Using Diet Analysis to Predict and prevent child malnutrition

Kar Woh Anchal Skehri Nafisa Jimoh

Zion Oluwasegun Aleksander Busz Rofiat Alimi

Abel Enagbonmhena

Abstract

Malnutrition remains one of the major global health challenges, particularly in low- and middle-income countries. It is an illness that is developed when the body is deprived of a balanced and nutritious diet. The lack of adequate nutrients impedes the body's growth and development in both children and adults. Undernutrition, a form of malnutrition has proven to be the major contributor to the deaths of children, accounting for an estimated 45% of all deaths among children under the age of five worldwide while 38.9 million children are considered overweight. Stunting and wasting are two forms of undernutrition that negatively affects the development of children. We study malnutrition at the country and region level, we further used the country-level data on the burden of malnutrition and child diet from Global Report 2000-2021 to build predictive models for malnutrition. Our analysis shows that Africa and Asia are the regions most affected by malnutrition. Deep learning ad machine learning techniques were used to build models, the results depict that polynomial regression model had a lower MSE value as compared to other regression models.

1. Introduction

Malnutrition is a serious health condition that is affecting all age groups in both developing and developed countries. According to the World Health Organization (WHO), it is described as the deficiencies, excesses, or imbalances in a person's intake of energy and/or nutrients. It can be in the form of undernutrition, overweight/obesity, diet-related incommunicable diseases, and micronutrient-related malnutrition [1]. Undernutrition which is caused by not having enough food that contains nutritious substances is the sole contributor to about 45% of deaths in children under the age of 5 in low and middle countries across the world. In 2020, a report by WHO showed that 149 million children under 5 years globally were estimated to be stunted, 45 million were estimated to be wasted and 38.9 million were overweight/obese [1].

Undernutrition consists of stunting, wasting, deficiencies in vitamins and minerals, and being underweight. These forms of malnutrition have made children vulnerable to diseases and have been the major cause of death among children. Children who suffer from acute malnutrition (wasting) have a higher risk of death from common childhood illnesses such as diarrhea, pneumonia, and malaria. During the COVID-19 pandemic, children under the age of 5 had 0.1% of the global deaths with a total of 2% of global cases [2]. The consequences of undernutrition in early childhood can be long-lasting and can affect a child's physical and cognitive development as well as their future productivity. To tackle and reduce the burden of malnutrition, the World Health Assembly in 2012 set up a list of global nutrition targets to be achieved in 2025. To monitor and track the progress made toward achieving these targets, an independent expert group in collaboration with other partners including UNICEF, WHO and the World Food Programme provides an annual report known as the Global Nutrition Report. The report provides a comprehensive assessment of the progress towards the global nutritional targets, highlights key challenges and opportunities in global nutrition, and includes data on nutritious indicators, child and adolescent diets, the burden of malnutrition, nutritional finances, and strategies [3].

Of the six target goals set to improve maternal, infant, and young children's nutrition and prevent malnutrition, five are heavily tied to children. They include achieving a 40% reduction in the number of children under the age of 5 who are stunted, achieving a 30% reduction in low birth weight, ensuring that there is no increase in childhood overweight, increasing the rate of exclusive breastfeeding in the first 6 months up to 50% and reducing the number of childhood wasting to less than 5% [4]. To identify if a child is suffering from malnutrition, a popular method known as

Anthropometric measurement is used. It is used as an indicator to determine acute (wasting) malnutrition or chronic (stunting) malnutrition. This measurement involves measuring the physical properties of a child. It provides the z-score of a child's weight for height, height for age, weight for age, body mass index (BMI), and mid-upper circumference. Nevertheless, all factors which include the socioeconomic status of the parents, poverty, and poor maternal health attributed to malnutrition results in denying the child access to a well-balanced and nutritious diet. This further contributes to the decline in the Anthropometric measurement of a child with respect to their age. A child who consumes all the adequate nutrients needed for efficient bodybuilding will not only win the war against malnutrition but will have a body that can fight against common childhood illnesses.

