NUTRITIONAL DEFICIENCY PREDICTION BY REGION USING MACHINE LEARNING

ANNE DASHINI KANNAN

UNIVERSITI TEKNOLOGI MALAYSIA

PSZ 19:16 (Pind. 1/23)

UNIVERSITI TEKNOLOGI MALAYSIA
DECLARATION OF THESIS

Author's full name	:	ANNE DASHINI KANNAN		
Student's Matric No.	:	MCS241013	Academic Session	: 20242025-01
Date of Birth	:	21ST OCTOBER 2000	UTM Email	: annedashini@graduate.utm.my
Thesis Title	:	NUTRITIONAL DEFICIENCY PREDICTION BY REGION USING MACHINE LEARNING		

I declare that this thesis is classified as:

☒ **OPEN ACCESS** — I agree that my report to be published as a hard copy or made available through online open access.

☐ **RESTRICTED** — Contains restricted information as specified by the organization/institution where research was done. *(The library will block access for up to three (3) years)*

☐ **CONFIDENTIAL** — Contains confidential information as specified in the Official (Secret Act 1972)

(If none of the options are selected, the first option will be chosen by default)

I acknowledged the intellectual property in the thesis belongs to Universiti Teknologi Malaysia, and I agree to allow this to be placed in the library under the following terms :

1. This is the property of Universiti Teknologi Malaysia
2. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.
3. The Library of Universiti Teknologi Malaysia is allowed to make copies of this thesis for academic exchange.

Signature of Student:
Signature :

Full Name : ANNE DASHINI KANNAN
Date : 17TH JANUARY 2025

Approved by Supervisor

Signature of Supervisor I:

Full Name of Supervisor I

Date : 17TH JANUARY 2025

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

This letter should be written by a supervisor and addressed to Perpustakaan UTM. A copy of this letter should be attached to the thesis.

Date: 17th January 2025

Librarian

Jabatan Perpustakann UTM,

University Teknologi Malaysia,

Johor Bahru, Johor

Sir,

CLASSIFICATION OF THESIS AS RESTRICTED/CONFIDENTIAL

TITLE: NUTRITIONAL DEFICIENCY PREDICTION BY REGION USING MACHINE LEARNING

AUTHOR'S FULL NAME: ANNE DASHINI KANNAN

Please be informed that the above -mentioned thesis titles Nutritional Deficiency Prediction by Region using Machine Learning should be classified as RESTRICTED/CONFIDENTIAL for a period of three (3) years from the date of this letter. The reason for these classifications are

(i)

(ii)

(iii)

Thank you.

Yours sincerely,

SIGNATURE:

NAME:

ADDRESS OF SUPERVISOR:

"I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in term of scope and quality for the award of the degree of Master in Data Science"

Signature : _____

Name of Supervisor I :

Date : 17TH JANUARY 2025

NUTRITIONAL DEFICIENCY PREDICTION BY REGION USING MACHINE LEARNING

ANNE DASHINI KANNAN

A thesis submitted in fulfilment of the requirements
for the award of the degree of Master in Data
Science

Faculty of Computing
Universiti Teknologi Malaysia

JANUARY 2025

DECLARATION

I declare that this thesis entitled *"Nutritional Deficiency Prediction By Region Using Machine Learning"* is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature	:	.....
Name	:	ANNE DASHINI KANNAN
Date	:	17TH JANUARY 2025

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. I wish to express my sincere appreciation to my subject lecturer, Dr Shahizan, for encouragement, guidance, critics and friendship. Without their continued support and interest, this thesis would not have been the same as presented here.

I am also indebted to Universiti Teknologi Malaysia (UTM) for funding my master study. Librarians at UTM also deserve special thanks for their assistance in supplying the relevant literatures.

My fellow postgraduate student should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family member.

ABSTRACT

Malnutrition is still a burning problem in the population affecting health and production, hampering growth and development in the most affected areas. These deficiencies differ greatly by regions depending on the presence or absence of certain foods, one's economic status, supplier of health care services, and the surrounding environment. It is crucial to predict nutrient inadequacy to design interventions and allocate resources for their enhancement by region accurately.The current paper puts forward a machine learning method to predict malnutrition more effectively at the regional level by incorporating a wider range of variables based on new demographic, dietary, health besides environmental variables. The data collection is done from different sources and cleaning of data where some features have missing values and some features have been normalized to improve the ability of the models that will be used in the research. A collection of supervised machine learning techniques such as random forest, support vector machines, gradient boosting, as well as deep neural networks are used to capture full interaction patterns of this data. Accuracy, Precision, Recall and F1-score are widely used for model performance measurement, while cross-validation confirms its efficiency and universality.The results indicate that it is possible to accurately identify deficiencies with the help of machine learning algorithms. In this research, feature importance analysis helps identify key factors influencing the next year's output that can be useful for decision-makers, including regional diet patterns, household income, and healthcare facility accessibility, and environmental characteristics. In addition, the authors focus on the applicability of the identified framework to other regions and age populations as well. As it can determine vulnerability hotspots, this study helps prevent/implement optimal nutritional interventions to sensitive populations. This work demonstrates the effectiveness of machine learning for the analysis of the given global public health problems and make an attempt to fill the gap between the technical approaches and tangible health intervention that can help to fight nutritional deficiencies.

ABSTRAK

Kekurangan zat makanan masih merupakan masalah besar dalam populasi yang menjejaskan kesihatan dan pengeluaran, menghalang pertumbuhan dan pembangunan di kawasan yang paling terjejas. Kekurangan ini berbeza dengan ketara mengikut kawasan bergantung kepada kehadiran atau ketiadaan makanan tertentu, status ekonomi seseorang, penyedia perkhidmatan penjagaan kesihatan, dan persekitaran sekeliling. Adalah penting untuk meramalkan kekurangan nutrien bagi merancang intervensi dan memperuntukkan sumber untuk peningkatannya mengikut kawasan dengan tepat.Kertas kerja semasa mencadangkan kaedah pembelajaran mesin untuk meramalkan kekurangan zat makanan dengan lebih berkesan di peringkat wilayah dengan menggabungkan pelbagai jenis pembolehubah berdasarkan pembolehubah demografi, pemakanan, kesihatan serta persekitaran yang baru. Pengumpulan data dilakukan dari pelbagai sumber dan pembersihan data di mana beberapa ciri mempunyai nilai yang hilang dan beberapa ciri telah dinormalisasi untuk meningkatkan keupayaan model yang akan digunakan dalam penyelidikan. Sekumpulan teknik pembelajaran mesin terawasi seperti hutan rawak, mesin vektor sokongan, pengukuhan kecerunan, serta rangkaian neural mendalam digunakan untuk menangkap corak interaksi penuh data ini. Ketepatan, Ketepatan, Ingatan dan Skor F1 digunakan secara meluas untuk pengukuran prestasi model, manakala pengesahan silang mengesahkan kecekapan dan keseragamannya.Keputusan menunjukkan bahawa adalah mungkin untuk mengenal pasti kekurangan dengan tepat dengan bantuan algoritma pembelajaran mesin. Dalam penyelidikan ini, analisis kepentingan ciri membantu mengenal pasti faktor-faktor utama yang mempengaruhi hasil tahun depan yang boleh berguna untuk pembuat keputusan, termasuk corak diet serantau, pendapatan isi rumah, dan aksesibiliti kemudahan penjagaan kesihatan, serta ciri-ciri alam sekitar. Selain itu, para penulis memberi tumpuan kepada kebolehlaksanaan rangka kerja yang dikenalpasti kepada kawasan lain dan populasi umur yang berbeza. Oleh kerana ia dapat menentukan hotspot kerentanan, kajian ini membantu mencegah/melaksanakan intervensi pemakanan yang optimum kepada populasi yang sensitif. Kerja ini menunjukkan keberkesanan pembelajaran mesin untuk analisis masalah kesihatan awam global yang diberikan dan berusaha untuk mengisi jurang antara pendekatan teknikal dan intervensi kesihatan yang nyata yang boleh membantu memerangi kekurangan nutrisi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

RCTs	-	Randomized Controlled Trials
NLP	-	Natural Language Processing
CNN	-	Convolution Neural Network
RNN	-	Recurrent Neural Network
EDA	-	Exploratory Data Analysis

NUTRITIONAL DEFICIENCY PREDICTION BY REGION USING MACHINE LEARNING

ANNE DASHINI KANNAN

UNIVERSITI TEKNOLOGI MALAYSIA

CHAPTER 1

INTRODUCTION

1.1 Introduction

Nutritional deficiencies are one of the major global issue in human health growth especially on specific reasons like low and middle income areas, where thousands of people, mainly kids and pregnant woman who suffer from insufficient intake of proper diet with full of vital micro nutrients.These nutrition deficiency are linked to various serious health issues such as stunted growth, health impairments, weak immune systems and high mortality rates. The usual traditional approaches to detect and identify nutritional deficiency, including surveys and interventions, tend to be expensive and inefficient.Consequently, there is a more pressure needed for more targeted, data-driven solutions to analyze regions and populations that are most harmful.This paper explains the application of machine learning (ML) techniques to predict and analyze nutritional deficiencies in different areas. By combining various data sets on health conditions, socioeconomic, environmental and dietary factors this paper goals to build model predictive models that can list out regions that are at high risk for deficiencies.Machine learning has the potential to define complex, non-linear patterns within large datasets,which been using often in traditional methods.These insights could help improve public health outcomes.

