MPPKI (November, 2024) Vol. 7 No. 11

ISSN 2597– 6052

DOI: https://doi.org/10.56338/mppki.v7i11.6289

MPPKI

Media Publikasi Promosi Kesehatan Indonesia

The Indonesian Journal of Health Promotion

Review Articles *Open Access*

Prediction of Anemia Using Machine Learning Algorithms: Scoping Review

Asrit Jessica Kario[1], Rico Kurniawan[2]*

[1]Departemen Biostatistik dan Ilmu Kependudukan, Fakultas Kesehatan Masyarakat, Universitas Indonesia | email asrit.jessica@ui.ac.id

[2]Departemen Biostatistik dan Ilmu Kependudukan, Fakultas Kesehatan Masyarakat, Universitas Indonesia | email ricokurniawan@ui.ac.id

* Corresponding Author: ricokurniawan@ui.ac.id

ABSTRACT

Introduction: One of the major public health problems is anemia, especially affecting newborn and infant children, adolescent girls, young women, pregnant women, and postpartum women. The cause of anemia is the reduced supply of red blood cells in the human body or the damage or weakening of the structure of red blood cells. One of the preferences of utilizing machine learning is the prediction of results.

Objective: The purpose of this study is to compare effective algorithms, related to the origin or source of the data set, data set size, metric evaluation and accuracy and produce predictors in predicting anemia using machine learning.

Method: This research uses a scoping review method on 4 databases, namely Scopus, EBSCO, PubMed, and IEEE Xplore from 2019 - 2024 with keywords anemia, algorithms, machine learning, and prediction. The results of screening articles on the Scopus, EBSCO, PubMed, and IEEE Xplore databases obtained 384 articles which were then selected through several stages and obtained 9 articles.

Result: The review found that the highest algorithm performance in anemia prediction, namely Penalized Regression (LASSO regression) accuracy above 64%, XGboost accuracy 100% and execution time 0.2404 seconds, Catboost accuracy 97.6%, Random Forest accuracy 95.49% and 72%, J48 algorithm accuracy of 97.7%, Logistic Regression accuracy 66% and AUC 69%, and SVM linear AUC 79.9%.

Conclusion: Machine learning can assist in the development of anemia prediction models by exploring large amounts of data and producing precise and fast predictors. The predictors obtained are determined by the selection of algorithms in the study.

Keywords: Anemia; Machine Learning; Algorithms; Prediction; Predictors

INTRODUCTION

One of the major public health problems is anemia, especially affecting newborn and infant children, adolescent girls, young women, pregnant women, and postpartum women (1). Globally, in 2021 the prevalence of anemia for all age groups is 24.3% (2). Meanwhile, the prevalence of anemia for all age groups in Indonesia in 2023 was 16.2% (3). The impact of anemia can affect school performance related to decreased motor activity, social interaction, decreased work performance, decreased work productivity that impacts the finances of the individual, family, community and country, inhibits growth, delays puberty, impairs overall physical development, and affects the general quality of life. In addition, anemia affects adolescents mental health, increases the likelihood of having complications during childbirth, including premature birth, trouble concentrating, diminished physical execution, and increments the probability of having an underweight child (1),(4).

The cause of anemia is the reduced supply of red blood cells in the human body or the damage or weakening of the structure of red blood cells (5). As a result, the capacity of oxygen circulated throughout the body decreases and the body will experience oxygen deficiency (6). Common and non-specific side effects of anemia include fatigue, dizziness, cold hands and feet, headache, and shortness of breath amid activities. In the meantime, indications of serious anemia incorporate pale mucous films (mouth, nose, etc.), quick breathing and pulse, pale skin and nails, dizziness when standing and bruising more easily (1).

