

Analysis of Risk Factors Associated with Acute Respiratory Infections among Under-Five Children in Uganda

Yassin Nshimiyimana

School of Statistics, East China Normal University

Yingchun Zhou (✉ yczhou@stat.ecnu.edu.cn)

KLATASDS-MOE, School of Statistics, East China Normal University

Research Article

Keywords: Acute Respiratory Infections (ARIs), Risk Factors, Under-Five Children, Uganda

Posted Date: September 13th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-685180/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Analysis of Risk Factors Associated with Acute Respiratory Infections among Under-Five Children in Uganda

Yassin Nshimiyimana¹ and Yingchun Zhou^{2*}

Abstract

Background: Globally, 7.4 million young children are being killed from infectious and treatable diseases, and Saharan Africa accounts for 90% of deaths. In Uganda, Acute Respiratory Infections (ARIs) remain the leading cause of childhood morbidity and mortality among under-five children. The study aims to identify and analyze contributing potential risk factors of childhood ARIs disease among under-five children in Uganda.

Methods: A case-control study was conducted using data for 13,493 sampled under-five children retrieved from a recent 2016 Uganda Demographic and Health Survey. We compared two supervised tree-like structure machine learning algorithms and two logistic regression methods in terms of classification performances in predicting ARIs disease outcomes and analysing various child and parental socio-demographic, behavioural, and environmental characteristics.

Results: The study results revealed that the ARIs prevalence among under-five children accounted 40.3% cases. The Logistic regression findings showed that the risk of developing childhood ARIs disease declined with increase in child's age where the risk of having ARIs was higher in children in one year of birth (AOR=1.27; $p < 0.001$) and lower in children aged four years old (AOR=0.69; $p < 0.001$) compared to the infants. Other factors such as the age of mother where children born from teen mothers (15-19 years) were high likely (AOR=1.28; $p < 0.001$) to have ARIs illness compared to those whose mothers were in the middle age groups, and children whose mothers breastfed showed a lower risk of ARIs disease (AOR=0.83; $p < 0.001$) compared to those who not breastfed. In the dry season, children were more likely to develop ARIs disease (AOR=1.34; $p < 0.001$) compared to the wet season, and factors such as the region of residences like central region, mother employment, and domestic cooking energy like wood were also potential risk factors of ARIs disease among under-five children in Uganda. In addition, three methods such as Decision Tree (Accuracy = 61.2%; AUC=0.610), Elastic Net Logistic Regression (Accuracy=61.7%; AUC=0.627), and Binary Logistic Regression (Accuracy=62.1%; AUC=0.638) showed approximately similar performances in predicting and classifying ARIs disease outcomes. However, the Random Forest (Accuracy=88.7%; AUC=0.951) showed superior difference in performance comparatively.

Conclusion: Government and healthcare stakeholders need to make effective programs to improve public health policy against childhood infectious diseases by targeting the proper provision of maternal and child health-related education to household heads and mothers to adopt and prioritize breastfeeding practices, childcare, and ensure proper dwelling places for families and young children particularly in crowded regions and geographic places where ARIs prevalence is high.

Keywords: Acute Respiratory Infections (ARIs); Risk Factors; Under-Five Children; Uganda

*Correspondence:

yczhou@stat.ecnu.edu.cn

²KLATASDS-MOE, School of

Statistics, East China Normal

University, Shanghai, China

Full list of author information is available at the end of the article

Background

Globally, infant and child mortality rate is an essential issue and fundamental indicator of health status, quality of life, and socio-economic situation of a country's population [1]. It has been a 60% remarkable decline in under-five mortality over the last three decades worldwide. However, 7.4 million young children are still dying

from preventable and treatable diseases every year across countries where 70% of deaths occurred among children under five years, and South-Asia and sub-Saharan Africa nations account for 95% of deaths. The under-five Mortality rate remained higher, with 76 deaths per 1,000 live births on average, i.e., 1 in 13 children died before reaching five years old in sub-Saharan Africa [2].

Various factors such as living conditions and socio-economic development of a country are significant contributing factors to the child mortality rate where Acute Respiratory Infections (ARIs) remain the leading causes of morbidity and mortality in young children in low-income countries, particularly sub-Saharan African nations, including Uganda [3]. The ARIs diseases are typically caused by a contagious virus and bacterial infections, which spread rapidly through droplets from either person-to-person or contaminated food or drinking water due to poor hygiene [4]. The World Health Organization (2019) ranked ARIs illness the fourth-deadliest disease and top among treatable and infectious childhood diseases where ARIs and Diarrhea communicable diseases kill as many children who died from other conditions such as Malaria, HIV/AIDS, measles, and all other illnesses combined [5–7]. Acute Respiratory Infections (ARI) are infections that affect the upper and lower respiratory system from the middle ear to the lungs accompanied by short and rapid breathing related to the chest, and when they coincide with other childhood illnesses like Diarrhoea or Malaria, they sometimes lead to death [8].

In Uganda, infant and under-five mortality rates also remained high where 1 in 16 young Ugandan children died before surviving five years old and 42% of the under-five mortality occurred in the neonatal period. The ARIs prevalence among under-five children in Uganda accounted for 39% of children and 81.3% in urban areas [9, 10].

In dealing with medical data, proficiency and experience were required in disease diagnosis in medical research [11]. Various diseases such as diabetes, liver disorder, heart disease, diabetes, and breast cancer have been diagnosed using different data mining and machine learning techniques in various studies. The appropriate application of these techniques and tools showed better results in disease diagnosis than the traditional methods for disease classification, and predictions [11, 12]. Childhood mortality information and estimates are relevant to the country's health assessment of its population. This information is an essential indicator of the country's quality of life and socioeconomic development. The preventable and communicable infectious diseases like ARIs take millions of young children's lives every year across the world where under-five years mortality rate account 70%, particularly in sub-Saharan Africa including Uganda [2]. Most studies found that families living in poverty and low living standards were most likely to be killed by ARIs diseases in countries like Uganda. Based on the various revealed findings, new methods of disease classifications performed better, but none of the techniques is more consistent than others in predicting disease using different datasets.

The Chi Chiao and Deji-Abiodun (2020) study about global analysis of the regional variation in the symptoms of ARIs during childhood conducted on 432,860 children across 48 low-income nations revealed that environmental vulnerability and the level of ambient air pollution were contributing factors of childhood ARIs disease [13]. Another study by Abdul A. et al. (2019) conducted across 28 sub-Saharan African countries showed that parental employment, domestic facilities,

and childhood vaccinations were associated with ARIs illness among children below five years old [14]. Other factors such as child age, season effect, and regional factors were also significantly associated with childhood ARIs disease in the study conducted in Rwanda in 2016 [15]. Related studies conducted in Uganda, such as studies by Edward Bbaale (2011) [16] and another conducted using clinical data about Influenza and several ARIs in Uganda in 2016 [17] showed that the mother breastfeeding, parental wealth status and employment, child nutrition and age, and environmental factors were contributing factors of under-five children ARIs diseases in Uganda.

Many related research and studies regarding risk factors associated with childhood Acute Respiratory Infections have been conducted using the most popular traditional analysis methods such as conventional logistic regression and the chi-square test. However, there is no other research that has investigated the childhood ARIs disease determinants using new penalized logistics regression and machine learning methods for disease classification, either in Uganda or globally. Therefore, this study used traditional and new analysis methods to analyze and identify potential risk factors associated with ARIs disease among under-five children in Uganda. The heavy loss of young lives from ARIs childhood mortalities poses a heavy burden to families and healthcare providers in Uganda. With this research, crucial and helpful analytical findings will play a great role in reducing childhood mortality rate, and thanks to this, effective precautions and measures addressing the children illness risk factors will be taken. Childhood disease information and estimates are relevant in disease management and health assessment of the country, especially in sub-Saharan Africa.

