ISSN 2458-8865 E-ISSN 2459-1505
www.fppc.com.tr
Family Practice & Palliative Care

Research Article

Analyzing of iron-deficiency anemia in pregnancy using rule-based intelligent classification models

Gebelikte demir eksikliği anemisinin kural tabanlı akıllı sınıflandırma modelleri kullanılarak incelenmesi

Mehmet Onur Kaya[a], Ruveyda Yildirim[a], Burkay Yakar[b] Bilal Alatas[c]

[a] Department of Biostatistics and Medical Informatics, Faculty of Medicine, Firat University, Elazığ, Türkiye

[b] Department of Family Medicine, Faculty of Medicine, Firat University, Elazığ, Türkiye

[c] Department of Software Engineering, Faculty of Engineering, Firat University, Elazığ, Türkiye

Abstract

Introduction: Iron deficiency anemia is the most common cause of anemia worldwide, and increased iron requirement during pregnancy increases the risk of anemia. Anemia in pregnancy is associated with adverse pregnancy outcomes such as low birth weight, preterm and intrauterine growth restriction. This study used a Rule-based Intelligent Classification Models to predict socio-demographic, nutritional, antenatal care and obstetric factors on iron deficiency anemia during pregnancy

Methods: This retrospective study was a secondary analysis of a community-based cross-sectional study conducted between January and June 2019 in the province of Elazig in eastern Turkey. Data of 495 pregnant women were included in the study iron deficiency anemia was defined as hemoglobin < 11 g/dl, and ferritin < 30 µg/L. Rule-based machine learning methods were used to predict factors associated with anemia during pregnancy.

Results: The mean age of 495 pregnant women were 30.06 ± 5.15 years. The prevalence of anemia was 27.9% in study population. Maternal age, educational status, occupation, nutrition education status, nutritional property, gravida, and parity were significantly related to anemia. Jrip, OneR, and PART algorithms estimated factors associated with anemia with 96.36%, 85.45%, and 97.98% accuracy, respectively.

Conclusion: Rule-based machine learning algorithm may offer a new approach to risk factors for iron deficiency anemia during pregnancy. With the use of this model, it is possible to predict the risk of anemia both before and during pregnancy and to take preventative measures.

Keywords: Pregnancy, Iron-Deficiency Anemia, Algorithms, Machine learning

Öz

Giriş: Demir eksikliği anemisi, dünya çapında aneminin en yaygın nedenidir ve hamilelik sırasında artan demir gereksinimi anemi riskini artırır. Gebelikte anemi, düşük doğum ağırlığı, preterm ve intrauterin gelişme geriliği gibi olumsuz gebelik sonuçları ile ilişkilidir. Bu çalışma, gebelik sırasında demir eksikliği anemisi üzerindeki sosyo-demografik, beslenme, antenatal bakım ve obstetrik faktörleri tahmin etmek için Kural Tabanlı Akıllı Sınıflandırma Modelleri kullanmıştır.

Yöntem: Bu retrospektif çalışma, Türkiye'nin doğusundaki Elazığ ilinde Ocak ve Haziran 2019 tarihleri arasında yürütülen toplum temelli kesitsel bir çalışmanın ikincil bir analiziydi. Çalışmaya 495 gebenin verileri dahil edildi. Demir eksikliği anemisi hemoglobin < 11,0 g/dl ve ferritin <30,0 µg/L olarak tanımlandı. Hamilelik sırasında anemi ile ilişkili faktörleri tahmin etmek için kural tabanlı makine öğrenimi yöntemleri kullanıldı.

Bulgular: 495 gebenin yaş ortalaması 30,06 ± 5,15 yıldı. Çalışma popülasyonunda anemi prevalansı %27,9 idi. Anne yaşı, eğitim durumu, meslek, beslenme eğitimi durumu, beslenme özelliği, gravida ve parite anemi ile anlamlı şekilde ilişkiliydi. Jrip, OneR ve PART algoritmaları anemi ile ilişkili faktörleri sırasıyla %96,36, %85,45 ve %97,98 doğrulukla tahmin etti.

Sonuç: Kural tabanlı makine öğrenimi algoritması, hamilelik sırasında demir eksikliği anemisi için risk faktörlerine yeni bir yaklaşım sunabilir. Bu model ile gebelik öncesi ve gebelik anında anemi riski tahmin edilebilir ve önleyici girişimler yapılabilir.

Anahtar kelimeler: Gebelik, Demir eksikliği anemisi, Algoritmalar, Makine öğrenme

Received	Accepted	Published Online	Corresponding Author	E-mail
August 22, 2023	December 13, 2023	December 28, 2023	Mehmet Onur Kaya, PhD	mokaya@firat.edu.tr
Correspondence	PhD Mehmet Onur Kaya. Department of Biostatistics and Medical Informatics, Faculty of Medicine, Firat University, 23119 Elazığ, Türkiye			
doi	https://doi.org/10.22391/fppc.1347373			

Key Points

1. Machine learning algorithms may used to construct a risk prediction model of anemia during pregnancy.
2. The major predictive factors were found to be dark tea, age, gravida, nutritional education status, and nutritional diversity.
3. Supervised machine learning methods predicted anemia precursors with 85.5% to 97.9% accuracy.
4. With existing algorithms, the risk of pregnancy anemia can be predicted and early preventive measures can be taken.

Introduction

Anemia during pregnancy is an important health problem because of its high prevalence, maternal and perinatal mortality, and morbidity. Anemia affects 16-62% of pregnant women in developing countries and 16-29% in developed countries. Anemia in pregnancy can reach a frightening prevalence of 57.1% and 48.2% in Africa and South East Asia, respectively. Anemia during pregnancy has been associated with maternal mortality, preterm birth, preeclampsia, low birth weight, intrauterine growth retardation, and increased cesarean section risk [1, 2].

Previous studies have reported that the prevalence of anemia in pregnancy varies in different cultures in women with different socio-economic conditions, lifestyles, or health-seeking behaviors. Pregnancy at an early age, short birth intervals, the high number of gravida, cesarean section, and iron prophylaxis treatment are the factors affecting the prevalence of anemia in pregnancy [3-5]. Strategies to prevent anemia in pregnancy include excluding underlying causes and iron supplementation during pregnancy. In this context, WHO recommends antenatal counseling at least 4 times and iron supplementation during pregnancy to all pregnant women [6].

As explained above, anemia during pregnancy is a high-prevalence and important public health problem. The fact that multiple factors play a role in its etiology makes it difficult to reduce the prevalence of pregnancy anemia. Machine learning algorithms can examine vast datasets, accept hundreds of various kinds of variables, and require just a few statistical presumptions. In order to gain fresh insights into the epidemiology of pregnant anemia, machine learning (ML) methods can also effectively and reliable quantify complicated correlations between factors. Additionally, ML algorithms are the most effective instruments for enhancing clinical performance since they may exceed conventional statistical approaches in individual-level predictions [7, 8]. The current study aimed to test the success of the supervised machine learning methods on the dataset obtained to examine the factors associated with anemia during pregnancy.

