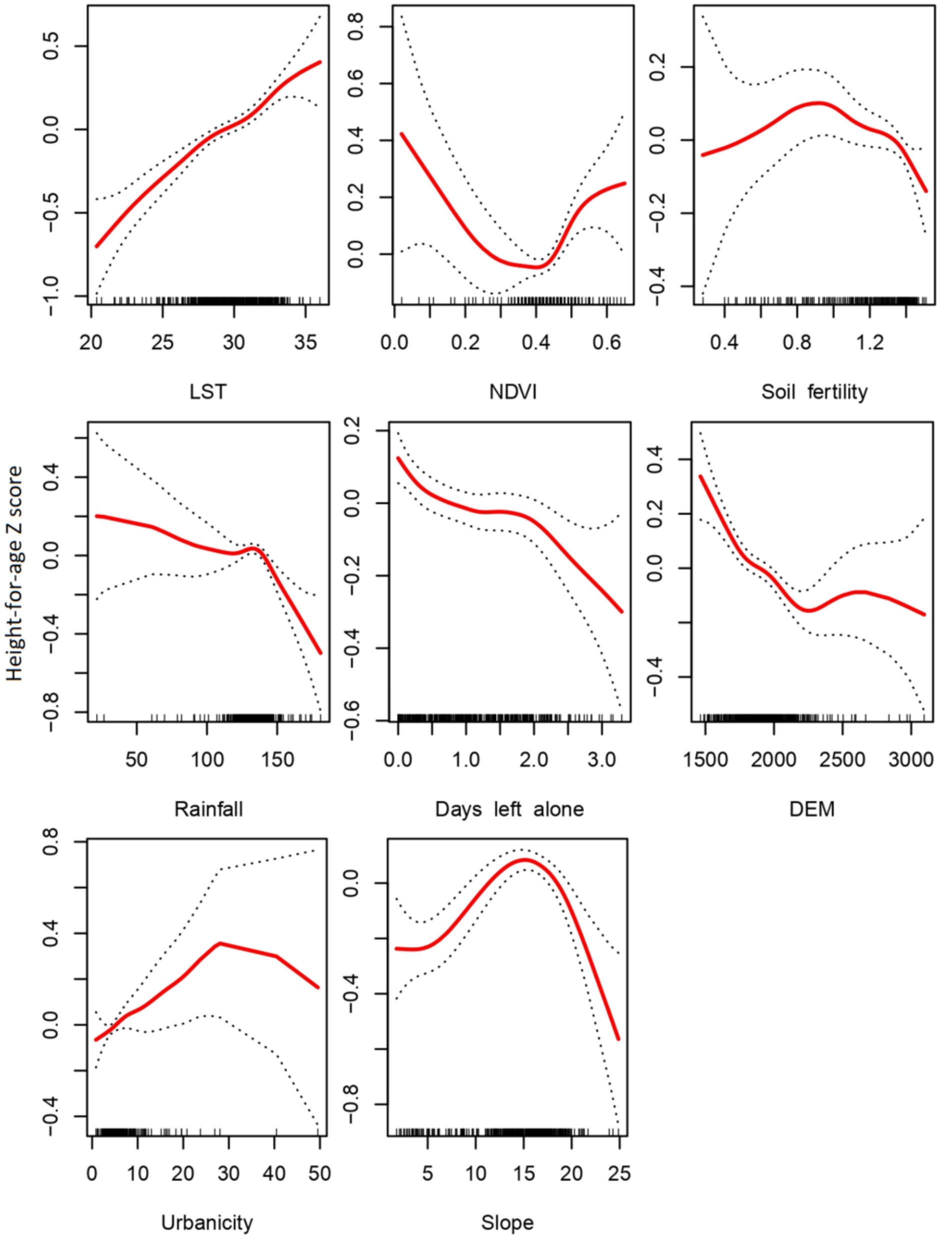


Figure 11: GAM plots of the nonlinear effects of continuous predictors on height-for-age Z score, with the center red lines representing the mean estimated effects, and the bands between upper and lower dashed lines indicating the 95% confidence intervals. A declining curve signifies increasing stunting prevalence, and a rising curve signifies lowering stunting prevalence.

MGWR reveals pronounced spatial heterogeneity in the drivers of child HAZ across Northern Province of Rwanda, complementing GAM's global nonlinear trends. Both models consistently identified overall trends, such as negative impacts of elevation and soil fertility Figure 12. While GAM consistently captured overall predictor HAZ relationships e.g., positive LST, nonlinear NDVI, and negative elevation effects MGWR revealed local variability hidden in these global trends, such as regions where rainfall decreased HAZ or urbanicity had negative local effects, complementing GAMs by adding spatial context essential for targeted interventions (Nduwayezu et al., 2024). Importantly, MGWR reveals local-scale variations, enabling spatial epidemiology to target interventions with geographic precision and context specificity. Although MGWR alone can robustly explain localized stunting risk, combining it with GAM ensures broad applicability (Stofer, 2024), where health professionals may find GAM's global trends easier to interpret than spatially complex MGWR maps, highlighting the importance of using intuitive color schemes to facilitate map interpretation (Stofer, 2016).

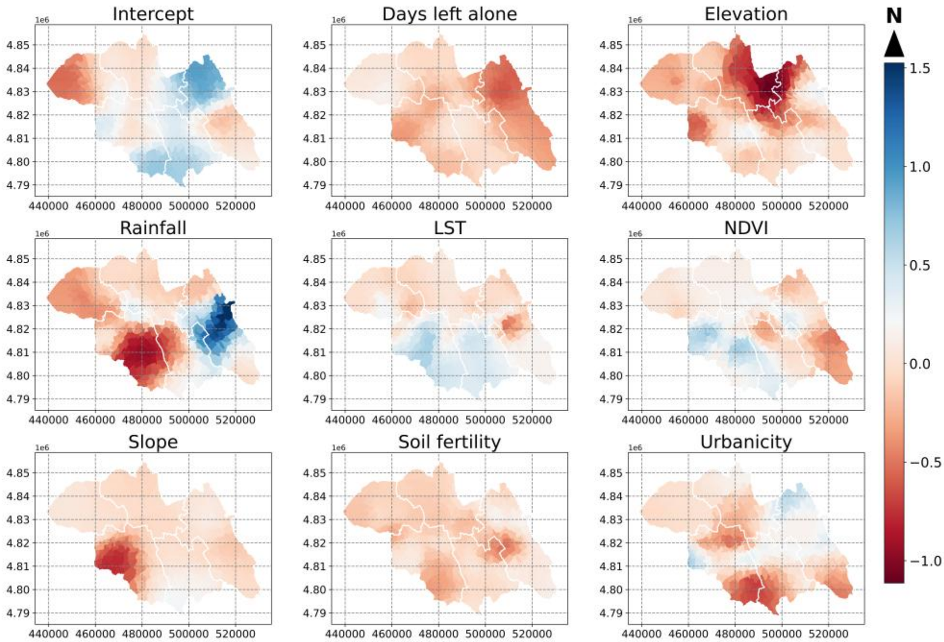


Figure 12: Spatial distribution of coefficients of the MGWR model. A positive sign denotes that the explanatory variable increases the probability of the outcome, whereas a negative sign indicates that the variable lowers the likelihood of the outcome.

Paper II provided further additional nuance by uncovering nonlinear and geographically specific effects. This **paper II** demonstrated that maternal literacy, antenatal care, sanitation, and facility-based deliveries were strongly linked to lower stunting prevalence, with partial dependency analyses (Molnar, 2022), showing threshold effects for example, child growth faltering marked reductions in risk once antenatal visits exceeded 50% or access to clean water surpassed 70%. **Paper II** also found that maternal literacy, antenatal care visits, and sanitation practices exerted strong but spatially varying effects, with partial dependency analysis showing threshold improvements in stunting risk once clean water access and facility-based deliveries surpassed certain levels (Figure 13).