Methods

Design and Sampling Procedures

A case-control study was adopted involving the use of a two-stage stratified sampling design. The sampling frame used in this study involved selecting geographical areas called Enumeration Areas (EAs) that cover an estimated number of 130 residential households for each of EAs. The first sampling design involved setting EAs in the urban and rural areas in the country, while the second involved selecting households to include in each cluster (EAs). This study used a probability proportional to segment size to choose one EA segment in places where EAs were large with more than 300 households. Thus, each EA segment included 30 male and female households selected randomly from each of the EA clusters.

Population and Sample Size

The target population of this study comprised male and female children below five years of age from different regions of the country who were at risk of ARIs disease in Uganda. Mothers in reproductive age (15-49 years old) who lived in the selected households the night before the survey have been surveyed. This study considered a representative sample size of 13,493 children below five years of age whose mothers reported regarding the child's illness status and other related issues. In this study, all the responses regarding children's illness were regarded as subjective since all the information was based on mothers' perceptions without validations by medical

personnel. We used only 10,120 (75%) of the sample size to analyze and train four methods considered in this study. The remaining 3,373 (25%) was used to evaluate the four methods' performances regarding predictions and classifications.

Data Source

In this study, we employed secondary data drawn from the most recent Uganda Demographic and Health Survey (UDHS) collected from 20 June to 16 December 2016. The 2016 UDHS surveys were conducted by the Uganda Bureau of Statistics (UBOS) in collaboration with the Government of Uganda through the DHS Program funded by the United States Agency for International Development (USAID). These surveys are national representative cross-sectional surveys conducted every five years in over 90 countries worldwide, especially in low-income countries, including Uganda. The purpose of the DHS surveys is to collect up-to-date data and estimates of fundamental demographic and health indicators relevant to policymakers and program managers in evaluating population health and nutritional programs of countries [9]. This study used data on children whose mothers of reproductive age were surveyed. The 2016 UDHS dataset can be found at DHS Program websites at www.dhsprogram.com after applying for data approval.

Variable Descriptions

Children and parental characteristics were used as the predictor variables. The information on mothers' perceptions regarding the child illness of the ARIs disease was considered as an outcome variable of interest in this study. The study used 15 selected child and mother characteristics such as socioeconomic, demographic, parental behaviors, and environmental risk factors of ARIs found in other previous related literature. Data features and variables descriptions used in this study were briefly explained in Table 1 and 2.

Outcome Variable The prevalence of Acute Respiratory Infections in children below five years of age in Uganda was the primary outcome measure of this study. Mothers were asked whether their children below five years of age had been ill with cough accompanied by short, rapid, or difficulty breathing in the last two weeks before being surveyed, i.e., symptoms of ARIs disease. The children whose mothers were reported to have symptoms were considered cases coded 1 for each case while those whose mothers reported that they did not experience any symptoms of ARIs illness were used as controls with values 0 for each control. The Childhood ARIs disease outcome was a binary outcome variable in nature i.e., occurrence or non-occurrence of the disease in the child. In this study, ARIs illness information was considered subjective since they were based on mothers' perceptions without a medical diagnosis.

Predictor Variables We used fifteen exposure variables based on the risk factors found in the previous findings as contributing factors of ARIs illnesses. Parental, children, and household characteristics were considered as the predictor variables in Table 1 and 2 of this study. Socioeconomic and demographic characteristics such as region of residence, gender, age, education, employment, and family wealth status were employed. Parental behaviors and environmental characteristics such as

Table 1 Description of Socioeconomic and Demographic Characteristics

Variable	Type	Description
Outcome Variable		
Childhood ARIs	dummy	ARIs responses were classified into control for a child reported as not had ARIs symptoms, and a case for a child reported as had ARIs symptoms. 1 = case, 0 = control.
Socioeconomic and Demographic Characteristics		
Child Age	categorized	Children ages were measured in terms of months from 0 to 59 months of birth categorized into five age groups.
Child Sex	Nominal	Children in both gender, males and females were used, 0 for male and 1 for female.
Mother Age	categorized	The ages for children mothers were included based on the reproductive ages (15 to 49 years old) categorized into six age groups.
Region	Nominal	The Uganda districts were classified into four regions (central, eastern, northern, and western region) based on their locations in the country.
Mother education	Nominal	This factor was used based on the levels of education such primary, secondary, higher, and considered also those who were not educated.
Mother employment	Nominal	Women whose occupation were based on agriculture and farming were considered as farmer while those worked in other occupation were used as employed category, not employed for otherwise.
Parental wealth	Nominal	Family income wealth background of the child was considered and grouped as were grouped as lower class for poorer and poor families, higher class for rich and richer families, and middle class for neither poor nor rich.

breastfeeding practices of the mother, delivery place, household size, vaccination status for the child, climate and season factor, and domestic facilities such as toilets, cooking energy, and drinking water source facilities were also considered. The variables such as ages for both child and mother were grouped into age groups, the household size was categorized into the crowding status where households with more than five people were considered as crowded, and other variables such as the region of residence, employment for mothers, and domestic facilities were combined into groups based on the categories in the initial 2016 DHS dataset. The season effect variable was classified according to the two Uganda annual season periods. The dry season occurs from December to February and June to August, while the wet season occurs from September to November and March to May. All the variables and features were analyzed and used to predict childhood ARIs disease outcomes among under-five children.

Statistical Analysis Methods

We employed four supervised analytical methods comprised of two regression methods and two tree-like structure data mining methods used to predict and classify childhood ARIs disease outcomes (i.e. cases and controls). Two regression methods were used to determine the effects and its 95% confidence intervals for potential risk factors contributing to the childhood ARIs outcomes in Uganda.

Table 2 Description of Behavioral and Environmental Characteristics

Variable	Type	Description
Household Size	categorized	Numbers of people in the Household were used as a crowding factor contributing to the child ARIs illness. The number of people less than 6 were considered as not crowded households, otherwise crowded households.
Breastfeeding	Nominal	Mothers were asked if they were breastfed their children. And “Yes” responses were used as breastfeed, otherwise not breastfeed.
Child given IP drug	Nominal	Mothers were asked if the child received a drug for intestinal parasites drugs in last 6 months. It was a “Yes” and “No” question. Coded 1 for Yes, and 0 otherwise.
Place of delivery	Nominal	The various possible places of delivery for children were also used in contributing factors of childhood ARIs illness such as private and public hospitals, centers and posts, and home places were considered.
Toilet facility	Nominal	Different household types of toilets were considered based on their protective likelihood. Household with slab, flush, improved pit latrine were grouped into one group while household with other types of toilets were in another group and Otherwise no facilities.
Cooking energy	Nominal	Since the majority of households used unimproved energy when cooking. The popular energy considered in this study were wood and charcoal that caused air pollution.
Use water source	Nominal	The household with water sources were from unprotected spring, rainwater, and river or dam were considered as unprotected sources while those from piped, protect well and spring were protected sources of drinking water.
Season effect	Nominal	The country climate risk factor was also considered. Ugandan two annual season periods such as the dry season from December to February and June to August and wet season from September to November and March to May were measured based on the time the mothers were interviewed.

Logistic Regression

We used logistic regression (LR) approach [18], which is a particular case of the generalized linear model used to estimate the relationship between predictor variables and a binary outcome variable of interest, i.e., cases and controls of childhood ARIs disease. The mathematical convenience of the LR model and its easier odds ratios interpretation made it a popular traditional approach and powerful supervised method for classification mostly used in medical and biomedical research. Since the LR model provides an easy way of finding the probabilities that given instances belong to certain classes, and the outcome lies between 0 and 1, these made the LR method, as a binary classifier, used to assign an input instance into two classes. It classified the instance with a probability higher than 0.5 into one class, say “a case”, and otherwise, “a control” Binary logistic regression was used to predict whether the risk factors were associated with the odds of being a case of ARIs disease versus odds of being control to the child.