Methods

The current study was a secondary analysis of a Community-based cross-sectional study conducted between January and June 2019 in the province of Elazig in eastern Turkey. The data of this study are available from the corresponding author upon reasonable request. The formula "$n=Nt^2pq/d^2(N-1)+t^2pq$" was used to calculate the sample size. The prevalence of anemia in pregnancy has been reported by the WHO as 28% in Turkey. The minimum sample size was found 282 pregnant women (N=3228, t=1.96, p=28%, q=72%, and d=0.05). The study population, sample size, and sampling method were also specified in the previous study by Yakar et al. [4]. Medical records of 495 pregnant women were retrospectively analyzed and included in the study [4]. The study's exclusion criteria are defined as; eclampsia, preeclampsia, diabetes, hypertension, sickle cell disease, any congenital hemoglobinopathy, hearing, and vision impairments, and mental disorders.

Variables

Independent variables included; Sociodemographic characteristics (Age, Educational Status, Occupation, Income), nutritional characteristics (Tea Preferences, Tea Consumption Time, Number of Meals Per Day, Intake of Red Meat, Intake of Egg, Intake of Cereals, Intake of Vegetables, Intake of Fruits), antenatal care characteristics (Iron Supplementation Time, Number of Prenatal Care Follow-Up, Nutritional Education Status) and pregnancy characteristics (Gravida, and trimester) of the participants. The dependent variable of study is iron deficiency anemia. Participants were grouped as anemic and non-anemic according to the WHO cut-off value of hemoglobin 11gr/dl for anemia in pregnancy. Iron deficiency anemia was defined as hemoglobin < 11 g/dl, and ferritin < 30 μg/L [2].

Rule-based supervised machine learning algorithms

Simple artificial intelligence models that make use of the if-then coding convention are systems that implement artificial intelligence using a rule-based approach. In contrast to other machine learning techniques, rule-based AI models can function with simple, fundamental data and knowledge. The system is interpreted, explicable, and complete thanks to these rules. The approach is more understandable than previous block-box machine learning models as a result of automatically summarizing and analyzing the data with found rules.

Jrip, a propositional rule learner, upgraded IREP to produce decision rules that are as effective as or more effective than decision trees [9]. Three steps are taken to achieve this approach. In the first stage (grow), criteria are added to a rule until it correctly classifies as a subset of data using the separate-and-conquer approach. Decision trees employ information acquired to determine the next splitting characteristic. The rule is pruned as a second step when increasing a rule's specificity no longer decreases entropy (prune). The grow and prune processes are continued until the termination criterion is reached, at which time the rules are optimized using heuristics, and this third step is referred to as optimization. Using the cat swarm optimization technique, accurate and understandable numerical classification rules may be automatically mined [10].

OneR is a straightforward yet reliable classification technique that generates a rule for every predictor in the data before choosing the rule with the fewest sum of errors as the "one rule". OneR returns an attribute with one or more stated decision rules. In essence, as a basic classifier, it identifies a single attribute (and a single or more attribute value for that attribute) to categorize data [11].

As a relatively straightforward algorithm, the PART method does not do global optimization to provide accurate rules. This technique retrieves rules one at a time and only constructs incomplete decision trees. When instances covered by a rule are eliminated, the algorithm continues to build a recursive rule for instances rest until no more instances are present [12].

Using the collected data, supervised machine learning techniques were used to build rule-based classification models that can identify anemia. In rule-based supervised machine learning, algorithms for the anemia detection system developed using Jrip, OneR, and Part were compared in all training data and test data using several success metrics including accuracy, precision, ROC area, and F-Measure. K-fold cross-validation techniques and the selection of a specific portion of the training set were employed for the test data [13].

Ethical approval, informed consent and permissions

The current study protocol was approved by Firat University's non-interventional research ethics committee (date: 08.04.2021, IRB number: 2021/05/05). Written consent of the participants was not required as it was a retrospective study. Necessary permissions were obtained from the corresponding author for the reanalysis of the data of the study entitled "Prevalence of Anemia and Associated Risk Factors among Pregnant Women, What is the Role of Antenatal Care in Prevention? A Cross-sectional Study".

Statistical analysis

The IBM SPSS version 25.0 (IBM Corp., Armonk, New York, USA) package application was used to statistically analyze the study's data. The Shapiro-Wilk test was used to assess how closely quantitative variables adhered to the normal distribution. Descriptive statistics of the data are indicated as Median (minimum-maximum) for continuous data, as frequency and percentage [n (%)] for categorical variables. The Mann-Whitney U test was used to compare quantitative data between two independent groups that did not meet the normal distribution, and the descriptive statistics were provided as Median (Min–Max). Pearson and Fisher Exact Chi-Square tests were used to compare qualitative variables between two independent groups, while frequency (n) and percentage (%) were used to show descriptive statistics. The significance threshold of 0.05 was approved.

Results

The mean age of 495 pregnant women included in the study was 30.06 ± 5.15 years. The prevalence of iron deficiency anemia in the study population was 27.87% (n=138). Demographic and nutritional characteristics of anemic and non-anemic participants were compared in Table 1.

The Jrip algorithm, which is one of the supervised machine learning methods, is presented in Table 2. The Jrip algorithm has determined 11 rules and the analysis time required for the model is 0.08 seconds. The first rule of the Jrip algorithm associated anemia with dark tea, 3 or more pregnancies, and age ≤ 38 years. The Jrip algorithm predicted correctly 52 out of 56 cases. In the Jrip algorithm, if the first rule conditions are not met, the second rule is taken into consideration and the algorithm proceeds in this way. All rules are presented in Table 2.

Another machine learning method, the rule-based OneR classification algorithm, is presented in Table 3. The time required for model setup was 0.01 seconds. The OneR algorithm used only participants' tea preference for anemia prediction. The OneR algorithm associated the preference for dark tea with anemia (Table 3).

The rule-based PART classification algorithm, which is another machine learning method, is presented in Table 4. The PART algorithm specified 20 rules and the time required for the algorithm was 0.03 seconds. The first rule of the PART algorithm associated non-anemic status with (light tea) and (iron supplementation time ≤ 6 months) and (Gravida ≤ 3) and (Age ≤ 34 years) and (Tea Consumption Time = Only for Breakfast). The first rule of the PART algorithm correctly predicted 171 cases out of 172 cases (Table 4).