In previous studies, one of the benefits of using machine learning is the prediction of outcomes (7). Prediction of anemia status in the community can use machine learning algorithms and secondary data or survey data to produce prediction models and predictors. Some algorithms that can be used: XGBoost (XGB), Support Vector Machine (SVM), k-nearest neighbors (k-NN), Decision Tree (DT), Gradient Boosting (GB), Classification and Regression Trees (CART), Linear Discriminant Analysis (LDA), and Logistic Regression (LR) (8), (9). There are many benefits to using machine learning algorithms for prediction, such as design acknowledgment, flexibility, objectivity, non-linearity taking care of, feature determination, and generalization (10). In addition, machine learning can assist in the development of anemic prediction models by exploring large amounts of data to discover trends, patterns and relationships not obtained through statistical methods (10), (11). Disease prediction analysis in healthcare can significantly reduce the burden on healthcare workers in terms of diagnosis and treatment of patients, thus bringing about great changes in the healthcare system (12).

The purpose of this study is to compare effective algorithms, related to data set size, metric evaluation and generating predictors in predicting anemia using machine learning. The findings are expected to expand knowledge and can provide evidence in planning health programs related to prevention and intervention in vulnerable groups of anemia in the community.

METHOD

This research uses the scoping review method, which is to conduct a literature review by collecting, selecting, analyzing, summarizing research results from several previous articles or literature. The databases used were health/medical databases: MEDLINE (EBSCO and PubMed), technology databases (IEEE Xplore) and multidisciplinary databases (Scopus). The eligibility criteria use the PICO framework, consisting of Population, Intervention, Comparative, and Outcome. The population in this study, namely anemia, the intervention is machine learning algorithms, there is no comparative, and the results are prediction models and predictors. Articles taken from 2019 - 2024 in English. The keywords used are anemia, algorithms, machine learning, and prediction. Articles in the form of SLR, LR were not taken in this study.

