Using Diet Analysis to Predict and Prevent Child Malnutrition

Nikhita Ramachandran Bertram Okwudili Nnalue Alphonse Emmanuel Y Boudouin Obiageli Jessica Okoli

Bilah Hajara Jimoh Saheed Olufemi Adejayan John Mikky

***Abstract*—Malnutrition in children remains a critical global issue affecting the physical and cognitive development of millions of young lives. Diet analysis plays a crucial role in identifying the nutritional deficiencies that contribute to malnutrition in young children. Nutrition data often comprises a wide range of variables, such as food intake, nutrient content, and socioeconomic factors. Machine learning algorithms can efficiently process and analyze these complex datasets to provide accurate predictions of malnutrition risk in children. This paper will apply machine learning regression analysis and deep learning techniques to the latest nutrition data and predict various indicators of Malnutrition. It also uses time series forecasting to predict the future values of these indicators.**

Keywords—Malnutrition, Machine learning, regression analysis, deep learning.

I. INTRODUCTION

Malnutrition refers to the lack of proper nutrition in a person's diet, leading to deficiency or excess of certain nutrients essential for growth and development. One of the most common types of malnutrition is protein-energy malnutrition (PEM). This occurs when individuals do not consume enough calories and protein to meet their daily needs. Another type of malnutrition is micronutrient deficiency, also known as hidden hunger. It occurs when individuals lack essential vitamins and minerals in their diet, such as vitamin A, iron, iodine, and zinc. Obesity can also be considered a form of malnutrition, commonly known as overnutrition or overeating. While malnutrition is typically associated with undernutrition, overnutrition has become a growing concern in many parts of the world.

Stunting, Wasting, Underweight, Overweight, and Obese are different aspects of malnutrition that reflect the various ways inadequate nutrition can impact an individual's health and well-being. Stunting is a condition that occurs when a child does not receive adequate nutrition in their early years, leading to impaired growth and development. It is characterized by the child being significantly shorter in height compared to their peers of the same age. Wasting, on the other hand, refers to a condition where an individual experiences rapid weight loss and muscle wasting due to severe malnutrition. It is often a consequence of acute malnutrition, where the individual's body is unable to absorb or utilize the nutrients consumed. Underweight individuals have a body mass index (BMI) below the healthy range and may have insufficient fat reserves and muscle mass. Contrary to underweight, overweight and obesity are conditions characterized by excess body weight and high levels of body fat. People become overweight or obese due to an energy imbalance caused by consuming more calories than they burn through physical activity and bodily processes.

This project aims to leverage machine learning, and deep learning techniques to analyse diet patterns in children under 5 years old, in order to predict various indicators of Malnutrition such as Stunting, Wasting, Underweight, Overweight, etc. In addition to this, it also aims to predict the micronutrient deficiency related indicators such as Iodized salt consumption from the analysis of diet data. It also aims to predict future values of the indicators of Malnutrition using time series forecasting. Instead of starting from scratch, this paper aims to build on the research conducted by Team Seaborn of the HDSC 2023 Winter Cohort.

II. LITERATURE REVIEW

In [1], the team Seaborn used deep learning regression and also machine learning regression-based techniques such as Linear regression, Random Forest, Decision tree, Polynomial Regression, Ridge Regression, and Lasso Regression for its model buildings and testing. Their 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. They analysed the diet variables such as exclusive breastfeeding, early initiation, solid foods, etc. to predict the burdens of malnutrition such as stunting, wasting and overweight of children. Malnutrition can be caused by lack or excess of both macronutrients as well as the micronutrients. However, the contribution of micronutrients towards malnutrition was not considered in the work done during the previous cohort.

