

# Predicting the Level of Anemia among Ethiopian Pregnant Women using Homogeneous Ensemble Machine Learning Algorithm

**Belayneh E Dejene**

University of Gondar

**Tesfamariam M Abuhay** (✉ [tesfamariam.mabuhay@uog.edu.et](mailto:tesfamariam.mabuhay@uog.edu.et))

University of Gondar

**Dawit S Bogale**

University of Gondar

---

## Research Article

**Keywords:** Homogeneous Ensemble Machine Learning, Health Informatics, Anemia, Maternal Healthcare

**Posted Date:** March 18th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1445780/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** More than 115,000 maternal deaths and 591,000 prenatal deaths occurred in the world per year because of anemia, the reduction of red blood cells or hemoglobin in the blood. The world health organization divides anemia in pregnancy into mild anemia (Hb 10- 10.9g/dl), moderate anemia (Hb 7.0-9.9g/dl), and severe anemia (Hb < 7g/dl). This study aims to identify risk factors and predict the level of anemia among pregnant women in the case of Ethiopia using homogeneous ensemble machine learning algorithms.

**Methods:** This study was conducted following a design science research approach. The data were gathered from the Ethiopian demographic health survey and preprocessed to get quality data that are suitable for the machine learning algorithm. Decision tree, random forest, cat boost, and extreme gradient boosting with class decomposition (one versus one and one versus rest) and without class decomposition were employed to build the predictive model. For constructing the proposed model, twelve experiments were conducted using a total of 29104 instances with 23 features, and a training and testing dataset split ratio of 80/20.

**Results:** The overall accuracy of random forest, extreme gradient boosting, and cat boost without class decompositions is 91.34%, 94.26%, and 97.08.90%, respectively. The overall accuracy of random forest, extreme gradient boosting, and cat boost with one versus one is 94.4%, 95.21%, and 97.44%, respectively. The overall accuracy of random forest, extreme gradient boosting, and cat boost with one versus the rest are 94.4%, 94.54%, and 97.6%, respectively.

**Conclusion:** A predictive model that was developed with cat boost algorithms with one versus the rest was selected to identify risk factors, generate rules, and develop a deployed artifact because it has registered better performance with 97.6% accuracy. The most determinant risk factors of anemia among pregnant women were identified using feature importance. Some of them are the duration of the current pregnancy, age, source of drinking water, respondent's (pregnant women) occupation, number of household members, wealth index, husband/partner's education level, birth history.

## I. Background

According to [1], Anemia is defined as a decrease in the number of RBC or hemoglobin in the blood that has significant adverse health consequences. Anemia is a public health problem among women of reproductive age, affecting both poor and rich countries overall the world [2]. It negatively affects the social and economic well-being of a country and all its communities [1]. According to [3] and [4], anemia during pregnancy is one of the risk factors for poor pregnancy outcomes such as low birth weight (LBW), preterm birth, prematurity stillbirth, intrauterine growth restriction, and impaired cognitive development.

Anemia in pregnant women can be caused by parasitic infestation, socio-demographic status, economic status, dietary practice, obstetric factors, reproductive health, and other health-related factors [5]. More than 115,000 maternal deaths and 591,000 prenatal deaths are caused by anemia disease in the world per year [6]. According to the World Health Organization (WHO) 1993–2005 report, anemia affects 41.8% of pregnant women worldwide, with Africa having the highest prevalence (57.1%) [7][8]. According to [4] and [9], anemia during pregnancy is the main cause of morbidity and mortality of pregnant women in developing countries like Ethiopia and has both maternal and fetal consequences such as impairment of the capacity of the blood to transport oxygen around the body, fatigue, poor work capacity, impaired immune function, increased risk of cardiac diseases, and mortality [4][10]. The burden and underlying factors of this disease varied even within a country [10]. Most of the women who live in the rural area of Ethiopia have been affected by this disease due to different factors including nutrition, parasites, socio-demographic, obstetric, reproductive characteristics, and the like [10]. According to WHO guidelines, the minimum acceptable hemoglobin level during pregnancy is 11 g/dl, during the first half, 10.5 g/dl, during the second half, and 12 g/dl for lactating women [6][10][11]. To understand and predict the level of anemia among pregnant women in the case of Ethiopia, several types of research have been conducted. For example, [3][6][7][8][9][10][11][12][13] and [14] investigated the status of anemia among pregnant women using cross-sectional statistical

