

AFRICAN CENTER OF EXCELLENCE
IN DATA SCIENCE



College of Business and Economics

ANEMIA PREDICTION AMONG RWANDAN CHILDREN AGED 6 TO 59 MONTHS USING MACHINE LEARNING TECHNIQUES

By:

Diane UWIZERA

Registration Number: 220000067

University of Rwanda,

College of Business and Economics

A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science in Data Science, specialization in Biostatistics.

Supervisor: Dr. Pierre Claver RUTAYISIRE

September 2022

DECLARATION

I, Diane UWIZERA, hereby declare that this dissertation entitled: **"ANEMIA PREDICTION AMONG RWANDAN CHILDREN AGED 6 TO 59 MONTHS USING MACHINE LEARNING TECHNIQUES"** is my creative work, and it should never have been submitted to any University or other institutions for any degree. The reference list includes all sources of information used in this research.

Names: Diane UWIZERA

Registration Number: 220000067

Signature: 

APPROVAL SHEET

This dissertation entitled "**ANEMIA PREDICTION AMONG RWANDAN CHILDREN AGED 6 TO 59 MONTHS USING MACHINE LEARNING TECHNIQUES",** was prepared and submitted by Diane UWIZERA in partial fulfillment of the requirements for the degree of Master of Science in Data Science specializing in Biostatistics, is hereby accepted and approved. It was checked for plagiarism using Turnitin and the rate was 16%, which is lower than the ACE-DS threshold of 20%.



_____ _____

Dr. Pierre Claver RUTAYISIRE

Supervisor

_____

Dr. KABANO Ignace

Head of Training

DEDICATION

This dissertation is dedicated to the Almighty God, who guided me at every step of my studies. It is dedicated to my beloved parents, my love, my sisters, and my classmates. Thank you for the encouragement, patience, understanding, and incredible support that have ever been directed towards my work.

May God bless you abundantly!!

ACKNOWLEDGEMENT

This study would not have been possible without the support of different persons, who merit acknowledgment.

I would like to thank the Almighty God for his wonderful gifts of life, ingenuity, and opportunities offered to pertain to this research project as part of my Master's program.

I am grateful to the entire staff of the University of Rwanda-African Center of excellence in Data science for allowing me to undertake this program and assist throughout the research process.

My supervisor Dr. Pierre Claver RUTAYISIRE deserves special recognition for his guidance, intellectual support, encouragement, advice, clarification of concepts not well understood, and supervision that helped me in completing this project within the limited time frame.

My gratitude is extended to my relatives, friends, and family for their rounded contribution and support during my research project. And, I would like to express my sincere gratitude and gratefulness to my love for his helpful remarks on the manuscript of this work up to the completion of this study.

I am grateful to my classmates for their contributions and encouragement throughout the long period of our studies.

May God bless them all!!

ABSTRACT

Anemia, especially in children aged 6 to 59 months, remains to be one of the problems affecting public health in Rwanda. Anemia at a young age has major effects on children's developmental delays and impaired cognitive development. Therefore, there is a need for an accurate, reliable, and effective system or model that provides the likelihood of anemia given factors and assists in diagnosis at an early stage of anemia in time for proper treatment. This study aimed at building the best predictive model of anemia among Rwandan children aged 6 to 59 months by using the risk factors through machine learning techniques.

The Data used in this research were mined from RDHS 2019-2020 as the nationwide cross-sectional survey, whereby the KR and PR datasets were merged to obtain the necessary information. The participants in this study were 3525 children 6-59 months old whose mothers were interviewed. Several models were developed based on the stratified 8-fold cross-validation using five different machine learning algorithms such as Random Forest, XGBoost classifier, Support Vector Machine, Logistic Regression, and Naïve Bayes for the anemia prediction based on significantly associated factors. After an exhaustive evaluation of the predictive models through various measures including accuracy, sensitivity, specificity, F1 score, and AUC then, an outstanding model was found.

Based on the performance measures of different models, the best model was generated by the XGBoost classifier, which was chosen as the best classification with an accuracy of 85.21%, sensitivity of 98.71%, specificity of 71.70%, F1 score of 86.97%, and the area under the curve of 0.98. After modeling, the results indicated that the age of a child, child malaria infections, maternal anemia, children breastfeeding status, deworming and vitamin A in the last 6 months for a child, the residence of a child, and the household wealth index were the most contributing predictive features of anemia status. Overall, this study has proved that machine learning techniques can be considered in building an appropriate model, which can be helpful in the early stage of anemia detection among children and figure out the most important predictive factors, which can assist in controlling and preventing anemia in Rwandan children 6-59 months old.

Keywords: Anemia Prediction, Machine learning, children, Hemoglobin, Rwanda, XGBoost classifier, accuracy

TABLE OF CONTENTS

DECLARATION i
APPROVAL SHEET ii
DEDICATION iii
ACKNOWLEDGEMENT iv
ABSTRACT v
TABLE OF CONTENTS vi
LIST OF TABLES viii
LIST OF FIGURES ix
LIST OF ABBREVIATIONS AND ACRONYMS x
CHAPTER I. INTRODUCTION 1
I.1. Introduction 1
I.2. Background to the study 1
I.3. Problem statement 3
I.4. Research objectives 4
I.4.1. General objective 4
I.4.2. Specific objectives 4
I.5. Research questions 4
I.6. Significance of the research 5
I.7. Definition of operational terms in the study 5
I.8. Outline of the thesis 6
CHAPTER II. LITERATURE REVIEW 7
II.1. Introduction 7
II.2. Theoretical literature review on anemia 7
II.2.1. Generalities about anemia in young children aged 6 to 59 months 7
II.2.2. Hemoglobin concentration and diagnosis of anemia 8
II.2.3. Categories of anemia as the public health problem 8
II.2.4. Etiology of anemia 9
II.2.5. Theoretical and conceptual framework 11

II.3. Empirical literature review on anemia among children 12
II.3.1. Risk factors related to anemia in children 6-59 months old 12
II.3.2. Review of anemia prediction among children 15
II.4. Summary of literature review 16
CHAPTER III. RESEARCH METHODOLOGY 17
III.1. Data source and study design 17
III.2. Sampling method and sample size 17
III.3. Ethical considerations 18
III.4. Variables description 18
III.4.1. Response variable 18
III.4.2. Independent variables 18
III.5. Data preprocessing 20
III.6. Data analysis 21
III.6.1. Machine learning algorithms 21
III.7. Model evaluation 23
CHAPTER IV. RESULTS PRESENTATION 25
IV.1 Introduction 25
IV.2. Background characteristics of participants 25
IV.3. Findings on factors associated with the child's anemic status 26
IV.4. Predictive model's performance based on various metrics 28
IV.4.1. Findings of predictive models 28
IV.4.2. The best performance model 32
CHAPTER V. DISCUSSION, CONCLUSION, AND RECOMMENDATIONS 35
V.1. Discussion of the results 35
V.2. Conclusion 37
V. 3. Recommendations 38
REFERENCES 39
APPENDIX 44

LIST OF TABLES

Table 1. The thresholds of HB levels that are used to detect anemia at altitude level in g/l in the various populations [28] 8

Table 2. The anemia classification as a public health concern adapted from the reference [2] 8

Table 3. Recording and measurements of the explanatory variables to be used in this study..... 19

Table 4. The distribution of respondent's characteristics and the bivariate analysis of anemia among children in all selected variables 27

LIST OF FIGURES

Figure 1. Conceptual framework adapted from Balarajan et al., 2011 [27]**.** 12

Figure 2. The illustration of the Support Vector Machine works for classification[63]..... 22

Figure 3. The simplified illustration of the confusion matrix for the two classes [66]. 23

Figure 4. The ROC curve and confusion matrix for the Logistic regression model 29

Figure 5. The random forest model's confusion matrix and ROC curve. 30

Figure 6. The ROC curve and confusion matrix of the support vector machine 30

Figure 7. The Naive Bayes model's confusion matrix and ROC curve. 31

Figure 8. The XGBoost classifier's ROC curve and confusion matrix 31

Figure 9. Models comparison 32

Figure 10. The feature importance from the best model (XGBoost classifier) 33

Figure 11. Feature importance from the best model in predicting anemic children..... 34

LIST OF ABBREVIATIONS AND ACRONYMS

HB: Hemoglobin

TB: Tuberculosis

HIV: Human Immunodeficiency Virus

g/dl: grams per deciliter

RDHS: Rwanda Demographic and Health Survey

ML: Machine Learning

KR: Kids Records

PR: Household members (or Persons) Recode

UNICEF: United Nations Children's Fund

WHO: World Health Organization

LR: Logistic Regression

XGBoost: Extreme Gradient Boosting

NB: Naïve Bayes

RF: Random Forest

SVM: Support Vector Machine

ROC: Receiver Operating Characteristics

AUC: Area under Curve

KNN: K- Nearest Neighbors

ITNs: Insecticide –Treated bed Nets

BDHS: Bangladesh Demographic and Health Survey

HHs: Households

RNEC: Rwanda National Ethics Committee

ICF: Institutional Review Committee

SMOTE: Synthetic Minority Over-Sampling Technique

%: Percentages.

CBC: Complete Blood Count

NGOs: Non-Governmental Organizations

CHAPTER I. INTRODUCTION

I.1. Introduction

This chapter describes the background, problem statement, significance, and definitions of the key terms of the study. It also covers the research objectives, including general as well as specific objectives and research questions that oriented the researcher to obtain adequate results.

I.2. Background to the study

Anemia refers to the state in which the amount of blood cells or hemoglobin (HB) in the blood is lower than normal [1], whereby hemoglobin is needed to transport oxygen to the body's tissues. The normal distribution of red blood cells varies with age, gender, biological status, and altitude (residential elevation above sea level), while anemia affects people at all phases of growth, it is especially prevalent in children under the age of five and pregnant women [2], [3].

The etiology of anemia is multifactorial, with iron deficiency accounting for up to 50% of anemia worldwide [2], but this rate may vary from region to region due to local conditions. Anemia can also be caused by the deficiencies of micronutrients such as vitamin A, folic acid, and vitamin B12, heavy blood loss, hemoglobin hereditary diseases, presence of communicable diseases like malaria, cancer, hookworms, Ascaris, schistosomiasis, HIV, and tuberculosis [1], [2], [4].

The HB threshold without anemia for 6-59 months old children is ≥11.0 grams per deciliter and the diagnosis of anemia is categorized as normal, mild, moderate, and severe according to the concentration of Hb levels [5]. Anemia is an indicator of malnutrition and low immune disorders

which has serious consequences for children as a result of growth retardation, poor academic performance, laziness, impaired cognitive development, and increased mortality [1], [4].

Anemia is a global health concern that affected 1.74 billion people (22.8%) in 2019, while 269 million (approximately 39.8 %) of children 6-59 months old were affected by anemia [6] with a decrease from 42.6% in 2011 [5]. This varies by country, while the greatest number of prevalent cases of anemia was found in South Asia, Western Sub-Saharan Africa, and Eastern Sub-Saharan Africa[6]. The African Region has the greatest rate of severe anemia (hemoglobin of 7g/dl), with 3.6 percent of children affected [5].

Malaria and the deficiency of iron were the most prevalent causes of anemia in Kenya, according to the population-based cross-sectional study [7]. A similar study in Ethiopia but different in age group showed that children between 9 and 11 months of age, insufficient food variety, malnutrition, the lowest wealth index, the starting of complementary foods earlier, and male gender, were the major factors of anemia in children under five years [8]. Meanwhile, severe anemia among Tanzanian children under five years was linked to unemployment among caregivers, the presence of sickle Hb, and malaria parasites [9].

The most prevalent causes of anemia in Rwandan children aged 6 to 59 months are inadequate intake of iron, hookworm infection, and malaria parasite based on RDHS report 2019-2020 [10]. The anemia among children declines with age, with increasing mother's education, family income grows and is highly prevalent in rural areas [10]. According to R.Nkulikiyinka et al., malaria status was the single strongest factor that was associated with anemia, being anemic was also substantially linked to the male sex, young children under 5 years, and being underweight [11]. A non-educated mother and a recent fever in children under five years were two other factors with a less significant relationship to anemia [11].

Although the Rwandan government has made significant progress in reducing the rate of infant mortality, children under five years who suffer from anemia remain a burden. The RDHS over the last 20 years has shown that anemia among young children 6-59 months old has dropped significantly, whereby in 2005, was 52% and fell to 37% in 2015 [12], but this has remained stable at 37% according to the 2019-2020 report (RDHS) with mild anemia accounting for 21%, moderate anemia for 15%, and severe anemia accounting for fewer than 1% [10]. Therefore, this is a moderate public health concern based on the WHO standards. As is often said, "prevention is better

than cure", so predicting diseases and spreading infections can lead to early disease prevention [13].

However, the proper intervention and early diagnosis of anemia in children can reduce the risk of anemia, various clinical tools can be developed to assess the risk of anemia, and the main point of the risk appraisal techniques is accurate disease prediction [14]. Machine learning is a discipline that mixes statistical learning and artificial intelligence research to discover hidden patterns in massive amounts of data [15]. Various ML techniques can assist in the building of predictive models and have been used to more accurately forecast diseases based on the established risk factors and to aid patients in their treatment [13].

I.3. Problem statement

The government of Rwanda has developed several initiatives to eradicate anemia in children under five years and women of childbearing age by promoting and introducing iron-bio fortified beans among farmers that release in 2010[16], and twice-yearly, a one-week campaign is launched against anemia and malnutrition of mothers and children by providing vitamin A supplements, deworming for children 6-59 months old, food safety strategies and improving the health of children and mothers [17].

Despite this implementation of the Rwandan government for eradicating anemia, its prevalence continues to be high in children 6-59 months old, which remains stable at 37% in 2020 [10], making this a public health concern for Rwanda. According to R. Nkulikiyinka et al. (2015) using RDHS 2010 found that malaria was the most important factor related to anemia among children aged 6-59 months [11]. Currently, further research is needed to update information on key factors contributing to anemia so that the correct and effective preventive actions may be taken to lower the anemia prevalence among children aged 6-59 months.

In the fields of medical research and health, machine learning techniques have become increasingly popular, and compared with traditional statistical methods, they have shown powerful predictive models in solving classification problems[18]. There is limited research using machine learning approaches to create predictive models for anemia in children [14], [19]. Based on the

current studies, the researchers have provided evidence that anemia can be predicted using machine learning algorithms based on existing risk factors, thereby assisting in the treatment and prevention of anemia in children [14], [19]. Whereby, the widest classifier (logistic regression) used in predicting disease status showed the least accuracy with 62.75% and 61.67% respectively compared to the Random Forest (RF), which was one of the machine learning methods with the accuracy (68.53% and 67.18 %, respectively) [14], [19].

The majority of health research in Rwanda relies on statistical methods [11], [20], but there is a gap in using vigorous techniques. Therefore, developing an accurate model that can predict anemia in Rwandan children is essential for making the right decisions concerning anemia diagnosis and its management. Furthermore, there is a scarcity of published research that uses machine learning approaches to predict childhood anemia in Rwanda using cross-sectional health survey data, necessitating the conduct of this research.

I.4. Research objectives

I.4.1. General objective

- Using machine learning techniques, create the best accurate predictive model of anemia in Rwandan children aged 6 to 59 months.

I.4.2. Specific objectives

- To identify the factors associated with anemia in Rwandan children 6-59 months old.
- Building the predictive models of anemia among Rwandan children 6-59 months old through various machine learning techniques and testing their performance with different measures.
- To generate the best-performing model from the developed predictive models.
- To find out the most predictive features that greatly contributed to the best model.

I.5. Research questions

- What are the factors related to anemia in Rwandan children aged 6 to 59 months?
- How do build different predictive models of anemia among young children using ML techniques?

- What is the best performing model for predicting anemia in Rwandan children aged 6-59 months?
- What are the main predictive factors that contribute to anemia in Rwandan children aged 6 - 59 months from the best model?

I.6. Significance of the research

This section illustrates the expected users of the research findings, as well as the impact on policy change. As the accurate predictive model was established:

- The generated information will facilitate the health planner and policymakers in designing and implementing appropriate interventions to eradicate anemia through better prevention and control of childhood anemia.
- The accurate model and the identified potential factors will help medical practitioners to easily diagnose anemia, prescribe treatment as early as possible, as well as teach people about the factors that contribute to anemia, hence the decrease in severe anemia prevalence.
- The feature importance of this study will help mothers/caregivers and the community to adopt the best strategies to protect themselves and their children from anemia.
- The researchers will use this model, with new data in the prediction of anemia for further research in the future.

I.7. Definition of operational terms in the study

Anemia: Anemia is a disorder in which the number of red blood cells or hemoglobin levels in the blood is insufficient to fulfill the physiological requirements of the body, according to the WHO [3]. In this study, anemia was found when the level of hemoglobin was less than 11.0 g/dl in young children 6-59 months old.

Hemoglobin (HB) is a protein found in red blood cells that distributes oxygen throughout the body while also returning carbon dioxide from organs and tissues to the lungs [21]. In our context of the study, the normality of hemoglobin in young children aged 6 to59 months old is ≥11.0g/dl otherwise it is abnormal (anemic).