In recent years, there has been growing concern about malnutrition among children under 5 years old, particularly in developing countries. Several studies have examined the trends and factors associated with childhood malnutrition. [2] worked on applying machine learning algorithms in predicting malnutrition in women from Bangladesh. Though a slight deviation on our target but of importance because it applied the same principles we hope to engage as well. The risk factors of malnutrition were identified using Multinomial Logistic Regression (MLR). For predicting the presence of malnutrition in these women, these ML algorithms were implemented. They include Naïve Bayes, Support Vector Machine (SVM), Decision tree, Artificial neural network, and random forest (RF) in which Random Forest model performed the best.[3] focused its work on predicting malnutrition in Indian kids through investigating the nutritional status of the kids. Nominal Logistic Regression was employed for the prediction. The work focused on four class variables of malnutrition namely the body mass index (BMI), Stunting, Wasting and Underweight.

Different methods of dietary pattern analysis such as the investigator-driven, data-driven, hybrid and compositional data analysis are discussed in [4]. The data-driven method of dietary pattern analysis from this research was useful in understanding how to perform diet analysis for our work. [5] used machine learning techniques like logistic regression, Linear discriminant analysis and random forest on NHANES dietary dataset to classify the population into four classes based on the evaluation metrics like accuracy, precision and recall score. The four classes are balanced, unbalanced, and nearly unbalanced. While this research performs dietary analysis, it does not relate it to the problem of Malnutrition. It simply evaluates whether a balanced diet is consumed or not.

Ghodsi et al. [6] conducted a comprehensive reanalysis of national surveys in Iran from 1998 to 2017, revealing a decreasing trend in the prevalence of stunting, underweight, and wasting. This study's findings underscore the need for continued efforts to address malnutrition, particularly in food-insecure regions.

Another avenue of research explored the prevention of malnutrition in children under 5 years old. The study by [7] focused on developing a prevention model for malnutrition in Iran. By proposing a comprehensive model involving various interventions and policy dimensions, the research aimed to tackle the underlying causes of malnutrition. The emphasis on key stakeholders, context-based policies, and regional considerations highlights the complexity of combating malnutrition effectively.

No work so far has adopted time series forecasting to predict the future values of the indicators of malnutrition. Also no research has studied the effect of micronutrient deficiency on malnutrition. No study has tried to predict the micronutrient deficiency related features that contribute to malnutrition. Every work has just predicted the standard stunting, wasting , underweight and overweight.

III. Methodology

A. Dataset

The dataset for this study is sourced from the UNICEF data warehouse. In this data warehouse, the Malnutrition data is available from the year 1970 to 2022. This data is available across 346 geographic regions and 608 indicators of Malnutrition are available for exploration. Since our study mainly focuses on diet analysis for malnutrition prediction, a customized dataset was sourced from this data warehouse containing only the Nutrition related features [8].

B. Data preprocessing

Initially in the dataset the columns were named as REF_AREA:Geographic area, INDICATOR:Indicator, etc. They were renamed with the help of lambda functions. Now, the dataset contained all the indicators of Malnutrition under the Indicator column and its corresponding value was present under the Observation value column. However, we required the individual indicators to be separate columns and their observation values to be present under that specific column. So, we used pivot_table method to achieve this and used mean as the aggregation function. We created two subsets of the dataset namely; country_wise_avg_df.csv and year_wise_avg_df.csv. The country_wise_avg_df.csv was used for the training and testing of machine learning and deep learning models while the time_wise_avg_df.csv was used for the time series forecasting model.