This research focuses on predicting and preventing malnutrition in children using the data obtained from the Global Nutrition Report [3]. The dataset comprises six fields relating to malnutrition at both the country and regional levels, this work entails a multivariable analysis of the burden of malnutrition and the diet of children under the age of 5 at the country level. In this study, predictive models to predict stunting, wasting, and overweight using diet features is developed. By carrying out these analyses, this study aims to contribute to efforts to reduce the burden of malnutrition in children under 5.

2. Literature Review

Over the years AI has greatly contributed to solving problems across the world. Companies and industries in various sectors apply AI to help carry out their daily activities. Systems developed using AI have been proven to be successful in meeting up with the desired expectations [5]. With the buildup of huge data, machine learning which is an aspect of AI has enabled the development of predictive models/algorithms that facilitate the performance of AI systems. With these advancements, machine learning-based algorithms being integrated into electronic health records can provide care providers with decision support systems that are used in identifying patients at higher risk of malnutrition as well as managing them [6].

In order to prevent the risk of death in malnourished children, [7] uses a convolutional neural network to identify if a child is malnourished or not. The system takes in images of children as input and uses AlexNet and Transfer learning to build a model that carries out the classification task. The system achieved a 96% accuracy in detecting if a child is suffering from malnutrition or not. [8] developed different machine learning models using different algorithms to predict malnutrition in children under the age of 5 in Bangladesh using data gotten from the 2014 Bangladesh Demographic and Health Survey (BDHS). 17,863 samples of children under the age of 5 were used in training and testing the models while various combinations of demographic, socioeconomic, and health-related variables were used. Amongst the four algorithms used, the model built using random forest performed the best with an accuracy of 86.3%, a sensitivity of 73.8%, and a specificity of 90.1%.

Another research from Bangladesh uses machine learning algorithms to detect the risk factors of malnutrition (stunting, wasting, and underweight) [9]. The authors identified the region, child’s age, father’s education, toilet types, and wealth index as significant risk factors for malnutrition by using logistic regression. These factors were fed into the model for prediction. Amongst the classifiers used; the random forest classifier had high accuracy for all forms of malnutrition. [10] carried out research to predict stunting in children under the age of 5 in Rwanda using machine-learning models. The gradient-boosting algorithm model performs the best in predicting stunting in children.

[11] carried out an analysis to study the factors that are associated with undernutrition in children under the age of 2 in Pakistan. The dataset comprised 984 participants with different variables ranging from the age of the child, the mother’s age, immunization records of the child, vitamin A and iron consumption, breastfeeding, BMI of the mother, and so on. Logistic regression was used to determine the factors responsible for malnutrition in children. From the analysis, it was recorded that the odds of stunting, wasting, and underweight increased as the child’s age increased.

This research focuses on other dietary intakes of children under the age of 5 not considered by other studies. Factors such as early initiation, solid foods, minimum diet diversity, minimum meal, and minimum-accept diet are analyzed.

3. Methodology

In this section, a detailed description is given of data cleaning, exploratory data analysis, model development, and model evaluation.

Dataset Description: The data obtained from [3] comprise 6 sub-data. For this study, only the 2 datasets are selected; they include the burden of malnutrition (wasting, stunting, overweight in children and adolescents, obesity track in males &

females, and others) and diet at the country level. There are 195 countries in total and are further disaggregated into sex, age, environment, education, location, and wealth status of both children and adults.

Burden of Malnutrition: This includes wasting, stunting, and overweight. The values are given in percentage (%) for the year 2000 - 2021.

Diet: features provided are early initiation, exclusive breastfeeding, minimum meal, the minimum accept diet, minimum diet diversity, continued breastfeeding for 1 year, and solid foods, and other adult food intake from 2000 - 2020. Values are also provided in percentage.

3.1 Data Cleaning

Data relating to the burden of malnutrition and diet in children under the age of 5 are selected. As disaggregation, age, and sex at the country level are obtained. All other variables (adult variables) not needed for this study are dropped using Python's drop method. All missing values in both data were filled with 0. The resulting shape for diet data is 3242 rows and 132 columns while the burden of malnutrition data comprises 9148 rows and 66 columns. This form of cleaning is done for data modeling.

