Journal of Information Systems and Informatics
Vol. 7, No. 3, September 2025 e-ISSN: 2656-4882 p-ISSN: 2656-5935
DOI: 10.51519/journalisi.v7i3.1248 Published By DRPM-UBD

Global Research Trends and Map on Machine Learning Applications in Stunting Detection in Vulnerable Populations: A Bibliometric Analysis

Otong Saeful Bachri[1], Catur Edi Widodo[2], Oky Dwi Nurhayati[3]

[1,2,3]Postgraduate School, Diponegoro University, Semarang, Indonesia
[1]Informatic Engineering, University Muhadi Setiabudi, Brebes, Indonesia
Email: [1]otongsaefulbachri@students.undip.ac.id, [2]caturediwidodo@lecturer.undip.ac.id

Abstract

Stunting and malnutrition continue to be significant public health challenges, particularly in low-income and rural populations. With the growing reliance on data-driven strategies in public health, machine learning (ML) has emerged as a promising tool for identifying, classifying, and predicting conditions related to undernutrition. This study presents a bibliometric analysis of global research from 2019 to 2025, focusing on the application of ML techniques—such as clustering, support vector machines (SVM), and random forest—in addressing malnutrition and stunting. A total of 417 Scopus-indexed publications were analyzed using Biblioshiny (R) to assess research trends, key themes, influential authors, prominent journals, and thematic evolution. The analysis reveals a consistent growth rate of 10.72% in publications, with notable contributions from China and other low- and middle-income countries. Keyword mapping highlights that "machine learning," "spatial analysis," and "stunting" are central to the research, although they remain areas for further development. Thematic evolution indicates a shift towards more integrated, context-aware approaches, with a growing focus on built environments and vulnerable populations. The study concludes that while ML holds significant promise for advancing decision-making in child health and nutrition, its impact will depend on continued methodological refinement and effective implementation within public health systems.

Keywords: Machine Learning, Stunting, Malnutrition, Public Health, Bibliometric Analysis

1. INTRODUCTION

Malnutrition and stunting continue to be among the most pressing public health challenges affecting children in developing countries [1], [2], [3]. These conditions, which are particularly widespread in low-income and rural areas, have significant implications for childhood morbidity, cognitive development, and long-term socioeconomic disadvantage [4], [5]. In recent years, the growing availability of health-related datasets, coupled with advancements in computational technology, has sparked the integration of machine learning (ML) techniques into child health research [6], [7], [8], [9]. ML offers powerful tools for

uncovering hidden patterns, classifying risk groups, and predicting malnutrition outcomes with remarkable accuracy and scalability [10], [11], [12].

A growing body of literature has explored the application of ML techniques—such as support vector machines (SVM), decision trees, clustering algorithms, and ensemble learning methods—to identify key risk factors associated with stunting and undernutrition [13], [14], [15]. These studies typically leverage data from national demographic and health surveys (DHS), satellite imagery, and administrative records to build predictive models and generate spatial risk maps [12], [13]. Additionally, clustering methods like K-means and hierarchical clustering have been utilized to analyze patterns of food insecurity, environmental exposure, and child health disparities across regions [16], [17].

While there has been increasing interest in applying ML to the study of child malnutrition and health, there remains a gap in comprehensive bibliometric analyses of this research area. Understanding the evolution of research trends, key contributors, collaboration networks, and publication outlets is crucial for shaping future research directions and policy recommendations. This study addresses this gap by conducting a bibliometric analysis of global scholarly publications from 2019 to 2025, focusing on the application of ML and clustering methods in child nutrition and health, using Scopus-indexed literature.

This bibliometric analysis is guided by several research questions: RQ1: What are the trends in the number of publications on machine learning and child health between 2019 and 2025? RQ2: What are the most common keywords and thematic areas in this body of literature? RQ3: Who are the most productive authors, and which institutions are the most active in this field? RQ4: Which journals and publication venues are most frequently chosen for disseminating this research? RQ5: What collaboration patterns exist among authors and institutions in this area? By answering these questions, this study aims to provide a comprehensive overview of how ML techniques are influencing the research landscape of child malnutrition and public health, thus helping to shape future academic inquiry and policy initiatives.

