

HDSC Winter 23 Capstone Project Presentation

A Project by Team SEABORN



Using Diet Analysis to Predict and Prevent Child Malnutrition

TEAM SEABORN

1. Project Lead: Kar Woh Chong
2. Assist Project Lead: Anchal Sekhri
3. Query Analysts: Nafisa Opemi Jimoh
4. Presenter: Kar Woh Chong
5. Slide designer: Abel Enagbonmhena
6. Data cleaning & preprocessing: Rofiat Adetutu Alimi
7. EDA: Zion Oluwasegun
8. Modelling: kar Woh Chong, Anchal Sekhri
9. Evaluation: Anchal and Kar Woh
10. Research Paper: Nafisa Jimoh, , Aleksander Busz Fabritius, kar Woh

The submission included: a [research article](#), [slides](#), [codes](#) & [Dataset](#)



Introduction

As stated by the World Health Organization (WHO), [malnutrition](#) remains one of the major global health challenges, particularly in low and middle-income countries. Malnutrition is an illness that is developed when the body is deprived of a balanced and nutritious diet which impedes the body's growth and development in both children and adults. [Undernutrition](#) is a significant contributor to child mortality, accounting for an estimated 45% of all deaths among children under the age of five worldwide. [Stunting, wasting and underweight](#) are forms of undernutrition which negatively affects children and puts them at a higher risk of death from common childhood illnesses such as diarrhoea, pneumonia, and malaria. This is as a result of the body not receiving the adequate nutrition needed to strengthen the immune system. It was estimated by [WHO in 2020](#) that [149 million children were stunted, 45 million were wasted while 38.9 million were obese](#) which is another form of malnutrition. In order to ensure that these amounts are reduced across the globe, the World Health Assembly has paid keen attention to the factors which contribute and are responsible for eliminating malnutrition in children and adults. The consumption of a nutritious diet is one of these factors and is highly paramount in ensuring that a child's body gets the right and sufficient nutrients required. In order to reduce the burden of malnutrition by preventing child stunting, wasting, and being overweight, it is important to analyse the elements of a child's diet.

Problem Statement

This project seeks to use machine/deep learning algorithms [to build models used for predicting malnutrition](#) (stunting, wasting, overweight) in children under the age of 5 using the child's diets as the predictor.

Aim and Objectives

The **aim** is to use the diet feature of children to predict and prevent malnutrition in children under the age of 5.

The objectives include:

1. Investigating the elements of the child's diet.
2. [Developing a machine/deep learning model](#) to predict malnutrition in children under 5.
3. [Evaluating and deploying the model](#).



Data Understanding

The dataset is obtained from the [global nutrition website](#). The Global Nutrition Report captures the state of nutrition and progress towards the global nutrition targets at the country, regional and global level. To predict and prevent malnutrition in children using diet analysis the diet, and the burden of malnutrition datasets are used.

Diet: The diet dataset comprises variables such as exclusive breastfeeding, early initiation, solid foods, minimum diet diversity, minimum accept diet and other nutrients which are important for both children and adults.

Burden of Malnutrition: contains variables such as stunting, wasting, low birth weight, overweight of children who are below 5 years and other measures of malnutrition for adults in both country and regional levels. These values are given in percentage.

Data Preparation

1. Reducing Cardinality (Number of Features)

The features from both the diet and burden of malnutrition were reduced to only have the features relating to children under 5. After cleaning, the diet dataset had 126 columns and 3242 rows while the burden of malnutrition comprised 67 columns and 5906 rows. The features from the burden of malnutrition were summed up together to get the overall wasting, stunting and overweight.

2. Missing Values, Duplicates and Outliers

All missing values were replaced with 0.

Exploratory Data Analysis

EDA is an important step in the data analysis process because it allows us to understand the data before applying any statistical models or making any decisions.. After carrying out EDA, the following observations were made.

- Africa and Asia have recorded the highest number of stunting and wasting cases from 2000 - 2021.
- Europe has the highest number of severe overweight cases, followed by North America.

