

RESEARCH ARTICLE

Machine learning algorithms' application to predict childhood vaccination among children aged 12–23 months in Ethiopia: Evidence 2016 Ethiopian Demographic and Health Survey dataset

Addisalem Workie Demsash^{1*}, Alex Ayenew Chereka¹, Agmasie Dامتew Walle¹, Sisay Yitayih Kassie¹, Firomsa Bekele², Teshome Bekana³

1 Department of Health Informatics, College of Health Science, Mettu University, Mettu, Ethiopia,

2 Department of Pharmacy, College of Health Science, Mettu University, Mettu, Ethiopia, **3** Biomedical Science Department, College of Health Science, Mettu University, Mettu, Ethiopia

* addisalemworkie599@gmail.com



OPEN ACCESS

Citation: Demsash AW, Chereka AA, Walle AD, Kassie SY, Bekele F, Bekana T (2023) Machine learning algorithms' application to predict childhood vaccination among children aged 12–23 months in Ethiopia: Evidence 2016 Ethiopian Demographic and Health Survey dataset. PLoS ONE 18(10): e0288867. <https://doi.org/10.1371/journal.pone.0288867>

Editor: Engidaw Fentahun Enyew, University of Gondar, ETHIOPIA

Received: April 6, 2023

Accepted: July 6, 2023

Published: October 18, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0288867>

Copyright: © 2023 Demsash et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The dataset used for analysis is available on the DHS program website.

Abstract

Introduction

Childhood vaccination is a cost-effective public health intervention to reduce child mortality and morbidity. But, vaccination coverage remains low, and previous similar studies have not focused on machine learning algorithms to predict childhood vaccination. Therefore, knowledge extraction, association rule formulation, and discovering insights from hidden patterns in vaccination data are limited. Therefore, this study aimed to predict childhood vaccination among children aged 12–23 months using the best machine learning algorithm.

Methods

A cross-sectional study design with a two-stage sampling technique was used. A total of 1617 samples of living children aged 12–23 months were used from the 2016 Ethiopian Demographic and Health Survey dataset. The data was pre-processed, and 70% and 30% of the observations were used for training, and evaluating the model, respectively. Eight machine learning algorithms were included for consideration of model building and comparison. All the included algorithms were evaluated using confusion matrix elements. The synthetic minority oversampling technique was used for imbalanced data management. Informational gain value was used to select important attributes to predict childhood vaccination. The If/ then logical association was used to generate rules based on relationships among attributes, and Weka version 3.8.6 software was used to perform all the prediction analyses.

Results

PART was the first best machine learning algorithm to predict childhood vaccination with 95.53% accuracy. J48, multilayer perceptron, and random forest models were the

All the data generated and analyzed are included in the study.

Funding: The authors received no specific funding for this work.

Competing interests: The author declared that there are no competing interests in this work.

Abbreviations: ANC, Antenatal care; AUC, Area under the recursive curve; BCG, Bacillus Calmette Guerin; EDHS, Ethiopian Demographic and Health Survey; ROC, Receiver operators' curve; WHO, World Health Organization; DTP, Diphtheria Pertussis Tetanus.

consecutively best machine learning algorithms to predict childhood vaccination with 89.24%, 87.20%, and 82.37% accuracy, respectively. ANC visits, institutional delivery, health facility visits, higher education, and being rich were the top five attributes to predict childhood vaccination. A total of seven rules were generated that could jointly determine the magnitude of childhood vaccination. Of these, if wealth status = 3 (Rich), adequate ANC visits = 1 (yes), and residency = 2 (Urban), then the probability of childhood vaccination would be 86.73%.

Conclusions

The PART, J48, multilayer perceptron, and random forest algorithms were important algorithms for predicting childhood vaccination. The findings would provide insight into childhood vaccination and serve as a framework for further studies. Strengthening mothers' ANC visits, institutional delivery, improving maternal education, and creating income opportunities for mothers could be important interventions to enhance childhood vaccination.

Introduction

Globally, nearly 44% of child deaths occurred under 28 days of birth [1]. Around 75% of child deaths occur within 12 months of birth, and an estimated 4.1 million infants are projected to die in 2017 [2]. The rate of child deaths in developing countries is the highest in the world [2,3]. Around 1.2 million children are predicted to have died in Africa in the first 28 days of birth [4], and nearly 49% of child deaths are predicted to have occurred in Sub-Saharan countries [3]. According to the World Health Organization (WHO), more than half of child deaths are caused by infectious diseases that are easily preventable and treatable through simple and affordable interventions [5]. Worldwide, childhood mortality and morbidity are caused by tuberculosis, diphtheria, pertussis, tetanus, polio, and measles [6].

Child deaths due to diphtheria, pertussis, tetanus, polio, and measles are easily preventable through vaccines. Childhood vaccination is one of the most successful and cost-effective public health interventions for common childhood illnesses like pneumonia, diphtheria, tetanus, whooping cough, and measles [7]. Nowadays, nearly 3 million child deaths due to diphtheria, tetanus, whooping cough, and measles are prevented through child vaccination [8]. Over the past decade, more than 1,000,000 children's lives have been saved by immunization programs and infectious and communicable diseases have been controlled through child vaccination [9].

Nonetheless, 12.9 million children did not receive recommended vaccines across the world [8]. Sufficient numbers of children did not complete their immunization schedule due to various challenges and barriers [10]. Nearly 21 million children have been projected to miss out on vaccines, and two-thirds of vaccination missing occurred in developing regions due to the outbreaks of new cases [11]. Nearly 21 million children have been projected to miss out on vaccines, and two-thirds of the missing vaccinations occurred in developing regions due to an outbreak of new cases [12]. In Ethiopia, infant vaccination doses were usually delayed, with 63.8 percent of Diphtheria Pertussis Tetanus (DTP) dose 1, 63.1 percent of Polio dose 1, and 68.5 percent of measles delivered after the recommended date [13]. According to the Ethiopia Demographic and Health Survey (EDHS), data on vaccination coverage among children aged 12–23 months who received specific vaccines at any time before the survey revealed that only four out of ten children (43%) had received all basic vaccinations [14]. According to the WHO, the mean dropout rates of Bacillus Calmette–Guérin (BCG) and measles are 34.6% and 28.6%, respectively [15].

According to a traditional and multilevel logistic regression analysis report, different factors are reported that could affect child vaccination and immunization coverage. Maternal education, knowledge of mothers about the vaccines and their schedule, maternal age, fear of side effects, antenatal care (ANC) visits, and giving birth at a health institution are some of the maternally related factors that affect childhood vaccinations [16–19]. Additionally, the availability of vaccines, migration of caregivers, household income level, and sex of household heads are factors that affect childhood vaccination status [20,21]. Moreover, the sex and age of children, their birth interval and order, multiple children born at a time, a mother's media exposure, being a rural resident, and having distant health facilities are also factors associated with childhood vaccination [22,23]. However, the odds ratio and relative risk of traditional and multilevel logistic regression do not meaningfully classify attributes and do not discover new insights [24].

Despite the efforts of the government to improve child vaccination, increase vaccination coverage, and reduce vaccine dropout rates, vaccine providers and health programmers lack available on-site information handling tools to target high-risk children for vaccine dropout, and late and incomplete vaccination [15]. Therefore, low-income countries would model and visualize the childhood vaccination risks on large datasets to identify attributes for childhood vaccination and target children who are at a high risk of dropping out or delaying the next vaccine dose.

Massive amounts of biomedical and public health data are categorized and predicted using a variety of predictive algorithms to gain new knowledge and reveal hidden relationships and trends [25]. Multidimensional data mining techniques were used to correctly forecast future immunization outcomes based on existing data and to predict features of typical childhood immunization schedules [26]. Predictive analytics tools are potent and widely applicable for learning. Numerous machine learning algorithms have reportedly been used in earlier studies to predict disease prevalence, the use of healthcare services, vaccination uptake [27], routine immunization [15], childhood vaccination, and mortality [28,29]. For automated detection, identifying connections that aren't learner, and identifying significant patterns in data, machine learning algorithms are essential [30].

Specifically, random forests, logistic regression, J48, logit boost, and Addaboost algorithms were used to predict under-five and neonatal mortality [29,31], undernutrition status of children [32], and malnutrition among children [33,34]. Additionally, Naïve Bayes and PART algorithms are also used to forecast and classify text documents [35]. Prediction of childhood vaccination based on machine learning techniques is insufficient. Currently, massive amounts of data are being generated. So, these must be presented with the best data analysis tools. Policymakers and stakeholders need accurate predictions on various aspects of immunization and other health parameters for effective actions. Researchers are needed to test and compare various prediction and classification algorithms that are needed to classify and predict childhood vaccination. Therefore, this study aimed to (1) evaluate different machine learning algorithms using model evaluation matrix parameters; (2) identify important attributes for childhood vaccination based on the best performance algorithm; and (3) generate association rules that predictor together determine the vaccination of children aged 12–23 months in Ethiopia.

Methods and materials

Study design and setting

A cross-sectional study design was conducted across the nine regions of Ethiopia. Ethiopia is located in the Horn of Africa and is bordered by Eritrea to the North, Djibouti and Somalia to the East, Sudan, and South Sudan to the West, and Kenya to the South. Ethiopia has nine

regional states with two administrative cities. These are subdivided into different administrative zones (817 Woredas and 16253 Kebeles) [36,37].