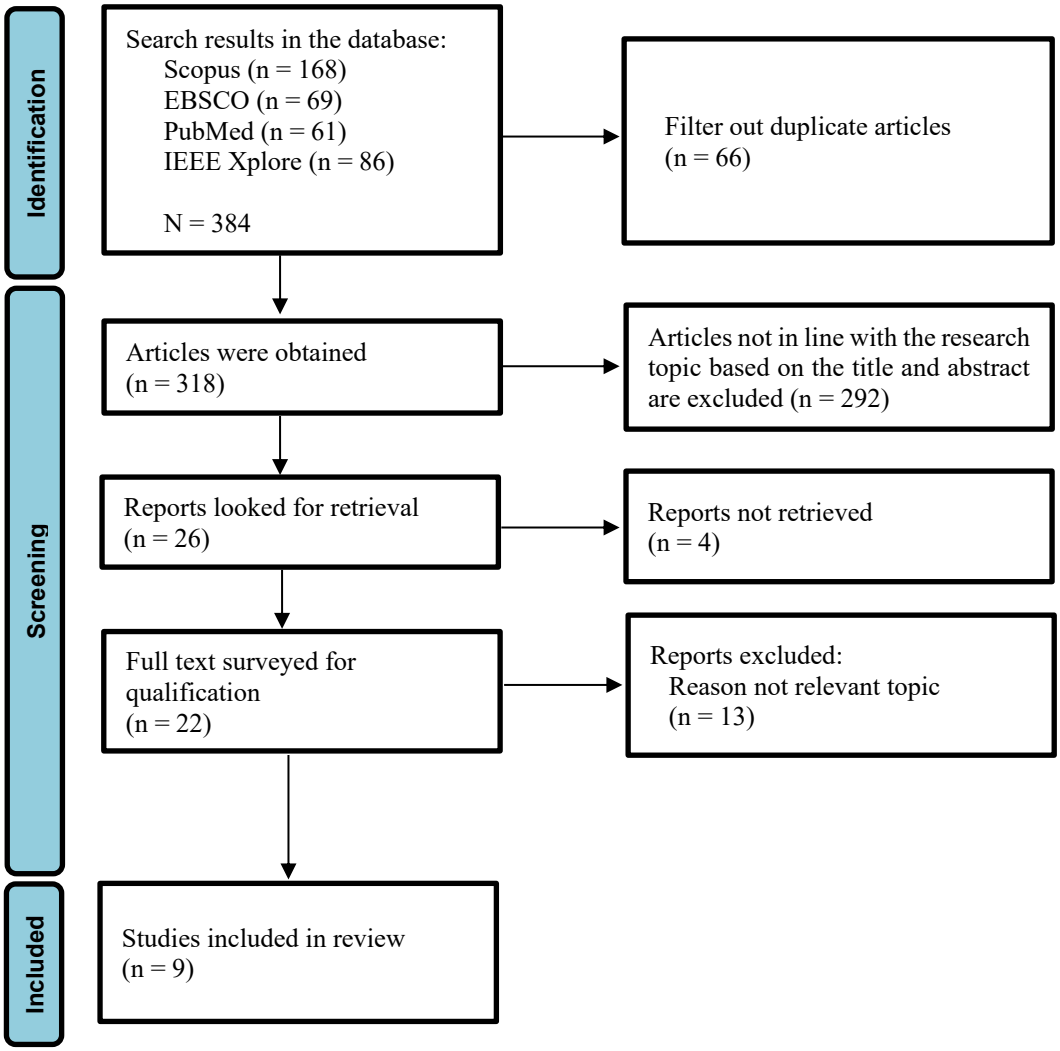

Figure 1. PRISMA Flow Diagram

Searching the Scopus, EBSCO, PubMed, and IEEE Xplore databases using advanced search techniques. In the advanced search, keywords and boolean were entered, namely anemia OR anaemia AND algorithms AND machine learning OR deep learning AND prediction OR predictors OR predicting, 384 articles were obtained. After searching for articles in 4 databases, all selected articles were checked if there were the same articles and filtered according to the criteria using the Mendeley application, 318 articles were obtained. The articles that had been filtered were then examined based on the title and abstract, those that did not match the research topic were not taken, and 26 articles were obtained. Then the articles were checked for fulltext, but 4 articles were found to be inaccessible. Next, 22 articles were examined. Only articles published in the past 5 years related to machine learning algorithm anemia prediction using secondary data or survey data were retrieved. As a result, 9 articles were found to be included in the review.

RESULTS

In the literature search in the Scopus, EBSCO, PubMed, and IEEE Xplore databases, 384 articles were obtained which were then selected through several stages and obtained 9 articles. The following table provides a summary of the articles that have been selected based on the established selection criteria.

Table 1. Article Data Extraction

Author, Year	Title	Subject	Size of Dataset	Predictors	Algorithm	Performance (Evaluation Metrics)	Findings
Meitei A Jiran et al., 2022 (11)	Predicting child anaemia in the North-Eastern states of India: a machine learning approach	Children (6 – 59 months)	21,000 children	Mother's education, social status, mother's anaemic status, religion, age of the child, and mother's age.	Penalized Regression (Ridge, LASSO and elastic net)	Cohen's Kappa value, accuracy, precision (positive predictive value), specificity, negative predictive value, F1 score and sensitivity.	Penalized Regression (LASSO regression) has a higher performance, accuracy above 64% which shows the suitability for anemia prediction in children aged 6 - 59 months.
Shweta & Pande, 2023 (13)	Prediction of anemia using various ensemble learning and boosting techniques	-	16,721 instances and 6 columns	Hemoglobin, MCHC, MCH, gender, and MCV.	Artificial Neural Network (ANN), Logistic Regression (LR), Support Vector Machine (SVM), hybridizing two algorithms and ensemble machine learning algorithms (AdaBoost classifier and Xgb classifier), and Random Forest (RF).	Accuracy, confusion matrix.	XGboost works best with an ideal execution time, XGboost with 100% accuracy and an execution time of 0.2404 seconds, indicating suitability for anemia prediction in certain demographics.
Dejene Belayneh Endalamaw et al., 2022 (14)	Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm	Pregnant women	11,174 instances with 34 features	Respondent's occupation, source of drinking water, duration of current pregnancy, history of contraceptive use, several household members and age in five years group.	Catboost, Random Forest (RF), and extreme gradient boosting with class decomposition (one versus one and one versus rest) and without class decomposition, and	Accuracy, precision, recall, F1 score, cross validation, ROC.	Catboost with course decomposition has way better execution with 97.6 % accuracy, which suggests the suitability for anemia prediction in pregnant women.

