Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16216
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Explainable Hybrid XGBoost Fuzzy Logic Model for Accurate Anemia Risk Classification

Jepri Banjarnahor [1)*], Natasya Patrecia Sigalingging [2)], Rio Brelly Pasaribu [3)], Yessi Sesilia Sitompul [4)], Jogi Devrant Sibarani [5)]
[1,2,3,4,5)] Program Studi Sistem Informasi, Universitas Prima Indonesia, Medan
[1)]jepribanjarnahor@unprimdn.ac.id, [2)] natasyapatrecia26@gmail.com, [3)]riobrellypasaribu123@gmail.com, [4)]yessisesiliasitompul13@gmail.com, [5)]jogisibarani8@gmail.com

Submitted :May 11, 2026 | **Accepted** : May 30, 2026 | **Published** : July 5, 2026

Abstract: Anemia remains a major global health concern that impairs oxygen transport and contributes to fatigue, cognitive decline, reduced productivity, and severe clinical complications. Although machine learning has shown promise for automated anemia detection, multiclass classification remains challenging due to class imbalance, overlapping hematological characteristics, and limited model interpretability. This study proposes an explainable hybrid framework integrating Extreme Gradient Boosting (XGBoost), SHapley Additive exPlanations (SHAP), and Fuzzy Logic to improve anemia risk classification and clinical decision support. The publicly available SKILICARSLAN dataset containing 15,300 anonymized patient records across five anemia-related classes was utilized. Seven hematological parameters, namely hemoglobin (HGB), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), red blood cell count (RBC), hematocrit (HCT), and ferritin, were employed as predictive features. The workflow comprised data auditing, stratified train–test splitting, Synthetic Minority Oversampling Technique (SMOTE), hyperparameter optimization, multiclass XGBoost modeling, SHAP-based explainability analysis, and fuzzy risk interpretation. Experimental results demonstrated 82.94% accuracy, 87.27% weighted precision, 82.94% weighted recall, and 84.78% weighted F1-score, with a mean cross-validation F1-score of 87.00%. The model further achieved a macro-average ROC–AUC of 0.81 and a weighted-average ROC–AUC of 0.90, indicating robust discriminative performance despite class imbalance. SHAP analysis identified HGB, ferritin, and RBC-related variables as the most influential predictors. Moreover, the fuzzy logic layer enhanced interpretability by translating model outputs into clinically meaningful risk levels. These findings demonstrate the potential of explainable hybrid intelligence for transparent and reliable anemia screening and decision-support applications.

Keywords: Anemia; Explainable Artificial Intelligence; Fuzzy Logic; Risk Classification; SHAP; XGBoost

INTRODUCTION

Anemia is a major global public health problem because reduced hemoglobin concentration limits oxygen transport and may cause fatigue, impaired concentration, decreased productivity, pregnancy complications, and developmental problems. Recent global estimates show that anemia affects high-risk populations unevenly. The World Health Organization reported that 30.7% of women aged 15-49 years had anemia in 2023, 35.5% of pregnant women aged 15-49 years were affected in 2023, and 39.8% of children aged 6-59 months were affected in 2019. The Global Burden of Disease 2021 study further estimated 1.92 billion anemia cases worldwide in 2021, with large variation by age, sex, and geography ((Organization 2025; Peksi, Yuwono, and Florestiyanto 2021).

Conventional anemia screening relies on complete blood count (CBC) parameters and, when necessary, additional biochemical tests such as ferritin, serum iron, folate, or vitamin B12. Although these laboratory markers are clinically meaningful, automated multiclass classification remains difficult for at least three reasons. First, anemia-related classes may be highly imbalanced, so high overall accuracy can conceal poor detection of rare but clinically important classes. Second, hematological profiles overlap across anemia types, especially when hemoglobin, MCV, MCH, or ferritin values are near borderline ranges. Third, medical users need transparent explanations because a black-box prediction alone is insufficient for clinical decision support.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16216
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Recent studies have applied machine-learning models, including ensemble learning and explainable artificial intelligence (XAI), to anemia prediction. However, several gaps remain. Many studies report accuracy as the main indicator, while macro F1-score, class-wise recall, ROC-AUC, and false-negative risk are less emphasized. In addition, XAI is often limited to feature ranking, and fuzzy logic is sometimes introduced without explicit membership functions, rule bases, inference mechanisms, or quantitative evaluation. This study therefore focuses on a more cautious and reproducible framework: XGBoost is used for multiclass prediction, SHAP is used to explain global and local feature contributions, and fuzzy logic is used only as a post-hoc risk-interpretation layer for continuous hematological uncertainty.