Given the childhood ARIs disease outcomes, Y , such that $y = 1$ for a case, and $y = 0$ for a control; And the selected k predictor variables, X_1, X_2, \dots, X_k . Then, the probability that an outcome was a case was given by $\pi = P\{y = 1|X_1, X_2, \dots, X_k\}$.

By considering the effects of predictor variables, b_1, b_2, \dots, b_k that each had on the ARIs disease outcome, Therefore, the LR model is given by:

$$\pi = \frac{\exp \{b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k\}}{1 + \exp \{b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k\}} \quad (1)$$

In general, the LR model in Equation (1) was expressed in terms of logits. Thus, the linear form of the log transformation of the log odds that ARIs outcome is a case versus a control outcome is was given by the Equation (2):

$$\begin{aligned} \text{LR} = \text{Logit}[\pi] &= \ln \left(\frac{\pi}{1 - \pi} \right) \\ &= b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k \end{aligned} \quad (2)$$

We used stepwise variable selection procedures to select the most influential predictor variables to include in the final LR model. The estimates of the adjusted odds ratios (AOR) and their corresponding 95% confidence intervals (C.I) for potential risk factors associated with childhood ARIs disease in children below five years of age were reported in Table 6.

Elastic Net Logistic Regression

The Elastic Net (EN) regression approach proposed by Zhou and Hastie (2005) [19] was also used as a regularization method that penalizes coefficients in the logistic regression (LR) model. The EN regression method is good in solving the overfitting problems from the high amount of input variables in the model, which differs from the previous LR model [2]. The method also involves shrinkage of the coefficients of non-informative predictor variables towards zero, and the model itself controlled all the correlated features that can present in the data analysis.

Given the observed data values of predictor variables considered in this study $X = x_1, x_2, \dots, x_n$ and the observed values of the ARIs outcome of interest $Y = y_1, y_2, \dots, y_n$ such that the tuning parameter λ and α were non-negative values, then the EN for logistic regression with all effect coefficients $b = (b_1, b_2, \dots, b_k)$ for the input variables is defined as:

$$\begin{aligned} \text{EN}_{(\widehat{b_0}, \widehat{b})} &= \arg \min \left\{ - \sum_{i=1}^n \left[y_i (b_0 + x_i^T b) \right. \right. \\ &\quad \left. \left. - \ln(1 + \exp \{b_0 + x_i^T b\}) \right] \right. \\ &\quad \left. + \lambda \left[\frac{1}{2} (1 - \alpha) \sum_{j=1}^k b_j^2 + \alpha \sum_{j=1}^k |b_j| \right] \right\}. \end{aligned} \quad (3)$$

In general, the parameter α in Equation (3) control the model and take values between 0 and 1. In contrast, parameter λ control the overall strength of the penalty. This study used ten-folds cross-validation (CV) to extract the best tunes of the parameter lambda values to predict childhood ARIs outcomes, case and control outcomes.

Decision Tree

We also considered two popular tree-like structure data mining methods for classifications used in this study to predict and classify the childhood ARIs disease outcomes in children below five years of age. The Decision Tree (DT) machine learning algorithm proposed by Quinlan J.R. (1986) [20] mostly applied in medical research to classify disease based on its tree-like structure that is quick and simple to learn and interpret [21]. The architecture of trees in Figure 1 for the DT method is constructed in top-down nodes structure starting with the most valuable node known as the root node of the tree and proceeding down in the internal nodes up to the leaves of the tree where final decisions are made according to the class outcome. In this study, each tree nodes represented each important input factor variable of the root input factor in the dataset from up to down importance sequence. The leaves of the tree corresponded to the ARIs outcomes, i.e. case and control outcomes. Every object in the training set was correctly classified according to the test outcome.

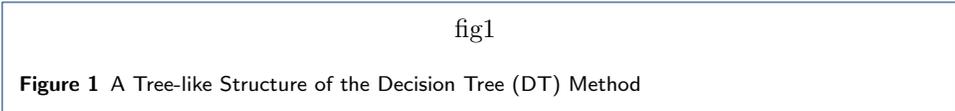


fig1

Figure 1 A Tree-like Structure of the Decision Tree (DT) Method

Random Forest

A Random Forest (RF) is also another tree-like structure method considered in this study. The RF method introduced by Leo Breiman (2001) [22] as one of the effective ensemble tools that produce accurate predictions in disease classifications without overfitting of the data compared to the DT method. As illustrated in Figure 2, the RF method involves growing a set of many DT trees from new random samples that are generated from the training dataset where a set of multiple independent and unpruned classification trees are constructed based on the combinations of randomly selected features with the most popular class in each node to grow a tree. The method itself is effective and more accurate in minimizing the variance resulting from using a single Decision Tree.

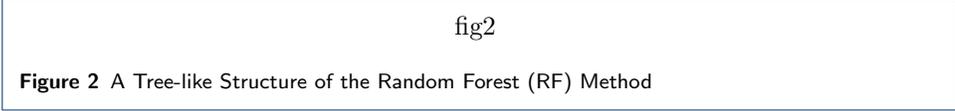


fig2

Figure 2 A Tree-like Structure of the Random Forest (RF) Method

Evaluation Metrics for Classification Performance

In this study, we considered five evaluation metrics to assess and measure the predictive and classification performances of the four analytical methods employed in this study. The metrics such as total accuracy, precision, specificity, recall, and the Area Under the Curve (AUC) were used and compared based on their predictive values. These five measures are all based on two by two matrix procedures known as confusion matrix strategy for classification to compare the predicted and actual numbers of cases and controls for ARIs disease class outcome values of each classifier, as shown in Table 3.

The matrix involves the use of metrics such as True Positives (TP) indicates the number of predicted case outcomes that are correctly classified, False Positives (FP) refers to the number of predicted case outcomes that are incorrectly classified, True Negatives (TN) shows the number of predicted control outcomes that are correctly, and False Negatives (FN) symbolizes the number of predicted control that are incorrectly classified by a given classifier method. These values are more important in the computation of the five evaluation measures considered in this study to evaluate the performance of each method.

Table 3 Confusion Matrix

		Predicted	
		Case	Control
Actual	Case	<i>TP</i>	<i>FN</i>
	Control	<i>FP</i>	<i>TN</i>

The predictive accuracy in Equation (4) was used to compare proportions of case and control outcomes correctly classified by the method.

$$\text{Total Accuracy} = \frac{(TP + TN)}{(TP + FP) + (TN + FN)}. \tag{4}$$

The precision metric in Equation (5) also known as positive predictive value measure was employed to measure and compare proportions of case outcomes that were correctly classified given all predicted cases similar to the specificity measure for control outcomes.

$$\text{Precision} = \frac{TP}{(TP + FP)}. \tag{5}$$

Finally, the recall metric in Equation (6) is a measure of sensitivity which is an important metric that is mostly applied in medical diagnosis to diagnose ill patients [23]. It was used to compare performances of the methods considered in this study based on the proportion of predicted case outcomes that were correctly classified in the real case outcomes.

$$\text{Recall} = \frac{TP}{(TP + FN)}. \tag{6}$$

The computations and analysis were done using a caret package in R statistical software using 75% of the sample to train various methods. The remaining 25% of the sample was used for predictions and performance evaluations of methods.