Data source

The 2016 Ethiopian Demographic and Health Survey (EDHS) dataset was used from the DHS program website (<https://dhsprogram.com>). The survey was conducted by the Ethiopian Public Health Institute (EPHI) in collaboration with the Central Statistical Agency (CSA). The actual data collection period was conducted from January 18, 2016, to June 27, 2016.

Sampling techniques and procedures

The sampling frame used for the 2016 EDHS is a frame of all Census Enumeration Areas (EAs) created for the 2016 Ethiopia Population and Housing Census (EPHC) and conducted by the Central Statistical Agency (CSA). The census frame is a complete list of the 84,915 EAs, covering an average of 181 households, created for the 2016 EPHC. The sample for the 2016 EDHS was designed to provide estimates of key indicators for the country as a whole, for urban and rural areas separately, and for each of the nine regions and the two administrative cities. Two-stage stratified cluster sampling was used. Each region was stratified into urban and rural areas. In the selected EAs, a household listing operation was done, and the results were used as a sampling frame for household selection in the second stage. Finally, a fixed number of households per cluster was selected. Samples of EAs were selected independently in each stratum through implicit stratification and equal proportional allocation.

Study populations

In this study, all living children aged 12–23 months were the source population, and all sampled living children aged 12–23 months living with their mothers were the study population. Details about the methodology of the data source, sampling procedure, and source population were presented in the 2016 EDHS report [38].

Study variables

Dependent variable

Childhood vaccination among children aged 12–23 months.

Independent variables

Socio-demographic characteristics of households, such as wealth status, educational status of mothers, age of mother, region, residency, sex, and age of children, birth interval and birth order, sex of households' heads, ANC visit, place of delivery, working status, visiting health facility, and media exposure were used as independent attributes to predict childhood vaccination among children aged 12–23 months in Ethiopia.

Operationalizations

Childhood vaccination

Childhood vaccination among children aged 12–23 months was assessed using one dose of BCG, three doses of polio vaccine, three doses of DPT vaccine, and one dose of measles vaccine. Accordingly, the children had basic childhood vaccination if the children received at least one dose BCG vaccine, three doses of the polio vaccine, three doses of the DPT vaccine, and one dose of the measles vaccine, else children did not receive basic childhood vaccination.

Information on basic childhood vaccination status was obtained from (1) written vaccination record that includes infant immunization card and other health cards, (2) the mothers' verbal reports, and (3) health facility records [38].

Birth interval

The period between two successive live births is a birth interval. For this study, a birth interval of <33 months between two consecutive live births is a **short birth interval**, whereas a birth interval of 33 and above is an **optimum birth interval** [39,40].

ANC visits

The pregnant women had visited a health facility during their pregnancy for ANC services. Accordingly, the women had adequate ANC visits when the women visited the health facility at least four times for ANC services, otherwise inadequate ANC visits [41,42].

Media exposure

If the mothers had access to either radio or television or both, then the mothers had media exposure; and if mothers did not have any means of media access then the mothers had no media exposure.

Data management and statically analysis

Data cleaning and labeling were performed using STATA version 15 software to prepare the data for analysis. Variables were recoded to meet the desired classification. To ensure the representativeness of survey results at the national level [43], sampling weights were applied during the analysis. The STATA version 15 software was used for data management and logistic regression analysis. Weka version 3.8.6 software was used for data pre-processing, important attribute selection that could predict childhood vaccination, and generating rules associated with childhood vaccination.

Ethical approval and consent to participate

Ethical clearance was not necessary for this study since it was based on publicly available data sources. Informed consent from the study participants was also not applicable to this study. There are no attributes that uniquely identify individuals or households in this study. As a result, specific individuals, and households cannot be identified uniquely in this study according to the clinical study checklist (S1 File).

Data pre-processing

Data pre-processing was used to manage missing and incomplete records, and duplicates. In the dataset, noise, outliers, and inconsistency are common. Therefore, all these unnecessary data values, including duplicate variables were managed. At this stage, all strings and categorical variables were transformed into nominal data types for ease of processing in Weka software.

Feature selection

In this study, there were two stages of variable selection in the machine learning algorithm. In the first stage, a logistic regression analysis was employed for a feature or independent variables selection. A variable with a p-value of less than 0.2 with backward stepwise logistic

regression analysis was selected as a candidate for further important attribute selection. During the first phase of variable selection, a variance inflation factor was performed to check the correlation between variables. As a result, a variance inflation factor's value for all possible variables was less than four. Hence, there was no significant correlation between the variables. The Hosmer and Lemeshow tests were also performed to assess the model's fitness. Consequently, the model was fitted with a p-value of 0.263. In the second stage, a best-performance machine learning algorithm with information gain values was used to find important features or attributes that have a major contribution to predicting childhood vaccination among children aged 12–23 months in Ethiopia. The highest information gain value of an independent attribute is the most important attribute to predict childhood vaccination [44]. Then the next important attributes were selected based on their order of highest information gain value.

Model building

Data split and model selection

In this step of the machine learning algorithm, 70% of the datasets were used for training the model, and 30% of the datasets were used for testing the performance of the algorithms. A total of 1617 instances/ observations were included to predict childhood vaccination. From a total of 1617 observations, 1132 observations (70% of total observations) were used for training the model, and the remaining 485 observations (30% of total observations) were used for testing or evaluating the model. Various machine-learning algorithms were used to predict child mortality and health service utilization [25,33,34]. For this study, the various appropriate machine learning algorithms such as Naïve Bayes, PART, logistic regression, multilayer perceptron, J48, logit Boost, random forest, and AdaBoost were used to predict childhood vaccination among children aged 12–23 months in Ethiopia.

Naïve Bayes

The Naïve Bayes algorithm is a supervised machine learning algorithm, which is based on the Bayes theorem and used for the classification and prediction of problems. In the Naïve Bayes algorithm, attributes are conditionally independent for the target class [25]. Naïve Bayes has a computational efficiency that several attributes and classification time is linear with several of several, and not affected by training time. Naive Bayes algorithms had an incremental learning behavior, could directly predict patterns with low variance, and their performance is measured by confusion matrix elements [45].

PART

PART is a hybrid approach of a rule-based classification algorithm, and it uses a separate and conquer classification process [35]. It creates a partial decision tree from all the iterations and considers the suitable leaf into a rule. So, it is best to perform if/ then rules to extract and build knowledge for childhood vaccination [46].

Logistic regression

Logistic regression is a type of regression model that is important to model the categorical dichotomous outcome variable or feature. Logistic regression is a statistical model used to classify and predict different parameters in health [47]. It might be a binary (Binary logistic) and (multiple) model used to predict binary (multiple) outcome variables. Logistic regression has different assumptions, of which the target variable is dichotomous, and independent variables that affect the target variable are indent of each other [48].

J48 classifier algorithm

A J48 classifier algorithm is one of the best machine learning algorithms that examine categorical data based on a top-down recursive divide and conquer strategy [49]. J48 classifier is a simple C4.5 decision tree for classification to create a binary tree. The algorithm is crucial for classifying the problems, and the J48 algorithm is important to ignore the missing values and able to predict the item of missing value based on what is known about the records of another attribute. The process is to divide the available data into ranges based on the attribute values for that item that are found in the training data, and then classification is done and rules are generated from the attributes [50].

Random forest

A random forest is a supervised machine-learning algorithm used to classify and problems health problems and health service utilization [51]. Random forest is the fastest to train and work with subsets of features, and it is important to detect complex relationships, including nonlinear and high-order interactions and yields the smallest prediction errors [52].

Addaboost and logit boost

Addaboost is an ensemble meta-learning method that enhances the efficiency of the binary classification tree. Addaboost uses an iterative approach to learn from the mistakes of weak classifiers and turn them into strong ones [53,54]. AdaBoost is critical to boosting the performance of decision trees based on binary classification problems [55]. Another very powerful boosting classifier algorithm (**logit boost**) was used to predict childhood vaccination in this study. The **logit boost** algorithm is designed as an alternative solution to address the limitations of Addaboost in handling noise and outliers [56].

Features of knowledge flow

The knowledge flow presents a "data-flow" inspired interface in Weka software for data processing and analysis. The knowledge flow can handle data either incrementally or in batches. The features of knowledge flow are initiative of the data flow layout, processing the data in batches or incrementally, processing multiple batches or streams in parallel, chain filtering together, and viewing and visualized model performance with a fold cross-validation [44]. The overall knowledge flow of model building for data processing, analyzing, and visualizing has been presented in Fig 1.

Imbalance data management

Data imbalance mainly occurs in medical diagnosis, pattern recognition, speech, and fraud detection. The dataset might have majority and minority classes in its observation [57]. Therefore, the classification and prediction might be certain to the majority class. In such a case, the minority class might not be considered, and classification and prediction might be inaccurate and biased. Therefore, the synthetic minority over-sampling technique (SMOTE) was used to manage imbalanced data [58]. SMOTE creates new synthetic samples for the minority class by interpolating linearity between the minority class [58,59], and it is critical to address underfitting and overfitting to reduce prediction errors [60]. As a result, a total of 359 additional records were generated and added to the minority class. Overall, the imbalanced data and balanced data are presented in Fig 2.