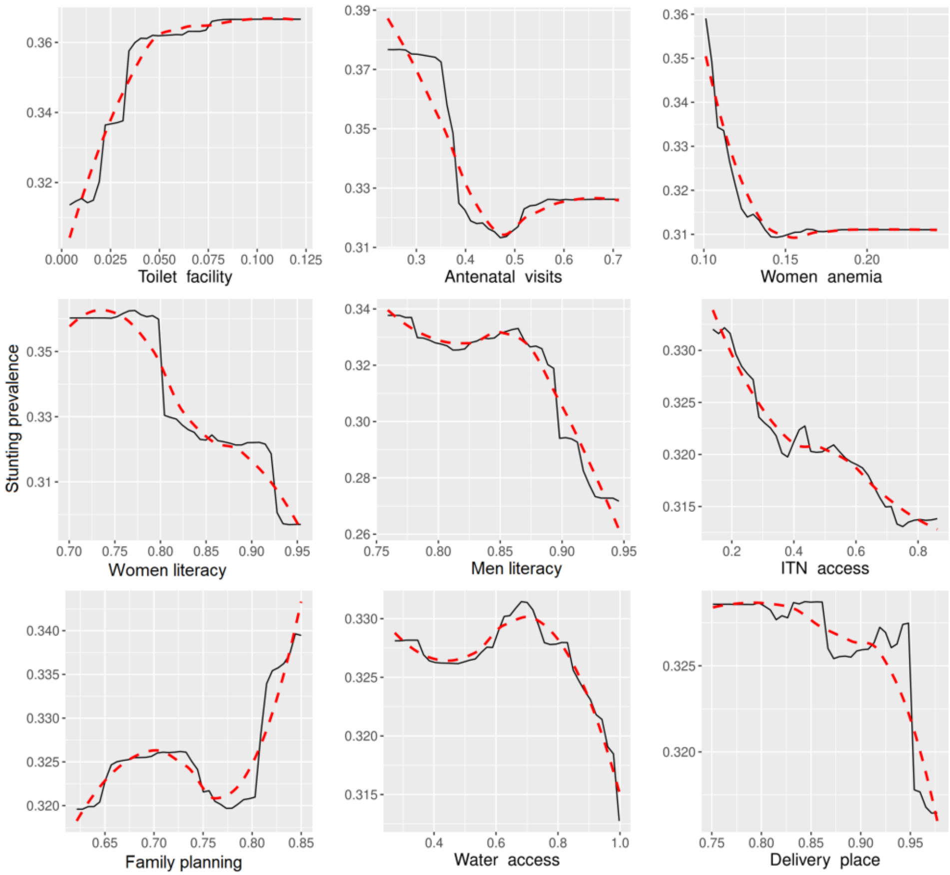


Figure 13: Partial dependence plots and smoothed response curves for the explanatory variables, selected using global RF. Partial plots show the dependence of the probability of stunting prevalence occurrence on one predictor variable after averaging out the effects of all other predictor variables in the model. The horizontal axis represents the values of the predictor, and the vertical axis represents the marginal effects of a predictor on the predicted target (the risk of child stunting).

The GWRF local variable importance maps in **Paper II**, also reveal marked spatial variation in the determinants of child stunting in Rwanda, underscoring that the influence of these factors varies considerably across regions. For example, in some northern and western sectors, stunting risk is most shaped by WASH-related barriers, while the east and south are more affected by maternal health service utilization and education, and some regions show high importance for reproductive health and malaria prevention (Figure 14). These fine-scale insights suggest targeted, context-specific interventions such as prioritizing improved sanitation where WASH factors predominate or boosting health education and antenatal care where literacy and maternal health are key can yield greater reductions in stunting than uniform, nation-level programs (Nduwayezu et al., 2025a). Prior studies confirm that stunting's determinants (poverty, water, education, health access) exhibit strong spatial variability, supporting the use of these maps for epidemiological decision making (Tamir et al., 2024).

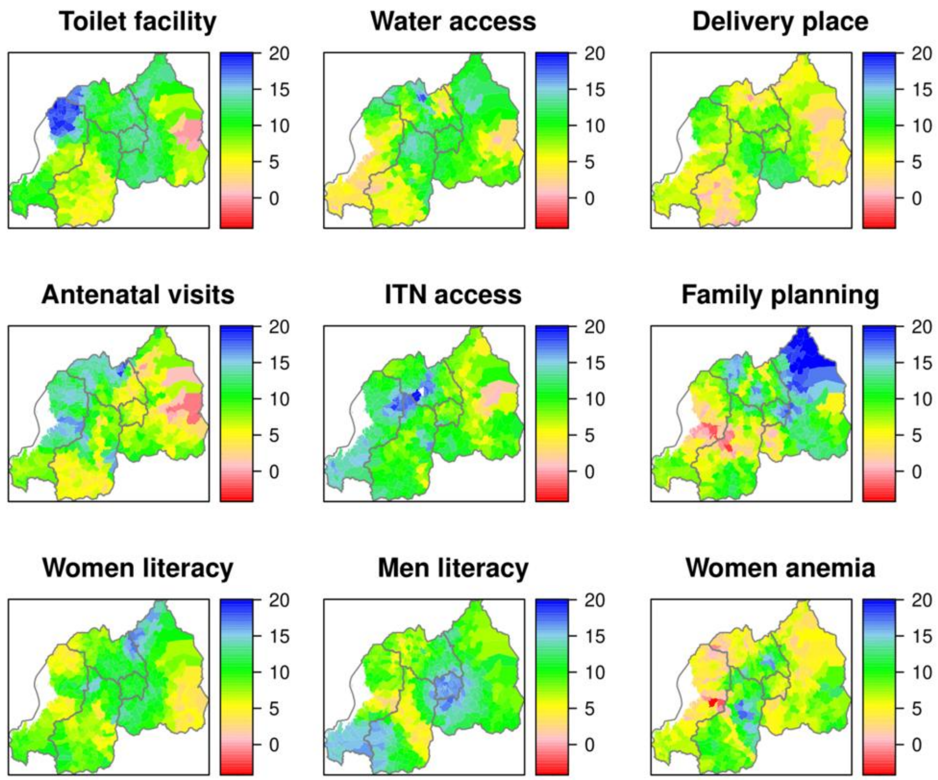


Figure 14: The spatial variation of the local feature importance from a GWRF model for child stunting in Rwanda, showing how the contribution of various health, demographic, and resource factors varies across regions. Spatial variation in %; local feature importance according to the mean decrease accuracy.

Paper III revealed additional complexity using GNNWR, which highlighted that child-level factors (age, male gender, birthweight) and household-level variables (sanitation, electricity, farmland size) explained much of the local heterogeneity of childhood stunting (Mertens et al., 2023). Local coefficient maps showed, for instance, that sanitation deficits had disproportionately large effects in rural areas, while electricity access was most relevant in peri-urban contexts. These findings underscore the value of geographically explicit models: the same variable does not carry the same influence everywhere. The influence of remotely sensed indicators warrants further discussion (Nduwayezu et al., 2024). **Papers I** and **IV** both included climatic and agroecological indicators such as NDVI, rainfall, and soil fertility. Although these variables were less predictive than maternal and household factors, they captured ecological vulnerability zones (Nduwayezu et al., 2025b). For example, **Paper I** highlighted that steep, high-rainfall areas with fragile soils coincided with higher stunting prevalence, suggesting that environmental stress interacts with poverty and childcare practices to exacerbate risk (Jones et al., 2016). Thus, RS is most useful when viewed not as a primary determinant (Yeboah et al., 2022), but as a contextual lens that situates household vulnerabilities within broader ecological systems (Grace et al., 2022). For policy, this has profound implications. A "one-size-fits-all" approach to nutrition interventions risks being inefficient or ineffective (Tamir et al., 2024). Instead, interventions should be tailored to the dominant local determinants (Nduwayezu et al., 2025a). In sectors where sanitation deficits are most pressing, investments in WASH (water, sanitation, hygiene) are likely to yield the greatest reductions in stunting (Uwiringiyimana et al., 2019). Where maternal health services lag, expanding antenatal care and facility-based deliveries should take priority (Kalinda et al., 2024). Electricity access, highlighted in **Paper III**, also emerged as a powerful determinant, pointing to the intersection of infrastructure development and nutrition outcomes (Davenport et al., 2017). The evidence from **Papers I**, **II**, and **III** thus supports a geo-targeted intervention strategy, aligning local risk profiles with tailored solutions, consistent with recommendations by Tamir et al. (2024) for context-specific approaches.

Predictive modeling and generalization to unsampled areas

The question of whether models can generalize to unsampled locations is addressed most directly by **Paper IV**. The SEL framework successfully combined LR, RF, several GBM variants, SVM, and ANNs to produce stunting prevalence maps at high spatial resolution (Figure 15). These maps filled critical data gaps by predicting stunting prevalence in unsampled locations, leveraging remotely sensed agroecological covariates alongside survey data.

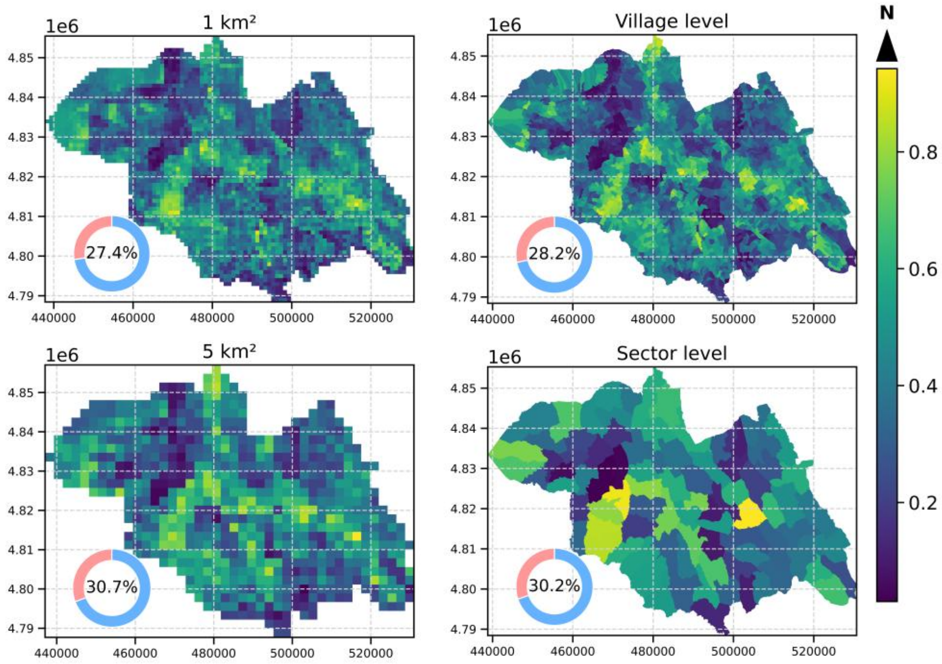


Figure 15: Maps of the predictive probability that HAZ lies below a threshold of -2 provided by the RF at 1 km^2, 5 km^2 spatial resolutions, at village and sector jurisdictional levels. The donut in the lower left corner illustrates the percentages of predicted stunting risk values with a probability exceeding 0.5.