Three methods are followed for the test data. In Test 1, all training data were used as test data. In Test 2, the 5-fold cross-validation method is used. In Test 3, 70% of all data were used as training data sets and the remaining as test data set. Accuracy, kappa statistic, mean absolute error, root mean squared error, relative absolute error, root relative squared error, False Positive rate, precision, recall, F-measure, MCC, ROC area, and PRC area metrics of the algorithms obtained using different test methods were listed in Table 5. (Table 5)

Table 1. Demographic characteristics of the variables included in the study

Variables		Anemic (n=138)	Non-Anemic (n=357)	p value
Age [Median (Min–Max)]		31 (20 – 45)	29 (20 – 48)	**0.002[a]**
Gravida [Median (Min–Max)]		3 (1 – 9)	2 (1 – 6)	**<0.001[a]**
Iron Supplementation Time [Median (Min–Max)]		0 (0 – 60)	0 (0 – 120)	0.776[a]
Educational Status n (%)	Illiterate	9 (6.5)	3 (0.8)	**<0.001[b]**
	Primary school	18 (13.0)	37 (10.4)	
	Middle school	22 (15.9)	56 (15.7)	
	High school	52 (37.7)	112 (31.4)	
	College	37 (26.8)	149 (41.7)	
Occupation n (%)	Housewife	107 (77.5)	231 (64.7)	**0.002[b]**
	Officer	20 (14.5)	94 (26.3)	
	Minimum wage employe	9 (6.5)	32 (9.0)	
	Other	2 (1.4)	0 (0.0)	
Income n (%)	Poor	13 (9.4)	20 (5.6)	0.276[b]
	Medium	102 (73.9)	268 (75.1)	
	Good	23 (16.7)	69 (19.3)	
Tea Preferences n (%)	Not drinking tea	9 (6.5)	33 (9.2)	**<0.001[b]**
	Light tea	38 (27.5)	299 (83.8)	
	Dark tea	91 (65.9)	25 (7.0)	
Tea Consumption Time n (%)	Not drinking tea	9 (6.5)	33 (9.2)	**<0.001[b]**
	Only for breakfast	70 (50.7)	244 (68.3)	
	Within 1 hour after meals	46 (33.3)	38 (10.6)	
	At least 1 hour after meals	13 (9.4)	42 (11.8)	

Trimester n (%)	1. Trimester	15 (10.9)	82 (23.0)	**<0.001[b]**
	2. Trimester	60 (43.5)	170 (47.6)	
	3. Trimester	63 (45.7)	105 (29.4)	
Number of Prenatal Care Follow-Up n (%)	Once	19 (13.8)	71 (19.9)	0.061[b]
	Two times	35 (25.4)	120 (33.6)	
	Three times	42 (30.4)	92 (25.8)	
	Four times	37 (26.8)	65 (18.2)	
	Five and more	5 (3.6)	9 (2.5)	
Nutrional Education Status n (%)	Yes	68 (49.3)	226 (63.3)	**0.004[b]**
	No	70 (50.7)	131 (36.7)	
Number of Meals Per Day n (%)	Two meals	28 (20.3)	49 (13.7)	0.097[b]
	Three meals	64 (46.4)	200 (56.0)	
	Four meals	36 (26.1)	93 (26.1)	
	Five and more	10 (7.2)	15 (4.2)	
Intake of Red Meat n (%)	Everyday	12 (8.7)	49 (13.7)	**<0.001[b]**
	Every other day	33 (23.9)	146 (40.9)	
	One a week	71 (51.4)	131 (36.7)	
	Less	21 (15.2)	29 (8.1)	
	None	1 (0.7)	2 (0.6)	
Intake of Egg n (%)	Everyday	67 (48.6)	156 (43.7)	**<0.001[b]**
	Every other day	39 (28.3)	154 (43.1)	
	One a week	21 (15.2)	21 (5.9)	
	Less	10 (7.2)	11 (3.1)	
	None	1 (0.7)	15 (4.2)	
Intake of Cereals n (%)	Everyday	33 (23.9)	116 (32.5)	0.356[b]
	Every other day	68 (49.3)	153 (42.9)	
	One a week	36 (26.1)	86 (24.1)	
	Less	0 (0.0)	1 (0.3)	
	None	1 (0.7)	1 (0.3)	
Intake of Vegetables n (%)	Everyday	67 (48.6)	264 (73.9)	**<0.001[b]**
	Every other day	60 (43.5)	70 (19.6)	
	One a week	9 (6.5)	21 (5.9)	
	Less	2 (1.4)	2 (0.6)	
Intake of Fruits n (%)	Everyday	23 (16.7)	48 (13.4)	**0.008[b]**
	Every other day	18 (13.0)	98 (27.5)	
	One a week	57 (41.3)	124 (34.7)	
	Less	36 (26.1)	69 (19.3)	
	None	4 (2.9)	18 (5.0)	

[a] Mann-Whitney U test, [b] Pearson or Fisher Exact Chi Square test

Table 2. JRIP rules list

Rules Number	The Left of The Rule	The Right of The Rule	The Correct Number	The Wrong Number
1	IF (Tea Preferences= Dark Tea) AND (Gravida ≥ 3) AND (Age ≤ 38)	THEN (Anemia Status = Anemic)	52	4
2	IF (Tea Preferences= Dark Tea) AND (Nutrional Education Status = Yes) AND (Tea Consumption Time = Within 1 Hour After Meals)	THEN (Anemia Status = Anemic)	10	0
3	IF (Tea Preferences= Dark Tea) AND (Intake of Red Meat = Every Other Day)	THEN (Anemia Status = Anemic)	18	3
4	IF (Age ≥ 34) AND (Intake of Fruits = Less) AND (Iron Supplementation Time ≤ 0) AND (Age ≤ 44)	THEN (Anemia Status = Anemic)	13	2
5	IF (Iron Supplementation Time ≥ 10) AND (Age ≤ 27)	THEN (Anemia Status=Anemic)	12	4
6	IF (Gravida ≥ 8)	THEN (Anemia Status = Anemic)	5	0
7	IF (Number of Prenatal Care Follow-Up = Five and More) AND (Educational Status = Primary)	THEN (Anemia Status = Anemic)	5	0
8	IF (İntake of Cereals = One a Week) AND (Intake of Egg = Everyday) AND (Educational Status = High School) AND (Nutrional Education Status = Yes)	THEN (Anemia Status = Anemic)	9	0
9	IF (Tea Consumption Time = Within 1 Houre After Meals) AND (İntake of Fruits = Less) AND (Age ≤ 28)	THEN Anemia Status = Anemic)	7	0
10	IF (Age ≥ 44) AND (Age ≤ 45)	THEN (Anemia Status = Anemic)	2	0
11	IF other	THEN (Anemia Status = Non-Anemic)	344	5

Table 3. OneR rules list

Rules Number	The Left of The Rule	The Right of The Rule
1	IF (Tea Preferences = Not Drinking Tea)	THEN (Anemia Status = Non-Anemic)
2	IF (Tea Preferences = Light Tea)	THEN (Anemia Status = Non-Anemic)
3	IF (Tea Preferences = Dark Tea)	THEN (Anemia Status = Anemic)