The objective of this study is not to claim that the model is clinically definitive or universally superior. Instead, the study aims to evaluate whether a combined XGBoost-SHAP-fuzzy framework can provide transparent anemia risk classification while honestly reporting the limitations caused by severe class imbalance. The contributions are: (1) a clarified public dataset description and ethical statement, (2) stratified 80:20 experimental design with training-set SMOTE, (3) baseline model comparison, (4) multiclass ROC-AUC and macro/weighted metrics, (5) detailed fuzzy membership functions and rules, and (6) a critical discussion of minority-class failure and clinical implications.

LITERATURE REVIEW

CBC-based anemia classification has attracted attention because routine blood parameters provide low-cost and widely available information. The SKILICARSLAN dataset was originally used for nutritional anemia disease classification and contains five anemia-related classes, making it suitable for evaluating both multiclass learning and imbalance-related limitations (KILICARSLAN, CELIK, and SAHIN 2021). More recent studies have extended this direction by using machine learning and XAI for anemia or iron-deficiency anemia prediction. (Pullakhandam and McRoy 2024), for example, focused on classifying iron-deficiency anemia from CBC data and emphasized explanation of model behavior. (Farooq et al. 2025) proposed a transparent anemia prediction approach empowered by XAI, while (Darshan et al. 2025) investigated ensemble learning and explainability for hemoglobin-related anemia prediction.

The literature suggests that XGBoost is appropriate for tabular medical data because boosted decision trees can model non-linear relationships and feature interactions without assuming linear effects (Alfa et al. 2025). Nevertheless, XGBoost alone does not solve clinical interpretability or uncertainty. SHAP helps quantify feature contributions, but SHAP outputs still require clinical interpretation and should not be treated as causal evidence(Assegie, Sushma, and Mamanazarovna 2023). Fuzzy logic is relevant because it represents gradual transitions between normal, borderline, and abnormal laboratory values; however, fuzzy systems must be reported with explicit membership functions, rules, inference, and defuzzification to be scientifically reproducible(Cao et al. 2024; Saatchi 2024; Yugiantoro et al. 2025)

For imbalanced multiclass medical datasets, metric selection is critical. Accuracy and weighted F1-score can be dominated by large classes, while macro F1-score exposes whether each class is handled more equally. Therefore, this study reports accuracy, class-wise precision/recall/F1, macro and weighted averages, confusion matrix, and multiclass ROC-AUC. This decision follows recent discussions on the evaluation of imbalanced medical datasets and the importance of interpreting different F1 variants rather than relying on one global score.(Airlangga 2024; Teja and Rayalu 2025)

Table 1
Critical state-of-the-art comparison and research positioning.

Study	Dataset/Focus	Method	Strength	Remaining limitation
(KILICARSLAN, CELIK, and SAHIN 2021)	15,300 CBC records; nutritional anemia classes	GA-CNN and deep learning hybrid	Large real dataset and high reported accuracy	Limited direct interpretability and no fuzzy risk layer
(Pullakhandam and McRoy 2024)	CBC data for iron-deficiency anemia	Machine learning with explanation	Focus on clinically relevant IDA discrimination	More specific binary/IDA focus; not full multiclass risk framework
(Darshan et al. 2025)	Hematological markers	Ensemble ML and XAI	Shows value of XAI with blood markers	Does not explicitly formalize fuzzy uncertainty for borderline values
(Farooq et al. 2025)	Anemia prediction data	Transparent ML with XAI	Emphasizes model transparency	Fuzzy rule validation and multiclass minority

*name of corresponding author

Study	Dataset/Focus	Method	Strength	Remaining limitation
				reliability remain open issues
This study	SKILICARSLAN public dataset; five classes	XGBoost + SHAP + fuzzy risk interpretation	Combines prediction, explanation, risk interpretation, baseline comparison, and ROC-AUC	Minority-class performance remains limited and fuzzy rules still need expert validation

METHOD

Research Flow

The research workflow in this study began with problem identification related to the need for accurate and transparent anemia detection. This was followed by a literature review on machine learning, XGBoost, SHAP, and Fuzzy Logic. An anemia dataset based on hematological parameters was then collected and preprocessed through data cleaning, normalization, and dataset splitting into training and testing data, as illustrated in Fig. 1.