The accuracy metrics summarized in Figure 16, confirmed that the ensemble approach improved predictive performance compared to single learners, with RF emerging as the strongest base model. The ability to generate fine-scale prevalence maps is particularly valuable in contexts like Rwanda, where national surveys are limited in coverage, and small-area prevalence is often unknown, has important public health policy relevance. By extending predictions to every grid cell, SEL outputs provide decision-ready surfaces that health planners can use to prioritize interventions (Reich & Haran, 2018). For instance, local governments can identify not only which districts have the highest stunting prevalence, but which specific sectors or villages require immediate action. Although **Papers I**, **II**, and **III** focused more on analyzing and interpreting local determinants, their methodological contributions complement predictive approaches. **Paper I** showed that MGWR achieved high explanatory power, while GNNWR in **Paper III** captured complex, non-linear effects with greater flexibility than traditional regression models, albeit with some limitations in generalization. However, both approaches face limitations in generalization: their strength lies in explaining local processes rather than predicting outcomes in entirely unsampled areas. **Paper IV**'s SEL, in contrast, prioritizes predictive accuracy and generalization, though it sacrifices some interpretability. Taken together, the four studies suggest a dual strategy: use interpretable local models to understand spatially varying determinants and use predictive SEL to fill data gaps and guide resource targeting.

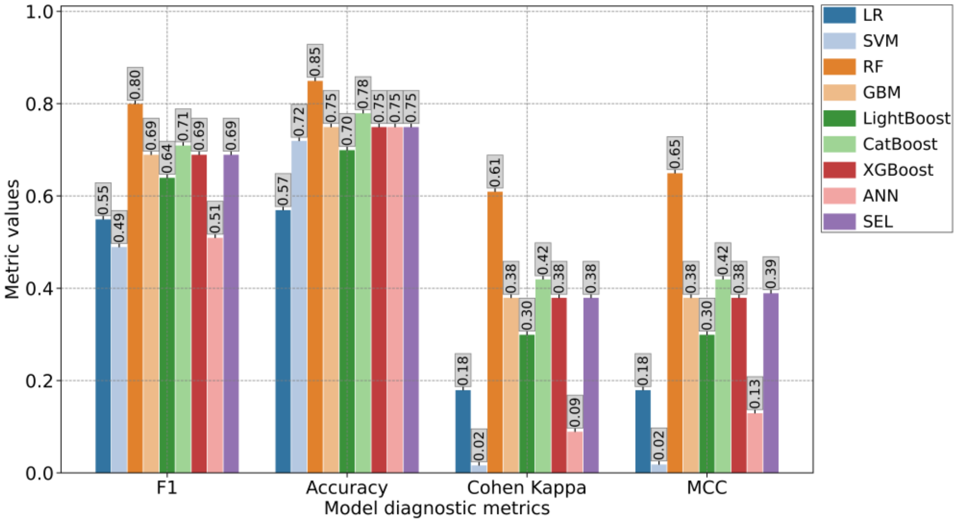


Figure 16: Performance metrics of the ML classifiers in predicting child stunting risk using four statistical metrics.

Synthesis

The integration of results across the four papers underscores three key insights. First, spatial clustering of stunting is robust, confirmed by both statistical measures and visual mapping across all methods and scales, with hotspots consistently observed in the Northern part of Rwanda. Second, localized nonlinear determinants matter, with maternal health, sanitation, and household infrastructure consistently emerging as key factors, while remotely sensed environmental variables provide essential ecological context, reflecting vegetation, climate, and topographic influences. Third, predictive models can generalize effectively to unsampled locations (Koldasbayeva et al., 2024), producing actionable risk maps that support geo-targeted interventions, except for its limited explainability, which remains an open challenge for future research. These findings demonstrate the power of hybrid spatial GeoAI approaches to move beyond descriptive national averages, offering fine-scale, evidence-based insights for reducing childhood stunting in Rwanda. The findings of this thesis reveal where stunting risks are concentrated, what drives them locally, and how predictions can be extended to unsampled areas, thereby providing a blueprint for spatially explicit public health policy that may extend beyond the Rwandan context.

Conclusion

This thesis presents an in-depth analysis of methodologies for detecting, quantifying, and predicting the spatially explicit patterns of childhood stunting in Rwanda and its underlying determinants. It provides insights that improve our understanding of how localized socioeconomic, climatic, and agroecological determinants interact to shape child nutrition outcomes across space. Trade-offs among modeling approaches, data sources, and spatial scales were central to all four studies on childhood stunting in Rwanda. Each paper applied different combinations of survey data, remotely sensed variables, and GeoAI methods, illustrating how methodological choices influence both explanatory power and predictive generalizability. While SEL approaches (**Paper IV**) excelled at producing continuous predictive surfaces across unsampled locations, spatially adaptive GeoAI methods (**Papers I**, **II**, and **III**) provided critical insight into how determinants of stunting vary across space. Together, these studies demonstrate that there is no single best approach; instead, models must be chosen and calibrated according to whether the priority is interpretability, generalization, or fine-scale sensitivity to local drivers. A key finding across the papers was that localized socioeconomic and maternal health variables consistently outperformed remotely sensed environmental features in explaining stunting prevalence. Nonetheless, climatic factors like rainfall, topographic features, and agroecological indicators such as NDVI provided valuable ecological context, refining spatial understanding of stunting risk (Nduwayezu et al., 2024).

This confirms that maternal literacy, antenatal care, sanitation, electricity access, and dietary diversity are the strongest immediate determinants of child nutrition outcomes (Nduwayezu et al., 2025a). However, environmental gradients such as rainfall, elevation, and NDVI still help explain why stunting clusters geographically, particularly in fragile, highland farming systems (Nduwayezu et al., 2025b). The integration of household-level and environmental predictors therefore provides a richer understanding of the nexus of population health and ecological vulnerabilities. Another important insight is that spatial heterogeneity is not only statistically detectable but also policy relevant (Kalinda et al., 2023). The clustering of stunting prevalence in Northern and Western Rwanda, identified in all four studies, demonstrates the limitations of relying on national averages. GNNWR revealed that the influence of individual predictors can shift dramatically across districts or villages, suggesting that uniform policy measures are unlikely to achieve maximum impact. Instead, these methods provide the evidence base for geo-targeted interventions such as prioritizing sanitation in rural clusters with poor WASH coverage or improving antenatal care in districts where maternal health services are weakest.

Finally, the studies collectively show that predictive modeling can extend beyond descriptive analysis to serve as a practical tool for intervention planning. SEL demonstrated that prevalence can be reliably mapped at fine spatial resolutions, even in areas without survey data. This capability is crucial for public health policy, as it enables governments to allocate resources more efficiently and equitably. At the same time, interpretable models such as MGWR and GWRF highlight not just where interventions should occur, but why certain locations are more vulnerable than others. This dual emphasis on both prediction and explanation marks a significant advancement in the application of GeoAI to public health challenges (Huang et al., 2025). In sum, the four papers underscore the potential of combining multi-source geospatial data with hybrid spatial ML to advance our understanding on child growth faltering. By statistically confirming clustering, revealing localized determinants, and enabling prediction in unsampled locations, these studies provide the foundations for evidence-based, geographically tailored nutrition interventions in Rwanda. More broadly, they illustrate the role of spatially explicit modeling in addressing nutritional deficits in childhood, where the interplay of environmental vulnerability and social determinants requires finely tuned, context-sensitive policy responses. Overall, these papers highlight that trade-offs between interpretability, generalizability, and spatial scale in the developed analytical and predictive models remain challenging to navigate and must be evaluated case by case according to research priorities, consistent with insights from Song et al. (2023). Consistently, studies show that no single SVC GeoAI model is universally optimal (Jemeļjanova et al., 2024); rather, selecting the most suitable model requires balancing accuracy, precision, computational cost (Janowicz et al., 2020), and ease of use in line with the goals of the decision-making context (Kmoch et al., 2025).

Limitations and outlook

Despite the merits of this study, several methodological limitations and challenges need to be addressed to ensure the effectiveness of localized-informed child stunting modeling. While this thesis leveraged recent spatially varying GeoAI methods to model the localized drivers of childhood stunting, it was fundamentally based on cross-sectional data, limiting its capacity to analyze temporal dynamics, trends, or the progression of stunting over time (Mertens et al., 2023). As a result, the predictive insights generated from the current models primarily reflect a static snapshot of existing conditions, without accounting for how the determinants of stunting may shift in response to seasonality, policy interventions, environmental changes, or socio-economic transitions (Bitew et al., 2023). Moreover, the proposed methods effectively analyze spatial associations between different geographical variables, overlooking spatial causality (Slater et al., 2025), which may restrict the depth of causal inference that can be drawn from the analyses (Molnar et al., 2022), a potential avenue for future research. This study integrates remotely sensed environmental data with census-based socioeconomic indicators to model climatic, and agroecological of child stunting. To address spatial misalignment, we applied harmonization techniques such as resampling and areal interpolation (Gribov & Krivoruchko, 2020), and assessed spatial consistency using global spatial autocorrelation metrics across different spatial scales. Similarly, the privacy and ethical challenges of cross-sectionally surveyed household childhood stunting data are linked to sensitive information on maternal violence, household decision-making, stigmatized illnesses such as HIV, child stature, and poverty (Polzin & Kounadi, 2021). These challenges necessitated varied geomasking methods (Wang et al., 2022), which may have introduced minor uncertainties not addressed by the rigorous model parameterization approach. Uncertainties associated with the models were also acknowledged to ensure transparency and support scalability, reproducibility (Delmelle et al., 2022), and replicability of the study (Li et al., 2024). This study balances the need for detailed insight with strict protections to ensure both scientific rigor and ethical concerns.