- Burundi, an African country has the highest percentage of stunted children over the past two decades while Ukraine is leading with the highest number of overweight cases in children under the age of 5.
- Regions in Europe have the highest percentage of children receiving food from all age groups while regions in Africa have the lowest percentage.
- Children within the age of 20-23 months have the highest percentage of receiving food
- Children within the age of 0-5 months are exclusively breastfed In different regions.
- Across all regions the percentage of children who are exclusively breastfed reduces as age increases.
- Globally boys are more affected by the burden of malnutrition than girls.

Prevalence of Malnutrition in Children under the age of 5 at Regional Level

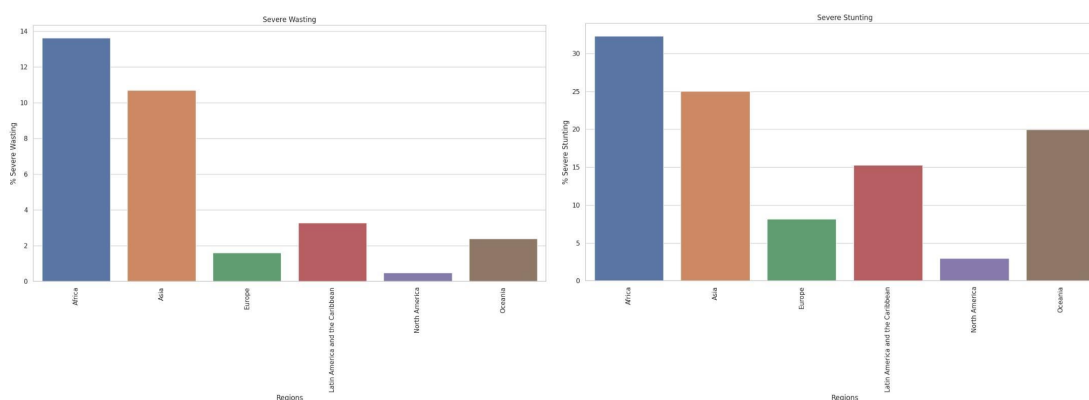


Figure 1: Prevalence of stunting, wasting, overweight across regions (2000-21)

Top 10 countries with the most cases of malnutrition from 2000 - 2021

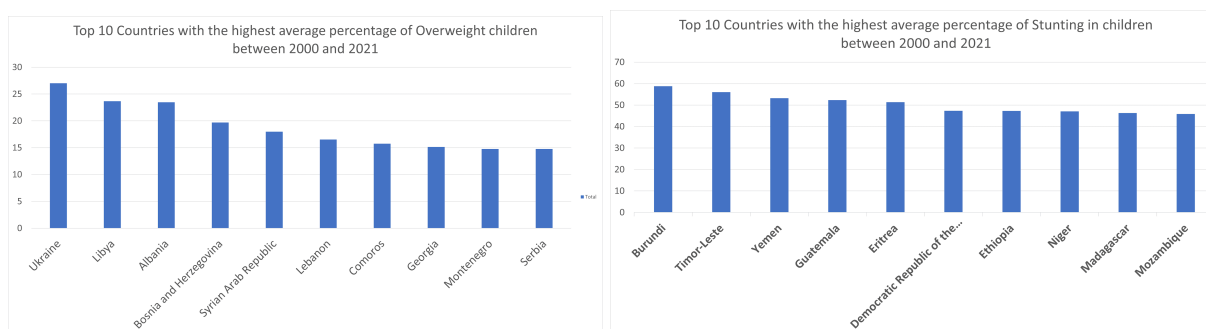
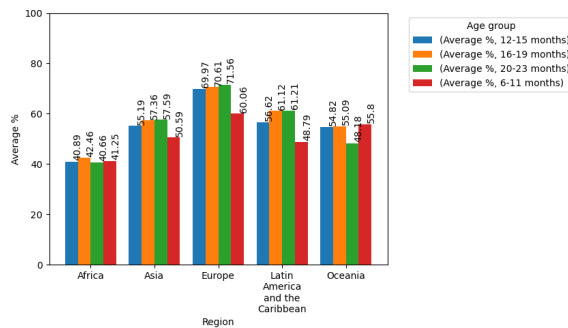
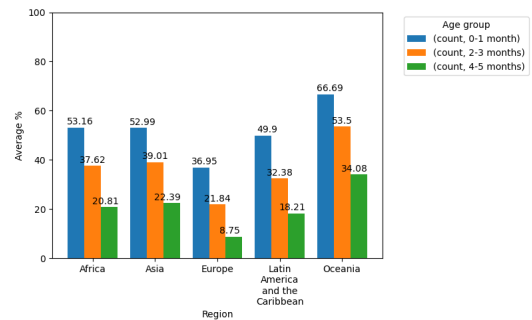