					Decision Tree (DT).		
Maasthi Manasvi Jagadeesh et al., 2023 (15)	Decision – making support system for predicting and eliminating malnutrition and anemia	Women and men	1,799 anemia datasets	Age, gender, red blood cells (RBC), packed cell volume (PCV), mean corpuscular volume (MCH), mean corpuscular hemoglobin concentration (MCHC), red cell distribution width (RDW), total leukocyte count (TLC), plateleds (PLT), and hemoglobin (HGB).	K-Nearest Neighbor (k-NN), naïve Bayes classifier (NBC), Random Forest (RF), and Decision Tree (DT).	Accuracy, sensitivity, specificity.	Random Forest provides efficient results for predicting anemia with accuracy of 95.49%, which shows the suitability for predicting anemia in men and women.
Zemariam Alemu Birara et al., 2024 (4)	Employing supervised machine learning algorithms for classification and prediction of anemia among youth girls in Ethiopia	Youth girls	5,642 weighted samples of young girls	Not used contraceptive method, region, religion, poor wealth index, having more than 5 family size, unimproved toilet facility, no media exposure, no formal education, occupation, age, and rural residence.	SVM, LR, GNB, XGB, LGB, RF, DT, KNN.	Receiver operating characteristic (ROC) curve with metrics such as accuracy, sensitivity (recall), specificity, F1 score, and area under the curve (AUC), and confusion matrix.	Random Forest classifier is the best for predicting anemia with an AUC value of 82% and an accuracy of 72%, indicating that suitability for anemia prediction in youth girls.
Göl Mehmet et., 2024 (16)	Predicting malnutrition - based anemia in geriatric patients using machine learning methods	Geriatric patients	438 patients	Weight, alcohol use, body mass index, muscle mass, and parameters that reveal the loss of strength and movement due to aging, eating attitude test, swallowing function test, timed up and go test, instrumental living activity score,	J48, DTNB, RF, BayesNet, Naïve Bayes, Decision Table, AVG.	Precision, recall, F-measure, ROC area, PRC area, and accuracy.	Anemia can be predicted with high accuracy in geriatric patients without hemogram data using the Random Forest algorithm with an accuracy of 85.39%. When hemogram data is involved, the J48 algorithm

				walking speed and hand strength, height, smoking use, and calf circumferenc e.			is most successful with an accuracy of 97.77%, indicates suitability for anemia prediction in geriatric patients.
Tesfaye Solomon Hailemaria m et al., 2024 (17)	Application of machine learning methods for predicting childhood anaemia: analysis of Ethiopian Demographic Health Survey of 2016	Childhood (6 – 59 months)	8,482 children	Stunting, uneducated mother, child age category between six and 23 months, unemployed mother, and poorest family.	Logistic regression (LR), Recursive Feature Elimination (RFE) and logistic regression, random forest (RF), Recursive Feature Elimination (RFE) and random forest classifier, K-nearest neighbours (KNN), and decision tree (DT), Linear support vector classifier (SVC), Lasso, chi-2.	AUC, accuracy, positive predictive value, sensitivity, negative predictive value, performance time, and specificity.	The logistic regression model is the best predictive model with an accuracy value of 66% and AUC of 69%, indicating its suitability for predicting anemia in children aged 6 - 59 months.
Kassaw Abdulaziz Kebede et al., 2023 (10)	The application of machine learning approaches to determine the predictors of anemia among under five children in Ethiopia	Under five children	9,501 children	Mothers' wealth index, number of children, mother's educational status, health insurance coverage, receiving rotavirus vaccine, youngest child's stool disposal, type of cooking fuel, residence, number of family members, and, distance to health facilities.	Boruta Algorism	-	Machine learning algorithm was applied to determine the predictors of anemia among under five year's age children in Ethiopia.

Pan Yunhui et al., 2023 (18)	Machine learning prediction of iron deficiency anemia in Chinese premenopausal women 12 months after Sleeve Gastrectomy	Chinese premenopausal women 12 months after sleeve gastrectomy	407 patients	Preoperative ferritin, age, hemoglobin, Cr, and FCP.	Linear SVM	ROC, AUC, calibration curve.	The model appeared direct separation in both sets (area under curve 0.858 and 0.799, separately, $p < 0.001$). The calibration curve demonstrated worthy consistency between watched and anticipated comes about in both sets. This suggests that accurate prediction of anemia in Chinese premenopausal women 12 months after sleeve gastrectomy.