Table 4. PART rules list

Rules Number	The Left of The Rule	The Right of The Rule	The Correct Number	The Wrong Number
1	IF (Tea Preferences = Light Tea) AND (Iron Supplementation Time ≤ 6) AND (Gravida ≤ 3) AND (Age ≤ 34) AND (Tea Consumption Time = Only for Breakfast)	THEN (Anemia Status = Non- Anemic)	171	1
2	IF (Tea Preferences = Not Drinking Tea) AND (Intake of Egg = Every Other Day)	THEN (Anemia Status = Non-Anemic)	24	0
3	IF (Tea Preferences = Dark Tea) AND (Number of Prenatal Care Follow-Up = Four Times)	THEN (Anemia Status = Anemic)	30	1
4	IF (Tea Preferences = Light Tea) AND (İntake of Vegetables = Everyday) AND (Iron Supplementation Time ≤ 10) AND (Educational Status = College)	THEN (Anemia Status = Non-Anemic)	50	1
5	IF (Tea Preferences = Dark Tea) AND (Age ≤ 38) AND (İntake of Red Meat = One a Week) AND (Income = Medium)	THEN (Anemia Status = Anemic)	34	3
6	IF (Educational Status = Middle School)	THEN (Anemia Status = Non-Anemic)	43	1
7	IF (Age > 38) AND (Tea Consumption Time = Only for Breakfast)	THEN (Anemia Status = Non-Anemic)	11	0
8	IF (Tea Preferences = Not Drinking Tea) AND (Occupation = Housewife)	THEN (Anemia Status = Anemic)	8	0
9	IF (Tea Preferences = Dark Tea) AND (İntake of Red Meat = Less)	THEN (Anemia Status = Anemic)	12	0
10	IF (Intake of Vegetables = Every Other Day) AND (Occupation = Housewife) AND (Iron Supplementation Time ≤ 20)	THEN (Anemia Status = Anemic)	15	0
11	IF (Gravida ≤ 6) AND (Tea Preferences = Light Tea) AND (Intake of Cereals = Every Other Day)	THEN (Anemia Status = Non-Anemic)	24	0
12	IF (Intake of Vegetables = One a Week)	THEN (Anemia Status = Non-Anemic)	5	1
13	IF (Income = Good) AND (Tea Preferences = Dark Tea) AND (Number of Prenatal Care Follow-Up = Three Times)	THEN (Anemia Status = Anemic)	6	0
14	IF (Income = Bad)	THEN (Anemia Status = Anemic)	5	0
15	IF (Gravida ≤ 5) AND (Number of Prenatal Care Follow-Up = Once) AND (Occupation = Housewife)	THEN (Anemia Status = Non-Anemic)	8	0
16	IF (Gravida ≤ 2) AND (Intake of Red Meat = Every Other Day) AND (Educational Status = College)	THEN (Anemia Status = Anemic)	9	0
17	IF (Income = Medium) AND (Intake of Red Meat = Every Other Day) AND (Number of Meals Per Day = Three Meals) AND (Intake of Vegetables = Everyday)	THEN (Anemia Status = Anemic)	6	2
18	IF (Intake of Red Meat = Every Other Day)	THEN (Anemia Status = Non-Anemic)	10	0
19	IF (Income = Medium)	THEN (Anemia Status = Anemic)	9	0
20	IF Other	THEN (Anemia Status = Non-Anemic)	5	0

Table 5. Metric values obtained from the methods

	Jrip Algorithm	OneR Algorithm	PART Algorithm
Accuracy (Test 1)	96.36%	85.45%	97.98%
Accuracy (Test 2)	91.92%	85.45%	91.92%
Accuracy (Test 3)	89.19%	86.49%	88.51%
Kappa statistic (Test 1)	0.911	0.620	0.950
Kappa statistic (Test 2)	0.805	0.620	0.803
Kappa statistic (Test 3)	0.726	0.642	0.707
Mean absolute error (Test 1)	0.064	0.146	0.036
Mean absolute error (Test 2)	0.107	0.146	0.099

Mean absolute error (Test 3)	0.133	0.135	0.126
Root mean squared error (Test 1)	0.180	0.381	0.135
Root mean squared error (Test 2)	0.255	0.381	0.273
Root mean squared error (Test 3)	0.322	0.368	0.326
Relative absolute error (Test 1)	16.01%	36.14%	9.04%
Relative absolute error (Test 2)	26.50%	36.13%	24.62%
Relative absolute error (Test 3)	32.97%	33.45%	31.23%
Root relative squared error (Test 1)	40.03%	85.05%	30.09%
Root relative squared error (Test 2)	56.95%	85.05%	60.82%
Root relative squared error (Test 3)	71.32%	81.53%	72.23%
Detailed Accuracy By Class			
TP Rate (Test 1)	0.964	0.855	0.980
TP Rate (Test 2)	0.919	0.855	0.919
TP Rate (Test 3)	0.892	0.865	0.885
FP Rate (Test 1)	0.036	0.265	0.026
FP Rate (Test 2)	0.089	0.265	0.098
FP Rate (Test 3)	0.187	0.269	0.204
Precision (Test 1)	0.965	0.850	0.980
Precision (Test 2)	0.923	0.850	0.922
Precision (Test 3)	0.890	0.863	0.883
Recall (Test 1)	0.964	0.855	0.980
Recall (Test 2)	0.919	0.855	0.919
Recall (Test 3)	0.892	0.865	0.885
F-Measure (Test 1)	0.964	0.850	0.980
F-Measure (Test 2)	0.920	0.850	0.920
F-Measure (Test 3)	0.890	0.859	0.883
MCC (Test 1)	0.912	0.624	0.950
MCC (Test 2)	0.807	0.624	0.804
MCC (Test 3)	0.728	0.652	0.710
ROC Area (Test 1)	0.972	0.795	0.991
ROC Area (Test 2)	0.947	0.795	0.942
ROC Area (Test 3)	0.848	0.798	0.816
PRC Area (Test 1)	0.969	0.795	0.990
PRC Area (Test 2)	0.947	0.795	0.933
PRC Area (Test 3)	0.838	0.801	0.850

TP Rate: true positive rate; **FP Rate**: false positive rate; **MCC**: Matthews Correlation Coefficient; **ROC:** Receiver Operating Characteristics; **PRC**: Precision Recall Curve

The accuracy values of the methods used are shown in Figure 1. In the whole training set method, the highest accuracy value belongs to the PART algorithm. In the 5-fold cross-validation method, the highest same value belongs to Jrip and PART algorithms. In the 70% training set method, the highest value belongs to the Jrip algorithm. When the accuracy values of these three methods are examined, the highest value belongs to the PART algorithm with 97.98%, while the lowest value of 85.45% belongs to the OneR algorithm (Figure 1).