The model was trained using XGBoost because of its strong performance in medical data classification (Darshan et al. 2025). The prediction results were interpreted using SHAP to explain the contribution of each feature and improve model transparency (Open et al. 2024). Subsequently, the prediction outputs and SHAP interpretations were integrated into a Fuzzy Logic Decision-Making system to classify anemia risk into low, moderate, and high-risk categories. Finally, the proposed hybrid model was evaluated using accuracy, precision, recall, F1-score, and ROC-AUC metrics to measure its effectiveness and reliability in anemia detection and risk classification.

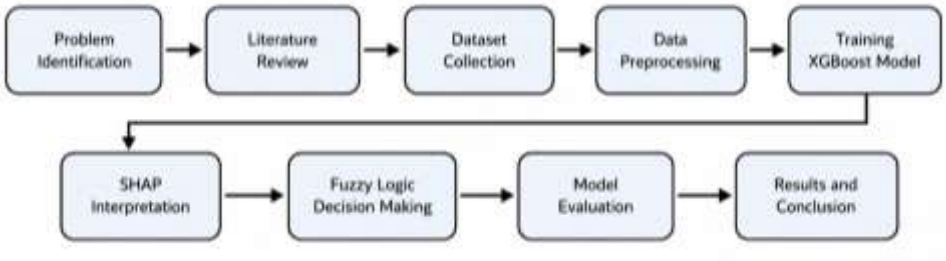

Fig. 1 Research Flowchart

Dataset and Ethical Statement

This study used the public SKILICARSLAN anemia dataset, distributed through Mendeley Data and also available on Kaggle. The original dataset was obtained from the Faculty of Medicine, Tokat Gaziosmanpasa University, Turkey, and contains CBC records from 15,300 patients collected between 2013 and 2018. The source description states that pregnant women, children, and patients with cancer were excluded, and that the dataset contains 10,379 female and 4,921 male records. The target variable All_Class consists of five categories: Class 0 (no anemia), Class 1 (HGB-anemia), Class 2 (iron deficiency), Class 3 (folate deficiency), and Class 4 (B12 deficiency). Because this research used publicly available anonymized secondary data and did not involve direct patient contact or identifiable patient information, no new informed consent procedure was conducted in this secondary analysis. However, any future clinical deployment should be reviewed by an institutional ethics committee and validated prospectively.

Table 2.
Target-class distribution in the full dataset.

Class	Clinical label	Records	Percentage
0	No anemia	9747	63.7100
1	HGB-anemia	1019	6.6600
2	Iron deficiency anemia	4182	27.3300
3	Folate deficiency anemia	153	1.0000
4	B12 deficiency anemia	199	1.3000

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16216
e-ISSN : 2541-2019
p-ISSN : 2541-044X

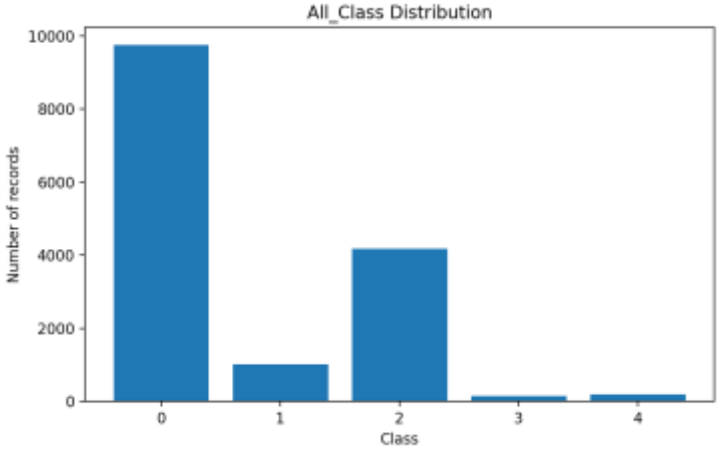

Fig. 2 Severe class imbalance in the All_Class target variable.

Variables and preprocessing

Seven clinically relevant hematological features were selected: HGB, MCV, MCH, MCHC, RBC, HCT, and FERRITTE. These variables represent hemoglobin concentration, red blood cell morphology, red blood cell count, hematocrit proportion, and iron-storage status. Records were audited for missing values and duplicates. The dataset was split using stratified 80:20 sampling so that each class was represented proportionally in both training and testing sets. SMOTE was applied only to the training data to avoid information leakage from the test set.