Results

Summaries and Descriptive Analysis

Distributions and descriptions of characteristics according to the ARIs illness prevalence among under-five children in Uganda were presented. Out of 13,493 sampled Ugandan under-five children, the prevalence of childhood Acute Respiratory Infections disease was 5,437 (40.3%) of the ARIs cases, and 8,056 (59.7%) of the ARIs controls as shown in Figure 3.

Results in Table 4 show that the ARIs cases among under-five children were approximately equal in males (49.3%) and females (50.7%). In the sampled children,

fig3

Figure 3 The ARIs Prevalence among Under-Five children in Uganda.

the Chi-square test results reveal significant differences in ARIs prevalence according to child ages, parental wealth classes, mothers' employment, education level, and child region of residence. The majority of ARIs cases were high in younger children 0-11 months (21.1%) and 12-24 months (23.4%) of age from parents in lower wealth classes (49.8%) whose employment status were farmers (43.8%) and employed (43.4%). The childhood ARIs cases were found high in the eastern region (27.9%) of the country in mothers in the middle age groups 20-24 years (27.1%) and 25-29 years old (27.0%) with education level was primary education (61.2%) compared to the other categories in this study.

Table 4 Distribution of Socioeconomic and Demographic Characteristics by ARIs Disease Prevalence

Characteristics	ARIs Prevalence n = 13,493			Sig.	
	Controls (%)	Cases (%)	Total (%)		
Child Age (months)	0-11	1,724 (21.4)	1,149 (21.1)	2,873 (21.3)	< 0.001
	12-23	1,440 (17.9)	1,273 (23.4)	2,713 (20.1)	
	24-35	1,534 (19.0)	1,124 (20.7)	2,658 (19.7)	
	36-47	1,608 (20.0)	1,008 (18.5)	2,616 (19.4)	
	48-59	1,750 (21.7)	883 (16.3)	2,633 (19.5)	
Child Sex	Male	3,991 (49.5)	2,757 (50.7)	6,748 (50.0)	0.183
	Female	4,065 (50.5)	2,680 (49.3)	6,745 (50.0)	
Region	Central	1,358 (16.9)	1,360 (25.0)	2,718 (20.1)	< 0.001
	Eastern	2,346 (29.1)	1,517 (27.9)	3,863 (28.6)	
	Northern	2,023 (25.1)	1,357 (25.0)	3,380 (25.1)	
	Western	2,329 (28.9)	1,203 (22.7)	3,532 (26.2)	
Mother Age (years)	15-19	441 (5.5)	364 (6.7)	805 (6.0)	< 0.001
	20-24	2,165 (26.9)	1,474 (27.1)	3,639 (27.0)	
	25-29	1,991 (24.7)	1,466 (27.0)	3,457 (25.6)	
	30-34	1,716 (21.3)	1,072 (19.7)	2,788 (20.7)	
	35-39	1,094 (13.6)	670 (12.3)	1,764 (13.1)	
	40-49	649 (8.1)	391 (7.2)	1,040 (7.7)	
Mother Education	No education	1,097 (13.6)	705 (13.0)	1,802 (13.4)	0.002
	Primary	5,110 (63.4)	3,328 (61.2)	8,438 (62.5)	
	Secondary	1,449 (18.0)	1,088 (20.0)	2,537 (18.8)	
	Tertiary	400 (5.0)	316 (5.8)	716 (5.3)	
Mother Employment	Unemployed	1,481 (18.4)	698 (12.8)	2,179 (16.1)	< 0.001
	Farmer	3,947 (49.0)	2,380 (43.8)	6,327 (46.9)	
	Employed	2,628 (32.6)	2,359 (43.4)	4,987 (37.0)	
Parental Wealth	Lower	4,012 (49.8)	2,579 (47.4)	6,591 (48.8)	0.004
	Middle	1,557 (19.3)	1,033 (19.0)	2,590 (19.2)	
	Higher	2,487 (30.9)	1,825 (33.6)	4,312 (32.0)	

Parental behaviors and environmental characteristics presented in Table 5 such as the breastfeeding of the child's mother, place of delivery, and effects from climatic season had relationships with childhood ARIs prevalence in Uganda. In the dry season, the majority of childhood ARIs cases were reported (53.4%) compared to

wet season (46.6%) in mothers who were currently breastfed (56.6%) and whose site of delivery was at health centers (40.6%) in Uganda.

Table 5 Distribution of Environmental and Behavioral Characteristics by ARIs Disease Prevalence

Characteristics		ARIs Prevalence n = 13,493			Sig.
		Controls (%)	Cases (%)	Total (%)	
Household Size	No crowd (≤ 5)	3,746 (46.5)	2,581 (47.5)	6,327 (46.9)	0.267
	Crowd (> 5)	4,310 (53.5)	2,856 (52.5)	7,166 (53.1)	
Breastfeeding	No Breastfeed	3,323 (41.3)	2,362 (43.4)	5,685 (42.1)	0.011
	Breastfeed	4,733 (58.7)	3,075 (56.6)	7,808 (57.9)	
Child given IP drug	No	3,615 (44.9)	2,363 (43.5)	5,978 (44.3)	0.105
	Yes	4,441 (55.1)	3,074 (56.5)	7,515 (55.7)	
Place of delivery	Home	2,257 (28.0)	1,383 (25.4)	3,640 (27.0)	0.005
	Public Hospital	1,464 (18.2)	1,030 (18.9)	2,494 (18.5)	
	Health Center	3,225 (40.0)	2,205 (40.6)	5,430 (40.2)	
	Private Hospital	1,110 (13.8)	819 (15.1)	1,929 (14.3)	
Toilet facility	With slab	2,320 (28.8)	1,645 (30.3)	3,965(29.4)	0.122
	Without slab	4,893 (60.7)	3,208 (59.0)	8,101(60.0)	
	No facility	843 (10.5)	584 (10.7)	1,427 (10.6)	
Cooking energy	Wood	6,370 (79.1)	4,237 (77.9)	10,607 (78.6)	0.112
	Charcoal	1,686 (20.9)	1,200 (22.1)	2,886 (21.4)	
Use water source	Protected	6,097 (75.7)	4,179 (76.9)	10,276 (76.2)	0.115
	Unprotected	1,959 (24.3)	1,258 (23.1)	3,217 (23.8)	
Season Effect	Dry	3,436 (42.6)	2,905 (53.4)	6,341 (47.0)	< 0.001
	Wet	4,620 (57.4)	2,532 (46.6)	7,152 (53.0)	

IP: Intestinal Parasite

LR Model Estimates for Risk Factors of Childhood ARIs Disease

We used all the characteristics described to fit four classifiers methods using ten-fold cross-validation on 75% of the sample size to find the potential risk factors associated with ARIs disease among children below five years of age in Uganda. The estimated results of the stepwise LR model and its corresponding Adjusted Odds Ratios (AOR) and 95% confidence intervals were presented in Table 6. By remaining other factors constant, the risk of developing ARIs illness was found higher (AOR=1.27; 95% CI: 1.12-1.44) in One year children compared to the infants (0-11 months old). As the child age increased, the risk of developing ARIs disease symptoms declined. Children in regions such as the eastern, northern, and western parts of Uganda showed less likely (AOR=0.63; AOR=0.65; AOR=0.51 respectively) to develop disease symptoms compared to children in the central region of the country while children born from teen mothers (15-19 years) showed high risk (AOR=1.28; 95% CI: 1.06-1.53) of developing ARIs illness compared to children born from mothers aged 20-24 years old.