C. Exploratory Data analysis

For the exploratory data analysis,a time period of 19 years observation starting from 2004 to 2022 was considered. As per the analysis, South Sudan is the country affected mostly by Severe wasting having 10.1% of under five children who are severely wasted, followed by India (7.0%) and Papua New Guinea (6.38%). The countries experiencing the least severe wasting are Poland, Latvia, and Belgium.The top countries having the most underweight cases over the time period are Timor-Leste with 40.05%, Eritrea with 38.5%, and Yemen with 38.07% followed by India with a score of 35.6%. Latvia is safe and has no case of Underweight at all. Turks and Caicos islands, Australia and Estonia come next in the list of countries not having underweight cases amongst their children with scores of 0.41%, 0.50% and 0.53% respectively.Again, South Sudan tops the chart of most countries with wasting cases with a score of 22.4%. India follows closely with a score of 19.4% then Sri lanka with a score of 15.9%. Some countries such as the United Kingdom, Australia and Chile boasts of having the least wasting cases with wasting scores 0f 0.17%, 0.20% and 0.31% respectively. Stunting cases are on the rise but the most affected countries happen to be Burundi (54.3%), Timor-Leste (52.6%) and then Eritrea (51.0%). Some countries seem less affected by the high rise of stunting cases. Latvia, Poland, Estonia and Germany happen to sit comfortably on the top of the least affected with scores of 0.54%, 0.94%, 1.17% and 1.21% respectively. In Libya,25.1% of under five children are overweight. Other countries most affected are Bosnia and Herzegovina (21.3% each) followed by Albania (21.2%). Overweight seems not to be popular amongst kids in Djibouti, Niger and Burkina Faso because their scores seems the lowest sitting at 0.74%, 0.75% and 1.07% respectively. Turkmenistan, Serbia and El Salvador happen to be countries that the highest scores for the minimum meal frequency. This implies that they do not miss meals. The scores are 94.8%, 90.7% and 86.5% respectively. There are countries that cannot boast of the fact that they do not miss meals. South Sudan, Central African Republic and Uzbekistan happen to be the top 3 of least countries with minimum meal frequency.No doubt that breastfeeding affects the nutritional growth of a child. Countries that boast of consistent breastfeeding are Democratic People's Republic of Korea (99.4%), followed by Sri Lanka (99.3%) and then Bhutan (99.1%). The United States sits on the top of countries that has reduced breastfeeding with a score of 72.6%, followed by Botswana with 80.5% and South Africa with score of 82.2%. Introduction to solid or semisolid foods within the early months is also a deciding factor to nutritional growth. These countries boast of early introduction to solid foods. They are Argentina (96.4%), Belarus (95.3%) and Costa Rica (94.5%). Countries like Somalia, Mali and South Sudan are among the ones that do not expose kids early to solid foods. Their scores are 18.0%, 40.6% and 42.8% respectively. Some of this visualized data is shown in Figure 1 to 6.

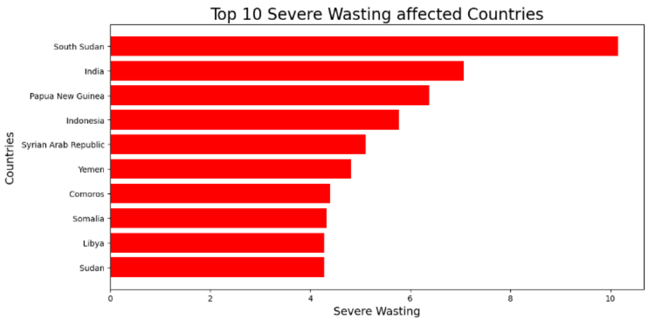

Fig. 1. Top 10 Severe Wasting affected countries.

