ACADEMY JOURNAL
OF
SCIENCE AND ENGINEERING
Available online @
www.academyjsekad.edu.ng
AJSE 19 (5) 2025
Original Research

DEVELOPMENT OF DATA-DRIVEN SYSTEM FOR EARLIER CHILDHOOD MALNUTRITION PREDICTION

Rebeka Samwel[1], Mrindoko Nicholaus[2]

[1] Department of Computer Science and Engineering Mbeya University of Science and Technology, Mbeya, Tanzania
[2] Department of Computer Science and Engineering Mbeya University of Science and Technology, Mbeya, Tanzania
Corresponding Authors E-mail: nicholausmrindoko@gmail.com
Corresponding Authors Mobile No: +255 714 427 241

Abstract

Childhood malnutrition remains a critical issue in Tanzania, impacting the health and development of children under five. Early detection and intervention are vital to mitigating its effects, yet they are often hampered by a lack of effective tools. This study addresses this challenge by developing a data-driven system that uses machine learning to predict malnutrition risks early in childhood. Unlike previous studies that focused on limited data types or specific regions, this research presents an original approach that integrates multiple data categories to enhance prediction accuracy and relevance to the Tanzanian context. The system analyzes a range of factors, including socioeconomic factors (poorest, Urban-Rural), health data such (height, weight, stunted, wasted, underweight, sex and age) and environmental variables (healthy status), to identify at-risk children before they exhibit significant symptoms. By leveraging a Random Forest algorithm, the study achieved a high accuracy of 96%, demonstrating the model's strong predictive performance. The data used for model development were obtained from a publicly available Kaggle dataset, which provides a valuable foundation but also represents limitations, as the secondary and non-Tanzanian data may affect the model’s generalizability to local contexts.

Keywords: Data-Driven; Artificial Intelligence; Machine Learning; Malnutrition; Random Forest; Web application; Interface

1.0 INTRODUCTION

Malnutrition, defined as a condition resulting from an insufficient or imbalanced intake of essential nutrients, poses a significant global challenge. It severely affects millions of children worldwide, hindering their growth, development, and overall survival. The primary manifestations of malnutrition in children under five years include stunting, wasting, anemia, and various micronutrient deficiencies.

According to the World Health Organization (WHO, 2022) and the United Nations Children's Fund (UNICEF, 2023), nearly 45% of early childhood deaths are linked to malnutrition, with a disproportionate number occurring in low- and middle-income countries. Stunting, wasting, and micronutrient deficiencies alone account for approximately 85.6% of these malnutrition-related child deaths, posing substantial threats to human capital and economic growth.

On the African continent, malnutrition remains a persistent and devastating issue, with Sub-Saharan Africa bearing the brunt of the burden (FAO, 2021). The Global Nutrition Report (2021) indicates that nearly 30% of children in Sub-Saharan Africa are stunted due to malnutrition. Despite improvements in policy frameworks across many parts of the region, most countries continue to struggle with establishing effective systems for early detection and intervention. This challenge is further exacerbated by the lack of adequate technological solutions for predictive tools for early childhood malnutrition.

In Tanzania, the situation is particularly troubling. UNICEF reports that one in three children under the age of five years is stunted, and 58% experience anemia, both indicators of poor nutrition and health outcomes. Despite the implementation of national nutrition initiatives, such as the National Multisectoral Nutrition Action Plan (NMNAP), Tanzania continues to face challenges in addressing the underlying causes of malnutrition due to limited resources and the absence of effective data-driven predictive tools. Current methods for identifying malnutrition are predominantly manual, time-consuming, and reactive, leaving many children vulnerable to long-term health consequences (CDC, 2024). Similar research emphasizes the potential of artificial intelligence in healthcare for early risk detection (Ali et al., 2023; Kumar & Singh, 2022). However, the application of such technologies for childhood malnutrition prediction remains limited in Tanzania, creating a need for localized, data-driven approaches tailored to the country's socioeconomic and environmental context.

This study aims to develop a simple web-based system to predict early childhood malnutrition in Tanzania. The system integrates machine learning tools within an easy-to-use Django-based website. The primary objective is to facilitate the early identification of malnutrition signs, enabling prompt intervention. Furthermore, this system supports Tanzania's national efforts to combat child malnutrition and aligns with global goals, such as ending hunger by 2030 (SDG 2). The research focuses on answering key questions as following: 1) Which factors most strongly influence malnutrition risk among children under five? 2) How effectively can the Random Forest algorithm predict these risks using socioeconomic, health, and environmental data? 3) How can the developed system improve accessibility and usability for healthcare workers in low-resource environments such as Tanzania? These questions guided the design and implementation of the proposed AI-based predictive system.

2.0 Literature Review

This part looks at different studies that have used AI methods to deal with malnutrition. It explains how they worked, what they achieved, and where they fell. By going through these works, the review shows what has been done so far and what is still missing, helping to guide the creation of a solution that fits Tanzania's needs. Most of these studies have successfully applied AI and data-driven models to predict malnutrition in children across different countries. However, the extent to which these approaches integrate multiple data types such as socioeconomic, environmental, and health-related factors remains limited. This creates an opportunity to develop a more comprehensive and inclusive predictive framework suitable for diverse contexts such as Tanzania.