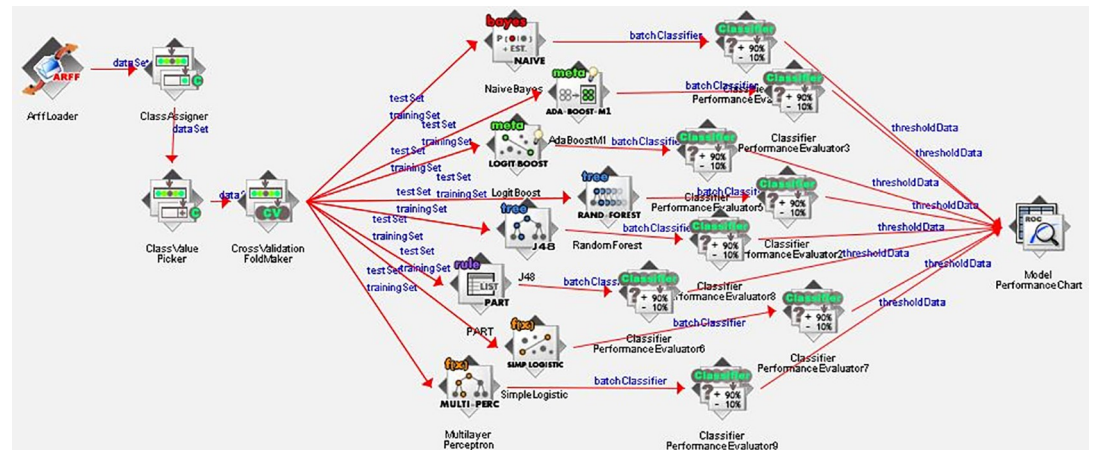


Fig 1. Features of knowledge flow of the included algorithms.

<https://doi.org/10.1371/journal.pone.0288867.g001>

Model evaluation

The performance of all the included algorithms has been evaluated using the confusion matrix. The accuracy of actual and predicted classes has been visualized by the confusion matrix model [61]. The predicted and actual classifications of under-five child mortality were compared using confusion matrix elements, such as true positive, false positive (FP), true negative, and false negative. The receiver operators' curve (ROC) was also used for model evaluation based on sensitivity, and specificity relationships. Since ROC is based on probability, the area under the ROC curve (AUC) is crucial to representing the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. Hence, the higher the AUC, the better the model is at predicting true classes as true and false classes as false. Usually, the AUC value is good if it is greater than 80%, fair if it is between 70% and 80%, poor if it is between 60% and 70%, and failed if it is less than 60% [62]. A metric of interrater agreement i.e. kappa statistics was used to measure the degree of agreement/ reliability and to evaluate the accuracy of a classification. If the Kappa statistics value is ≤ 0 indicating the agreement is worse than random agreement, 0.01–0.20 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and 0.81–1.00 almost perfect agreement [63].

The formula for the confusion matrix's element is presented in **Box 1**.

Box 1. Formula for the element of the confusion matrix.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Sensitivity} = TP / (TP + FN)$$

Not that Sensitivity = Recall = True Positive Rate (TPR)

$$\text{Specificity} = TN / (TN + FP) \quad \text{False positive rate} = FP / (FP + TN)$$

$$\text{F_measure} = 2TP / (2TP + FP + FN)$$

$$\text{Precision} = \text{Postive predictive value} = TP / (TP + FP)$$

Whereas, **TP**: True positive, **TN**: True negative, **FP**: False positive, **FN**: False negative

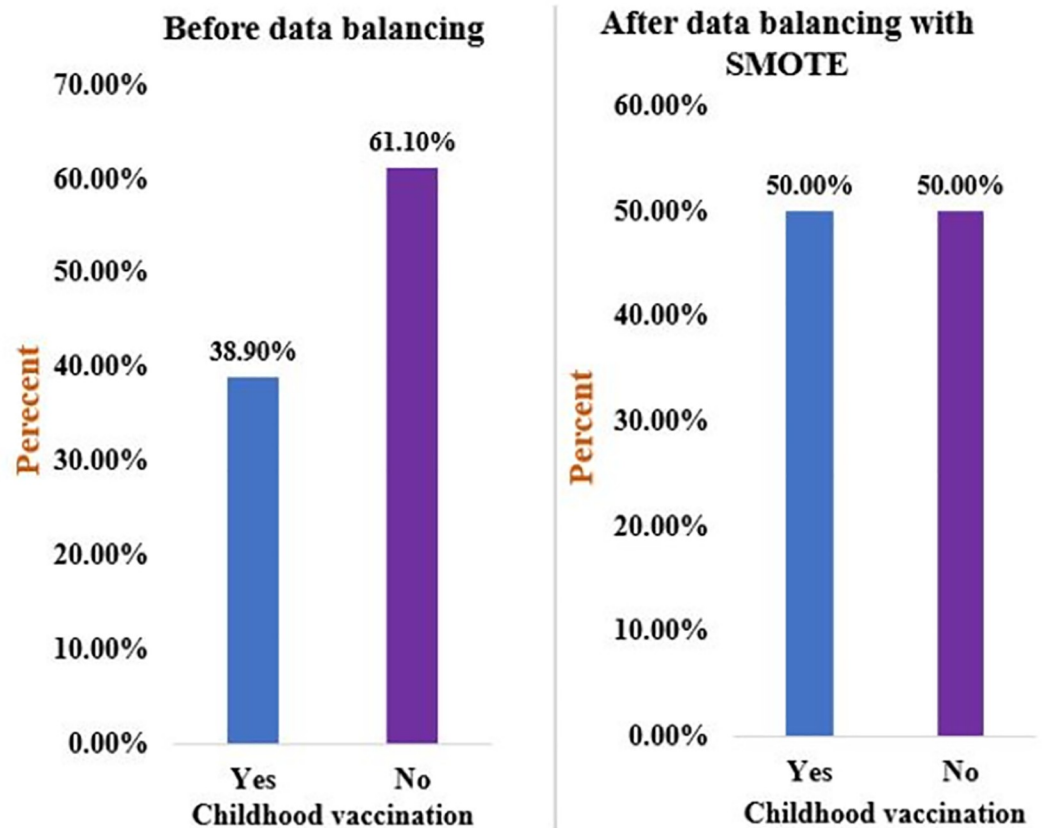


Fig 2. Overall childhood vaccination status among children aged 12–23 months in Ethiopia, before and after data balancing, using the 2019 EDHS dataset.

<https://doi.org/10.1371/journal.pone.0288867.g002>

True positive: The model correctly predicts a positive class of response outcome.

False positive: The model incorrectly predicts a positive class in the response outcome.

True negative: The model correctly predicts a negative class in the response outcome.

False-negative: The model incorrectly predicts a negative class in the response outcome.

Sensitivity: Sensitivity is the test to measure correctly positive predicted events out of a total number of positive events, and it shows the value of how many positives are predicted out of total positive classes.

Specificity: Specificity is the proportion of real negative cases that were predicted as negative. This indicates that there will be another proportion of real negative cases, which would be predicted as positive and could be termed as false positives.

Precision: Precision is a positive predictive value, and it is the correct events divided by the total number of positive events that the classifier predicts.

F_measure: F measure is the inverse relationship between accuracy and recall. The higher value of the F-measure score predicts a better model.

Prediction and association rule mining

Once the model is built and its performance assessed, childhood vaccination among children aged 12–23 months is predicted based on the predictors. Important variables selected based on a best-performance model were used to predict childhood vaccination. Although important

variables are used to predict childhood vaccination, the predictive model does not show which nominal variables are jointly associated with childhood vaccination among children aged 12–23 months.

Therefore, association rule mining analysis (the **If** (antecedent)/ **then** (consequent) statements) is used to discover relationships between seemingly relational attributes. Association rule mining analysis is important for non-numerical and categorical types of data attributes. It is important to observe frequently occurring patterns and identify the dependencies between attributes by supporting how frequently the if/then relationship appears in the observations and confidence in the number of times the relationships are true. The if/ then association rule mining analysis is critically important to select important features that jointly determine childhood vaccination and is the easiest way to interpret [64].

For the association rule mining analysis, the apriori algorithm method was used to identify strong and frequently related attributes. The **If then** association rule is the pair of X and Y (X, Y) attributes expressed as $X \rightarrow Y$, where X is an antecedent and Y a consequent that is as X happens Y would also happen [65]. These rules are critically important for the prevention and control of health problems and crucial for health policymakers' proactive decision-making purposes. Various studies have widely used if/then rules in healthcare research, such as predicting childhood care and child mortality [66], predicting parasite infection [67], the pattern of new cases and stroke [68,69], and maternal healthcare service utilization discontinuation to identify important features [70]. The relationship between X and Y attributes is expressed in the following way [69].

If the left attribute >1 | X and Y are positively associated to determine childhood vaccination. if the left attribute <1 | X and Y negatively associated to determine childhood vaccination.

If the left attribute = 1 | No relation between X and Y to determine childhood vaccination.

The detail of data preparation, model building, important variable selection, and analysis workflow is presented in Fig 3.