methods. They also used bivariate and multivariate logistic regression methods and identified the most determinant risk factors. Most of these studies, however, used local clinical data that covered limited geographical areas like a single city or town only, small data set less than 500 instances/records, and only focused on one of the risk factors such as socioeconomic, demographic, nutritional, and reproductive, apart from health-related variables. Previous studies including [3][6][7][8][9][10][11][12][13] and [14] also focused on identifying the determinant risk factors of anemia among pregnant women who followed first antenatal care only using descriptive statistical models. Besides, [3][6][7][8][9][10][11][12][13] and [14] were conducted using cross-sectional statistical methods which generally have limited capacity to discover new and unanticipated patterns that are hidden in data and identify cause and effect relationships [6][10][15]. These studies did not also include features that lead to anemia such as the history of birth, history of abortion, history of the place of delivery, history of malaria, and nutritional variables. i.e. the factors that contribute to the occurrence of anemia among pregnant women weren't thoroughly studied. In such situations, new technologies like machine learning algorithms may help to discover hidden patterns [16]. There were machine learning-related works such as [17][18][19] and [20]. However, these studies aimed at developing a predictive model, but did not identify the most determinant risk factors, and generate rules that allow the development of evidence-based strategies and policies towards preventing and/or reducing anemia among pregnant women. This study, hence, aims to develop a model that predicts the level of anemia among pregnant women using homogeneous ensemble machine learning algorithms by investigating the following research questions: (1) what is the underlying structure of anemia among pregnant women in Ethiopia? (2) Which homogeneous ensemble of machine learning algorithms is suitable for predicting the level of anemia among pregnant women in Ethiopia? (3) What are the associated risk factors that influence the occurrence of anemia among pregnant women in Ethiopia? (4) What are the important rules that may shape strategies and policies towards preventing and/or reducing anemia among pregnant women in Ethiopia?

The rest of this document is organized as follows: Section II presents related works, Section III discusses materials and methods used, Section IV mentions experimental setup and result discussion, and Section V presents the conclusion.

## **II. Related Works**

Several studies such as [3][6][7][8][9][10][11][12][13] and [14] investigated the status of anemia among pregnant women and its determinant factors in different parts of Ethiopia using cross-sectional statistical methods. They used bivariate and multivariate logistic regression methods. However, cross-sectional statistical methods usually have limited capacity to discover new and unanticipated patterns and identify cause and effect relationships that are hidden in data [6][10][15]. Most of these previous studies used local clinical data that covered limited geographical areas like a single city or town only, employed small data set less than 500 instances/records, and focused one of the risk factors like socioeconomic, demographic, nutritional, and reproductive, apart from health-related variables. Some of them also identified the determinant risk factors of anemia among pregnant women who followed first antenatal care. These studies did not include features, such as history of birth, history of abortion, history of place of delivery, history of malaria, and nutritional variables. i.e. the factors that contribute to the occurrence of anemia among pregnant women weren't thoroughly studied. Dithy and Krishnapriya [17] predicted anemia among pregnant women using ANN and gaussian classification algorithm with an accuracy of 0.65% and 0.74%, respectively. Dithy and Krishnapriya [18] tried to classify anemia in pregnant women using random prediction (Rp) classification algorithm and achieved an accuracy of 0.65%, 0.76%, 0.826%, and 0.92% with ANN, gaussian, vector neighbor, and random, respectively. Nevertheless, these studies did not consider all potential features that are discussed in section I, which helps to take holistic interventions. Furthermore, [17][18][19] and [20] aimed to construct a predictive model, but they did not identify risk factors, and extract rules which are important to make evidence-based strategies, policies and interventions towards preventing and/or reducing anemia among pregnant women in Ethiopia. This study, hence, motivated to fill these gaps by constructing a predictive model, identifying risk factors, extracting relevant rules, designing an innovative artifact and deploying the predictive model for potential users.