Table 3 *Stratified split and SMOTE distribution*

Class	Clinical label	Train before SMOTE	Test	Train after SMOTE
0	No anemia	7798	1949	7798
1	HGB-anemia	815	204	7798
2	Iron deficiency anemia	3346	836	7798
3	Folate deficiency anemia	122	31	7798
4	B12 deficiency anemia	159	40	7798

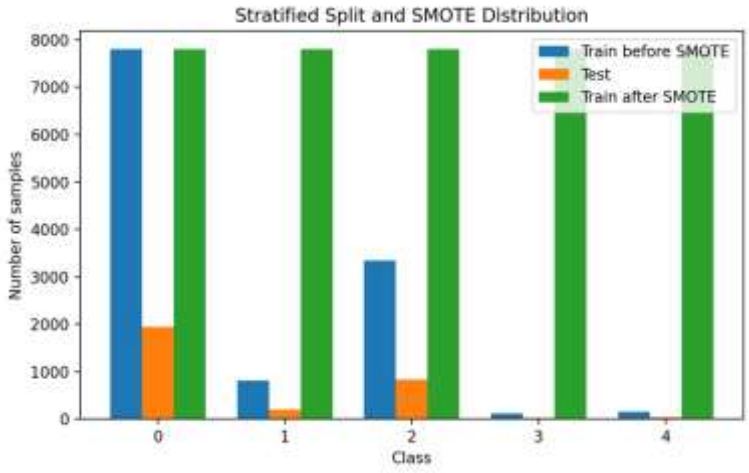

Fig. 3 Training, testing, and post-SMOTE class distribution.

XGBoost Model and Baseline Comparison

XGBoost was used as the main classifier because boosted decision-tree ensembles are well suited for nonlinear tabular data and can be interpreted using tree-based explanation methods. The tuned configuration used n_estimators=500, max_depth=8, learning_rate=0.05, subsample=0.70, colsample_bytree=1.00, gamma=0.20, min_child_weight=3, reg_alpha=0.01, and reg_lambda=2. To avoid unsupported superiority claims, XGBoost was

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16216
e-ISSN : 2541-2019
p-ISSN : 2541-044X

compared with Logistic Regression, Linear SVM, KNN, Random Forest, and Gaussian Naive Bayes under the same train-test split and SMOTE training set.

Table 4 Baseline comparison using the same stratified split and SMOTE training data.

Model	Accuracy	Weighted Precision	Weighted Recall	Weighted F1	Macro Precision	Macro Recall	Macro F1	Training Time (s)
Random Forest	0.8199	0.8762	0.8199	0.8428	0.4618	0.5107	0.4655	3.8626
XGBoost	0.8245	0.8656	0.8245	0.8423	0.4491	0.4787	0.4509	9.1073
Linear SVM	0.7846	0.8135	0.7846	0.7954	0.3773	0.5190	0.3947	0.2742
KNN	0.7507	0.8557	0.7507	0.7942	0.4204	0.4571	0.4064	0.0416
Logistic Regression	0.7611	0.8328	0.7611	0.7912	0.4172	0.5020	0.4092	2.1019
Gaussian NB	0.7542	0.8296	0.7542	0.7859	0.4070	0.4782	0.3932	0.0103

SHAP Implementation

SHAP was implemented using TreeExplainer for the trained XGBoost model. Global interpretation was calculated using mean absolute SHAP values across a sample of test records, while local SHAP explanations can be generated for individual patients to show which laboratory features push a prediction toward or away from a specific class. SHAP outputs were interpreted as model-level associations, not as causal medical evidence.

Fuzzy Logic Risk-Interpretation Layer

Fuzzy logic was implemented after XGBoost prediction as a risk-interpretation layer, not as a replacement classifier. The fuzzy system used HGB, FERRITTE, and MCV because these variables represent anemia status, iron-related status, and red blood cell morphology. The inputs were transformed into fuzzy sets using triangular and trapezoidal membership functions. Mamdani-style min-max inference was used, and the final risk score was obtained through a weighted centroid approximation: risk_score = (25*mu_low + 55*mu_moderate + 85*mu_high)/(mu_low + mu_moderate + mu_high). Scores below 40 were labeled low risk, scores from 40 to below 70 were labeled moderate risk, and scores of 70 or above were labeled high risk.