Figure 2: Top 10 leading countries with highest cases of stunting and overweight from 2000-2021

Average percentage of children 6-23 months that receive food for the minimum number of times by age group and region



Average percentage of regions that Exclusively breast feed infants by age group



Modelling & Evaluation

The **deep learning regression** and **machine learning regression** algorithms were used to build models. The **response variables** are **stunting, wasting and overweight** while the **predictors** are **all the elements of the diet dataset**.

The following procedures are carried out.

1. Train and Test Sets

The training and testing sets were formed from an 80/20 split (respectively) of the dataset.

2. Deep Regression Model

Three layers were used, the input, hidden and output layer comprising 100, 10 and 1 neuron respectively. To compile the model, Adam optimizer and mean absolute error was used. The model was trained for 100 and 500 epochs.

```
[ ] import tensorflow as tf

# Set random seed
tf.random.set_seed(42)

# Add an extra layer and increase number of units
model_1 = tf.keras.Sequential([
    tf.keras.layers.Dense(100), # 100 units
    tf.keras.layers.Dense(10), # 10 units
    tf.keras.layers.Dense(1) # 1 unit (important for output layer)
])

# Compile the model
model_1.compile(loss=tf.keras.losses.mae,
                 optimizer=tf.keras.optimizers.Adam(), # Adam works but SGD doesn't
                 metrics=['mae'])

# Fit the model and save the history (we can plot this)
history_1= model_1.fit(X_train, y_train, epochs=100, verbose=0)
```

Deep Learning Regression Model Building (1)

```
[ ] import tensorflow as tf

# Assume X_train_padded has a shape of (num_samples, max_length)
# and y_train has a shape of (num_samples,)

# Reshape X_train_padded to add a third dimension
max_length = 100

# Define your model as usual
model = tf.keras.Sequential([
    tf.keras.layers.Conv1D(32, 5, activation='relu', input_shape=(124, 1)),
    tf.keras.layers.MaxPooling1D(3),
    tf.keras.layers.Conv1D(32, 5, activation='relu'),
    tf.keras.layers.MaxPooling1D(3),
    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(1, activation='sigmoid')
])

# Compile the model as usual
model.compile(optimizer='adam', loss='mean_absolute_error', metrics=['mae'])

# Fit the model with the reshaped input data
model.fit(X_train, y_train, epochs=100)
```

Deep Learning Regression Model Building (2)

```
model_1.evaluate(X_test, y_test)

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

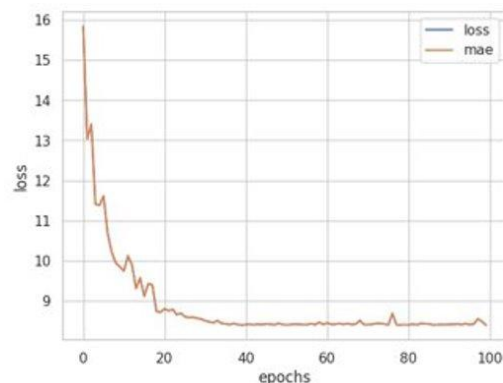
Evaluation Result (1): 100 epochs