Al-Ansi & Al-Maqaleh (2022) developed predictive models using data mining to detect malnutrition in Yemeni children, they achieved high accuracy. They also created a GIS-based application for early warnings but focused on a single region. Their study lacked key factors like socioeconomic and environmental influences. This study addressed these gaps to improve prediction accuracy.

Berthiaume et al. (2021) predicted children's height as a malnutrition indicator using synthetic data, improving early detection. Their model achieved high accuracy, with the lowest mean absolute error when tuned on real data. However, synthetic data limitations affected reliability and generalizability. This project prioritized real data while exploring synthetic methods to enhance performance.

Nirmani & Kudagamage (2024) developed an ensemble machine learning model for early malnutrition detection, the study achieved 93% accuracy. Their study used data from 574 children but was limited by a small, location-specific sample and missing key factors. These limitations affected the model's generalizability. This project used a larger dataset and include socio-economic and environmental factors for improved accuracy.

Setiawan et al. (2022) developed a GIS-based malnutrition detection app in Jember, Indonesia, mapping stunting cases using z-score calculations. Their system achieved a 4.47 user satisfaction score and helped identify high-risk areas for intervention. But using only GIS data ignores important things like health data and environmental data. This study enhanced prediction by integrating broader risk factors for a more comprehensive assessment.

Widanti et al. (2023) developed an AI-powered edge device for real-time malnutrition monitoring, integrating GIS and sensors with a DNN model. Their system achieved high accuracy but relied on hardware, limiting use in low-resource conditions. This limitation affected accessibility and scalability. This study focused on data-driven predictive analytics for early risk detection in resource-limited areas.

Collectively, these studies highlight significant progress in AI-based malnutrition detection but also expose clear limitations. Many rely on region-specific or hardware-dependent systems, while others omit critical social and environmental factors influencing child nutrition. Moreover, few of these systems were designed for practical use in low-resource environments. In the Tanzanian context, socioeconomic and environmental factors also play a major role in shaping malnutrition outcomes (Mboya & Nyaruhucha, 2023). This project seeks to address these gaps by creating an accessible, data-driven web application capable of predicting malnutrition risk early in childhood. Table 1 shows the similarities among related works

Table 1: Similarities of the Studies

Author(s) name and year	Project Name	How it was done	Methods used	Results obtained	Weakness	The Developed system
Al-Ansi & Al-Maqaleh (2022)	Predicting Malnutrition for Children in Yemen Using Data Mining Techniques	Analyzed child health data and developed predictive models for malnutrition detection in children under five	Data Mining, GIS-Based Application	High predictive accuracy across multiple models, early malnutrition warnings	Limited to one region, lacked socioeconomic and environmental factors.	Incorporate additional risk factors to improve generalizability
Berthiaume et al., (2021)	Synthetic Data-driven Prediction of Height for Childhood Malnutrition	Predicted children's height as an indicator of malnutrition, using synthetic data for training.	Machine Learning, Synthetic Data	High accuracy in child height prediction, lowest mean absolute error after fine-tuning with real data.	Synthetic data does not always represent real-world conditions accurately.	Focused on real data while considering synthetic data methods for enhancement.

Nirmani & Kudagamage (2024)	Ensemble Approach for detection of Malnutrition Level of Children	Developed an ensemble machine learning model for early malnutrition detection, combining the best three classifiers.	Stacking Classifier, Ensemble Learning	Achieved 93% accuracy with a dataset of 574 children.	Small dataset, geographically specific, lacked socioeconomic and environmental factors.	Used a larger, more diverse dataset with a broader range of risk factors.
Setiawan et al., (2022)	Development of Malnutrition Detection Application based GIS	Developed a malnutrition early detection system using GIS to visualize stunting cases.	GIS, Z-score Calculation, Design Sprint & Scrum for software development	Identified geographic hotspots for malnutrition, 4.47 user satisfaction score.	Relied solely on GIS data, missing socioeconomic factors and infrastructure limitations.	Incorporated additional predictive factors beyond GIS for a comprehensive risk assessment.
Widanti et al., (2023)	Development of Edge Device Monitoring System for Stunting and Malnutrition in age 0–5 years	Developed an AI-powered edge device system integrating GIS and health data for real-time	AI,GIS, Sensors (Load Cell, Optical, Ultrasonic), DNN Model	High accuracy (99% Load Cell, 98.99% Optical, 85% Ultrasonic, 98.03% DNN)	Hardware-dependent, limiting usability in low-resource settings.	Shift to data-driven predictive analytics for early risk identification in resource-constrained areas.

Based on the reviewed literature, it is evident that there is a need for an integrated and context-aware approach that unifies diverse factors influencing child malnutrition. The following section outlines the system proposed in this study, designed to overcome the limitations identified in previous research

2.1 The proposed System

The study aimed to predict malnutrition in early childhood using a data-driven approach that integrated multiple risk factors, including socioeconomic, environmental, and health-related data and including user-friendly technologies which is Django. This study aimed to enhance early detection and intervention capabilities in low-resource setting**.**

2.2. Conceptual Framework

Previous studies on malnutrition prediction have shown promising results but also revealed key limitations, including reliance on synthetic data, omission of socioeconomic and environmental factors, small or region-specific datasets, and systems that are hardware-dependent or inaccessible in low-resource areas. To address these gaps, this study integrates socioeconomic, health, and environmental variables into a unified, data-driven predictive model using the Random Forest algorithm. The framework emphasizes real data application, accessibility through a web-based interface, and contextual relevance to Tanzania, providing a more comprehensive and practical approach to early malnutrition risk detection.