From a modelling perspective, some potential challenges of hybrid models include computational complexities, parameter tuning difficulties, and scalability constraints, which can be overlooked to some extent as these spatially informed AI models provide significant improvements in spatial epidemiology compared to traditional non-hybrid approaches (Koldasbayeva et al., 2024). The future research directions may finally extend the developed framework to incorporate longitudinal analyses and temporal causal modeling (Zhu et al., 2022a), to allow the development of spatiotemporal models capable of capturing evolving patterns of risk and providing dynamic predictions, designed to forecast emerging risk zones, monitor the effectiveness of interventions over time, and enable proactive public health planning. Finally, this study did not formally engage health professionals, caregivers, and community stakeholders regarding the interpretability of exploratory, analytical and prediction outputs, their utility for decision-making, and preferred delivery formats, which could form future directions to ensure results are actionable for addressing child stunting.

References

Aheto, J. M. K., & Dagne, G. A. (2021). Geostatistical analysis, web-based mapping, and environmental determinants of under-5 stunting: evidence from the 2014 Ghana Demographic and Health Survey. *The Lancet Planetary Health*, *5*(6), e347–e355. https://doi.org/10.1016/S2542-5196(21)00080-2

Ahmed, Z. U., Krupnik, T. J., Alam, M. S., Timsina, J., Hossain, K. H., Rahman, G. K. M. M., & McDonald, A. J. (2025). *Ensemble Machine Learning with Limited Data: Feature Selection and Predicting Wheat Yield in Bangladesh*. https://doi.org/10.2139/SSRN.5401541

Ahmed, Z. U., Sun, K., Shelly, M., & Mu, L. (2021). Explainable artificial intelligence (XAI) for exploring spatial variability of lung and bronchus cancer (LBC) mortality rates in the contiguous USA. *Scientific Reports*, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-03198-8

Ahn, S., Ryu, D. W., & Lee, S. (2020). A Machine Learning-Based Approach for Spatial Estimation Using the Spatial Features of Coordinate Information. *ISPRS International Journal of Geo-Information 2020, Vol. 9, Page 587*, *9*(10), 587. https://doi.org/10.3390/IJGI9100587

Alahmadi, H., & Moraga, P. (2025). Bayesian modelling for the integration of spatially misaligned health and environmental data. *Stochastic Environmental Research and Risk Assessment*, *39*(4), 1485–1499. https://doi.org/10.1007/S00477-025-02927-Z/FIGURES/7

Annan, K. (2018). Data can help to end malnutrition across Africa. *Nature*, *555*(7694), 7. https://doi.org/10.1038/D41586-018-02386-3

Anselin, L. (1988a). *Spatial Heterogeneity*. 119–136. https://doi.org/10.1007/978-94-015-7799-1_9

Anselin, L. (1988b). Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, *20*(1), 1–17. https://doi.org/10.1111/J.1538-4632.1988.TB00159.X

Anselin, L. (1989). *What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis* . https://escholarship.org/uc/item/3ph5k0d4

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/J.1538-4632.1995.TB00338.X

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, *89*(1), 3–26. https://doi.org/10.1111/J.1435-5957.2010.00279.X

Anselin, L., & Amaral, P. (2024). Endogenous spatial regimes. *Journal of Geographical Systems*, *26*(2), 209–234. https://doi.org/10.1007/S10109-023-00411-2/TABLES/8

Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research*, *20*(1), 40. https://doi.org/10.1002/MPR.329

Baffour, B., Aheto, J. M. K., Das, S., Godwin, P., & Richardson, A. (2023). Geostatistical modelling of child undernutrition in developing countries using remote-sensed data: evidence from Bangladesh and Ghana demographic and health surveys. *Scientific Reports 2023 13:1*, *13*(1), 1–13. https://doi.org/10.1038/s41598-023-48980-y

Bharti, R., Dhillon, P., & Narzary, P. K. (2019). A spatial analysis of childhood stunting and its contextual correlates in India. *Clinical Epidemiology and Global Health*, *7*(3), 488–495. https://doi.org/10.1016/J.CEGH.2019.04.005

Binbin, L., Yong, G., Kun, Q., Jianghua, Z., Binbin, L., Yong, G., Kun, Q., & Jianghua, Z. (2020). A Review on Geographically Weighted Regression. *Geomatics and Information Science of Wuhan University, 2020, Vol. 45, Issue 9, Pages: 1356-1366*, *45*(9), 1356–1366. https://doi.org/10.13203/J.WHUGIS20190346

Bitew, F. H., Sparks, C. S., Nyarko, S. H., & Apgar, L. (2023). Spatiotemporal Variations and Determinants of Under-Five Stunting in Ethiopia. *Food and Nutrition Bulletin*, *44*(1), 27. https://doi.org/10.1177/03795721231158503

Blanford, J. (2025). *Geographic Information, Geospatial Technologies and Spatial Data Science for Health*. https://books.google.rw/books?hl=en&lr=&id=vDMWEQAAQBAJ&oi=fnd&pg=PP1&dq=Justine+I+Blanford&ots=e8rwIQ5u3h&sig=jyVG6BcXq_Yq3D7jLw-5G_PlmgI&redir_esc=y#v=onepage&q&f=false

Boah, M., Issah, A. N., Yeboah, D., Kpordoxah, M. R., & Sira, J. (2022). Association Between Compliance With the New WHO-Recommended Frequency and Timing of Antenatal Care Contacts and Receiving Quality Antenatal Care in Cameroon. *SAGE Open*, *12*(3). https://doi.org/10.1177/21582440221117807/ASSET/IMAGES/LARGE/10.1177_21582440221117807-FIG1.JPEG

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Brunsdon, C., Fotheringham, A. S., & Charlton, M. (2002). Geographically weighted summary statistics — a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems*, *26*(6), 501–524. https://doi.org/10.1016/S0198-9715(01)00009-6

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, *28*(4), 281–298. https://doi.org/10.1111/J.1538-4632.1996.TB00936.X

Burgert-Brucker, C. R., Dontamsetti, T., & Gething, P. W. (2018). The DHS Program's Modeled Surfaces Spatial Datasets. *Studies in Family Planning*, *49*(1), 87–92. https://doi.org/10.1111/SIFP.12050

Carlin, J. B., & Hocking, J. (1999). Design of cross-sectional surveys using cluster sampling: an overview with Australian case studies. *Australian and New Zealand Journal of Public Health*, *23*(5), 546–551. https://doi.org/10.1111/J.1467-842X.1999.TB01317.X

Christian, A. K., & Dake, F. A. A. (2022). Profiling household double and triple burden of malnutrition in sub-Saharan Africa: prevalence and influencing household factors. *Public Health Nutrition*, *25*(6), 1563–1576. https://doi.org/10.1017/S1368980021001750

Clark, L. P., Zilber, D., Schmitt, C., Fargo, D. C., Reif, D. M., Motsinger-Reif, A. A., & Messier, K. P. (2024). A review of geospatial exposure models and approaches for health data integration. *Journal of Exposure Science & Environmental Epidemiology 2024 35:2*, *35*(2), 131–148. https://doi.org/10.1038/s41370-024-00712-8

Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Yihe, L., Murakami, D., Nakaya, T., Wang, Y., & Harris, P. (2023). A Route Map for Successful Applications of Geographically Weighted Regression. *Geographical Analysis*, *55*(1), 155–178. https://doi.org/10.1111/GEAN.12316

Comber, A., & Harris, P. (2022). The Importance of Scale and the MAUP for Robust Ecosystem Service Evaluations and Landscape Decisions. *Land 2022, Vol. 11, Page 399*, *11*(3), 399. https://doi.org/10.3390/LAND11030399

Comber, A., Harris, P., Bratkova, K., Phe, H. H., Kieu, M., Bui, Q. T., Nguyen, T. T. H., Wanjau, E., & Malleson, N. (2022). Handling the MAUP: methods for identifying appropriate scales of aggregation based on measures on spatial and non-spatial variance. *AGILE: GIScience Series*, *3*, 1–5. https://doi.org/10.5194/AGILE-GISS-3-30-2022

Comber, A., Harris, P., & Brunsdon, C. (2024). Multiscale spatially varying coefficient modelling using a Geographical Gaussian Process GAM. *International Journal of Geographical Information Science*, *38*(1), 27–47. https://doi.org/10.1080/13658816.2023.2270285/ASSET/028D982D-354D-4536-92F0-29A107D7BA04/ASSETS/GRAPHIC/TGIS_A_2270285_F0006_C.JPG

Credit, K. (2022). Spatial Models or Random Forest? Evaluating the Use of Spatially Explicit Machine Learning Methods to Predict Employment Density around New Transit Stations in Los Angeles. *Geographical Analysis*, *54*(1), 58–83. https://doi.org/10.1111/GEAN.12273

Davenport, F., Grace, K., Funk, C., & Shukla, S. (2017). Child health outcomes in sub-Saharan Africa: A comparison of changes in climate and socio-economic factors. *Global Environmental Change*, *46*, 72–87. https://doi.org/10.1016/J.GLOENVCHA.2017.04.009