A prediction is the expected result of a test that is derived from a hypothesis or theory [22].

Predictive modeling is the process of predicting new or future observations using a statistical model, machine learning, or data mining approaches [23].

Machine learning (ML) was initially defined by Arthur Samuel as a "*field of study that gives computers the ability to learn without being explicitly programmed*" and according to Tom Michel, ML is a computer program that is deemed to be learned by experience if it improves its performance in a task, according to Tom Michel [24].

I.8. Outline of the thesis

The study is sectioned into five chapters: the introduction, the literature review, the methodology, the presentation of results, the discussion of findings, and the conclusion with recommendations.

CHAPTER II. LITERATURE REVIEW

II.1. Introduction

This chapter covers the literature on the theoretical and empirical review of the scientific research findings. This literature review explored in-depth the factors that were linked to anemia in children 6-59 months old and the machine learning-based predictive models from available national and international studies through various sources such as reports, journals, articles, and other published materials.

II.2. Theoretical literature review on anemia

II.2.1. Generalities about anemia in young children aged 6 to 59 months

Anemia is a major health concern that impacts people in both developed and developing countries worldwide, with severe effects for children such as impaired cognitive, motor development, growth retardation, and mortality [1], [4], [25].

Anemia affects approximately a quarter of the worldwide population, with young children under the age of five being the most impacted, accounting for 43 percent of all cases, whereas South Asia was the most affected, with 102 million people, next is Central and West Africa that has 53 million people [26]. In a population, the prevalence and distribution of anemia are influenced by a complex interaction of political, biological, ecological, and social factors, with anemia prevalence being inversely associated with economic development at the country level and is patterned by demographic and socioeconomic factors [27].

Significantly, it is important to mention that anemia is extremely common in Sub-Saharan Africa, with some nations reporting prevalence rates of over 70% in the general population [11], [26]. However, the RDHS revealed that anemia prevalence in Rwandan young children aged 6 to 59 months dropped remarkably from 52% in 2005 to 38% in 2010 [12] and slowed from 38% in 2010 to 37% in 2020 [10].

II.2.2. Hemoglobin concentration and diagnosis of anemia

The World Health Organization established the following standards for detecting anemia and categorizing it in various populations.

Table 1. The thresholds of HB levels that are used to detect anemia at altitude level in g/l in the various populations [28]

Populations	Non-Anemic(g/l)	Anemic		
		Mild	Moderate	Severe
Children between 6-59 months old	≥110	100-109	70-99	<70
Children between 5 to 11 years old	≥115	100-114	80-109	<80
Children 12 to 14 years old	≥120	110-119	80-109	<80
Women who are not pregnant (≥ 15 years)	≥120	110-119	80-109	<80
Women who are pregnant	≥110	100-109	70-99	<70
Men between the ages of 15 years and above	≥130	110-129	80-109	<80

II.2.3. Categories of anemia as the public health problem

The WHO classifies anemia prevalence for public health monitoring purposes to evaluate whether anemia is a major health concern and the severity of intervention challenges. The table below shows the global health relevance of anemia based on its prevalence.

Table 2. The anemia classification as a public health concern adapted from the reference [2]

Classification of public health problems	Anemia prevalence (%)
No public health problem	<4.9
Mild public health problem	Between 5.0 to 19.9
Moderate health problem	Between 20.0- 39.9
Severe health problem	≥40.0

II.2.4. Etiology of anemia

Anemia is developed through three main mechanisms that depend on physiological abnormalities: hemolysis (when red tissues are damaged), inactive erythropoiesis (when the body produces an insufficient number of red blood cells), and blood loss [29]. Nutritional deficiencies, genetic abnormalities, and infections are the main contributors to anemia, according to the World Health Organization [29].

II.2.4.1. Nutritional deficiencies

Iron is an important nutritious mineral required for the synthesis of hemoglobin in red blood cells. Deficiency of iron is the most prevalent type of malnutrition leading to anemia [2] and accounts for 42% of all cases of anemia in children under five years and 50% in women [5], [26]. The terms: anemia, iron deficiency, and iron deficiency anemia are frequently used interchangeably in the literature, but they are not identical. Iron deficiency anemia is a disorder in which the body develops insufficient red blood cells due to a lack of iron. It's the third and final phase of iron deficiency, which begins with a reduction in serum ferritin content [7].

Another micronutrient affects red blood cell metabolism, either directly or indirectly. While vitamins (like vitamin A, folic acid, vitamin B12, riboflavin (vitamin B2), and vitamin B6) are essential for optimal red blood cell synthesis, vitamin C and vitamin E protect mature red blood cells from early damage to free radicals[30]. Vitamin A deficiency is linked with chronic anemia through impaired immunity to infection, impacting iron storage mobilization and replacement, reducing the incidence of anemia [29], [31].

Low levels of vitamin B12 lead to pernicious anemia due to a lack of production of intrinsic factors by gastric parietal cells. It is necessary for the absorption of B12 in the intestine and ultimately leads to the development of anemia or severe neurological symptoms. Pernicious anemia can affect people of all ages [32], [33]. Other malnutrition conditions that can lead to anemia include deficiency of copper, vitamin B6, vitamin D, vitamin C, vitamin E, and riboflavin deficiency, but a few of these malnutrition cases are scarce and may not contribute significantly to the global anemia burden [29].

2.2.4.2. Diseases

Chronic and acute infections such as cancer, malaria, tuberculosis, HIV, chronic heart failure, and chronic kidney disease can limit nutrition absorption and metabolism resulting in the loss of nutrients which leads to anemia known as “Chronic disease anemia/inflammation anemia. It is the second most prevalent anemia after iron deficiency anemia [29]. Malaria is the most frequent cause of anemia in East and West Africa [5] with malaria accounting for 15% of anemia among preschool children in West Africa [34].

Anemia is prevalent in TB patients, and it may be even more common in individuals who have both TB and HIV. Anemia is associated with HIV both indirectly and directly opportunistic infections, which are frequent in HIV patients, are an indirect effect of HIV infection on anemia (such as malaria, hookworm, and antiretroviral therapy harms erythropoiesis. HIV directly affects anemia, which damages hematopoietic progenitor cells and reduces sensitiveness to erythropoietin [35].

Blood loss and iron deficiency anemia are linked to hookworm and schistosomiasis. Tropical and subtropical regions, particularly in Sub-Saharan Africa and Southeast Asia, are the most prevalent. Hookworm infections are aggravated by poverty, sanitation, hygiene, poor water, and infrastructure. Maternal anemia is commonly caused by hookworm infections whereby hookworm infection is responsible for one-third to half of the anemia in young children and pregnant women in Asia and Africa [36].

II.2.4.3. Genetic hemoglobin disorders

Anemia can also be caused by structural abnormalities or a decrease in the generation of hemoglobin chains. Every year, more than 300,000 newborns are born with a genetic hemoglobin disease around the world, with over 80% of these babies occurring in low and middle-income nations [37].

Sickle cell disorder is the most prevalent hereditary hemoglobin abnormality, which is mostly found in Sub-Saharan Africa, followed by β- and α- thalassemia, which are mostly found in Southeast Asia. Nearly 5% of the world's population is thought to have a major hemoglobin variation, with the percentages being substantially higher in Africa (18%) and Asia (7%) [29].

II.2.5. Theoretical and conceptual framework

II.2.5.1. Theoretical framework

This theoretical framework (appendix1), has been used to refer to the determinants of anemia and has been used by different researchers in their studies to analyze and understand the causes of anemia, especially in women and children [38], [39]. In its conceptual framework, there are many determinants by various means leading to anemia, including political, environmental, social, and biological factors. Anemia is usually caused by a decrease in the synthesis of erythrocytes or an increase in the loss of erythrocytes at the biological level [27], [29].

There are several complex behavioral, social, and environmental features that contribute to the occurrence of anemia, particularly in children, in addition to such biological factors. These factors are frequently categorized into three groups in this theory wherein the first instance, there are immediate causes, such as insufficient food intake and absorption, as well as infectious disorders (fever, malaria, and intestinal parasites). Second, at the family/household level, there are underlying causes such as sanitation, access to safe drinking water, health care, a variety of food sources, the usage of insecticide-treated bed nets, and anemia preventive skills [27].

In addition, this theory showed that more distant characteristics like wealth, cultural norms, education, and behavior are linked to a larger social and cultural background, whereby children living in families with the lowest wealth index were more likely to be anemic compared to those in the highest wealth index. Anemia was also found to be significantly connected to mothers' education level, and maternal anemia was found to be strongly and positively linked to anemia in children [27].

II.2.5.2. Conceptual framework

The conceptual model of this study is based on the Balarajan et *al*. model of anemia determinants. But some variables are not used in this model because the study is based on secondary data (RDHS 2019-2020), where all aspects of the framework are not available for analysis. It explains the risk factors that are strongly linked to anemia in Rwandan children 6 to 59 months old.

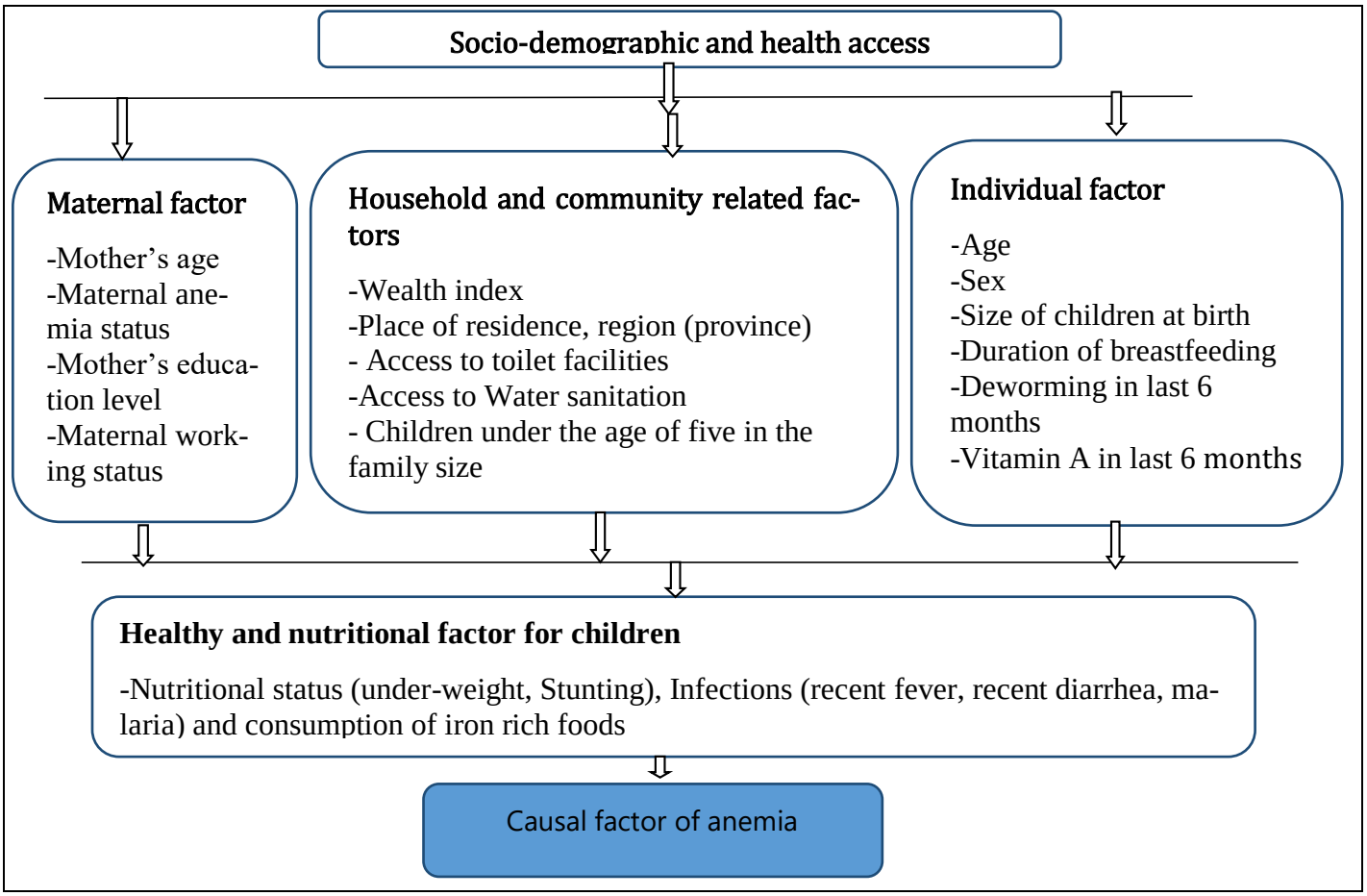


Figure 1. Conceptual framework adapted from Balarajan et al., 2011 [27]**.**

II.3. Empirical literature review on anemia among children

The empirical review of this literature showed the risk factors related to anemia in the prior studies and the previous machine learning models in anemia prediction among children to respond to the current research questions of this study.

II.3.1. Risk factors related to anemia in children 6-59 months old

The factors contributing to anemia are multifactorial and vary from country to country and within a country's population groups, they are ranging from demographics and socio-economic factors including individual factors and health conditions, mother's health status, household, and community-related factors.

II.3.1.1. Individual factors and health conditions

Among individual factors and health conditions include age, sex, recent fever, diarrhea episodes, malaria, and nutritional status of the child. The risk of having anemia in children decreases with age because they are developing and growing rapidly. After all, the mother's stored iron may be insufficient as early as six months of age [40]. Referring to the studies conducted in Bangladesh and Ethiopia whereby children in the range of 6 to 23 months old were more likely to be anemic than children aged 24 to59 months due to the nutritional imbalance during the first 2 years of life, the iron concentration or other micronutrients in breast milk was low and insufficient for a child's daily growth requirements [41], [42].

Previous studies revealed that sex was the risk factor for anemia in children. Male children under the age of five were shown to be more likely to be anemic. This disparity could be due to male children's rapid growth in the initial months of life, which raises their micronutrient requirements, particularly iron [8]; similar studies were done in Kenya [7], and Ethiopia [43]. Contrary to the one conducted in Cape Verde and another study in Ethiopia found that sex difference was not associated with anemia [44], [45].

As far as the nutritional status is concerned, various studies found that undernourished children have a higher risk of getting anemia than nourished children. Stunting is an indicator of chronic malnutrition, whereby anemia was more common in stunted children [1], [43]. Furthermore, wasted children were at a higher risk of becoming anemic. [42]. However, the study conducted in Namibia, could not show a link between anemia and the nutritional factors of stunting, under-weight, and wasting [46].

According to WHO, children who are solely breastfed for the first 4 to 6 months of life do not require supplements, however, after 6 months, children require an additional source to maintain adequate iron nutrition, control parasitic disease (deworming), and a vitamin A complement to prevent iron deficiency anemia [47].

Children who did not take iron-fortified foods (meat, beans, dark green leafy vegetables, seafood, and so on) were more prone to be anemic than their counterparts [48]. This contrasts with the study

conducted in Rwanda and Togo where their findings did not show the association between anemia and iron intake (iron-rich foods) [11], [25]. Regarding vitamin A supplements and deworming, previous studies have shown an association of them with anemia among children [11], [43] contrary to the study conducted in Namibia [46].

Fever is a common sign of both acute and chronic disorders, according to the study done in Bangladesh [41], and it has been linked to low hemoglobin concentration and anemia, the children who had diarrhea within the previous two weeks of the survey interview were more prone to manifest anemia [8], [44]. Regarding the child's malaria status, children with malaria were more likely to suffer from anemia than children without malaria, due to its lower red blood cell production rates. [7], [25].

II.3.1.2. Maternal factors and health status

Previous studies have found that maternal factors are related to the occurrence of childhood anemia. The children whose mothers were aged < 20 years were more susceptible to anemia than those with mothers above the age of 20 [41], [49]. Furthermore, some studies have shown that anemic mothers had a higher chance of having anemic children, possibly because both mother and child were usually in the same environment, which may involve the same risk factors [25], [41], [42].

Maternal education is considered as one of the factors which greatly influence the status of anemia in a child, the majority of research found that anemia was less common in children with mothers who had at least a secondary school diploma [7], [8], [25], [41], [50], this is because they have knowledge of good nutritional habits and take into consideration the quality of nutritious foods. In addition, they have a better understanding of and provide a clean and healthy diet to their children, but this differs from studies in [46].

II.3.1.3. Household and community-related factors

Various studies revealed that household income is associated with food insecurity[51], where children in low-income households were more prone to be anemic than children in high-income families [7], [44], [46]. Studies conducted in different areas found that anemia is influenced by where

you live. Children in rural areas were more likely to be anemic than children in urban areas [25], [41] differs from the study in Namibia [46].

Regarding hygiene, even if it is just indirectly, poor sanitation is associated with anemia, which is related to numerous infectious and parasitic diseases and contributes to lower hemoglobin levels [42], anemic children were more prevalent in households without access to improved water and toilet facilities [41], [52].

II.3.2. Review of anemia prediction among children

Previous studies have been conducted on anemia prediction given the risk factors where machine learning algorithms were used in different countries, but very few researchers have explored the application of machine learning approaches to the development of predictive models for childhood anemia [14], [19].