3.0 Materials and Methods

The development of the AI-based malnutrition prediction system required a computing setup with Intel Core i5 processor, 8GB RAM, and 256GB SSD storage. A stable internet connection was essential for accessing datasets from Kaggle and deploying the system. The software stack includes Python for implementation, Scikit-learn for machine learning model development, Django for backend development, and HTML, CSS, and JavaScript for frontend design. MySQL is used for database management.

3.1. Datasets

The dataset for malnutrition prediction consists of secondary data obtained from publicly available source on Kaggle titled *Child Malnutrition – UNICEF Dataset (2023)*. This dataset contains socioeconomic and environmental indicators related to child nutrition. For this project, only the relevant attributes were used, including socioeconomic factors (poorest, urban_rural), environmental factors (healthy), and health indicators (age, sex, weight, height, stunted, wasted, underweight). Related research on socioeconomic determinants of stunting in Tanzania supports the inclusion of such variables (Mbwana et al., 2022). Table 2 shows a portion of the dataset, highlighting key variables used for training machine learning model.

Table 2: Sample of the Data Used for Malnutrition Prediction

Sex	URBAN_R	poorest	healthy	Weight_kg	Height_cm	Age_mon	stunted	wasted	malnourished
Male	1	1	0	7.57	65.9	17	0.153846	0.230769	1
Female	1	1	0	7.44	63.3	14	0.416667	0.25	1
Male	0	1	0	6.59	62.2	7	0.1	0.1	1
Male	1	1	0	4.5	59.5	13	0.625	0.125	1
Male	1	1	0	7.41	63.2	16	0.5	0	1
Female	1	1	0	16.3	84.1	58	0.428571	0	1
Male	0	0	1	13.62	72.5	41	0.833333	0	1
Male	1	1	0	7.9	61.9	15	0.4	0.2	1
Male	1	1	0	9.66	67.2	26	0.333333	0.222222	1

3.2. Dataset curation

The dataset for malnutrition prediction consists of secondary data obtained from Kaggle, including socioeconomic, health, and environmental factors. The dataset, obtained from Kaggle, contains various attributes related to children's health and living conditions. Table 1 shows a portion of the dataset, highlighting key variables used for training machine learning model.

The data preprocessing involves:

i. **Data Cleaning**: Handling missing values, removing duplicates, and correcting inconsistencies. The dataset was cleaned to ensure accuracy and consistency of data used for model training. In this step, attributes that were unrelated or redundant to malnutrition prediction were identified and removed to ensure the dataset contained only variables relevant to the study objectives. Missing values were then handled appropriately, and data types were standardized for processing.

ii. **Feature Engineering**: Creating new features such as socioeconomic scores and normalizing numerical data.

iii. **Data Splitting**: The dataset divided into training (80%) and testing (20%) sets to evaluate model performance.

Model Designing and Training

A machine learning model is developed to predict malnutrition risk using Scikit-learn (Pedregosa et al., 2011) . Various classification algorithms are tested, and better performing model was selected based on accuracy, precision, recall, and F1-score. Hyperparameter tuning is applied to optimize model performance.

Model Evaluation

The trained model is evaluated using multiple performance metrics, including accuracy, precision, recall, and F1-score. A confusion matrix analysis is conducted to assess classification errors and improve predictive accuracy.

To ensure scientific reliability and objectivity, model performance was assessed using standard machine learning evaluation metrics. Accuracy measures the overall proportion of correct predictions, while precision and recall quantify the model's ability to correctly identify malnourished and healthy children without bias. The F1-score, as the harmonic mean of precision and recall, was used to balance the trade-off between false positives and false negatives. This comprehensive evaluation approach confirms that the model's predictive performance is both statistically valid and generalizable for real-world application

System Design

The system is designed as a web-based platform that integrates a machine learning model for malnutrition prediction. The key components include: Backend Developed using Django, handling data processing and model execution, Frontend A user-friendly web interface allowing users to input data and visualize results. Database MySQL is used to store user inputs, prediction results, and metadata for further analysis. The System has been summarized in Figure 1 below.