1.2 Problem Background

Nutritional deficiencies are spread widely over the global health issue that affect regions that are affected by various factors. Despite a significant progress in providing malnutrition over the years, millions of individuals mainly women and children continue to suffer from nutrition deficiency such as iron, vitamins, iodine, and zinc. These deficiencies are linked to various severe health outcomes, including stunted growth, impaired growth function, weakened immune systems, and high risk of infections and mortality. These complex issues need the use of advance algorithms and techniques to analyze the datasets and to aid the situation to prevent from getting worst. Machine learning provides ways to analyze data from various resources using health records, surveys, genetic studies. These were used to develop prediction models for nutritional deficiencies. By monitoring the prediction process, machine learning boost the efficiency, reduces the expenses and improve the access to the health care centre to prevent this nutrition deficiencies. However, the application of machine learning in this domain is still in early stages and there are more significant challenges that need to be sort out including data quality and ethical considerations.

1.3 Problem Statement

The current approaches to pin out the nutritional deficiencies are limited due to the reliance on sources with intensive processes and their inability to integrate data sources effectively. This causes the prevention and diagnosis of these deficiencies constrain in settings.There need to be and accurate ,calculable and interpret able prediction that can facilitate the targeted interventions and informed decision-making. Morever a clear platform and visualization is needed to display a clear view of prediction of nutritional deficiencies on specific regions for health providers ,NGOs, and stakeholders in order to help them to provide a better solution to overcome this deficiencies more effectively.

1.4 Research Question

1. How can machine learning techniques be utilized to predict the nutritional deficiencies effectively?

2. What are the data sets required to develop accurate predictive models on these nutritional deficiencies?

3. What are the potential challenges and limitations in developing the machine learning based prediction systems in real-world healthcare settings?

4. Who are the targeted groups focused in this prediction to develop an accurate prediction model?

1.5 Aim and Objectives

The aim of these project is to develop a machine learning based dashboard to display and predict the risk of nutritional deficiencies across different regions by analyzing datasets including health, diet, socioeconomic and the environmental factors.With the help of this prediction, this system enable the healthcare takers and organizations to develop targeted inventions to optimize resource allocation and to improve public health outcomes.

The objectives of this research are:

(a) To identify the key factors that contribute to nutritional deficiencies, such as dietary habits, incomes and environmental factors.

(b) To implement machine learning models to predict the risk of nutritional deficiencies in specific regions.

(c) To demonstrate an interactive platform to present the prediction and visualizations to the health providers and stakeholders.

1.6 Scope of Study

This project provides the prediction needed to solve the nutrition deficiencies under few circumstances. It prioritizes the prediction for certain regions only, especially focuses on rural areas where nutrient deficiencies may occur a lot due to limited access to food sources and healthcare services. This research also designs narrow predictions only to collect highly insights and it helps stakeholders to design more effective intervention on specific regions which allocates the solutions more efficiently. Its important to gather the main key features that impact the nutrition deficiencies. Here the critical nutrient minerals such as iron, zinc, and essential vitamins were listed depending on the availability of data sets. Current data and previous data from healthcare records , surveys related to food consumption, socioeconomic indicators are used to create the model trends and anticipate future deficiencies. Predictions based on specific physical needs and risk factors for each group are one of key points to calculate the prediction. However, to ease for a smoother predictions only targeted groups are focused suck children, pregnant women, low-income households and those who are at higher risk of nutrient deficiencies. This dashboard is developed to present an interactive report that can be communicate the data effectively to the healthcare providers and stakeholders.

1.7 Significance of Research

This research on nutritional deficiencies by region using machine learning is quite significance to the population as it allows people to understand the seriousness of this problem occurs in several areas. It also provides a data-driven approach to stakeholders and providers to understanding and combating malnutrition. Machine Learning can predict regional trends,and can identify future deficiencies by analyzing diverse datasets. It uncovers the hidden correlations between health factors, socioeconomic and environmental factors.This research also helps the targeted populations such as children, women, pregnant ladies and low income households to ensure they receive resources efficiently which allocated for them by providers based on the data predicted. Additionally with the use of machine learning models an interactive dashboards can be created which facilitates to decision-making and effective communication to the stakeholders.This project not only help to improvise this deficiency issue but also

contributes to global toward achieving sustainable development goals, particularly in reducing hunger in rural areas and improving people well-being.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter discusses into the challenges faced I previous research study related to this topic and provides as in-dept explanations details of their research findings. It writes and overview of the application of machine learning withing the health care department and areas, focussing its functionality and predictive insights. Key factors and datasets are important to developing prediction models are outlined to provide clear view on their development and creating prediction models. Furthermore, this chapter go in details on how innovative strategies to address the prediction of nutritional deficiencies.

2.2 Overview

Nutritional deficiency has become the global challenge especially on those who has low-income regions, remains a consistent issue affecting tons of lives mainly on women and children. The traditional method of collecting surveys was valuable but often left behind and short in efficiency causing unable to reach the stakeholders to overcome this issue as soon as possible (Ahmed & Kamalakkannan, 2022). Everyone focusing on children, women, elderly people are suggested to have a proper dietary plan with enough nutrition on it which needs in daily basis.

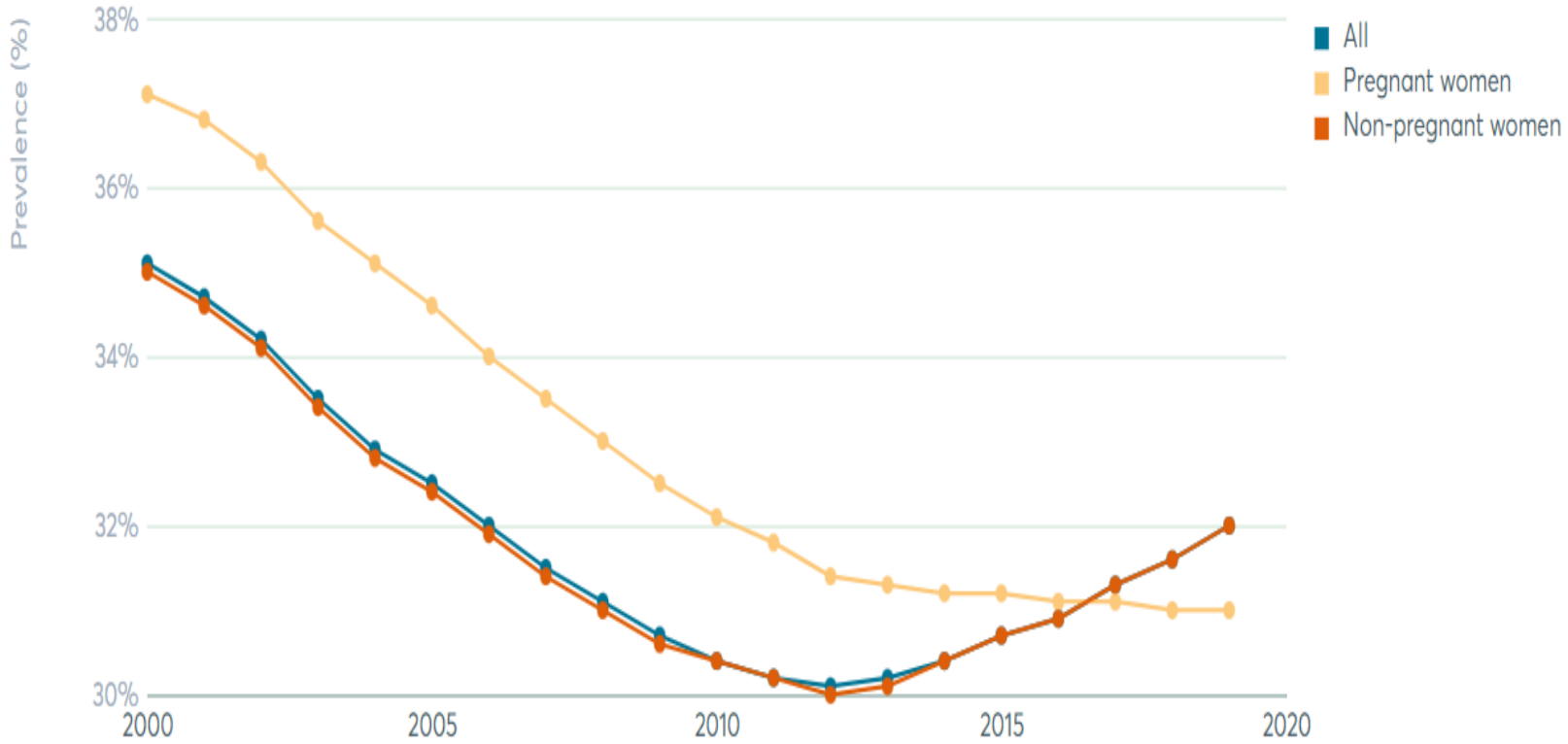


Figure 2.2.1 shows the decrease in value of nutrition intake over the year by targeted people according to the research by (Khudri et al., 2023)

This limitation has catalyzed this research into machine learning prediction as a potential step to find out the solution and analyze further on these mitigate nutrient deficiencies. Machine Learning provides a transformative approach to enhance and overcome this prediction analysis and intervention strategies in combating nutritional deficiencies. It is important to all age group to compulsory have their design nutrition pattern to avoid high risk in health issues. Due to some factors such as socioeconomics, income, environmental factors several places were constrained to have proper nutrition diet in their daily intake. The application of machine learning (ML) techniques to predict and analyse nutritional deficiencies in different areas.(Kananura, 2022) By combining various data sets on health conditions, socioeconomic, environmental and dietary factors this paper goals to build model predictive models that can list out regions that are at high risk for deficiencies. Machine learning has the potential to define complex, non-linear patterns within large datasets, which been using often in traditional methods(Zhou et al., 2019). These insights could help improve public health outcomes. Nutritional deficiencies have impacted the rural regions leading to serious health outcomes and changes in human developments. The challenge can predict the risk in nutritional deficiencies across various region by analysing diverse datasets that include health, dietary, socioeconomic and environmental factors. The aim is to develop a model with help of machine learning to identify the high-risk areas and the factors that has been contributing to the nutritional deficiencies.

