Mapping The Intellectual Landscape of Stunting Prediction Using Satellite Imagery and Machine Learning: A Bibliometric Analysis

Tigus Juni Betri tigusjuni.betri@staff.uinsaid.ac.id
Haya Nur Fadhilah fadhilahaya01@gmail.com
Anwar Rohmadi anwarrohmadi111@gmail.com
Sandy Fredella Elvaretta sandyfredella0@gmail.com
Deanova Cella Fadila deanovacellafadila27@gmail.com

Affiliation (Data Science Department, Islamic State University of Raden Mas Said, Indonesia)

ABSTRACT

Research on stunting risk prediction has become increasingly interdisciplinary, integrating demographic, environmental, and computational approaches. This study maps the global research landscape on stunting prediction using demographic data, satellite imagery, and machine learning, employing a hybrid methodology that combines a systematic scoping review and bibliometric analysis. Data from Scopus and Web of Science (248 publications, 2009–2025) were analyzed using Bibliometrix to examine publication trends, thematic structures, and collaboration patterns. Results show rapid growth after 2019, driven by the availability of demographic and remote-sensing data and advances in machine learning. Two dominant clusters were identified: a health–data cluster (machine learning, public health, malnutrition) and a socio-geospatial cluster (geography, remote sensing, poverty, environmental health). Thematic mapping indicates that medicine remains a basic theme, geography functions as a motor theme, and machine learning occupies a transitional position signifying ongoing methodological development. The scoping synthesis reveals a clear shift toward data-driven, spatially explicit research that integrates socioeconomic and environmental factors. However, gaps remain in model generalization, cross-regional comparison, and the inclusion of behavioral and climatic dimensions. This study provides the first comprehensive hybrid mapping of the field, illustrating its transition from health-based analyses to an interdisciplinary, data-science-driven paradigm. The findings offer a roadmap for researchers and policymakers to enhance collaboration, methodological rigor, and evidence-based actions aligned with Sustainable Development Goal 2 (Zero Hunger).

Keywords: Bibliometric Analysis; Global Public Health; Machine Learning; Satellite Imagery; Stunting; Scoping Review.

* Corresponding author: tigusjuni.betri@staff.uinsaid.ac.id

Introduction

Stunting remains one of the most pressing public health issues worldwide, particularly in lower-middle-income countries such as Indonesia. This condition is characterized by children's height being below the standard for their age due to chronic malnutrition during the first 1,000 days of life (HPK). Its impacts are multidimensional—limiting physical growth, cognitive development, productivity, and ultimately the quality of human capital in the future (WHO, 2021; Victora et al., 2017). According to the 2023 Indonesian Health Survey (SKI), the national prevalence of stunting remains high at 21.5 percent, underscoring the urgent need for data-driven innovations to accelerate targeted interventions and improve policy formulation.

In the past decade, research on stunting has shifted from conventional statistical analysis toward data-intensive and spatially explicit approaches. The integration of demographic data and satellite imagery offers a new opportunity to understand risk factors more comprehensively by linking household-level socioeconomic characteristics with environmental and infrastructural contexts (Engstrom et al., 2020; Shen et al., 2023). Combining these datasets with machine-learning algorithms enables more accurate spatial prediction of stunting risk and supports evidence-based intervention planning.

Conceptual Framework

This study is built on four conceptual pillars: stunting as the target variable, demographic data and satellite imagery as predictor data sources, and machine learning as the analytical framework.

Stunting is defined as chronic growth failure in children caused by prolonged undernutrition and related socioeconomic and environmental factors (Anjani, Nurhayati & Immawati, 2022). In this research, it serves as the dependent variable for predictive analysis. Demographic data capture socioeconomic characteristics such as education, employment, income, household size, and access to basic services (Beal, 2024; Adhila et al., 2023). These variables act as proximal predictors that represent household social conditions. Satellite imagery provides spatial information about land cover, vegetation, settlements, and infrastructure, functioning as a distal predictor that reflects environmental conditions influencing nutrition outcomes (Lillesand et al., 2015; Engstrom et al., 2020).