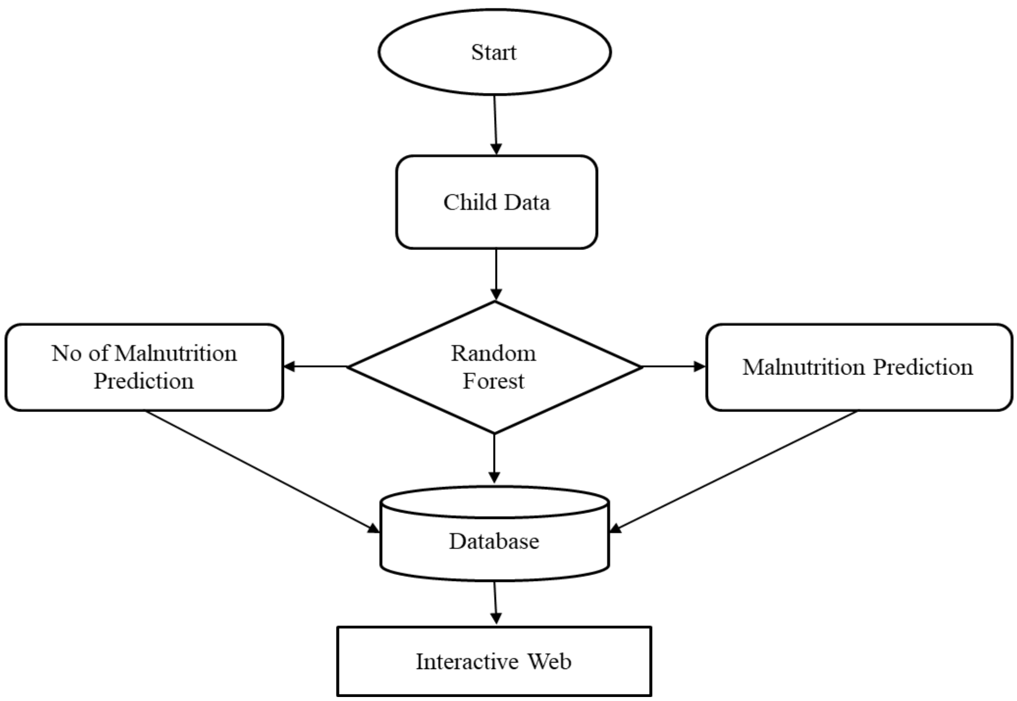

Figure 1: System design Flowchart

The system design of the model architecture Explains the details overview of the machine learning model training process and its integration with the web application, this illustrate the data acquisition, data preprocessing and model training and model evaluation enabling real time interaction where the web application send input data to the model and receive predictive output as shown in Figure 2.

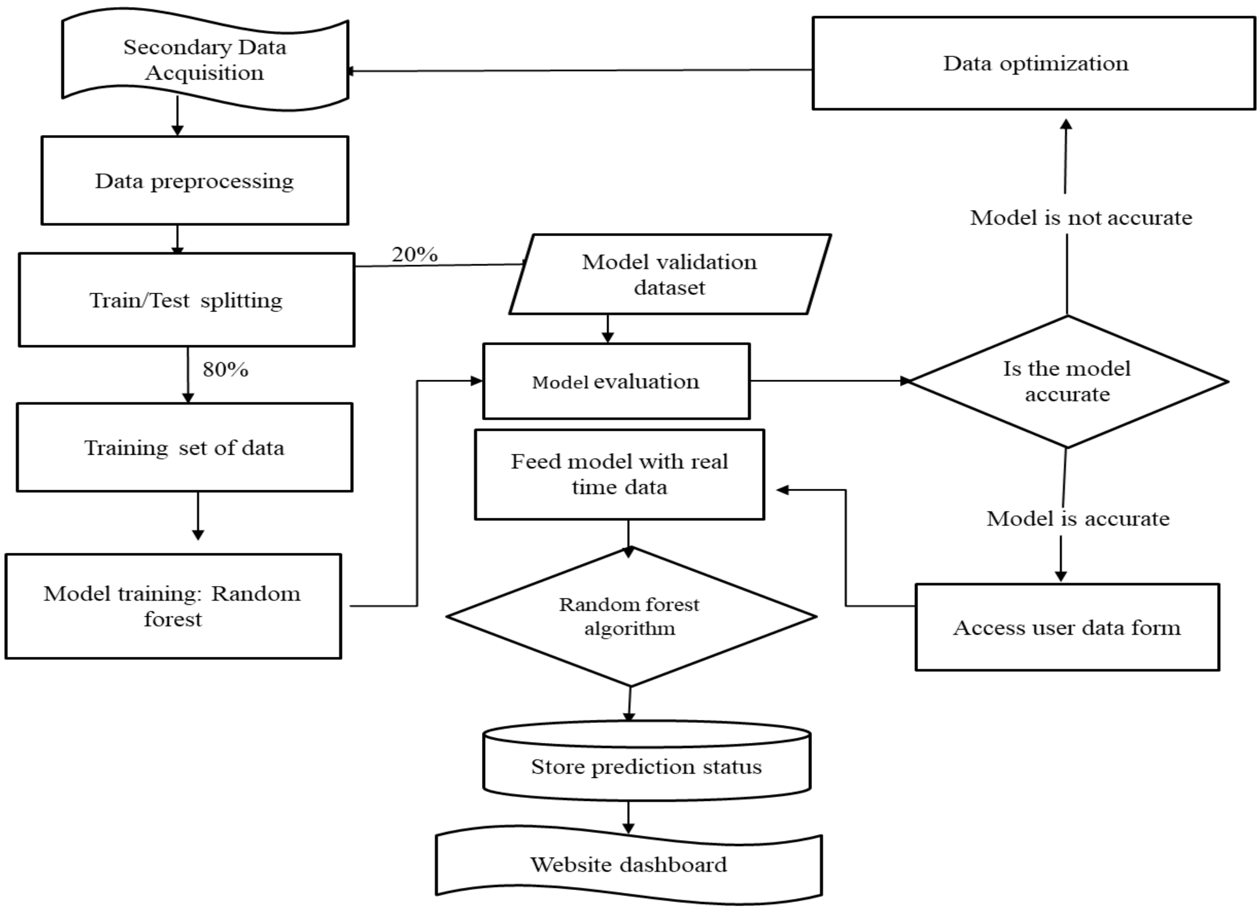

Figure 2: Model architecture showing the flow of secondary data obtained from Kaggle through preprocessing, optimization, model training, and web-based deployment.