2. METHODOLOGY

This study employs a bibliometric analysis approach to systematically map and quantify the research landscape concerning the application of machine learning (ML) in addressing malnutrition and stunting among children in vulnerable settings. Bibliometric analysis offers a robust methodology for identifying publication trends, research hotspots, influential contributors, and collaboration networks, thereby providing insights into the development and evolution of research within a specific domain.

2.1 Data Source and Search Strategy

The bibliographic data for this study were retrieved from the Scopus database, renowned for its comprehensive coverage of peer-reviewed scientific literature across various disciplines. The search query was designed to capture relevant publications by combining three thematic keyword clusters using Boolean operators.

1) Group A (Machine Learning Methods): Keywords such as "machine learning", "clustering", "k-means", "SVM" (support vector machines), and "particle swarm optimization" were included to capture the range of ML techniques applied in the context of malnutrition and stunting research.
2) Group B (Health Issues): Keywords like "malnutrition", "stunting", "child nutrition", and "public health" were included to ensure that the publications related to the specific health issues of interest were included in the dataset.
3) Group C (Target Population): Terms such as "developing countries", "low-income population", "rural areas", and "vulnerable population" were used to focus on studies relevant to the target populations in need of intervention.
4) Using this strategy, a dataset of 417 Scopus-indexed publications was retrieved, which were then exported in CSV format for further analysis and processing. This search strategy ensured that the data encompassed a broad but relevant range of publications in the intersection of machine learning, child malnutrition, and vulnerable populations.

2.2 Analysis Tool and Procedure

The analysis was conducted using Biblioshiny, the graphical interface for the Bibliometrix R package [18], which allows for advanced science mapping and citation analysis. Biblioshiny was chosen due to its capability to perform detailed co-authorship network visualizations, keyword co-occurrence clustering, and thematic mapping, making it ideal for this bibliometric analysis. The steps followed for the analysis are as follows:

1) Data Import and Cleaning: The exported CSV file was imported into Biblioshiny, where initial data cleaning steps were conducted. This included deduplication to ensure that only unique records were retained and keyword normalization to account for variations in terminology (e.g., singular vs. plural terms or synonyms).
2) Descriptive Indicators Extraction: Descriptive analysis was performed to extract key indicators, including:
    a) Annual scientific production: Number of publications per year to identify trends in research output over time.
    b) Most relevant sources: Identification of the leading journals and conferences where related research is published.

c) Top authors and key institutions: A ranking of the most productive authors and research institutions contributing to the field.

3) Co-occurrence Analysis of Keywords: Co-occurrence analysis was performed to identify thematic structures within the research literature. This analysis assessed the frequency with which specific keywords appeared together in publications, providing insights into the key themes and emerging areas of focus in the research on ML and child malnutrition.
4) Trend and Relationship Visualization: To gain a deeper understanding of the relationships between various aspects of the research, thematic maps, three-field plots, and network diagrams were generated. These visualizations helped illustrate the evolving research trends, thematic clusters, and collaboration patterns within the field.

2.3 Data Inclusion and Exclusion Criteria

For the purposes of this study, only original research articles and conference papers published in English were included in the analysis. Reviews, editorials, and incomplete entries were excluded to maintain the quality and relevance of the dataset. This selection process ensures that the analysis focuses on primary research that directly contributes to the understanding of machine learning applications in addressing childhood malnutrition and stunting in low-resource settings.

2.4 Reproducibility and Transparency

The methodology used in this study is designed to be both replicable and transparent. The clear steps involved in the data retrieval, cleaning, and analysis processes ensure that the findings can be independently verified and extended by other researchers. This approach not only provides a comprehensive overview of the current state of research but also sets the groundwork for future studies exploring the intersection of ML, child health, and vulnerable populations.