Machine learning algorithms have been used and are effective and efficient for predicting anemia in anemia-prone groups in communities within Ethiopia, China, India, Turkey. The determination of the algorithm used and effective in predicting anemia varies in each study depending on data characteristics, scalability, interpretability, accuracy, number of features, computational efficiency, expertise, problem type, noise/dependent tolerance for outliers, and biased variance tradeoffs (4).

DISCUSSION

One of the major public health problems is anemia, especially affecting newborn and infant children, adolescent girls, young women, pregnant women, and postpartum women (1). Globally, in 2021 the prevalence of anemia for all age groups is 24.3% (2). Indicators of malnutrition and poor health, namely anemia. Poor cognitive and motor development in children, premature birth, low birth weight babies, and decreased iron stores for infants lead to developmental disorders. In addition, anemia can affect adult work capacity and quality of life which can affect a country's economic development (19). Anemia is caused by a reduced amount of hemoglobin, a protein found in red blood cells, which carries oxygen throughout the body (13). Factors that can contribute to anemia include biological, socioeconomic, environmental, health and nutritional factors (20). In some cases, in addition to being caused by malnutrition and blood loss, anemia can be the result of chronic or congenital conditions such as genetic disorders, autoimmune and other diseases. While many cases of anemia are mild and easily treatable, there are some cases of anemia that are severe, chronic, and life-threatening (21). Therefore, risk prediction, early discovery, diagnosis and fitting treatment of anemia are required to decrease the impact on public health.

Improvements in technology and science have made machine learning one of the most exciting technologies in healthcare. Machine learning innovations and breakthroughs using large data sets such as demographic data, healthcare reports (clinical laboratory reports, physical examinations, surgery records) in creating disease prediction models can save diagnostic time, money, and find methods that can reduce the impact of incurable diseases by predicting their root causes (12), (22), (23). Machine learning can assist in the development of anemic prediction models by exploring large amounts of data to find patterns and relationships that are not obtained through statistical methods (10). Research from (8), (24) and (25) claim that machine learning can better predict anemia than statistical methods.

Feature selection or algorithm type is the most important part of building a prediction model because including irrelevant features can complicate the process of learning from samples. In addition, feature selection can enable

better visualization and interpretation of the data (26). Evaluation metrics are used in this study to assess the performance of anemia prediction models. The choice of the leading assessment metric ought to be driven by context-specific prerequisites, trade-offs between different assessment measurements, benchmarks and measures within the same field, demonstrate interpretability, issue sort, information characteristics, and the targets of the errand at hand (4). One of the evaluation metrics is the accuracy value. The accuracy value is the basis for evaluating the performance of a model (27). Accuracy is a measure that shows the level of similarity between the measurement results and the actual value (11). The most widely used accuracy values in classification are expressed in percentages (26). In addition, the AUC value can also be an evaluation metric. The assessment of model performance can be seen in the AUC (Area Under the Curve) because of its assessment on different classification parameters (28).

The penalized regression algorithm is an improvement or refinement of the Logistic Regression (LR) algorithm that allows handling complex regression problems by providing greater accuracy (29). One of the prominent penalized regression algorithms is LASSO regression (11). The LASSO (Least Absolute Shrinkage and Selection Operator) regression serves as a variable selection method. The LASSO regression can eliminate the slightest vital factors by driving their coefficients to zero so that it can not only improve classification accuracy, but can make interpretation of the prediction model easier (30). In a study in East India, LASSO accuracy was found to be above 64%, which means that it is acceptable and efficient in analyzing and drawing conclusions from demographic data. Important factors that can predict anemia in children aged 6 - 59 months are maternal anemia status, child age, social status, mother's age, mother's education, and religion (11).

The XGboost (Extreme Gradient Boosting) algorithm belongs to ensemble learning, which is an algorithm that combines predictions from multiple models to improve results. XGboost is based on a boosting technique which combines the results of weak learners sequentially (31). In research (13), XGboost can be implemented because it has the best performance with 100% accuracy and execution time of 0.2404 seconds. In this prediction model, predictors such as gender, hemoglobin, MCV, MCH, MCHC were found. Prediction models using the XGboost algorithm can be precise and fast in generating predictors.