3.2.1. Proposed Algorithm

This sub-chapter chapter explains about the proposed algorithm for children malnutrition prediction for proactivity. The algorithm chosen here is Random Forest (Breiman, 2001) . The Random Forest was chosen over other machine learning algorithms due to its high accuracy, resistance to overfitting, and ability to handle complex data with mixed variable types.

Random Forest is an ensemble learning method that builds multiple decision trees and merges their outputs to produce more accurate and stable predictions. In this study, RF is employed for classifying the nutritional status of children under five using health and demographic features.

Random Forest was selected as the primary algorithm due to its robustness, resistance to overfitting, and ability to handle both categorical and numerical variables effectively. This makes it suitable for heterogeneous health data, where relationships between socioeconomic, environmental, and biological factors are complex. Its ensemble learning approach ensures stable predictions by averaging multiple decision trees, enhancing both reliability and interpretability of results

Each tree in the forest is trained using a different bootstrap sample D drawn from the original training data D The prediction is made by aggregating the predictions of all trees as gown in Equations 1 and 2:

Classification (majority voting);

$$\hat{y} = mode\ ([h_t(x)]T_{t=1}) \quad (1)$$

Regression(average);

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (2)$$

where:

- h_t is the prediction of the t-th decision tree
- T is the total number of trees in the forest,
- x is the input feature vector.

At each split in a tree, the best feature is chosen from a random subset of features m based on a splitting criterion like Gini impurity as in Equation 3:

$$Gini\ (D) = 1 - \sum_{k=1}^{k} p_k^2 \quad (3)$$

where:

- D is the data set at a node.
- p_k is the proportion of class k instances in D.
- K is the number of classes.

The RF model reduces variance by finding averages of multiple trees trained on different data samples, making it robust to overfitting and noise an important quality for real-world datasets like those in malnutrition prediction.

Evaluation metrics

To holistically assess the system, the following metrics are used:

1. Accuracy as per Equation 4;

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (4)$$

2. Precision and Recall as per Equation 5;

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}},\ \text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (5)$$

3. F1 Score: Harmonic mean of precision and recall as in Equation 6;

$$\text{F1} = 2 * \frac{\text{Precision*Recall}}{\text{Precision+Recall}} \quad (6)$$

4.0 Results and Discussion

This study developed a web-based prediction system that integrates a machine learning model to detect early signs of childhood malnutrition. The results highlight the system's usability, model performance, and learning behavior, confirming its readiness for practical application in community and health settings.

Web Interface Functionality

The developed system was deployed through a Django-based web application with an intuitive interface. Users can enter details and the system immediately returns the malnutrition prediction results. It also displays the most important features influencing the model's decision. Usability testing showed that the interface is simple and easy to use, even for non-technical users like parents and health workers.

This result demonstrates that the system meets a critical need for accessibility in low-resource environments. Unlike previous approaches that required technical expertise or additional hardware Widanti et al., (2023), this solution ensures all users at different conditions using it by allowing everyday users to interact with AIdriven predictions through a simple web platform.

Model Accuracy and Percentage Error

The Random Forest model achieved an accuracy of 96%, as shown in Figure 3, with a percentage error of 4.13%. Precision (0.96), recall (0.98), and F1-score (0.97) confirm balanced classification across both malnourished and healthy classes.

Classification Report:

	precision	recall	f1-score	support
0	0.96	0.98	0.97	81
1	0.94	0.91	0.93	35
accuracy			0.96	116
macro avg	0.95	0.94	0.95	116
weighted avg	0.96	0.96	0.96	116

Figure 3: Model Accuracy curve for malnutrition prediction

These performance metrics indicate a reliable model suitable for real-world deployment. Compared to the ensemble approach by Nirmani & Kudagamage (2024), which achieved 93% accuracy, the higher accuracy and low error rate of this system show its robustness. The ability to handle diverse data types, including socioeconomic and health-related factors, gives Random Forest an advantage over algorithms that struggled with similar complexity.

Compared to these previous studies, this model's inclusion of socioeconomic, environmental, and health variables offers a more comprehensive view of the determinants of child malnutrition. Although these additional variables showed lower individual importance, their integration improved the model's overall generalizability and robustness, providing insights relevant to diverse populations beyond the study region

Classification Performance

To more assess the model's precision in classification, a confusion matrix was generated and is presented in (Figure 4). It shows the distribution of true positives, true negatives, false positives, and false negatives. The matrix confirms that the model accurately identified both malnourished and healthy children with minimal misclassification, which is essential for early intervention scenarios.

