

# Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning Techniques

Similien Ndagijimana<sup>1</sup>, Ignace Habimana Kabano<sup>1</sup>, Emmanuel Masabo<sup>1</sup>, Jean Marie Ntaganda<sup>2</sup>

<sup>1</sup>African Centre of Excellence in Data Science, University of Rwanda, Kigali, Rwanda; <sup>2</sup>University of Rwanda College of Science and Technology, Kigali, Rwanda

**Objectives:** Rwanda reported a stunting rate of 33% in 2020, decreasing from 38% in 2015; however, stunting remains an issue. Globally, child deaths from malnutrition stand at 45%. The best options for the early detection and treatment of stunting should be made a community policy priority, and health services remain an issue. Hence, this research aimed to develop a model for predicting stunting in Rwandan children.

**Methods:** The Rwanda Demographic and Health Survey 2019-2020 was used as secondary data. Stratified 10-fold cross-validation was used, and different machine learning classifiers were trained to predict stunting status. The prediction models were compared using different metrics, and the best model was chosen.

**Results:** The best model was developed with the gradient boosting classifier algorithm, with a training accuracy of 80.49% based on the performance indicators of several models. Based on a confusion matrix, the test accuracy, sensitivity, specificity, and F1 were calculated, yielding the model's ability to classify stunting cases correctly at 79.33%, identify stunted children accurately at 72.51%, and categorize non-stunted children correctly at 94.49%, with an area under the curve of 0.89. The model found that the mother's height, television, the child's age, province, mother's education, birth weight, and childbirth size were the most important predictors of stunting status.

**Conclusions:** Therefore, machine-learning techniques may be used in Rwanda to construct an accurate model that can detect the early stages of stunting and offer the best predictive attributes to help prevent and control stunting in under five Rwandan children.

**Key words:** Machine learning, Prediction, Under-5 children, Stunting, Rwanda

## INTRODUCTION

In 2020, 149.2 million under-5 children were stunted globally, and Africa accounted for 40% of them [1]. Although the global number of stunted children has decreased considerably

since 1990, Africa is the only region where the number of stunted children has increased despite a drop in the prevalence of stunting [2,3]. Stunting in children under the age of 5 remains an issue in Rwanda, despite the government's tremendous success in reducing infant mortality. Stunting in children under the age of 5 has decreased significantly over the last 15 years, according to the Rwanda Demographic and Health Survey (RDHS), with the prevalence of stunting decreasing from 51% in 2005 to 38% in 2015 and 33% in the most recent RDHS 2019-2020 reports; however, the prevalence of stunting remains high [4].

Educational level, maternal demographics, nutrition and health condition, family size, and assets have all been shown

**Received:** August 30, 2022 **Accepted:** December 15, 2022

**Corresponding author:** Similien Ndagijimana  
African Centre of Excellence in Data Science, University of Rwanda,  
P.O Box 4285 Kigali, Rwanda

**E-mail:** similienn@gmail.com

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

to influence nutrition and health outcomes for children, including stunting [5]. Sex, underweight, infections, low-quality breastfeeding from impoverished mothers, and a lack of dietary diversity appear to be individual variables that contribute to child stunting [6]. Childhood stunting has been demonstrated to be influenced by geophysical circumstances. The distance to the main road and the market have both been used to signify how well a property is linked to markets where food may be purchased and sold. Household altitude has also been linked to the risk of childhood stunting [7]. The surge in stunting is due to the appalling rate of undernutrition in the region. Agriculture, education, social protection, water, sanitation, and hygiene are just a few of the sectors that have long been recognised as essential for better nutrition. Therefore, the “all hands on deck” strategy was employed [3].

Most of the relevant literature in Rwanda has used statistical analysis methods, probability strategies, nearest neighbour techniques, logical analysis, comparison techniques, logistic regression, descriptive analysis, and association rules [5,8]. The literature on the use of data science approaches is scarce. These strategies are used to analyse traits or qualities that are chosen based on prior information or logical inference. Hence, it would be fascinating to investigate whether machine learning (ML) techniques can duplicate earlier traits and/or generate new features that require the attention of academics and policy-makers [9].

Several data science and ML approaches have recently been employed in the healthcare industry to detect hidden patterns in nutritional data. The main purpose of these techniques is to offer relevant information to healthcare decision-makers and policymakers so that the nutritional status of children under the age of 5 may be properly determined. ML and data science are currently key contributors to predicting and monitoring patient outcomes [10]. Hence, developing a model capable of predicting stunting in Rwandan children is crucial, so that the best alternatives for early diagnosis, detection, and treatment of stunting can be made a policy priority for the community and health services. Furthermore, few published studies have used ML approaches to predict childhood stunting in Rwanda based on cross-sectional health survey data, necessitating the conduct of this research [11]. Factors influencing stunting have been identified in different countries, even in Rwanda, by using existing statistical approaches; thus, there is a need to identify the factors influencing malnutrition using the ML approach [11].