For the exploratory data analysis, the sex of the children is selected as disaggregation. The average of each variable for a country is calculated. The missing values are filled with 0. The dataset comprises 382 rows and 17 columns.

3.2 Exploratory Data Analysis

The prevalence of stunting, wasting, and being overweight globally currently is 17.8%, 4.5%, and 5.5%. Region analysis in Figure 1 shows that Africa has had the highest prevalence of stunting and wasting over the past two decades. Children from North America are rarely affected by the burden of malnutrition. Figure 2 depicts the percentage of malnutrition in both boys and girls under the age of 5. From the analysis, boys are mostly affected by all cases of malnutrition than girls. As of 2021, further analysis shows that Burundi has had the highest number of stunted children (56.51%) followed by Timor-Lester and Yemen.

Across all regions, children aged 12-15 months are the most likely to receive a diverse diet food for a minimum number of times the previous day, with an average percentage ranging from 40.89% in Africa to 69.97% in Europe. The percentage decreases for older age groups, with children aged 20-23 months consistently having the lowest percentage across all regions. The highest percentages for each age group are observed in Europe, with Latin America and the Caribbean also having relatively high percentages for children aged 16-19 months and 20-23 months.

Figure 10 shows the average percentage of regions with children that receive food from 5 different groups by age group and by region. Looking at the data, we can see that regions in Europe have the highest percentage of children receiving food from all 5 groups across all age groups, while regions in Africa have the lowest percentages. The age group with the highest percentage of children receiving food from all 5 groups is 20-23 months, while the age group with the lowest percentage is 6-11 months. Overall, this data highlights the need for increased efforts to improve dietary diversity in regions with lower percentages of children receiving food from multiple groups, particularly in younger age groups.

The percentage of infants aged 0-5 months who are exclusively breastfed in different regions of the world is shown in Figure 9.The image shows that, generally, a higher percentage of infants in Oceania and Africa are exclusively breastfed in the first few months of life compared to other regions. Specifically, over 66% of infants in Oceania are exclusively breastfed in the first month, compared to 37% in Africa and 36% in Europe. Across all regions, the percentage of infants exclusively breastfed declines as they age, with the highest percentage in the 0-1 month age group and the lowest in the 4-5 month age group. These findings suggest that breastfeeding practices vary widely across regions and highlight the need for targeted interventions to promote and support exclusive breastfeeding in the first few months of an infant's life.

3.3 Data Visualization

This section is divided into three parts, and the burden of malnutrition (wasting, stunting, overweight) is analyzed. The second part consists of images from analyzing the diet data, features such as exclusive breastfeeding, continued breastfeeding for 1 year, solid foods based on gender and a child's early initiation are considered. The last part visualizes the diet data for different child age groups.

Visualization of the Burden on Malnutrition

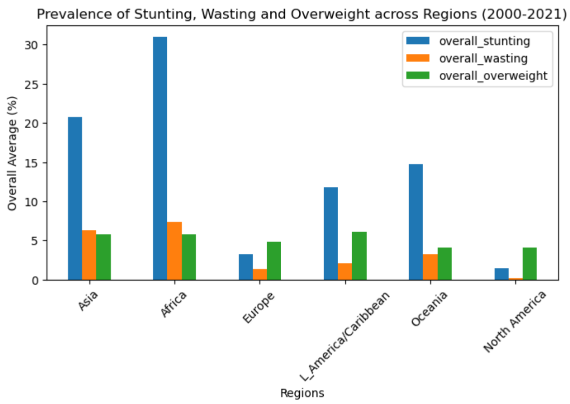

Figure 1: Prevalence of Malnutrition in children under the age of 5 regionally

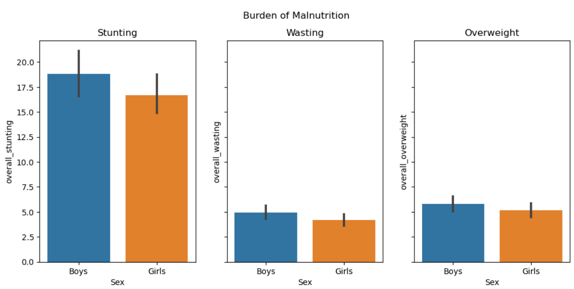

Figure 2: Burden of Malnutrition amongst girls and boys globally.