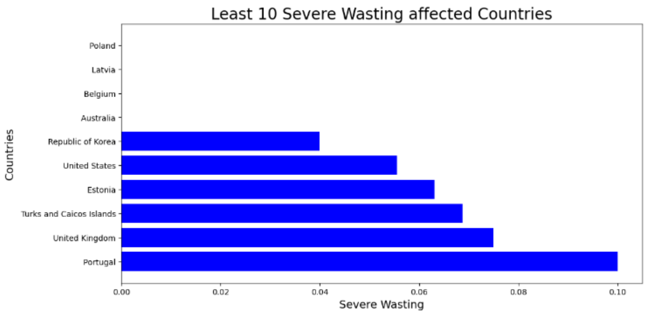

Fig. 2. Least 10 Severe wasting affected countries

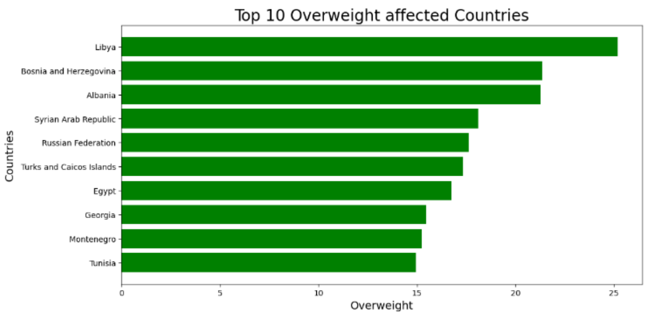

Fig. 3. Top 10 Overweight affected countries

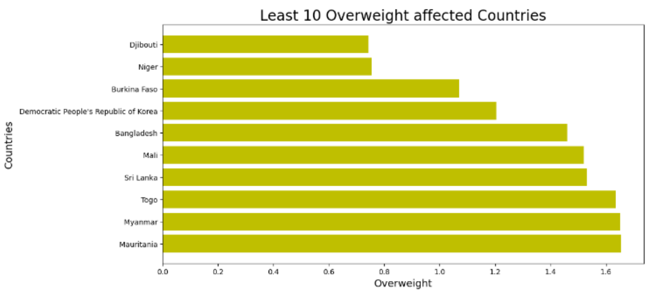

Fig. 4. Least 10 Overweight affected countries

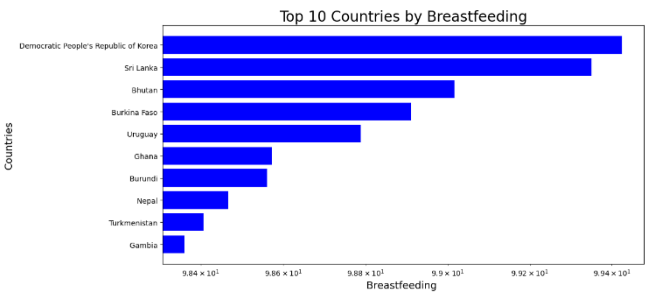

Fig. 5. Top 10 countries by breastfeeding.

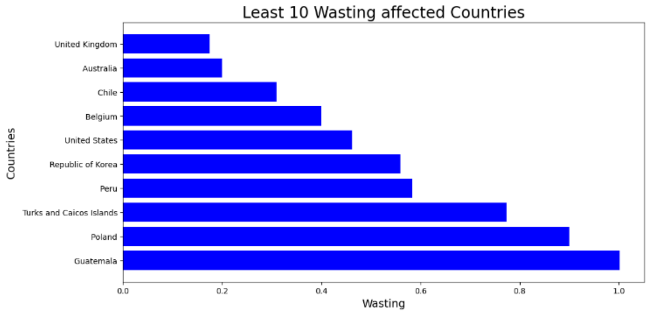

Fig. 6. Least 10 wasting affected countries.

Some other visualizations that was done during data exploration is shown in Figure 7 and 8.