The main goal of this article is to use ML approaches to develop the most accurate predictive model of stunting among Rwandan children under the age of 5 by identifying the risk factors for stunting among Rwandan children under the age of 5, determining the association between stunting status and the identified risk factors among children under the age of 5, and developing predictive models of stunting among children using different ML algorithms.

## METHODS

### Study Design and Source of Data

The RDHS is a household survey that is nationally representative and collects data for a wide range of population, health, and nutrition-related monitoring and impact evaluation indicators. Therefore, this research is a secondary data analysis of the RDHS 2019-2020, which included a total of 4052 children under the age of 5. Height and weight measurements were obtained from 3814 eligible children under the age of 5, with a validity of height for age measurements of 99.7% [4].

### Description of Variables

#### *Outcome variable*

In this study, the outcome variable was stunting status, which was classified according to World Health Organization (WHO) guidelines. Children's nutritional status was categorised into three ordinal categories based on height for age z-scores, as follows: -3.0 standard deviations (SD): severely stunted; between 3.0 SD and -2.0 SD: moderately stunted, and above -2.0 SD: nourished. These values were derived using the median values and SDs of the WHO child growth guidelines [12]. Generally, children with z-scores less than -2.0 SD are classified as stunted.

#### *Explanatory variables*

The explanatory variables that were associated with stunting were child, parental and household characteristics, as summarized in Table 1. The categories of explanatory variables provided by the RDHS were used [4].

### Data Pre-processing and Transformation

Data pre-processing is the practice of preparing raw data to be comprehensible and usable for analysis. It is made up of several phases, including data cleaning, integration, transformation, and reduction [13]. The process of dealing with miss-

**Table 1.** Explanatory factors measured and recorded for use in the analysis

Variables	Description	Categories
Variables related to a child's characteristics		
Baby_age	Age of the child (mo)	0: <6, 1: 6-11, 2: 12-23, 3: 24-35, 4: 36-47, 5: 48-59
Sex	Sex of the child	0: female, 1: male
Size of a child	Size of the child at birth	0: large, 1: average, 2: small
Birthweight	The weight of the child at birth (kg)	0: ≥2.5, 1: <2.5
Breastfeeding start	Time when the child starts breastfeeding	0: within the first hour, 1: 1-24 hr, 1-2: 30 day
Presence of diarrhoea	The child had diarrhoea in the last 2 wk	0: no, 1: yes
Duration of breastfeeding	Duration of breastfeeding	0: ever breastfed, not currently breastfeeding, 1: never breastfed, 2: still breastfeeding
Intestinal parasite	Drugs for intestinal parasites in the last 6 mo	0: no, 1: yes
Variables related to the child's mother		
Maternal age	Age of the mother (y)	0: < 18, 1: 19-35, 2: >35
Maternal education	Education level of the mother	0: no education, 1: primary, 2: secondary or higher
Maternal anaemia	Anaemia status the of mother	0: not anaemic, 1: anaemic
Mother_occupation	Mother currently working	0: no, 1: yes
Marital_status	Mother's marital status	0: single, 1: married, 2: separated
Mother_height	Respondent's height in centimetres (cm)	0: <160, 1: ≥160
Mothers_BMI	Mothers body mass index	0: underweight, 1: healthy, 2: overweight, 3: obese
Antenatal	No. of antenatal visits during pregnancy	0: no antenatal care, 1: 1-4 antenatal care visits, 2: more than 5 antenatal care visits
Variables related to households		
Residence	Type of place of residence of the child	0: rural, 1: urban
Wealth index	Combined wealth index	1: poor, 2: middle: 3: rich
Source of drinking water	Source of drinking water in the household	0: unimproved, 1: improved
Toilet facilities	Type of toilet facilities in the household	0: unimproved, 1: improved
No. of children under 5 y in the household	All children in the household under 5 y old	0: 1, 1: 2, 2: 3+
Place of delivery	Distribution of live births by place of delivery	0: other, 1: delivery at home, 2: delivery at health facility
Province	Region	0: Kigali, 1: south, 2: east, 3: west, 4: north
Television	The household has a television	0: no, 1: yes
Reading	Frequency of reading newspapers or magazines	0: ever, 1: reading
Radio	The household has a radio	0: no, 1: yes
Altitudes	Cluster altitude (m)	0: ≤2000, 1: >2000

BMI, body mass index.

ing values or incomplete data within a dataset was characterised in this study as data cleaning. Missing values were calculated with the k-nearest-neighbours imputer, which reduces data similarity when compared to the Euclidean distance [14]. The process of changing the format or structure of data in order to prepare it for analysis is known as data transformation. Furthermore, the synthetic minority over-sampling technique (SMOTE) was used to address class imbalance in the target variable by oversampling the minority class by taking each sample from it and providing synthetic examples along the line segments connecting any or all of the minority class k's nearest neighbours [15].