The study conducted at the University of Allahabad used the data collected from pathology centers and laboratory test centers for anemia prediction by using CBC (complete blood count). They used the supervised machine learning algorithms such as Naive Bayes, random forest, and decision tree algorithm, and the result suggested that with an average accuracy of 96.09 percent, the Naive Bayes classification algorithm outperformed the decision tree and Random Forest classification algorithms in terms of accuracy [53].

Referring to the study conducted in the hospital of Jammu (India) to predict anemia in young children aged 6-36 months, they developed predictive models by applying machine learning methods such as Classification and Regression Trees (CART), Random Forest, K- Nearest Neighbors, Logistic Regression and Linear Discriminant Analysis. The results suggested that the Random Forest algorithm performed the best in accuracy compared to the rest of the classifiers in forecasting whether a child is anemic based on risk variables, and the findings should aid decision-makers in detecting children at greater risk of anemia in the future, with a focus on the child's nutritional and mother’s education [19].

According to the study done in Bangladesh, ML techniques including Classification and Regression Trees, Linear Discriminant Analysis, Support Vector Machines, K-Nearest Neighbors, Logistic Regression, and Random Forest were used to develop multiple predictive models utilizing identified risk variables for anemia in children aged 6-59 months based on the 2011 BDHS. Therefore, a systematic evaluation of the algorithms was conducted and discovered that the RF model provided the best results. Thus, the use of ML techniques in this study has shown that interfering with certain sociodemographic and health-related factors can significantly reduce the risk of anemia in children [14].

II.4. Summary of literature review

This literature review describes the majority of information about anemia in children 6-59 months old from various studies. It also summarizes and analyzes the cause and the risk factors associated with anemia, which were identified by the researchers in their respective areas. Anemia in children 6-59 months old has often proven to be a global burden, but it is prominent in developing countries, especially in Central West Africa and South Asia [26].

In addition to understanding the likelihood of causal anemia, this study is intended to know the prediction of anemia status which is crucial for community and healthcare policy and resource planning prediction [14]. This was the reason why the ML techniques were applied to create an accurate model that predicts anemia status in Rwandan children 6-59 months old using the common risk factors like features as part of the research gap, hence the conduct of this study.

CHAPTER III. RESEARCH METHODOLOGY

III.1. Data source and study design

This research used the secondary data mined from the Rwanda Demographic and Health Survey (RDHS) conducted in 2019-2020. It is a quantitative, nationwide survey based on a cross-sectional study design, conducted every five years.

III.2. Sampling method and sample size

The data used in this research was conducted by using a two-stage stratified sampling procedure, with the first stage selecting 112 enumeration areas (also known as villages or clusters) in urban regions and 388 in rural areas, with the likelihood proportional to the number of household's resident in the village. A systematic sample of households was conducted in the second stage, where twenty-six (26) households (HHs) were randomly chosen on average per cluster in a total of 12,949 households interviewed, which had 14634 women aged between 15 and 49 years and 8092 children under the age of five [10]. The RDHS report for 2019/2020 covers extensive information on sampling methodology and the whole survey [10].

For all young children aged from 6 to59 months and women aged 15 to 49 years who gave their consent, blood specimens for hemoglobin measurement were collected immediately using the battery-operated portable HemoCue 201+ analyzer. A drop of blood obtained by pricking a child's finger was put into a micro cuvette and examined using a photometer that showed the hemoglobin concentration [10]. In the 2019-2020 RDHS, 4087 children 6-59 months old were selected for anemia testing from those whose mothers were interviewed, however, only 3525 children were included in the study since they received anemia test results.

Survey weights were applied to adjust the sample scheme's unequal probability of selecting households and non-response. For this study, the area of interest covers data on aspects of children's demographic, socio-economic, and health factors, and related variables were obtained by merging the KR and PR datasets to provide all the necessary information. The Demographic and Health Survey methodology report has a complete explanation of the weighting techniques [54].

III.3. Ethical considerations

The RDHS 2019-2020 survey was appraised and authorized by the Rwanda National Ethics Committee (RNEC) and the Institutional Review Committee (ICF). The informed consent was obtained from each respondent or parents/ from adults responsible for the child before interviewing. The RDHS program removed the defendants' personal information from the database before making it available to the public [10]. Therefore, the request for using RDHS data was made by the researcher and approved by the Demographic and Health Survey (DHS) program and authorized the access to download the dataset from the DHS program website (appendix2).

III.4. Variables description

III.4.1. Response variable

In this research, the response variable was anemia status which was categorized according to World Health Organization's criteria whereby the 6-59 months old children with hemoglobin levels below 11.0 g/dl were considered as "anemic" otherwise "non-anemic".

III.4.2. Independent variables

The independent factors linked to the child's features, the child's mother, and household features were summarized in Table 1. As explained below, some variables related to demographic and socioeconomic were computed using other variables from the dataset:

The iron-rich foods were used in defining children who took at least one of these foods in the last 24 hours before the date of the interview: i) meat, ii) pumpkin, carrots, squash, iii) any dark green leafy vegetables, iv) liver, heart, other organs, v) fish or shellfish, vi) food made from beans, peas, lentils, or nuts, and vii) eggs otherwise did not take it [49].

According to the DHS program, a household's source of drinkable water and the kind of toilet facility were categorized as improved and unimproved based on the WHO/UNICEF guidelines. The piped into dwelling, piped to yard/plot, public tap/standpipe, tube well borehole, protected well, protected spring, as well as rainwater, were among the improved sources of drinkable water, while the unprotected well, unprotected spring, cart with a small tank and surface water (river/dam/lake/pond/stream/canal/irrigation channel), were among the unimproved sources [54].

For the type of toilet facilities, they classified flush - to the piped sewer system, flush - to the septic tank, flush to pit latrine, pit latrine - ventilated improved pit (VIP), pit latrine - with slab, composting toilet as improved and flush - to somewhere else, flush - don't know where, pit latrine - without slab / open pit, no facility/bush/field, and others as unimproved [54].

Table 3. Recording and measurements of the explanatory variables to be used in this study

Variable name	Description of variable	measurement
Variables related to child's characteristics		
Age (hw1)	A child's age	0:"6-23", 1:"24-42", 2:"43-59"
Sex (b4)	Sex of child	0="Male", 1="Female"
Child's birth size (m18)	The size of the child at birth	0: "Large",1: "Average",2: "Small"
Malaria history (hml32g)	A blood smear test revealed the final result of a malaria test	0:"Negative", 1: "Positive"
Presence of fever (h22)	The child had a fever in the last 2 weeks	0:"No", 1:"Yes"
Presence of diarrhea (h11)	The child had diarrhea in the last 2 weeks	0:"No", 1:"Yes"
Deworming in the last six months (h43)	Medications for intestinal parasites in the last six months	0: "No" 1: "Yes"
Vitamin A in the last six months (h34)	Vitamin A in the last six months	0: "No",1: "Yes"
Duration of breastfeeding (m4)	A child's breastfeeding duration	0: "ever breastfed, not currently breastfeed",1: "never breastfed",2: "Still breastfeeding"
Child stunting status (hw70) computed from DHS	Height /age standard deviation (new WHO), stunted if <= -200 otherwise not stunted	0: "No",1: "Yes"
Child underweight status (hw71) computed from DHS	Weight /age standard deviation, underweight if <= -200 otherwise normal	0: "No",1: "Yes"
Iron-rich foods	Source of iron supplements foods within 24 hours from the date of interview	0: "No"1: "Yes"
Variables relating to the mother of children		
Maternal age (v013)	Mother's age	0: "<20",1: "20-29", 2: "30-39",3: ">=40"
Maternal education (v106)	Mother's educational level	0:"No education", 1:"Primary", 2:"Secondary or Higher"

Maternal working status (v714)	Working status of a mother	0: "No",1: "Yes"
Maternal anemia (v457)	Anemia status of a mother	0: "Not anemic", 1:"Anemic"
Variables related to Households		
Residence (v025)	Residential location of the child	0:"Rural",1:"Urban"
Wealth index (v190)	Households' socio-economic status	0:"poor", 1:"Middle": 2:"Rich"
Region (v024)	Region (province) of a child residence	0: "Kigali",1: "South",2: "West", 3: " North", 4: "East"
Source of Drinking Water (v113)	A household's source of drinking water	0: "unimproved',1:"improved"
Type of toilet facilities (v116)	A household's type of toilet facilities	0: "unimproved', 1: "improved"
Children under five years in the family size (v137)	all children under the age of five in a household	-

III.5. Data preprocessing

Data preprocessing is the practice of turning raw data into an understandable and readable format to make it suitable for analysis. Some of the techniques involved are data cleaning, data integration, data transformation, and data reduction [55]. In this study, Data cleaning was performed as the process of handling missing values or incomplete data within the dataset. The KNN imputer was used to fill the missing values, which minimizes the data similarities by applying Euclidean distance [56].

Data transformation refers to the procedure of changing data format into the required one for analysis preparation. The map function was used to encode categorical data and pandas. get _dummies were used to transform them into dummy (0 and 1) variables that treat categories of variables separately.

The numerical data were then normalized using the Minimax Scaler, which is a function that spans all data values between 0 and 1 without modifying the original data. In addition, the Synthetic Minority Over-Sampling Technique (SMOTE) was utilized to solve the target variable's class imbalance, which over-sampled the minority class k in the dataset by generating new synthetic instances of the minority class along the line segments linking the minority instance and its identified

nearest neighbors of minority class k to increase the number of instances without any additional information to the model [57].

III.6. Data analysis

In this research, Python software was used as the tool for data analysis. After data preprocessing, the point biserial correlation and Pearson chi-square test with 95% confidence levels were performed to determine if there is a link between the dependent variable which is anemia, and the independent continuous and categorical variables. Through machine learning techniques, models were developed via the stratified K-fold cross-validation method to assess their overall performance when used to make predictions on data that was not utilized during the model training.

The learning dataset is split into K distinct subgroups/ folds of about equal size in stratified K-fold cross-validation, which is done by stratification random sampling by ensuring that each subgroup has the same proportion of observations with a specific categorical value as the class outcome value. The model's training set consisted of K–1 subset, while the remaining subset was served to evaluate the model's performance as the validation set. This process is done till every K subset is treated as a training and validation set, indicating that all observations of the whole dataset underwent the training and testing [58]. In addition, the feature importance was used to describe the contribution of features in the best model.

III.6.1. Machine learning algorithms

The ML approaches are classified into two categories: supervised and unsupervised based on algorithms[59]. Supervised learning refers to the techniques in which a model is trained on a variety of factors with a specific outcome, whereas unsupervised learning does not require a predetermined outcome and is used to find unspecific patterns that occur in datasets [59]. The supervised learning algorithms can predict the discrete or continuous outcomes when applied to new data after being effectively trained, and the models which generate discrete categories are the classification algorithms otherwise the regression algorithms [59].

In this study, the following classification algorithms for supervised machine learning have been used:

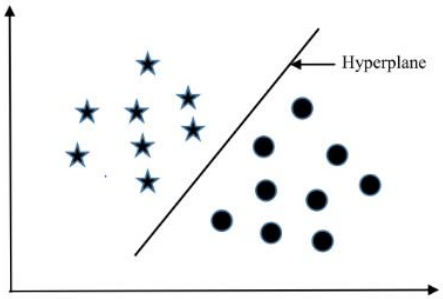
Logistic Regression (LR): is one of the most common classification algorithms broadly used in medical, business, economics, and other areas. LR is a quick, relatively straightforward method for determining the association between the explanatory variables and the dichotomous outcome variable [60] with logic and simple interpretation in medical research because of using odds or probability rather than the risk in the link function [61].

$$\boldsymbol{Logit(p) = \ln\left(\frac{p}{p-1}\right) = \alpha + \beta x}$$

Whereby p stands for the probability of success at value x, ∝= value of p when β = 0 and β: defines the rate of change**.**

Random Forest (RF) is a well-known monitoring algorithm for machine learning. This approach consists of a series of linked decision trees to create a more stable and accurate prediction, and it may be used for both classification and regression problems. The attributes of each tree in the forest are utilized to classify a person. Trees are planted at random and assigned a class marker based on their votes, with the decision made by the majority of the forest's trees [14].

Support Vector Machines (SVM) is a common supervised ML method that can be utilized as classifiers and predictors. The SVM algorithm creates a hyperplane by maximizing class margins and completely separates training observations according to the class label. The SVM assigns test observations to classes relying on which part of the hyperplane it is on [62].



For example, in this figure, the SVM identified the right hyperplane to classify 2 classes without misclassification by selecting the maximum distance between the nearest data point (margin).

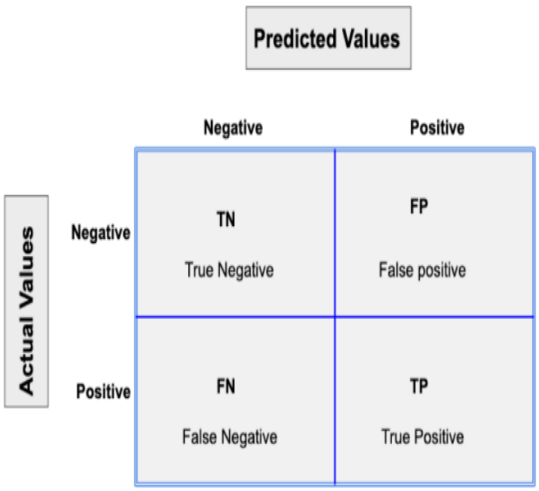
Figure 2. The illustration of the Support Vector Machine works for classification[63]

Naïve Bayes is a simple and effective classification ML technique based on the Bayes principle. It refers to the independence between the factors which are not related to each other or should be unrelated in any manner. Regardless of the dependency, it is called Naïve because all of these traits contribute to the probability. This requires very little training data [53].

Extreme Gradient Boosting (XGBoost Classifier): is the ML algorithm that utilizes the ensemble rule to transform a group of weak learners into strong learners, improving the model's efficacy, processing capacity, and predictability. It belongs to the boosting algorithm family as an upgraded form of the gradient boosting algorithm. It is a member of the Distributed Machine Learning Community and an open-source library. XGBoost is a unique mixture of software and hardware features aimed to improve conventional boosting methods with accuracy and control [64].

III.7. Model evaluation

Model evaluation was used to find out the effectiveness of the classification model and the model's ability to classify data. In this research, the Receiver Operating Characteristics (ROC) and confusion matrix were utilized to evaluate the algorithm's performance with various measures including Area under Curve (AUC), accuracy, sensitivity, specificity, and F1 score.



The confusion matrix is a summarized table that is used to assess a classifier's performance by comparing the predicted and the true /actual values[65]. This study used the confusion matrix of two classes whereby, True Positive (TP) indicates that the model correctly predicts the positive class, while False Positive (FP) implies that the model incorrectly predicts the positive class. The True Negative (TN) indicates that the model correctly predicts a negative value, while the False Negative (FN) means the model incorrectly predicts the negative class [66].

Figure 3. The simplified illustration of the confusion matrix for the two classes [66].

From the confusion matrix, the following metrics were generated:

Accuracy: the accuracy of the model is a metric that generally describes how the model performs across all classes, it is calculated as the proportions of observations that are classified correctly to the total number of predictions [65]. $\boldsymbol{Accuracy = \frac{TN+TP}{TN+FN+TP+FP}}$

The recall or sensitivity (True positive rate) is the metric that evaluates the ability of the model to predict true positives correctly. It is calculated as the proportions of observations classified correctly as positive to the total positive observations[65]. $\boldsymbol{Recall = \frac{TP}{TP+FN}}$

The specificity is the metric that evaluates the ability of the model to predict true negatives correctly. $\mathbf{Specificity = \frac{TN}{FP+TN}}$

The F1-Score illustrates a combination of precision and recall. The F1-Score is the symmetrical mean of sensitivity and precision, with a value range of 0 (worst score) to 1 (highest score), where precision is the proportion of correctly identified positive observations to total positive observations (either correctly or incorrectly) [65].

$$\boldsymbol{F1 = 2 * \frac{sensitivity * precision}{sensitivity + precision}}$$

A Receiver Operating Characteristics chart is a graphical representation that illustrates a classification model's performance at all possible decision limits, where the X-axis shows the rate of false-positive (1-specificity) and the Y-axis shows the rate of true-positive (sensitivity) [67]. Theoretically, the Area under Curve varies from 0 to 1, with a classifier that achieves a maximum value of 1 being a perfect classifier. The true lower limit for random classification is 0.5, and classifiers with an AUC greater than 0.5 may differentiate cases from non-cases at least on certain levels [67].

CHAPTER IV. RESULTS PRESENTATION

IV.1 Introduction

This chapter illustrates the research findings which were presented according to the research objectives. The exploratory analysis summarized the background characteristics of the participants. The point biserial and the Pearson chi-square were used to determine the relationship between the target variable and the independent continuous and categorical variables at the significance level of 5%. Furthermore, the machine learning algorithms were used to find out the best model to predict anemia among children 6 -59 months old, and the feature importance was performed to find out the most predictive features that significantly contributed to anemia among children.