Catboost algorithm with class decomposition belongs to ensemble learning that works with categorical data and uses gradient improvement. Research on pregnant women in Ethiopia found Catboost performance better in predicting the level of anemia in pregnant women with 97.6% accuracy with important factors that can predict anemia in pregnant women in the form of respondent's occupation, source of drinking water, duration of current pregnancy, history of contraceptive use, several household members and age in five years group (14).

Random Forest algorithm is a collection of decision trees with the purpose of classification, regression, and dimension reduction that can be used with large amounts of data. Research on anemia prediction in men and women in India obtained Random Forest accuracy of 95.49% with parameters in anemia prediction, namely age, sex, red blood cells (RBCs), TLC, packed cell volume (PCV), hemoglobin, MCH, RDW, PLT, and MCHC (15). While a study on adolescent girls in Ethiopia, the accuracy of Random Forest was 72% in predicting anemia with the main predictors, namely not used contraceptive method, region, religion, poor wealth index, having more than 5 family size, unimproved toilet facility, no media exposure, no formal education, occupation, age, and rural residence (4).

The J48 algorithm is an implementation of the C4.5 algorithm that generates a Decision Tree (DT) with the aim of separating objects. Tree or decision tree is widely known as a part of Graph and Tree itself is a connected undirected graph, and does not contain circuits (32). Research on Geriatric patients in Turkey found that the accuracy of J48 algorithm was 97.7% in predicting anemia. Predictors of anemia in Geriatric patients in Turkey, namely weight, alcohol use, body mass index, muscle mass, and parameters that reveal the loss of strength and movement due to aging, eating attitude test, swallowing function test, timed up and go test, instrumental living activity score, walking speed and hand strength, height, smoking use, and calf circumference (16).

The Logistic Regression (LR) algorithm is a simple method of analyzing binary data where the output is formed as a probability of belonging to a particular class (13). Research on children aged 6 - 59 months in Ethiopia used the Logistic Regression (LR) algorithm with an accuracy value of 66%, this value is better than other algorithms. Predictors of anemia in children aged 6 - 59 months, namely stunting, uneducated mother, child age category between six and 23 months, unemployed mother, and poorest family (17). While (10) also conducted a study on children under 5 years old in Ethiopia found different predictors of anemia, namely mothers' wealth index, number of children, mother's educational status, health insurance coverage, receiving rotavirus vaccine, youngest child's stool disposal, type of cooking fuel, residence, number of family members, and, distance to health facilities. This difference is due to the different prediction models used. Research (10) used the Boruta algorithm. The Boruta algorithm is a technique that contains a randomized forest classifier algorithm and is highly selective in choosing the variable (27).

The linear SVM (Support Vector Machine) algorithm is an algorithm for classifying data with small sample sizes and good accuracy. Research on premenopausal women 12 months after sleeve gastrectomy in China using linear SVM algorithm, AUC value 79.9% with anemia predictors age, preoperative ferritin, FCP (Fasting C-Peptide) hemoglobin, and Creatinine (18).

Limitations to this study in terms of English language articles and the type of data used in the articles that have been used, such as secondary data or demographic data or survey data or health records in predicting anemia. Articles using medical images were not included in this study.

CONCLUSION

Machine learning can assist the development of anemia prediction models by exploring large amounts of data and generating precise and fast predictors. The predictors obtained are determined by the selection of algorithms in the study. Machine learning can also be implemented into anemia screening programs. Risk prediction, early discovery, diagnosis, and appropriate treatment of anemia are needed to reduce its impact on public health.

SUGGESTION

Suggestions for future researchers on the adaptability of anemia prediction models for other demographics or countries, as well as comparative studies between machine learning and traditional statistical methods for anemia prediction.