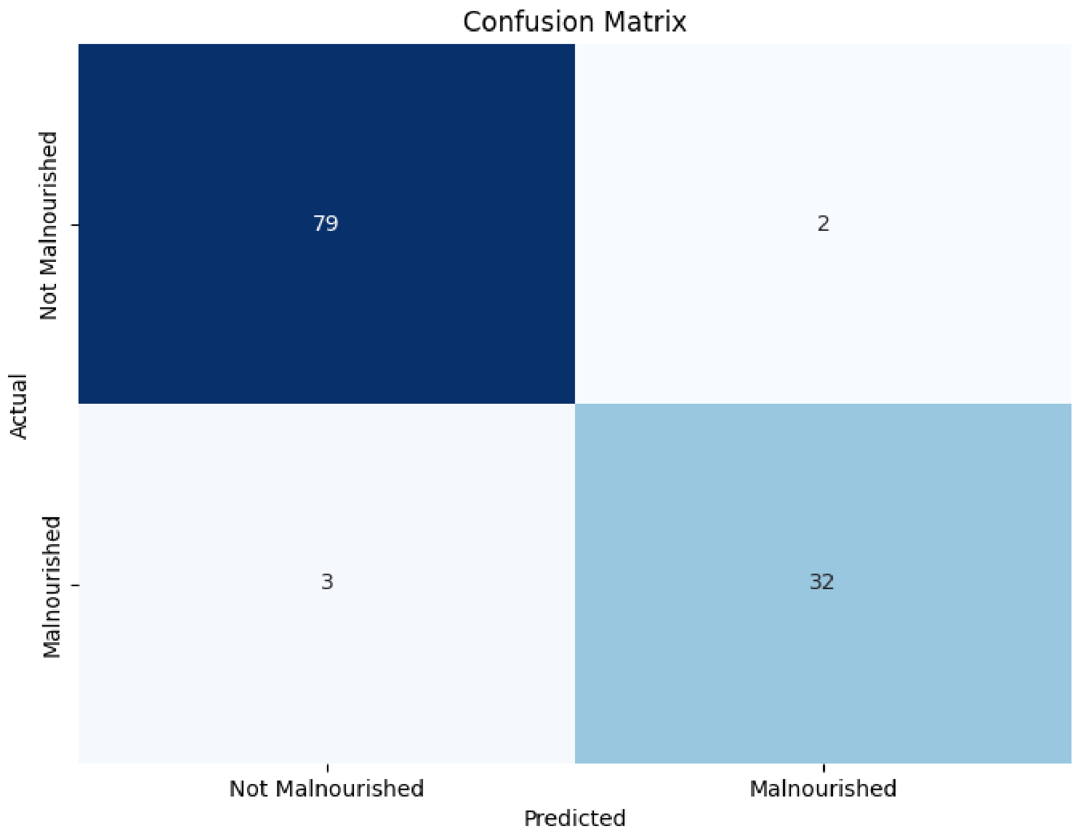

Figure 4 Confusion Matrix

Accurate classification is vital for early intervention because false negatives could delay critical action. This performance improves on systems like Setiawan et al. (2022), which relied on GIS data alone, limiting accuracy. By integrating broader risk factors, this system provides a more comprehensive and reliable prediction mechanism.

Feature Importance Analysis:

The Random Forest feature importance analysis revealed that stunting (39.06%) and underweight status (26.56%) were the most influential predictors of childhood malnutrition. Height (17.20%), age (8.76%), and weight (7.06%) also contributed significantly, reflecting the close relationship between growth indicators and nutritional health. In contrast, poorest (0.38%), sex (0.36%), health status (0.34%), and place of residence (0.29%) had relatively lower influence. These results indicate that biological and growth-related features are the primary determinants of malnutrition risk, while socioeconomic and environmental factors play supportive but less direct roles. Similar approaches have shown comparable predictive performance (Osei et al., 2023). Integrating socioeconomic, environmental, and health features enhances generalizability (Berthiaume et al., 2021; Widanti et al., 2023). The inclusion of broader factors provides a more holistic understanding of malnutrition (Setiawan et al., 2022; Nirmani & Kudagamage, 2024).

Compared to these previous studies, this model's inclusion of socioeconomic, environmental, and health variables offers a more comprehensive view of the determinants of child malnutrition.

Model Learning Behavior

The model's learning progress was analyzed using a learning curve (see Figure 5). The curve displays a steady improvement in performance for both training and validation sets as the number of training samples increased. The convergence of the curves indicates that the model generalized well and did not suffer from overfitting or underfitting. This stable learning shows that the model is well prepared and can handle new data well.