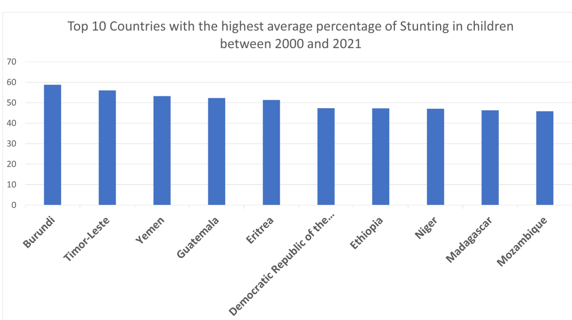

Figure 3: Top 10 countries with the highest average percentage of stunting in children between 2000-2021

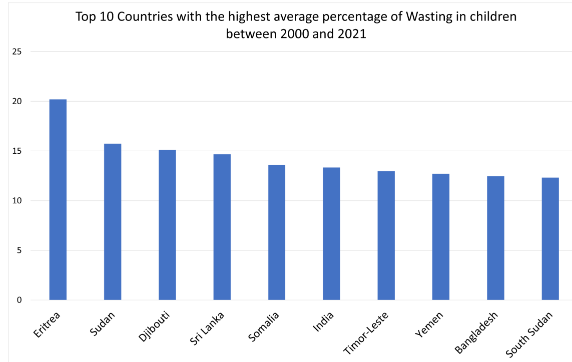

Figure 4: Top 10 countries with the highest average percentage of wasting in children between 2000-2021

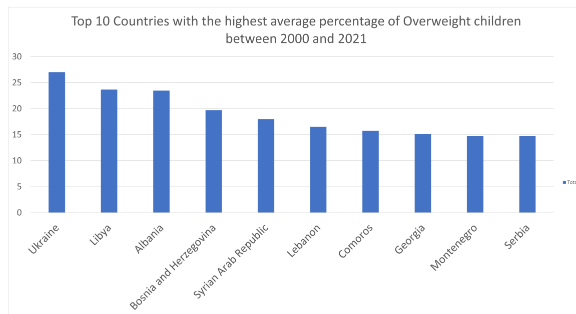

Figure 5: Top 10 countries with the highest average percentage of Overweight in children between 2000-2021

Diet Analysis Based on Gender

The following plots are analyzed based on the sex group of the children under 5 years.

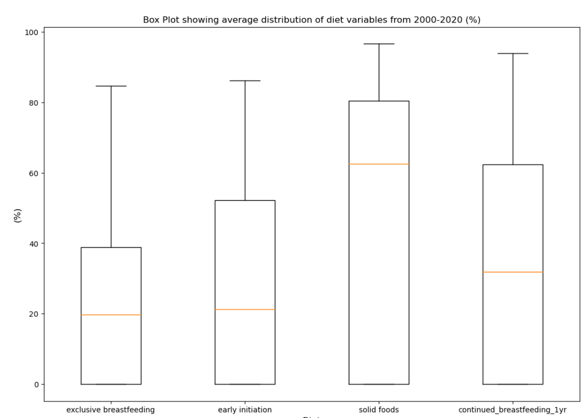

Figure 6: Box Plot showing the distribution between different diet variables across countries

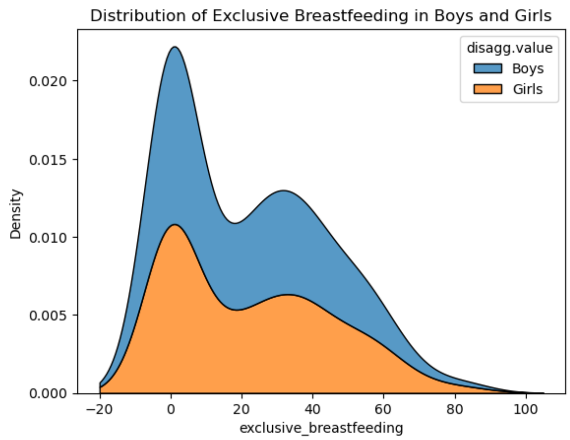

Figure 7: Probability density curve showing the distribution of exclusive breastfeeding among boys and girls globally.