REFERENCES

1. World Health Organization. Anaemia [Internet]. 2023 [cited 2024 Sep 16]. Available from: https://www.who.int/news-room/fact-sheets/detail/anaemia
2. Gardner WM, Razo C, McHugh TA, Hagins H, Vilchis-Tella VM, Hennessy C, et al. Prevalence, years lived with disability, and trends in anaemia burden by severity and cause, 1990–2021: findings from the Global Burden of Disease Study 2021. Lancet Haematol. 2023 Sep;10(9):e713–34. https://doi.org/10.1016/S2352-3026(23)00160-6
3. Kementerian Kesehatan RI. Survey Kesehatan Indonesia 2023 Dalam Angka [Internet]. 2023 [cited 2024 Sep 16]. Available from: https://www.badankebijakan.kemkes.go.id/ski-2023-dalam-angka/
4. Zemariam AB, Yimer A, Abebe GK, Wondie WT, Abate BB, Alamaw AW, et al. Employing supervised machine learning algorithms for classification and prediction of anemia among youth girls in Ethiopia. Sci Rep. 2024 Apr 20;14(1):9080. https://doi.org/10.1038/s41598-024-60027-4
5. Tartan EO, Berkol A, Ekici Y. Anemia Diagnosis By Using Artificial Neural Networks. Int J Multidiscip Stud Innov Technol. 2020;4(1):14–7.
6. Ortiz-Prado E, Dunn JF, Vasconez J, Castillo D, Viscor G. Review Article: Partial pressure of oxygen in the human body: a general review. Am J Blood Res. 2019;9(1):1–14.
7. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019 Dec;19(1):64. https://doi.org/10.1186/s12874-019-0681-4
8. Khan JR, Chowdhury S, Islam H, Raheem E. Machine Learning Algorithms To Predict The Childhood Anemia In Bangladesh. J Data Sci. 2021 Feb 24;17(1):195–218. https://doi.org/10.6339/JDS.201901_17(1).0009
9. Pullakhandam S, McRoy S. Classification and Explanation of Iron Deficiency Anemia from Complete Blood Count Data Using Machine Learning. BioMedInformatics. 2024 Mar 1;4(1):661–72. https://doi.org/10.3390/biomedinformatics4010036
10. Kassaw AK, Yimer A, Abey W, Molla TL, Zemariam AB. The application of machine learning approaches to determine the predictors of anemia among under five children in Ethiopia. Sci Rep. 2023 Dec 21;13(1):22919. https://doi.org/10.1038/s41598-023-50128-x
11. Meitei AJ, Saini A, Mohapatra BB, Singh KhJ. Predicting child anaemia in the North-Eastern states of India: a machine learning approach. Int J Syst Assur Eng Manag. 2022 Dec;13(6):2949–62. https://doi.org/10.1007/s13198-022-01765-4
12. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak. 2019 Dec;19(1):281. https://doi.org/10.1186/s12911-019-1004-8
13. Shweta N, Pande SD. Prediction of Anemia using various Ensemble Learning and Boosting Techniques. EAI Endorsed Trans Pervasive Health Technol [Internet]. 2023 Oct 20;9. Available from: https://publications.eai.eu/index.php/phat/article/view/4197. https://doi.org/10.4108/eetpht.9.4197
14. Dejene BE, Abuhay TM, Bogale DS. Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm. BMC Med Inform Decis Mak. 2022 Sep 22;22(1):247. https://doi.org/10.1186/s12911-022-01992-6
15. Maasthi MJ, Gururaj HL, Ravi V, D B, Almeshari M, Alzamil Y. Decision-making Support System for Predicting and Eliminating Malnutrition and Anemia. Open Bioinforma J. 2023 Oct 27;16(1):e18750362246898. https://doi.org/10.2174/0118750362246898230921054021