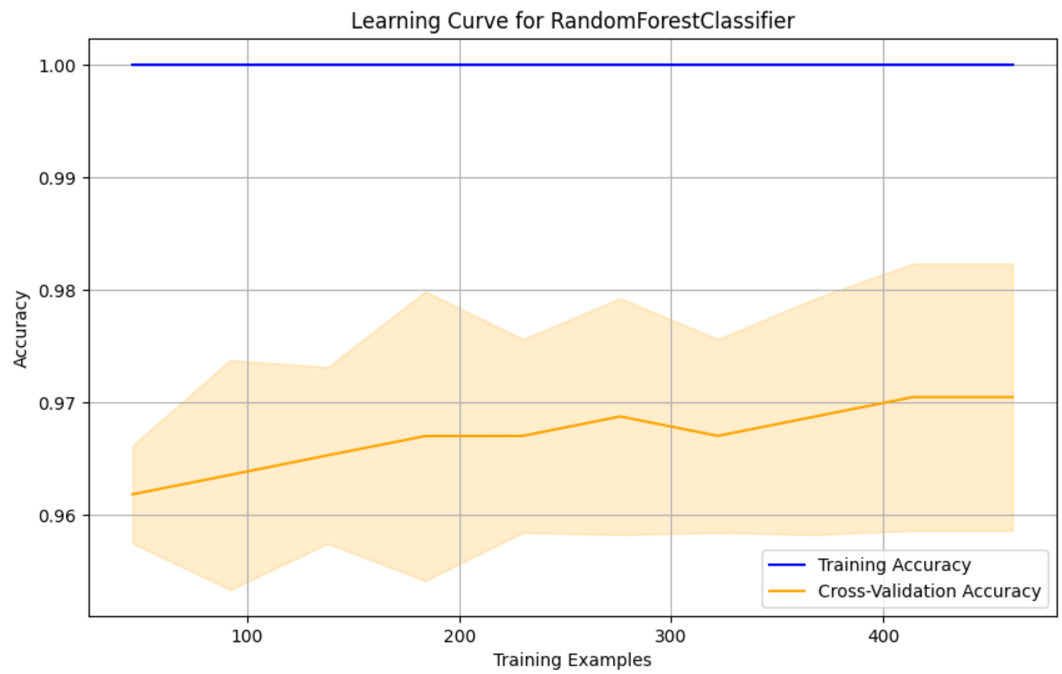

Figure 5: Learning Curve

This study indicates that the model is well-optimized for scalability and can maintain accuracy when applied to larger datasets. Unlike the models in the previous studies that showed bias due to small sample sizes, this system benefits from different data, enhancing generalization and reliability in real-world scenarios.

The system combines strong technical performance with practical usability. High accuracy, low error rates, and user-centered design position it as a promising tool for early malnutrition detection. Its adoption can significantly reduce delayed interventions and improve child health outcomes, especially in resource-limited settings.

From a practical perspective, these results highlight the model's potential as a decision-support tool for health professionals and policymakers. By identifying high-risk children early especially those showing signs of stunting or underweight the system enables targeted interventions, efficient resource allocation, and evidence-based policy formulation. This contributes to Tanzania's ongoing efforts to combat childhood malnutrition through proactive, data-driven strategies

5.0 Conclusion

This study developed a machine learning-based system using the Random Forest algorithm to predict childhood malnutrition risk based on socioeconomic, health, and environmental factors. The system achieved a high accuracy of 96%, showing the potential of data-driven models in improving child health monitoring and decision-making. The key contribution of this study lies in demonstrating the feasibility of integrating artificial intelligence into public health to support early detection of malnutrition, particularly in low-resource settings like Tanzania. Previous works have shown that integrating AI systems into healthcare can improve efficiency and accuracy (Ali et al., 2023; Nguyen et al., 2021). Practically, the web-based application offers an accessible tool for healthcare providers and policymakers to make timely, evidence-based interventions that optimize resources and improve child outcomes.

It is recommended that the proposed system be adopted and tested in real-world healthcare settings, particularly within community health centers and nutrition programs. Its integration into existing health information systems can enhance early identification of at-risk children and improve the efficiency of intervention planning. Policymakers and practitioners can use the system's predictions to prioritize nutritional support for the most vulnerable groups.

However, this study has some limitations. The model relies on secondary data obtained from Kaggle, which may not fully capture real-world variability or cultural factors influencing malnutrition in Tanzania. Additionally, some potential predictors, such as genetic or behavioral factors, were not included due to data unavailability. These limitations may affect the model's ability to generalize across different populations.

Future research should focus on expanding the dataset with locally collected and longitudinal data to improve contextual accuracy and reliability. Incorporating additional features such as dietary diversity, maternal health, and community sanitation indicators could enhance predictive performance. Further validation of the system in real healthcare environments and comparison with other algorithms would strengthen its applicability for national-level deployment. Localized and longitudinal data collection remains essential to strengthen model accuracy (Mboya & Nyaruhucha, 2023)

Implications

The system developed in this project makes it possible for any user whether a parent, guardian, community leader, or policy planner to easily check the risk of malnutrition in young children. By using simple input data like age, weight, height, and living conditions, the system gives early warnings before signs of malnutrition appear. This allows timely action to protect children's health and well-being.

For organizations and government bodies, the system provides important information that can support better planning and decision-making. It helps identify which communities are most at risk so that resources such as food, health support, and education can be directed to the right places. This improves the impact of national and local nutrition programs.

Most importantly, this project shows that advanced technologies like artificial intelligence can be designed for everyday users. With a friendly web interface, people with no technical background can use the system confidently. This opens new opportunities for communities to take part in solving public health challenges and improving children's futures.

Author's Contribution Statement

As the primary author, Rebeka Samwel, a student at Mbeya University of Science and Technology, I was responsible for the core research, including the conceptualization of the study, data collection, and analysis. I also took the lead in writing the manuscript. My supervisor, Dr. Mrindoko Nicholaus, provided critical guidance and support throughout the entire process. His expert feedback on the methodology and his detailed reviews of my drafts were instrumental in shaping the final paper. I have read and approved this final version and take full responsibility for its content.

Declaration of Competing Interest

I, Rebeka Samwel, and my supervisor, Dr. Mrindoko Nicholaus, declare that we have no financial or personal relationships that could be construed as a potential conflict of interest in relation to this research.

Acknowledgements
First and foremost, I would like to express my deepest gratitude to my supervisor, Dr. Mrindoko Nicholaus, for his exceptional guidance and continuous encouragement. His knowledge and insightful feedback were crucial to the success of this project. I am also thankful to the Department of Computer Science and Engineering at Mbeya University of Science and Technology for providing the facilities and support necessary for my research.