## iii. Materials And Methods

### A. Data Collection

The data used in this research was extracted from the Ethiopian Demographic Health Survey (EDHS) which was collected by the Ethiopian central statistical agency in 2005, 2011, and 2016, in the five-year interval.

### B. Data Preprocessing

The extracted datasets consist of a total of 11174 instances with 34 features. As all these features are not relevant for developing a predictive model that can predict the level of anemia among pregnant women in the case of Ethiopia, data preprocessing techniques such as data cleaning, data transformation, handling class imbalance, removal of quasi constant features, and feature selection methods were applied. The missing values were handled using mode imputation techniques for categorical data. Redundant data were removed manually. The quasi constant features were not directly removed, but we have constructed one feature and combined them into one. There were features which have several distinct values and need to be transformed for mining purposes; such as features with more categorical values such as the source of drinking water, body mass index, wealth index, marital status, and household members were transformed into discrete values using binning discretization mechanisms. Then, features selection methods were applied to select the relevant features which are important for the further process [23]. In this study, two types of feature selection methods (filter, and wrapper) were employed to see which one produces better results. As a result, the step forward feature selection method performs better than others, see Table 1 which shows the list of features ordered based on their importance to predicting anemia among pregnant women. Besides, domain expert's recommended additional seven features, see Table 2. After conducting all the required data preprocessing tasks, a total of 29104 instances with 23 features were considered for further analysis and prediction model development. Finally, the dataset was divided into training and testing datasets following an 80/20% ratio. The class level of the training dataset was imbalanced which was treated using the synthetic minority over-sampling technique (SMOTE) to avoid loss of valuable information [21][22].

Table 1  
Feature selection results

	<b>Mutual information feature selection</b>	<b>Chi2 feature selection</b>	<b>F class if feature selection</b>	<b>Step forward feature selection</b>	<b>Step backward feature selection</b>
0	Age in 5-year groups	Region	Region	Age in 5-year groups	Age in 5-year groups
1	Region	Highest educational level	Type of place of residence	Region	Region
2	Number of antenatal care visits	Source of drinking water	Highest educational level	Number of antenatal care visits	Number of antenatal care visits
3	Highest educational level	Religion	Source of drinking water	Source of drinking water	Highest educational level
4	Religion	Frequency of reading newspaper or magazine	Religion	Religion	Source of drinking water
5	Frequency of watching television	Frequency of listening to radio	Frequency of watching television	Number of household members	Religion
6	Duration of current pregnancy	Frequency of watching television	Duration of current pregnancy	Frequency of listening to radio	Number of household members
7	Birth history	Currently breastfeeding	Current pregnancy wanted	Duration of current pregnancy	Frequency of listening to radio
8	History of contraceptive use	Mosquito bed net	History of contraceptive use	birth history	Duration of current pregnancy
9	Body mass index	Husband/partner's education level	Husband/partner's education level	Current pregnancy wanted	birth history
10	Husband/partner's education level	Respondent's occupation	Respondent's occupation	History of contraceptive use	Current pregnancy wanted
11	Husband/partner's occupation	History of the place of delivery	History of the place of delivery	Body mass index	Body mass index
12	Respondent's occupation	Iron tablet during pregnancy	Iron tablet during pregnancy	Husband/partner's education level	Husband/partner's education level
13	History of the place of delivery	Had diarrhea recently	Had diarrhea recently	Husband/partner's occupation	Husband/partner's occupation
14	Vitamin a in last 6 months	Vitamin a in last 6 months	Vitamin a in last 6 months	Respondent's occupation	Respondent's occupation
15	Wealth index combined	Wealth index combined	Wealth index combined	Wealth index combined	Wealth index combined
Accuracy with RF	89.091221	76.120941	82.85518	0.91813755	0.917751321

Table 2  
Features selected with domain experts

No	Features	Feature descriptions
1	m49a	Take drug for malaria during pregnancy
2	H34	Take Vitamin A
3	V106	Highest educational level
4	M15	History of Place of delivery
5	m45	Iron tablet during pregnancy
6	V228	History of terminating a pregnancy
7	V404	Breastfeeding status