Machine learning is a branch of artificial intelligence that enables computers to learn from data without explicit instructions (Sarker, 2021). With a foundation in mathematics, statistics, and data mining, ML is capable of modeling nonlinear relationships among heterogeneous variables. This approach is considered ideal for integrating demographic data with satellite imagery to accurately predict stunting risk (Shen et al., 2023). Within this framework, demographic and satellite-based indicators are conceptualized as integrated inputs for data-driven models that can produce more precise, spatially explicit risk assessments of stunting.

Previous Research and Research Gaps

Machine learning applications for stunting prediction have developed rapidly. Sharma et al. (2024), through the *KidSat* project, introduced a global benchmark combining household surveys with high-resolution satellite imagery to predict child welfare indicators. Hasdyna et al. (2024) developed a hybrid ML framework in Aceh, Indonesia, improving classification accuracy and facilitating priority-area identification. Nduwayezu et al. (2024) applied a spatial ML approach in Rwanda to assess socioeconomic and climatic determinants of child growth, highlighting the influence of rainfall and vegetation.

Despite these advances, three major research gaps remain. First, methodological standards for integrating demographic and spatial data are still inconsistent, complicating cross-study comparison (Sharma et al., 2024). Second, most analyses remain localized and lack model generalization across countries (Hasdyna et al., 2024). Third, existing reviews rarely combine thematic and methodological perspectives, leaving limited understanding of how this interdisciplinary field evolves (Nduwayezu et al., 2024).

To address these gaps, this study applies a bibliometric analysis using R (Bibliometrix package) to quantitatively map global research trends, thematic structures, and collaboration networks related to stunting prediction through demographic and satellite data integrated with machine learning. The results aim to provide a global-scale knowledge map and evidence to support data-driven strategies for achieving Sustainable Development Goals (SDG 2 and 3).

Method

This research adopts a bibliometric analysis approach using the *Bibliometrix* package in R to explore the intellectual structure, thematic evolution, and publication trends in studies on stunting-risk prediction involving demographic data, satellite imagery, and machine learning. Bibliometric analysis is well-established for quantitatively examining large bodies of literature and identifying emerging research patterns and influential contributions (Donthu et al., 2021).

Research Design

Bibliographic data were retrieved from the Scopus and Web of Science databases, covering the period 2009 to 2025. The search string included keywords such as *"stunting," "malnutrition," "machine learning," "satellite imagery," "remote sensing,"* and *"demographic data."* All retrieved records were exported in BibTeX format and merged using R. After data cleaning and deduplication, a total of 248 publications were retained for analysis.

Data Analysis Procedures

The cleaned dataset was analyzed using the *Bibliometrix* and *Biblioshiny* packages in R (version 4.x). Several bibliometric indicators were calculated to quantitatively describe and map the structure of the research field. The analysis began with the computation of descriptive metrics that included annual scientific production, the most-cited authors, and the main journals contributing to the topic. Subsequently, a keyword co-occurrence analysis was conducted to identify thematic clusters and dominant research topics, while a thematic map and evolution analysis were used to observe the development and centrality of research themes over time. In addition, a collaboration network analysis was performed to examine the relationships among authors and the distribution of research across different countries.

All visualizations, including trend graphs, keyword co-occurrence networks, and thematic maps, were generated directly within the R environment. This quantitative mapping approach provides a comprehensive understanding of both the macro-level evolution of the field and the micro-level linkages between themes, authors, and institutions. The analytical workflow in this study followed a systematic sequence consisting of data retrieval, data cleaning, bibliometric computation, visualization, and interpretation.

Results and Discussion: Mapping the Intellectual Landscape of Stunting Prediction

Publication growth and geographic distribution

The cleaned bibliometric dataset contained 248 articles on stunting-risk prediction published between 2009 and 2025. Publication output remained minimal until 2016 (≤2 papers per year) but began to increase rapidly after 2017. From just 15 articles in 2019, output more than doubled to 24 papers in 2020 and reached 39 papers in 2021. The growth continued through 2024 (48 papers) and 2025 (46 papers), illustrating how research interest has accelerated since the late 2010s (Figure 1). This surge coincides with greater availability of demographic surveys, high-resolution satellite imagery, and accessible machine-learning software.