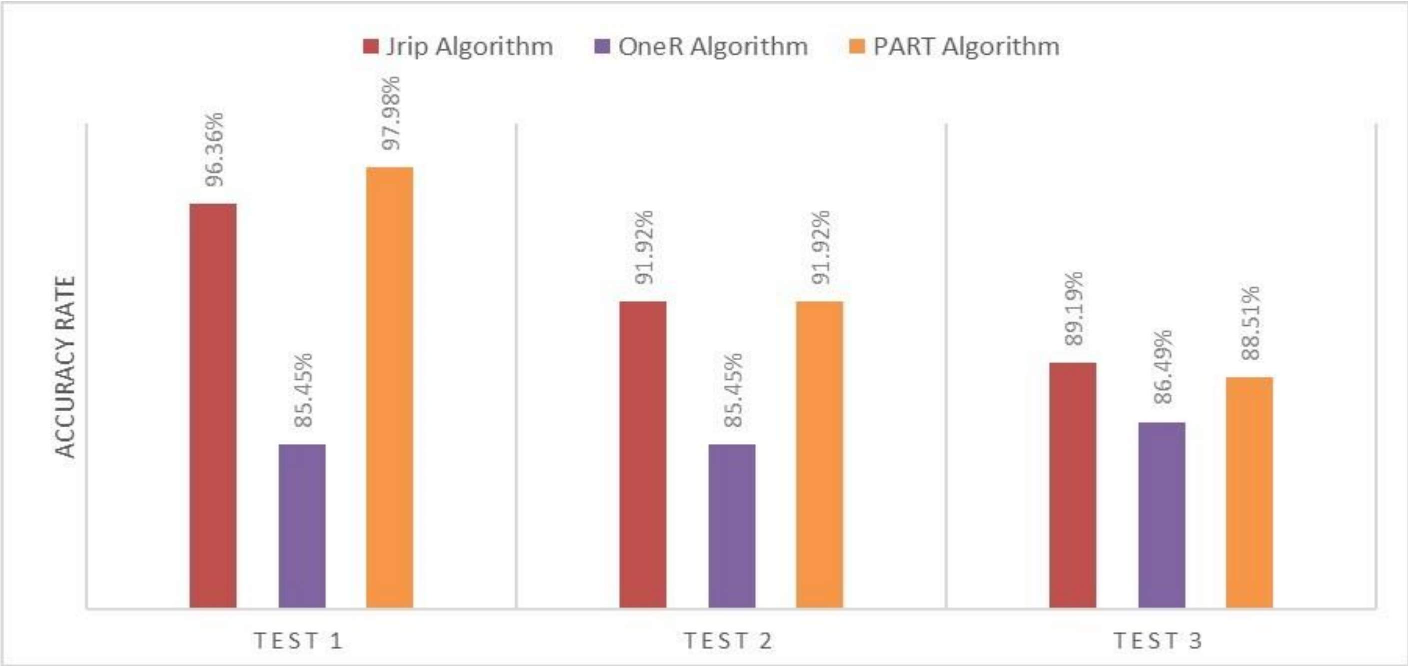

Figure 1. Accuracy rates obtained from different test types for the classification algorithms

The precision values of the methods used are shown in Figure 2. The highest precision value in the whole training set method belongs to the PART algorithm. In the 5-fold cross-validation method, the highest value belongs to the Jrip algorithm. The highest value in the 70% training set method belongs to the Jrip algorithm. When the precision values of these three methods are examined, the highest value of 0.980 belongs to the PART algorithm, while the lowest value of 0.850 belongs to the OneR algorithm (Figure 2).

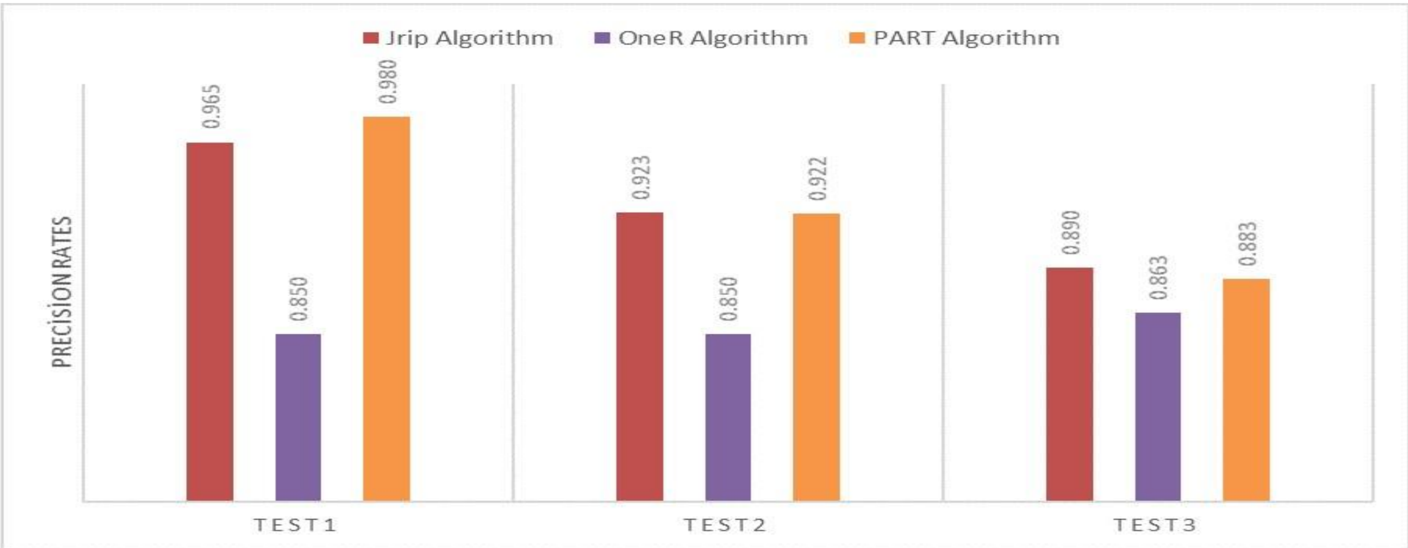

Figure 2. Precision rates of test types by all algorithms

The F-measure values of the methods used are shown in Figure 3. In the whole training set method, the highest F-measure value belongs to the PART algorithm. 5- In the fold cross-validation method, the highest same value belongs to the Jrip algorithm and the PART algorithm. The highest value in the 70% training set method belongs to the Jrip algorithm. When the F-measure values of these three methods are examined, the highest value of 0.980 belongs to the PART algorithm, while the lowest value of 0.850 belongs to the OneR algorithm (Figure 3).