IV.2. Background characteristics of participants

The results of the exploratory analysis were based on both descriptive and cross-tabulation analysis and were summarized in Table4. The representative of 3525 children 6 -59 months old (weighted N) was included in this research, and out of those children, 36.79% were anemic. The children aged 6-23 months were the most represented at 35.31% with a high prevalence of 52.77% of anemic children. The sample had roughly similar numbers of males and females, with 50.30 % and 49.70 %, respectively, while the males were more prevalent for anemia with 38.35% than females with 35.21%. The children with the average size at birth were more represented at 52.27% and the anemia prevalence was higher in those who had a small size at birth at 38.51% than among those of large and average size.

During the survey, some children seemed to have infections such as malaria, diarrhea, and fever. The anemia prevalence was higher in children who tested positive for malaria at 75.86% compared to those who tested negative; and higher among children who had diarrhea in the previous two weeks at 46.61% compared to children who did not have diarrhea. The anemia prevalence was also higher in children who had a fever in the previous two weeks, at 44.16% than in children who did not have it.

The majority of participants took vitamin A and drugs for intestinal parasites in the last 6 months from the date of the interview with 85.99%, and 80.30% respectively. The anemia was most prev-

alent among children who did not take drugs for intestinal parasites at 54.38% and also most common among those who did not take vitamin A in the previous 6 months 51.42% compared to those who did. Concerning nutrition, anemia was more common in children who did not eat iron-rich foods the night before the interview, at 42.13 % than among children who did. It was also more common in children who still breastfeed, with 47.27% than in those who had never breastfed or did not. The anemia prevalence was higher in stunted children at 39.19% than in non-stunted children, and it was more prevalent in underweight children at 41.97% than in non-underweight children at 36.36%.

Furthermore, the anemia also varied with the household wealth index in Rwanda. The results revealed that the anemia was higher in children from poor families at 40.08% and lower among children from rich families at 32.72%, and anemia prevalence was higher in children from rural areas at 37.43% compared to the children from urban areas at 33.72%. However, children from the Eastern province were most represented at 26.46% whereas those from the western region had a high prevalence of 40.88% compared to children from other regions. The results revealed that the anemia prevalence varied with the type of toilets used. It was higher in children who used unimproved toilets at 38.77% than children who used improved toilets, and the mean of the children under five years in the family size was 1.65 ± 0.12.

Concerning maternal characteristics, the anemia prevalence was higher at 63.41% in children whose mothers were less than 20 years old, and it decreased as the mother's age increased. The findings also revealed that the anemia prevalence among children 6-59 months old decreased as their mothers' educational levels increased. The anemia prevalence was higher in children whose mothers were unemployed, at 40.08% than in children of employed mothers. The prevalence of anemia was also higher in children whose mothers were anemic, at 47.52% than in those who were not anemic.

IV.3. Findings on factors associated with the child's anemic status

The Bivariate analysis between the outcome variable and the specific characteristics of participants were shown in Table 4 below. The chi-square test with a confidence level of 95% showed that the categorical variables except for the child's sex, the child's birth size, and the source of drinking water were statistically significantly related to anemia in children 6-59 months old. This means

that those variables showed statistical differences between the two classes of anemia status (not anemic and anemic). And the point biserial correlation showed that the number of children under the age of five in the family was not significantly related to anemia status.

Table 4. The distribution of respondent's characteristics and the bivariate analysis of anemia among children in all selected variables

Variables		Frequency	Anemia Status		P-Value
		N (%)	Not Anemic (%)	Anemic (%)	
Age of child	6-23 months	1245(35.31)	588(47.23)	657(52.77)	**0.000**
	24-42 months	1222(34.67)	827(67.68)	395(32.32)	
	43-59 months	1058(30.02)	813(76.84)	245(23.16)	
Sex of the child	Male	1773(50.30)	1093(61.65)	680(38.35)	**0.06**
	Female	1752(49.70)	1135(64.78)	617(35.21)	
Child's birth size	Large	1077(30.56)	687(63.79)	390(36.21)	**0.61**
	Average	1843(52.27)	1170(63.48)	673(36.52)	
	Small	605(17.17)	372(61.49)	233(38.51)	
Child malaria infection	Negative	3496(99.17)	2221(63.53)	1275(36.47)	**0.000**
	Positive	29 (0.83)	7 (24.1)	22 (75.86)	
Presence of fever in last 2 weeks	No	2823(80.09)	1836(65.04)	987(34.96)	**0.000**
	Yes	702(19.91)	392(55.84)	310 (44.16)	
Presence of diarrhea in last 2 weeks	No	2950(83.69)	1921(65.12)	1029(34.88)	**0.000**
	Yes	575(16.31)	307(53.39)	268(46.61)	
Vitamin A in the last 6 months	No	494(14.01)	240(48.58)	254(51.42)	**0.000**
	Yes	3031(85.99)	1988(65.59)	1043(34.41)	
Deworming in last 6 months	No	695(19.70)	317 (45.62)	378(54.38)	**0.000**
	Yes	2830(80.30)	1911(67.53)	919(32.47)	
Iron-rich foods	No	432(12.25)	250(57.87)	182(42.13)	**0.006**
	Yes	3093(87.75)	1978(63.95)	1115(36.05)	
Child breast-feeding status	Not currently breastfeed	2075(58.88)	1460(70.36)	615(29.64)	**0.000**
	never breastfed	22(0.62)	15(68.18)	7(31.82)	
	Still breastfeed	1428(40.50)	753(52.73)	675(47.27)	
Child stunting status	No	2295(65.12)	1480(64.49)	815(35.51)	**0.0197**
	Yes	1230(34.88)	748(60.81)	482(39.19)	

Child underweight	No	3251(92.23)	2069(63.64)	1182(36.36)	**0.05**
	Yes	274(7.77)	159(58.03)	115(41.97)	
Maternal age	<20	41(1.15)	15(36.59)	26(63.41)	**0.0002**
	20-29	1235(35.03)	738(59.76)	497(40.24)	
	30-39	1768(50.18)	1145(64.76)	623(35.23)	
	>=40	481(13.64)	329(68.40)	152(31.60)	
Maternal education	No education	482(11.99)	287(59.54)	195(40.46)	**0.017**
	Primary	2404(65.18)	1527(63.52)	877(36.48)	
	Secondary or Higher	396(22.83)	255(64.39)	141(35.6)	
Maternal anemia	No	3102(87.99)	2006(64.67)	1096(35.33)	**0.000**
	Yes	423(12.01)	222(52.48)	201(47.52)	
Maternal working status	No	853(24.19)	510(59.79)	343(40.21)	**0.0115**
	Yes	2672(75.81)	1718(64.30)	954(35.70)	
Wealth index	poor	1537(43.60)	921(59.92)	616(40.08)	**0.002**
	Middle	689(19.56)	433(62.84)	256(37.15)	
	Rich	1299(36.84)	874(67.28)	425(32.72)	
Source of drinking water	Unimproved	748(21.23)	474(63.37)	274(36.63))	**0.92**
	Improved	2777(78.77)	1754(63.16)	1023(36.84)	
Toilet facility	Unimproved	1024(29.05)	627(61.23)	397(38.77)	**0.014**
	Improved	2501(70.95)	1601(64.01)	900(35.99)	
Type of place of residence	Rural	2923(82.92)	1829(62.57)	1094(37.43)	**0.0199**
	Urban	602(17.08)	399(66.28)	203(33.72)	
Region of residence	Kigali	490(13.91)	314(64.08)	176(35.92)	**0.0073**
	South	713(20.23)	482(67.60)	231(32.40)	
	West	844(23.94)	499(59.12)	345(40.88)	
	North	545(15.46)	317(58.17)	228(41.83)	
	East	933(26.46)	616(66.03)	316(33.87)	
Children under 5 years in a household	-	-	-	-	**0.31**
Total		**3525(100)**	**2228(63.21)**	**1297(36.79)**	

IV.4. Predictive model's performance based on various metrics

IV.4.1. Findings of predictive models

Based on the factors that revealed statistical differences between the two classes of anemia status, the five different machine learning algorithms were utilized to build models and to classify the

anemia state (whether a child was anemic or non-anemic). The Receiver Operating Characteristics (ROC) and the confusion matrix were used to present the prediction results with the performance metrics for every machine learning model on the test data set.

IV.4.1.1. Logistic regression model

The training accuracy of the logistic regression model in anemia prediction for young children 6-59 months old was 65.07%. The confusion matrix was used to calculate the test accuracy, recall, specificity, and F1 score, which revealed that the model's ability to classify all observations accurately was 63.34 %, predicted anemic children correctly at 64.63 %, identified non-anemic children correctly at 62.05 %, and had an F1 score of 63.81 %. For the ROC curve, the AUC was executed and indicated that the model can distinguish the children with and without anemia at 0.68.

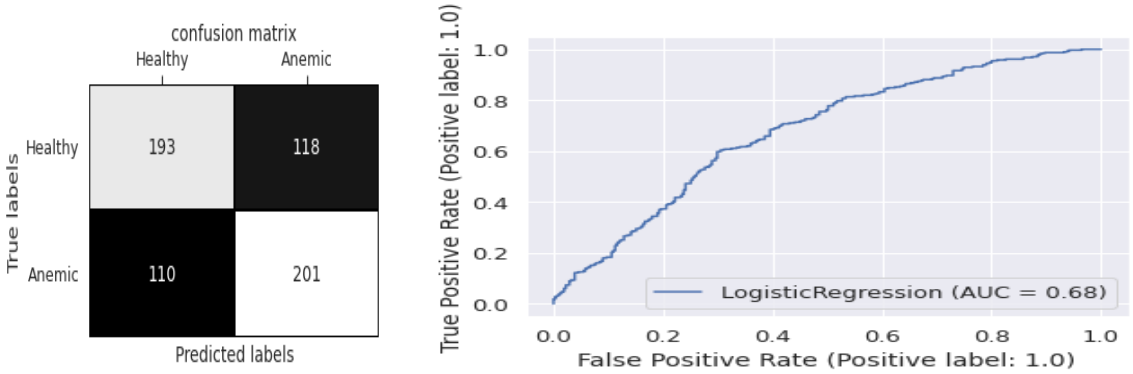


Figure 4. The ROC curve and confusion matrix for the Logistic regression model

IV.4.1.2. Random forest classifier

For the random forest classifier in the anemia prediction, the training accuracy was 81.75%. Based on its confusing matrix, the test accuracy, the sensitivity, the specificity, and the F1_score were executed and indicated that the classifier could show all observations accurately at 81.51%, identified the anemic children correctly at 91.96%, and identified the non_anemic children correctly at 71.06% and F1 score of 83.26%. For the ROC curve, the AUC was executed and indicated that

the model can distinguish the children with anemia and those who don't at 0.91

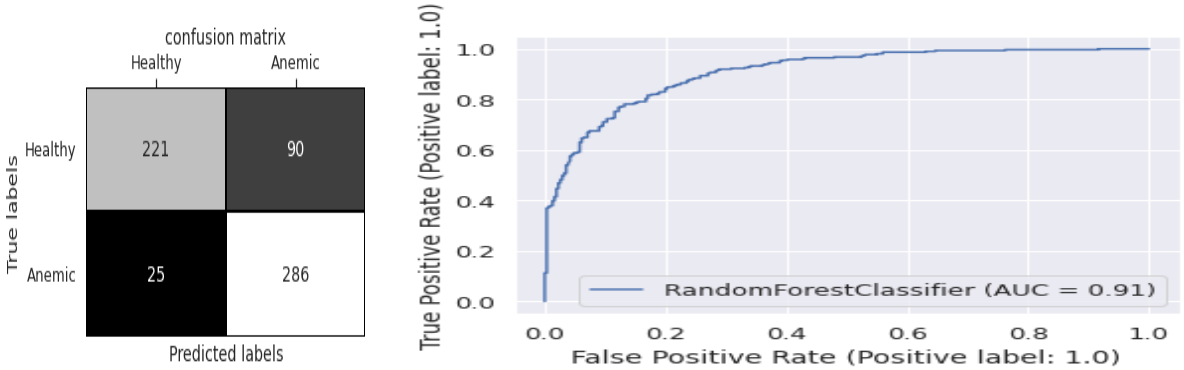


Figure 5. The random forest model's confusion matrix and ROC curve.

IV.4.1.3. Support Vector Machine(SVM)

For the SVM in this study, the training accuracy was 81.99%. Based on its confusing matrix, the test accuracy, recall, specificity, and F1 score were executed and demonstrated that the model's ability to classify all cases accurately was 80.39%, identified the anemic children correctly at 92.60%, and identified the healthy children (non-anemic) correctly at 68.16 % and F1 score of 82.52%. For the ROC curve, the AUC was executed and indicated that the model can distinguish the children with and without anemia at 0.90.

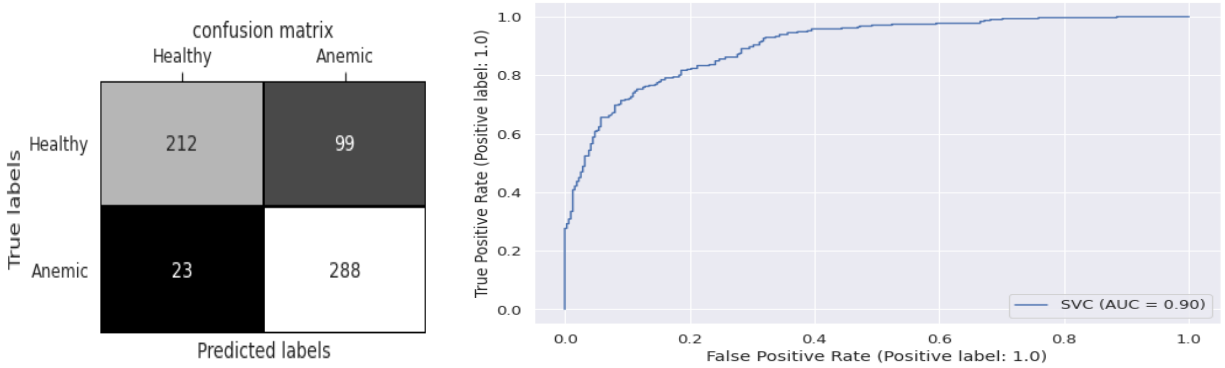


Figure 6. The ROC curve and confusion matrix of the support vector machine

IV.4.1.4. Naïve Bayes

The Naïve Bayes in this research indicated that the training accuracy was 62.23%. Based on its confusing matrix, the test accuracy, recall, specificity, and F1 score were derived and illustrated that the ability of the model to classify all cases accurately was 61.74 %, identified the anemic

children correctly at 52.09%, and identified the healthy children correctly at 71.38% and F1 score of 57.65%. For the ROC curve, the AUC was executed and indicated that the model can distinguish children with and without anemia at 0.66.

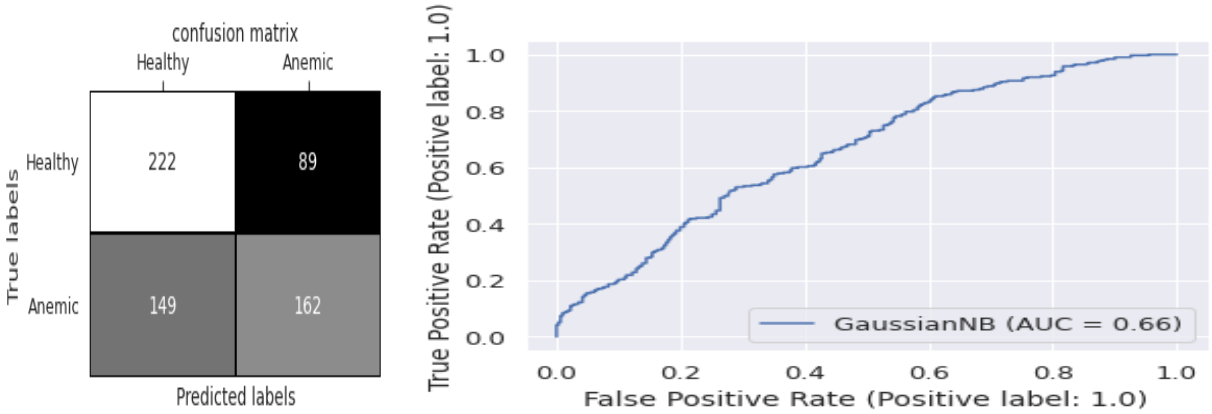


Figure 7. The Naive Bayes model's confusion matrix and ROC curve.