Diet Analysis Based on Children's Age Group

Children under the age of 5 are further grouped into sub-groups.

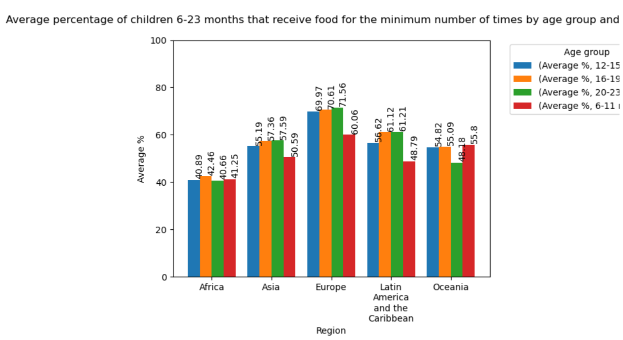

Figure 8: Bar plot showing the average percentage of children that receive food the minimum number of times by age group and region

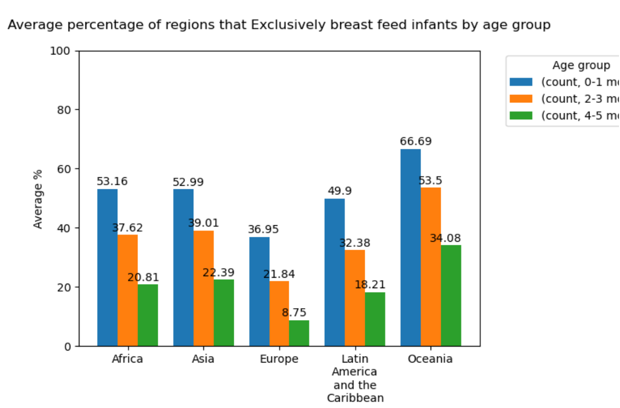

Figure 9: Average percentage of regions that exclusively breastfeed infants by age group

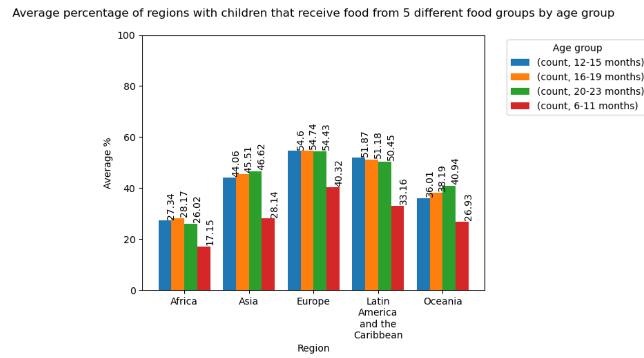

Figure 10: Average percentage of regions with children who received 5 different food groups by age group

3.4 Modeling and Evaluation

Model training uses deep learning regression and machine learning regression techniques. All models are trained on the Google Colab platform.

Regression Based on Deep Learning: a neural network that learns to map out input features to continuous output values is designed. The architecture consists of three layers. The input layer has 100 neurons; the hidden layers have 10 neurons and the output layer has 1 neuron. Before training, the dataset is combined together and split into train and test sets on an 80:20 ratio. The model's parameters (weights and biases) are updated using the Adam optimizer which computes the gradients and adapts the learning rate to the individual parameter. Parameters are updated to reduce loss error. To check and evaluate the model, the mean absolute error is used. The model is trained for 100 epochs and 200 epochs.