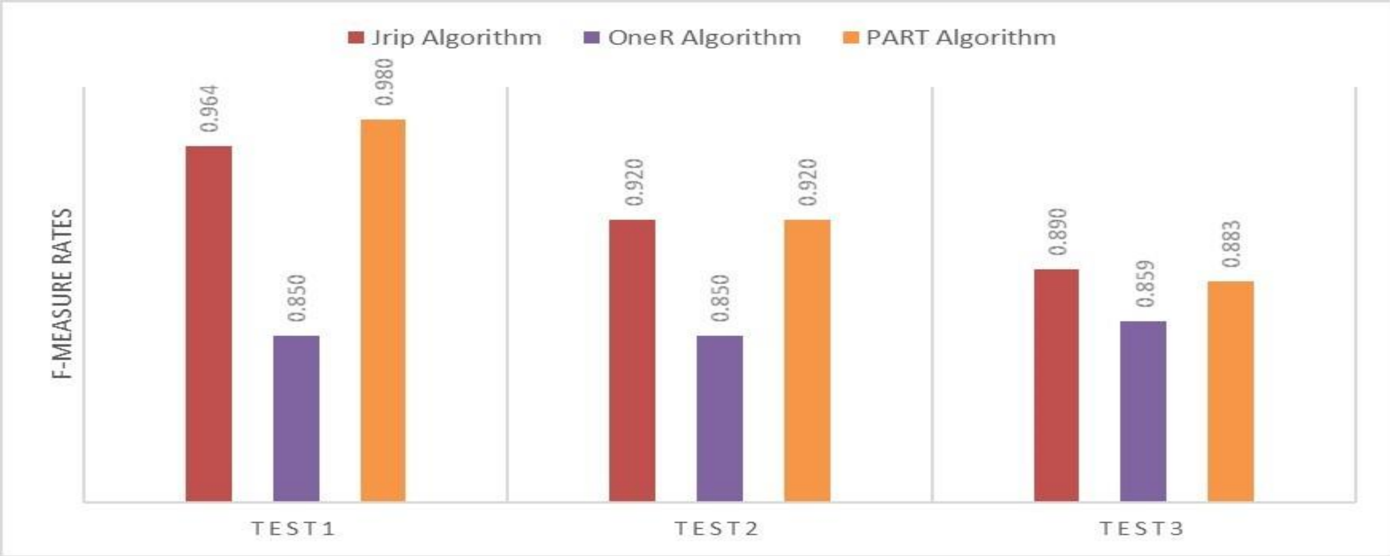

Figure 3. F-Measure rates of test types by all algorithms

ROC area values of the methods used are shown in Figure 4. In the whole training set method, the highest ROC area value belongs to the PART algorithm. 5- In the fold cross-validation method, the highest value belongs to the Jrip algorithm. The highest value in the 70% training set method belongs to the Jrip algorithm. When the ROC area values of these three methods are examined, the highest value of 0.991 belongs to the PART algorithm, while the lowest value of 0.795 belongs to the OneR algorithm (Figure 4).

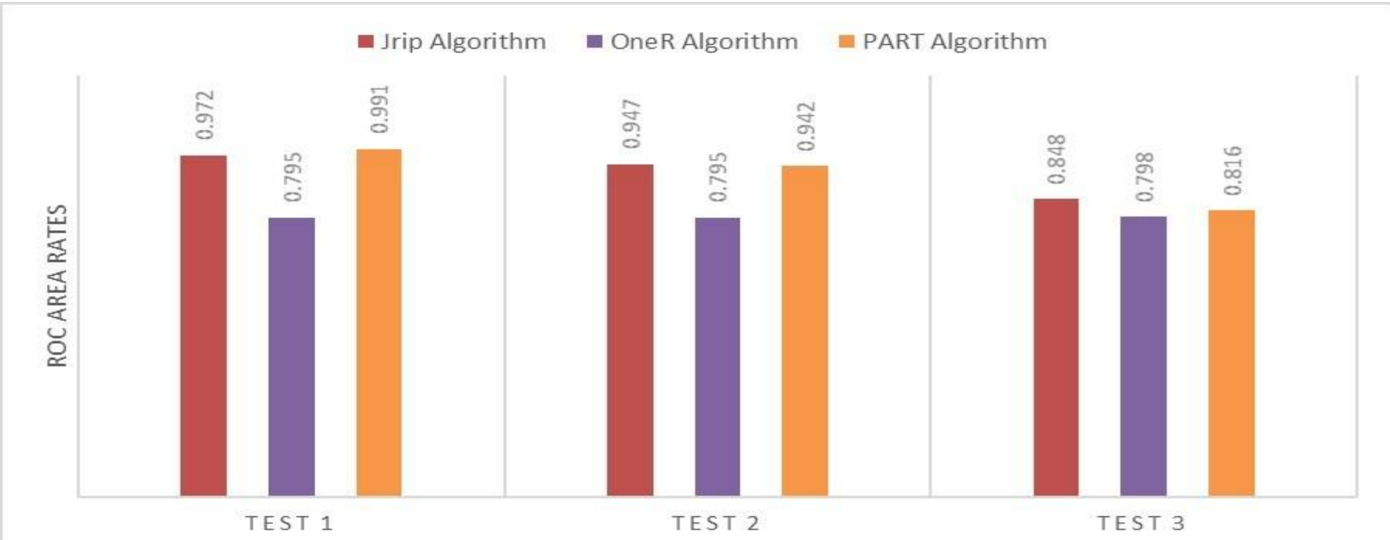

Figure 4. ROC Area rates of test types by all algorithms

Discussion

The current study investigated the relationship between anemia during pregnancy and the sociodemographic, obstetric, and nutritional characteristics of pregnant women. The current study tested whether factors associated with pregnancy anemia can be detected with rule based supervised machine learning methods. The current study showed that sociodemographic factors such as age, gravida, education level, and occupation were associated with anemia during pregnancy. Previous literature data reported that low socioeconomic status is closely related to anemia. The low socioeconomic status may be caused by increased gravida, inadequate health care, and lack of access to quality and sufficient food [14, 15]. It has been reported that these factors cause anemia. The current study findings were consistent with previous literature and sociodemographic factors were related to anemia during pregnancy.

The current study showed that advanced maternal age and gestational age are related to anemia. Anemia was found more frequently in increased maternal age and the third trimester. Gupta et al. reported that pregnant women aged 35-49 had a significantly higher prevalence of anemia compared to pregnant women aged 20-34 [16]. Wu et al. reported that the prevalence of anemia was significantly higher in women aged 18-20 and over 35 years of age compared to other age groups [15]. Literature data reported that another factor associated with anemia is gestational age. It has been shown that women in the second and third trimesters are 3.09 and 3.68 times more likely to be anemic than those in the first trimester [14]. In 2 different studies conducted in Ethiopia and India, it was reported that women in the second and third trimesters have a higher risk of developing anemia compared to those in the first trimester [17, 18]. The current study findings are similar to previous literature data, and advanced age and gestational age were found to be associated with anemia. The increasing number of pregnancies in elderly individuals, the inability to meet the iron and vitamin needs arising from their previous pregnancies, and the increasing iron need in the later stages of pregnancy may be the reasons for the increased prevalence of anemia.