Davies, M. M., & Van Der Laan, M. J. (2016). Optimal Spatial Prediction Using Ensemble Machine Learning. *International Journal of Biostatistics*, *12*(1), 179–201. https://doi.org/10.1515/IJB-2014-0060/MACHINEREADABLECITATION/RIS

Delmelle, E. M., Desjardins, M. R., Jung, P., Owusu, C., Lan, Y., Hohl, A., & Dony, C. (2022). Uncertainty in geospatial health: challenges and opportunities ahead. *Annals of Epidemiology*, *65*, 15–30. https://doi.org/10.1016/J.ANNEPIDEM.2021.10.002

Deng, R., Li, Z., & Wang, M. (2025). GeoAggregator: An Efficient Transformer Model for Geo-Spatial Tabular Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, *39*(11), 11572–11580. https://doi.org/10.1609/AAAI.V39I11.33259

Diggle, P. J., Menezes, R., & Su, T. li. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *59*(2), 191–232. https://doi.org/10.1111/J.1467-9876.2009.00701.X

Du, Z., Wang, Z., Wu, S., Zhang, F., & Liu, R. (2020). Geographically neural network weighted regression for the accurate estimation of spatial non-stationarity. *International Journal of Geographical Information Science*, *34*(7), 1353–1377. https://doi.org/10.1080/13658816.2019.1707834

Dusingizimana, T., Nduwayezu, G., & Kjelqvist, T. (2024). Women's dietary diversity is associated with homestead production and market access: A cross-sectional study in rural Rwanda. *Maternal & Child Nutrition*, e13755. https://doi.org/10.1111/MCN.13755

Eberth, J. M., Kramer, M. R., Delmelle, E. M., & Kirby, R. S. (2021). What is the place for space in epidemiology? *Annals of Epidemiology*, *64*, 41–46. https://doi.org/10.1016/J.ANNEPIDEM.2021.08.022

Egaña, A. F., Valenzuela, M. J., Maleki, M., Sánchez-Pérez, J. F., & Díaz, G. (2025). Adaptive ensemble spatial analysis. *Scientific Reports 2025 15:1*, *15*(1), 1–24. https://doi.org/10.1038/s41598-025-08844-z

ESRI. (2024). *How Incremental Spatial Autocorrelation works—ArcGIS Pro | Documentation*. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/how-incremental-spatial-autocorrelation-works.htm

Fang, G., Pang, W., Zhao, L., Cui, W., Zhu, L., Cao, S., & Ge, Y. (2021). Extreme Typhoon Wind Speed Mapping for Coastal Region of China: Geographically Weighted Regression–Based Circular Subregion Algorithm. *Journal of Structural Engineering*, *147*(10), 04021146. https://doi.org/10.1061/(ASCE)ST.1943-541X.0003122

Finley, A. O. (2011). Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, *2*(2), 143–154. https://doi.org/10.1111/J.2041-210X.2010.00060.X

Fotheringham, A. S. (2009). “The problem of spatial autocorrelation” and local spatial statistics. *Geographical Analysis*, *41*(4), 398–403. https://doi.org/10.1111/J.1538-4632.2009.00767.X

Fotheringham, A. S., Crespo, R., & Yao, J. (2015). Geographical and Temporal Weighted Regression (GTWR). *Geographical Analysis*, *47*(4), 431–452. https://doi.org/10.1111/GEAN.12071

Fotheringham, A. S., Oshan, T. M., & Li, Z. (2023). Multiscale Geographically Weighted Regression: Theory and Practice. *Multiscale Geographically Weighted Regression: Theory and Practice*, 1–176. https://doi.org/10.1201/9781003435464

Fotheringham, A. S., & Sachdeva, M. (2022). Scale and local modeling: new perspectives on the modifiable areal unit problem and Simpson’s paradox. *Journal of Geographical Systems*, *24*(3), 475–499. https://doi.org/10.1007/S10109-021-00371-5/FIGURES/18

Fotheringham, A. S., Yang, W., & Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, *107*(6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480

Fotheringham, A. S., Yu, H., Wolf, L. J., Oshan, T. M., & Li, Z. (2022). On the notion of ‘bandwidth’ in geographically weighted regression models of spatially varying processes. *Https://Doi.Org/10.1080/13658816.2022.2034829*, *36*(8), 1485–1502. https://doi.org/10.1080/13658816.2022.2034829

Fouedjio, F., & Arya, E. (2024). Locally varying geostatistical machine learning for spatial prediction. *Artificial Intelligence in Geosciences*, *5*, 100081. https://doi.org/10.1016/J.AIIG.2024.100081

Fratesi, U., Abreu, M., Bond-Smith, S., Corrado, L., Ditzen, J., Felsenstein, D., Fuerst, F., Ioramashvili, C., Kopczewska, K., Monastiriotis, V., Piras, G., Quatraro, F., Ravazzolo, F., Tranos, E., Tsiotas, D., & Yu, J. (2025). Spillovers and small spatial scale analyses: contributions from spatial economics. *Spatial Economic Analysis*, *20*(1), 1–7. https://doi.org/10.1080/17421772.2025.2466304

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., & Kalogirou, S. (2021). Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, *36*(2), 121–136. https://doi.org/10.1080/10106049.2019.1595177/FORMAT/EPUB

Georganos, S., & Kalogirou, S. (2022). A Forest of Forests: A Spatially Weighted and Computationally Efficient Formulation of Geographical Random Forests. *ISPRS International Journal of Geo-Information 2022, Vol. 11, Page 471*, *11*(9), 471. https://doi.org/10.3390/IJGI11090471

Getis, A. (2010). Spatial Autocorrelation. *Handbook of Applied Spatial Analysis*, 255–278. https://doi.org/10.1007/978-3-642-03647-7_14

Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, *24*(3), 189–206. https://doi.org/10.1111/J.1538-4632.1992.TB00261.X

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *Journal of Statistical Software*, *63*(17), 1–50. https://doi.org/10.18637/JSS.V063.I17

Goodchild, M. F. (2004). GIScience, Geography, Form, and Process. *Annals of the Association of American Geographers*, *94*(4), 709–714. https://doi.org/10.1111/J.1467-8306.2004.00424.X

Goodchild, M. F., & Li, W. (2021). Replication across space and time must be weak in the social and environmental sciences. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(35), e2015759118. https://doi.org/10.1073/PNAS.2015759118/ASSET/9AFAE61C-E626-4060-B5D1-551D7E7D889C/ASSETS/IMAGES/LARGE/PNAS.2015759118FIG02.JPG

Grace, K., Verdin, A., Brown, M., Bakhtsiyarava, M., Backer, D., & Billing, T. (2022). Conflict and Climate Factors and the Risk of Child Acute Malnutrition Among Children Aged 24–59 Months: A Comparative Analysis of Kenya, Nigeria, and Uganda. *Spatial Demography*. https://doi.org/10.1007/S40980-021-00102-W

Gribov, A., & Krivoruchko, K. (2020). Empirical Bayesian kriging implementation and usage. *Science of The Total Environment*, *722*, 137290. https://doi.org/10.1016/J.SCITOTENV.2020.137290

Guo, H., Dong, L., Wu, L., & Liu, Y. (2025). Tackling spatial heterogeneity in geographical analysis: an overview. *Acta Geographica Sinica*, *80*(3), 567–585. https://doi.org/10.11821/DLXB202503001

Habtu, M., Agena, A. G., Umugwaneza, M., Mochama, M., & Munyanshongore, C. (2022). Effect and Challenges of an Integrated Nutrition-Intervention Package Utilization among Pregnant Women and Lactating Mothers in Rwanda: An Exploratory Qualitative Study. *Current Developments in Nutrition*, *7*(1), 100018. https://doi.org/10.1016/J.CDNUT.2022.100018

Haileselassie, M., Redae, G., Berhe, G., Henry, C. J., Nickerson, M. T., Tyler, B., & Mulugeta, A. (2020). Why are animal source foods rarely consumed by 6-23 months old children in rural communities of Northern Ethiopia? A qualitative study. *PLOS ONE*, *15*(1), e0225707. https://doi.org/10.1371/JOURNAL.PONE.0225707

Hampton, K. H., Fitch, M. K., Allshouse, W. B., Doherty, I. A., Gesink, D. C., Leone, P. A., Serre, M. L., & Miller, W. C. (2010). Mapping Health Data: Improved Privacy Protection With Donut Method Geomasking. *American Journal of Epidemiology*, *172*(9), 1062. https://doi.org/10.1093/AJE/KWQ248

Hastie, T. J. (2017). Generalized Additive Models. *Statistical Models in S*, 249–307. https://doi.org/10.1201/9780203738535-7/GENERALIZED-ADDITIVE-MODELS-