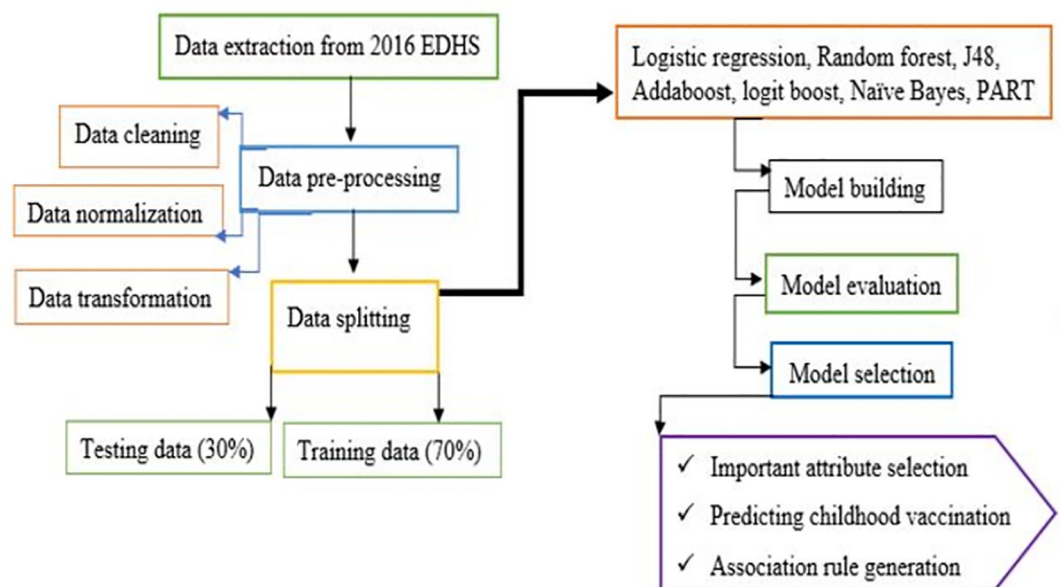


Fig 3. Workflow for data pre-processing, and childhood vaccination prediction processing.

<https://doi.org/10.1371/journal.pone.0288867.g003>

Results

Children's and mothers' characteristics

A total of 1617 weighted samples of children aged from 12–23 months were included for analysis. The majority (62.52%) of children's mothers were under the age of 35 years. The majority (72.5%) of children were born from mothers who had not had formal education. Seven hundred thirty (45.1%) and two hundred eighty-eight (17.8%) children were from the Oromia and Amhara regions, respectively. The majority (91.2%) of the children were born to rural residents' mothers. Five hundred fifty-three (34.2%) and four out of ten (40.3%) of children were born from mothers whose religious were Orthodox and Muslim, respectively. Seven hundred sixty (47%) of children's mothers were poor. Nearest to half (52.9%) and the majority (86.2%) of children and household heads were female and male, respectively. Six hundred seventy-five (41.7%) of children were under the age of 12–15 months ([Table 1](#)).

Children's and mothers' characteristics

Less than half (47.2%) of children visited a health facility in the last 12 months after birth, and the majority (70.6%) of children's mothers had not worked during the time of the interview. Only 29.8% of children's mothers had media exposure, and 29% of mothers had given birth to health institutions. The majority (70.6%) of the mothers did not adequate ANC visits during their pregnancy period. The majority (64.5%) of children had a birth order of less than five, and 65.1% of children had an optimal birth interval ([Fig 4](#)).

Vaccination coverage among children aged 12–23 months in Ethiopia

In Ethiopia, the overall vaccination coverage of children aged 12–23 months was 38.9% (95% CI: 36.52%-41.28%). Specifically, more than half (54.1%) of children received the measles vaccination, and seven out of ten (68.6%) of children had received the BCG vaccine. The majority (72.9%), nearly two-thirds (65.2%), and more than half (53%) of children had received DPT1, DPT2, and DPT3 vaccines, respectively. The majority (80.8%), nearly seven out of ten (72%), and more than half (55.9%) of children aged 12–23 months had received POLIO 1, POLIO 2, and POLIO 3 respectively ([Fig 5](#)).

Models performance to predict childhood vaccination in Ethiopia using 2016 EDHS data

Eight machine learning algorithms were used to predict childhood vaccination in Ethiopia. The PART, Naïve Bayes, logit boost, J48, random forest, addaboost, logistic regression, and multilayer perceptron algorithms were included to predict childhood vaccination. The confusion matrix parameter elements (TPR, FNR, precision, F-measure, AUR, and accuracy) were used to evaluate the performance of the included algorithms. Accordingly, the PART algorithm was the first best performance algorithm to predict childhood vaccination with 95.53% accuracy, and 91.89% of AUC. The Kappa statistics value also confirmed that the classification accuracy of the PART algorithm was almost perfect with 86.57% of accuracy. The j48 algorithm was the second-best machine learning algorithm to predict childhood vaccination with 89.24% accuracy. The 86.01% AUR value also confirmed that the j48 algorithm was the best model next to the PART algorithm, and the classification accuracy of the j48 algorithm had a substantial agreement with 79.27% of kappa statistics. The overall machine learning algorithms comparison for childhood vaccination are presented in [Table 2](#) and [Fig 6](#).

Table 1. Children's and mothers' characteristics, 2016 EDHS data (n = 1617).

Variable	Category	Frequency (n)	Percent (%)
Mothers' educational status	No formal education	1172	72.5
	Primary	377	23.3
	Secondary	39	2.4
	Higher	29	1.8
Region	Tigray	116	7.2
	Afar	16	1.0
	Amhara	288	17.8
	Oromia	729	45.1
	Somali	60	3.7
	Benishangul	18	1.1
	SNNPR	347	21.4
	Gambela	4	.2
	Harari	3	.2
	Addis Ababa	29	1.8
	Dire Dawa	7	.4
Mothers' age (year)	15–34	1011	62.52
	> = 35	606	37.48
Family's wealth index	Poor	760	47.0
	Middle	339	20.9
	Rich	518	32.1
Mother/caregiver religion	Orthodox	553	34.2
	Catholic	26	1.6
	Protestant	346	21.4
	Muslim	652	40.3
	Traditional, and other	40	2.5
Place of residency	Urban	142	8.8
	Rural	1475	91.2
Sex of children	Male	761	47.1
	Female	856	52.9
Sex of household head	Male	1394	86.2
	Female	223	13.8
The current age of children (months)	12–15 months	675	41.7
	16–19 months	519	32.1
	20–23 months	423	26.2

<https://doi.org/10.1371/journal.pone.0288867.t001>

Importance attributes of childhood vaccination in Ethiopia

The information gain coefficients with a 10-cross-fold validation process were used to select important attributes of childhood vaccination in Ethiopia. The best performance model (PART algorithm) was used to select important attributes for childhood vaccination. According to the PART algorithm report, having adequate ANC visits, institutional delivery, visiting health facilities in the last 12 months, higher educational status of mothers, children whose mothers were rich, being of urban residents, female household heads, mothers' age greater than 35 years, having birth order less than five, and mothers currently working were important attributes for childhood vaccination among children aged 12–23 months. The important attributes and their information gain values are presented in [Table 3](#) and [Fig 7](#).

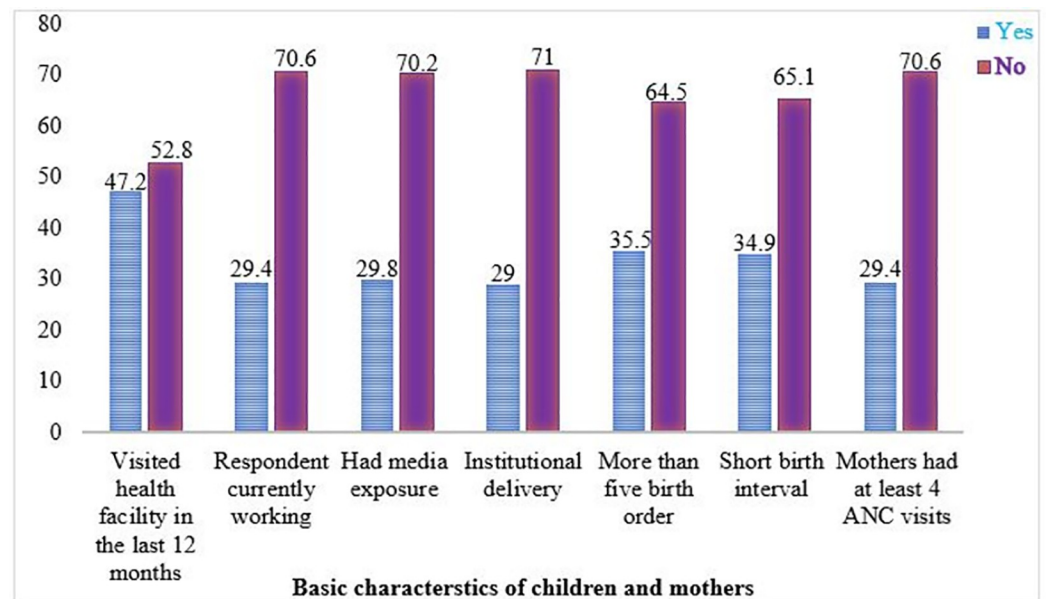


Fig 4. Children's and mothers' characteristics.

<https://doi.org/10.1371/journal.pone.0288867.g004>

Association rule building

The association rule generation process was done based on important attributes selected by performing the best-performing machine learning model (PART). A total of seven association rules were generated, and the details of the rules were presented in [Box 2](#).

Box 2. Association rule generation and knowledge extraction

Rule 1: If wealth status = 3 (Rich), adequate ANC visits = 1 (Yes), and residency = 2 (Urban), then the probability of childhood vaccination would be 86.73% (left = 1.87).

Rule 2: If institutional delivery = 1 (Yes), mothers' educational status = 4 (Higher), and household heads' sex = 0 (Female), then the probability of childhood vaccination would be 82.14% (left = 1.67).

Rule 3: If adequate ANC visit = 1 (Yes), mothers' age = 1 (>35 years), and institutional delivery = 1 (Yes), then the probability of childhood vaccination would be 79.21% (left = 1.47).