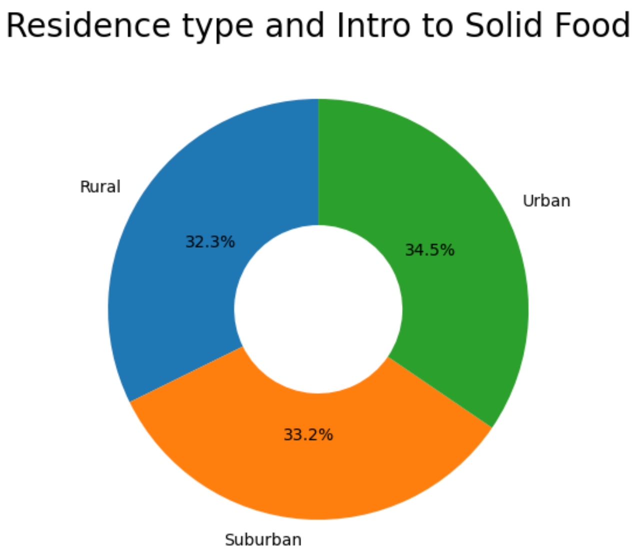

Fig. 7. Residence type and introduction to solid food

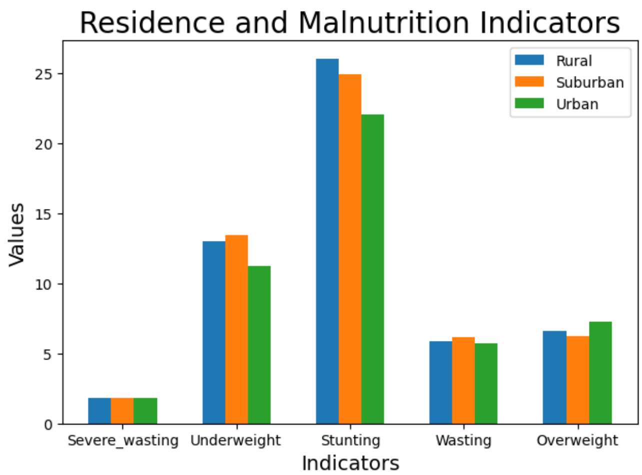

Fig. 8. Residence and Malnutrition indicators.

D. *Feature engineering and Selection*

Place Standardization of data was done using StandardScaler. Important features from the decision tree were selected by first training a decision tree model and then getting the important features from it. Significance threshold of 0.001 was selected and only the features that had their feature importance score above this threshold were selected for further processing. Among the selected features , Correlation was checked using a correlation matrix. The features were checked for multicollinearity keeping the correlation threshold at 0.9. The variance Inflation factor for the features was calculated and it was found that Multicollinearity was present among the features. To solve this we transformed the correlated features using Principal Component Analysis.The number of components was set to 5. Later this PCA transformed features were concatenated with the uncorrelated features. This combined data was then used for model training and evaluation.

E. Model Training using Machine Learning

The unicef_country_wise_avg.csv file was split into X and y. The response variables i.e., y are Underweight, Severe Underweight, wasting, severe wasting, overweight, Obese, stunting, Severe Stunting, Mean BMI-for-age, Mean Height-for-age, Mean Weight-for-height,indicators related to Vitamin A dosage and Iodized salt consumption and some other miscellaneous indicators. The remaining indicators such as breastfeeding data, various food group consumption data constitute the predictors i.e., X. An 80:20 split of data for training and testing purposes is done. Models are built using ElasticNet, XGBoost, Linear Regression and Gradient Boosting.

F. Model Training using Deep Learning

A convolutional neural network with relu activation layer and 14 neurons in the output layer was designed. The reason behind having 14 neurons in the output layer is because we have 14 target variables or regressors. A recurrent neural network consisting of LSTM layers was also designed. Adam optimizer was used, training was done for 50 epochs and the loss was measured in the form of mean squared error in case of both the models. The trained model was then used for making the predictions.

G. Time series forecasting

Before doing the time series forecasting, the date column was converted to the data type datetime. Facebook prophet was used for performing the time series forecasting as it is robust to outliers and does not get affected by missing data. Predictions for next five years were done and the predictions were plotted along with the observed data points and the uncertainty interval.

H. Dashboard

Interactive dashboard for visualizing Malnutrition data was developed. This dashboard was developed in python using plotly and dash. Governments, healthcare professionals, and communities can use this dashboard to understand the trends and combat malnutrition.