Age Group	RNI 2005 (µg/Day)	RNI 2005 (IU/Day)	RNI 2017 (µg/Day)	RNI 2017 (IU/Day)	Upper Limit (µg/Day)
Infants					
0–5 months	5	200	10	400	25
6–11 months	5	200	10	400	37.5
Children					
1–3 years	5	200	15	600	
4–6 years	5	200	15	600	100
7–9 years	5	200	15	600	
Boys					
10–18 years	5	200	15	600	100
Girls					
10–18 years	5	200	15	600	100

Figure 2.2.2 above shows the findings of a proper nutrition intake that an infant should at least be taking to avoid health issues in early stages.(Jain et al., 2022)

2.3 Application of Machine Learning in Healthcare Industry

Machine Learning has always proven as a transformative element in the healthcare sector, offering robust frameworks to analyze large and complex data. Surveys, papers and studies have visualized the application of machine learning in determining disease outbreaks, medical treatments and optimizing healthcare delivery systems. (Ali et al., 2022) mentioned that few ML models such as supervised learning models, including decision trees and vector machines have been successfully used in predicting anemia in certain populations to showcase the feasibility of such technologies in nutrition focused studies. These methods have always been effective in handling structured data where labeled examples are available enabling accurate prediction of outcomes like malnutrition or any other specific deficiencies. Advanced machine learning techniques such as shown in table 1 below are particularly successful due to their robustness and ability to model complex interactions among different variables.

2.3.1 Supervised Learning Approaches

Studies have used the supervised leaning technique to predict the deficiency model for this nutritional deficiency in each region. Logistic regression and decision trees have been utilized to predict the anemia prevalence using the income list from each household together with their dietary plan. (Mpakairi et al., 2023) proposed random forest and support vector machines have shown robustness in dealing with complex, non-linear relationships between variables in nutritional deficiency predictions.

A notable study by Ling et al. (2020) applies gradient boosting models to predict the vitamin deficiency in achieving high accuracy by studying dietary intake data along with the region datasets taken from and with regional climate variables. The study highlighted the importance of feature selection and preprocessing in improving a prediction model for future references. Singh eth al. (2019) applied these clustering techniques to demographic and health surveys (DHS) data, showing the regional issues in malnutrition across the Sub-Saharan Africa.(Khudri et al., 2023)

2.3.2 Deep Learning Techniques

These model types, especially (CNNs) convolutional neural networks and recurrent neural networks (RNNs), have been explored for malnutrition predictions. This model excels in preprocessing large scale of date sets and complex inputs including imagery and time-series data sets. These been used to predict the productivity of the region according to their socioeconomic directly linking to the regional dietary deficiencies.

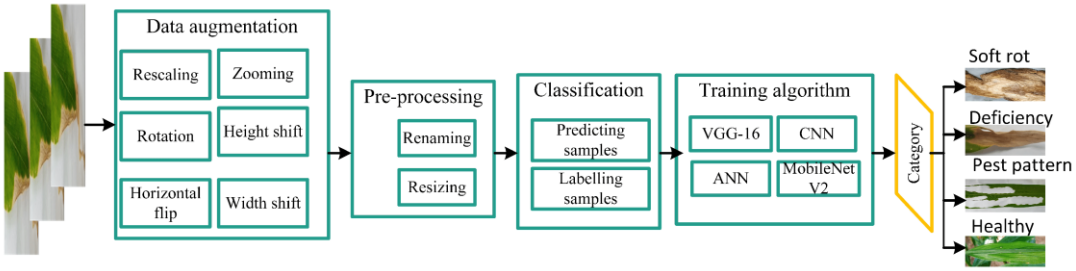


Figure 2.3.2.1 shows the proposed method of deep learning techniques used by (Velásquez et al., 2020) to identify the nutritional deficiencies among plantations.

2.3.3 Unsupervised Learning and Clustering

Unsupervised Learning methods such K-means clustering and hierarchical clustering, have been used to group regions with similar profiles. This approach helps in identifying the high-risk areas and underlying patterns in nutritional deficiency. Moreover, using unsupervised algorithm helps to demonstrate the pattern of the deficiency issue on the region. These patterns enhance the monitoring for the deficiency trend.

Supervised Learning	Unsupervised Learning	Semi-Supervised Learning	Reinforcement Learning
Models are built according to the definitions and findings	Deep Learning is used to arrive at the conclusions and patterns through unlabeled data.	Build model through a mix of labeled and unlabeled data.	Self-interpreting is based on a system learned through trial and error seeking findings.
• Linear Regressions • Support Vector Machines • Decision Trees	• Apriori • K-means clustering • Artificial Neural networks	• Generative networks • Naïve bayes Classifier	• Q-learning • Model based estimation
These algorithms are used to demonstrate risk assessment, predictive analysis, and image classification.	These algorithms are suitable for describing the functions, perform monitoring, data mining and pattern recognition.	These algorithms describe more about manipulation, data visualization and natural language processing.	These algorithms are more self-paced prediction modelling which need linear tasks and estimating parameters.

Table 1 shows the differences between various machine learning algorithms.(Gollapalli, 2022)

2.4 Datasets and Key Factors in Prediction Models

The quality and diversity of a datasets is important and critical in developing a prediction machine learning model. A real time data sets gives more accurate value to the scenarios as it will be taken from real data source. (Meshram et al., 2021) says the traditional way of prediction model has been taken with the use of comprehensive surveys and feedback that provide demographic, health, and nutrition data. Datasets also collected on food supply and consumption organizations to know the amount of food been consumed and distributer to each region. Datasets plays crucial role in finding to justify the findings with proof and evidence.

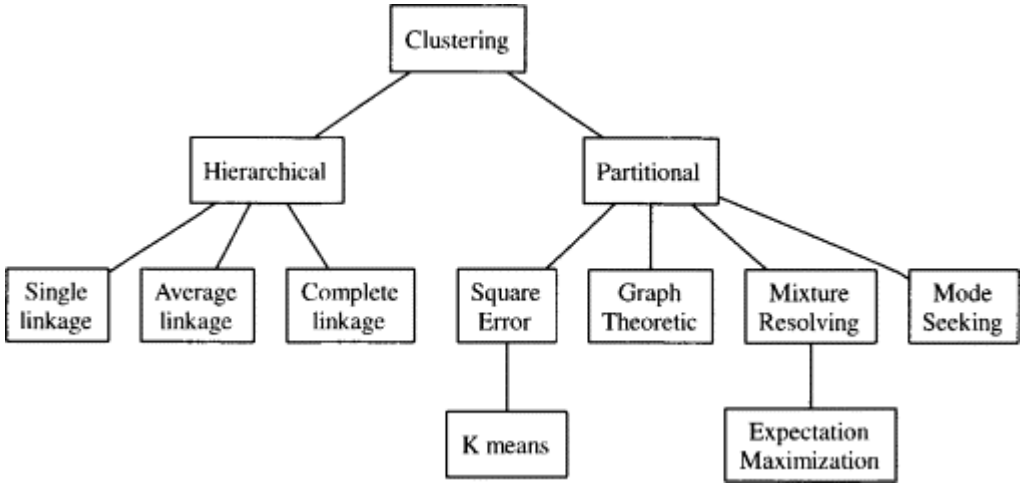


Figure 2.4.1 explains how clustering method works in mode prediction using machine learning algorithms.

2.5 Innovation in Nutritional Deficiency Prediction

Developing standardized and comprehensive datasets for nutritional deficiency prediction helps to monitor more in details which could leave the stakeholders to a proper decision making. Integrating explainable data sets along with machine learning techniques will enhance the model interpretability. Incorporating real time data streams improve the accuracy in the findings which can justify the prediction model. Exploring hybrid models that combines both the statical and machine learning techniques approaches will lead to a better generalization. In this paper the provided data driven predictions of nutritional deficiencies across regions enables the healthcare providers and stakeholders to identify the risk in the populations efficiently. It's also able to highlight the main key factors that been affecting to the nutrition deficiencies and help to reduce it. This project contributes by providing evidence based on live data to guide and develop for long term strategies to combat malnutrition. All these contributions are aim to create lasting impact and provide practical solutions for improving health.

References	Experiment	Strength	Limitations
(Ahmed & Kamalakkannan, 2022)	Micronutrient classification in IoT based agriculture using machine learning (ML) Algorithm	• Machine learning algorithms are used to predict the pattern for the micronutrient	• Prediction pattern only shows the future effects.
(Ali et al., 2022)	Machine Learning Approaches for Prediction of Nutrition Deficiency among Women of Different Age Groups	• Able to find the root cause of deficiency mostly among women with prediction models	• Unable to visualize the findings to the providers
(Gollapalli, 2022)	Ensemble machine learning model to predict the waterborne syndrome	• Elaborates the factors and demonstrate the solutions with a prediction model	• No accurate data shown to justify the prediction model.
(Jain et al., 2022)	Efficient Machine Learning for Malnutrition	• Stakeholders manage to visualize the future	• Visualizations are well delivered.

	Prediction among under-five children in India	effects from the predictions model created	
(Kananura, 2022)	Machine learning predictive modelling for identification of predictors of acute respiratory infection and diarrhoea in Uganda's rural and urban settings	• Algorithms are used to identify the accurate cause of infections	• No real time data are shown to have a clear view of the accuracy
(Khudri et al., 2023)	Predicting nutritional status for women of childbearing age from their economic, health, and demographic features: A supervised machine learning approach	• Random forests are used to gather the data from various variable and telecast the outcome in one prediction model	• Noisy data should have been cleaned.