## Statistical Analysis

The Pearson chi-square test and the point biserial correlation matching 95% confidence levels were used to establish the link between independent factors and the dependent variable using Python as statistical software. A *p*-value less than 0.05 was considered as statistically significant in this study. Models were constructed using ML methods, and all models were trained and assessed using a stratified k-fold cross-validation methodology to measure the performance of the ML algorithms on the dataset. Finally, feature importance was used to define the contribution of features in the best model. The learning dataset is partitioned into k disjoint subgroups of ap-

proximately equal size using stratified random sampling to maintain the same class distribution in each subset in stratified k-fold cross-validation [16]. The model was trained on a k-1 subset of the training set, and the remaining subset was utilised to test performance. This procedure was repeated until each of the k subsets served as both a validation and a training set, implying that all possible observations from the entire dataset had been trained and tested, resulting in lower variance within the set estimator and, as a result, less bias in the true rate estimator [17].

### Feature selection

Feature selection is a key phase in data science pre-processing that involves selecting a subset of the original feature spaces based on discriminating capabilities to improve data quality. To acquire significant information in a dataset for a certain objective, feature selection approaches have been widely employed. There are 2 types of feature selection: supervised feature selection and unsupervised feature selection [18]. This study used chi-square supervised feature selection.

### ML algorithms

Based on algorithms, which are sets of mathematical methods that define the relationships between variables, ML approaches are divided into 2 categories: supervised and unsupervised [19]. The researcher used supervised learning techniques to train the model on a variety of risk variables. When the algorithm was effectively trained, it could predict outcomes when applied to new data. A classification algorithm is a form of model that generates discrete categories [19]. The following supervised ML algorithms were used in this study: support vector machines, naïve Bayes, random forest, logistic regression, and extreme gradient boosting.

### Model evaluation

Model evaluation was performed to see how well the classification model worked and how well it classified data. In this study, a confusion matrix and receiver operating characteristic (ROC) curve were used to assess algorithm performance using various metrics such as accuracy, sensitivity, specificity, F1 score, and area under the curve (AUC). The confusion matrix is a condensed table that is used to assess a classifier's performance using data about actual and expected values [20,21].

**Table 2.** The bivariate study of stunting and the distribution of respondents' characteristics

Characteristics	Total	Stunting status		p-value
		Not stunted	Stunted	
Baby_age (mo)				<0.001
<6	376 (9.9)	311 (82.7)	65 (17.3)	
6-11	417 (10.9)	327 (78.4)	90 (21.6)	
12-23	784 (20.6)	499 (63.6)	285 (36.4)	
24-35	786 (20.6)	464 (59.0)	322 (41.0)	
36-47	731 (19.2)	454 (62.1)	277 (37.9)	
48-59	720 (18.9)	487 (67.6)	233 (32.4)	
Size of the child at a birth				<0.001
Large	1211 (31.6)	885 (73.0)	326 (27.0)	
Average	2037 (53.4)	1350 (66.3)	687 (33.7)	
Small	556 (14.8)	307 (55.2)	249 (44.8)	
Birthweight (kg)				<0.001
≥2.5	3590 (90.5)	2446 (68.1)	1144 (31.9)	
<2.5	224 (5.9)	96 (42.8)	128 (57.2)	
Maternal education				<0.001
No education	456 (11.9)	244 (53.5)	212 (46.5)	
Primary	2489 (65.3)	1610 (64.7)	879 (35.3)	
Secondary or higher	872 (22.9)	690 (79.1)	182 (20.9)	
Mother_height (cm)				<0.001
<160	2729 (71.6)	1652 (60.5)	1077 (39.5)	
≥160	1085 (28.5)	890 (82.0)	195 (18.0)	
Type of place of residence				<0.001
Rural	3024 (79.3)	1911 (63.2)	1113 (36.8)	
Urban	790 (20.7)	631 (79.9)	159 (20.1)	
No. of children under 5 y in a household				0.003
1	2109 (55.3)	1444 (68.5)	665 (31.5)	
2	1507 (39.5)	984 (62.9)	523 (37.1)	
≥3	198 (5.2)	114 (57.6)	84 (42.4)	
Province				<0.001
Kigali	430 (11.3)	339 (78.8)	91 (21.2)	
South	878 (23.0)	598 (68.1)	280 (31.9)	
East	920 (24.1)	647 (70.3)	273 (29.7)	
West	975 (25.6)	590 (60.5)	385 (39.5)	
North	611 (16.0)	368 (60.2)	243 (39.8)	
Television				<0.001
No	3281 (86.0)	2068 (63.0)	1213 (37.0)	
Yes	533 (14.0)	474 (88.9)	59 (11.1)	
Altitude (m)				<0.001
≤2000	3143 (82.4)	2178 (69.3)	965 (30.7)	
>2000	671 (17.6)	364 (54.2)	307 (45.8)	
Total	3814 (100)	2542 (66.6)	1272 (33.3)	

Values are presented as number (%).