IV. Results

The results of the machine learning and deep learning models are shown in Table 1. The evaluation metric used for evaluating the performance of our model is Mean squared Error. The lower the Mean squared error, the better is the model performance. From the comparison of the Mean squared error values we can see that ElasticNet has performed the best with the least Mean squared value followed by CNN and Gradient Boosting Regressor. The mean squared of both the Deep learning and Machine learning models are less than 0.3.

Sr. No	Model	Mean squared error
1	ElasticNet	0.17895505880978693
2	XGBoost	0.2355057192012161
3	Linear Regression	0.20068247400023106
4	Gradient Boosting	0.18564662790538358
5	CNN	0.18418040820425466
6	RNN	0.2705734372138977

Table 1. Machine Learning and deep learning model results

The results of time series forecasting are shown in Figures 9 to 15. But taking a look at the predictions, one can observe that there will likely be a slight spike in the global stunting case in the year 2024 to 2025 before normalcy returns and the decline cases continue again. We expect a slight increase in severe stunting cases globally from 2022 to 2023 but the number will be steady all through till 2024. From hence we experienced a drop before becoming slightly constant once again till 2026. Globally, we shall be expecting a massive drop in underweight cases starting from 2022. Although it tried to raise its ugly head in 2025 but it was defeated and it continued its downward trend.Cases of severe underweight have been haphazardly according to records but we expect to see a more constant stretch of its occurrence globally from 2025 upwards from our predictions.We predict a massive downward trend of Wasting cases globally starting from 2022 except for a slight increase that occurred around 2025 but it was quickly curtailed. The prediction for the global severe wasting cases is similar to that of wasting cases. There is a downward trend which tried to increase in the year 2023 to 2024 but quickly reverted to its downward movement. This indicates lesser cases of severe wasting worldwide.The overweight trend shows an upward movement from the year 2022 though it dropped a bit in 2024/2025, it picked up its upward motion immediately. So it is expected that the world will experience more cases of overweight.We expect a huge spike in cases of obesity globally starting from 2023 then it settles for an almost constant number for the successive years.That sudden spike is a matter of great concern.