The top contributing countries were the United States (49 papers), Indonesia (41 papers) and Ethiopia (29 papers), followed by China (24), the United Kingdom (23) and Australia (15). The combination of high-income countries with lower- and middle-income countries underlines the global relevance of stunting and the collaborative nature of this field. Open-access journals dominated the corpus: PLOS ONE (33 papers), BMC Public Health (14) and Nutrients (9) were the most frequent outlets, indicating that researchers seek wide dissemination for studies addressing public health and development issues.

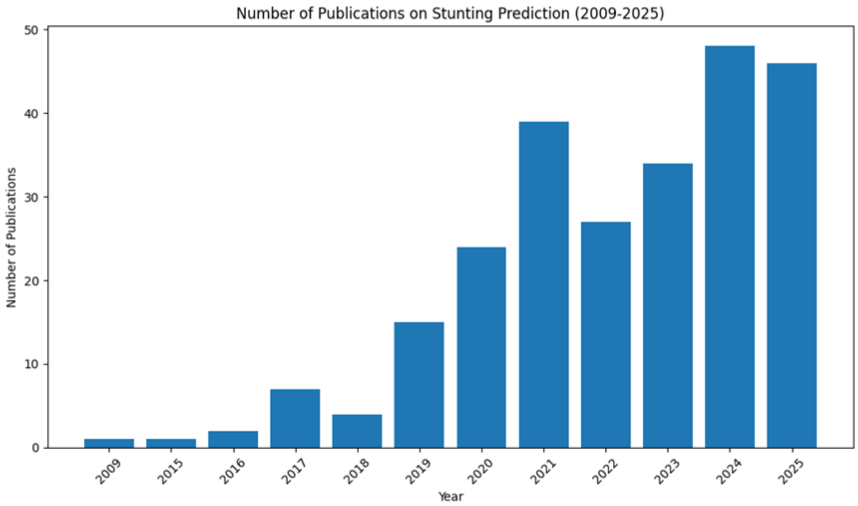
Descriptive trends in the literature

Growth of publications

Analysis of the bibliographic dataset (248 records from 2009–2025) shows rapid growth after 2019. Figure 1 summarises the yearly output. Only one article appeared before 2015, but the annual count reached 48 papers in 2024 and 46 papers in 2025. This surge reflects the increasing availability of demographic data, remote-sensing imagery, and computational tools, enabling interdisciplinary studies of malnutrition prediction. The underlying paper notes that stunting risk prediction has evolved toward a multidisciplinary approach that combines demographic health data with environmental information from satellite imagery, analyzed using machine learning.

Figure 1. Growth of Publications (2009–2025)

The number of publications per year shows a significant increase after 2019, indicating rapid expansion of interdisciplinary research in stunting prediction.



Evolving thematic emphasis

Trends in core keywords

The selected keywords (Figure 2) illustrate a shift from purely health-oriented terms to computational and spatial terms. In the early years, keywords such as malnutrition, demography and environmental health appeared sporadically. Starting in 2019, there is a pronounced rise in machine learning, artificial intelligence and computer science, reflecting the adoption of predictive algorithms. By 2024–2025, machine-learning keywords occur in more than 15 papers per year, overtaking many traditional health terms. Meanwhile, geography—representing the spatial context—is increasingly prominent; its frequency jumps from a handful of papers in 2017 to ~25 papers in 2024. These trajectories indicate that research on stunting prediction has evolved from biomedical analyses to interdisciplinary, data-science-oriented approaches that integrate socioeconomic and geospatial determinants.

Keyword co-occurrence network

The co-occurrence network (Figure 3) reveals two dominant clusters:

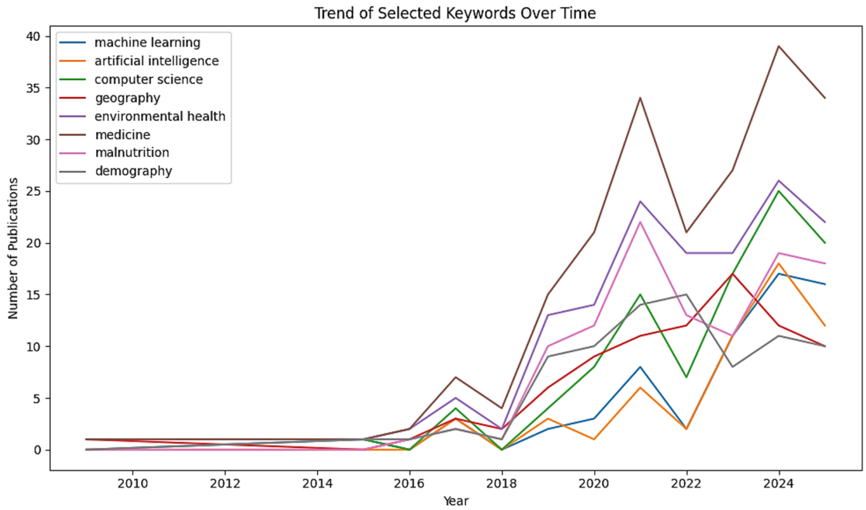
1. Health–data cluster (red nodes). This dense cluster contains terms related to public health (e.g., *public health*, *underweight*), clinical measurement (e.g., *body mass index*, *anthropometry*), and methodological tools (e.g., *logistic regression*, *receiver operating characteristic*). Keywords such as medicine, malnutrition and machine learning act as hubs. Their central positions show that quantitative modelling techniques now intersect with traditional health metrics.

2. Socio-geospatial cluster (blue nodes). This cluster includes geography, remote sensing, poverty, developing country and political science. The presence of *economics*, *economic growth* and *computer security* highlights the expanding scope of stunting research, which now considers socioeconomic drivers, environmental variables, and even infrastructure vulnerability.

The grey edges linking the clusters indicate that studies increasingly bridge health and spatial/economic domains. For example, many recent papers combine household-survey variables with satellite-derived vegetation indices to model child growth outcomes. This interconnectivity reflects the interdisciplinary nature of emerging research programs.

Figure 2. Trend of Selected Keywords

The temporal development of key terms that mark a shift in focus from clinical studies to data- and spatial-based approaches



Co-Occurrence Network

Figure 3 shows the keyword co-occurrence network generated using Bibliometrix R package (corresponding to the user-supplied "Figure 3. Co-occurrence Network"). Nodes represent keywords and edges indicate that two terms co-occur in at least one publication. The size of a node corresponds to its overall frequency, while colors denote clusters detected by the software.

Two major clusters emerge:

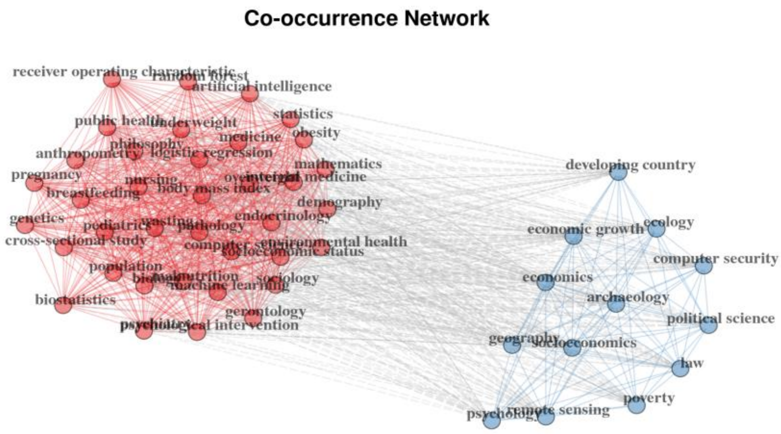
1. Health–data cluster (red). This dominant cluster includes keywords such as *machine learning*, *random forest*, *support vector machine*, *public health*, *nutritional status*, *logistic regression* and *malnutrition*. The strong interconnections indicate that researchers are applying advanced algorithms to clinical and epidemiological data. The dataset's abstract emphasizes that the rapid growth of literature has led to a rich but fragmented knowledge landscape, motivating the need to identify key trends and research gaps.

2. Socio-geospatial cluster (blue). This secondary cluster contains terms like *geography*, *remote sensing*, *economics*, *poverty*, *socio-economics*, *developing country*, *law* and *political science*. These words reflect the integration of socioeconomic and spatial perspectives. The network shows that geospatial terms link to health terms via edges, underscoring collaboration across fields to map stunting risk in developing regions.