In this study, children whose working mothers were found to be at high risk (AOR=1.25; 95% CI: 1.11-1.42 farmers and AOR=1.93; 95% CI: 1.71-2.19 other employment) of developing ARIs disease compared to children whose mothers had enough time for caring (unemployed) while those whose mothers breastfed were lower likely (AOR=0.83; 95% CI: 0.76-0.92) to develop ARIs illness compared to

those who were not breastfed. The analysis showed that in the wet season, the risk of having ARIs disease in children declined (AOR=0.66; 95% CI: 0.61-0.72) compared to the dry season. Also, children from households whose domestic cooking energy type was charcoal showed a lower risk of developing ARIs symptoms (AOR=0.77; 95% CI: 0.69-0.87) compared to households that used wood as their cooking energy.

Table 6 LR Model Estimates of Potential Risk Factors Associated with ARIs among UnderFive Children in Uganda

Risk Factors		n (%)	AOR [95% C.I]	Sig.
Socio-economic and Demographic Factors				
Child Age (months)	0-11	2,131 (21.1)	Ref	-
	12-23	2,070 (20.4)	1.27 [1.12 - 1.44]	< 0.001
	24-35	2,003 (19.8)	0.98 [0.85 - 1.12]	0.727
	36-47	1,939 (19.2)	0.88 [0.76 - 1.01]	0.065
	48-59	1,977 (19.5)	0.69 [0.60 - 0.80]	< 0.001
	Region	Central	2,064 (20.4)	Ref
Eastern		2,877 (28.4)	0.63 [0.56 - 0.72]	< 0.001
Northern		2,535 (25.1)	0.65 [0.57 - 0.74]	< 0.001
Western		2,644 (26.1)	0.51 [0.44 - 0.57]	< 0.001
Mother Age (years)	15-19	595 (5.9)	1.28 [1.06 - 1.53]	0.009
	20-24	2,736 (27.0)	Ref	-
	25-29	2,586 (25.6)	1.07 [0.95 - 1.19]	0.268
	30-34	2,067 (20.4)	0.94 [0.83 - 1.06]	0.313
	35-39	1,334 (13.2)	0.96 [0.84 - 1.11]	0.600
	40-49	802 (7.9)	0.95 [0.80 - 1.12]	0.546
Mother Employment	Unemployed	1,624 (16.1)	Ref	-
	Farmer	4,780 (47.2)	1.25 [1.11 - 1.42]	< 0.001
	Employed	3,716 (36.7)	1.93 [1.71 - 2.19]	< 0.001
Behavioral and Environmental Factors				
Breastfeeding	No Breastfeed	4,247 (42.0)	Ref	-
	Breastfeed	5,873 (58.0)	0.83 [0.76 - 0.92]	< 0.001
Cooking energy	Wood	7,955 (78.6)	Ref	-
	Charcoal	2,165 (21.4)	0.77 [0.69 - 0.87]	< 0.001
Season Effect	Dry	4,769 (47.1)	Ref	-
	Wet	5,351 (52.9)	0.66 [0.61 - 0.72]	< 0.001
Model Performance Metrics:				
Training Sample Size = 10,120 (75%)				
AIC = 13,218				
Hosmer-Lemeshow Test = 8.951 ($p = 0.346$)				
Likelihood Ratio Test = 464.05 ($p < 0.001$)				

EN Model Estimates for Risk Factors of childhood ARIs disease

As shown in Table 7, the analysis results for EN model estimates revealed similar potential risk factors of ARIs disease as the LR model in Table 6. Risk factors contributing to the childhood ARIs such as the child and mother ages, the regional effect, the employment of mother, and seasonal effect were significantly associated with ARIs disease cases among under-five children in Uganda. In the EN regression model, the coefficients or effects of none contributing factors were shrunk to zero using a ten-fold cross-validation strategy to select the best tune to impose the penalty to the model. In this study, we preferred to consider the effects of predicted

risk factors in the LR regression model since the LR method showed high predictive performance than the EN method as presented in Table 8.

Table 7 EN Model Estimates of Potential Risk Factors Associated with ARIs among Under-Five Children in Uganda

Risk Factors		n (%)	Coeff.	AOR	Sig.
Child Age (months)	0-11	2,131 (21.1)	Ref	-	-
	12-23	2,070 (20.4)	0.16	1.18	< 0.001
	24-35	2,003 (19.8)	0	1	-
	36-47	1,939 (19.2)	0	1	-
	48-59	1,977 (19.5)	-0.11	0.89	< 0.001
Region	Central	2,064 (20.4)	Ref	-	-
	Eastern	2,877 (28.4)	0	1	-
	Northern	2,535 (25.1)	0	1	-
	Western	2,644 (26.1)	-0.22	0.80	< 0.001
Mother Age (years)	15-19	595 (5.9)	0.03	1.03	< 0.001
	20-24	2,736 (27.0)	Ref	-	-
	25-29	2,586 (25.6)	0	1	-
	30-34	2,067 (20.4)	0	1	-
	35-39	1,334 (13.2)	0	1	-
	40-49	802 (7.9)	0	1	-
Mother Employment	Unemployed	1,624 (16.1)	Ref	-	-
	Farmer	4,780 (47.2)	0	1	-
	Employed	3,716 (36.7)	0.27	1.31	< 0.001
Season Effect	Dry	4,769 (47.1)	Ref	-	-
	Wet	5,351 (52.9)	-0.25	0.78	< 0.001

Comparison of Methods Performances

Two recent supervised tree-like structure machine learning (ML) algorithms, such as Decision Tree (DT) and Random Forest (RF), were also analyzed and compared with LR and EN models according to their classification performances. Table 8 shows that three methods such as LR, EN, and DT had approximately the same performances in various measures. However, the RF method reveals better performances in predicting ARIs disease outcomes compared to the other three classifiers considered in this study. The RF technique showed a total accuracy of 88.7% (AUC=0.951) where 88.6% of ARIs cases (recall) and 88.9% of ARIs controls (selectivity) were correctly diagnosed with a precision performance of 93.1%. The LR method showed less performance 26.6% (AUC=0.639) and 5.1% in total accuracy and precision, respectively, compared to the RF technique in classifying childhood ARIs disease outcomes. Only 63.0% of ARIs cases and 57.0% of ARIs controls were correctly diagnosed by the LR method. The EN (AUC=0.627) and DT (AUC=0.610) methods of classifications revealed lower classification performance in predicting under-five children ARIs outcomes in Uganda as shown in Figure 4.

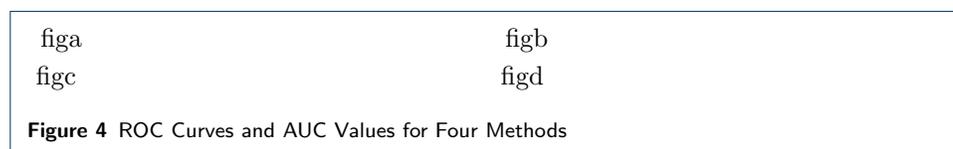


Table 8 Comparison of Classification Performances for Methods

Method	Recall	Selectivity	Precision	Accuracy	AUC
Logistic Regression (LR)	63.05%	57.04%	88.03%	62.05%	0.638
Elastic Net (NE)	63.14%	55.47%	86.25%	61.73%	0.627
Decision Tree (DT)	63.34%	53.33%	82.97%	61.16%	0.610
Random Forest (RF)	88.57%	88.93%	93.10%	88.70%	0.951

Predictive Potential Risk Factors of Childhood ARIs Disease

Table 9 considered potential and important risk factors most contributing in the RF and LR classification methods in predicting childhood ARIs disease among under-five children in Uganda. The risk factors such as the employment or occupation of the mother, the cooking energy materials like wood, the climate season effect, and the region of residence for the child were considered important risk factors in classifying ARIs disease outcomes in both methods that showed higher performances in this study.