Rule 4: If birth order >5 = 0 (No), visited HF in the last 12 months = 1 (Yes), residency = 2 (Urban) and mothers' current working = 1 (Yes), then the probability of childhood vaccination would be 66.81% (left = 1.32).

Rule 5: If institutional delivery = 1 (Yes), mothers' wealth status = 2 (Middle), and visited HF in the last 12 months = 1 (Yes), then the probability of childhood vaccination would be 62.45% (left = 1.25).

Rule 6: If residency = 2 (Urban), birth order >5 = 1 (Yes), wealth status = 2 (Middle), and adequate ANC visits = 1 (Yes), then the probability of childhood vaccination would be 57.16% (left = 1.17).

Rule 7: If mothers' educational status = 3 (Secondary), institutional delivery = 1 (Yes), and mothers currently working = 1 (Yes), then the probability of childhood vaccination would be 51.92% (left = .12).

Table 2. Model accuracy of the included machine learning algorithms Based on confusion matrix parameters.

Confusion matrix Parameters (%)	The included machine-learning algorithms							
	PART	Naïve Bayes	Random forest	Logit Boost	J48	AdaBoost	Multilayer perceptron	LR
True positive rate (%)	89.90	64.00	88.50	72.10	88.40	69.80	83.60	70.90
False positive rate (%)	18.20	30.80	1.50	38.90	32.40	35.21	12.10	36.30
Precision (%)	93.80	73.90	87.80	70.60	76.00	72.00	81.00	71.70
F-measure (%)	94.30	68.20	88.60	71.30	77.71	70.90	82.30	71.30
Relative absolute error (%)	51.78	75.33	34.05	83.68	75.05	85.10	23.19	84.21
AUC (%)	91.89	72.30	82.70	73.20	86.01	72.90	83.29	65.60
Kappa statistics (%)	86.57	32.66	78.68	33.26	79.27	36.50	72.02	35.00
Accuracy (%)	95.53	66.29	82.37	77.28	89.24	68.60	87.20	65.80

Note that LR, stands for Logistic Regression

<https://doi.org/10.1371/journal.pone.0288867.t002>

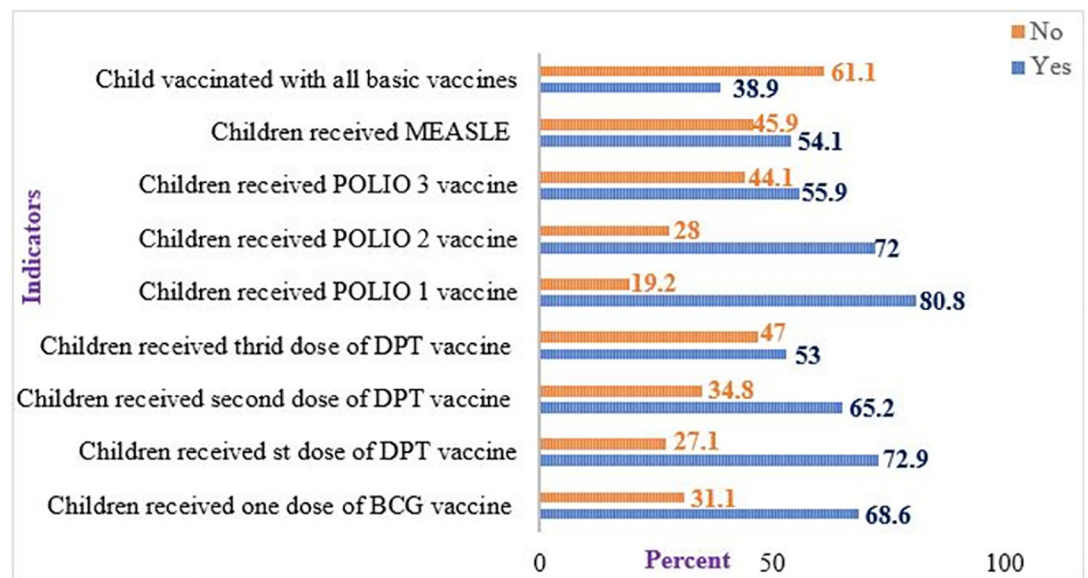


Fig 5. The vaccination status of children aged 12–23 months with recommended vaccination types.

<https://doi.org/10.1371/journal.pone.0288867.g005>

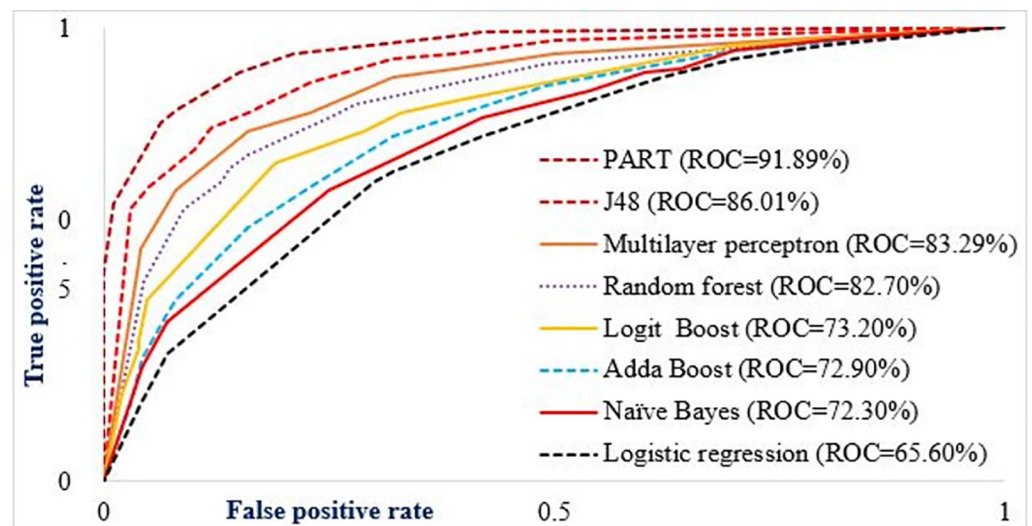


Fig 6. Comparison of machine learning algorithms using the area under ROC value.

<https://doi.org/10.1371/journal.pone.0288867.g006>

Table 3. Information gain value for each predictor.

Predictor variables	Type	Measurement	Information gain value
Adequate ANC visit, (Yes)	Nominal	Scale	0.087
Institutional delivery, (Yes)	Nominal	Scale	0.084
Visited HF in the last 12 months, (Yes)	Nominal	Scale	0.076
Educational status, (Higher)	Nominal	Scale	0.071
Wealth status, (Rich)	Nominal	Scale	0.071
Residency, (Urban)	Nominal	Scale	0.062
Household sex, (Female)	Nominal	Scale	0.059
Mothers' age, (>35 years)	Nominal	Scale	0.037
Birth order >5, (No)	Nominal	Scale	0.023
Mothers currently working, (Yes)	Nominal	Scale	0.023

<https://doi.org/10.1371/journal.pone.0288867.t003>

Discussion

The 2016 EDHS dataset was used, with a total of 1617 sampled observations. The childhood vaccination status of children aged 12–23 months was assessed. As a result, nearly four out of ten (38.9%) of children had received at least one dose of the BCG vaccine, three doses of the polio vaccine, three doses of the DPT vaccine, and one dose of the measles vaccine. The current finding was higher than the study done in the Dabat demographic and health survey site, in Ethiopia [71]. According to the World Health Organization vaccination estimation, the current finding was inadequate since below 90%. Plus, the finding was lower than the study done in East Africa, 69% [72], and in Gondar City, 98% [73]. This might be due to disparities in vaccination program access, and mothers might not understand the value of childhood vaccinations, and not remember when the children had been appointed [74]. Additionally, different natural and human-made factors might limit the uptake of childhood vaccination [75,76]. Moreover, women might face problems with health service access (70.2%), might have poor health-seeking behavior, high transportation costs, and inaccessibility of health facilities might be significant reasons for low coverage of childhood vaccination in Ethiopia [77].

70% and 30% of total observations were set for model training, and model evaluation, respectively. The objectives were to evaluate machine learning algorithms and to identify the

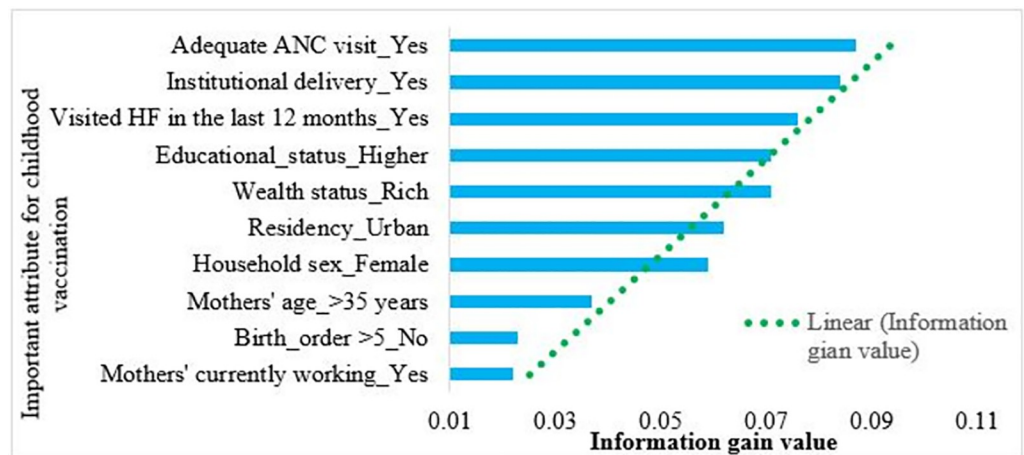


Fig 7. Important attributes selection, based on best performance algorithm (PART), to predict childhood vaccination among children aged 12–23 months in Ethiopia.