### C. Predictive Model Development

To construct a model that predicts the level of anemia among pregnant women in the case of Ethiopia, homogeneous ensemble machine learning algorithms such as extreme gradient boosting, random forest, and cat boost algorithms without applying class decomposition and with applying one versus one and one versus rest class decomposition were selected for an experiment. To show that homogeneous ensemble algorithms can perform better than other supervised machine learning algorithms, another model was developed using decision tree algorithms. For developing the predictive models, 23 features selected by the step forward feature selection method and domain experts were used. Grid search was implemented to tune the hyperparameters of each algorithm, as the performance of the algorithm highly depends on the selection of hyperparameter, which has always been a crucial step in the process of machine learning model development [24][25][26]. The performance of each predictive model was evaluated using accuracy, precision, recall, F1- score, and roc auc.

Figure 1 represents the proposed model architecture that was implemented in this study to develop a predictive model, select the best-performed model, identify risk factors, generate relevant rules, design artifacts, and deploy the final model for the potential set of users.

## IV. Experimental Setup And Results Discussion

Here below results are discussed based on the research questions.

### A) What is the underlying structure of anemia among pregnant women in Ethiopia?

To show the underlying structure of anemia among pregnant in the case of Ethiopia, a descriptive statistical technique was used by considering the age, place of residence, region, antenatal care visit, history of the place of delivery, history of terminating the pregnancy, and wealth index with the anemia level. See Fig. 1 below which represents that pregnant women who live in the rural area of Ethiopia are highly affected by anemia, and in the rural area of Ethiopia the level of anemia shows that 57.2%, 14.1%, 2.5%, and 14.7% of non-anemic mild, severe, and moderate respectively. In Fig. 1 we conclude that every level of anemia in a rural area of Ethiopia was higher than the urban area of Ethiopia.

Figure 2 here below represents that pregnant women with poor economic status were highly affected by anemia. As we see in Fig. 2 here below pregnant women with Poor wealth index status were higher than other wealth index status in every level of anemia.

Figure 3 here below represents that the pregnant women which didn't follow or follow one time only during pregnancy were highly affected by anemia, and pregnant women who follows antenatal care visits repeatedly reduced the level of anemia.

See in Fig. 4 here above, which represents the anemia level distribution among pregnant women with different age groups and, the pregnant women in the age between 30–34 were severely affected by anemia.

As we see in Fig. 5 the Ethiopian regions, like Somalia, afar, Dire Dawa, and snnpr were highly affected by anemia.

*B) Which homogeneous ensemble of machine learning algorithms is suitable for predicting the level of anemia among pregnant women in Ethiopia?*

To answer this question, twelve experiments using three homogeneous ensemble machine learning algorithms namely random forest, extreme gradient boosting, and cat boost with class decomposition (by using one versus one and one versus the rest), and without class decomposition was conducted. To show that homogeneous ensemble algorithms can perform better than other supervised machine learning algorithms, we have also conducted an experiment using decision tree algorithms. The experiments showed that the model that was developed using the cat boost algorithm with one versus the rest class decomposition performs better in predicting the level of anemia among pregnant in the case of Ethiopia with 97.6% of accuracy, 97.59% of precision, 97.57% of recall, 97.58% of f1\_score, and, 99.9% of roc see Table 3 below, using all the tuning parameters of (depth = 10, iterations = 300, l2\_leaf\_reg = 1, learning\_rate = 0.15) extracted with grid search. To develop a model using random forest algorithm also uses (criterion='entropy', max\_features='sqrt', min\_samples\_split = 3, n\_estimators = 500, random\_state = 0, max\_depth = 20, max\_leaf\_nodes = 400, n\_jobs=-1) parameters and performs less than cat boost algorithms, extreme gradient boosting algorithms use a default parameters, and decision tree algorithms use (criterion='entropy',max\_features='sqrt',min\_samples\_split = 12,random\_state = 0,max\_depth = 30, max\_leaf\_nodes = 600) parameters and performs less performance than all other algorithms.