16. Göl M, Aktürk C, Talan T, Vural MS, Türkbeyler İH. Predicting malnutrition-based anemia in geriatric patients using machine learning methods. J Eval Clin Pract. 2024 Sep 23;jep.14142. https://doi.org/10.1111/jep.14142
17. Tesfaye SH, Seboka BT, Sisay D. Application of machine learning methods for predicting childhood anaemia: Analysis of Ethiopian Demographic Health Survey of 2016. Moinuddin M, editor. PLOS ONE. 2024 Apr 11;19(4):e0300172. https://doi.org/10.1371/journal.pone.0300172
18. Pan Y, Du R, Han X, Zhu W, Peng D, Tu Y, et al. Machine Learning Prediction of Iron Deficiency Anemia in Chinese Premenopausal Women 12 Months after Sleeve Gastrectomy. Nutrients. 2023 Jul 30;15(15):3385. https://doi.org/10.3390/nu15153385
19. World Health Organization. Anaemia in women and children [Internet]. 2021 [cited 2024 Oct 9]. Available from: https://www.who.int/data/gho/data/themes/topics/anaemia_in_women_and_children
20. Singh S, Parihar S. Prevalence of anemia in under five-year-old children: a hospital-based study. Int J Contemp Pediatr. 2019 Feb 23;6(2):842. https://doi.org/10.18203/2349-3291.ijcp20190740
21. Anand P, Gupta R, Sharma A. Prediction of Anaemia among children using Machine Learning Algorithms. Int J Electron Eng. 2019 Jun;11(2):469–80.
22. Kumar K, Kumar P, Deb D, Unguresan ML, Muresan V. Artificial Intelligence and Machine Learning Based Intervention in Medical Infrastructure: A Review and Future Trends. Healthcare. 2023 Jan 10;11(2):207. https://doi.org/10.3390/healthcare11020207
23. Gaurav K, Kumar A, Singh P, Kumari A, Kasar M, Suryawanshi T. Human Disease Prediction using Machine Learning Techniques and Real-life Parameters. Int J Eng. 2023;36(6):1092–8. https://doi.org/10.5829/IJE.2023.36.06C.07
24. Dukhi N, Sewpaul R, Derrick Sekgala M, Olawale Awe O. Artificial Intelligence Approach for Analyzing Anaemia Prevalence in Children and Adolescents in BRICS Countries: A Review. Curr Res Nutr Food Sci J. 2021 Apr 30;9(1):01–10. https://doi.org/10.12944/CRNFSJ.9.1.01
25. Wallner C, Hurst J, Behr B, Rony MAT, Barabás A, Smith G. Fanconi Anemia: Examining Guidelines for Testing All Patients with Hand Anomalies Using a Machine Learning Approach. Children. 2022 Jan 7;9(1):85. https://doi.org/10.3390/children9010085
26. Zahirzada A, Zaheer N, Shahpoor MA. Machine Learning Algorithms to Predict Anemia in Children Under the Age of Five Years in Afghanistan: A Case of Kunduz Province. J Surv Fish Sci. 2023;10(4S):752–62.
27. Chen RC, Dewi C, Huang SW, Caraka RE. Selecting critical features for data classification based on machine learning methods. J Big Data. 2020 Dec;7(1):52. https://doi.org/10.1186/s40537-020-00327-4
28. Luque A, Carrasco A, Martín A, De Las Heras A. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. Pattern Recognit. 2019 Jul;91:216–31. https://doi.org/10.1016/j.patcog.2019.02.023
29. Greenwood CJ, Youssef GJ, Letcher P, Macdonald JA, Hagg LJ, Sanson A, et al. A comparison of penalised regression methods for informing the selection of predictive markers. Kotozaki Y, editor. PLOS ONE. 2020 Nov 20;15(11):e0242730. https://doi.org/10.1371/journal.pone.0242730
30. Liu J, Ma Y, Xie W, Li X, Wang Y, Xu Z, et al. Lasso-Based Machine Learning Algorithm for Predicting Postoperative Lung Complications in Elderly: A Single-Center Retrospective Study from China. Clin Interv Aging. 2023 Apr;Volume 18:597–606. https://doi.org/10.2147/CIA.S406735
31. Yeruva S, Varalakshmi MS, Gowtham BP, Chandana YH, Prasad PesnK. Identification of Sickle Cell Anemia Using Deep Neural Networks. Emerg Sci J. 2021 Apr 1;5(2):200–10. https://doi.org/10.28991/esj-2021-01270
32. Christian Y. Comparison of Machine Learning Algorithms Using WEKA and Sci-Kit Learn in Classifying Online Shopper Intention. J Inform Telecommun Eng. 2019 Jul 25;3(1):58–66. https://doi.org/10.31289/jite.v3i1.2599