## Ethics Statement

No ethics approval was required since we conducted an analysis using publicly available data.

## RESULTS

### Descriptive Results

The supervised feature selection using the chi-square method was performed before the analysis. Based on the  $p$ -value, 10 factors were chosen that contributed to stunting and had the highest relative importance out of the 29 factors initially included. Table 2 summarizes the results of the participants' background characteristics. A total of 3814 children aged 0-59 months were enrolled in the study, and 33.35% of those children were stunted. In general, stunting was distributed by age, with the most afflicted age groups being 24-35 months, 36-47 months, and 12-23 months, with respective prevalence rates of 41.0%, 37.9%, and 36.4%. The child-size factor was statistically significant; children born tiny were more likely to be stunted, with a frequency of 44.8%. Birth weight also affected children's stunting status; children born with birth weights less than 2.5 kg were more likely to be stunted, with a prevalence of 57.2%, whereas children born with birth weights more than 2.5 kg had a stunting prevalence of 31.9%.

Mothers with no formal education were more likely to give birth to stunted children, at a prevalence of 46.5%, and the primary and secondary or higher education of a mother also showed an association with the stunting prevalence (35.3 and 20.9%, respectively). The height of the mother influenced the stunting status of children; mothers with a height greater than 160 cm were less likely to have stunted children, with a prevalence of 18.0%, while those whose height was less than 160 cm were more likely to have stunted children, with a prevalence of 39.5%. Children from rural areas were more likely to be stunted than those from urban areas, with a prevalence of

36.8% and 20.1% respectively.

The city of Kigali had a lower prevalence of stunting (21.2%) than other provinces. Media use was also identified as the risk factor for stunting in this study; for example, factors such as having a television were statistically significant. Of children who did not watch television, 37.0% were stunted, versus 11.1% of those who watched television. Altitude was a statistically significant factor, as households located at an altitude less than 2000 m had fewer stunted children (30.7%), while those who lived at an altitude greater than 2000 m were more likely to be stunted, at a prevalence of 45.8%.

### The Findings of Predictive Models and Comparisons Using Different Metrics

#### *The results of predictive models*

To develop models for identifying stunting status based on factors that showed statistical differences between the two groups of stunting status, different ML methods were used. A confusion matrix and ROC analysis were used to present the prediction results with performance parameters for each of the ML algorithms on the test dataset. Before the analysis, a class imbalance was found between the stunted children and non-stunted children (1272 and 2542, respectively). The classes were balanced to 2542 each using SMOTE. The results are summarised in Table 3, which shows the performance of different predictive models.

#### *The best performance model*

The comparative performance of classification algorithms was evaluated in terms of the training accuracy, the test accuracy, the sensitivity, specificity, and the F1. The gradient boosting classifier algorithm yielded the best model, with the highest performance measures for the prediction of stunting status among children aged 0-59 months compared to the other algorithms used in this study, as seen in Table 3.

**Table 3.** The prediction performance results with machine learning algorithms

Variables	Training accuracy	Test accuracy	Precision	Recall	F1	AUC
Logistic	68.27	66.34	63.43	77.17	69.63	71
Random forest classifier	76.92	76.18	69.28	94.09	79.08	84
Support vector machine	75.98	73.03	66.67	92.13	77.36	84
Naïve Bayes	67.07	67.32	63.10	83.46	77.36	74
Extreme gradient boosting	81.62	79.13	72.29	94.49	81.91	89
Gradient boosting	80.49	79.33	72.51	94.49	82.05	89

Values are presented as percentage.  
AUC, area under the curve.

### The most predictive features

The next step in the modelling process was to determine the feature relevance in the prediction of stunting in children aged 0-59 months. The chi-square method was used for feature selection since the literature showed it to be the most successful technique for aggressive term reduction, removing up to 90% of terms while maintaining category accuracy and was among the best feature selection methods [22].