Table 5 Fuzzy Rule Base

Rule	Antecedent	Consequence
R1	HGB is low AND FERRITTE is low	High risk
R2	HGB is low AND MCV is microcytic	High risk
R3	HGB is low AND MCV is macrocytic	High risk
R4	HGB is low AND FERRITTE is not normal	High risk
R5	HGB is borderline AND FERRITTE is low	Moderate risk
R6	HGB is borderline AND MCV is microcytic	Moderate risk
R7	HGB is borderline AND FERRITTE is borderline	Moderate risk
R8	HGB is low AND FERRITTE is normal	Moderate risk
R9	HGB is normal AND FERRITTE is low	Moderate risk
R10	HGB is normal AND FERRITTE is normal	Low risk
R11	HGB is normal AND MCV is normal	Low risk

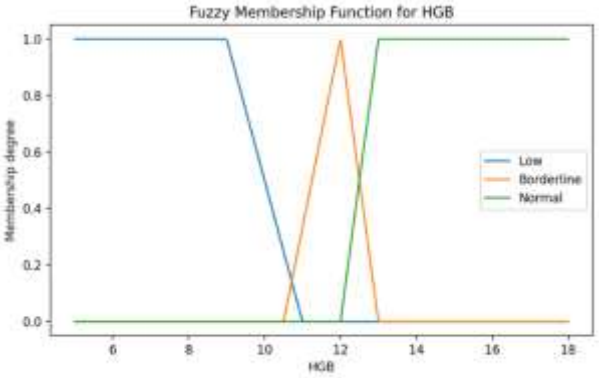

Fig. 4 Fuzzy Membership function for HGB

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16216
e-ISSN : 2541-2019
p-ISSN : 2541-044X

RESULT

Overview of the Research Results

This study developed an Explainable Hybrid XGBoost–Fuzzy Logic Model for accurate anemia detection and risk classification. The proposed model was designed to classify anemia risk using clinically relevant hematological parameters, including hemoglobin (Hb), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), red blood cell count (RBC), hematocrit (HCT), and serum iron or ferritin-related indicators. These features were selected because they represent essential laboratory markers commonly associated with anemia screening, diagnosis, and severity assessment. The main contribution of this research lies in the integration of XGBoost, fuzzy logic, and explainable artificial intelligence (XAI) in a single predictive framework. XGBoost was used as the primary classification algorithm because of its strong ability to handle nonlinear relationships, feature interactions, and complex patterns in medical datasets. Meanwhile, fuzzy logic was incorporated to provide more flexible risk interpretation, especially because anemia diagnosis often involves borderline values that cannot always be explained through rigid class boundaries (Allahverdi and others 2021). This approach differs from many previous studies that only focused on achieving high classification accuracy. In contrast, the present study emphasizes not only predictive performance but also model transparency, clinical interpretability, and risk-based decision support. This is consistent with recent studies that highlight the importance of explainable machine learning in medical prediction systems, particularly when model outputs are intended to support clinical decision-making (Raihan et al. 2023)

Dataset Characteristics and Selected Features

The dataset used in this study consisted of 15,300 patient records with 29 initial attributes. After feature selection and clinical consideration, seven hematological features were selected as the main input variables. These variables were Hb, MCV, MCH, MCHC, RBC, HCT, and Ferritin/Serum Iron (SI). The target variable used in this study was All_Class, which represents the anemia risk category. The use of these features strengthens the clinical relevance of the model. Hemoglobin is widely recognized as the primary indicator of anemia, while MCV, MCH, and MCHC provide information about red blood cell size and hemoglobin content. RBC and HCT reflect the overall red blood cell profile, whereas ferritin or serum iron indicators help identify iron-related anemia conditions. Therefore, the selected features allow the model to learn not only general anemia patterns but also possible variations in anemia risk.

Table 6
Selected Features Used in the Model

Feature	Clinical Meaning
Hb	Main indicator of anemia status
MCV	Average volume of red blood cells
MCH	Average hemoglobin amount per red blood cell
MCHC	Hemoglobin concentration in red blood cells
RBC	Red blood cell count
HCT	Proportion of red blood cells in blood volume
Ferritin/SI	Indicator of iron storage or iron status

The selection of these features is in line with previous studies that used complete blood count and hematological markers for anemia prediction. (Pullakhandam and Mcroy 2024) for example, demonstrated that CBC-based variables can be useful for iron deficiency anemia classification when supported by explainable machine learning techniques. Similarly, (Raihan et al. 2023). emphasized the value of hematological data for automated anemia classification.