Smaller clusters (not coloured due to overlap) include *anthropometry*, *nursing*, *psychology*, *computer security* and *ecology*. Their connections suggest emerging niches combining biological anthropology, behavioural science and computing. Overall, the network underscores that stunting-prediction research is inherently multidisciplinary, with machine learning techniques bridging health and socio-environmental domains.

Figure 3. Co-Occurrence Network of Keywords

The keyword network shows two main clusters: health–data and socio-geospatial, with increasingly strong cross-field connections.



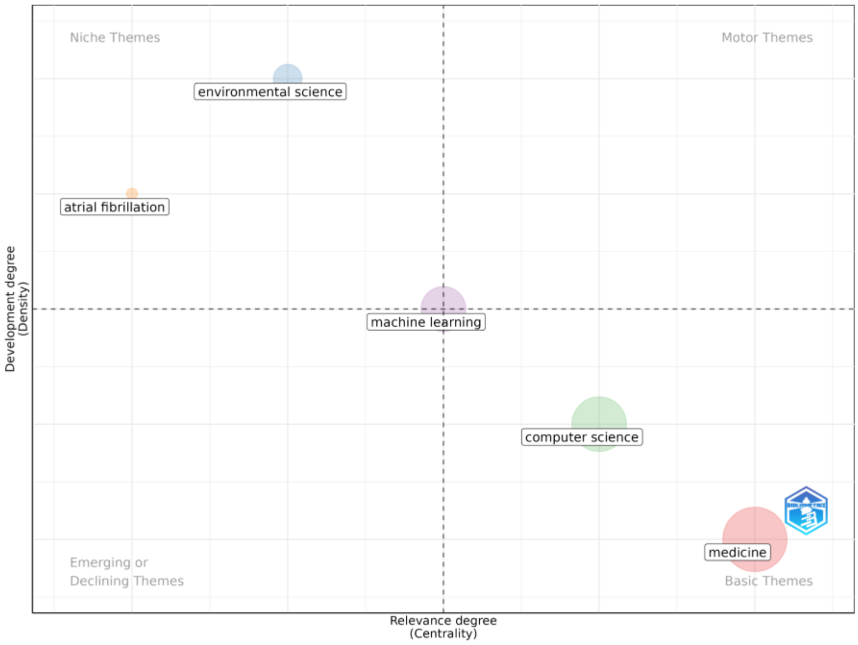
Thematic map and maturity of topics

The thematic map (Figure 4) positions topics according to their relevance (centrality) and degree of development (density). Four quadrants represent different roles in the intellectual structure:

1. Basic themes (lower right) refer to themes with high centrality and low density. The term *medicine* lies in this quadrant, meaning it is a basic but less internally developed theme—an essential foundation for the field.
2. Motor themes (upper right) represent themes with both high centrality and high density. *Computer science* sits closer to this quadrant, indicating that computational methods are not only central but also increasingly cohesive.
3. Niche themes (upper left) are characterized by high density but low centrality. *Environmental science* appears in this quadrant, implying a specialized but well-developed subfield that may inspire cross-fertilisation. A few isolated topics (e.g., *atrial fibrillation*) also appear here with limited ties to other themes.

4. Emerging or declining themes (lower left) generally have low centrality and density. *Machine learning* occupies a transitional position near the midlines in this area; it is central to the field but still evolving, suggesting active methodological experimentation.

Figure 4. Thematic Map of Research Areas



The thematic map (Figure 4) positions topics according to their relevance (centrality) and degree of development (density). Medicine lies in the lower-right quadrant, meaning it is a basic but less internally developed theme—an essential foundation for the field. Computer science sits closer to the motor-themes quadrant, indicating that computational methods are not only central but also increasingly cohesive. Machine learning occupies a transitional position near the midlines; it is central to the field but still evolving, suggesting active methodological experimentation. Environmental science appears in the niche-themes quadrant (high density, low centrality), implying a specialized but well-developed subfield that may inspire cross-fertilisation. A few isolated topics (e.g., *atrial fibrillation*) appear in the niche area with limited ties to other themes.