The feature significance rates demonstrated the role of each feature in the prediction of stunting. In this study, the best method (gradient boosting classifier) demonstrated that all

characteristics in the model contributed to the prediction of stunting status, with each category of variables contributing positively or negatively to predicting the risk of children being stunted or non-stunted. This model determined that the mother's height, watching television, the baby's age, province, size at childbirth, and mother's education were the most important predictors of stunting status. Figure 1 depicts the role of each trait in predicting stunting status in both stunted and non-stunted children.

Figure 2 shows the characteristics that had a greater effect in predicting stunted children than in predicting non-stunted children, as well as descriptions of these features. It also describes the contribution of features that were useful for predicting stunted children, such as the mother's height being less than 160 cm, living in the Southern and Western provinces, and living in households without television.

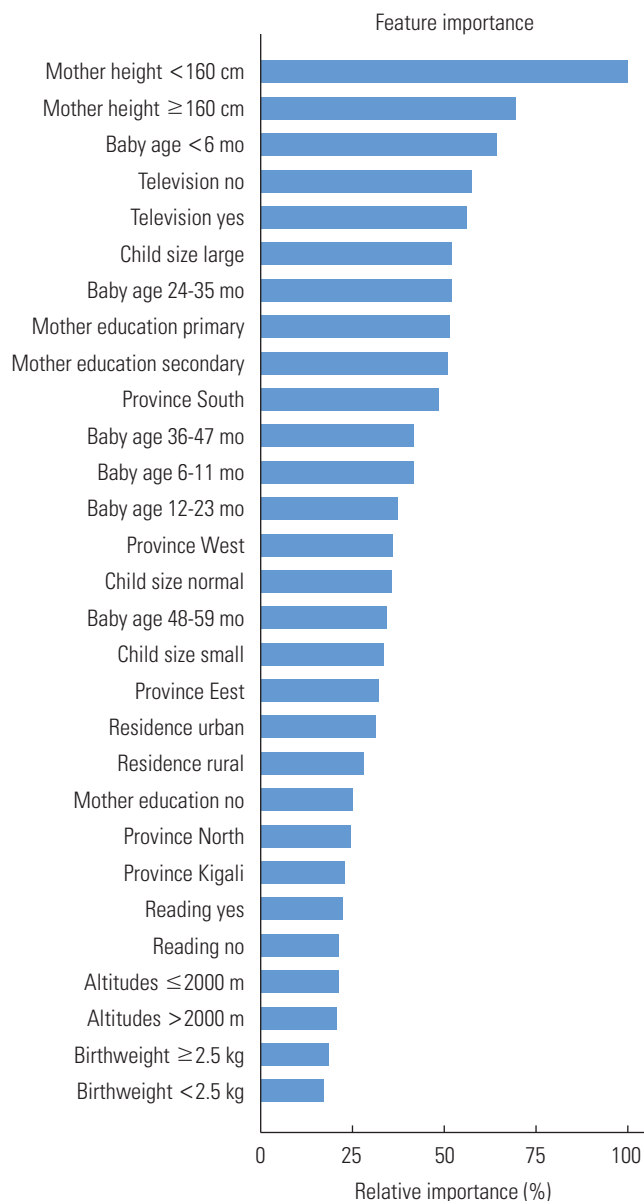


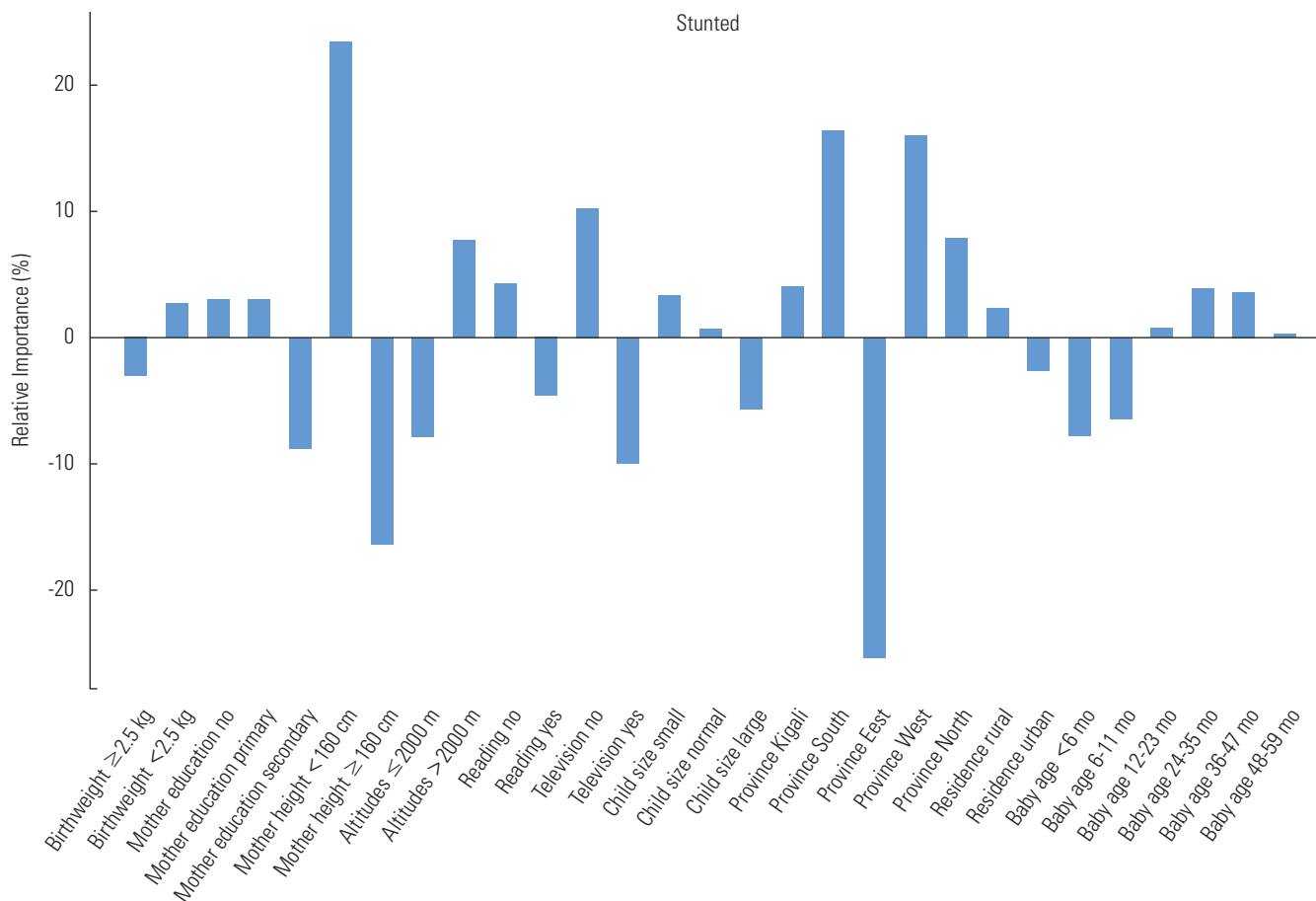
Figure 1. Features importance from the best model (gradient boosting classifier).

### DISCUSSION

This work contributed to the advancement of knowledge about the prediction of stunting in children aged 0-59 months. Using chi-square feature selection, risk factors for stunting were identified; among the top 10 were birth weight, the child's age, and the child's size at birth. The mother's height, and mother's education, the presence of a television, the place of residence, and the province were observed as risk factors for stunting, as shown in Figure 1.

In this study, 10-fold cross-validation was used to train models, and 6 different ML methods (logistic regression, random forest, support vector machine, naïve Bayes, extreme gradient boosting, and gradient boosting) were used to predict stunting status based on substantially linked features, since 10-fold cross-validation was identified as the best method [16]. The predictive models were also evaluated based on their 5 related performance characteristics, with the top-performing model chosen, as discussed above in model evaluation section.