TREVOR-HASTIE
Hastie, T. J., & Tibshirani, R. (1995). *Generalized additive models for medical research*. *4*(3), 187–196. https://doi.org/10.1177/096228029500400302
Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *2018*(8), e5518. https://doi.org/10.7717/PEERJ.5518/SUPP-1
Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8
Hu, G., & Tang, Y. (2023). *GERPM: A Geographically Weighted Stacking EnsembleLearning-Based Urban Residential Rents Prediction Model*. https://doi.org/10.3390/math11143160
Huang, X., Wang, S., Wilson, J., & Kedron, P. (2025). *GeoAI and Human Geography: The Dawn of a New Spatial Intelligence Era* (X. Huang, S. Wang, J. Wilson, & P. Kedron (eds.)). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-87421-5
IHME. (2020). Global Child Growth Failure Geospatial Estimates 2000-2019. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2020. In *Institute for Health Metrics and Evaluation (IHME)*. https://www.healthdata.org/data-tools-practices/interactive-visuals/child-growth-failure
Janocha, B., Donohue, R. E., Fish, T. D., Mayala, B. K., & Croft, T. N. (2021). *Guidance and recommendations for the use of indicator estimates at subnational administrative level 2*. DHS Spatial Analysis Reports No. 20. Rockville, Maryland, USA: ICF. https://dhsprogram.com/publications/publication-sar20-spatial-analysis-reports.cfm
Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, *34*(4), 625–636. https://doi.org/10.1080/13658816.2019.1684500
Jemeļjanova, M., Kmoch, A., & Uuemaa, E. (2024). Adapting machine learning for environmental spatial data - A review. *Ecological Informatics*, *81*, 102634. https://doi.org/10.1016/J.ECOINF.2024.102634
Jensen, H. A. R., Lau, C. J., Davidsen, M., Feveile, H. B., Christensen, A. I., & Ekholm, O. (2022). The impact of non-response weighting in health surveys for estimates on primary health care utilization. *European Journal of Public Health*, *32*(3), 450–455. https://doi.org/10.1093/EURPUB/CKAC032
Jiang, S., Sweet, L. belle, Blougouras, G., Brenning, A., Li, W., Reichstein, M., Denzler, J., Shangguan, W., Yu, G., Huang, F., & Zscheischler, J. (2024). How Interpretable Machine Learning Can Benefit Process Understanding in the Geosciences. *Earth's Future*, *12*(7), e2024EF004540. https://doi.org/10.1029/2024EF004540
Jiao, Z., & Tao, R. (2025). Geographical Gaussian Process Regression: A Spatial Machine-Learning Model Based on Spatial Similarity. *Geographical Analysis*, *57*(3), 507–520. https://doi.org/10.1111/GEAN.12423
Jones, A. D., Acharya, Y., & Galway, L. P. (2016). Urbanicity Gradients Are Associated with the Household- and Individual-Level Double Burden of Malnutrition in Sub-Saharan Africa. *The Journal of Nutrition*, *146*(6), 1257–1267. https://doi.org/10.3945/JN.115.226654
Kalinda, C., Phiri, M., Simona, S. J., Banda, A., Wong, R., Qambayot, M. A., Ishimwe, S.

M. C., Amberbir, A., Abebe, B., Gebremariam, A., & Nyerere, J. O. (2023). Understanding factors associated with rural-urban disparities of stunting among under-five children in Rwanda: A decomposition analysis approach. *Maternal & Child Nutrition*, *19*(3). https://doi.org/10.1111/MCN.13511

Kalinda, C., Qambayot, M. A., Ishimwe, S. M. C., Regnier, D., Bazimya, D., Uwizeyimana, T., Desie, S., Rudert, C., Gebremariam, A., Brennan, E., Karumba, S., Wong, R., & Bekele, A. (2024). Leveraging multisectoral approach to understand the determinants of childhood stunting in Rwanda: a systematic review and meta-analysis. *Systematic Reviews 2024 13:1*, *13*(1), 1–12. https://doi.org/10.1186/S13643-023-02438-4

Kamgno, J., Adeleke, M., Basáñez, M. G., Coulibaly, Y., de Souza, D. K., Debrah, L. B., Debrah, A. Y., Diggle, P. J., Nana-Djeunga, H. C., Domché, A., Gass, K., Hoerauf, A., Hopkins, A., Klion, A., Mackenzie, C. D., Mwingira, U., Njenga, S. M., Nutman, T. B., Nwane, P., … Kelly-Hope, L. A. (2025). Vector-borne helminthiases: a road map for current and future research to support control and elimination in sub-Saharan Africa. *The Lancet Infectious Diseases*. https://doi.org/10.1016/S1473-3099(25)00084-2

Kanevski, M. (2025). On machine learning from environmental data. *Cahiers Du Centre de Linguistique et Des Sciences Du Langage*, *69*, 127–142. https://doi.org/10.26034/LA.CDCLSL.2025.8351

Katz, J. (1995). Sample-size implications for population-based cluster surveys of nutritional status. *The American Journal of Clinical Nutrition*, *61*(1), 155–160. https://doi.org/10.1093/AJCN/61.1.155

Khaki, J. J., Macharia, P. M., Bĕová, L., Giorgi, E., & Semaan, A. (2024). Prevalence and determinants of double and triple burden of malnutrition among mother-child pairs in Malawi: a mapping and multilevel modelling study. *Public Health Nutrition*, 1–28. https://doi.org/10.1017/S1368980024002064

Kianfar, N., Sartorius, B., Lau, C. L., Bergquist, R., & Kiani, B. (2025). The future of spatial epidemiology in the AI era: enhancing machine learning approaches with explicit spatial structure. *Geospatial Health*, *20*(1). https://doi.org/10.4081/GH.2025.1386

Kirby, R. S., Delmelle, E., & Eberth, J. M. (2017). Advances in spatial epidemiology and geographic information systems. *Annals of Epidemiology*, *27*(1), 1–9. https://doi.org/10.1016/J.ANNEPIDEM.2016.12.001

Kismul, H., Acharya, P., Mapatano, M. A., & Hatløy, A. (2017). Determinants of childhood stunting in the Democratic Republic of Congo: Further analysis of Demographic and Health Survey 2013-14. *BMC Public Health*, *18*(1), 1–14. https://doi.org/10.1186/S12889-017-4621-0/TABLES/2

Kmoch, A., Harrison, C. T., Choi, J., & Uuemaa, E. (2025). Spatial autocorrelation in machine learning for modelling soil organic carbon. *Ecological Informatics*, *86*, 103057. https://doi.org/10.1016/J.ECOINF.2025.103057

Koldasbayeva, D., Tregubova, P., Gasanov, M., Zaytsev, A., Petrovskaia, A., & Burnaev, E. (2024). Challenges in data-driven geospatial modeling for environmental research and practice. *Nature Communications 2024 15:1*, *15*(1), 1–16. https://doi.org/10.1038/s41467-024-55240-8

Konstantinoudis, G., Schuhmacher, D., Ammann, R. A., Diesch, T., Kuehni, C. E., & Spycher, B. D. (2020). Bayesian spatial modelling of childhood cancer incidence in Switzerland using exact point data: A nationwide study during 1985-2015. *International Journal of Health Geographics*, *19*(1), 1–13.

https://doi.org/10.1186/S12942-020-00211-7/FIGURES/3

Lee, S. J., Yeatts, K. B., & Serre, M. L. (2009). A Bayesian Maximum Entropy approach to address the change of support problem in the spatial analysis of childhood asthma prevalence across North Carolina. *Spatial and Spatio-Temporal Epidemiology*, *1*(1), 49–60. https://doi.org/10.1016/J.SSTE.2009.07.005

Li, W. (2020). GeoAI: Where machine learning and big data converge in GIScience. *Journal of Spatial Information Science*, *20*(20), 71–77. https://doi.org/10.5311/JOSIS.2020.20.658

Li, W., Hsu, C. Y., Wang, S., & Kedron, P. (2024). GeoAI Reproducibility and Replicability: A Computational and Spatial Perspective. *Annals of the American Association of Geographers*, *114*(9), 2085–2103. https://doi.org/10.1080/24694452.2024.2373787

Li, Z. (2024). GeoShapley: A Game Theory Approach to Measuring Spatial Effects in Machine Learning Models. *Annals of the American Association of Geographers*, 1–21. https://doi.org/10.1080/24694452.2024.2350982

Lin, C. H., & Wen, T. H. (2022). How Spatial Epidemiology Helps Understand Infectious Human Disease Transmission. *Tropical Medicine and Infectious Disease 2022, Vol. 7, Page 164*, *7*(8), 164. https://doi.org/10.3390/TROPICALMED7080164

Liu, L. (2024). An ensemble framework for explainable geospatial machine learning models. *International Journal of Applied Earth Observation and Geoinformation*, *132*, 104036. https://doi.org/10.1016/J.JAG.2024.104036

Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine*, *226*. https://doi.org/10.1016/J.CMPB.2022.107161

Lu, B., Brunsdon, C., Charlton, M., & Harris, P. (2017). Geographically weighted regression with parameter-specific distance metrics. *International Journal of Geographical Information Science*, *31*(5), 982–998. https://doi.org/10.1080/13658816.2016.1263731

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence 2020 2:1*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Neural Information Processing Systems*.