Table 3  
Model performance

	Evaluation metrics	Without class decompositions	With one vs. one class decomposition	With one vs. rest class decomposition
Decision tree	Accuracy	79.38%	89.88%	89.09%
	precision	79.09%	89.81%	89.01%
	Recall	79.21%	89.77%	88.98%
	F1_score	79.03%	89.71%	88.96%
	Cross validation	68.48%	84.27%	83.17%
Random forest	Accuracy	91.34%	94.4%	94.4%
	Precision	91.32%	94.36%	94.37%
	Recall	91.28%	94.35%	94.35%
	F1_score	91.25%	94.34%	94.34%
	Cross validation	81.23%	89.37%	88.18%
	ROC	99%	-	99.43%
Cat Boost	Accuracy	97.08%	97.44%	<b>97.595%</b>
	Precision	97.09%	97.438%	<b>97.596%</b>
	Recall	97.05%	97.418%	<b>97.574%</b>
	F1_score	97.06%	97.422%	<b>97.58%</b>
	Cross validation	95.94%	96.478%	<b>96.482%</b>
	ROC	99.9%	-	<b>99.9%</b>
Extreme gradient Boost	Accuracy	94.26%	95.21%	94.54%
	Precision	94.27%	95.20%	94.53%
	Recall	94.20%	95.16%	94.48%
	F1_score	94.20%	95.16%	94.48%
	Cross validation	88.86%	91.73%	89.72%
	ROC	99.53%	-	99.54%

C) *What are the associated risk factors that influence the occurrence of anemia among pregnant women in the case of Ethiopia?*

To answer this question, feature importance analysis was performed using the model that was developed with the best performing algorithm which is cat boost. Table 4 presents the most important risk factors that determines the level of anemia among pregnant women in Ethiopia.

Table 4  
Identified risk factors with best fit model and feature importance

Feature	Values	Feature	Values
Duration of current pregnancy	10.3953193	Current pregnancy wanted	3.838873474
Age in 5-year groups	9.69394377	Body mass index	2.787116569
Source of drinking water	8.99369175	Number of ANC visits	2.600944933
History of contraceptive use	6.61405164	Highest educational level	2.419310637
Respondent's occupation	6.12946203	History of terminating a pregnancy	0.849814164
Number of household members	5.85914199	Currently breastfeeding	0.732357678
Wealth index	5.63211101	Type of place of residence	0.576997215
Frequency of listening to the radio	5.16045505	Vitamin A in last 6 months	0.356953114
Husband/partner's education level	5.02943094	During pregnancy, given or bought iron tablets/syrup	0.046775106
Region	4.3314029	History of Place of delivery	0.010932682
Husband/partner's occupation	3.96855455	During pregnancy took: sp/ fansidar for malaria	0.00058328
Birth history	3.87177534		

D) *What are the important rules that can be generated from the predictive model?*

To answer this question, we used all the features that we used to develop the predictive model and generate all the important rules by using the best-performed algorithms (cat boost algorithms with one versus rest class decompositions) for the level of anemia among pregnant in the case of Ethiopia. The most important rules that were also validated by domain experts are presented here below:

**RULE1**, IF given iron tablet or syrup during pregnancy == 'No' AND vitamin A in last 6 months == 'No' AND during pregnancy took sp fansidar for malaria== 'No' AND region == 'Somali' AND currently breastfeeding == 'No' AND place of residence == 'rural' AND Duration of current pregnancy == 'seven-nine-week' AND current pregnancy wanted == 'Yes' AND respondents occupation == 'did not work' AND history of place of delivery == 'Home' AND age == 'thirty - thirty four' AND educational level == 'no education' AND husband educational level == 'no education' AND number of household== 'six-ten' AND history of terminating pregnancy== 'No' AND body mass index == 'normal' AND husband occupation == 'did not work' THEN anemia level== 'sever'.