<https://doi.org/10.1371/journal.pone.0288867.g007>

best algorithm to select important attributes to predict childhood vaccination in Ethiopia. Hence, eight machine learning algorithms were considered for comparison. Different confusion matrix elements were used to compare the candidate machine learning algorithms.

The included eight machine learning algorithms were evaluated and compared by classification matrix elements accuracy and AUR score values. Hence, the accuracy and AUR value of the PART algorithms were 95.53% and 91.89% with 10-fold cross-validations, respectively. Hence, the PAR algorithm was the first accurate model to predict childhood vaccination among children aged 12–23 months in this study. This finding was agreed with studies done about data classification and terms of association [35], and the application of data mining for the prediction of patients' CD4 count [78]. The j48, multilayer perceptron, and random forest algorithms were the second, third, and fourth best machine learning algorithms to predict childhood vaccination with 89.24%, 87.20%, and 82.37% accuracy, respectively. This finding was supported by various studies conducted to predict under-five child mortality [29,32,44,79], contraceptive discontinuation [70], stunting, and malnutrition among children [80–82].

The second objective of the study was to select important attributes that could predict childhood vaccination among children aged 12–23 months in Ethiopia. From the attributes selected to predict childhood vaccination, adequate ANC visits, institutional delivery, health facility visits, higher education of mothers, rich wealth status, children from urban areas, female household heads, a mother's age greater than 35 years, a child's birth order less than five, and mothers currently working were important attributes to predict vaccination of children aged 12–23 months in Ethiopia.

Adequate ANC visits were the top-ranked attribute to predict childhood vaccination among children aged 12–23 months in Ethiopia, with a 0.087 information gain value. This finding was agreed upon with the previous similar studies done in Ethiopia [14,83], and Zimbabwe [84]. This might be due to women who attend ANC follow-up might get counseling services about child immunization [85], and mothers might receive adequate education about the importance of postnatal visits [86]. Moreover, an adequate number of ANC visits is associated with a greater likelihood of having a child vaccinated [87].

Institutional delivery was the second-most important attribute in predicting childhood vaccination. This finding is supported by similar studies done in Ethiopia [8,14], and Nigeria [88]. This might be because children who were born at health facilities might be more likely to get BCG and OPV 0 vaccines at birth than children who were born elsewhere [85]. Plus, institutional delivery might create an opportunity for children's mothers to communicate with health professionals about the importance and side effects of immunization, and the vaccine initiation time [85]. Moreover, children's mothers who gave birth at health facilities might get information about the basic childhood vaccination services for the current and the next vaccination appointment schedules [89].

Visiting HF was the third most important attribute in predicting childhood vaccination in Ethiopia. This finding was in line with studies done in Ethiopia [14], and similar resource-limited settings [90,91]. This might be because mothers who visit a health facility might receive adequate education and counseling about child immunization, and mothers after birth are recommended to visit a health facility for postnatal check-ups and services [85].

The higher educational status of mothers was the fourth important attribute to predict childhood vaccination among children aged 12–23 months. This study is also similar to a study done in Bangladesh in that maternal education is an important feature in predicting anemia among under-five children [92]. Another study done in India also supports the current findings of the study [93]. This might be due to educated mothers knowing the importance of

vaccines for child care, and educated mothers empowering them and feel free to make decisions to visit the health facility for child health services [94].

Being rich and being urban residents were the fifth and sixth important attributes to predict childhood vaccination among children aged 12–23 months in Ethiopia. This finding was similar to a study done in Bangladesh [92], and Ethiopia [8]. This might be because mothers from urban areas might have more access to media, which plays a vital role in disseminating educational information and creating awareness [95,96]. Plus, children's mothers in an urban area might have adequate information communication technology infrastructure that enables them to receive short message services for health information services access [1]. Therefore, children in urban areas might be more likely to get and uptake vaccines. Moreover, wealthier people might have media access and afford to cover the transport cost of health facilities, so they might have access to information and better health-seeking behavior, and good childcare practices [71].

Generating rules for childhood vaccination was the third objective of the study. Previous studies have assessed the joint effect of independent predictors on the outcome of interest [44,70,78]. Consequently, seven association rules were generated to determine vaccination status among children aged 12–23 months in Ethiopia. According to association rule 1, the probability of a childhood vaccination would be 86.73%, if and only if the mothers' wealth status was rich, mothers had adequate ANC visits, and the children were urban residents. This might be because women with rich wealth status might be able to afford to pay any costs needed for vaccination, mothers who had adequate ANC visits might have adequate awareness and knowledge about child vaccination during their health facility visits during their pregnancy period, and health facilities in urban areas might be easily accessible for mothers to vaccinate their children. The effects of these three attributes are critical for childhood vaccination, and the combination of these factors might make it particularly important for children to be vaccinated when they are under 12 to 23 months. Based on Rule 2, childhood vaccination would be 82.14%, if mothers gave birth at health institutions, mothers' educational status was higher, and if the household heads' sex was female. The if/ then rules are critical to discovering hidden relationships between attributes, extracting knowledge from a set of data, and accurately representing knowledge and information about the vaccination of children. The findings presented in this study are critically important for policymakers and stakeholders to support public health action, decision-making purposes, and the storage of knowledge regarding child vaccination status.

Strengths and limitations of the study

In this study, machine learning algorithms are used to classify, and predict childhood vaccination. This study used nationally representative data, and the findings might be representative of the study populations. However, machine learning algorithms do not have coefficients like odds and incident rate ratios. Therefore, the strength and direction of associations are unknown.

Conclusions

In this study, PART, J48, multilayer perceptron, and random forest algorithms were the first, second, third, and fourth best performance machine learning algorithms to predict childhood vaccination in Ethiopia. Adequate ANC visits, institutional delivery, health facility visits, higher educational status, and rich mothers were the top five important attributes to predict childhood vaccination in Ethiopia. Moreover, seven rules were generated that attributes together can determine the magnitude of childhood vaccination.

The findings of this study would support policymakers and stakeholders in developing childcare intervention mechanisms and early preparedness for caring for children through child immunization, and the findings would serve as input for immunization coverage and reduction of vaccine dropouts. The generated rule would be important for knowledge creation and representation. Specifically, stakeholders are recommended to enhance mothers' ANC visits and institutional delivery by constructing nearby health facilities. Creating income opportunities and awareness of mothers would be also critical interventions for childhood vaccination. Moreover, the current study would serve as a baseline for future studies.

Supporting information

S1 File. Checklist for the current study.

(PDF)

Acknowledgments

The authors would like to express their deepest appreciation to the DHS program for permitting data access and use for this study.

Author Contributions

Conceptualization: Addisalem Workie Demsash, Teshome Bekana.

Data curation: Addisalem Workie Demsash, Teshome Bekana.

Formal analysis: Addisalem Workie Demsash, Teshome Bekana.

Investigation: Addisalem Workie Demsash.

Methodology: Addisalem Workie Demsash, Teshome Bekana.

Validation: Addisalem Workie Demsash.

Writing – original draft: Addisalem Workie Demsash, Alex Ayenew Chereka, Agmasie Damtew Walle, Sisay Yitayih Kassie, Firomsa Bekele.

Writing – review & editing: Addisalem Workie Demsash, Alex Ayenew Chereka, Agmasie Damtew Walle, Sisay Yitayih Kassie, Firomsa Bekele.

Reference

1. Unicef, statistical snapshot. Child mortality: Accessed from <https://data.unicef.org/resources/2013-statistical-snapshot-child-mortality/>. New York, 2013.
2. Organization, W.H., Meeting report: WHO technical consultation: nutrition-related health products and the World Health Organization model list of essential medicines—practical considerations and feasibility: Geneva, Switzerland, 20–21 September 2018. 2019, World Health Organization.
3. UNICEF and W.H. Organization, Levels & trends in child mortality estimates developed by the UN Inter-Agency Group for Child Mortality Estimation. 2015.
4. Sakelo A.N., et al., Newborn care practice and associated factors among mothers of one-month-old infants in Southwest Ethiopia. *International Journal of Pediatrics*, 2020. 2020: p. 1–7. <https://doi.org/10.1155/2020/3897427> PMID: 33133199
5. Organization, W.H., World health statistics 2016: Monitoring health for the SDGs sustainable development goals. 2016: World Health Organization.
6. Meleko A., Geremew M., and Birhanu F., Assessment of child immunization coverage and associated factors with full vaccination among children aged 12–23 months at Mizan Aman town, Bench Maji zone, Southwest Ethiopia. *International Journal of Pediatrics*, 2017. 2017. <https://doi.org/10.1155/2017/7976587> PMID: 29434643