3. RESULTS AND DISCUSSION

From 2019 to 2025, a total of 417 documents were published across 239 sources, reflecting an annual growth rate of 10.72%. This consistent increase in publications indicates a growing academic interest in applying machine learning (ML) techniques to the study of child malnutrition and public health. As shown in Table 1, this trend demonstrates the expanding scope of research in this field.

Table 1. Overview of Main Bibliometric Indicators

Description	Results
Timespan	2019–2025
Sources (Journals, Books, etc.)	239

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Description	Results
Documents	417
Annual Growth Rate %	10.72
Document Average Age	2.55
Average Citations per Document	11.45
Author's Keywords (DE)	1,380
Authors	2,462
Authors of Single-Authored Docs	6
Co-Authors per Doc	6.5

The documents are relatively recent, with an average age of 2.55 years, and have received an average of 11.45 citations each, suggesting moderate scholarly impact. The field is highly collaborative, involving 2,462 authors, with an average of 6.5 co-authors per paper and only 6 single-authored publications. This highlights the interdisciplinary nature of research in this area. Additionally, the presence of 1,380 unique author keywords reflects a broad thematic coverage, encompassing both methodological terms (e.g., ML, clustering) and domain-specific keywords (e.g., stunting, health), further demonstrating the integrative character of the research in this domain.

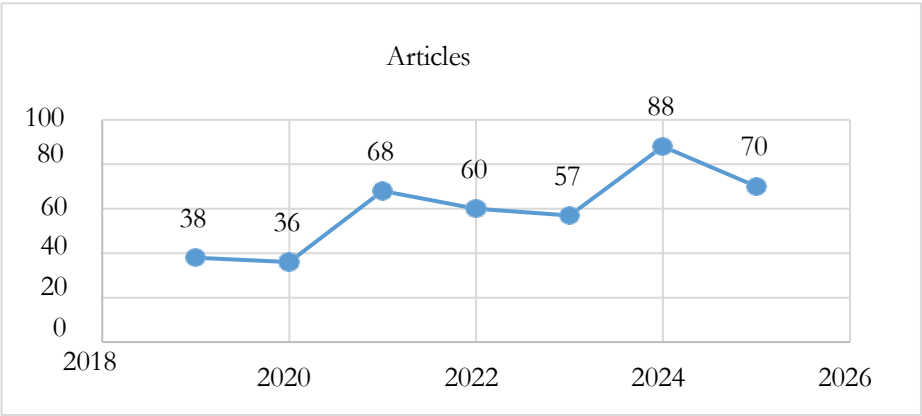

Figure 1. Annual number of publications (2019–2025)

3.1. Publication Trends

The annual number of publications reveals a clear upward trend (Figure 1). The number of publications increased from 38 in 2019 to a peak of 88 in 2024, with 70 publications in 2025 (partial year), indicating sustained interest and growth in this research area. As depicted in Figure 1, the growth pattern reflects an increasing recognition of the relevance of ML applications in addressing health issues in low-resource settings. For instance, Carrillo-Larco et al. highlight the rise of ML techniques in global health research, particularly in low- and middle-

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

income countries [19]. This sustained growth is supported by the increasing relevance of ML to public health challenges, as evidenced by the rising number of publications year on year.

3.2. Top Keywords

A total of 1,380 unique author keywords were identified across the publications. As shown in Figure 2, the most frequently occurring keywords include "machine learning" (83 occurrences), "public health," "COVID-19," "epidemiology," and "random forest." The prominence of terms such as "machine learning" and "random forest" underscores the methodological focus of these studies. Notably, health-related keywords like "malnutrition," "stunting," "anemia," and "obesity" are also highly prevalent, indicating the domain-specific applications of these methods. These results highlight the dual focus on both the technical aspects of machine learning and its application to pressing health issues. The co-occurrence network (Figure 2) further illustrates how keywords related to ML methods are intertwined with key health-related topics, emphasizing the interdisciplinary nature of the research.