Discussion and interpretation

The bibliometric evidence demonstrates a rapid transition of stunting-risk prediction research from isolated health studies to a data-driven, interdisciplinary paradigm. Several factors drive this shift:

1. Technological advances: The ready availability of satellite imagery, global positioning, and computing resources has enabled researchers to incorporate environmental and infrastructural variables into malnutrition studies. Recent work shows that machine-learning algorithms can handle high-dimensional and non-linear relationships, outperforming traditional statistical models. Techniques such as random forests, support-vector machines, and gradient

boosting are now commonly applied alongside feature-selection methods to optimize predictive performance.

2. Epidemiological urgency: Despite slight improvements, the prevalence of stunting in Indonesia remains high. A national study reported that stunting decreased only 0.1 percentage point from 2022 to 2023, reaching 21.5%. This slow progress, attributed partly to the COVID-19 pandemic, has spurred policy makers and researchers to explore innovative approaches for identifying high-risk communities and targeting interventions more effectively.
3. Interdisciplinary collaboration: The strong contribution from both high-income (e.g., the United States) and low- to middle-income countries (e.g., Indonesia, Ethiopia) highlights extensive international cooperation. Integration of remote-sensing specialists, demographers, and machine-learning experts with public-health researchers has allowed for a more comprehensive understanding of stunting's multifactorial causes.

The specific bibliometric dimensions of this transition are detailed in the following sections.

Multidisciplinarity and integration

Both the co-occurrence network and thematic map reveal a high degree of multidisciplinarity. Keywords from medicine, public health, and epidemiology dominate the corpus, yet they are closely linked to geography, environmental science, and data-science terms. This alignment mirrors the interdisciplinary approach combining demographic health data with ecological information and machine learning. The integration of socioeconomic factors (e.g., poverty, economic growth) indicates that stunting is studied not merely as a medical condition but as a complex phenomenon influenced by social determinants and spatial contexts.

Rise of data-driven methodologies

The temporal analysis shows a sharp increase in publications containing machine learning and artificial intelligence after 2017 (Figure 2). This corresponds to the network's health data cluster, where ML algorithms such as random forests and support vector machines are prominent. These methods are increasingly used to model non-linear relationships among demographic, environmental, and socioeconomic variables, enabling predictive mapping of stunting risk. However, the thematic map positions machine learning as a transitional theme, reflecting that methodological rigor and reporting standards are still developing. Many studies adopt existing algorithms but provide limited innovation in model design or validation. Further work is needed to establish reproducible workflows, assess model generalisability across regions, and incorporate fairness considerations.

Emerging geospatial and environmental emphasis

The appearance of geography as a motor theme suggests that spatial analysis has matured into a key driver of research. Remote sensing, geographic information systems, and spatial statistics enable researchers to map malnutrition prevalence at fine scales and relate it to environmental factors such as vegetation, climate, and infrastructure. The presence of environmental science as a niche theme indicates that ecological determinants are receiving focused attention. Integrating these data sources can identify environmental risks and inform targeted interventions.

Socio-economic and policy dimensions

The blue cluster in the co-occurrence network connects poverty, economic growth, law, political science, and developing country to stunting. This shows that scholars are examining how economic development, governance, and social policies influence malnutrition. Such analyses are vital for designing context-appropriate interventions, aligning with Sustainable Development Goal 2 (Zero Hunger).

Emerging patterns and gaps

The bibliometric analysis highlights several critical directions for future research:

1. Integration of demographic and spatial data: Many post-2019 studies link household characteristics with satellite-derived indices (e.g., vegetation cover, access to roads and water), enabling spatially explicit risk mapping. However, methodological standards vary widely, complicating cross-study comparability. Harmonising variable definitions and model-evaluation metrics will be crucial for generalising results across regions.
2. Cross-regional generalisation: Most models are built on data from a single country or sub-region; relatively few papers conduct comparative analyses across countries. Given the heterogeneity of socioeconomic and climatic conditions, transferring models without adaptation may produce biased estimates. Future research should explore domain adaptation and meta-learning techniques to improve generalisability.
3. Behavioural and climatic dimensions: Thematic clusters highlight environmental and economic factors, yet behavioural variables (e.g., maternal feeding practices) and extreme climate events receive limited attention in the bibliometric corpus. Incorporating seasonal climate indicators and behavioural surveys could enrich models and enhance early-warning capabilities.
4. Policy-relevant communication: Although machine-learning models can achieve high accuracy, they must translate into actionable insights. Explainable AI tools (e.g., SHAP values) can help identify the most influential predictors and facilitate communication with non-technical stakeholders.