7. WHO, U., World Bank. State of the World's Vaccines and Immunization. Geneva, Switzerland: World Health Organization; 2009: Accessed from <https://www.tandfonline.com/doi/abs/10.4161/hv.6.2.11326>. 2010.
8. Tesfaye T.D., Temesgen W.A., and Kasa A.S., Vaccination coverage and associated factors among children aged 12–23 months in Northwest Ethiopia. *Human vaccines & immunotherapeutics*, 2018. 14 (10): p. 2348–2354. <https://doi.org/10.1080/21645515.2018.1502528> PMID: 30118398
9. Touray E., et al., Childhood vaccination uptake and associated factors among children 12–23 months in rural settings of the Gambia: a community-based cross-sectional study. *BMC Public Health*, 2021. 21 (1): p. 1–10.
10. Taiwo L., et al., Factors affecting access to information on routine immunization among mothers of under 5 children in Kaduna State Nigeria, 2015. *Pan African Medical Journal*, 2017. 27(1). <https://doi.org/10.11604/pamj.2017.27.186.11191> PMID: 29187919
11. WHO, Vaccines, and immunization. 2023. https://www.who.int/health-topics/vaccines-and-immunization#tab=tab_1.
12. Payne S., et al., Achieving comprehensive childhood immunization: an analysis of obstacles and opportunities in The Gambia. *Health policy and planning*, 2014. 29(2): p. 193–203. <https://doi.org/10.1093/heapol/czt004> PMID: 23426974
13. Ndiritu M., et al., Immunization coverage and risk factors for failure to immunize within the Expanded Programme on Immunization in Kenya after the introduction of new Haemophilus influenzae type b and hepatitis b virus antigens. *BMC public health*, 2006. 6(1): p. 1–8.
14. Dirirsa K., et al., Assessment of vaccination timeliness and associated factors among children in Toke Kutaye district, central Ethiopia: A Mixed study. *Plos one*, 2022. 17(1): p. e0262320. <https://doi.org/10.1371/journal.pone.0262320> PMID: 35085296
15. Chandir S., et al., Using predictive analytics to identify children at high risk of defaulting from a routine immunization program: a feasibility study. *JMIR public health and Surveillance*, 2018. 4(3): p. e9681.
16. Animaw W., et al., an Expanded program of immunization coverage and associated factors among children age 12–23 months in Arba Minch town and Zuria District, Southern Ethiopia, 2013. *BMC public health*, 2014. 14(1): p. 1–10.
17. Collishaw N.E., The millennium development goals, and tobacco control. *Global Health Promotion*, 2010. 17(1_suppl): p. 51–59. <https://doi.org/10.1177/1757975909358250> PMID: 20595354
18. Debie A. and Taye B., Assessment of full vaccination coverage and associated factors among children aged 12–23 months in Mecha District, north West Ethiopia: a cross-sectional study. *Sci J Public Health*, 2014. 2(4): p. 342–8.
19. Mohammed H. and Atomsa A., Assessment of child immunization coverage and associated factors in Oromia regional state, eastern Ethiopia. *Science, Technology, and Arts Research Journal*, 2013. 2(1): p. 36–41.
20. Negussie A., et al., Factors associated with incomplete childhood immunization in Arbogona district, southern Ethiopia: a case–control study. *BMC public health*, 2015. 16(1): p. 1–9.
21. Ekouevi D.K., et al., Incomplete immunization among children aged 12–23 months in Togo: a multilevel analysis of individual and contextual factors. *BMC public health*, 2018. 18: p. 1–10.
22. Budu E., et al., Trend and determinants of complete vaccination coverage among children aged 12–23 months in Ghana: analysis of data from the 1998 to 2014 Ghana demographic and health surveys. *Plos one*, 2020. 15(10): p. e0239754. <https://doi.org/10.1371/journal.pone.0239754> PMID: 33002092
23. Tegene T., et al., Newborn care practice and associated factors among mothers who gave birth within one year in Mandura District, Northwest Ethiopia. *Clinics in Mother and Child Health*, 2015. 12(1).
24. Pepe M.S., et al., Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology*, 2004. 159(9): p. 882–890. <https://doi.org/10.1093/aje/kwh101> PMID: 15105181
25. Saroj R.K., et al., Machine Learning Algorithms for understanding the determinants of under-five Mortality. *BioData mining*, 2022. 15(1): p. 1–22.
26. Cody S. and Asher A., Smarter, better, faster: The potential for predictive analytics and rapid-cycle evaluation to improve program development and outcomes. 2014, Mathematica Policy Research.
27. Cheong Q., et al., Predictive modeling of vaccination uptake in US counties: A machine learning–based approach. *Journal of Medical Internet Research*, 2021. 23(11): p. e33231. <https://doi.org/10.2196/33231> PMID: 34751650
28. Mannion N., Predictions of changes in child immunization rates using an automated approach: USA. 2020, Dublin, National College of Ireland.

29. Tesfaye B., et al., Determinants and development of a web-based child mortality prediction model in resource-limited settings: a data mining approach. *Computer methods and programs in biomedicine*, 2017. 140: p. 45–51. <https://doi.org/10.1016/j.cmpb.2016.11.013> PMID: 28254089
30. Osisanwo F., et al., Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 2017. 48(3): p. 128–138.
31. Jaskari J., et al., Machine learning methods for neonatal mortality and morbidity classification. *IEEE Access*, 2020. 8: p. 123347–123358.
32. Fenta H.M., Zewotir T., and Muluneh E.K., A machine learning classifier approach for identifying the determinants of under-five child undernutrition in Ethiopian administrative zones. *BMC Medical Informatics and Decision Making*, 2021. 21(1): p. 1–12.
33. Thangamani D. and Sudha P., Identification of malnutrition with use of supervised data mining techniques—decision trees and artificial neural networks. *Int J Eng Comput Sci*, 2014. 3(09).
34. Kuttiyapillai D. and Ramachandran R., Improved text analysis approach for predicting effects of nutrient on human health using machine learning techniques. *IOSR J Comput Eng*, 2014. 16(3): p. 86–91.
35. Dhar, A., N.S. Dash, and K. Roy. An innovative method of feature extraction for text classification using the part classifier. in *Information, Communication and Computing Technology: Third International Conference, ICICCT 2018, New Delhi, India, May 12, 2018, Revised Selected Papers 3*. 2019. Springer.
36. Demsash A.W., et al., Spatial and multilevel analysis of sanitation service access and related factors among households in Ethiopia: using 2019 Ethiopian national dataset. *PLOS Global Public Health*, 2023. 3(4): p. e0001752. <https://doi.org/10.1371/journal.pgph.0001752> PMID: 37014843
37. Geography of Ethiopia. https://en.wikipedia.org/wiki/Geography_of_Ethiopia.
38. The 2016 Ethiopian Demography and Health Survey. <https://dhsprogram.com/methodology/survey/survey-display-478.cfm>.
39. Wakeyo M.M., et al., Short birth interval and its associated factors among multiparous women in Mieso agro-pastoralist district, Eastern Ethiopia: A community-based cross-sectional study. *Front Glob Womens Health*, 2022. 3: p. 801394. <https://doi.org/10.3389/fgwh.2022.801394> PMID: 36159883
40. Kassie S.Y., et al., Spatial distribution of short birth interval and associated factors among reproductive age women in Ethiopia: a spatial and multilevel analysis of 2019 Ethiopian mini demographic and health survey. *BMC Pregnancy and Childbirth*, 2023. 23(1): p. 1–14.
41. Demsash A.W., et al., Spatial distribution of vitamin A rich foods intake and associated factors among children aged 6–23 months in Ethiopia: a spatial and multilevel analysis of 2019 Ethiopian mini demographic and health survey. *BMC Nutrition*, 2022. 8(1): p. 1–14.
42. Muhwava L.S., Morojele N., and London L., Psychosocial factors associated with early initiation and frequency of antenatal care (ANC) visits in a rural and urban setting in South Africa: a cross-sectional survey. *BMC pregnancy and childbirth*, 2016. 16(1): p. 1–9. <https://doi.org/10.1186/s12884-016-0807-1> PMID: 26810320
43. Levy P.S. and Lemeshow S., *Sampling of populations: methods and applications*. 2013: John Wiley & Sons.
44. Demsash A.W., Using best performance machine learning algorithm to predict child death before celebrating their fifth birthday. *Informatics in Medicine Unlocked*, 2023: p. 101298.
45. Webb G.I., Keogh E., and Miikkulainen R., Naïve Bayes. *Encyclopedia of machine learning*, 2010. 15: p. 713–714.
46. Hall M., et al., The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 2009. 11(1): p. 10–18.
47. Hosmer D.W. Jr, Lemeshow S., and Sturdivant R.X., *Applied logistic regression*. Vol. 398. 2013: John Wiley & Sons.
48. Uddin S., et al., Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, 2019. 19(1): p. 1–16.
49. Kaur G. and Chhabra A., Improved J48 classification algorithm for the prediction of diabetes. *International journal of computer applications*, 2014. 98(22).
50. Sharma A.K. and Sahni S., A comparative study of classification algorithms for spam email data analysis. *International Journal on Computer Science and Engineering*, 2011. 3(5): p. 1890–1895.
51. Saroj R.K., et al., Machine Learning Algorithms for understanding the determinants of under-five Mortality. *BioData Min*, 2022. 15(1): p. 20. <https://doi.org/10.1186/s13040-022-00308-8> PMID: 36153553
52. Fu G., Dai X., and Liang Y., Functional random forests for curve response. *Sci Rep*, 2021. 11(1): p. 24159. <https://doi.org/10.1038/s41598-021-02265-4> PMID: 34921167