(Qasrawi et al., 2024)	Machine Learning Approach for Predicting the Impact of Food Insecurity on Nutrient Consumption and Malnutrition in Children Aged 6 Months to 5 Years	• Demonstrate risk assessment and perform monitoring form the prediction theory	• Risk assessment are predicted without solutions to overcome it.
(Velásquez et al., 2020)	A method for detecting coffee leaf rust through wireless sensor networks, remote sensing, and deep learning	• Integrates the data finely with the algorithms and transfer into machine learning concepts to visualize it	• Data are finely integrated and but not real time data are used to prove the findings
(Zhou et al., 2019)	Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize	• Provides image classification • Predictive analysis demonstrates the clear view of classification.	• Still exist irrelevant image with not accurate to the findings

	yield in China using machine learning approaches		
(Zhang et al., 2019)	Application of deep learning in food	• Deep learning techniques are used to filter out the main cause of food deficiency	• Fewer number of clusters

Table 2 shows the summary of findings from various resources and research paper s used in this literature review

2.6 Research Gap

Despite significant advancements, several research gaps are to be remain in the application of machine learning for nutritional deficiency prediction. The region data will be limited according to several factors. Including many regions even though will enlarge the findings there will be lack in producing high quality predictions especially involving low in come Ares(Zhang et al., 2019). Limiting model development and validation can enhance the precision in the findings(Qasrawi et al., 2024). This research will be focussing on few deficiencies only limiting to the main five important nutrition deficiencies to helps cover the overall nutrition deficiencies. Most of the studies concentrate on one type of deficiency, while comprehensive models addressing multiple deficiencies are scarce. While multimodal approaches exist it is important to verify that diverse datasets are remain underexplored to interpret and study while encountering this research. Most of the datasets from different regions lead to fail vales due to differences in their culture and economic state or dietary plans. Addressing these gaps earlier before carrying out the research will require collaborative efforts between researchers, policymakers and data providers to enhance the utility of machine learning based solution.

2.7 **Summary**

Machine learning holds crucial role for prediction nutritional deficiencies by region, offering a data driven approach to addressing global health disparities. By interpreting diverse data sets and advanced algorithms, machine learning leads the understanding on how important the nutrition is related to our human organizations and helps to support the target people by region. However, addressing challenges related to data quality, interpretability, and ethical concerns is significant for a complete implementation of these technologies in health policies. This chapter has included the literature review of this ongoing research regarding the prediction model based on nutrition deficiency with the findings of previous work. Apart from it, it also explains the machine learning algorithm will be used throughout this research.

CHAPTER 3

RESEARCH METHODOLOGY

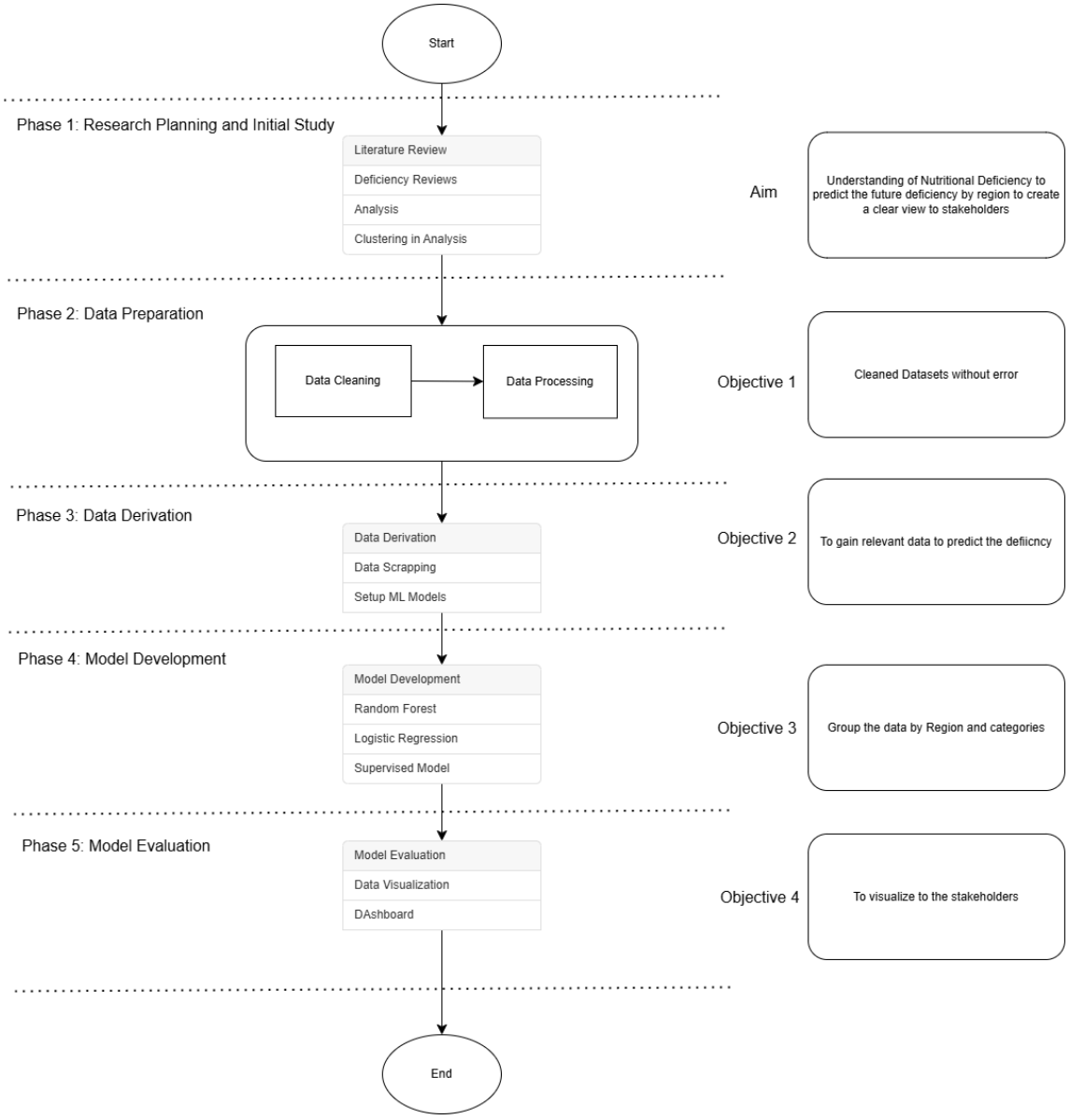
3.1 Introduction

This chapter outlines the methodology used to predict nutrition deficiencies across regions using machine learning techniques. The research workflow, from the initial topic exploration to model evaluation, is described in detail. The dataset and performance evaluation criteria employed are also elaborated.

3.2 Research Framework

The structure of research is going to strive to ensure all the challenges faced when predicting nutrition issues across different areas. It is divided into five key phases: The main steps of the proposed framework include research planning and initial study, data preparation, feature extraction, model development, and model evaluation. Every stage in this process supports the development of an overall understanding of nutrition profiles and main drivers in a region. During formal system analysis phase, the research problem was comprehensively defined in addition to setting achievable objectives. This was then followed by a data description where real data sets were cleaned and preprocessed. During the feature extraction phase, the extracted information in the datasets was analyzed by employing the Machine Learning models to establish relevant patterns. The fourth phase aimed at using clustering methods to cluster regions according to their distinct nutritional aspect. Lastly, in phase five, the authors assessed the validity and accuracy of the developed clustering model through method including ML models.

Figure 3.2.1 explains the research framework.



3.2.1 Phase 1: Research Planning and Initial Study

The first step of the study was primarily exploratory concerned with identifying the issue of malnutrition and defining research questions. This included a synthesis of the literature on past efforts on malnutrition and lack of nutrients. Literature reviews by Black et al. (2013) and other studies by UNICEF and WHO contributed their invaluable findings regarding the frequency, underlying causes, and consequences of malnutrition. These studies revealed that regional differences regarding diets, standards of living and accessibility to health facilities influenced malnutrition. The literature on methods of assessment also revealed a number of limitations of the conventional approaches like national health surveys and RCT trials where the scientific sample Investigation also exposed a number of weaknesses and shortcomings of the prevalent methods of assessment; for example, national health surveys and RCT trials do not capture the full array of nutrition problems at micro level.

To address these challenges, machine learning was adopted as a cost-effective method for analysing large and large datasets and locating regional characteristics. This phase also defined the purpose of the study, to identify the characteristics associated with poor nutrition and to use machine learning to discover relevant patterns to support public health interventions.

3.2.2 Phase 2: Data Preparation

Before analysis, data has to go through data preparation process to make sure that the collected dataset is accurate and can be used for analysis. The study used different data sources ranging from national health surveys, global nutritional databases, demographic records of regions among others. These datasets offered finer details on aspects including dietary preferences, health status, environmental aspects, and population characteristics.

It must be mentioned that before the analysis the data was cleaned to handle missing values and duplicates as well as inconsistencies. Data that were skipped out were either removed or removed whenever the data was not relevant in assessing the result. The dataset was cleaned up to ensure that the availability of observations was across a different set of variables. Duplicates were deleted to ensure data validity. Secondly, applying text preprocessing techniques, any self-reported data, for example, survey responses, were used. Text was normalized by first converting it to lowercase and secondly all irrelevant terms have been stripped off the text data. Where needed, supplementary relevant data sources included measures of agricultural yields and climate characterizations were added to the dataset to increase the efficacy of the machine learning algorithms.

3.2.3 Phase 3: Feature Extraction

In the process of data analysis, an important step entails selection of relevant, distinguishing features from the dataset. Using regression analyses, the study found that the likelihood of nutritional deficiencies was positively predicted by the dietary diversity scores, perceived prevalence of infectious diseases, mother's education and accessibility of fortified foods. The selection of these variables follows previous studies which confirmed their role in influencing nutritional health.

Textual data mining and, specifically, Large Language Models such as OpenAI's GPT were used to investigate prominent patterns from the collected textual data. For example, the models found words and phrases linked to diet and nutrition issues, health

issues, and geographical aspects of life. Feature engineering was also used to convert raw data into derived features including normalized dietary diversity score and Health Risk Indices. These enriched features upset the clustering algorithm to differentiate the regions efficiently cay the nutritional categories they represent.