Anemia during pregnancy is usually caused by nutrient deficiencies as a result of not meeting the increased nutritional needs. Inadequate intake of nutrients such as iron, folic acid, B 12, C, and A vitamins, and protein have been associated with anemia during pregnancy [19]. In the current study, it was determined that the consumption of red meat, eggs, green vegetables, and fruits in anemic pregnant women was lower than in those without anemia. Zerfu et al. reported that the risk of anemia is higher in pregnant women with insufficient maternal diet diversity than in those with sufficient maternal diet [20]. The World Health Organization recommends the use of 400 μg folic acid and 30-60 mg elemental iron supplementation per day during pregnancy to reduce the risk of maternal anemia and iron deficiency. In line with these recommendations, studies have shown that the prevalence of anemia can be reduced by the iron and folic acid intake of pregnant women [18-21]. The current study focused on the association of tea consumption and preference with anemia in addition to nutritional diversity. The anemia prevalence is higher in those who consumed dark tea and drank tea with meals. Literature data suggest that polyphenols in tea may cause anemia by reducing dietary iron absorption. Lazrak et al. reported that tea consumption reduces iron absorption by 85%, and tea consumption, especially with meals, caused anemia by impairing iron absorption [22]. The data we obtained were compatible with the literature and showed that tea consumption was associated with anemia.

The main purpose of the current study is to determine the factors associated with anemia during pregnancy and to develop a new approach by combining their role in clinical diagnosis with supervised machine learning methods. In this context, we investigated the relationship between 3 different algorithms and factors related to anemia in pregnancy and anemia status. The first algorithm, the Jrip algorithm, created 11 rules. In the first rule of the Jrip algorithm, dark tea consumption was associated with multigravida and anemia in pregnant women. The rule correctly predicted 52 cases out of 56 cases that met this rule. In the second rule, he correctly guessed all 10 cases who consumed dark tea and usually consumed the tea within 1 hour after the meal. In a multicenter study, pregnancy anemia was associated with maternal age 35 and over and multigravida [23]. It has also been shown that tea consumption causes anemia by negatively affecting iron absorption [22, 23]. The first rule of the algorithm, the consumption of dark tea, the number of pregnancies being 3 and above, and the relationship between advanced maternal age and anemia is consistent with the literature data. In addition, other rules of the Jrip algorithm correlated nutritional education, red meat consumption, fruit and vegetable consumption, iron supplementation therapy, and maternal education level with anemia. In our article, a statistically significant relationship was found between some sociodemographic, obstetric, and nutritional characteristics of pregnant women and anemia. The Jrip algorithm quickly and consistently predicted the factors associated with anemia.

OneR algorithm, which is another supervised machine learning method correlated dark tea preference and anemia. In this study, while 65.9% of anemic pregnant women preferred dark tea, this rate was 7.0% in non-anemic women. In addition, 33.3% of anemic pregnant consumed tea within 1 hour after meals, and this rate was 10.6% in non-anemic pregnant women. Both the consumption of dark tea and the consumption of tea within 1 hour after meals of anemic pregnant women were statistically higher than those of non-anemic women. Black tea consumption negatively affects the absorption of iron taken with food, so it is not recommended to consume tea with food [22, 24]. The consumption of black tea in our country is quite high, and in the study population, the tea consumption of anemic pregnant women was found to be much higher than that of non-anemic women. The OneR model correctly classified 423 samples from a total of 495 study populations. Considering the frequency and type of black tea consumption in our country and the findings of our study, the OneR model detected a significant relationship between black tea consumption and tea preference.

The PART algorithm which is another supervised machine learning method determined 19 rules. The PART algorithm determined that anemia prevalence is lower in iron supplementation, gravida <3, maternal age under 34 years, and when tea consumption is low. In previous studies, it has been shown that an increased number of pregnancies, advanced maternal age, and tea consumption are risk factors for anemia during pregnancy [15, 18-24]. It is known that iron supplementation treatment given during pregnancy is the most important protective factor in preventing anemia [25]. The findings obtained as a result of the Part algorithm are compatible with the literature, and the Part algorithm quickly and consistently determined the protective factors from anemia. The PART algorithm found high consistency and accuracy in all learning rules that included factors such as income level, occupation, education level, number of meals, meat, legumes, green vegetables and fruit consumption, number of pregnancies, and maternal age. The rules determined in the part algorithm are compatible with the statistical significance obtained in the study data.

The current study aimed to create a clinical decision support system that will determine the frequency of anemia during pregnancy and related factors and the factors associated with pregnancy anemia using supervised machine learning in the Turkish population. The study findings were consistent with the previous literature data. Algorithms created by supervised machine learning, Jrip 96.4%, OneR 85.5%, and Part algorithm 97.9% quickly and consistently detected anemia and related factors. Sow et al. reported that they detected malaria and anemia estimation with supervised machine learning techniques with an accuracy of 94.74% and 84.17%, respectively [26]. Dauvin et al., on the other hand, reported that they predicted anemia with 89% accuracy in patients admitted to the intensive care unit using the machine learning model [27]. They emphasized that the findings obtained in both studies are consistent and important in the prediction of disease-related factors. Li et al., in their study to evaluate the important risk factors for mortality after hip fracture of the machine learning algorithm, reported that the machine learning model can create a risk classification model for patients who have undergone hip fracture surgery to identify those with a high risk of long-term mortality [28]. In our current study, supervised machine learning algorithms detected factors associated with anemia with high accuracy. With the machine learning models to be created for anemia, which is very common in daily practice, the risk of anemia can be predicted quickly and with high accuracy. For this reason, we recommend further studies with larger populations.

Limitations

This study is a retrospective analysis of a single-center cross-sectional study. The study cohort may have been insufficient for machine learning model performance. Relatively large clinical datasets and multicenter validation are required to improve the performance of the model and make precise comparisons between machine learning-based models. Prospective studies are required to confirm and improve algorithm performance in future studies.

Conclusion

The present study showed that those who drink tea with meals consume dark tea, multigravida, advanced maternal age, advanced gestational age, insufficient iron support, and nutritional diversity cause an increase in the frequency of anemia. Three different algorithms created in the study predicted anemia with 85.5% to 97.9% accuracy. The decision support systems used in our study can be used to predict anemia among pregnant women. It is thought that more research is needed to transform the findings we have obtained into a clinical decision support system and to apply them in daily practice.

Conflict of interest: The authors have no conflicts of interest to declare for this study

Author Contributions		Author Initials
SCD	Study Conception and Design	MOK, RY, BY
AD	Acquisition of Data	BY, RK
AID	Analysis and Interpretation of Data	MOK, BA
DM	Drafting of Manuscript	MOK, BY, RK, BA
CR	Critical Revision	MOK, BY, RK, BA

Financial support: No financial support was received for this study.

Acknowledgments: We thank all authors for allowing us to use the data of the study "Prevalence of Anemia and Associated Risk Factors among Pregnant Women, What is the Role of Antenatal Care in Prevention? A Cross-sectional Study".

Prior publication: This study presented as an oral presentation at XXII. NATIONAL and V. INTERNATIONAL BIOSTATISTICS CONGRESS in 2021.