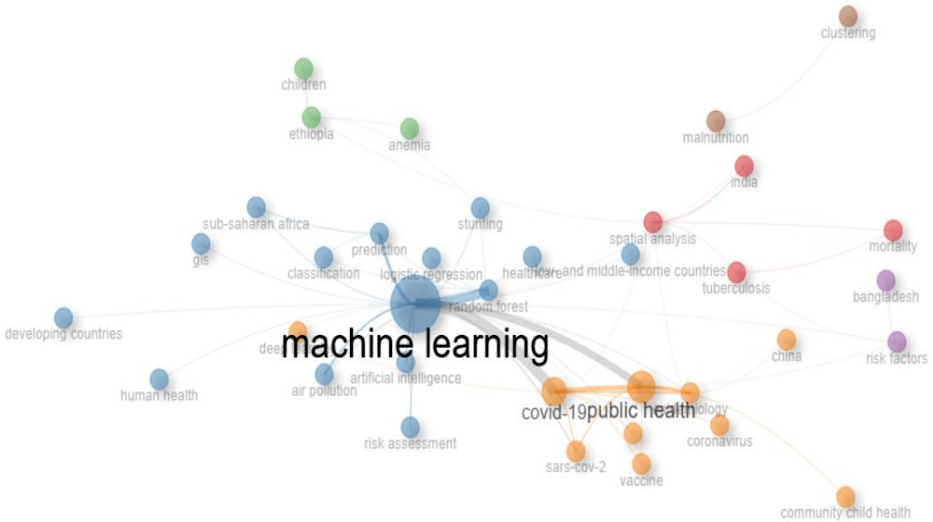

Figure 2. Co-occurrence Network

3.3. Author-Country-Journal Relationship (Three-Field Plot Analysis)

To better understand the relationships among contributing authors, countries, and publication outlets, a three-field plot was generated (Figure 3). This plot reveals the connections between countries (AU_CO), authors (AU), and source journals (SO), highlighting the global spread and academic collaboration in this research field. As shown in Figure 3, China dominates the research landscape in this domain, contributing the highest number of publications. Other prominent

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

contributors include the United States, Canada, Australia, the United Kingdom, Spain, South Korea, and Bangladesh. The central part of the plot reveals key authors such as Li Z, Wang J, Liu Y, Liu X, and Zhang Y, who are frequently associated with publications from China. This pattern suggests that a significant portion of scholarly output on ML applications in child health and malnutrition is concentrated in Chinese research clusters.

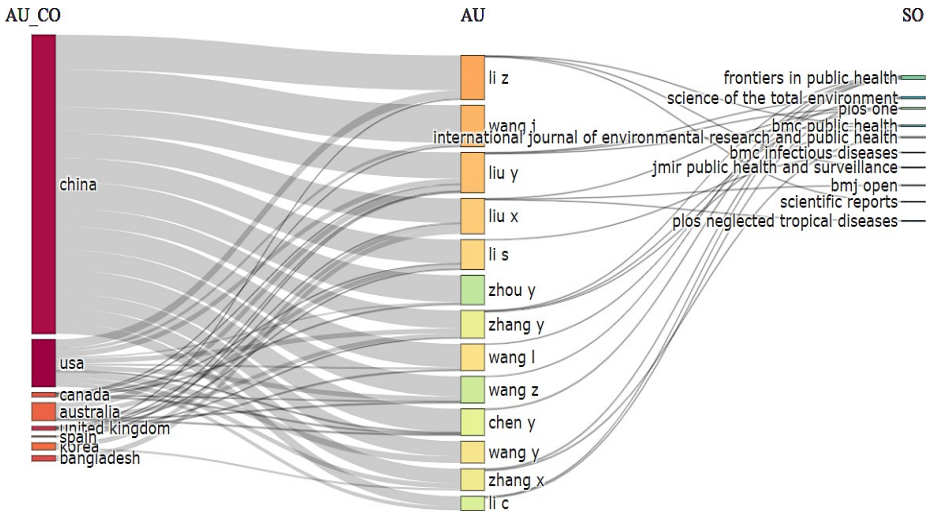

Figure 3. Three-Field Plot Analysis

On the right side of the plot, the distribution of publications across journals is visible. The *International Journal of Environmental Research and Public Health* (IJERPH) stands out as the leading outlet for these studies, with other notable journals including *Frontiers in Public Health*, *Science of the Total Environment*, *BMJ Public Health*, *BMC Public Health*, *BMC Infectious Diseases*, and *JMIR Public Health and Surveillance*. These journals focus on public health, environmental health, and interdisciplinary research, making them well-suited for disseminating studies that integrate ML with nutrition and epidemiology.