Luo, Y., Yan, J., & McClure, S. (2021). Distribution of the environmental and socioeconomic risk factors on COVID-19 death rate across continental USA: a spatial nonlinear analysis. *Environmental Science and Pollution Research International*, *28*(6), 1. https://doi.org/10.1007/S11356-020-10962-2

Maitra, I., Lin, R., Chen, E., Donnelly, J., Scepanovic, S., & Rudin, C. (2025). How Your Location Relates to Health: Variable Importance and Interpretable Machine Learning for Environmental and Sociodemographic Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, *39*(27), 28240–28248. https://doi.org/10.1609/AAAI.V39I27.35044

McLean, J., Northrup-Lyons, M., Reid, R. J., Smith, L., Ho, K., Mucumbitsi, A., Kayumba, J., Omwega, A., McDonald, C., Schauer, C., & Zlotkin, S. (2018). From evidence to national scale: An implementation framework for micronutrient powders in Rwanda. *Maternal & Child Nutrition*, *15*(3), e12752. https://doi.org/10.1111/MCN.12752

Mertens, A., Benjamin-Chung, J., Colford, J. M., Coyle, J., van der Laan, M. J., Hubbard, A. E., Rosete, S., Malenica, I., Hejazi, N., Sofrygin, O., Cai, W., Li, H., Nguyen, A., Pokpongkiat, N. N., Djajadi, S., Seth, A., Jung, E., Chung, E. O., Jilek, W., … Yori, P. P. (2023). Causes and consequences of child growth faltering in low-resource settings. *Nature 2023 621:7979*, *621*(7979), 568–576. https://doi.org/10.1038/s41586-023-06501-x

Milà, C., Ludwig, M., Pebesma, E., Tonne, C., & Meyer, H. (2024). Random forests with spatial proxies for environmental modelling: opportunities and pitfalls. *Geoscientific Model Development*, *17*(15), 6007–6033. https://doi.org/10.5194/GMD-17-6007-2024

Molnar, C. (2022). *Interpretable Machine Learning:A Guide for Making Black Box Models Explainable*.

Molnar, C., & Freiesleben, T. (2024). *Supervised Machine Learning for Science: How to stop worrying and love your black box - Causality*. https://ml-science-book.com/

Molnar, C., König, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., & Bischl, B. (2022). General Pitfalls of Model-Agnostic Interpretation Methods for Machine Learning Models. *Lecture Notes in Computer Science*, *13200 LNAI*, 39–68. https://doi.org/10.1007/978-3-031-04083-2_4

Moore, D. A., & Carpenter, T. E. (1999). Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiologic Reviews*, *21*(2), 143–161. https://doi.org/10.1093/OXFORDJOURNALS.EPIREV.A017993

Moraga, P. (2019). *Geospatial Health Data, Modeling and Visualization with R-INLA and Shiny*. Chapman and Hall/CRC.

Nduwayezu, G., Kagoyire, C., Zhao, P., Eklund, L., Pilesjo, P., Bizimana, J. P., & Mansourian, A. (2024). Spatial Machine Learning for Exploring the Variability in Low Height-For-Age From Socioeconomic, Agroecological, and Climate Features in the Northern Province of Rwanda. *GeoHealth*, *8*(9), e2024GH001027. https://doi.org/10.1029/2024GH001027

Nduwayezu, G., Mansourian, A., Bizimana, J. P., & Pilesjö, P. (2025a). Hybridizing spatial machine learning to explore the fine-scale heterogeneity between stunting prevalence and its associated risk determinants in Rwanda. *Geo-Spatial Information Science*, 1–21. https://doi.org/10.1080/10095020.2025.2459133

Nduwayezu, G., Zhao, P., Pilesjö, P., Bizimana, J. P., & Mansourian, A. (2025b). Multilevel small-area childhood stunting risk estimation: Insights from spatial ensemble learning, agro-ecological and environmentally remotely sensed indicators. *Environmental and Sustainability Indicators*, *27*, 100822. https://doi.org/10.1016/J.INDIC.2025.100822

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *135*(3), 370–384. https://doi.org/10.2307/2344614

NISR, MoH-Rwanda, & ICF. (2021). *Rwanda Demographic and Health Survey 2019-2020. Rockville, Maryland, USA: NISR, MOH, and ICF International*.

Olatunji, J., Fronterre, C., Amoah, B., Montresor, A., Giorgi, E., Midzi, N., Mutsaka-Makuvaza, M. J., Kargbo-Labor, I., Hodges, M. H., Zhang, Y., Okoyo, C., Mwandawiro, C., Minnery, M., & DIggle, P. J. (2021). Model-Based Geostatistical Methods Enable Efficient Design and Analysis of Prevalence Surveys for Soil-

Transmitted Helminth Infection and Other Neglected Tropical Diseases. *Clinical Infectious Diseases*, *72*(Supplement_3), S172–S179. https://doi.org/10.1093/CID/CIAB192

Openshaw, S., & Taylor, P. (1979). A million or so correlated coefficients : three experiment on the modifiable areal unit problem. *Statistical Applications in the Spatial Sciences*, *7*(1), 19–26. https://doi.org/10.5638/THAGIS.7.19

Ord, J. K., & Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, *27*(4), 286–306. https://doi.org/10.1111/J.1538-4632.1995.TB00912.X

Osgood-Zimmerman, A., Millear, A. I., Stubbs, R. W., Shields, C., Pickering, B. V., Earl, L., Graetz, N., Kinyoki, D. K., Ray, S. E., Bhatt, S., Browne, A. J., Burstein, R., Cameron, E., Casey, D. C., Deshpande, A., Fullman, N., Gething, P. W., Gibson, H. S., Henry, N. J., … Hay, S. I. (2018). Mapping child growth failure in Africa between 2000 and 2015. *Nature 2018 555:7694*, *555*(7694), 41–47. https://doi.org/10.1038/nature25760

Oshan, T. M., Li, Z., Kang, W., Wolf, L. J., & Stewart Fotheringham, A. (2019). mgwr: A Python Implementation of Multiscale Geographically Weighted Regression for Investigating Process Spatial Heterogeneity and Scale. *ISPRS International Journal of Geo-Information*, *8*(6), 269. https://doi.org/10.3390/IJGI8060269

Oulaid, B., Harris, P., Maas, E., Fakeye, I. A., & Baker, C. (2025). Geographically Weighted Quantile Machine Learning for Probabilistic Soil Moisture Prediction from Spatially Resolved Remote Sensing. *Remote Sensing 2025, Vol. 17, Page 2907*, *17*(16), 2907. https://doi.org/10.3390/RS17162907

Pattnaik, A., Mohan, D., Tsui, A., Chipokosa, S., Katengeza, H., Ndawala, J., & Marx, M. A. (2021). The aggregate effect of implementation strength of family planning programs on modern contraceptive use at the health systems level in rural Malawi. *PLoS ONE*, *16*(11). https://doi.org/10.1371/JOURNAL.PONE.0232504

Polzin, F., & Kounadi, O. (2021). Adaptive Voronoi Masking: A Method to Protect Confidential Discrete Spatial Data. *Leibniz International Proceedings in Informatics, LIPIcs*, *208*, 1:1-1:17. https://doi.org/10.4230/LIPICS.GISCIENCE.2021.II.1

Pradhan, N., Agrawal, A., By, E., & Kimbrough, E. (2025). Mapping fine-scale socioeconomic inequality using machine learning and remotely sensed data. *PNAS Nexus*, *4*(2). https://doi.org/10.1093/PNASNEXUS/PGAF040

Prince, M. (2012). Epidemiology. *Core Psychiatry: Third Edition*, 115–129. https://doi.org/10.1016/B978-0-7020-3397-1.00009-4

Reich, B. J., & Haran, M. (2018). Precision maps for public health. *Nature 2021 555:7694*, *555*(7694), 32–33. https://doi.org/10.1038/d41586-018-02096-w

Sahana, K. S., Karthik JRF, C. B., Builder Project, D., Aher, J., Mounika Sree, M. M., Authors S Sahana, A. K., Karthik, C. B., Sree, M. M., Chaithra, M., Abhinandithe, S. K., Snehalatha, B., Sridhar, R., Balasubramanian, S., Madhu, B., Sahana, K., Karthik, C., Mounika, S., & Stavelin Abhinandithe, K. (2023). Geospatial Artificial Intelligence (GeoAI): Applications in Health Care. *International Journal of Health and Allied Sciences*, *11*(4), 4. https://doi.org/10.55691/2278-344X.1044

Santos, F., Graw, V., & Bonilla, S. (2019). A geographically weighted random forest approach for evaluate forest change drivers in the Northern Ecuadorian Amazon. *PLOS ONE*, *14*(12), e0226224. https://doi.org/10.1371/JOURNAL.PONE.0226224

Seidl, D. E. (2025). Geomasking. *International Encyclopedia of Geography*, 1–4.