**RULE2**, IF given iron tablet or syrup during pregnancy == 'No' AND vitamin A in last 6 months == 'No' AND during pregnancy took sp fansidar for malaria== 'No' AND region == 'Somali' AND currently breastfeeding == 'No' AND place of residence == 'rural' AND Duration of current pregnancy == 'seven-nine-week' AND current pregnancy wanted == 'Yes' AND respondents occupation == 'did not work' AND place of delivery == 'Home' AND age == 'thirty - thirty four' AND educational level == 'no education' AND husband educational level == 'no education' AND number of household== 'six-ten' AND History of terminating pregnancy== 'No' AND body mass index == 'normal' AND husband occupation == ' agricultural - employee' AND source of water == 'pure' AND history of contraceptive use == 'Yes' THEN anemia level== 'none anemic'.

**RULE3**, IF given iron tablet or syrup during pregnancy == 'No' AND vitamin A in last 6 months == 'No' AND during pregnancy took sp fansidar for malaria== 'No' AND region == 'Somali' AND currently breastfeeding == 'No' AND place of residence == 'rural' AND Duration of current pregnancy == 'seven-nine-week' AND current pregnancy wanted == 'Yes' AND respondents

occupation == 'did not work' AND history of place of delivery == 'Home' AND age == 'thirty - thirty four' AND educational level == 'no education' AND husband educational level == 'no education' AND number of household== 'six-ten' AND history of terminating pregnancy== 'No' AND body mass index == 'normal' AND husband occupation == ' agricultural - employee' AND source of water == 'not pure' AND history of contraceptive use == 'Yes' THEN anemia level== 'Moderate'.

Finally, the predictive model was deployed on cloud for potential users. The artifact was designed using a Python module called flask framework with HTML and deployed on Heroku. All potential users can access (<https://anemia-level-prediction-model.herokuapp.com/>) the predictive model to evaluate a pregnant woman's level of anemia.

## V. Conclusion

Anemia is a global public health issue that affects a wide range of people of all ages. Anemia during pregnancy is one of the risk factors for poor pregnancy outcomes, such as low birth weight, preterm birth, prematurity stillbirth, intrauterine growth restriction, and impaired cognitive development. This study aimed to develop a predictive model for the level of anemia among pregnant women in the case of Ethiopia by using homogeneous ensemble machine learning algorithms. This study was conducted using design science methodology. The proposed model was constructed using homogeneous ensemble machine learning algorithms namely random forest, extreme gradient boosting, and cat boost algorithms with class decomposition methods and without class decomposition methods. To conduct this study we have done a total of twelve experiments. The cat boost algorithm with one versus all class decomposition has registered the highest performance with 97.6% of accuracy, 97.59% of precision, 97.57% of recall, 97.58% of f1\_score, and 96.48% of cross-validation. We have identified the determinant risk factors by conducting feature importance analysis on the best-performed algorithms. Some of the most determinant risk factors were duration of current pregnancy, age in five years group, source of drinking water, history of contraceptive use, respondent's occupation, and several household members. The most important rules were also generated using the best fit model for developing policies and interventions towards maintaining anemia among pregnant women.

Finally, we recommend that future researchers conduct a predictive model for pregnant women that predicts which type (Vitamin deficiency anemia, Anemia of inflammation, Aplastic anemia, or iron-deficiency anemia) of anemia is occurred within the pregnant women. A predictive model that can predict the level of anemia among neonatal based on maternal determinants during pregnancy. The determinant risk factors over time.

## Abbreviations

ANN: Artificial Neural Network; EDHS: Ethiopian Demographic and Health Survey; Hb: Hemoglobin; HTML: Hypertext Markup Language; LBW: Low Birth Weight; RP: Random Prediction; SMOTE: Synthetic Minority Over-sampling Technique; snnpr: Southern nation and nationality of people.

## Declarations

### Acknowledgment

We would like to acknowledge the Ethiopian central statistics for providing us the data with a data set description.

### Authors' contributions

Belayneh conceived and designed the study, participated in data analysis, wrote the report, finished the model refinements, carried out deep analysis of the experiment results, drafted and revised the initial manuscript, and revised the manuscript; Tesfamariam managed the quality and progress of the whole study, and revised the manuscript; Dawit revised the manuscript; all authors read and approved the final manuscript.