Regression Based on Machine Learning Techniques:

Six machine-learning regression models are trained in different algorithms the from sci-kit library using their base parameters. Linear regression, Random forest, Decision tree, Polynomial Regression, Ridge Regression, and Lasso Regression. Eighty percent of the datasets are used for training and twenty percent for testing. The mean squared error (MSE) is used as an evaluation metric. The model with the lowest MSE value is selected as the final model.
By using both deep learning and machine learning, we can compare the performance of the models.

4. Results

Results from the regression models developed are shown below.

Deep Learning Regression Model

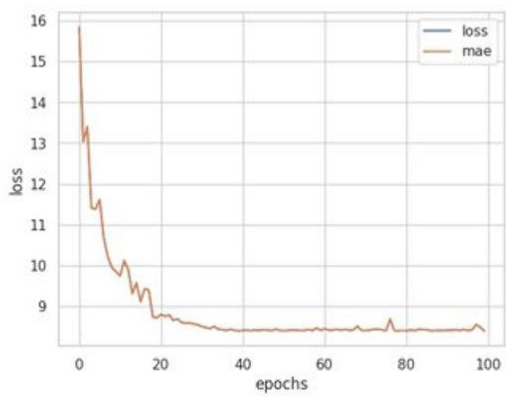

Figure 11: Result after training for 100 epochs

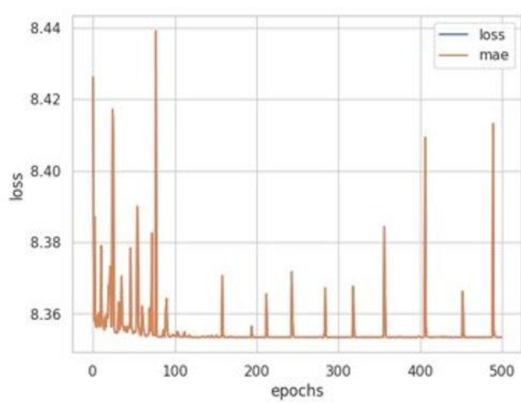

Figure 12: Result after the model is trained for 500 epochs, epochs shown only for 100 are since the history variable is overridden.

```
model_1.evaluate(X_test, y_test)

37/37 [==============================] - 0s 2ms/step - loss: 7.9043 - mae: 7.9043
[7.9043192863464355, 7.9043192863464355]
```

```
# Evaluate the model trained for 500 total epochs
model_2_loss, model_2_mae = model_1.evaluate(X_test, y_test)
model_2_loss, model_2_mae

37/37 [==============================] - 0s 3ms/step - loss: 7.8388 - mae: 7.8388
(7.838763236999512, 7.838763236999512)
```

Figure 13: The evaluation of model trained for Deep Learning Model 1

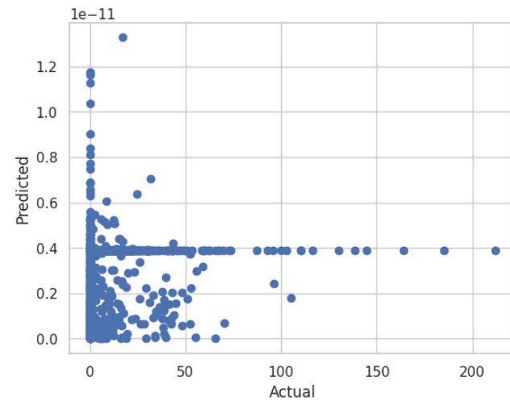

Figure 14: The plot actual test labels against the predicted labels in the Deep Learning Model 2

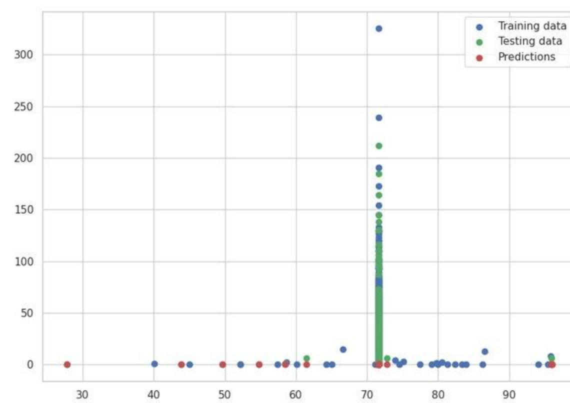

Figure 15: Plot comparison of training data, test data and predictions for Deep Learning Model 2