3.4. Thematic Analysis

To gain deeper insights into the conceptual structure of the research, author keywords were analyzed using co-word analysis, and the results were visualized through a thematic map (Figure 4) and thematic evolution diagram (Figure 5). The thematic map (Figure 4) identifies clusters of keywords based on their centrality (relevance) and density (development level), revealing the key thematic areas within the literature. The thematic map categorizes the literature into four major themes:

1) Basic Themes (bottom-right): These are core but still developing topics such as "machine learning," "random forest," "spatial analysis," "stunting,"

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

and "anemia." These foundational areas are well-established but open to further methodological expansion.

2) Motor Themes (top-right): These themes are both well-developed and highly relevant, such as "public health," "epidemiology," and "low- and middle-income countries," indicating the maturity and interdisciplinary relevance of these topics.
3) Niche Themes (top-left): These specialized areas include "community health," "spatial clustering," and "breast cancer," which are methodologically rich but have limited broader impact.
4) Emerging or Declining Themes (bottom-left): Topics such as "health equity," "large language models," and "accessibility" appear to be emerging or declining in focus, with low centrality and density.

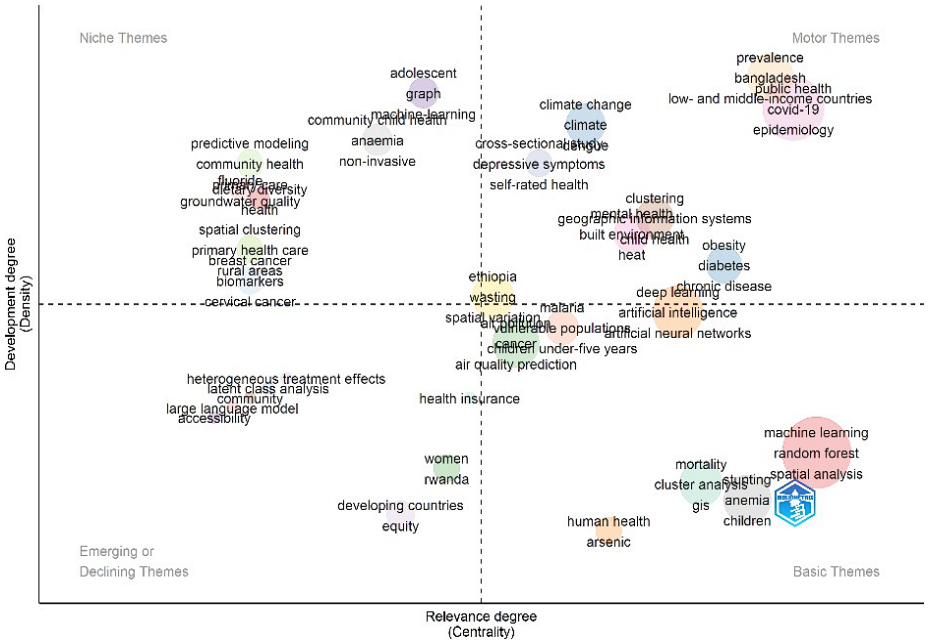

Figure 4. Thematic Map

The thematic evolution diagram (Figure 5) tracks the progression of major keywords from 2019 to 2025. In the earlier period (2019-2020), dominant themes included "epidemiology," "malnutrition," "mortality," "stunting," and "random forest." These foundational concerns represent early applications of ML in public health. In the later period (2021-2025), there was an expansion of topics such as "cluster analysis," "developing countries," "low- and middle-income countries," and "built environment." This shift suggests a broadening of the research focus, with increased attention on the contextual factors influencing health outcomes in underserved populations. The continuity of topics like "machine learning" and "malnutrition" across both periods highlights their sustained importance, while

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

the emerging focus on "public health" and "risk factors" suggests a trend toward more applied and policy-relevant research. This evolution reflects the growing recognition of the need for ML to address complex, real-world health disparities.