## Funding

The research was supported by the University of Gondar research and community service vice president's office.

## Availability of data and materials

The datasets generated and/or analysed during the current study are available in the 'Anemia level' repository, [https://github.com/belzman/Anemia\\_level](https://github.com/belzman/Anemia_level).

## Ethics approval and consent to participate

All methods used in this study followed guidelines and regulations. Health care professionals who work on antenatal care services from the University of Gondar specialized hospital approved this study.

## Consent for publication

Not applicable.

## Competing interests

The authors report that they have no conflicts.

## Author Details

All authors of this study are affiliated to the University of Gondar, College of Informatics, Gondar, Ethiopia.

## References

1. A. R. Kavsaolu, K. Polat, and M. Hariharan, "Non-invasive prediction of hemoglobin level using machine learning techniques with the PPG signal's characteristics features," *Appl. Soft Comput. J.*, vol. 37, pp. 983–991, 2015, doi: 10.1016/j.asoc.2015.04.008.
2. F. Habyarimana, T. Zewotir, and S. Ramroop, "Prevalence and risk factors associated with anemia among women of childbearing age in Rwanda," *Afr. J. Reprod. Health*, vol. 24, no. 2, pp. 141–151, 2020, doi: 10.29063/ajrh2020/v24i2.14.
3. W. Worku Takele, A. Tariku, F. Wagnew Shiferaw, A. Demsie, W. G. Alemu, and D. Zelalem Anlay, "Anemia among Women Attending Antenatal Care at the University of Gondar Comprehensive Specialized Referral Hospital, Northwest Ethiopia, 2017," *Anemia*, vol. 2018, 2018, doi: 10.1155/2018/7618959.
4. G. Stephen, M. Mgongo, T. H. Hashim, J. Katanga, B. Stray-pedersen, and S. E. Msuya, "Anaemia in Pregnancy: Prevalence, Risk Factors, and Adverse Perinatal Outcomes in Northern Tanzania," vol. 2018, 2018.
5. S. K. Ndegwa and S. K. Ndegwa, "Anemia & Its Associated Factors Among Pregnant Women Attending Antenatal Clinic At Mbagathi County Hospital, Nairobi County, Kenya," vol. 32, no. 1, pp. 59–73, 2019.
6. W. Gari, A. Tsegaye, and T. Ketema, "Magnitude of anemia and its associated factors among pregnant women attending antenatal care at Najjo General Hospital, northwest Ethiopia," *Anemia*, vol. 2020, pp. 1–8, 2020, doi: 10.1155/2020/8851997.
7. T. A. Gudeta, T. M. Regassa, and A. S. Belay, "Magnitude and factors associated with anemia among pregnant women attending antenatal care in Bench Maji, Keffa and Sheka zones of public hospitals, Southwest, Ethiopia, 2018: A cross-sectional study," *PLoS One*, vol. 14, no. 11, pp. 30–34, 2019, doi: 10.1371/journal.pone.0225148.
8. A. Gebreweld and A. Tsegaye, "Prevalence and Factors Associated with Anemia among Pregnant Women Attending Antenatal Clinic at St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia," *Adv. Hematol.*, vol. 2018,

2018, doi: 10.1155/2018/3942301.