Table 9 Potential Risk Factors of Childhood ARIs Disease by RF and LR Methods

Place of Importance	Important Variables and Features	
	Random Forest (RF)	Logistic Regression (LR)
1	Mother Employment (*)	Region (*)
2	Season Effect (*)	Mother Employment (*)
3	Region (*)	Season Effect (*)
4	Cooking Energy (*)	Child Age
5	Parental Wealth	Cooking Energy (*)
6	Place of Delivery	Breastfeeding
7	Mother Education	Mother Age
Accuracy Performance	88.70%	62.05%

(*) Common in both classifiers

Discussion

Our study aimed to identify and analyze the child and parental socioeconomic, demographic, behavioral, and environmental-related risk factors of acute respiratory infections among under-five children in Uganda by using traditional and new analysis methods. Our study and other studies' findings revealed that various risk factors of acute respiratory infections in childhood could differ across geographic, social life, and demographic reasons, and they apply traditional methods like the Logistic Regression method to analyze ARIs disease and predictions. The findings of the study showed that various risk factors were associated with childhood ARIs disease. Other studies conducted in different places and countries across the world also found similar results.

A finding such as a season factor is found associated with childhood ARIs and is also found consistent with other studies conducted in neighboring countries of Uganda. A study conducted in Rwanda [15] found that the rainy season was a risk factor of ARIs similar to the study conducted about the tropical climates where also rain season can increase the risk of ARIs due to the exposure to other people [24]. The effect of the season risk factor is consistent with another study conducted about influenza-associated severe ARIs risk factors in Uganda [17]. The risk factor of ARIs illness, like geographical region factor, is in concurrence with the finding found in other studies conducted in Rwanda and Uganda [15, 17] where the child's

place of residence was associated with ARI disease. In addition, it corroborates with findings found in the study conducted in Gulu district, northern Uganda, about the use of antibacterial in managing ARI symptoms [25].

The employment or occupation of the child's mother and the cooking materials can influence the child's risk of having ARI illness. Our study results showed to be consistent with results found in Rwanda [15] since the majority in both countries' households share a similar way of living and use firewood or charcoal as their primary sources of energy in cooking. In addition, the majority of mothers are being employed where they do not have enough time to take care of their children. Similar results were reported in Cameroun country where the smoke from exposure to wood was highly associated with childhood ARI [26]. The results also compromise with the finding in the earlier study conducted in Ethiopia about risk factors of ARI in children below five years of age [27].

A recent cross-sectional study by Lanyero H. et al. conducted on 856 under-five children in Gulu, northern Uganda, found that risk factors such as health facility type, peri-urban area, and cough in children were associated with symptoms of ARI [25]. A 2019 case-control study conducted in southern Tigray in Ethiopia revealed that malnutrition, presence of a smoker in the family, maternal literacy, and cow dung use could contribute to the childhood ARI illness in children age under five years [28]. In addition, some studies showed that peoples with diseases like HIV infection and other uncommunicable diseases were at high risk of being killed by ARI [26]. The study also showed that contributing factors such as maternal education (below secondary level), passive smoking, and exposure to wood smoke were associated with childhood ARI in Cameroon.

In 2019, a study conducted by M.I.K. Imran et al. [29] on 7,032 Bangladesh under-five children showed that stunted under-five girls children in both infants and 24-35 months old from lower and middle wealth families were highly associated with ARI prevalence in Bangladesh. A 2020 matched case-control study conducted in rural areas of Legambo District, South Wollo Zone, Ethiopia, also revealed that factors such as the mother's age and occupation, the family wealth status, child nutritional status, the fuel type, and smokes contributed to the ARI disease by using multivariable conditional logistic regression analysis [27]. Another multilevel logistic regression study of lifestyles in Nigeria confirmed that lifestyle factors such as hand-washing, cooking materials, and country regions contribute to ARI illness among under-five children [30].

In our study, we also compared the various methods performances in predicting and classifying ARI disease outcomes by considering traditional supervised regression and new machine learning (ML) algorithms approaches where we found that the Random Forest (RF) and Logistic Regression (LR) methods performed well in classifying childhood ARI disease compared to the other techniques. These machine learning (ML) techniques are commonly applied in disease diagnosis and prediction of disease outcomes such as heart diseases, diabetes, liver diseases, lung cancer, etc. However, no study has been conducted yet using the DHS dataset and these new ML algorithms to classify and predicting ARI disease. The RF method performance found in this study compromise with various medical research findings for diseases classification.

A comparative study conducted for predicting the effective diagnosis of liver diseases using comparisons of methods such as Neural Network, Support Vector Machine, Random Forest, and Decision Tree, revealed that the Neural Network and Random Forest methods achieved the highest accuracy rate [31]. Another study carried out for predicting liver disorders showed that the Random Forest algorithm also performed well compared to the other methods employed in the study [32]. A study conducted by Marikani and Shyamala (2017) using clinical and demographic data revealed that RF (Accuracy = 96.3%) performed better in predicting heart disease [33], also similar to the Mani et al. (2012) study conducted using Demographic and clinical test results in predicting type 2 diabetes [34]. However, few studies showed that the Random Forest and Logistics Regression techniques were not the best methods in disease classification like the Islam et al. (2018) study conducted on liver disease classification [35], and another Tapak et al. (2013) study conducted in Iran for predicting diabetes using real dataset comprised of demographic, anthropometric, diagnostic and clinical lab measurement data [36].

The results and findings for our study and various literature reviews showed that the performance accuracies of classification methods varied according to disease of interest and the type of dataset used. However, Random Forest and Support Vector Machine data mining techniques revealed consistency in better performances than other classification methods used in disease classifications and predictions.

Conclusion

This case-control study identified and analyzed various child and parental risk factors of childhood ARIs disease by studying comparative predictive performances of different supervised regressions and machine learning algorithms in ARIs disease prediction and classification. In dealing with medical data, proficiency and experience were required in disease diagnosis in medical research. The appropriate application of the new methods and tools like the Random Forest technique showed better results in disease diagnosis than the traditional methods for disease classification and predictions. In Uganda, Acute Respiratory Infections (ARIs) remain the leading cause of childhood morbidity and mortality among under-five children. The ARIs prevalence accounted for 40.3% of cases that burden the country's health-care providers and families. Many studies and literature found that families living in poverty and low living standards were more likely to be killed by ARIs diseases in countries like Uganda, particularly in the sub-Saharan Africa region where health facilities and development were still under. Identifying potential risk factors contributing to the country's ARIs illness among under-five children help decision-makers in taking measures and precautions to reduce the ARIs disease prevalence associated with childhood morbidity and mortality.

Acknowledgements

Firstly, we express our sincere appreciation and gratitude to the Demographic and Health Survey (DHS) program for its data provision used to conduct this study. We also gratefully thank the East China Normal University advisory committee members for their continuous support, motivation, and advice from the study's progress to completion.

Funding

The work was supported by National Natural Science Foundation of China (project number: 11771146, 11831008), which sponsored the data analysis part of the study.

Abbreviations

ARIs: Acute Respiratory Infections; UDHS: Uganda Demographic and Health Survey; EA: Enumeration Area; LR: Logistic Regression; EN: Elastic Net Regression; DT: Decision Tree; RF: Random Forest; AOR: Adjusted Odds Ratio; CI: Confidence Interval; TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative.

Availability of data and materials

The 2016 Uganda DHS dataset used in this study are openly available and can be found in the Demographic and Health Surveys (DHS) repository, please login at:
https://www.dhsprogram.com/data/dataset_admin/login_main.cfm

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of the School of Statistics, East China Normal University (ECNU), China. The access and use permission of the dataset used during the current study are approved by The Demographic and Health Surveys (DHS) Program.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Authors' contributions

First Author: Analysed and extracted the data, interpreted results and drafted the manuscript. Second Author: Designed and proposed the study methodology, and supervised the whole study process. All authors have read and approved the final version of the manuscript.