3.2.4 Phase 4: Model Development

To propose a model for predicting and analysing the states of nutritional deficiencies together, the model consisted of supervised and unsupervised machine learning. Logistic regression was also applied in supervised models to predict region specific probability of the presence of nutritional deficiencies. High accuracy of results and interpretability make these models appropriate and efficient for sets of ordered data. These models were trained under labelled data where scarcity was the elements or variables of concern.

When designing the clusters to partition the regions according to nutritional similarity, two algorithms were used: K-Means and Hierarchal Clustering. The following models partitioned geographical area into different categories according to a common nutritional characteristic, including dietary scarcity, prevalence of deficiency anaemia, nutritional status, and socioeconomic status. K-Means was preferred for its ability to deal with big data while Hierarchical Clustering offered a layering structure of the regions.

There were several stages during the model development. First, hyperparameters of the predictive models were selected by using grid search for their optimal parameters. For instance, features such as the number of trees and the model's maximum depth were optimized for Random Forest, as was the learning rate and the number boosting rounds. The same applies to K-Means where the Number of clusters was decided and to Hierarchical Clustering where clusters were searched.

The numerical results of the feature selection were used to determine crucial predictors associated with nutritional deficiencies and guide policymaking. For instance, dietary diversity scores, the household's access to fortified foods, and maternal education were depicted as significant predictors in the studies. The

clustering results were depicted by geographic heat maps and 3D scatter plots for easy interpretation of the regionality.

3.2.5 Phase 5: Model Evaluation

The conclusion of the assessment of the machine learning models incorporated supervised and unsupervised measures for the validity and accuracy of the models in anticipating nutrient depletion and determining geographical zones. To assess the predictive performance of the supervised models, accuracy, precision, recall and F1-score were used. They offered an understanding of how accurately our models were able to assign regions depending on their risk of possible deficiencies.

For example, countries with high prevalence of vitamin A deficiency and low dietary diversity matrix were clustered together, indicating the potential areas suitable for food fortification programs. Data obtained from the clustering were further illustrated using heatmap and geographic maps, which can be easily understood by policy makers and other stakeholders. To strengthen the generalizability of the developed predictive models, the process of cross-validation was used. The dataset was divided into training and testing sets and performance was tested on new data not used for training. For clustering, validation with the regional health reports was conducted to fit the identified clusters with WHO's Global Nutrition Reports for effectiveness of the malnutrition patterns known.

Refinement of the questions took place continuously during the evaluation stage. Main characteristics were included or excluded according to the importance values, and parameters were tuned for better results. This iterative was definitely a strength since it helped in the creation of final models that would be more accurate and most importantly usable for giving out recommendations concerning the nutritional needs of the various regions.

3.3 Dataset

The dataset contained more than 185 000 rows and eight main predictors: geographic variables, demographic data, nutritional status, health status, and environmental information. The overall data set of these subsequent products offered a strong baseline for assessing malnutrition in relationships to regions. This approach allowed the study to get an appreciation of the various factors that precipitate malnutrition.

3.4 Performance Measurement

Evaluation of accuracy, reliability and potential usability of the machine learning models created for this study required performance measurement. In the case of the supervised predictive models, the evaluation metrics concerned the classification accuracy, while in the case of the clustering models the quality of the formed clusters was of concern.

To deem, the supervised models of classification such as Random Forest were evaluated using basic classification performance measures. Accuracy defined a ratio of the number of instances which were correctly predicted out of the total number of cases. However, since nutritional deficiency data may have class imbalance problem, using precision, recall and the F1 score offered more valuable information. Among all the positive predictions made by the model, precision Stayed high, indicating that the model did not make many false positives. Recall focused on how many of all actual positives were true positives, showing that the model captured all areas where a risk of nutritional deficiencies existed. Since the F-measure is defined as the harmonic mean between the precision and the recall, their weaknesses were compensated which provided an objective assessment of the system's performance.

3.5 Summary

This chapter provided a detailed analysis of how the nutritional deficiencies at the regional level can be estimated by using machine learning. To promote the reliability and applicability of the obtained results, the framework of the research included data pre-processing, feature extraction, clustering model construction, and assessment. The use of multiple data sources and appropriate analytical methods allowed the researchers to develop practical recommendations concerning deficiency and possible actions in the field of nutrition.

CHAPTER 4

INITIAL FINDINGS

4.1 Introduction

This chapter explores the dataset collected and provide a visualization on analysing the nutritional deficiencies globally. Exploratory Data Analysis (EDA) is used to discover the patterns in the dataset collected and identifies the meaningful insights. Various techniques and methods are used to visualize the datasets such as the summary statistics, preparations of the datasets, analysis and comparison.

4.2 Exploratory Data Analysis (EDA)

Exploratory Data Analysis or EDA is a fundamental stage in data analysis in ensuring that one will receive valuable insights from data. It is used to find occasional patterns, aberrations, and research assumptions and hypothesis making by applying basic statistical techniques and graphical representations. EDA requires a number of steps in order to gain a solid understanding of the data. The initial step involves problem formulation together with analysis of data that are at the disposal of the researcher. The next step is to make structures and determine in which extent there are missing values or in-equalities in the imported data. Any data gaps should be handled rightly as either imputing the missing value or excluding this record, in order not to create an imbalance in the results.

Subsequently the distribution, mean, standard deviation of the data is analysed to look for patterns and outliners. Sometimes the raw data is converted by scaling, encoding, or combining other data sets to facilitate analysis of data sets. Graphs, charts, and any other graphical methods are used to amplify gist and trends established in the data. It is also important to handle with outliers to make the analysis results more reliable.

Finally, the conclusions of the findings are briefly described in informative visuals and highlighted summaries that also point out the next steps for the research.

4.3 Steps of Exploratory Data Analysis

4.3.1 Understand the problem and the data

The variables in the dataset include country, age group, intake levels of nutrient, diseases of malnutrition and the population that suffers from it. This dataset has the purpose of studying the causes of malnutrition around the world, differentiated by regions, age, and socioeconomic differences between rural and urban areas. For this problem of nutritional health, each variable in the dataset goes a long way in helping to create dimensions of the problem.

The sample size comprises 187,534 records and 8 variables. Figure 4.3.1.1 below shows the columns include `Country` which denotes some of the affected countries; `Age_Group` which splits people into some number of standard age groups; `Nutrient_Intake` which represents the level of nutrient intake; `Malnutrition_Disease` which denotes some diseases associated with malnutrition. Other subfields like `Region_Type` subdivides data by rural/urban, while `Household_Income`, `Affected_Children`, and `Affected_Women` represent monetary health and impinged people. This structure makes it possible to carry out a multidimensional assessment of the causes of nutritional deficiencies.

	Country	Age_Group	Nutrient_Intake	Malnutrition_Disease	Region_Type	Household_Income	Affected_Children	Affected_Women
0	Nigeria	51+	High	Vitamin D Deficiency	Rural	922.957740	376	18
1	Peru	31-50	Moderate	Iron-Deficiency Anemia	Rural	2054.202346	441	271
2	Malaysia	0-5	Moderate	Scurvy	Rural	1052.049571	217	120
3	Mozambique	51+	Low	Scurvy	City	1168.316693	201	254
4	Kenya	31-50	Moderate	Iron-Deficiency Anemia	Rural	1194.207944	362	240
5	Ethiopia	0-5	Low	Scurvy	Rural	1456.667771	432	71
6	Malaysia	51+	Moderate	Marasmus	City	1327.633271	197	41
7	Indonesia	0-5	Low	Vitamin D Deficiency	Rural	1177.309544	288	31
8	Nigeria	0-5	Moderate	Vitamin D Deficiency	Rural	4931.089823	166	57
9	Vietnam	51+	Moderate	Kwashiorkor	City	1039.534410	200	243

Figure 4.3.1.1: Nutritional Deficiency Datasets

4.3.2 Import and Inspect Data

The first process of EDA incorporated included importing the dataset and verifying the range of data types for the dataset. These include `Household_Income`, `Affected_Children`, and `Affected_Women` which were normalized for precision and then compressed for memory. In the same way, category variables such as Country and `Malnutrition_Disease` were checked in terms of category consensus to exclude categories that are not necessary or entries that are wrongly classified. Here everything was done during the cleaning step, so that they do not become a problem when analysing it.

```
Country                 object
Age_Group               object
Nutrient_Intake         object
Malnutrition_Disease    object
Region_Type             object
Household_Income       float64
Affected_Children        int64
Affected_Women           int64
dtype: object
      Country Age_Group Nutrient_Intake    Malnutrition_Disease Region_Type
\
0     Nigeria       51+            High    Vitamin D Deficiency       Rural
1        Peru     31-50        Moderate  Iron-Deficiency Anemia       Rural
2    Malaysia       0-5        Moderate                  Scurvy       Rural
3  Mozambique       51+             Low                  Scurvy        City
4       Kenya     31-50        Moderate  Iron-Deficiency Anemia       Rural

   Household_Income  Affected_Children  Affected_Women
0        922.957740                376              18
1       2054.202346                441             271
2       1052.049571                217             120
3       1168.316693                201             254
4       1194.207944                362             240
```

Figure 4.3.2.1

For instance, the value for the numeric columns were scaled to give correct calculations and comparison of the variables. The data was cleaned in this stage by checking for cases such as missing or erroneously entered values and recoding of the variable types appropriately as shown in figure 4.3.2.1 above. It was carried out to provide a high-quality data set free of mistakes that would have been formed and could be used in later steps of data analysis and data visualization.

4.3.3 Handle Missing Data

This was a crucial step in as far as data cleaning was concerned because of proper management of missing data. Among the variables, gaps in `Nutrient_Intake` and `Malnutrition_Disease` were detected. To fill these gaps, mode imputation was used for the categorical variables by replacing missing values by the most popular values in every column. By following this approach, equal distribution of data and bias was achieved in the best manner.