Machine Learning Models

S/N	Model	Mean Squared Error (MSE)	R-Squared Error (R^2)
1.	Linear regression	14.47	0.06
2.	Decision tree	14.53	0.002
3.	Lasso	14.47	0.00
4.	Ridge	14.47	0.01
5.	Random Forest	14.41	0.01
6.	Polynomial Features	2.85	-17926906880569209163612 6.00

Figure 13: Tabular results of all the machine learning models with their evaluation performance metrics.

5. Conclusion

The research aims to predict and prevent malnutrition in children under the age of 5 using data obtained from the Global Nutrition Report. The dataset comprises six fields relating to malnutrition at both the country and regional levels. The study develops predictive models to predict stunting, wasting, and overweight using diet features. The literature review discusses the application of AI and machine learning in solving problems across different sectors, including healthcare. Previous studies have used machine learning models to predict malnutrition in children, achieving high accuracy. The study contributes to efforts to reduce the burden of malnutrition in children under 5 by providing insights into the factors influencing malnutrition and developing predictive models to identify high-risk children.

On the basis of the findings, we can presume that the Polynomial Regression was moderately superior to any other ML algorithms used in this study to predict malnutrition status among under children. Using Deep Learning Models, to compile the model, Adam optimizer and mean absolute error was used.This research focused on the identification and prediction of major risk factors for stunting, wasting, and underweight using ML algorithms which will aid in reducing malnutrition among children.

6. References

[1] World Health Organization, "Malnutrition," https://www.who.int/en/news-room/fact-sheets/detail/malnutrition.

[2] World Health Organization, "Child mortality (under 5 years)," https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-under-5-mortality-in-2020.

[3] Global Nutrition, "Global Nutrition Report," https://globalnutritionreport.org/.

[4] World Health Assembly, "Global Nutrition Targets," https://www.who.int/publications/i/item/WHO-NMH-NHD-14.2.

[5] C. Zhang and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," J Ind Inf Integr, vol. 23, Sep. 2021, doi: 10.1016/j.jii.2021.100224.

[6] B. N. Limketkai, K. Mauldin, N. Manitius, L. Jalilian, and B. R. Salonen, "The Age of Artificial Intelligence: Use of Digital Technology in Clinical Nutrition," Current Surgery Reports, vol. 9, no. 7. Springer, Jul. 01, 2021. doi: 10.1007/s40137-021-00297-3.

[7] A. R. Lakshminarayanan, B. Pavani, V Rajeswari, S. Parthasarathy, A. A. Azeez Khan, and K. Javubar Sathick, "Malnutrition Detection using Convolutional Neural Network," in Proceedings of 2021 IEEE 7th International Conference on Bio Signals, Images and Instrumentation, ICBSII 2021, Institute of Electrical and Electronics Engineers Inc., Mar. 2021. doi: 10.1109/ICBSII51839.2021.9445188.

[8] A. Talukder and B. Ahammed, "Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh," Nutrition, vol. 78, Oct. 2020, doi: 10.1016/j.nut.2020.110861.

[9] S. M. J. Rahman et al., "Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach," PLoS One, vol. 16, no. 6 June 2021, Jun. 2021, doi: 10.1371/journal.pone.0253172.

[10] S. Ndagijimana, I. H. Kabano, E. Masabo, and J. M. Ntaganda, "Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning Techniques," J Prev Med Public Health, vol. 56, no. 1, pp. 41–49, Jan. 2023, doi: 10.3961/jpmph.22.388.

[11] J. Tariq, A. Sajjad, R. Zakar, M. Z. Zakar, and F. Fischer, "Factors associated with undernutrition in children under the age of two years: Secondary data analysis based on the Pakistan demographic and health survey 2012–2013," Nutrients, vol. 10, no. 6, Jun. 2018, doi: 10.3390/nu10060676.