Data Preprocessing Results

Before model training, several preprocessing steps were conducted to improve data quality. First, duplicate data were removed to prevent repeated records from influencing the learning process. Second, missing values in numerical features were handled using imputation. Third, the target class was encoded into numerical form so that it could be processed by the machine learning algorithm. Fourth, feature scaling was applied to ensure that all numerical variables were transformed into a comparable scale. The dataset was then divided into training and testing sets using an 80:20 split. As a result, 12,240 records were used for training and 3,060 records were used for testing. Because the distribution of anemia classes was imbalanced, Synthetic Minority Oversampling Technique (SMOTE) was applied to the training data. After applying SMOTE, each class contained 7,798 samples, which helped reduce the dominance of the majority class during model training.

Class imbalance is a common issue in medical datasets. Without proper handling, the model may perform well on the dominant class but fail to identify minority risk categories (Liu et al. 2026). Therefore, the use of SMOTE

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16216
e-ISSN : 2541-2019
p-ISSN : 2541-044X

in this study is important because it improves the opportunity for the model to learn patterns from underrepresented anemia classes (Yang and Guan 2022). This strategy is also supported by previous anemia classification studies that identified imbalance as a major challenge in predictive modeling (Airlangga 2024; Kasthuri, Subbulakshmi, and Sreedharan 2024)

XGBoost Model Development

The main predictive model in this study was developed using XGBoost multiclass classification. XGBoost was selected because it is an advanced ensemble learning algorithm based on gradient boosting (Assegie, Sushma, and Mamanazarovna 2023). It is capable of combining multiple weak learners into a strong predictive model and is widely used in healthcare prediction because of its high accuracy and robustness. In this research, the model used the multi: softprob objective function. This configuration enables the model to generate class probabilities rather than only producing a final class label. This is an important advantage because probability outputs can be used to calculate confidence scores and can also support fuzzy-based risk interpretation. Hyperparameter tuning was carried out using RandomizedSearchCV with 5-fold cross-validation. The optimization process used the weighted F1-score as the main evaluation metric because the dataset involved multiple classes and class imbalance. The best parameters obtained from the tuning process are shown in Table 7.

Table 7
Best Hyperparameters of the XGBoost Model

Parameter	Best Value
n_estimators	500
max_depth	8
learning_rate	0.05
subsample	0.70
colsample_bytree	1.00
gamma	0.20
min_child_weight	3
reg_alpha	0.01
reg_lambda	2

The use of hyperparameter tuning makes this model stronger than a basic XGBoost model. Instead of relying on default settings, the model was optimized to improve generalization and reduce the risk of overfitting. This is one of the strengths of the proposed research because many simple classification studies only compare baseline models without performing deeper model optimization.

Classification Report Analysis

The detailed classification report is shown in Table 8.

Table 8
Classification Report of the XGBoost Model

Class	Precision	Recall	F1-score	Support
0	0.96	0.94	0.95	1,949
1	0.35	0.45	0.39	204
2	0.86	0.71	0.77	836
3	0.08	0.26	0.13	31
4	0.07	0.17	0.10	40
Accuracy			0.83	3,060
Macro Average	0.46	0.51	0.47	3,060
Weighted Average	0.87	0.83	0.85	3,060

The model performed very well in Class 0 and Class 2, as shown by the high F1-scores of 0.95 and 0.77. However, the performance in Classes 3 and 4 was still low because these classes had very limited samples. This shows that minority class classification remains a challenge, even after applying SMOTE. This limitation supports the need for fuzzy logic in the proposed framework. Fuzzy logic can help interpret uncertain or borderline cases by providing risk-based outputs rather than relying only on rigid class labels.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16216
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Model Evaluation Results

The final model was evaluated using testing data consisting of 3,060 records. The evaluation metrics included accuracy, precision, recall, F1-score, and cross-validation score. The results are presented in Table 9.