IV.4.1.5. Extreme gradient boosting (XGBoost classifier)

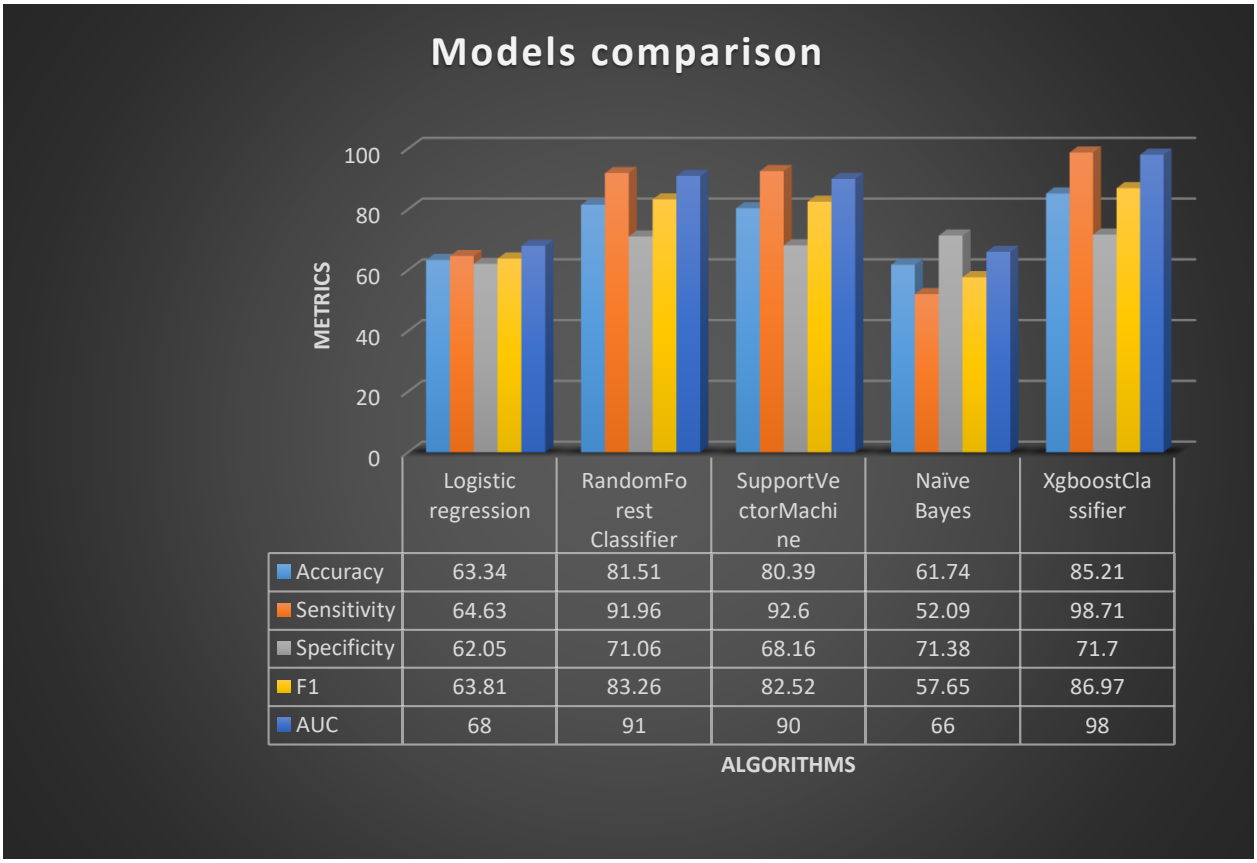
The XGBoost classifier in this study indicated that the accuracy in the training set was 85.71%. According to its confusing matrix, the test accuracy, recall, specificity, and F1 were executed, and the model showed the ability to classify all observations accurately at 85.21%, identify the anemic children correctly at 98.71%, identified the non-anemic children correctly at 71.70% and F1 score of 86.97%. For the ROC curve, the AUC was executed and indicated that the model can discriminate between the children with anemia and those without anemia at 0.98.



Figure 8. The XGBoost classifier's ROC curve and confusion matrix

IV.4.2. The best performance model

The models of different algorithms applied in this research were evaluated using the five previous measures and the best model with the extreme performance metrics was the XGBoost classifier in predicting anemia status among young children aged 6 to 59 months. The results are seen in figure 8 below:



	Logistic regression	RandomForest Classifier	SupportVectorMachine	Naïve Bayes	XgboostClassifier
Accuracy	63.34	81.51	80.39	61.74	85.21
Sensitivity	64.63	91.96	92.6	52.09	98.71
Specificity	62.05	71.06	68.16	71.38	71.7
F1	63.81	83.26	82.52	57.65	86.97
AUC	68	91	90	66	98

Figure 9. Models comparison

IV.5. Findings of the most predictive features

Following the process of modeling, the researcher was interested in identifying the feature importance in anemia prediction among children 6-59 months old. The feature importance rates showed the contribution of each factor in predicting anemia status. The best model of the XGBoost classifier in this study showed that each feature contributed to the anemia prediction, whereby

every category of variables contributed positively or negatively in predicting the risk of being anemic or non_anemic children.

This model figured out that, the child's age, the child malaria infection, maternal anemia, the child breastfeeding status, region of residence, the deworming within 6 months of a child, the recent diarrhea status of a child, and the wealth index for households were the most contributing predictive features in the prediction of anemia status.

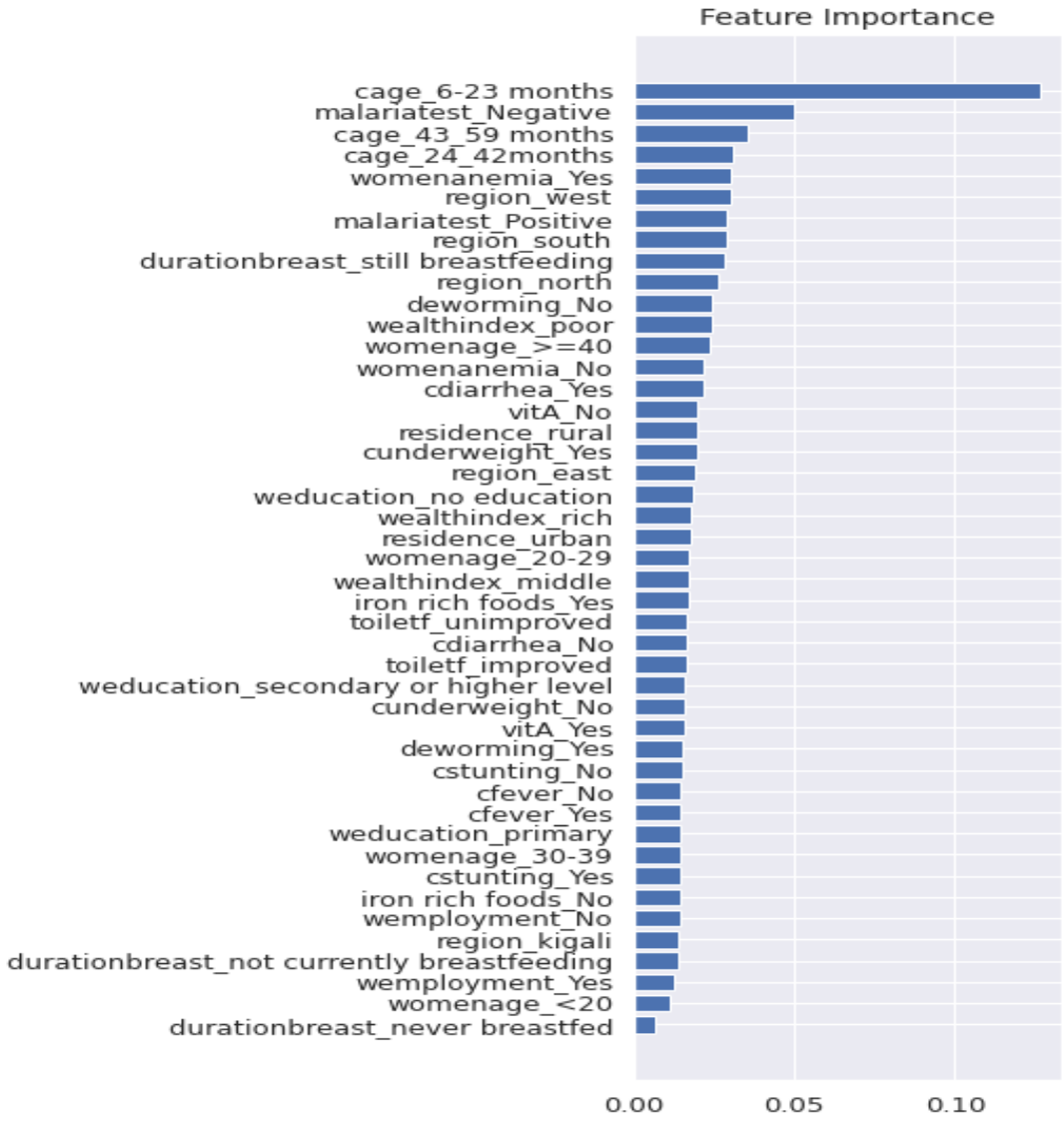


Figure 10. The feature importance from the best model (XGBoost classifier)

Figure 9 above demonstrates the contribution of each factor in the anemia prediction for both anemic and non-anemic children. Some of them influenced much in predicting anemic children other than non-anemic children, and the description of features is shown in the figures below.

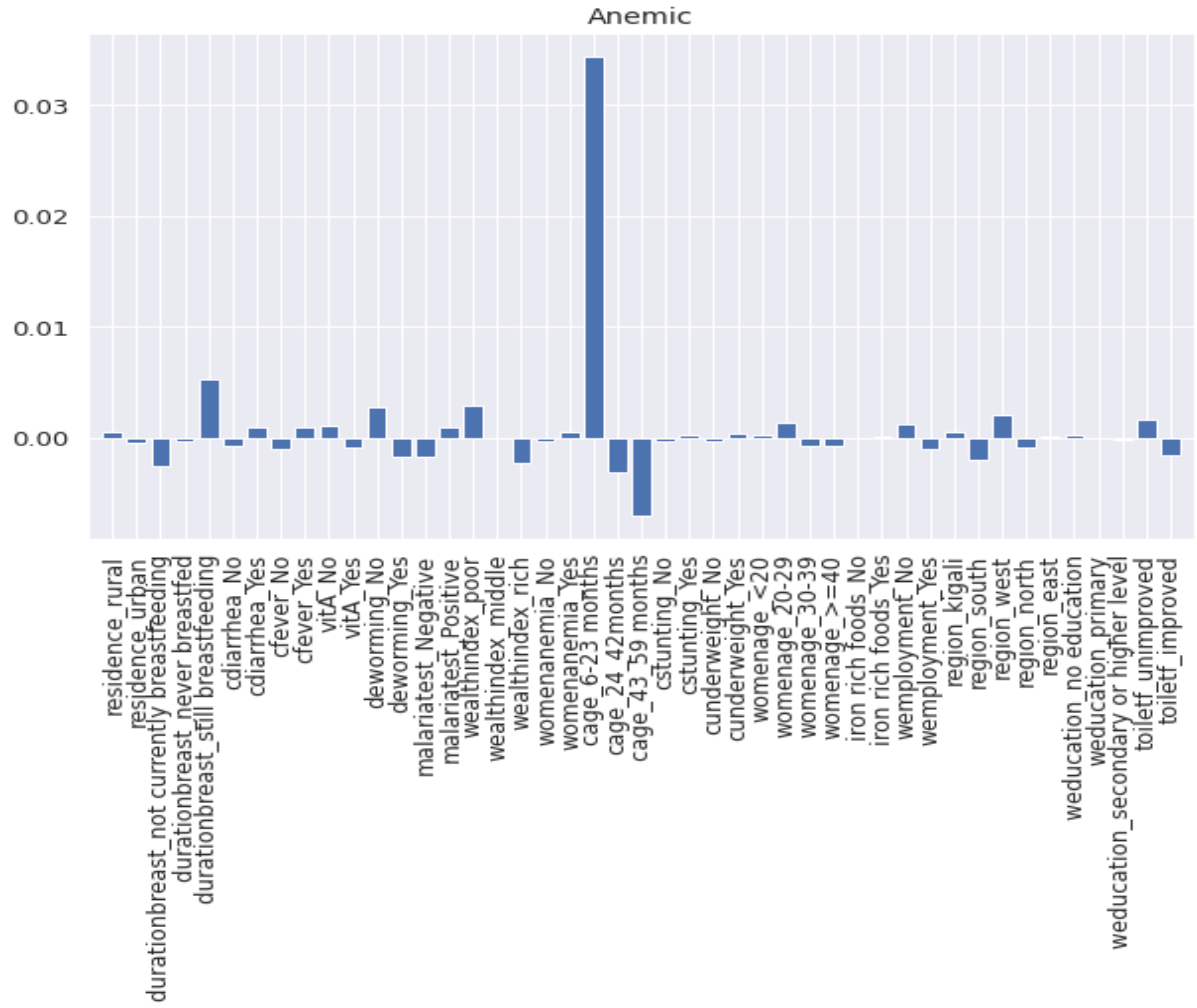


Figure 11. Feature importance from the best model in predicting anemic children

The Figure10 above illustrates the contribution of features in predicting anemic children, whereby the factors like the age of children range between 6-23months, the positive test for malaria infection, the children still breastfeeding, those who did not take the deworming and vitamin A in the last 6 months, the recent fever and diarrhea for a child, the children whose mothers were unemployed, the children living in a poor family, the children from a family used the unimproved toilet and those who live in the western province were the most factors contributing highly in predicting the risk of being anemic in this study.

CHAPTER V. DISCUSSION, CONCLUSION, AND RECOMMENDATIONS

In this chapter, the researcher discussed the findings and came up with the conclusions and recommendations for anemia prediction in Rwandan children 6-59 months old.

V.1. Discussion of the results

This research contributes to the literature concerning anemia prediction in young children 6-59 months old nationally, in Sub-Saharan African countries, and around the globe. The XGboost classifier was found to be the best accurate predictive model in this study with the ability to classify all children correctly at 85.21%, identified the anemic children correctly at 98.71%, and had the highest discriminating performance of 98% to differentiate the anemic and non-anemic children.

Similar studies were conducted, where in the hospital of Jammu (India), Random Forest showed the best prediction accuracy of 67.18% and sensitivity of 69.43% among all the predictive models, whereas, in Bangladesh, the Random Forest algorithm achieved the best classification accuracy of 68.53% with a sensitivity of 70.73% [14] [19]. The XGboost classifier outperforms models used in these studies of Bangladesh and India. This was due to its ability to convert the weak learners to strong learners as an ensemble learning algorithm, overcome overfitting, minimize the training losses, and reduce the classification errors in developing the final model with strong regularization [68].

Furthermore, the findings of this study showed that the children aged 6 to 23 months and children who were still breastfed, contributed greatly to predicting the risk of being anemic. This may be because children at that age are not being breastfed exclusively due to the introduction to the complementary foods, they are unable to consume a more diversified and full diet with adequate iron consumption that could protect against anemia at this age, and they are more prone to infections or diseases that cause poor iron absorption [25]. These findings are consistent with prior research from Bangladesh and Ethiopia [41], [42].

Malaria infection influenced highly the model in predicting the anemic for the individual child. This is similar to the previous studies from different countries that have reported the clear influence

of malaria on anemia among children, and it's because of the increase of hemolysis (destruction of red blood cells) in malaria patients which leads to anemia [7], [11], [25].

The results of this research also revealed that not receiving anti-helminths (drugs of parasites) and vitamin A for children contributed more to the prediction of anemic children, which explained that both deworming and vitamin A for children every six months were the protective factors of anemia. These findings were in line with those of earlier research [11], [43].

As far as maternal factors are concerned, maternal anemia was also contributed to predicting anemic children and this can be driven by the fact that most women and their children experienced the same standard of living i.e. the same socio-economic, environmental, and dietary foods which involved the same risk factors that lead to anemia among children 6-59 months old. Similar results were reported with the studies in Togo, Bangladesh, and Ethiopia respectively [25], [41], [42].

In addition, the children from the lowest wealth index (poor) influenced the model in predicting anemic children and these findings were also consistent with other studies conducted in Kenya [7], Bangladesh [41], Namibia [46], and Ethiopia [8]. Families with the lowest wealth index may lack the means to afford nutrient-rich foods including iron supplements foods and minimum diversity dietary foods which return to the low concentrations of hemoglobin as well as anemia.

This study met some limitations that should be noticed and taken into consideration. As the predicted models were developed using cross-sectional data (DHS), this design did not allow any causal relationships to be established, which is hard to distinguish cause from effect variables. For instance, anemia can be either a cause or an effect of stunting. The health status, clinical and deep variables related to the dietary intake among children and their mothers were unavailable in DHS. Since some attributes were self-reported such as fever and diarrhea in the last 2 weeks, deworming and vitamin A in the last 6 months, and supplements foods last night from the day before the interview, the likelihood of recall bias is high.

There is no explicit selection of algorithms in ML modeling; as they are selected -based on subjective judgment. Despite the highlighted limitations, machine learning algorithms are useful in the prediction of anemia status among children based on socio-demographic and health-related factors due to their advanced computation power.

V.2. Conclusion

Anemia is a severe public health concern that affects people all over the world, and it is currently classified as moderate among Rwandan children 6 -59 months old. The proper intervention and early detection/diagnosis of anemia in children can reduce the risk of anemia. For diagnostic purposes, various clinical tools can be developed to assess the risk of anemia, and the main point of the risk management approach is the accurate prediction of the disease.

Among the numerous models developed using machine learning approaches based on the risk factors, the XGboost model performed well, with the highest discriminating power between anemic and non-anemic children and the highest classification accuracy in the prediction of anemia among Rwandan children. This supports the health care sector to increase the accurate diagnosis and successful disease treatment by focusing on specific factors that may help in improving the prevention of anemia among children in Rwanda. The best prediction model established, showed that child malaria infections, age of a child, maternal anemia, children breastfeeding status, the children underweight, the deworming within 6 months for a child, and the household wealth index are the major predictive features contributing to the anemia prediction among Rwandan children 6-59 months old.