Conclusion

This hybrid bibliometric–scoping review underscores how stunting-risk prediction research is undergoing a paradigm shift. The field has expanded from small-scale, health-centric analyses to an interdisciplinary space that leverages demographic surveys, satellite imagery and machine-learning algorithms. While publication output has grown rapidly, further methodological harmonisation, cross-regional validation and integration of behavioural and climate data are needed. By addressing these gaps, researchers and policy makers can develop more accurate and equitable early-warning systems, supporting evidence-based interventions that advance progress toward Sustainable Development Goal 2 (Zero Hunger).

References

Adhila, F., Yulia, Y., & Aris, F. (2023). A systematic review of socioeconomic and demographic determinants of stunting in developing countries. *PLOS ONE, 18*(9), e0290962. https://doi.org/10.1371/journal.pone.0290962

Anjani, D. M., Nurhayati, S., & Immawati. (2024). Penerapan pendidikan kesehatan terhadap pengetahuan ibu tentang stunting pada balita di wilayah kerja UPTD Puskesmas Rawat Inap Banjarsari Metro Utara. *Jurnal Cendikia Muda, 4*(1), 62-69. Retrieved from https://jurnal.akperdharmawacana.ac.id/index.php/JWC/article/view/564

Beal, V. (2024). *Demographic Data*. Techopedia. Diperoleh dari https://www.techopedia.com/definition/demographic-data

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research, 133*, 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Engstrom, R., Hersh, J., & Newhouse, D. (2020). The Use of Satellite Data to Measure and Understand Development Outcomes. *The World Bank Research Observer, 36*(1), 1-32. https://doi.org/10.1093/wbro/lkaa007

Hasdyna, N., Dinata, R. K., Rahmi, & Fajri, T.I. (2024). Hybrid Machine Learning for Stunting Prevalence: A Novel Comprehensive Approach to Its Classification, Prediction, and Clustering Optimization in Aceh, Indonesia. *Informatics, 11*(4), 89. https://www.mdpi.com/2227-9709/11/4/89

Lillesand, T., Kiefer, R. W., & Chipman, J. (2015). *Remote Sensing and Image Interpretation* (7th ed.). John Wiley & Sons. https://www.wiley.com/en-us/Remote+Sensing+and+Image+Interpretation%2C+7th+Edition-p-9781118343289

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *El Profesional de la Información, 29*(1), e290103. https://doi.org/10.3145/epi.2020.ene.03

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, Rwanda. *GeoHealth, 8*(1), e2023GH001234. https://doi.org/10.1029/2024GH001027

Sarker, I. H. (2021). Machine learning: Algorithms, real-world application,s and research directions. *SN Computer Science, 2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Sharma, M., Yang, F., Vo, D.-N., Suel, E., Mishra, S., Bhatt, S., Fiala, O., Rudgard, W., & Flaxman, S. (2024). KidSat: Satellite imagery to map childhood poverty dataset and benchmark. arXiv Preprint, arXiv:2407.05986. https://arxiv.org/abs/2407.05986

Shen, H., Zhao, H., & Jiang, Y. (2023). Machine Learning Algorithms for Predicting Stunting among Under-Five Children in Papua New Guinea. *Children, 10*(10), 1638. https://www.mdpi.com/2227-9067/10/10/1638

Victora, C. G., Christian, P., Vidaletti, L. P., Gatica-Domínguez, G., Gubert, M. B., & Black, R. E. (2017). Revisiting maternal and child undernutrition in low-income and middle-income countries: variable progress and district-level disparities. *The Lancet, 391*(10118), P344-359. https://doi.org/10.1016/S0140-6736(17)31544-X

World Health Organization (WHO). (2021). *Levels and trends in child malnutrition: Key findings of the 2021 edition of the joint child malnutrition estimates*. WHO/UNICEF/World Bank Group. https://www.who.int/publications/i/item/9789240025257