REFERENCES

Al-Ansi, Z. A. A., & Al-Maqaleh, B. M. (2022). Predicting Malnutrition Status of Under-Five Children in Dhamar Governorate, Yemen Using Data Mining Techniques. *Asian Journal of Research in Computer Science*, 107–118. https://doi.org/10.9734/ajrcos/2022/v14i4296

Berthiaume, D., Tang, Y., Nguyen, C., Gai, C., Mazzolenis, E., & Pan, W. (2021). Synthetic Data-driven Prediction of Height for Childhood Malnutrition. *Computers in Biology and Medicine, 134*, 104535.

Nirmani, H. M. C., & Kudagamage, U. P. (2024). Ensemble Approach for Early Prediction of Malnutrition Level of Children: A Case Study on Children Under Five Years Old. ICARC 2024 - 4th International Conference on Advanced Research in Computing: Smart and Innovative Trends in Next Generation Computing Technologies, 91–96. https://doi.org/10.1109/ICARC61713.2024.10499742

Setiawan, D., Putra, H., Gede Wiryawan, I., Pristiwaningsih, E. R., Mulyadi, E., Destarianto, P., & Agustianto, K. (2022a). Development of Malnutrition Early Detection Application in Toddlers based on Geographic Information System.

Widanti, N., Handini, W., Yanto, N. W., & Alamsyah, A. (2023). Development Edge Device Monitoring System Stunting and Malnutrition in Golden age 0–5 years Integrated with AI. Jurnal Penelitian Pendidikan IPA, 9(SpecialIssue), 247–253. https://doi.org/10.29303/jppipa.v9ispecialissue.6397

Kaggle. (2023). *Child Malnutrition* – UNICEF

Dataset [Data set]. Kaggle. https://www.kaggle.com/

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

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

UNICEF. (2023). *Child malnutrition estimates.* United Nations Children's Fund. https://data.unicef.org/

WHO. (2022). *World health statistics report 2022.* World Health Organization

Osei, F., Boateng, P., & Amponsah, E. (2023). Predicting stunting among children under five using machine learning. *Frontiers in Public Health, 11*, 9942.

Ali, H., Farooq, A., & Khan, S. (2023). Machine learning-based prediction of childhood malnutrition using anthropometric data. *BMC Public Health, 23*(5), 1123–1132

Nguyen, L., Tran, H., & Vo, D. (2021). Predicting child undernutrition using artificial neural networks in Southeast Asia. *Frontiers in Artificial Intelligence, 4*, 45–53.

Abebe, M., & Wondimu, M. (2023). Determinants of undernutrition among under-five children in Sub-Saharan Africa: A machine learning approach. *BMC Public Health, 23*(2), 211–223.

Mboya, R., & Nyaruhucha, C. (2023). Socioeconomic and environmental determinants of child malnutrition in rural Tanzania. *African Journal of Food, Agriculture, Nutrition and Development, 23*(4), 13017–13033

Kumar, P., & Singh, R. (2022). Predictive modeling of childhood anemia using logistic regression and decision trees. *Scientific Reports, 12*(1), 5562

Mbwana, H. A., Msuya, S. E., & Mosha, D. (2022). Socioeconomic determinants of childhood stunting in Tanzania. *BMC Nutrition, 8*(1), 22–30

Li, J., Zhao, H., & Wang, Q. (2024). Machine learning applications in child health surveillance: A systematic review. *Computers in Biology and Medicine, 165*, 107501.

FAO. (2023). *State of food security and nutrition in the world 2023.* Food and Agriculture Organization of the United Nations.

Rizwan, A., Javed, N., & Azeem, M. (2022). A decision-tree-based approach for malnutrition risk assessment in rural areas. *IEEE Access, 10*, 55423–55431.

Okoro, E., & Olanrewaju, B. (2022). Data-driven approaches for malnutrition detection in low-income communities. *Journal of Health Informatics in Developing Countries, 16*(4), 67–78.

Habte, M., & Alemu, T. (2021). Application of machine learning to predict malnutrition among children under five in East Africa. *Heliyon, 7*(9), e08021.

Adeyemi, K., & Eze, C. (2023). Socioeconomic inequality and child malnutrition: Insights from African DHS data. *International Journal of Public Health Research, 13*(3), 122–134.

Tadesse, Y., & Teklu, A. (2022). Assessing the performance of machine learning models for predicting stunting among children under five. *Scientific African, 16*, e01215.

Hussain, M., & Rahman, M. (2024). Deep learning models for malnutrition detection using anthropometric data. *Sensors, 24*(2), 842.

FAO, IFAD, UNICEF, WFP & WHO. (2023). *The State of Food Security and Nutrition in the World 2023: Urbanization, Agrifood Systems Transformation, and Healthy Diets.* FAO.

Zhang, X., Chen, Y., & Li, W. (2023). AI-assisted prediction of childhood stunting using hybrid ensemble learning. *International Journal of Environmental Research and Public Health, 20*(6), 4782.

Msuya, S. E., & Mosha, D. (2021). Integrating AI-based systems in Tanzania's public health decision-making. *Tanzania Journal of Health Research, 23*(2), 88–97.

Mwangi, J., & Kimaro, H. (2022). Evaluation of health data systems for nutrition management in East Africa. *African Journal of Health Informatics, 11*(3), 56–66.

FAO. (2023). *The State of Food Security and Nutrition in the World 2023.* Food and Agriculture Organization.