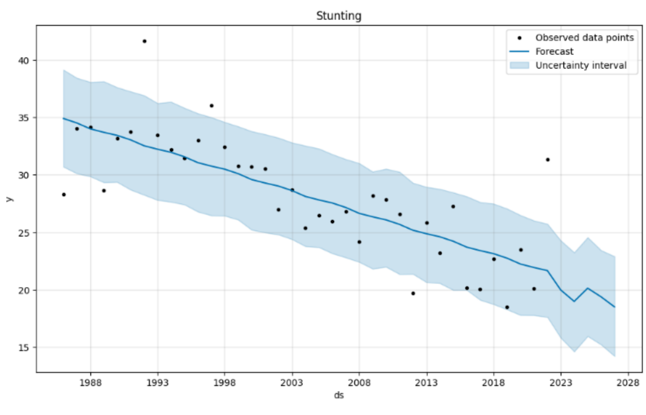

Fig. 9. Global annual Stunting forecast

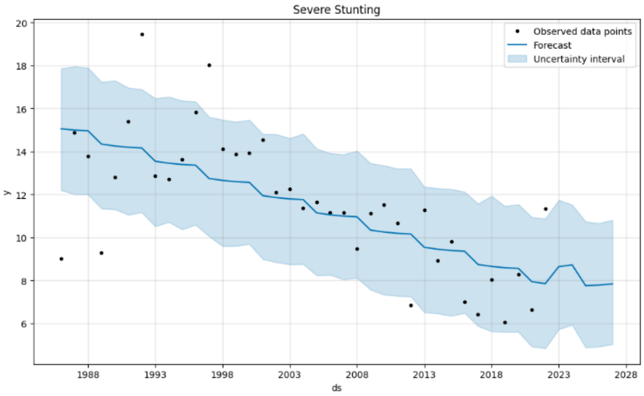

Fig. 10. Global Annual Severe Stunting

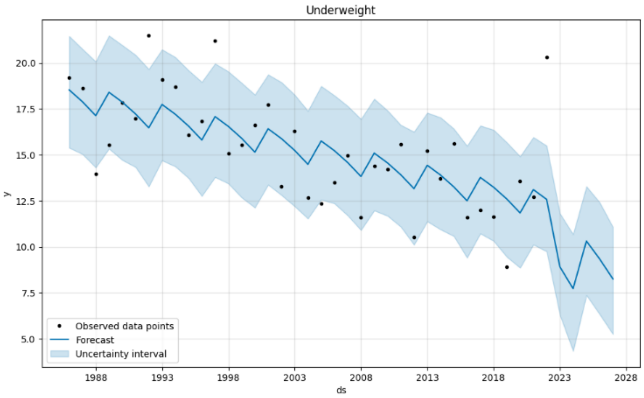

Fig. 11. Global Annual Underweight forecast

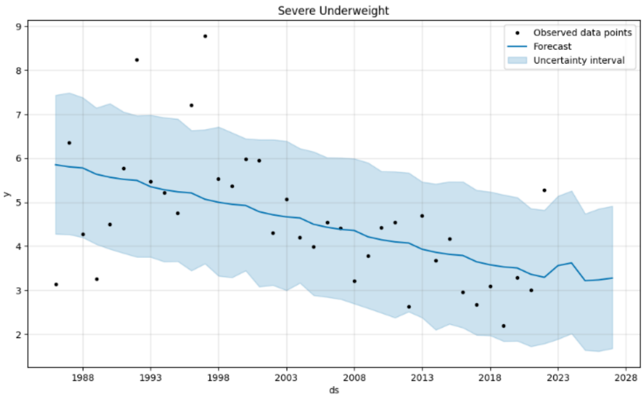

Fig. 12. Global Annual Severe Underweight forecast

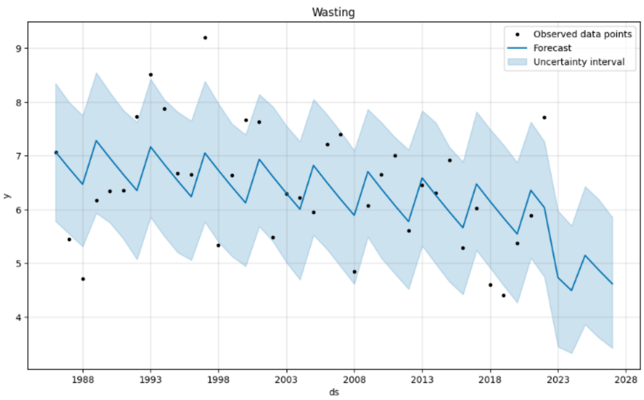

Fig. 13. Global Annual Wasting forecast

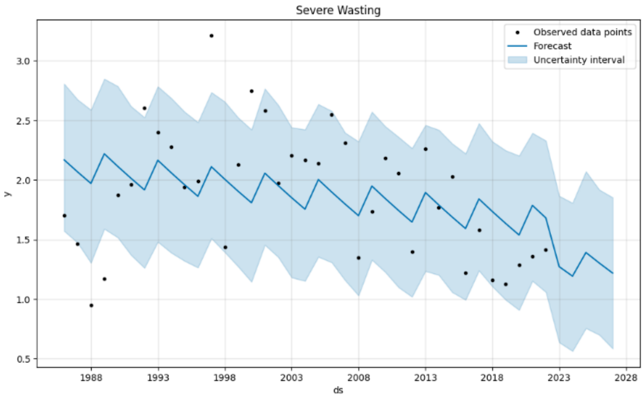

Fig. 14. Global Annual Severe wasting forecast

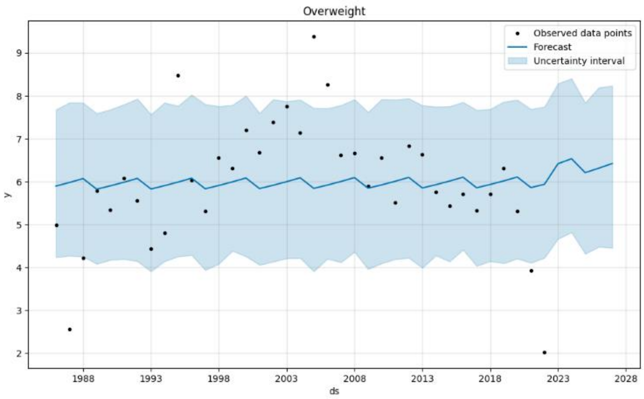

Fig. 15. Global Annual Overweight forecast

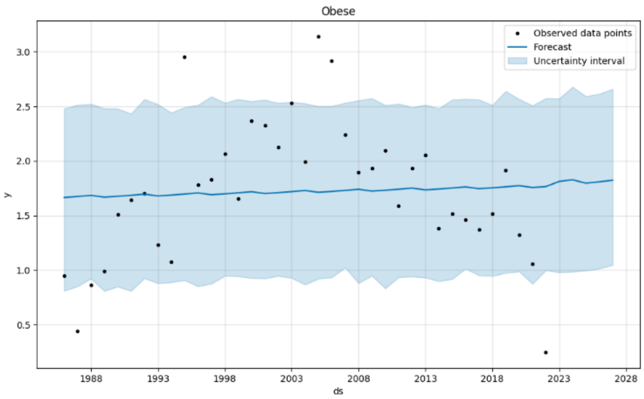

Fig. 16. Global Annual Obesity Forecast

V. Conclusion

Our machine learning and deep learning models have performed better compared to the work done by the Seaborn team in the previous. Also, we have carried out extensive diet analysis by considering a wider range of diet related features. Also we have predicted indicators of malnutrition that are related to micronutrient deficiencies. With interactive dashboards, the work has become more user friendly.

To combat malnutrition, interventions should focus not only on increasing access to nutritious food but also on educating individuals about the importance of a balanced diet and healthy lifestyle choices. Governments, healthcare professionals, and communities can use these dashboards to better understand the situation and educate people.

References