9. M. S. Teshome, D. H. Meskel, and B. Wondafrash, "Determinants of anemia among pregnant women attending antenatal care clinic at public health facilities in kacha birra district, southern ethiopia," *J. Multidiscip. Healthc.*, vol. 13, pp. 1007–1015, 2020, doi: 10.2147/JMDH.S259882.
10. B. Zekarias, A. Meleko, A. Hayder, A. Nigatu, and T. Yetagessu, "Prevalence of Anemia and its Associated Factors among Pregnant Women Attending Antenatal Care (ANC) In Mizan Tepi University Teaching Hospital, South West Ethiopia," *Heal. Sci. J.*, vol. 11, no. 5, pp. 1–8, 2017, doi: 10.21767/1791-809x.1000529.
11. F. Weldekidan, M. Kote, M. Girma, N. Boti, and T. Gultie, "Determinants of Anemia among Pregnant Women Attending Antenatal Clinic in Public Health Facilities at Durame Town: Unmatched Case Control Study," vol. 2018, 2018.
12. M. O. Osman, T. Y. Nour, H. M. Bashir, A. K. Roble, A. M. Nur, and A. O. Abdilahi, "Risk factors for anemia among pregnant women attending the antenatal care unit in selected jigjiga public health facilities, somali region, east ethiopia 2019: Unmatched case–control study," *J. Multidiscip. Healthc.*, vol. 13, pp. 769–777, 2020, doi: 10.2147/JMDH.S260398.
13. B. Berhe, F. Mardu, H. Legese, A. Gebrewahd, G. Gebremariam, and K. Tesfay, "Prevalence of anemia and associated factors among pregnant women in Adigrat General," *BMC Res. Notes*, pp. 1–6, 2019, doi: 10.1186/s13104-019-4347-4.
14. D. Getaneh, A. Bayeh, B. Belay, T. Tsehaye, and Z. Mekonnen, "Assessment of the Prevalence of Anemia and Its Associated Factors among Pregnant Women in Bahir Dar City Administration, North-West Ethiopia," *J. Pregnancy Child Heal.*, vol. 05, no. 02, 2018, doi: 10.4172/2376-127x.1000367.
15. R. C. Solem, "Limitation of a cross-sectional study," *Am. J. Orthod. Dentofac. Orthop.*, vol. 148, no. 2, p. 205, 2015, doi: 10.1016/j.ajodo.2015.05.006.
16. A. M. Abaidullah, N. Ahmed, and E. Ali, "Identifying Hidden Patterns in Students' Feedback through Cluster Analysis," *Int. J. Comput. Theory Eng.*, vol. 7, no. 1, pp. 16–20, 2014, doi: 10.7763/ijcte.2015.v7.923.
17. M. D. Dithy and V. Krishnapriya, "Predicting Anemia in Pregnant Women By Using Gausnominal," vol. 118, no. 20, pp. 3343–3349, 2018.
18. M. D. Dithy and V. Krishnapriya, "Anemia selection in pregnant women by using random prediction (Rp) classification algorithm," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2, pp. 2623–2630, 2019, doi: 10.35940/ijrte.B3016.078219.
19. S. S. Yadav and S. M. Jadhav, "Machine learning algorithms for disease prediction using lot environment," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 4303–4307, 2019, doi: 10.35940/ijeat.F8914.088619.
20. P. Anand, R. Gupta, and A. Sharma, "Prediction of Anaemia among children using Machine Learning Algorithms," no. June, pp. 469–480, 2020.
21. I. Journal and C. Science, "Class Imbalance Problem in Data Mining: Review," vol. 2, no. 1, 2013.
22. R. P. Ribeiro, "SMOTE for Regression," no. October 2015, 2013, doi: 10.1007/978-3-642-40669-0.
23. S. Wang, J. Tang, H. Liu, and E. Lansing, "Encyclopedia of Machine Learning and Data Mining," *Enycl. Mach. Learn. Data Min.*, pp. 1–9, 2016, doi: 10.1007/978-1-4899-7502-7.
24. M. J. Healy, "Statistics from the inside. 15. Multiple regression (1).," *Arch. Dis. Child.*, vol. 73, no. 2, pp. 177–181, 1995, doi: 10.1136/adc.73.2.177.
25. R. G. Mantovani, A. L. D. Rossi, E. Alcobaça, J. C. Gertrudes, S. B. Junior, and A. C. P. de L. F. de Carvalho, "Rethinking Defaults Values: a Low Cost and Efficient Strategy to Define Hyperparameters," 2020, [Online]. Available: <http://arxiv.org/abs/2008.00025>.
26. m. M. Ramadhan, i. S. Sitanggang, f. R. Nasution, and a. Ghifari, "Parameter Tuning in Random Forest Based on Grid Search Method for Gender Classification Based on Voice Frequency," *DEStech Trans. Comput. Sci. Eng.*, no. cece, 2017, doi: 10.12783/dtcse/cece2017/14611.



prevalence of anemia among pregnant women Based on region