When some rows had many missing values, they were deleted from the process entirely. For example, if entire row was blank in major important fields, then exclusion made sure that results were not influenced from any incomplete values. These steps helped me to make the dataset ready with proper structure and cleaning and then allowed to focus more investigation over the findings.

4.3.4 Explore Data Characteristics

Since the aim of the study was to explore the data and its features, the descriptive statistics for the numeric variables was computed. These were mean, median and standard deviation for `Household_Income`, `Affected_Children`, `Affected_Women` and other similar variables. Box plots were created and examined for issues regarding outliers when assessing descriptive data of categorical variables.

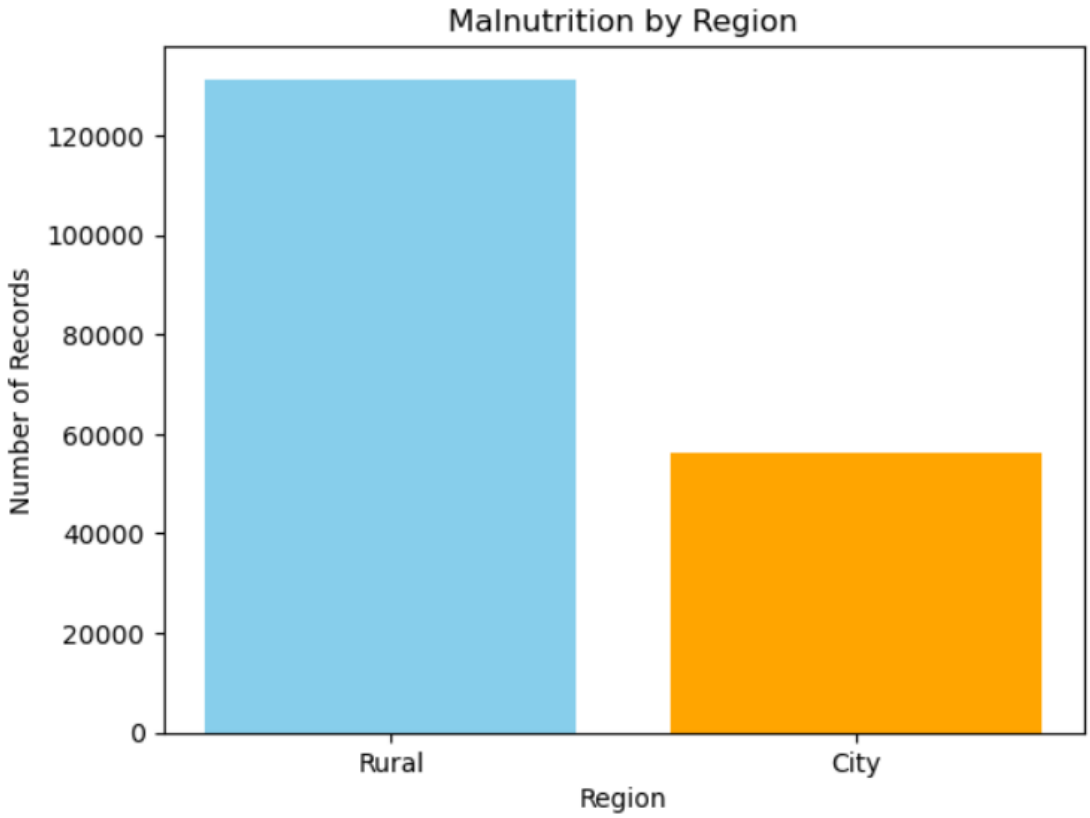
For instance, the studied rural community residents' average income was estimated to be $ 1,200 while that of the urban counterparts was estimated to be $ 3,000. Likewise, each record involved an average of 250 children and women affected by violence. Theoretical distribution of `Nutrient_Intake` CI showed that 10% of entries fit into the High category while 40% and 50% fell into the Low and Moderate category respectively. These findings point at the social problem of the population with low nutrient intake.

4.3.5 Perform Data Transform

Despite the data being mainly structured, there was normalization made on attributes like `Household_Income` so that all values could be processed under the same standard. To perform this, several changes were made through data normalization so that changes in the income level of the two areas would be captured. No more modifications were needed anymore, because the dataset had been made clean in this stage.

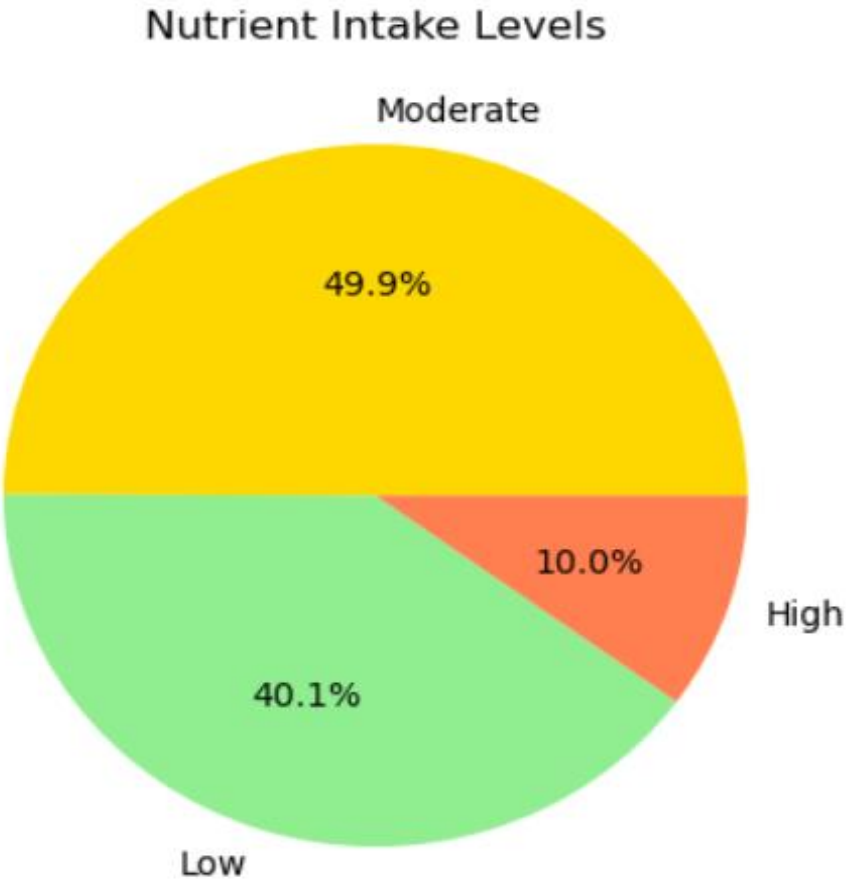
4.3.6 Visualize Data Relationships

Analysis of the data was greatly facilitated with the help of a data visualization technique. Here are leads on some important visualisations along with the clarifications on what they depict on the larger plane.



Bar Chart 4.3.6.1: Malnutrition by Region in the form of a Bar Chart

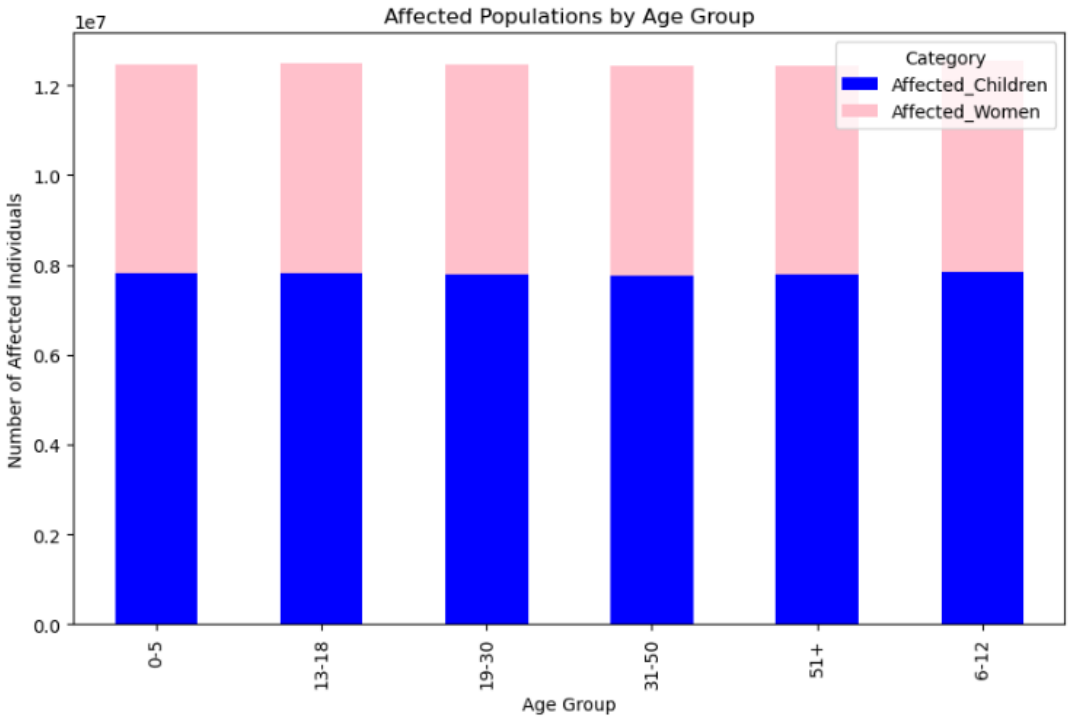
Bar Chart 4.3.6.1 present the level of malnutrition that was observed, a bar Chart was developed with rural and urban zones analysed. This made the visualization show that rural areas had higher absorption rates of malnutrition than the urban areas. It emphasizes the necessity for focusing on populations inhabiting rural areas because they still can have few opportunities to receive necessary resources, including healthcare.