In this study, the gradient boosting classifier significantly outperformed other ML algorithms used in the prediction of stunting status, correctly classifying all children at 80.49%, identifying stunted children at 79.33%, and having the highest discriminating power to differentiate stunted and non-stunted children at 94.49% (Table 3). This is confirmed by the findings from other related studies, which revealed that gradient boosting was the best-performing model [16,23]. In addition, the gradient boosting classifier identified the important



**Figure 2.** Feature importance from the best model in predicting stunted children.

predictive factors that contributed to the prediction of stunting status among children aged 0-59 months through the feature importance methodology as shown in Figure 1.

Generally, as shown in Figure 2, several features contributed to the prediction, but the most strongly contributing factor was the mother's height being less than 160 cm. The mother's height is an important indicator that may reflect a combination of the mother's genetics and the nutritional and environmental factors she experienced during her childhood [24]. Television exposure in young children has been found to affect their language and cognitive development, and television is an important tool for most people, young and old, as most information is now delivered to the public via this technology [25]. Stunting was most common in babies from 24 months to 35 months of age, and stunting has been associated with poor nutrition and recurrent episodes of illness within the first 1000 days of a child's life. Recent studies in South Africa, however, revealed that the risk of stunting among preschool children rose with age, with older children (12-59 months) being more

likely to be stunted than their younger counterparts [25]. The mother's education is also very important in predicting stunting; research has shown that when a mother's education level increases, so does her child's nutritional quality [26].