This research revealed that machine learning techniques with the most predictive features can assist in the control and prevention of anemia in Rwandan children 6-59 months old by substantially reducing physical efforts in the screening and facilitating early detection of anemia. However, machine learning algorithms cannot change physician perception and interpretive abilities.

V. 3. Recommendations

The following recommendations are made based on the study's findings:

The health planners and policymakers should design and implement appropriate interventions for eradicating anemia among children based on the most predictive features as the strategies of prevention and control of childhood anemia.

The health practitioners are recommended to use this model in the screening of anemia at an early stage so that effective treatment options can be made to reduce severe anemia cases.

All public health institutions, private and NGOs of health-related interventions should educate the community on the potential risk factors to eradicate childhood anemia as well as conduct regular supervision and follow-up for identified cases of children and mothers at risk.

Further researchers should conduct other related studies or longitudinal studies with clinical and dietary intake data for both mothers and children to increase the performance and the reliability of the model.

REFERENCES

[1] L. L. S. Da Silva, W. W. Fawzi, and M. A. Cardoso, "Factors associated with anemia in young children in Brazil," *PLoS One*, vol. 13, no. 9, pp. 1–12, 2018, doi: 10.1371/journal.pone.0204504.

[2] E. McLean, M. Cogswell, I. Egli, D. Wojdyla, and B. De Benoist, "Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993-2005," *Public Health Nutr.*, vol. 12, no. 4, pp. 444–454, 2009, doi: 10.1017/S1368980008002401.

[3] Who and M. Chan, "Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity," *Geneva, Switz. World Heal. Organ.*, pp. 1–6, 2011, doi: 2011.

[4] M. Melku *et al.*, "Anemia severity among children aged 6-59 months in Gondar town, Ethiopia: A community-based cross-sectional study," *Ital. J. Pediatr.*, vol. 44, no. 1, pp. 1–12, 2018, doi: 10.1186/s13052-018-0547-0.

[5] World Health Organization(WHO), "The global prevalence of anaemia in 2011," *Who*, pp. 1–48, 2011.

[6] W. . K. . Gardner, "The Global Burden o f Anemia. Global, Regional, and National Prevalence of Anemia and Its Causes in 204 Countries and Territories, 1990–2019.," *Hematol. Clin. NA*, vol. 30, no. 2, pp. 247–308, 2020, doi: 10.1016/j.hoc.2015.11.002.

[7] O. Ngesa and H. Mwambi, "Prevalence and risk factors of anaemia among children aged between 6 months and 14 years in Kenya," *PLoS One*, vol. 9, no. 11, 2014, doi: 10.1371/journal.pone.0113756.

[8] H. Woldie, Y. Kebede, and A. Tariku, "Factors Associated with Anemia among Children Aged 6-23 Months Attending Growth Monitoring at Tsitsika Health Center, Wag-Himra Zone, Northeast Ethiopia," *J. Nutr. Metab.*, vol. 2015, 2015, doi: 10.1155/2015/928632.

[9] R. H. Simbauranga, E. Kamugisha, A. Hokororo, B. R. Kidenya, and J. Makani, "Prevalence and factors associated with severe anaemia amongst under-five children hospitalized at Bugando Medical Centre, Mwanza, Tanzania," *BMC Hematol.*, vol. 15, no. 1, pp. 1–9, 2015, doi: 10.1186/s12878-015-0033-5.

[10] N. I. of S. of Rwanda and Kigali, *Rwanda Demographic and Health Survey 2019-20 Final Report*. 2019.

[11] R. Nkulikiyinka, A. Binagwaho, and K. Palmer, "The changing importance of key factors associated with anaemia in 6- to 59-month-old children in a sub-Saharan African setting where malaria is on the decline: Analysis of the Rwanda Demographic and Health Survey 2010," *Trop. Med. Int. Heal.*, vol. 20, no. 12, pp. 1722–1732, 2015, doi: 10.1111/tmi.12604.

[12] National Institute of Statistics of Rwanda (NISR) [Rwanda], Ministry of Health (MOH) [Rwanda], and ICF International, *Rwanda Demographic and Health Survey 2014-2015*. 2015.

[13] A. Agrawal, H. Agrawal, S. Mittal, and M. Sharma, "Disease Prediction Using Machine Learning," *SSRN Electron. J.*, no. May, pp. 6937–6938, 2018, doi: 10.2139/ssrn.3167431.

[14] J. R. Khan, S. Chowdhury, H. Islam, and E. Raheem, "Machine Learning Algorithms To Predict the Childhood Anemia in Bangladesh.," *J. Data Sci.*, vol. 17, no. 1, pp. 195–217, 2019, doi: 10.6339/JDS.201901.

[15] A. Talukder and B. Ahammed, "Machine learning algorithms for predicting malnutrition

among under-five children in Bangladesh," *Nutrition*, vol. 78, 2020, doi: 10.1016/j.nut.2020.110861.
[16] J. Mulambu, "Iron beans in Rwanda: Crop development and delivery experience," *African J. Food, Agric. Nutr. Dev.*, vol. 17, no. 02, pp. 12026–12050, 2017, doi: 10.18697/ajfand.78.harvestplus10.
[17] N. C. Paper, "Unité – Egalité – Paix Comprehensive Africa Agriculture Development Programme (CAADP) East and Central Africa Regional CAADP Nutrition Program Development Workshop Nutrition Country Paper – République de Djibouti DRAFT – VERSION FRANCAISE Février 2013," 2013.
[18] M. Alghamdi, M. Al-Mallah, S. Keteyian, C. Brawner, J. Ehrman, and S. Sakr, "Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford ExercIse Testing (FIT) project," *PLoS One*, vol. 12, no. 7, pp. 1–15, 2017, doi: 10.1371/journal.pone.0179805.
[19] P. Anand, R. Gupta, and A. Sharma, "Prediction of Anaemia among children using Machine Learning Algorithms," no. June, pp. 469–480, 2020.
[20] M. Donahue Angel, P. Berti, K. Siekmans, P. L. Tugirimana, and E. Boy, "Prevalence of Iron Deficiency and Iron Deficiency Anemia in the Northern and Southern Provinces of Rwanda," *Food Nutr. Bull.*, vol. 38, no. 4, pp. 554–563, 2017, doi: 10.1177/0379572117723134.
[21] L. Panawala, "What is the Function of Hemoglobin in the Human Body," *Pediaa*, no. February, pp. 1–7, 2017.
[22] P. Eastwell, "Understanding Hypotheses, Predictions, Laws, and Theories.," *Sci. Educ. Rev.*, vol. 13, no. 1, pp. 16–21, 2014.
[23] G. Shmueli, "To explain or to predict?," *Stat. Sci.*, vol. 25, no. 3, pp. 289–310, 2010, doi: 10.1214/10-STS330.
[24] R. Shouval, O. Bondi, H. Mishan, A. Shimoni, R. Unger, and A. Nagler, "Application of machine learning algorithms for clinical predictive modeling: A data-mining approach in SCT," *Bone Marrow Transplant.*, vol. 49, no. 3, pp. 332–337, 2014, doi: 10.1038/bmt.2013.146.
[25] A. Nambiema, A. Robert, and I. Yaya, "Prevalence and risk factors of anemia in children aged from 6 to," *BMC Public Health*, vol. 19, no. 1, pp. 1–10, 2019, doi: 10.1186/s12889-019-6547-1.
[26] G. A. Stevens *et al.*, "Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995-2011: A systematic analysis of population-representative data," *Lancet Glob. Heal.*, vol. 1Prevalenc, no. 1, pp. 16–25, 2013, doi: 10.1016/S2214-109X(13)70001-9.
[27] Y. Balarajan, U. Ramakrishnan, E. Özaltin, A. H. Shankar, and S. V. Subramanian, "Anaemia in low-income and middle-income countries," *Lancet*, vol. 378, no. 9809, pp. 2123–2135, 2011, doi: 10.1016/S0140-6736(10)62304-5.
[28] D. C. R. Sastry, "Study on clinical and hematological profile of Anemia in children aged 5 t0 12 years in rural Telangana," *Pediatr. Rev. Int. J. Pediatr. Res.*, vol. 4, no. 7, pp. 488–493, 2017, doi: 10.17511/ijpr.2017.i07.09.
[29] World Health Organization, *Nutritional Anaemias : Tools for Effective Prevention*. 2017.
[30] S. M. Fishman, P. Christian, and J. West, "The role of vitamins in the prevention and control anaemia," *Public Health Nutr.*, vol. 3, no. 2, pp. 125–150, 2000, doi:

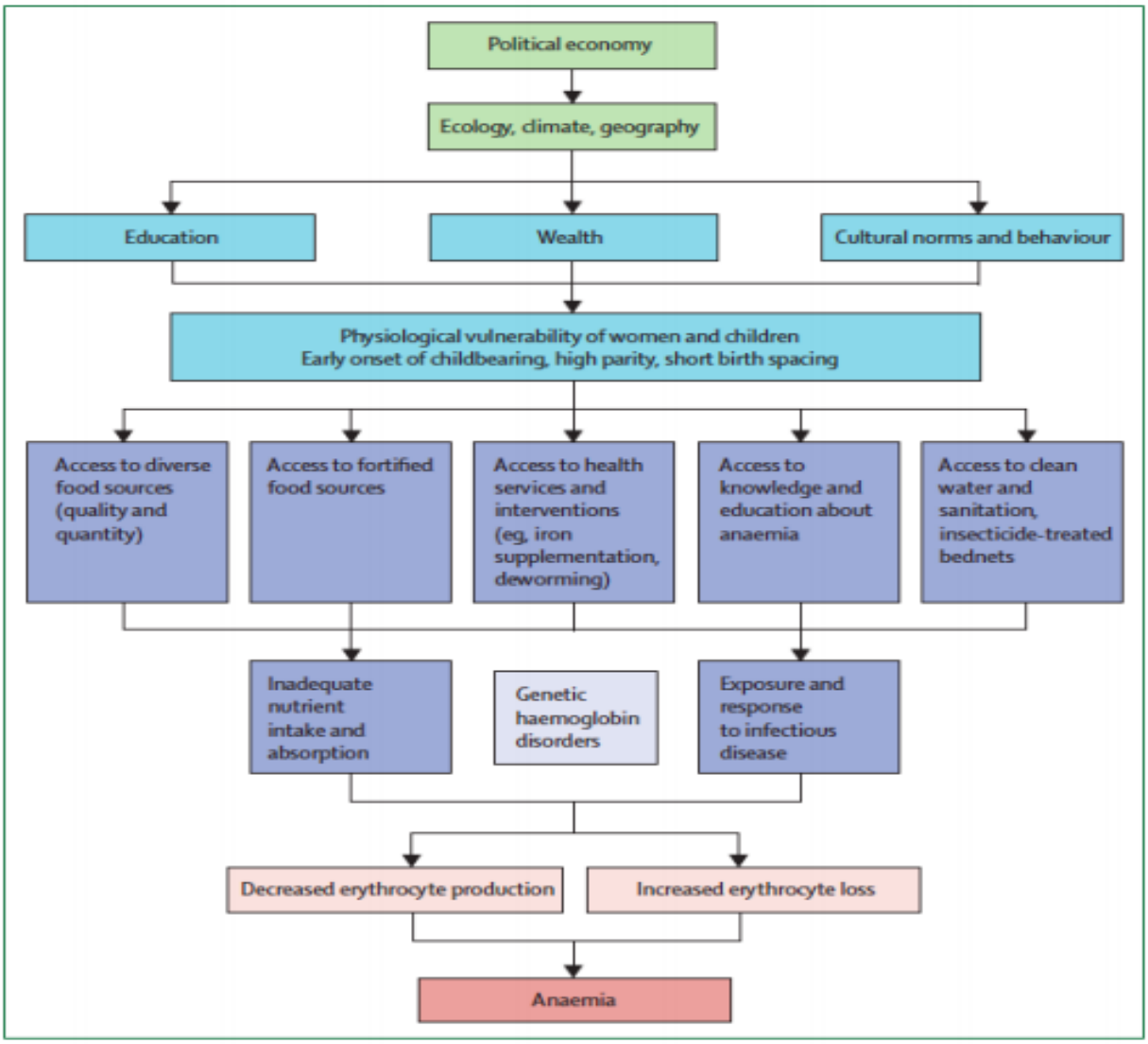
10.1017/s1368980000000173.
[31] R. D. Semba and M. W. Bloem, “The anemia of vitamin a deficiency: Epidemiology and pathogenesis,” *Eur. J. Clin. Nutr.*, vol. 56, no. 4, pp. 271–281, 2002, doi: 10.1038/sj.ejcn.1601320.
[32] B. H. Toh, I. R. Van Driel, and P. A. Gleeson, “Mechanisms of disease: Pernicious anemia,” *N. Engl. J. Med.*, vol. 337, no. 20, pp. 1441–1448, 1997, doi: 10.1056/NEJM199711133372007.
[33] R. Green *et al.*, “Vitamin B12 deficiency,” *Nat. Rev. Dis. Prim.*, vol. 3, no. June, pp. 1–19, 2017, doi: 10.1038/nrdp.2017.40.
[34] R. J. S. Magalhães and A. C. A. Clements, “Mapping the risk of anaemia in preschool-age children: The contribution of malnutrition, malaria, and helminth infections in West Africa,” *PLoS Med.*, vol. 8, no. 6, 2011, doi: 10.1371/journal.pmed.1000438.
[35] A. J. Redig and N. Berliner, “Pathogenesis and clinical implications of HIV-related anemia in 2013.,” *Hematology Am. Soc. Hematol. Educ. Program*, vol. 2013, pp. 377–381, 2013, doi: 10.1182/asheducation-2013.1.377.
[36] J. Bethony *et al.*, “Soil-transmitted helminth infections : ascariasis , trichuriasis , and hookworm,” vol. 367, 2006.
[37] D. J. Weatherall, “The inherited diseases of hemoglobin are an emerging global health burden,” *Blood*, vol. 115, no. 22, pp. 4331–4336, 2010, doi: 10.1182/blood-2010-01-251348.
[38] L. M. Shenton, A. D. Jones, and M. L. Wilson, “Factors Associated with Anemia Status Among Children Aged 6–59 months in Ghana, 2003–2014,” *Matern. Child Health J.*, vol. 24, no. 4, pp. 483–502, 2020, doi: 10.1007/s10995-019-02865-7.
[39] H. Henriksson, D. Project, and I. Maternal, “P revalence of anemia and its association with socio-demographic factors and micronutrient deficiencies in 4 . 5-year old children in Matlab , Bangladesh : a cross-sectional follow-up study Secondary analysis of data from the MINIMat randomized trial,” no. May, 2015.
[40] Q. Li, F. Liang, W. Liang, W. Shi, and Y. Han, “Prevalence of anemia and its associated risk factors among 6-months-old infants in Beijing,” *Front. Pediatr.*, vol. 7, no. JULY, pp. 1–7, 2019, doi: 10.3389/fped.2019.00286.
[41] J. R. Khan, N. Awan, and F. Misu, “Determinants of anemia among 6-59 months aged children in Bangladesh: Evidence from nationally representative data,” *BMC Pediatr.*, vol. 16, no. 1, pp. 1–12, 2016, doi: 10.1186/s12887-015-0536-z.
[42] K. F. Muchie, “Determinants of severity levels of anemia among children aged 6-59 months in Ethiopia: Further analysis of the 2011 Ethiopian demographic and health survey,” *BMC Nutr.*, vol. 2, no. 1, pp. 1–8, 2016, doi: 10.1186/s40795-016-0093-3.
[43] A. Belachew and T. Tewabe, “Under-five anemia and its associated factors with dietary diversity, food security, stunted, and deworming in Ethiopia: Systematic review and meta-analysis,” *Syst. Rev.*, vol. 9, no. 1, pp. 10–12, 2020, doi: 10.1186/s13643-020-01289-7.
[44] R. M. L. Semedo, M. M. A. S. Santos, M. R. Baião, R. R. Luiz, and G. V. Da Veiga, “Prevalence of Anaemia and Associated Factors among Children below Five Years of Age in Cape Verde, West Africa,” *J. Heal. Popul. Nutr.*, vol. 32, no. 4, pp. 646–657, 2014, doi: 10.3329/jhpn.v32i4.3021.
[45] G. Gebreegziabiher, B. Etana, and D. Niggusie, “Determinants of Anemia among Children Aged 6 – 59 Months Living in Kilte Awulaelo Woreda , Northern Ethiopia,” vol. 2014, 2014.