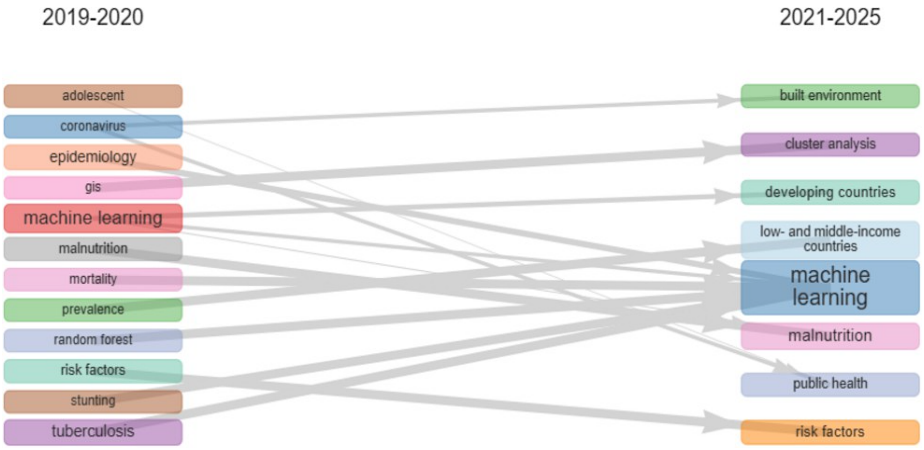

Figure 5. Thematic Evolution

3.5. Discussion

This bibliometric analysis underscores the growing significance of machine learning (ML) in addressing global public health challenges, particularly in tackling malnutrition and stunting among children. The steady annual growth rate of 10.72% in publications and the substantial co-authorship levels highlight an increasing academic interest and collaboration in this domain. The prominence of core keywords such as “machine learning,” “random forest,” “spatial analysis,” and “stunting” further emphasizes the centrality of computational approaches in diagnosing, predicting, and addressing undernutrition risks, reflecting a maturing field of study. The concentration of these keywords suggests that ML is rapidly becoming a foundational element in malnutrition research, providing both methodological tools and conceptual frameworks for understanding the complex, multifactorial nature of child health.

As indicated by the thematic map, ML-related terms such as "machine learning" and "random forest" are categorized as basic themes, which are still foundational but ripe for further exploration. These terms are central to the research landscape, signaling that ML methods have already gained widespread adoption in the field. However, there remains significant potential for deeper development, particularly in refining algorithms and integrating them with more comprehensive datasets. The recent thematic evolution, which highlights the increasing relevance of terms like “cluster analysis,” “built environment,” and “developing countries,” suggests a noteworthy shift from purely algorithm-focused research to more contextually aware models. This evolution points to the growing recognition of

the need to incorporate environmental, socioeconomic, and geographic factors into predictive models, making them more applicable to the complexities of malnutrition settings in diverse regions. This shift emphasizes the importance of adapting ML techniques to address real-world disparities in child health.

The dominance of authors and research institutions from China, along with a significant contribution from low- and middle-income countries (LMICs), signals a positive trend toward the democratization of AI-driven health research. This growing representation from LMICs highlights the global commitment to leveraging machine learning to address child malnutrition in resource-constrained environments, where the need for effective, scalable solutions is most urgent. This trend also points to the increasing inclusivity of AI-driven research, as it allows researchers from diverse geographic and economic contexts to contribute to the global discourse on malnutrition. However, despite the impressive growth in publications, the relatively moderate citation impact suggests that while the field is gaining momentum, it has yet to fully translate into widespread influence and implementation within policy and intervention systems. The concentration of publications in a small number of core journals further suggests that the field may benefit from greater diversification in publication venues to reach a broader audience.