Pie Chart 4.3.6.2: Dietary Energy Distribution

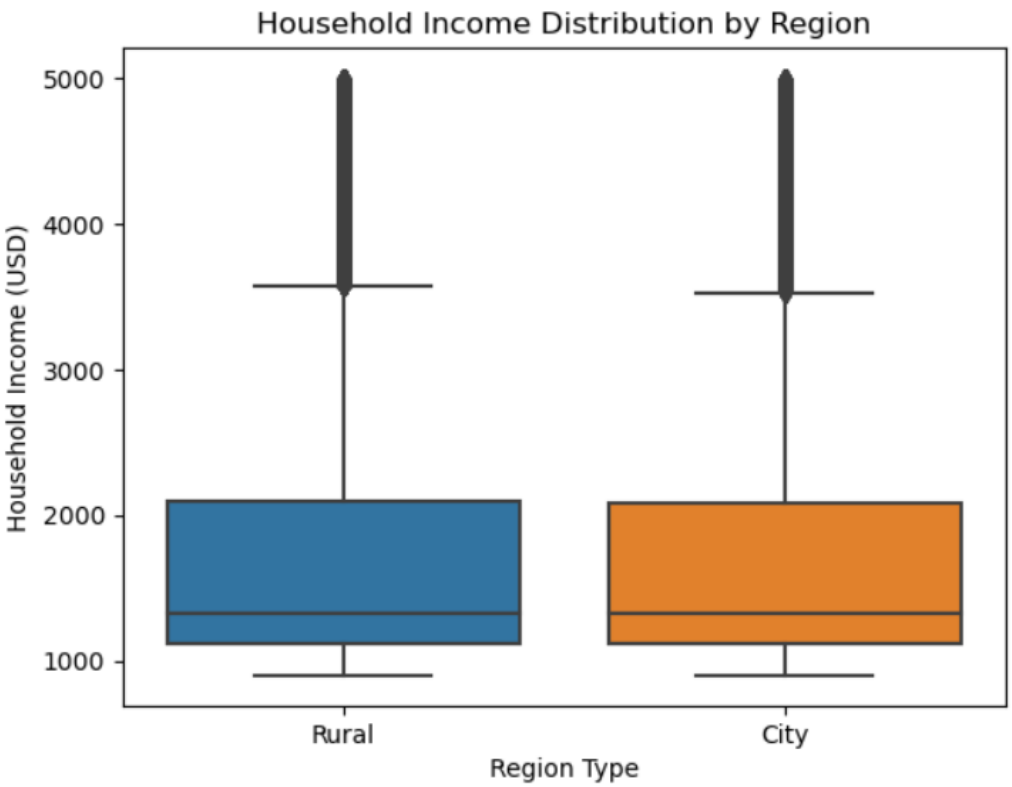
The above pie chart 4.3.6.2 is presenting the portion of the total the Nutrient Intake Distribution Chart was used in the form of pie chart. The distribution of the nutrient intakes was also determined, showing a high percentage of Low nutrient values obtained from the chart, which was 40%, Moderate which was 50% and High nutrient

intake came out 10%. This visualization clearly shows how people still do not consume the recommended daily nutrients the report indicates majority of the population.



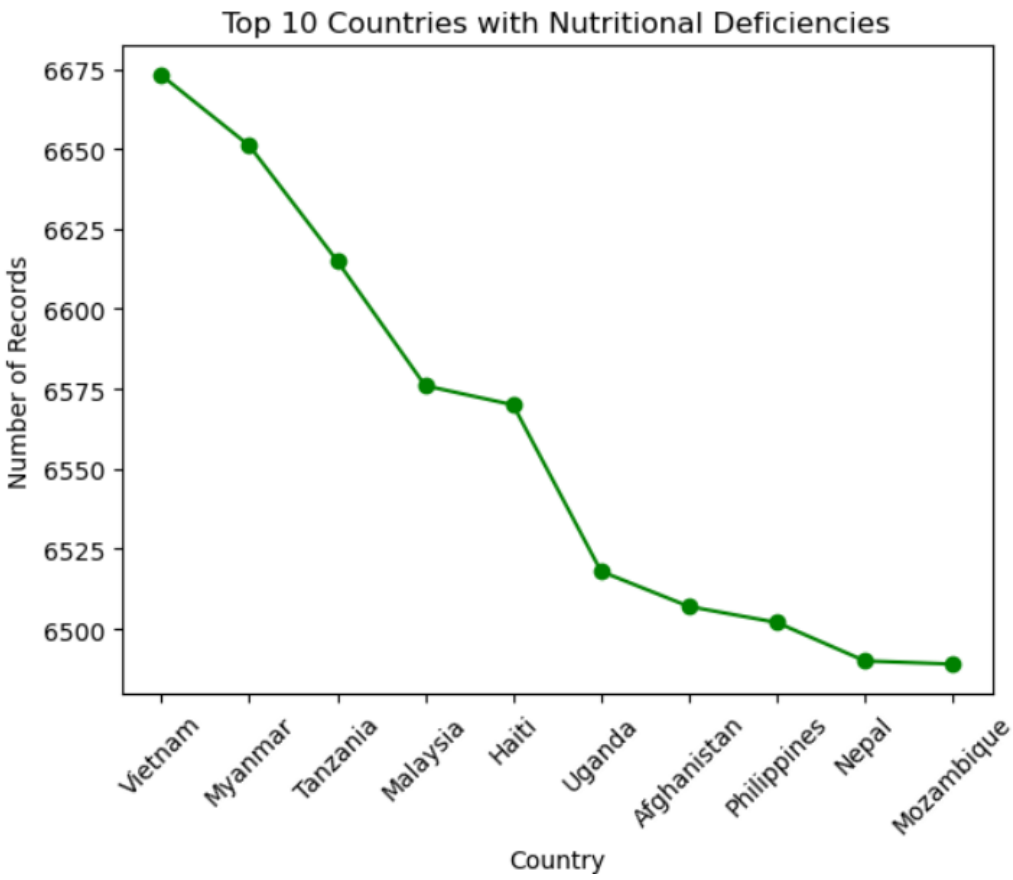
The Stacked Bar Chart 4.3.6.3 of samples on affect populations by age group

Bar chart 4.3.6.3 is present to comparatively show the number of affected children and women categorized by their age, a stacked bar chart was produced. From the chart, the "0-5" age bracket, as the most vulnerable population had the highest number of affected children, pointing to the need to feed these children. It also revealed increased cases of affected women with special focus to the reproductive age groups.



The Box Plot 4.3.6.4 of Household Income Distribution

To represent the distribution of income in households in rural and urban areas, a box plot was used as shown above in figure 4.3.6.4. More so, this plot showed that the rural households had lower and more dispersed medians than the urban households who had superior and more aggregated median incomes. This map raises questions on the effect of spatial inequities between rural and urban areas on nutrition status.



Over the countries – Line chart 4.3.6.5 representing the trends of Nutritional Deficiency

Among different chart types, line chart 4.3.6.5 was selected to determine of the ten countries that had the highest number of cases of nutritional deficiencies. Chart A also showed that the two counties that contributed higher values were India and Nigeria, due to the high number of people suffering from the problem and poor efforts towards eliminating malnutrition. It further supports need for region-based policies and programs.

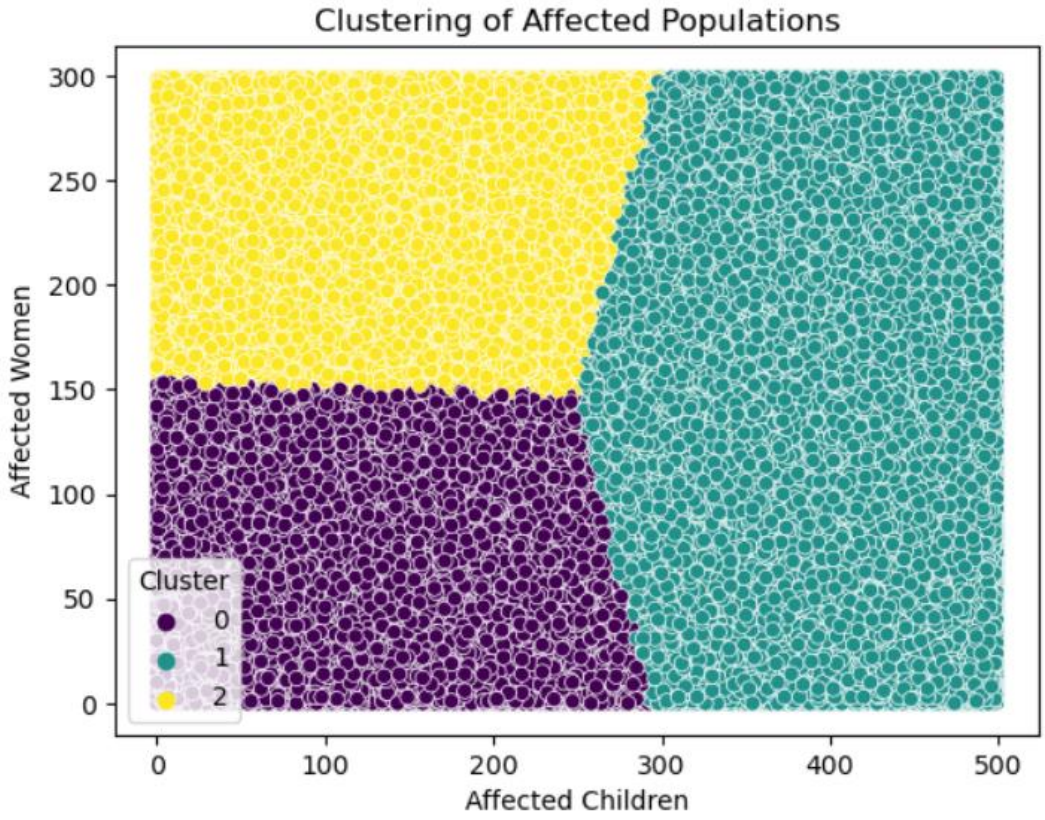


Chart 4.3.6.6: Among the intervention models, one can identify the Clustering of Affected Populations (Clustering).