The size of the child at birth was identified as a risk factor for stunting in children, and monitoring the nutritional status of the mother during pregnancy is therefore highly recommended [26]. Several studies have revealed that children from high altitudes are most likely to be stunted, as was also confirmed by this study [27]. Birth weight is among the risk factors contributing to stunting. The birth weight for children with low birth weight are viable strategies to solve Rwanda's childhood stunting problem. The ML techniques were revealed as the most useful methods of predicting the presence or absence of stunting among children. This study also confirmed that ML approaches outperformed the usual logistic regression model in terms of predictive performance [23].

It is important to be aware of and to take into account the limitations of this study. It was impossible to demonstrate any

causal linkages since the prediction models relied on cross-sectional data (from the RDHS), making it difficult to separate cause from effect. For instance, stunting may be both a cause and a consequence of anaemia. There were no accessible clinical and in-depth factors pertaining to mothers' and children's nutritional consumption or personal hygiene habits. Due to the lack of a clear selection of algorithms in ML modelling, the researcher made decisions based on the literature and subjective judgment. Despite the aforementioned drawbacks, there is evidence that, because of ML algorithms' superior processing capability, they can be used to predict children's stunting status based on socioeconomic and health-related factors.

In conclusion, certain notable strengths are highlighted in this study, which identified a collection of critical predictors of stunting status among under-5 Rwandan children. Among the models developed using ML techniques, the gradient boosting classifier performed best, with the highest classification accuracy in predicting stunting in Rwandan children and the highest discriminating power of stunted versus non-stunted. Eliminating treatment-related errors aids the healthcare industry in improving diagnostic accuracy and treatment success. ML algorithms can be useful in screening children's stunting status, reducing manual efforts in diagnosis, and facilitating the early diagnosis of stunting, according to this study. However, it cannot replace physicians' perceptions and interpretive abilities, although the most predictive features can aid in the prevention and control of stunting among Rwandan children aged 0-59 months.

The findings might be used and implemented by policymakers and healthcare practitioners in the treatment of stunted children. Policymakers should develop policies that assist mothers in receiving vital nutrition education about how to feed their children a balanced diet, as well as programs that encourage females to exclusively breastfeed their infants. Therefore, ML techniques may be used in Rwanda to construct an accurate model that can detect the early stages of stunting and offer the best predictive attributes to help in the prevention and control of stunting in Rwandan children under the age of 5. More studies using ML to predict children's outcomes are recommended.

## CONFLICT OF INTEREST

The authors have no conflicts of interest associated with the material presented in this paper.

## FUNDING

This work was supported by the World Bank funding (ID: ESC 91) through African Centre of Excellence in Data Science, University of Rwanda.

## ACKNOWLEDGEMENTS

This research could not have been completed without the contribution of different people who deserve acknowledgment: Deborah Mukandamage, Diane Uwizera, Prosper Twiringiyimana, and the African Centre of Excellence in Data Science at the University of Rwanda

## AUTHOR CONTRIBUTIONS

Conceptualization: Ndagijimana S. Data curation: Ndagijimana S, Masabo E. Formal analysis: Masabo E, Ndagijimana S. Funding acquisition: Ndagijimana S. Methodology: Ndagijimana S, Ntaganda JM. Project administration: Ndagijimana S, Kabano IH. Visualization: Ntaganda JM, Masabo E. Writing – original draft: Ndagijimana S, Masabo E, Ntaganda JM. Writing – review & editing: Ndagijimana S, Ntaganda JM, Kabano IH.

## ORCID

Similien Ndagijimana <https://orcid.org/0000-0002-5636-3551>  
Ignace Habimana Kabano

<https://orcid.org/0000-0003-3473-4558>

Emmanuel Masabo <https://orcid.org/0000-0002-0894-4909>

Jean Marie Ntaganda <https://orcid.org/0000-0003-2464-2377>

## REFERENCES

1. World Health Organization. Levels and trends in child malnutrition: UNICEF/WHO/the World Bank Group joint child malnutrition estimates: key findings of the 2020 edition; 2020 [cited 2022 Aug 1]. Available from: <https://www.who.int/publications/i/item/9789240003576>.
2. Skoufias E, Vinha K, Sato R. Reducing stunting through multi-sectoral efforts in sub-Saharan Africa. *J Afr Econ* 2021;30(4): 324-348.
3. Skoufias E. All hands on deck: halting the vicious circle of stunting in sub-Saharan Africa; 2018 [cited 2021 Dec 19]. Available from: <https://blogs.worldbank.org/health/all-hands-deck>