Author details

¹School of Statistics, East China Normal University, Shanghai, China. ²KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, China.

References

1. UNDP: Measuring Human Development: a Primer. United Nations Development Programme (UNDP), New York, USA (2007)
2. UNIGME: Levels Trends in Child Mortality: Report 2020, Estimates Developed by the United Nations Inter-agency Group for Child Mortality Estimation. United Nations Inter-agency Group for Child Mortality Estimation (UNIGME), New York, USA (2020)
3. WHO: The World Health Report 2003: Shaping the Future. World Health Organization (WHO), Geneva 27, Switzerland (2003)
4. WHO: Infection Prevention and Control of Epidemic- and Pandemic-prone Acute Respiratory Infections in Health Care. World Health Organization (WHO), Geneva 27, Switzerland (2014)
5. WHO: The top 10 causes of death. World Health Organisation (WHO) (2020).
<https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> Accessed 25 Feb 2021
6. WHO: Diarrhoeal disease. World Health Organisation (WHO) (2017).
<https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease> Accessed 03 March 2021
7. WHO, et al.: Ending preventable child deaths from pneumonia and diarrhoea by 2025: the integrated global action plan for pneumonia and diarrhoea (gappd) (2013)
8. Wardlaw, T.M., Johansson, E.W., Hodge, M.J.: Pneumonia: the Forgotten Killer of Children. Unicef, Geneva 27, Switzerland (2006)
9. UBOS, ICF: Uganda Demographic and Health Survey 2016. Demographic and Health Survey (DHS) & Uganda Bureau of Statistics (UBOS), Kampala, Maryland (2018)
10. Dicker, D., Nguyen, G., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., et al.: Global, regional, and national age-sex-specific mortality and life expectancy, 1950–2017: a systematic analysis for the global burden of disease study 2017. *The lancet* **392**(10159), 1684–1735 (2018)
11. Sridevi Radhakrishnan, D.D.: A critical study on data mining techniques in healthcare dataset (2015)
12. Kirubha, V., Priya, S.M.: Survey on data mining algorithms in disease prediction. *International Journal of Computer Trends and Technology* **38**(3), 124–128 (2016)
13. Chiao, C., Deji-Abiodun, O.: A global analysis of the regional variation in the symptoms of acute respiratory infection during childhood: Epidemics and their association with environmental vulnerability. *Health & Place* **65**, 102400 (2020)
14. Seidu, A.-A., Dickson, K.S., Ahinkorah, B.O., Amu, H., Darteh, E.K.M., Kumi-Kyereme, A.: Prevalence and determinants of acute lower respiratory infections among children under-five years in sub-saharan africa: evidence from demographic and health surveys. *SSM-population health* **8**, 100443 (2019)
15. Harerimana, J.-M., Nyirazinyoye, L., Thomson, D.R., Ntaganira, J.: Social, economic and environmental risk factors for acute lower respiratory infections among children under five years of age in rwanda. *Archives of Public Health* **74**(1), 1–7 (2016)
16. Bbaale, E.: Determinants of diarrhoea and acute respiratory infection among under-fives in uganda. *The Australasian medical journal* **4**(7), 400 (2011)
17. Cummings, M.J., Bakamutumaho, B., Kayiwa, J., Byaruhanga, T., Owor, N., Namagambo, B., Wolf, A., Wamala, J.F., Morse, S.S., Lutwama, J.J., et al.: Epidemiologic and spatiotemporal characterization of influenza and severe acute respiratory infection in uganda, 2010–2015. *Annals of the American Thoracic Society* **13**(12), 2159–2168 (2016)

18. Hosmer Jr, D.W., Lemeshow, S., Sturdivant, R.X.: Applied Logistic Regression vol. 398. John Wiley & Sons, New Jersey, USA (2013)
19. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)* **67**(2), 301–320 (2005)
20. Quinlan, J.R.: Induction of decision trees. *Machine learning* **1**(1), 81–106 (1986)
21. Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V., Fotiadis, D.I.: Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal* **13**, 8–17 (2015)
22. Breiman, L.: Random forests. *Machine learning* **45**(1), 5–32 (2001)
23. Spencer, R., Thabtah, F., Abdelhamid, N., Thompson, M.: Exploring feature selection and classification methods for predicting heart disease. *Digital health* **6**, 2055207620914777 (2020)
24. Murray, E., Klein, M., Brondi, L., McGowan, J., Van Mels, C., Brooks, W.A., Kleinbaum, D., Goswami, D., Ryan, P., Bridges, C.: Rainfall, household crowding, and acute respiratory infections in the tropics. *Epidemiology & Infection* **140**(1), 78–86 (2012)
25. Lanyero, H., Eriksen, J., Obua, C., Stålsby Lundborg, C., Nanzigu, S., Katureebe, A., Kalyango, J.N., Ocan, M.: Use of antibacterials in the management of symptoms of acute respiratory tract infections among children under five years in gulu, northern uganda: Prevalence and determinants. *PLoS one* **15**(6), 0235164 (2020)
26. Tazinya, A.A., Halle-Ekane, G.E., Mbuagbaw, L.T., Abanda, M., Atashili, J., Obama, M.T.: Risk factors for acute respiratory infections in children under five years attending the bamenda regional hospital in cameroon. *BMC pulmonary medicine* **18**(1), 1–8 (2018)
27. Hassen, S., Getachew, M., Eneyew, B., Keleb, A., Ademas, A., Berihun, G., Berhanu, L., Yenuss, M., Natnael, T., Kebede, A.B., *et al.*: Determinants of acute respiratory infection (ari) among under-five children in rural areas of legambo district, south wollo zone, ethiopia: A matched case–control study. *International Journal of Infectious Diseases* **96**, 688–695 (2020)
28. Alemayehu, S., Kidanu, K., Kahsay, T., Kassa, M.: Risk factors of acute respiratory infections among under five children attending public hospitals in southern tigray, ethiopia, 2016/2017. *BMC pediatrics* **19**(1), 1–8 (2019)
29. Imran, M., Inshafi, M., Sheikh, R., Chowdhury, M., Uddin, M.: Risk factors for acute respiratory infection in children younger than five years in bangladesh. *Public health* **173**, 112–119 (2019)
30. Adesanya, O.A., Chiao, C.: A multilevel analysis of lifestyle variations in symptoms of acute respiratory infection among young children under five in nigeria. *BMC public health* **16**(1), 1–11 (2016)
31. Bihter, D.: A comparative study on the performance of classification algorithms for effective diagnosis of liver diseases. *Sakarya University Journal of Computer and Information Sciences* **3**(3), 366–375 (2020)
32. Pathan, A., Mhaske, D., Jadhav, S., Bhondave, R., Rajeswari, K.: Comparative study of different classification algorithms on ilpd dataset to predict liver disorder. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)* **6**(2), 388–394 (2018)
33. Marikani, T., Shyamala, K.: Prediction of heart disease using supervised learning algorithms. *Int J Comput Appl* **165**(5), 41–4 (2017)
34. Mani, S., Chen, Y., Elasy, T., Clayton, W., Denny, J.: Type 2 diabetes risk forecasting from emr data using machine learning. In: *AMIA Annual Symposium Proceedings*, vol. 2012, p. 606 (2012). American Medical Informatics Association
35. Islam, M.M., Wu, C.-C., Poly, T.N., Yang, H.-C., Li, Y.-C.J.: Applications of machine learning in fatty liver disease prediction. In: *MIE*, pp. 166–170 (2018)
36. Tapak, L., Mahjub, H., Hamidi, O., Poorolajal, J.: Real-data comparison of data mining methods in prediction of diabetes in iran. *Healthcare informatics research* **19**(3), 177 (2013)