53. Tkachev V., et al., Flexible Data Trimming Improves Performance of Global Machine Learning Methods in Omics-Based Personalized Oncology. *Int J Mol Sci*, 2020. 21(3). <https://doi.org/10.3390/ijms21030713> PMID: 31979006
54. Yu Y., et al., Machine Learning Methods for Predicting Long-Term Mortality in Patients After Cardiac Surgery. *Front Cardiovasc Med*, 2022. 9: p. 831390. <https://doi.org/10.3389/fcvm.2022.831390> PMID: 35592400
55. What is AdaBoost Algorithm Model?: Accessed from <https://data-flair.training/blogs/adaboost-algorithm/>.
56. Kamarudin M.H., et al., A logit boost-based algorithm for detecting known and unknown web attacks. *IEEE Access*, 2017. 5: p. 26190–26200.
57. Alghamdi M., et al., Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project. *PloS One*, 2017. 12(7): p. e0179805. <https://doi.org/10.1371/journal.pone.0179805> PMID: 28738059
58. Handling Imbalanced Datasets in Machine Learning. 2020. <https://www.section.io/engineering-education/imbalanced-data-in-ml/>.
59. Zenu S., et al., Determinants of first-line antiretroviral treatment failure among adult patients on treatment in Mettu Karl Specialized Hospital, South West Ethiopia; a case-control study. *Plos one*, 2021. 16(10): p. e0258930. <https://doi.org/10.1371/journal.pone.0258930> PMID: 34679085
60. Elhassan T. and Aljurf M., Classification of imbalance data using tome link (t-link) combined with random under-sampling (rus) as a data reduction method. *Global J Technol Optim S*, 2016. 1: p. 2016.
61. Narkhede S., Understanding auc-roc curve. *Towards Data Science*, 2018. 26(1): p. 220–227.
62. El Khouli R.H., et al., Relationship of temporal resolution to diagnostic performance for dynamic contrast-enhanced MRI of the breast. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 2009. 30(5): p. 999–1004. <https://doi.org/10.1002/jmri.21947> PMID: 19856413
63. McHugh M.L., Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*, 2012. 22(3): p. 276–82. <https://doi.org/10.1016/j.jocd.2012.03.005> PMID: 23092060
64. Molnar, C., *Interpretable machine learning*. 2020: Lulu. com.
65. Shi R., et al., Obesity is negatively associated with dental caries among children and adolescents in Huizhou: a cross-sectional study. *BMC Oral Health*, 2022. 22(1): p. 76. <https://doi.org/10.1186/s12903-022-02105-5> PMID: 35300666
66. Ivančević V., et al., Using association rule mining to identify risk factors for early childhood caries. *Computer methods and programs in biomedicine*, 2015. 122(2): p. 175–181. <https://doi.org/10.1016/j.cmpb.2015.07.008> PMID: 26271408
67. Zafar A., et al., Machine learning-based risk factor analysis and prevalence prediction of intestinal parasitic infections using epidemiological survey data. *PLOS Neglected Tropical Diseases*, 2022. 16(6): p. e0010517. <https://doi.org/10.1371/journal.pntd.0010517> PMID: 35700192
68. Tandan M., et al., Discovering symptom patterns of COVID-19 patients using association rule mining. *Computers in biology and medicine*, 2021. 131: p. 104249. <https://doi.org/10.1016/j.combiomed.2021.104249> PMID: 33561673
69. Li Q., et al., Mining association rules between stroke risk factors based on the Apriori algorithm. *Technology and Health Care*, 2017. 25(S1): p. 197–205. <https://doi.org/10.3233/THC-171322> PMID: 28582907
70. Kebede S.D., et al., Prediction of contraceptive discontinuation among reproductive-age women in Ethiopia using Ethiopian Demographic and Health Survey 2016 Dataset: A Machine Learning Approach. *BMC Medical Informatics and Decision Making*, 2023. 23(1): p. 1–17.
71. Gelagay A.A., et al., Complete childhood vaccination and associated factors among children aged 12–23 months in Dabat demographic and health survey site, Ethiopia, 2022. *BMC Public Health*, 2023. 23(1): p. 1–9.
72. Tesema G.A., et al., Complete basic childhood vaccination and associated factors among children aged 12–23 months in East Africa: a multilevel analysis of recent demographic and health surveys. *BMC Public Health*, 2020. 20(1): p. 1–14.
73. Yismaw A.E., et al., Incomplete childhood vaccination and associated factors among children aged 12–23 months in Gondar city administration, Northwest, Ethiopia 2018. *BMC research notes*, 2019. 12(1): p. 1–7.
74. Ozawa S., et al., Return on investment from childhood immunization in low-and middle-income countries, 2011–20. *Health Affairs*, 2016. 35(2): p. 199–207. <https://doi.org/10.1377/hlthaff.2015.1086> PMID: 26858370

75. Tugumisirize F., Tumwine J., and Mworoza E., Missed opportunities and caretaker constraints to childhood vaccination in rural areas of Uganda. *East African medical journal*, 2002. 79(7): p. 347–354.
76. Demsash A.W., Emanu M.D., and Walle A.D., Exploring spatial patterns, and identifying factors associated with insufficient cash or food received from a productive safety net program among eligible households in Ethiopia: a spatial and multilevel analysis as an input for international food aid programmers. *BMC Public Health*, 2023. 23(1): p. 1141.
77. Demsash A.W., and Walle A.D., Women's health service access and associated factors in Ethiopia: application of geographical information system and multilevel analysis. *BMJ Health & Care Informatics*, 2023. 30(1). <https://doi.org/10.1136/bmjhci-2022-100720> PMID: 37116949
78. Mariam B.G., and Mariam T.H., Application of data mining techniques for predicting CD4 status of patients on ART in Jimma and Bonga Hospitals, Ethiopia. *Journal of Health & Medical Informatics*, 2015. 6(6): p. 1–9.
79. Bitew F.H., et al., Machine learning approach for predicting under-five mortality determinants in Ethiopia: evidence from the 2016 Ethiopian Demographic and Health Survey. *Genus*, 2020. 76: p. 1–16.
80. Fenta H.M., et al., Determinants of stunting among under-five years children in Ethiopia from the 2016 Ethiopia Demographic and Health Survey: Application of ordinal logistic regression model using complex sampling designs. *Clinical Epidemiology and Global Health*, 2020. 8(2): p. 404–413.
81. Talukder A. and Ahammed B., Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. *Nutrition*, 2020. 78: p. 110861. <https://doi.org/10.1016/j.nut.2020.110861> PMID: 32592978
82. Kassie G.W. and Workie D.L., Determinants of under-nutrition among children under five years of age in Ethiopia. *BMC Public Health*, 2020. 20(1): p. 1–11.
83. Tesfa G.A., et al., Spatial distribution and associated factors of measles vaccination among children aged 12–23 months in Ethiopia. A spatial and multilevel analysis. *Human Vaccines & Immunotherapeutics*, 2022. 18(1): p. 2035558. <https://doi.org/10.1080/21645515.2022.2035558> PMID: 35148252
84. Mukungwa T., Factors associated with full immunization coverage amongst children aged 12–23 months in Zimbabwe. *African Population Studies*, 2015. 29(2).
85. Tamirat K.S. and Sisay M.M., Full immunization coverage and its associated factors among children aged 12–23 months in Ethiopia: further analysis from the 2016 Ethiopia demographic and health survey. *BMC public health*, 2019. 19: p. 1–7.
86. Gualu T. and Dilie A., Vaccination coverage and associated factors among children aged 12–23 months in debre markos town, Amhara regional state, Ethiopia. *Advances in Public Health*, 2017. 2017.
87. Tefera Y.A., et al., Predictors and barriers to full vaccination among children in Ethiopia. *Vaccines*, 2018. 6(2): p. 22. <https://doi.org/10.3390/vaccines6020022> PMID: 29642596
88. Antai D., Regional inequalities in under-5 mortality in Nigeria: a population-based analysis of individual- and community-level determinants. *Population health metrics*, 2011. 9: p. 1–10.
89. Mutua M.K., Kimani-Murage E., and Ettarh R.R., Childhood vaccination in informal urban settlements in Nairobi, Kenya: who gets vaccinated? *BMC public health*, 2011. 11: p. 1–11.
90. Darroch J.E., Sedgh G., and Ball H., *Contraceptive technologies: responding to women's needs*. New York: Guttmacher Institute, 2011. 201(1): p. 1–51.
91. Kozuki N. and Walker N., Exploring the association between short/long preceding birth intervals and child mortality: using reference birth interval children of the same mother as a comparison. *BMC public health*, 2013. 13(3): p. 1–10.
92. Khan J.R., et al., Machine learning algorithms to predict childhood anemia in Bangladesh. *Journal of Data Science*, 2019. 17(1): p. 195–218.
93. Khare S., et al., Investigation of nutritional status of children based on machine learning techniques using Indian demographic and health survey data. *Procedia computer science*, 2017. 115: p. 338–349.
94. Bekele Y.A. and Fekadu G.A., Factors associated with HIV testing among young females; further analysis of the 2016 Ethiopian demographic and health survey data. *PLoS One*, 2020. 15(2): p. e0228783. <https://doi.org/10.1371/journal.pone.0228783> PMID: 32045460
95. Oginni A.B., Adebajo S.B., and Ahonsi B.A., Trends, and determinants of comprehensive knowledge of HIV among adolescents and young adults in Nigeria: 2003–2013. *African Journal of reproductive health*, 2017. 21(1): p. 26–34. <https://doi.org/10.29063/ajrh2017/v21i2.4> PMID: 29624937
96. Haque M.A., et al., Factors associated with knowledge and awareness of HIV/AIDS among married women in Bangladesh: evidence from a nationally representative survey. *SAHARA-J: Journal of Social Aspects of HIV/AIDS*, 2018. 15(1): p. 121–127. <https://doi.org/10.1080/17290376.2018.1523022> PMID: 30249174