References

1. James AH. Iron deficiency anemia in pregnancy. Obstet Gynecol. 2021;138(4):663-74. https://doi.org/10.1097/aog.0000000000004559
2. WHO Global anaemia estimates, 2021 Edition. Access date: 27.07.2022 available at: https://www.who.int/data/gho/data/themes/topics/anaemia_in_women_and_children#:~:text=Summary%20findings&text=In%202019%2C%20global%20anaemia%20prevalence,women%20aged%2015%2D49%20years. (Access Date: December 22, 2023)
3. Ullah A, Sohaib M, Saeed F, Iqbal S. Prevalence of anemia and associated risk factors among pregnant women in Lahore, Pakistan. Women Health 2019;59(6):660-71. https://doi.org/10.1080/03630242.2018.1544966
4. Yakar B, Pirincci E, Kaya MO, Onalan E. Prevalence of anemia and associated risk factors among pregnant women, what is the role of antenatal care in prevention? A cross-sectional study. J Coll Physicians Surg Pak. 2021;31(11):1341-45. https://doi.org/10.29271/jcpsp.2021.11.1341
5. Mahamoud NK, Mwambi B, Oyet C, Segujja F, Webbo F, Okiria JC, et al. Prevalence of anemia and its associated socio-demographic factors among pregnant women attending an antenatal care clinic at Kisugu health center IV, Makindye division, Kampala, Uganda. J Blood Med 2020;11:13-8. https://doi.org/10.2147/jbm.s231262
6. World Health Organization. Meeting to develop a global consensus on preconception care to reduce maternal and childhood mortality and morbidity; 6-7 February 2012 Meeting Report; World Health Organization: Geneva, Switzerland, 2013. Available at: https://apps.who.int/iris/handle/10665/78067 (Access Date: December 22, 2023)
7. Deo RC. Machine Learning in Medicine. Circulation. 2015;132(20):1920-30. https://doi.org/10.1161/circulationaha.115.001593
8. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. J Intern Med. 2018;284(6):603-19. https://doi.org/10.1111/joim.12822
9. Cohen WW. Fast effective rule induction, twelfth international conference on machine learning. Morgan Kaufmann 1995:115-23.
10. Akyol S, Alatas B. Automatic mining of accurate and comprehensible numerical classification rules with cat swarm optimization algorithm. J Fac Eng Arch Gazi Uni 2016;31:839-57.
11. Holte, RC. Very simple classification rules perform well on most commonly used datasets. Machine Learning 1993;1(1):63-90.
12. Frank E, Witten IH. Generating accurate rule sets without global optimization. Machine Learning: Proceedings of the fifteenth international conference, Madison, Wisconsin, 1998:144-51.
13. Balaban M, Kartal E. Data mining and machine learning basic algorithms and applications with R programming. 2sd Ed. İstanbul: Caglayan Publishing House, 2018.
14. Lebso M, Anato A, Loha E. Prevalence of anemia and associated factors among pregnant women in Southern Ethiopia: A community based cross-sectional study. PLoS One. 2017;12(12):e0188783. https://doi.org/10.1371/journal.pone.0188783
15. Wu Y, Ye H, Liu J, Ma Q, Yuan Y, Pang Q, et al. Prevalence of anemia and sociodemographic characteristics among pregnant and non-pregnant women in southwest China: a longitudinal observational study. BMC Pregnancy Childbirth. 2020;20(1):535. https://doi.org/10.1186/s12884-020-03222-1
16. Gupta PM, Hamner HC, Suchdev PS, Flores-Ayala R, Mei Z. Iron status of toddlers, nonpregnant females, and pregnant females in the United States. Am J Clin Nutr. 2017;106(Suppl 6):1640-46. https://doi.org/10.3945/ajcn.117.155978
17. Addis Alene K, Mohamed Dohe A. Prevalence of anemia and associated factors among pregnant women in an urban area of eastern Ethiopia. Anemia. 2014;2014:561567. https://doi.org/10.1155/2014/561567
18. Viveki RG, Halappanavar AB, Vivek PR, Halki SB, Maled VS, Deshpande PS. Prevalence of anemia and its epidemiological determinants in pregnant women. Al Ameen J Med Sci. 2012;5(3):216–23.
19. Kangalgil M, Sahinler A, Kirkbir IB, Ozcelik AO. Associations of maternal characteristics and dietary factors with anemia and iron-deficiency in pregnancy. J Gynecol Obstet Hum Reprod. 2021;50(8):102137. https://doi.org/10.1016/j.jogoh.2021.102137

20. Zerfu TA, Umeta M, Baye K. Dietary diversity during pregnancy is associated with reduced risk of maternal anemia, preterm delivery, and low birth weight in a prospective cohort study in rural Ethiopia. Am J Clin Nutr. 2016;103(6):1482-88. https://doi.org/10.3945/ajcn.115.116798
21. Chakrabarti S, George N, Majumder M, Raykar N, Scott S. Identifying sociodemographic, programmatic and dietary drivers of anaemia reduction in pregnant Indian women over 10 years. Public Health Nutr. 2018;21(13):2424-33. https://doi.org/10.1017/s1368980018000903
22. Lazrak M, El Kari K, Stoffel NU, Elammari L, Al-Jawaldeh A, Loechl CU, et al. Tea consumption reduces iron bioavailability from NaFeEDTA in nonanemic women and women with iron deficiency anemia: Stable iron isotope studies in Morocco. J Nutr. 2021; 151(9):2714-20. https://doi.org/10.1093/jn/nxab159
23. Machmud P, Hatma R, Syafiq A. Tea consumption and iron-deficiency anemia among pregnant woman in Bogor district, Indonesia. Media Gizi Mikro Indonesia 2019;10(2):91-100. https://doi.org/10.22435/mgmi.v10i2.1384
24. Zijp IM, Korver O, Tijburg LB. Effect of tea and other dietary factors on iron absorption. Crit Rev Food Sci Nutr. 2000;40(5):371-98. https://doi.org/10.1080/10408690091189194
25. Ugwu NI, Uneke CJ. Iron deficiency anemia in pregnancy in Nigeria-A systematic review. Niger J Clin Pract. 2020;23(7):889-96. https://doi.org/10.4103/njcp.njcp_197_19
26. Sow B, Mukhtar H, Ahmad HF, Suguri H. Assessing the relative importance of social determinants of health in malaria and anemia classification based on machine learning techniques. Inform Health Soc Care. 2020;45(3):229-41. https://doi.org/10.1080/17538157.2019.1582056
27. Dauvin A, Donado C, Bachtiger P, Huang KC, Sauer CM, Ramazzotti D, et al. Machine learning can accurately predict pre-admission baseline hemoglobin and creatinine in intensive care patients. NPJ Digit Med. 2019;2:116. https://doi.org/10.1038/s41746-019-0192-z
28. Li Y, Chen M, Lv H, Yin P, Zhang L, Tang P. A novel machine-learning algorithm for predicting mortality risk after hip fracture surgery. Injury. 2021;52(6):1487-93. https://doi.org/10.1016/j.injury.2020.12.008