[46] P. P. Shimanda, H. J. Amukugo, and F. Norström, "Socioeconomic factors associated with anaemia among children aged 6–59 months in Namibia," *J. Public Health Africa*, vol. 11, no. 1, pp. 20–24, 2020, doi: 10.4081/jphia.2020.1131.
[47] World Health Organization(WHO), *Iron Deficiency Anaemia*. 2001.
[48] C. R. K. Zuffo, M. M. Osório, C. A. Taconeli, S. T. Schmidt, B. H. C. da Silva, and C. C. B. Almeida, "Prevalence and risk factors of anemia in children," *J. Pediatr. (Rio. J).*, vol. 92, no. 4, pp. 353–360, 2016, doi: 10.1016/j.jped.2015.09.007.
[49] P. A. Parbey *et al.*, "Risk Factors of Anaemia among Children under Five Years in the Hohoe Municipality, Ghana: A Case Control Study," *Anemia*, vol. 2019, 2019, doi: 10.1155/2019/2139717.
[50] I. Ngnie-teta, O. Receveur, and B. Kuate-defo, "Risk factors for moderate to severe anemia among children in Benin and Mali : Insights from a multilevel analysis," vol. 28, no. 1, pp. 76–89, 2007.
[51] B. T., S. S., H. S., and M. K.S., "Urban food insecurity in the context of high food prices: a community based cross sectional study in Addis Ababa, Ethiopia," *BMC Public Health*, vol. 14, p. 680, 2014.
[52] P. P. Moschovis *et al.*, "Individual, maternal and household risk factors for anaemia among young children in sub-Saharan Africa: A cross-sectional study," *BMJ Open*, vol. 8, no. 5, pp. 1–14, 2018, doi: 10.1136/bmjopen-2017-019654.
[53] M. Jaiswal, A. Srivastava, and T. J. Siddiqui, "Machine learning algorithms for anemia disease prediction," *Lect. Notes Electr. Eng.*, vol. 524, pp. 463–469, 2019, doi: 10.1007/978-981-13-2685-1_44.
[54] S. O. Rutstein and G. Rojas, "Guide to DHS statistics," *Calverton, MD ORC Macro*, no. September, pp. 1–161, 2018.
[55] B. Malley, D. Ramazzotti, and J. T. Wu, "Secondary Analysis of Electronic Health Records," *Second. Anal. Electron. Heal. Rec.*, pp. 1–427, 2016, doi: 10.1007/978-3-319-43742-2.
[56] P. Jönsson and C. Wohlin, "An evaluation of k-nearest neighbour imputation using lIkert data," *Proc. - Int. Softw. Metrics Symp.*, pp. 108–118, 2004, doi: 10.1109/METRIC.2004.1357895.
[57] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. February 2017, pp. 321–357, 2002, doi: 10.1613/jair.953.
[58] D. Berrar, "Cross-validation," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, pp. 542–545, 2018, doi: 10.1016/B978-0-12-809633-8.20349-X.
[59] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, "Machine learning in medicine: a practical introduction," *BMC Med. Res. Methodol.*, vol. 19, no. 1, pp. 1–18, 2019, doi: 10.1186/s12874-019-0681-4.
[60] M. R. Abonazel, M. Gamal Ghallab, M. Reda Abonazel, and M. Gamal Ibrahim, "On Estimation Methods for Binary Logistic Regression Model with Missing Values," *Int. J. Math. Comput. Sci.*, vol. 4, no. 3, pp. 79–85, 2018.
[61] P. Golpour *et al.*, "Comparison of support vector machine, naïve bayes and logistic regression for assessing the necessity for coronary angiography," *Int. J. Environ. Res. Public Health*, vol. 17, no. 18, pp. 1–9, 2020, doi: 10.3390/ijerph17186449.
[62] H. A. Le Thi, V. V. Nguyen, and S. Ouchani, "Gene selection for cancer classification using DCA," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect.*

Notes Bioinformatics), vol. 5139 LNAI, pp. 62–72, 2008, doi: 10.1007/978-3-540-88192-6_8.
[63] S. Uddin, A. Khan, E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," vol. 8, pp. 1–16, 2019.
[64] S. Malik and R. Harode, "XGBoost : A Deep Dive into Boosting," no. February, 2020, doi: 10.13140/RG.2.2.15243.64803.
[65] F. R. Lumbanraja, E. Fitri, Ardiansyah, A. Junaidi, and R. Prabowo, "Abstract Classification Using Support Vector Machine Algorithm (Case Study: Abstract in a Computer Science Journal)," *J. Phys. Conf. Ser.*, vol. 1751, no. 1, 2021, doi: 10.1088/1742-6596/1751/1/012042.
[66] K. M. Ting, "Confusion Matrix," *Encycl. Mach. Learn. Data Min.*, no. October, pp. 260–260, 2017, doi: 10.1007/978-1-4899-7687-1_50.
[67] N. Lachiche, "Receiver Operating Characteristic (ROC) Analysis," *Encycl. Data Warehous. Mining, Second Ed.*, vol. 5, no. 3, pp. 1675–1681, 2017, doi: 10.4018/978-1-60566-010-3.ch255.
[68] R. Wieland, T. Lakes, and C. Nendel, "Using SHAP to interpret XGBoost predictions of grassland degradation in Xilingol, China," *Geosci. Model Dev. Discuss.*, no. June, pp. 1–28, 2020, doi: 10.5194/gmd-2020-59.

APPENDIX

Appendix 1. Conceptual framework Illustrating determinants of anemia by Balarajan et al.[24]



Appendix2. Authorization letter for DHS data

Mar 08, 2021

Diane UWIZERA
university of Rwanda
Rwanda
Phone: +250785520334
Email: uwizeradiane112015@gmail.com
Request Date: 03/06/2021

Dear Diane UWIZERA:

This is to confirm that you are approved to use the following Survey Datasets for your registered research paper titled: "PREDICTION OF ANEMIA AMONG CHILDREN AGED 6-59 MONTHS IN RWANDA ":

Rwanda

To access the datasets, please login at: https://www.dhsprogram.com/data/dataset_admin/login_main.cfm. The user name is the registered email address, and the password is the one selected during registration.

The IRB-approved procedures for DHS public-use datasets do not in any way allow respondents, households, or sample communities to be identified. There are no names of individuals or household addresses in the data files. The geographic identifiers only go down to the regional level (where regions are typically very large geographical areas encompassing several states/provinces). Each enumeration area (Primary Sampling Unit) has a PSU number in the data file, but the PSU numbers do not have any labels to indicate their names or locations. In surveys that collect GIS coordinates in the field, the coordinates are only for the enumeration area (EA) as a whole, and not for individual households, and the measured coordinates are randomly displaced within a large geographic area so that specific enumeration areas cannot be identified.

The DHS Data may be used only for the purpose of statistical reporting and analysis, and only for your registered research. To use the data for another purpose, a new research project must be registered. All DHS data should be treated as confidential, and no effort should be made to identify any household or individual respondent interviewed in the survey. Please reference the complete terms of use at: https://dhsprogram.com/Data/terms-of-use.cfm.

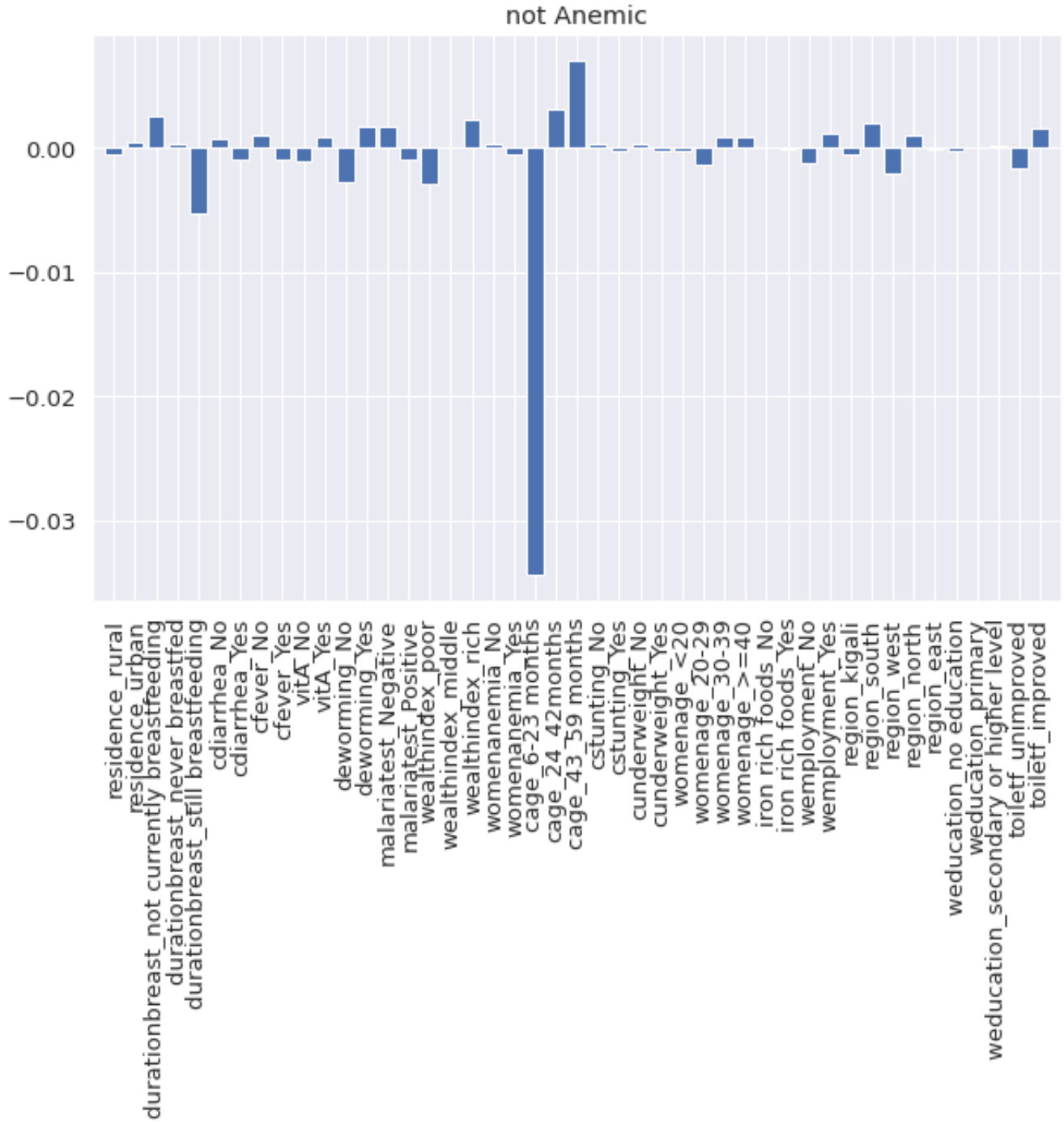
The data must not be passed on to other researchers without the written consent of DHS. However, if you have coresearchers registered in your account for this research paper, you are authorized to share the data with them. All data users are required to submit an electronic copy (pdf) of any reports/publications resulting from using the DHS data files to: references@dhsprogram.com.

Sincerely,

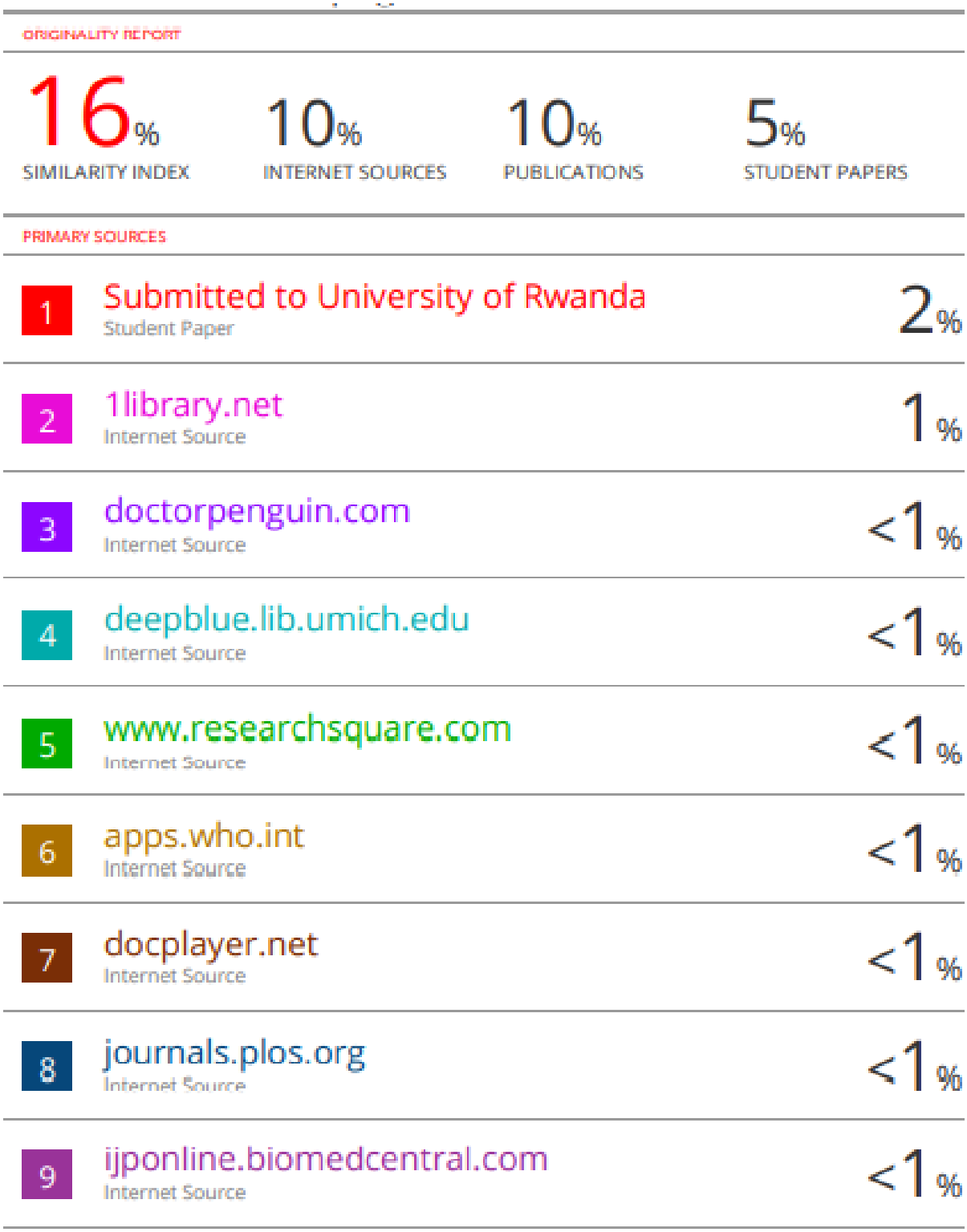
Bridgette Wellington

Bridgette Wellington
Data Archivist
The Demographic and Health Surveys (DHS) Program

Appendix3.Features importance from the best model in predicting the non_anemic children



Appendix 4. Detection of plagiarism: ANEMIA PREDICTION AMONG RWANDAN CHILDREN AGED 6 TO 59 MONTHS USING MACHINE LEARNING TECHNIQUES



ORIGINALITY REPORT

16%	10%	10%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Rwanda Student Paper	2%
2	1library.net Internet Source	1%
3	doctorpenguin.com Internet Source	<1%
4	deepblue.lib.umich.edu Internet Source	<1%
5	www.researchsquare.com Internet Source	<1%
6	apps.who.int Internet Source	<1%
7	docplayer.net Internet Source	<1%
8	journals.plos.org Internet Source	<1%
9	ijponline.biomedcentral.com Internet Source	<1%

10	lume.ufrgs.br Internet Source	<1 %
11	Submitted to University of Arizona Student Paper	<1 %
12	repository.kemu.ac.ke:8080 Internet Source	<1 %
13	www.measuredhs.com Internet Source	<1 %
14	Submitted to University of KwaZulu-Natal Student Paper	<1 %
15	Fabian P Mghanga, Christopher M Genge, Leonia Yeyeye, Zainab Twalib, Wilfred Kibopile, Fredrick J Rutalemba, Tito M Shengena. "Magnitude, Severity, and Morphological Types of Anemia in Hospitalized Children Under the Age of Five in Southern Tanzania", Cureus, 2017 Publication	<1 %
16	Rebecca A. Heidkamp, Ismael Ngnie-Teta, Mohamed Ag Ayoya, Rebecca J. Stoltzfus et al. "Predictors of Anemia among Haitian Children Aged 6 to 59 Months and Women of Childbearing Age and Their Implications for Programming", Food and Nutrition Bulletin, 2013 Publication	<1 %

17	irgu.unigoa.ac.in Internet Source	<1 %
18	Amina Z Khambalia, Ashley M Aimone, Stanley H Zlotkin. "Burden of anemia among indigenous populations", Nutrition Reviews, 2011 Publication	<1 %
19	Jahidur Rahman Khan, Nabil Awan, Farjana Misu. "Determinants of anemia among 6–59 months aged children in Bangladesh: evidence from nationally representative data", BMC Pediatrics, 2016 Publication	<1 %
20	Submitted to Liverpool John Moores University Student Paper	<1 %
21	erepository.uonbi.ac.ke Internet Source	<1 %
22	Submitted to University of Limerick Student Paper	<1 %
23	bmcpediatr.biomedcentral.com Internet Source	<1 %
24	etd.aau.edu.et Internet Source	<1 %
25	etd.astu.edu.et Internet Source	<1 %