Looking ahead, the field of ML applications in child malnutrition and public health offers promising opportunities for future development in at least three key areas. First, interdisciplinary integration stands out as a critical avenue for growth. Combining ML with data from social determinants of health, nutrition surveys, and geospatial analysis could result in more holistic, multi-dimensional models of malnutrition. Such models would account for the interconnectedness of health, environment, and socio-economic factors, providing a more comprehensive understanding of malnutrition in diverse contexts. Second, local adaptability is essential. There is a significant need to customize predictive tools for specific national health systems and sub-populations, particularly in rural and high-risk regions where the burden of malnutrition is most acute. Tailoring ML models to local contexts will enhance their relevance and effectiveness in real-world applications. Finally, explainability and decision support will be crucial to fostering trust in ML-driven solutions. Developing interpretable models that provide clear, actionable insights will ensure that machine learning is not only useful from a technical standpoint but also accessible and actionable for policymakers, healthcare providers, and other stakeholders working to improve child health outcomes.

The potential of machine learning to revolutionize our understanding, monitoring, and response to child undernutrition and stunting is immense, realizing this promise requires more than just technical innovation. It necessitates a concerted, interdisciplinary effort to bridge the gap between algorithmic

advancements and real-world applications. Collaboration across disciplines, institutions, and geographies is essential to ensure that the insights generated from ML models are effectively integrated into public health workflows and intervention systems. Only through sustained global cooperation can we ensure that these technological innovations translate into tangible improvements in health equity, particularly for vulnerable populations in low-resource settings.

4. CONSLUSION

This bibliometric analysis demonstrates the growing significance of machine learning (ML) in addressing child malnutrition and stunting, particularly in low-resource settings. The steady increase in publications, coupled with the interdisciplinary nature of the research, highlights the potential of ML to revolutionize our understanding and response to these critical public health issues. However, to fully realize this potential, it is essential to prioritize the integration of ML with contextual factors, ensure local adaptability, and enhance the explainability of models to facilitate their real-world application. Continued collaboration across disciplines, institutions, and geographies will be key to translating algorithmic insights into tangible health outcomes, advancing global efforts toward reducing child undernutrition and achieving greater health equity.

REFERENCES

[1] E. Darnila, Maryana, K. Mawardi, M. Sinambela, and I. Pahendra, "Supervised models to predict the stunting in East Aceh," *Int. J. Eng. Sci. Inf. Technol.*, vol. 2, no. 3, pp. 34–39, 2022, doi: 10.52088/ijesty.v1i4.280.

[2] H. H. Amare and B. Lindtjorn, "Concurrent anemia and stunting among schoolchildren in Wonago district in southern Ethiopia: A cross-sectional multilevel analysis," *PeerJ*, vol. 9, 2021, doi: 10.7717/peerj.11158.

[3] E. K. Anku and H. O. Duah, "Predicting and identifying factors associated with undernutrition among children under five years in Ghana using machine learning algorithms," *PLoS ONE*, vol. 19, no. 2, 2024, doi: 10.1371/journal.pone.0296625.

[4] C. Browne et al., "Multivariate random forest prediction of poverty and malnutrition prevalence," *PLoS ONE*, vol. 16, no. 9, 2021, doi: 10.1371/journal.pone.0255519.

[5] D. Brahma and D. Mukherjee, "Infant malnutrition, clean-water access and government interventions in India: a machine learning approach towards causal inference," *Appl. Econ. Lett.*, vol. 28, no. 16, pp. 1426–1431, 2021, doi: 10.1080/13504851.2020.1822507.

[6] M. S. Ali et al., "Spatial variation and determinants of underweight among children under 5 y of age in Ethiopia: A multilevel and spatial analysis based on data from the 2019 Ethiopian Demographic and Health Survey," *Nutrition*, vol. 102, 2022, doi: 10.1016/j.nut.2022.111743.