```
# 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)
```

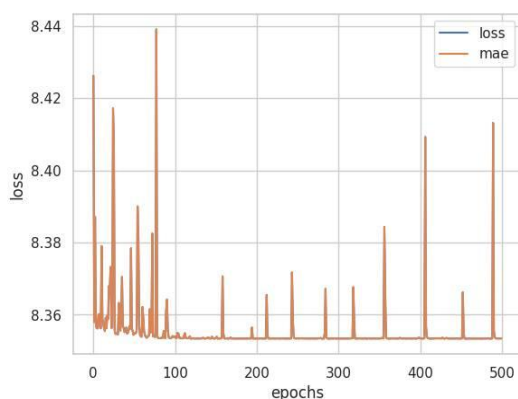
Evaluation Result (1): 500 epochs

Loss Curve Plots after 100 epochs



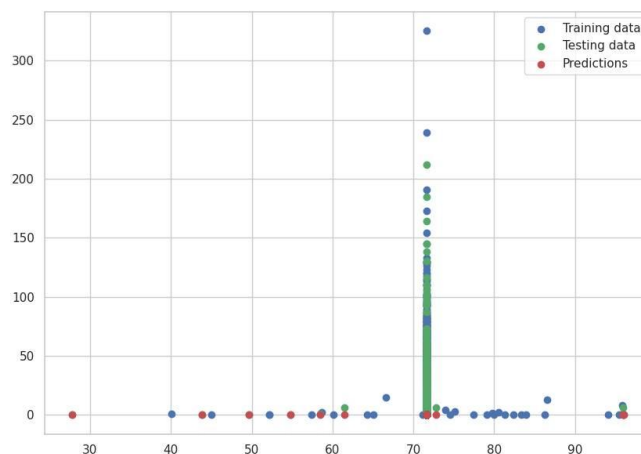
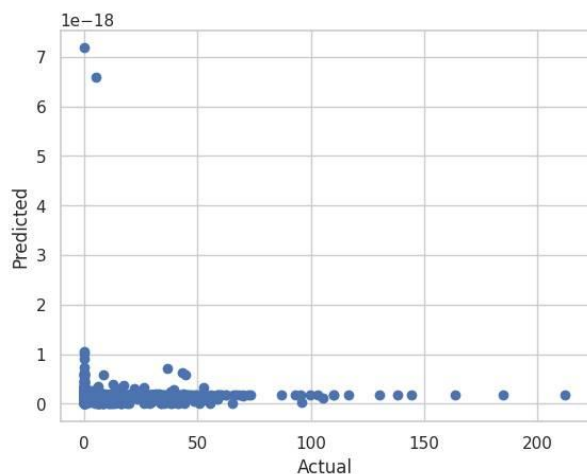
Loss Curve after 100 epochs

Based on the plot above, the model's loss experiences stability after 20 epochs. There was no further improvement to the decrease in the loss value as the number of epochs increased.



Loss Curve after 500 epochs

Evaluation:



3. Machine Learning Models

Six machine learning algorithms were used to develop the predictive models. Mean absolute error and R2 metrics were used for evaluation. The [Polynomial Regression](#) outperformed other models with [lowest MSE of 2.85](#).

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	-179269068805692091636121 6.00

Results from Machine Learning Models

Recommendations

It is recommended that:

- targeted interventions be set up to encourage and support exclusive breastfeeding in the first five months of an infant's life.
- increased efforts be made to improve the dietary diversity in regions with lower percentages of children receiving food from multiple groups, particularly in younger age groups.

Data on more dietary nutrients are needed for further analysis.

Conclusion

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.

References

- Investigate the risk factors of stunting, wasting, and underweight among under-five Bangladeshi children and its prediction based on machine learning approach. [link](#)
- World Health Organization, "Malnutrition," [link](#)
- 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. [link](#)