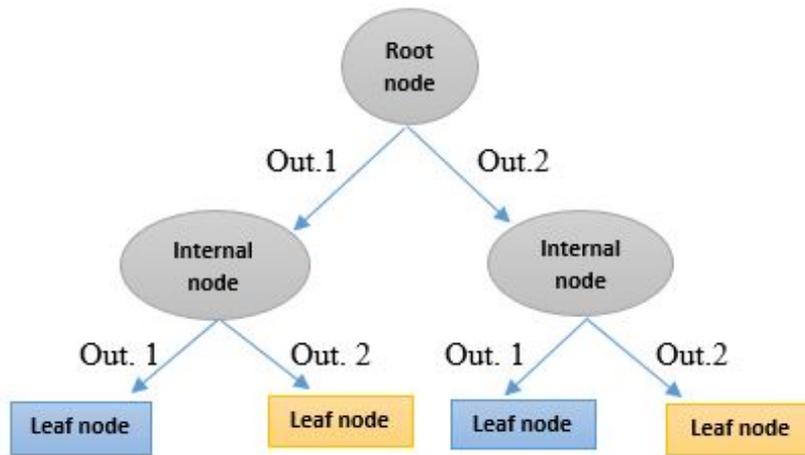


Figure 1: A Tree-like Structure of the Decision Tree Method

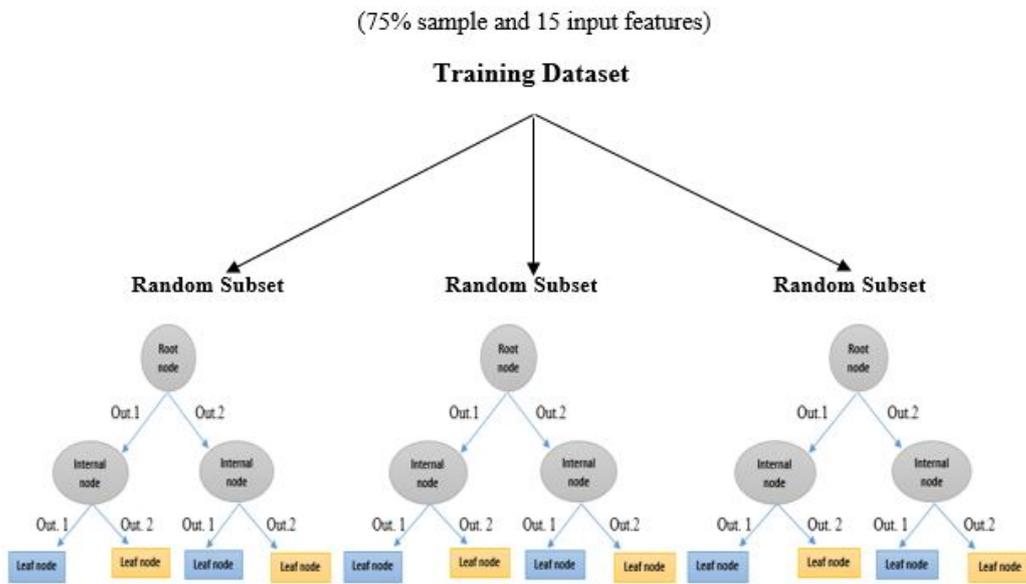


Figure 2: A Tree-like Structure of the Random Forest (RF) Method

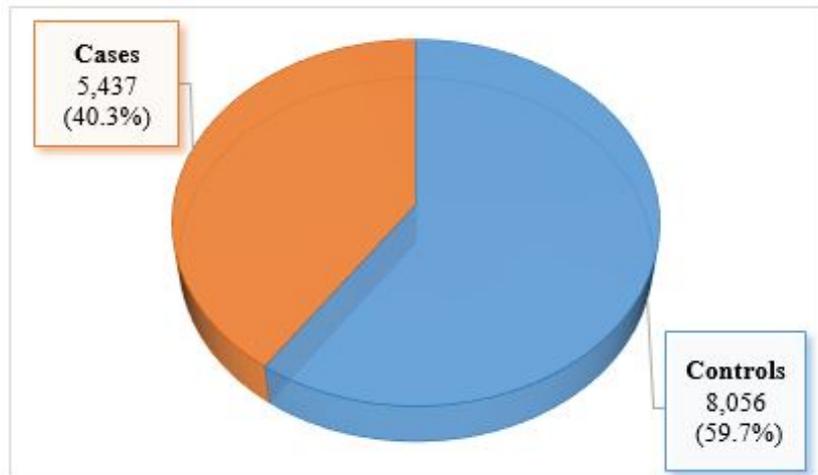


Figure 3: The ARIs Prevalence among Under-Five children in Uganda

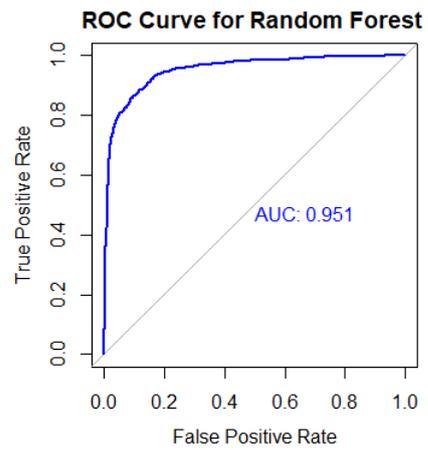
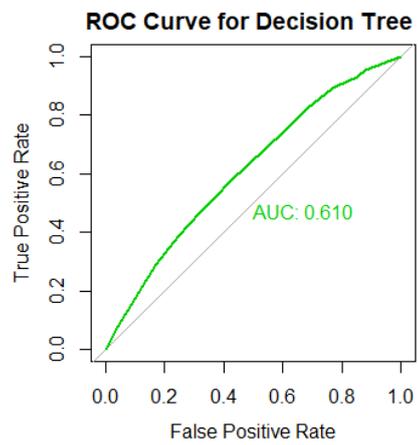
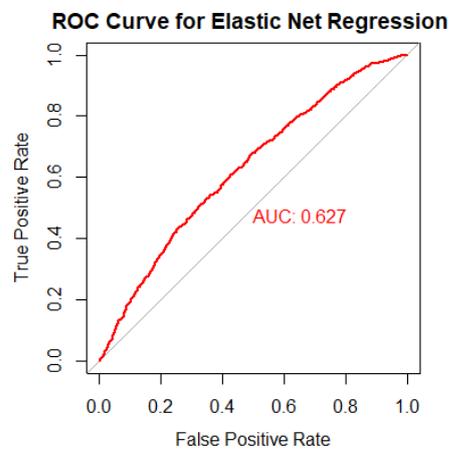
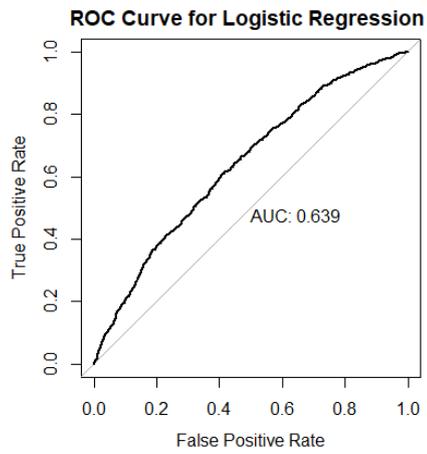


Figure 4: ROC Curves and AUC Values for Four Methods

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [bmcarticle.bib](#)
- [bmcarticle.tex](#)
- [bmcartbiblio.sty](#)
- [bmcart.cls](#)