[7] H. Chen, J. Xing, X. Yang, and K. Zhan, "Heterogeneous effects of health insurance on rural children's health in China: A causal machine learning approach," *Int. J. Environ. Res. Public Health*, vol. 18, no. 18, 2021, doi: 10.3390/ijerph18189616.

[8] M. M. Khudri, K. K. Rhee, M. S. Hasan, and K. Z. Ahsan, "Predicting nutritional status for women of childbearing age from their economic, health, and demographic features: A supervised machine learning approach," *PLoS ONE*, vol. 18, no. 5, 2023, doi: 10.1371/journal.pone.0277738.

[9] T. Z. Yehuala et al., "Machine learning algorithms to predict healthcare-seeking behaviors of mothers for acute respiratory infections and their determinants among children under five in sub-Saharan Africa," *Front. Public Health*, vol. 12, 2024, doi: 10.3389/fpubh.2024.1362392.

[10] A. Y. Ambriola Oku, G. A. Zimeo Morais, A. P. Arantes Bueno, A. Fujita, and J. R. Sato, "Potential confounders in the analysis of Brazilian adolescent's health: A combination of machine learning and graph theory," *Int. J. Environ. Res. Public Health*, vol. 17, no. 1, 2019, doi: 10.3390/ijerph17010090.

[11] H. M. Fenta, T. Zewotir, and E. K. Muluneh, "A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, 2021, doi: 10.1186/s12911-021-01652-1.

[12] Md. M. Islam, N. Md. Shoukot Jahan Kibria, S. Kumar, D. C. Roy, and Md. R. Karim, "Prediction of undernutrition and identification of its influencing predictors among under-five children in Bangladesh using explainable machine learning algorithms," *PLoS ONE*, vol. 19, no. 12, 2024, doi: 10.1371/journal.pone.0315393.

[13] Md. M. Alam, A. I. Khan, A. Zafar, M. Sohail, M. T. Ahmad, and R. Azim, "Advancing nutritional status classification with hybrid artificial intelligence: A novel methodological approach," *Brain Behav.*, vol. 15, no. 5, 2025, doi: 10.1002/brb3.70548.

[14] T. Tamanna, S. Mahmud, N. Salma, Md. M. Hossain, and Md. R. Karim, "Identifying determinants of malnutrition in under-five children in Bangladesh: insights from the BDHS-2022 cross-sectional study," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-99288-y.

[15] C. Mullally, M. Rivas, and T. McArthur, "Using machine learning to estimate the heterogeneous effects of livestock transfers," *Am. J. Agric. Econ.*, vol. 103, no. 3, pp. 1058–1081, 2021, doi: 10.1111/ajae.12194.

[16] A. Mitra, "Selection of khadi fabrics for optimal comfort properties using multi-criteria decision-making technique," *Res. J. Text. Appar.*, vol. 27, no. 1, pp. 118–140, Jan. 2023, doi: 10.1108/RJTA-08-2021-0108.

[17] M. Cheung, A. Dimitrova, and T. Benmarhnia, "An overview of modern machine learning methods for effect measure modification analyses in high-dimensional settings," *SSM Popul. Health*, vol. 29, 2025, doi: 10.1016/j.ssmph.2025.101764.

[18] M. Aria and C. Cuccurullo, "Bibliometrix: An R-tool for comprehensive science mapping analysis," *J. Informetr.*, vol. 11, no. 4, pp. 959–975, Nov. 2017, doi: 10.1016/j.joi.2017.08.007.

[19] R. M. Carrillo-Larco, L. Tudor Car, J. Pearson-Stuttard, T. Panch, J. J. Miranda, and R. Atun, "Machine learning health-related applications in low-income and middle-income countries: A scoping review protocol," *BMJ Open*, vol. 10, no. 5, 2020, doi: 10.1136/bmjopen-2019-035983.

[20] T. Sugihartono, B. Wijaya, Marini, A. F. Alkayes, and H. A. Anugrah, "Optimizing stunting detection through SMOTE and machine learning: A comparative study of XGBoost, random forest, SVM, and k-NN," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 667–682, 2025, doi: 10.47738/jads.v6i1.494.