26	vtechworks.lib.vt.edu Internet Source	<1 %
27	arno.uvt.nl Internet Source	<1 %
28	link.springer.com Internet Source	<1 %
29	mts.intechopen.com Internet Source	<1 %
30	Mahama Saaka, Sylvester Zakaria Galaa. "How is dietary diversity related to haematological status of preschool children in Ghana?", Food & Nutrition Research, 2017 Publication	<1 %
31	Manal Alghamdi, Mouaz Al-Mallah, Steven Keteyian, Clinton Brawner, Jonathan Ehrman, Sherif Sakr. "Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project", PLOS ONE, 2017 Publication	<1 %
32	Monica M Pasqualino, Andrew L Thorne-Lyman, Swetha Manohar, Angela KC et al. "The Risk Factors for Child Anemia Are Consistent across 3 National Surveys in Nepal", Current Developments in Nutrition, 2021 Publication	<1 %

33	hdl.handle.net Internet Source	<1 %
34	open.library.ubc.ca Internet Source	<1 %
35	repository.president.ac.id Internet Source	<1 %
36	Sohani Afroja, Md Rasel Kabir, Md Akhtarul Islam. "Analysis of determinants of severity levels of childhood anemia in Bangladesh using a proportional odds model", Clinical Epidemiology and Global Health, 2020 Publication	<1 %
37	escholarship.org Internet Source	<1 %
38	www.memoireonline.com Internet Source	<1 %
39	Submitted to British University In Dubai Student Paper	<1 %
40	Kamalesh Kumar Patel, Jyoti Vijay, Abha Mangal, Daya Krishan Mangal, Shiv Dutt Gupta. "Burden of anaemia among children aged 6–59 months and its associated risk factors in India – Are there gender differences?", Children and Youth Services Review, 2021 Publication	<1 %

41	Mohamed Ag Ayoya, Ismael Ngnie-Teta, Marie Nancy Séraphin, Aissa Mamadoultaibou et al. "Prevalence and Risk Factors of Anemia among Children 6–59 Months Old in Haiti", Anemia, 2013 Publication	<1 %
42	scholarworks.waldenu.edu Internet Source	<1 %
43	Biruk Beletew Abate, Ayelign kassie, Mohammed Ahmed. "Determinants of anemia among HIV positive children on Highly Active Antiretroviral Therapy (HAART) attending hospitals of North Wollo Zone, Amhara Region, Ethiopia, 2019: A case-control study", Research Square, 2019 Publication	<1 %
44	Dia Sanou, Ismael Ngnie-Tet. "Chapter 11 Risk Factors for Anemia in Preschool Children in Sub-Saharan Africa", IntechOpen, 2012 Publication	<1 %
45	Submitted to Eiffel Corporation Student Paper	<1 %
46	Jeetendra Yadav, Nilima Nilima. "Geographic variation and factors associated with anemia among under-fives in India: A multilevel approach", Clinical Epidemiology and Global Health, 2021 Publication	<1 %

47	pubmed.ncbi.nlm.nih.gov Internet Source	<1 %
48	Submitted to College of Banking and Financial Studies Student Paper	<1 %
49	Submitted to ESC Rennes Student Paper	<1 %
50	Sengtavanh Keokenchanh, Sengchanh Kounnavong, Kaoru Midorikawa, Wakaha Ikeda, Akemi Morita, Takumi Kitajima, Shigeru Sokejima. "Prevalence of anemia and its associated factors among children aged 6–59 months in the Lao People's Democratic Republic: A multilevel analysis", PLOS ONE, 2021 Publication	<1 %
51	imanagerpublications.com Internet Source	<1 %
52	mjm.usm.my Internet Source	<1 %
53	Ashenafi Abate Woya, Abay Kassa Tekile. "Prevalence and associated factors of anemia among children aged 6-59 months in Ethiopia: Evidence from the Ethiopia Demographic and Health Survey", Research Square, 2020 Publication	<1 %

54	Edouard Tuaillon. "Close association of $CD8^+/CD38^{bright}$ with HIV-1 replication and complex relationship with $CD4^+$ T-cell count", Cytometry Part B Clinical Cytometry, 2008 Publication	<1 %
55	Kassandra L. Harding, Victor M. Aguayo, Grace Namirembe, Patrick Webb. "Determinants of anemia among women and children in Nepal and Pakistan: An analysis of recent national survey data", Maternal & Child Nutrition, 2017 Publication	<1 %
56	Submitted to Malayan Colleges Mindanao, A Mapua School Student Paper	<1 %
57	Richard Nkulikiyinka, Agnes Binagwaho, Katie Palmer. "The changing importance of key factors associated with anaemia in 6 - to 59 - month - old children in a sub - Saharan African setting where malaria is on the decline: analysis of the Rwanda Demographic and Health Survey 2010", Tropical Medicine & International Health, 2015 Publication	<1 %
58	Submitted to University of New South Wales Student Paper	<1 %
59	erepository.uonbi.ac.ke:8080	

	Internet Source	<1 %
60	researchspace.ukzn.ac.za Internet Source	<1 %
61	assets.researchsquare.com Internet Source	<1 %
62	vuir.vu.edu.au Internet Source	<1 %
63	www.researchgate.net Internet Source	<1 %
64	www.scielo.br Internet Source	<1 %
65	A. Zemirline, L. Lecornu, B. Solaiman. "Data Mining System applied to Endoscopic Image Base", 2006 2nd International Conference on Information & Communication Technologies, 2006 Publication	<1 %
66	Messina, Jane P., Kashamuka Mwandagalirwa, Steve M. Taylor, Michael Emch, and Steven R. Meshnick. "Spatial and social factors drive anemia in Congolese women", Health & Place, 2013. Publication	<1 %
67	Mohd Nazrul Effendy Mohd Idrus, Kim Seng Chia. "Partial Least Square with Savitzky Golay	<1 %

Derivative in Predicting Blood Hemoglobin Using Near Infrared Spectrum", MATEC Web of Conferences, 2018
Publication

68	Narmeen Jamal Al-Awwad, Jennifer Ayoub, Rawhieh Barham, Wafaa Sarhan et al. "Review of the Nutrition Situation in Jordan: Trends and Way Forward", Nutrients, 2021 Publication	<1 %
69	Peter Austin Morton Ntenda, Kun-Yang Chuang, Fentanesh Nibret Tiruneh, Ying-Chih Chuang. "Multilevel Analysis of the Effects of Individual- and Community-Level Factors on Childhood Anemia, Severe Anemia, and Hemoglobin Concentration in Malawi", Journal of Tropical Pediatrics, 2018 Publication	<1 %
70	annalsofrscb.ro Internet Source	<1 %
71	es.scribd.com Internet Source	<1 %
72	mdpi.com Internet Source	<1 %
73	pv.glenbrook.com Internet Source	<1 %
74	www.iosrjournals.org Internet Source	<1 %

75	www.thefreelibrary.com Internet Source	<1 %
76	www.unicef.org Internet Source	<1 %
77	Alinoor Mohamed Farah, Tahir Yousuf Nour, Bilal Shikur Endris, Seifu Hagos Gebreyesus. "Concurrent of stunting and overweight/obesity among children: evidence from Ethiopia", Cold Spring Harbor Laboratory, 2020 Publication	<1 %
78	Angesom Gebreweld, Neima Ali, Radiya Ali, Temesgen Fisha. "Prevalence of anemia and its associated factors among children under five years of age attending at Guguftu health center, South Wollo, Northeast Ethiopia", PLOS ONE, 2019 Publication	<1 %
79	Belayneh E Dejene, Tesfamariam M Abuhay, Dawit S Bogale. "Predicting the Level of Anemia among Ethiopian Pregnant Women using Homogeneous Ensemble Machine Learning Algorithm", Research Square Platform LLC, 2022 Publication	<1 %
80	Dickson A. Amugsi. "Determinants of normal haemoglobin concentration among children in	<1 %

Ghana: a positive deviance analysis of nationally representative cross-sectional survey data", Scientific Reports, 2020
Publication

---

81 Gebremedhin Gebreegziabiher, Belachew Etana, Daniel Niggusie. "Determinants of Anemia among Children Aged 6–59 Months Living in Kilte Awulaelo Woreda, Northern Ethiopia", Anemia, 2014 <1 %
Publication

---

82 James Nyirenda, Gershom B. Lwanga, Kaampwe M. Muzandu, David K. Chuba, Gibson M. Sijumbila. "In Vivo Anti-Anemic Effect of an Aqueous Root Extract of Phyllanthus Muellerianus (Kuntze) Exell in Model Rats.", Research Square Platform LLC, 2021 <1 %
Publication

---

83 Jeanine Ahishakiye, Laura Bouwman, Inge D. Brouwer, Eric Matsiko, Margaret Armar-Klemesu, Maria Koelen. "Challenges and responses to infant and young child feeding in rural Rwanda: a qualitative study", Journal of Health, Population and Nutrition, 2019 <1 %
Publication

---

84 Negasa Eshete Soboksa, Sirak Robele Gari, Abebe Beyene Hailu, Bezatu Mengistie Alemu. "Association of childhood undernutrition with <1 %

water supply, sanitation and hygiene interventions in Kersa and Omo Nada districts of Jimma Zone, Ethiopia: A case-control study", Research Square, 2020
Publication

85	Nuzhat Choudhury, Towfida Jahan Siddiqua, S. M. Tanvir Ahmed, Md. Ahshanul Haque et al. "Iron content of drinking water is associated with anemia status among children in high ground water iron areas in Bangladesh", Tropical Medicine & International Health, 2021 Publication	<1 %
86	S. Thapa. "Nepal's vitamin A supplementation programme, 15 years on: Sustained growth in coverage and equity and children still missed", Global Public Health, 2010 Publication	<1 %
87	article.sciencepublishinggroup.com Internet Source	<1 %
88	bmcpregnancychildbirth.biomedcentral.com Internet Source	<1 %
89	bmcpublichealth.biomedcentral.com Internet Source	<1 %
90	dhsprogram.com Internet Source	<1 %

impactfactor.org

91	Internet Source	<1 %
92	researchonline.lshtm.ac.uk Internet Source	<1 %
93	www.antibioticstreatment.com Internet Source	<1 %
94	www.mdpi.com Internet Source	<1 %
95	150.214.191.180 Internet Source	<1 %
96	Aboubakari Nambiema, Alexie Robert, Issifou Yaya. "Prevalence and risk factors of anemia in children aged from 6 to 59 months in Togo: analysis from Togo demographic and health survey data, 2013–2014", BMC Public Health, 2019 Publication	<1 %
97	Agha Haider Raza, Krishnadas Nanath. "Predicting a Hit Song with Machine Learning: Is there an apriori secret formula?", 2020 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA), 2020 Publication	<1 %
98	Anil Gupta. "Nutritional Anemia in Preschool Children", Springer Science and Business	<1 %

Media LLC, 2017
Publication

99	Ashis Talukder, Benojir Ahammed. "Machine Learning Algorithms for Predicting Malnutrition among Under-Five Children in Bangladesh", Nutrition, 2020 Publication	<1 %
100	Dube, K.. "Iron intake and iron status in breastfed infants during the first year of life", Clinical Nutrition, 201012 Publication	<1 %
101	Getayeneh Antehunegn Tesema, Misganaw Gebrie Worku, Zemenu Tadesse Tessema, Achamyeleh Birhanu Teshale et al. "Prevalence and determinants of severity levels of anemia among children aged 6–59 months in sub-Saharan Africa: A multilevel ordinal logistic regression analysis", PLOS ONE, 2021 Publication	<1 %
102	Halimatou Alaofè, Jennifer Burney, Rosamond Naylor, Douglas Taren. "Prevalence of anaemia, deficiencies of iron and vitamin A and their determinants in rural women and young children: a cross-sectional study in Kalalé district of northern Benin", Public Health Nutrition, 2017 Publication	<1 %

103	Hassan A. Alamri, Vijey Thayananthan. "Bandwidth Control Mechanism and Extreme Gradient Boosting Algorithm for Protecting Software-Defined Networks Against DDoS Attacks", IEEE Access, 2020 Publication	<1 %
104	Md. Mehedi Hasan, Ricardo J Soares Magalhaes, Saifuddin Ahmed, Sonia Pervin, Md. Tariqujjaman, Yaqoot Fatima, Abdullah A Mamun. "Geographical variation and temporal trend in anemia among children 6-59 months of age in low- and middle-income countries during 2000-2018: forecasting the 2030 SDG target", Public Health Nutrition, 2021 Publication	<1 %
105	Md. Mehedi Hasan, Ricardo J Soares Magalhaes, Saifuddin Ahmed, Sonia Pervin, Md. Tariqujjaman, Yaqoot Fatima, Abdullah A Mamun. "Geographical variation and temporal trend in anemia among children aged 6–59 months in low- and middle-income countries during 2000–2018: forecasting the 2030 SDG target", Public Health Nutrition, 2021 Publication	<1 %
106	Menaseb Gebrehaweria Gebremeskel, Afework Mulugeta, Abate Bekele, Lire Lemma	<1 %

et al. "Individual and community level factors associated with anemia among children 6—59 months of age in Ethiopia: A further analysis of 2016 Ethiopia demographic and health survey", PLOS ONE, 2020
Publication

107	Nkechi G. Onyeneho, Daniel J. Corsi, Anura Kurpad, SV Subramanian. "Intergenerational influences on childhood anaemia", Maternal & Child Nutrition, 2018 Publication	<1 %
108	Shegaw Mulu Tarekegn, Leslie Sue Lieberman, Vincentas Giedraitis. "Determinants of maternal health service utilization in Ethiopia: analysis of the 2011 Ethiopian Demographic and Health Survey", BMC Pregnancy and Childbirth, 2014 Publication	<1 %
109	archive.unu.edu Internet Source	<1 %
110	bmcnutr.biomedcentral.com Internet Source	<1 %
111	downloads.hindawi.com Internet Source	<1 %
112	dspace.bracu.ac.bd:8080 Internet Source	<1 %

eprints.gouni.edu.ng

113	Internet Source	<1 %
114	ir.mu.ac.ke:8080 Internet Source	<1 %
115	journals.sagepub.com Internet Source	<1 %
116	nyaspubs.onlinelibrary.wiley.com Internet Source	<1 %
117	qdoc.tips Internet Source	<1 %
118	systematicreviewsjournal.biomedcentral.com Internet Source	<1 %
119	www.childinfo.org Internet Source	<1 %
120	www.coursehero.com Internet Source	<1 %
121	www.hindawi.com Internet Source	<1 %
122	www.ijert.org Internet Source	<1 %
123	"Recent Trends in Communication, Computing, and Electronics", Springer Nature, 2019 Publication	<1 %

124	Md. Sakib Bin Alam, Muhammed J. A. Patwary, Maruf Hassan. "Birth Mode Prediction Using Bagging Ensemble Classifier: A Case Study of Bangladesh", 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021 Publication	<1 %
125	Menaseb Gebrehaweria Gebremeskel, Lire Lemma Tirore. " Factors Associated with Anemia Among Children 6–23 Months of Age in Ethiopia: A Multilevel Analysis of Data from the 2016 Ethiopia Demographic and Health Survey ", Pediatric Health, Medicine and Therapeutics, 2020 Publication	<1 %
126	Mihiretu Alemayehu, Mengistu Meskele, Bereket Alemayehu, Bereket Yakob. "Prevalence and correlates of anemia among children aged 6-23 months in Wolaita Zone, Southern Ethiopia", PLOS ONE, 2019 Publication	<1 %
127	Nils Milman. "Anemia—still a major health problem in many parts of the world!", Annals of Hematology, 2011 Publication	<1 %

128 Ebenezer Opoku, Shirley Crankson, Nana Kwame Anokye. "Prevalence and Risk Factors of Anaemia Among Children Under Five Years in Ghana: Analysis From the Ghana Demographic and Health Survey", Research Square Platform LLC, 2021
Publication
<1 %

129 Huixia Li, Juan Xiao, Minghui Liao, Guangwen Huang, Jianfei Zheng, Hua Wang, Qun Huang, Aihua Wang. "Anemia prevalence, severity and associated factors among children aged 6–71 months in rural Hunan Province, China: a community-based cross-sectional study", BMC Public Health, 2020
Publication
<1 %

130 Innocent B. Mboya, Redempta Mamseri, Beatrice J. Leyaro, Johnston George, Sia E. Msuya, Melina Mgongo. "Prevalence and factors associated with anemia among children under five years of age in Rombo district, Kilimanjaro region, Northern Tanzania", F1000Research, 2020
Publication
<1 %

131 Ropo Ebenezer Ogunsakin, Bayowa Teniola Babalola, Oludare Akinyemi. "Statistical Modeling of Determinants of Anemia Prevalence among Children Aged 6–59
<1 %

	Months in Nigeria: A Cross-Sectional Study", Anemia, 2020 Publication	
132	repository.out.ac.tz Internet Source	<1 %
133	Hayelom Gebrekirstos Mengesha, Hassan Vatanparast, Cindy Feng, Pammla Petrucka. "Modeling the Predictors of Stunting in Ethiopia: Analysis of 2016 Ethiopian Demographic Health Survey Data (EDHS)", Research Square, 2019 Publication	<1 %
134	www.osapublishing.org Internet Source	<1 %