https://doi.org/10.1002/9781118786352.WBIEG2223
Seidler, V., Utazi, E. C., Finaret, A. B., Luckeneder, S., Zens, G., Bodarenko, M., Smith, A. W., Bradley, S. E. K., Tatem, A. J., & Webb, P. (2025). Subnational variations in the quality of household survey data in sub-Saharan Africa. *Nature Communications 2025 16:1*, *16*(1), 1–12. https://doi.org/10.1038/s41467-025-58776-5
Slater, L., Blougouras, G., Deng, L., Deng, Q., Ford, E., Hoek Van Dijke, A., Huang, F., Jiang, S., Liu, Y., Moulds, S., Schepen, A., Yin, J., & Zhang, B. (2025). Challenges and opportunities of ML and explainable AI in large-sample hydrology. *Philosophical Transactions A*, *383*(2302). https://doi.org/10.1098/RSTA.2024.0287
Song, Y., Kalacska, M., Gašparović, M., Yao, J., & Najibi, N. (2023). Advances in geocomputation and geospatial artificial intelligence (GeoAI) for mapping. *International Journal of Applied Earth Observation and Geoinformation*, *120*, 103300. https://doi.org/10.1016/J.JAG.2023.103300
Stofer, K. A. (2016). When a Picture Isn't Worth 1000 Words: Learners Struggle to Find Meaning in Data Visualizations. *Journal of Geoscience Education*, *64*(3), 231–241. https://doi.org/10.5408/14-053.1
Stofer, K. A. (2024). Visualizing Spatially Based Data for Various Stakeholder Audiences. *UF/IFAS*. https://edis.ifas.ufl.edu/publication/WC163
Tamir, T. T., Tekeba, B., Mekonen, E. G., Zegeye, A. F., & Gebrehana, D. A. (2024). Spatial heterogeneity and predictors of stunting among under five children in Mozambique: a geographically weighted regression. *Frontiers in Public Health*, *12*, 1502018. https://doi.org/10.3389/FPUBH.2024.1502018
Tangena, J. A. A., Mategula, D., Sedda, L., & Atkinson, P. M. (2023). Unravelling the impact of insecticide-treated bed nets on childhood malaria in Malawi. *Malaria Journal*, *22*(1), 1–12. https://doi.org/10.1186/S12936-023-04448-Y/TABLES/3
Tiwari, A., Ahmad, S., Qurunflah, E., Helmi, M., Almaimani, A., Alaidroos, A., & Hallawani, M. M. (2023). Exploring geomasking methods for geoprivacy: a pilot study in an environment with built features. *Geospatial Health*, *18*(2). https://doi.org/10.4081/GH.2023.1205
Tuson, M., Yap, M., Kok, M. R., Boruff, B., Murray, K., Vickery, A., Turlach, B. A., & Whyatt, D. (2020). Overcoming inefficiencies arising due to the impact of the modifiable areal unit problem on single-Aggregation disease maps. *International Journal of Health Geographics*, *19*(1), 1–18. https://doi.org/10.1186/S12942-020-00236-Y/FIGURES/10
UN-HABITAT. (2018). Transforming Our World: The 2030 Agenda for Sustainable Development. *A New Era in Global Health*. https://doi.org/10.1891/9780826190123.ap02
UNICEF, WHO, & World Bank. (2025). *Levels and trends in child malnutrition: Key findings of the 2025 Edition of the Joint Child Malnutrition Estimates*. https://apps.who.int/iris/handle/10665/341135
Uwiringiyimana, V., Veldkamp, A., & Amer, S. (2019). Stunting spatial pattern in Rwanda: An examination of the demographic, socio-economic and environmental determinants. *Geospatial Health*, *14*(2), 329–339. https://doi.org/10.4081/GH.2019.820
Vapnik, V. N. (1995). The Nature of Statistical Learning Theory. *The Nature of Statistical Learning Theory*. https://doi.org/10.1007/978-1-4757-2440-0
Vollmer, S., Bommer, C., Krishna, A., Harttgen, K., & Subramanian, S. V. (2017). The association of parental education with childhood undernutrition in low- and middle-

income countries: comparing the role of paternal and maternal education. *International Journal of Epidemiology*, *46*(1), 312–323. https://doi.org/10.1093/IJE/DYW133

Wang, C., Zhang, M., Wang, C., & Zhang, M. (2025). *GeoAI and Health Geography*. 193–206. https://doi.org/10.1007/978-3-031-87421-5_14

Wang, J. (2024a). Geomasking to Safeguard Geoprivacy in Geospatial Health Data. *Encyclopedia 2024, Vol. 4, Pages 1581-1589*, *4*(4), 1581–1589. https://doi.org/10.3390/ENCYCLOPEDIA4040103

Wang, J. F., Li, X. H., Christakos, G., Liao, Y. L., Zhang, T., Gu, X., & Zheng, X. Y. (2010). Geographical Detectors-Based Health Risk Assessment and its Application in the Neural Tube Defects Study of the Heshun Region, China. *International Journal of Geographical Information Science*, *24*(1), 107–127. https://doi.org/10.1080/13658810802443457

Wang, J., Haining, R., Zhang, T., Xu, C., Hu, M., Yin, Q., Li, L., Zhou, C., Li, G., & Chen, H. (2024b). Statistical Modeling of Spatially Stratified Heterogeneous Data. *Annals of the American Association of Geographers*, *114*(3), 499–519. https://doi.org/10.1080/24694452.2023.2289982

Wang, J., Kim, J., & Kwan, M. P. (2022). An exploratory assessment of the effectiveness of geomasking methods on privacy protection and analytical accuracy for individual-level geospatial data. *Cartography and Geographic Information Science*, *49*(5), 385–406. https://doi.org/10.1080/15230406.2022.2056510

Wang, M., Burchard-Levine, V., Hu, T., Roujean, J.-L., Wang, J., Zhang, N., Zhang, L., Jing, H., Yan, Y., Wu, S., & Liu, R. (2024c). Spatial Downscaling of Nighttime Land Surface Temperature Based on Geographically Neural Network Weighted Regression Kriging. *Remote Sensing 2024, Vol. 16, Page 2542*, *16*(14), 2542. https://doi.org/10.3390/RS16142542

WHO. (2006). WHO Child Growth Standards based on length/height, weight and age. *Acta Paediatrica (Oslo, Norway : 1992). Supplement*, *450*, 76–85. https://doi.org/10.1111/J.1651-2227.2006.TB02378.X

Wiedemann, N., Martin, H., & Westerholt, R. (2023). Benchmarking Regression Models Under Spatial Heterogeneity. *Leibniz International Proceedings in Informatics, LIPIcs*, *277*, 11:1-11:15. https://doi.org/10.4230/LIPICS.GISCIENCE.2023.11/-/STATS

Wilson, K., & Wakefield, J. (2020). Pointless spatial modeling. *Biostatistics*, *21*(2), e17–e32. https://doi.org/10.1093/BIOSTATISTICS/KXY041

Wolf, L. J., Oshan, T. M., & Fotheringham, A. S. (2018). Single and Multiscale Models of Process Spatial Heterogeneity. *Geographical Analysis*, *50*(3), 223–246. https://doi.org/10.1111/GEAN.12147

Wu, S., Wang, Z., Du, Z., Huang, B., Zhang, F., & Liu, R. (2021). Geographically and temporally neural network weighted regression for modeling spatiotemporal non-stationary relationships. *International Journal of Geographical Information Science*, *35*(3), 582–608. https://doi.org/10.1080/13658816.2020.1775836

Yeboah, E., Kuunibe, N., Mank, I., Parisi, D., Bonnet, E., Lohmann, J., Hamadou, S., Picbougoum, B. T., Belesova, K., Sauerborn, R., Baernighausen, T., Danquah, I., & De Allegri, M. (2022). Every drop matters: combining population-based and satellite data to investigate the link between lifetime rainfall exposure and chronic undernutrition in children under five years in rural Burkina Faso. *Environmental*

Research Letters, *17*(5), 054027. https://doi.org/10.1088/1748-9326/AC661C
Yin, Z., Ding, J., Liu, Y., Wang, R., Wang, Y., Chen, Y., Qi, J., Wu, S., & Du, Z. (2024). GNNWR: An open-source package of spatiotemporal intelligent regression methods for modeling spatial and temporal nonstationarity. *Geoscientific Model Development*, *17*(22), 8455–8468. https://doi.org/10.5194/GMD-17-8455-2024
Yoneshiro, T., Matsushita, M., Fuse-Hamaoka, S., Kuroiwa, M., Kurosawa, Y., Yamada, Y., Arai, M., Wei, Y., Iida, M., Kuma, K., Kameya, T., Harada, T., Matsumura, Y., Osawa, T., Aoki, Y., Nakamura, H., Hamaoka, T., Sakai, J., & Saito, M. (2025). Pre-fertilization-origin preservation of brown fat-mediated energy expenditure in humans. *Nature Metabolism 2025 7:4*, *7*(4), 778–791. https://doi.org/10.1038/s42255-025-01249-2
Zhang, Z., Li, Z., & Song, Y. (2024a). On ignoring the heterogeneity in spatial autocorrelation: consequences and solutions. *International Journal of Geographical Information Science*. https://doi.org/10.1080/13658816.2024.2391981/ASSET/921CEC13-BA39-432E-9DBE-0A6FD073C8E5/ASSETS/GRAPHIC/TGIS_A_2391981_F0013_C.JPG
Zhang, Z., Wu, S., Zhang, B., Du, Z., & Xia, Q. (2024b). The Distribution of Surface Heat Flow on the Tibetan Plateau Revealed by Data-Driven Methods. *Journal of Geophysical Research: Solid Earth*, *129*(10). https://doi.org/10.1029/2023JB028491
Zhao, Z., Xu, Z., Hu, C., Wang, K., & Ding, X. (2024). Geographically weighted neural network considering spatial heterogeneity for landslide susceptibility mapping: A case study of Yichang City, China. *CATENA*, *234*, 107590. https://doi.org/10.1016/J.CATENA.2023.107590
Zhu, D., Gao, S., & Cao, G. (2022a). Towards the intelligent era of spatial analysis and modeling. *Proceedings of the 5th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, GeoAI 2022*, 10–13. https://doi.org/10.1145/3557918.3565863
Zhu, D., Liu, Y., Yao, X., & Fischer, M. M. (2022b). Spatial regression graph convolutional neural networks: A deep learning paradigm for spatial multivariate distributions. *GeoInformatica*, *26*(4), 645–676. https://doi.org/10.1007/S10707-021-00454-X/FIGURES/13

Printed by Media-Tryck, Lund 2025 NORDIC SWAN ECOLABEL 3041 0903

Lund University
Faculty of Science
Department of Physical Geography and Ecosystem Science

ISBN 978-91-89187-65-8