Logically, clustering techniques that were followed categorized the entries depending on the number of affected children and women. An X and a y-axis scatter presented these clusters and showed that each of them was quite like the others, constituting three groups in total. These clustering patterns as shown above in chart 4.3.6.6 assist in understanding patterns and place priority for reactionary measures to regions or populations.

4.3.7 Handling Outliers

For categorical variables such as `Household_Income`, `Affected_Children`, and `Affected_Women` outliers were detected using the Interquartile Range (IQR). Some values were truncated to opposite extremes, to bring the data into line. For instance, household's required income was limited to a lower limit of $500 and upper limit of $5,000. This method made it possible to exclude the impact of outlying observations while keeping the overall shape of the data intact.

4.3.8 Feature Engineering

Feature engineering involved feature extraction which entails generating new features and transforming existing ones for enhanced capability to be useful for further analysis and modelling. The following enhancements were made:

1. Income Grouping: The `Household_Income` feature was further split into bins: "Low income" for a value < $1000; "Middle income" for the value of $1000 to $3000; "High income" for the value >$3000. This categorization makes it easier to compare one socioeconomic class to the other.

2. Malnutrition Risk Score: A composite score had then arrived by benchmarking `Nutrient_Intake` with weight 'W1', `Affected_Children` with weight 'W2' and `Affected_Women` with weight 'W3' as 0·575, 0·263 and 0·162 respectively. This score raises the precariousness of malnutrition for each entry and permits a detailed examination of the most vulnerable groups.

3. Age Group Indicator: Specific for each age range of patients, binary indicators were included for easier filtering based on age: `Is_Child` and `Is_Adult`.

4. Region and Disease Interaction: The data interactive feature was developed to trace the relation between `Region_Type` and `Malnutrition_Disease`, with a focus on how disease incidence differs between rural and urban communities.
5. Normalized Population Impact: The ratio of `Affected_Children` and `Affected_Women` was subsequently divided by the total population of each country so that it can be compared with other nations as they have different population bases.

4.3.9 Communicate Findings and Insights

This assessment led to the identification of several important findings in the analysis. Lifestyle diseases/ risky factors: Again, nutrient deficiencies that were considered were more prevalent among rural inhabitants than the urban dwellers. The "0-5" age group was identified as being the most vulnerable; evidence that underscores the need to target this age group more. Further, this was followed by poor results regarding nutrition where both the prevalence and incidence of malnutrition were found to be higher among the low-income population indicating SES as a major determinant of nutritional status. Thus, these results give a clear picture of the entire dataset and can be used as the basis for additional research and practical activities.

4.4 Conclusion

This chapter has also endeavoured to present a detailed exploratory analysis of the nutritional deficiency dataset. As a result of using descriptive statistics, data visualization, and missing values and outliers' treatment the general and specific patterns and associations have been identified. The results presented herein underscore important directions for practice including rural-urban disparities, young children, and income related issues. It is an insight that opens up possibilities for enhancing easier strategies in combating malnutrition on an international level.

CHAPTER 5

DISCUSSION AND FUTURE WORKS

5.1 Introduction

This project deals with the prediction of likely deficiencies by region and with the application of pattern recognition to gain greater insight into regional dietary and health patterns. Many processes were done to achieve high reliability and accuracy level of the predictors such as data preprocessing and exploratory data analysis (EDA). Cleaning the data meant that the data was free from any inconsistencies that might have secluded it from further analysis. A machine learning model is then utilized for the nutritional analysis of the data obtained and for making nutritional deficiency predictions for different regions. These efforts involved the utilization of the feature of machine learning capacity to identify intricate patterns as a means by which considerable information on the status of regional nutritional health was attained, devoid of noise and features that added no value. Division of the regions into groups according to identified need predictions was also helpful to visualize the need to improve nutrition and to develop intervention strategies.

5.2 Achievements

The success of this project includes the accomplishment of cleaning and preparation on the dataset since removing the duplicates and the irrelevant information and dealing with the missing value issue to attain data completeness. Deficiency factors were determined based on nutritional and demographic information, and data was standardized so that Southern and Northern regions could be compared. A classification model of supervised machine learning was adopted for the prediction of nutritional deficiencies, meanwhile the regions were clustered according to the similar pattern of the deficiency in order to facilitate the visualization of the prediction. The generated predictions were then assessed with other measures such as accuracy, precision, and

recall and displayed in an intuitive table for the stakeholder to see which regions require high intervention.

5.3 Future Work

The current research study creates a platform for viewing regional nutritional deficiencies while launching several options for more discussion. Extensions of the current study may involve the use of a bigger sample size and diverse data set that will capture more regional and or seasonal adjustments while feeding these data sets with real time information cross sectional survey of health status of the population, supply side data on availability of foods of concern. The form of feature engineering that might come in handy is the addition of environmental, economic, and cultural variables that may have a bearing on malnutrition, and or using tools such as NLP to create a lot of health data from the unstructured data form. Increasing model performance could be realized through the optimization of other advanced machine learning methods including ensemble or deep learning to increase the prediction capabilities of the model and use the other AI methods to explain why the model is making these predictions. Methods could be improved by employing algorithms, such as silhouette coefficients, to validate clusters and compare them with the ground truth; In addition, further segmentation of clusters to look for sub-regional specifics and to pinpoint populations of interest could be made. Further, policy and intervention design may entail consultation with policy makers or health organizations to develop specific nutrition programs using predictive results of interventions addressing and following up on the effectiveness of implemented interventions based on re-surveys and evaluation.Through these areas, the future study can support the improvement of nutritional health equity and reduce dietary lacks on the local and international levels.

5.4 Conclusion

Overall, this chapter offers a summary of the development of machine learning methods for the prediction of nutritional deficiencies by region. The goals of the research were accomplished through successful preprocessing of the data and cleaning of datasets, feature extraction and a viable method aimed at the predictions. Further, the findings were cross-verified and discussed in form of dynamic data-analytics dashboards to inform decision-making process that would address malnutrition. But the limitations of the work described herein are followed by a great number of opportunities for the work improvement by enlarging datasets, increasing models, and including factors to increase the predictive value. The analysis of such areas in the future will continue to increase the appreciation of nutritional health disparities and enable better ways of proving the intervention, and hence enhance the health of the global community.

References

Ahmed, G. N., & Kamalakkannan, S. (2022). Micronutrient classification in IoT based agriculture using machine learning (ML) Algorithm. *2022 4th International …*. https://ieeexplore.ieee.org/abstract/document/9716293/

Ali, J., Khan, M. A., & Khan, N. A. (2022). Machine Learning Approaches for Prediction of Nutrition Deficiency among Women of Different Age Groups. *2022 3rd International Conference …*. https://ieeexplore.ieee.org/abstract/document/10100340/

Gollapalli, M. (2022). Ensemble machine learning model to predict the waterborne syndrome. In *Algorithms*. mdpi.com. https://www.mdpi.com/1999-4893/15/3/93

Jain, S., Khanam, T., Abedi, A. J., & ... (2022). Efficient Machine Learning for Malnutrition Prediction among under-five children in India. *2022 IEEE Delhi Section …*. https://ieeexplore.ieee.org/abstract/document/9753080/

Kananura, R. M. (2022). Machine learning predictive modelling for identification of predictors of acute respiratory infection and diarrhoea in Uganda's rural and urban settings. In *PLOS Global Public Health*. journals.plos.org. https://journals.plos.org/globalpublichealth/article?id=10.1371/journal.pgph.0000430

Khudri, M. M., Rhee, K. K., Hasan, M. S., & Ahsan, K. Z. (2023). Predicting nutritional status for women of childbearing age from their economic, health, and demographic features: A supervised machine learning approach. In *Plos one*. journals.plos.org. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0277738

Meshram, V., Patil, K., Meshram, V., Hanchate, D., & ... (2021). Machine learning in agriculture domain: A state-of-art survey. In *... Intelligence in the Life …*. Elsevier. https://www.sciencedirect.com/science/article/pii/S2667318521000106

Mpakairi, K. S., Dube, T., Sibanda, M., & Mutanga, O. (2023). Fine-scale characterization of irrigated and rainfed croplands at national scale using multi-source data, random forest, and deep learning algorithms. In *ISPRS Journal of …*. Elsevier. https://www.sciencedirect.com/science/article/pii/S0924271623002460

Qasrawi, R., Sgahir, S., Nemer, M., Halaikah, M., & ... (2024). Machine Learning Approach for Predicting the Impact of Food Insecurity on Nutrient Consumption and Malnutrition in Children Aged 6 Months to 5 Years. In *Children*. mdpi.com. https://www.mdpi.com/2227-9067/11/7/810

Velásquez, D., Sánchez, A., Sarmiento, S., Toro, M., & ... (2020). A method for detecting coffee leaf rust through wireless sensor networks, remote sensing, and deep learning: Case study of the Caturra variety in Colombia. In *Applied Sciences*. mdpi.com. https://www.mdpi.com/2076-3417/10/2/697

Zhang, L., Zhang, Z., Luo, Y., Cao, J., & Tao, F. (2019). Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using machine learning approaches. In *Remote Sensing*. mdpi.com. https://www.mdpi.com/2072-4292/12/1/21

Zhou, L., Zhang, C., Liu, F., Qiu, Z., & ... (2019). Application of deep learning in food: A review. *... in Food Science and Food ...*. https://doi.org/10.1111/1541-4337.12492