- halting-vicious-circle-stunting-sub-saharan-africa.
4. National Institute of Statistics of Rwanda. Demographic and Health Survey (2019/20); 2021 [cited 2022 Aug 1]. Available from: <https://www.statistics.gov.rw/datasource/demographic-and-health-survey-201920>.
  5. Weatherspoon DD, Miller S, Ngabitsinze JC, Weatherspoon LJ, Oehmke JF. Stunting, food security, markets and food policy in Rwanda. *BMC Public Health* 2019;19(1):882.
  6. Wamani H, Astrøm AN, Peterson S, Tumwine JK, Tylleskär T. Boys are more stunted than girls in sub-Saharan Africa: a meta-analysis of 16 demographic and health surveys. *BMC Pediatr* 2007;7:17.
  7. Dang S, Yan H, Yamamoto S. High altitude and early childhood growth retardation: new evidence from Tibet. *Eur J Clin Nutr* 2008;62(3):342-348.
  8. Binagwaho A, Agbonyitor M, Rukundo A, Ratnayake N, Ngabo F, Kayumba J, et al. Underdiagnosis of malnutrition in infants and young children in Rwanda: implications for attainment of the Millennium Development Goal to end poverty and hunger. *Int J Equity Health* 2011;10(1):61.
  9. Khare S, Kavyashree S, Gupta D, Jyotishi A. Investigation of nutritional status of children based on machine learning techniques using Indian Demographic and Health Survey data. *Procedia Comput Sci* 2017;115:338-349.
  10. Momand Z, Mongkolnam P, Kositpanthavong P, Chan JH. Data mining based prediction of malnutrition in Afghan children. In: 12th International Conference on Knowledge and Smart Technology (KST). Pattaya: IEEE; 2020, p. 12-17.
  11. Agrawal A, Agrawal H, Mittal S, Sharma M. Disease prediction using machine learning. In: Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT) [cited 2022 Aug 1]. Available from: <http://dx.doi.org/10.2139/ssrn.3167431>.
  12. World Health Organization. Guideline: assessing and managing children at primary health-care facilities to prevent overweight and obesity in the context of the double burden of malnutrition; 2017 [cited 2022 Aug 1]. Available from: <https://www.who.int/publications/i/item/9789241550123>.
  13. MIT Critical Data. Secondary analysis of electronic health records; 2016 [cited 2023 Jan 20]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK543630/>.
  14. Jonsson P, Wohlin C. An evaluation of k-nearest neighbour imputation using Likert data. In: 10th International Symposium on Software Metrics, 2004. Chicago: IEEE; 2004, p. 108-118.
  15. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 2002; 16:321-357.
  16. Bitew FH, Sparks CS, Nyarko SH. Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. *Public Health Nutr* 2022;25(2):269-280.
  17. Berrar D. Cross-validation. *Encycl Bioinform Comput Biol* 2019; 1:542-545.
  18. Sulistiani H, Tjahyanto A. Comparative analysis of feature selection method to predict customer loyalty. *IPTeK J Eng* 2017; 3(1):1-5.
  19. Sidey-Gibbons JA, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* 2019; 19(1):64.
  20. Lumbanraja FR, Fitri E, Junaidi A, Prabowo R. Abstract classification using support vector machine algorithm (case study: abstract in a Computer Science Journal). 2021 *J Phys Conf Ser* 2021;1751(1):012042.
  21. Lachiche N. Receiver operating characteristic (ROC) analysis. In: John W, editor. *Encyclopedia of data warehousing and mining*. 2nd ed. Hershey: IGI Global; 2017, p. 1675-1681.
  22. Sun C, Wang X, Xu J. Study on feature selection in finance text categorization. In: 2009 IEEE International Conference on Systems, Man and Cybernetics. 2009 Oct 11-14; San Antonio: IEEE; 2009, p. 5077-5082.
  23. Khan JR, Tomal JH, Raheem E. Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh. *Inform Health Soc Care* 2021; 46(4):425-442.
  24. Saleh A, Syahrul S, Hadju V, Andriani I, Restika I. Role of maternal in preventing stunting: a systematic review. *Gac Sanit* 2021; 35 Suppl 2:S576-S582.
  25. Jusoff K, Sahimi NN. Television and media literacy in young children: issues and effects in early childhood. *Int Educ Stud* 2009;2(3):151-157.
  26. Makoka D. The impact of maternal education on child nutrition: evidence from Malawi, Tanzania, and Zimbabwe; 2013 [cited 2022 Aug 1]. Available from: <https://www.dhsprogram.com/pubs/pdf/WP84/WP84.pdf>.
  27. Nshimiyiryo A, Hedt-Gauthier B, Mutaganzwa C, Kirk CM, Beck K, Ndayisaba A, et al. Risk factors for stunting among children under five years: a cross-sectional population-based study in Rwanda using the 2015 Demographic and Health Survey. *BMC Public Health* 2019;19(1):175.