[1] Team Seaborn, " Using Diet Analysis to Predict and prevent child malnutrition."

[2] Md. Merajul Islam, Md. Jahanur Rahman, Md Moidul Islam, Dulal Chandra Roy, N.A.M. Faisal Ahmed, Sadiq Hussain, Md Amanullah, Md. Menhazul Abedin, Md. Maniruzzaman, "*Application of machine learning based algorithm for prediction of malnutrition among women in Bangladesh*", International Journal of Cognitive Computing in Engineering, Volume 3, 2022, doi: https://doi.org/10.1016/j.ijcce.2022.02.002.

[3] Sangita Khare, S Kavyashree, Deepa Gupta, Amalendu Jyotishi, "*Investigation of Nutritional Status of Children based on Machine Learning Techniques using Indian Demographic and Health Survey Data*", Procedia Computer Science, Volume 115, 2017,.doi: https://doi.org/10.1016/j.procs.2017.09.087.

[4] S. Ghosh , "Factors Responsible for Childhood Malnutrition: A Review of the Literature .Curr Res Nutr Food Sci 2020, doi : http://dx.doi.org/10.12944/CRNFSJ.8.2.01

[5] Zhao, J., Li, Z., Gao, Q. et al. "*A review of statistical methods for dietary pattern analysis".* Nutr J 20, 37 (2021). https://doi.org/10.1186/s12937-021-00692-7

[6] D. Ghodsi et al., "Prediction and analysis of trends in the nutritional status of children under 5 years in Iran: reanalysis of the results of national surveys conducted between 1998 and 2020,"

[7] M. Mosheni, A. Aryyankhaesal, "A model for prevention of malnutrition among children under 5 years old in Iran,"

[8] Data Warehouse - UNICEF DATA