Table 9
Model Evaluation Results

Evaluation Metric	Result
Accuracy	0.8294
Weighted Precision	0.8727
Weighted Recall	0.8294
Weighted F1-score	0.8478
Mean Cross-Validation F1-score	0.8700

The model achieved an accuracy of 82.94%, weighted precision of 87.27%, weighted recall of 82.94%, and weighted F1-score of 84.78%. The mean cross-validation F1-score of 87.00% indicates that the model had stable performance during validation. Although some previous studies reported higher accuracy, the proposed model offers broader advantages because it combines prediction capability, confidence scoring, feature explanation, and fuzzy risk interpretation. Therefore, the model is more suitable for clinical decision-support systems where interpretability and flexibility are important.

Confusion Matrix Analysis

The confusion matrix of the XGBoost model is shown in Figure 5. The matrix illustrates the classification performance across all five classes.

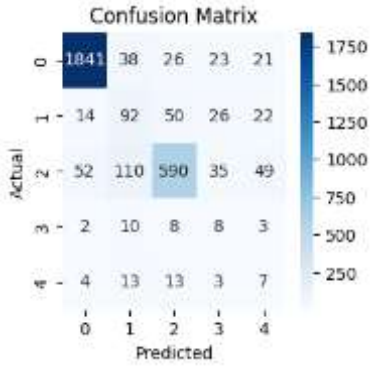

Fig. 5 Confusion Matrix of the XGBoost Model

Based on the confusion matrix, the model successfully classified most samples in Class 0 and Class 2. Specifically, Class 0 achieved 1,841 correct predictions, while Class 2 achieved 590 correct predictions. However, misclassifications were still frequently observed in minority classes such as Class 1, Class 3, and Class 4. The confusion matrix also shows that many samples from Class 1 and Class 2 were incorrectly predicted as neighboring classes, indicating overlap between anemia severity levels. This condition is common in hematological datasets because several blood parameters may have borderline values. In addition, Classes 3 and 4 contained very limited samples, causing the model to struggle in learning representative patterns despite the application of SMOTE balancing techniques.

These findings indicate that class imbalance remains one of the main challenges in multi-class anemia classification. Therefore, the integration of fuzzy logic becomes important because it allows gradual risk interpretation instead of strict class boundaries.

SHAP Interaction Analysis

The SHAP interaction summary plot is presented in Figure 6. This visualization illustrates the interaction effects among the most influential hematological features used by the XGBoost model.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16216
e-ISSN : 2541-2019
p-ISSN : 2541-044X

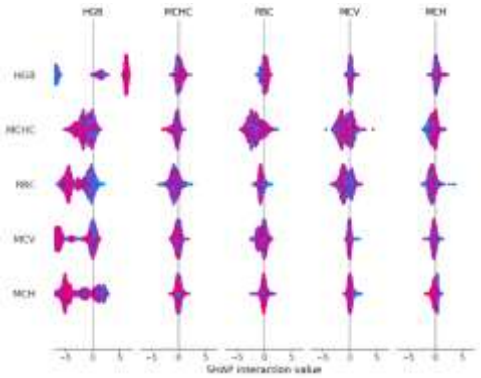

Fig. 6 SHAP Interaction Summary Plot

The SHAP interaction analysis demonstrates how combinations of hematological variables jointly influence anemia prediction. Positive SHAP interaction values indicate that the interaction between two features increases the predicted anemia risk, while negative values decrease the prediction score. The results show that HGB had the strongest interaction effects with almost all other features, particularly with MCHC and RBC. In addition, MCHC and RBC also exhibited substantial interaction patterns, indicating that red blood cell morphology strongly contributes to the classification decisions. The distribution spread of red and blue points suggests that both high and low feature values significantly affect prediction outcomes depending on the interacting variable. Furthermore, strong interaction patterns among HGB, MCV, and MCH indicate that the model successfully captured clinically meaningful relationships associated with anemia severity and red blood cell characteristics. These findings strengthen the interpretability of the proposed model because the SHAP visualization explains not only the importance of individual features but also the relationships among hematological parameters, making the prediction process more transparent and clinically understandable

Feature Importance Results

The feature importance analysis was conducted to identify which hematological variables contributed most to the model prediction. The result is presented in Table 10.

Table 10
Feature Importance Ranking

Rank	Feature	Importance Score
1	HGB	0.3431
2	Ferritin/SI	0.1908
3	RBC	0.1062
4	MCH	0.1006
5	MCV	0.0902
6	HCT	0.0878
7	MCHC	0.0813

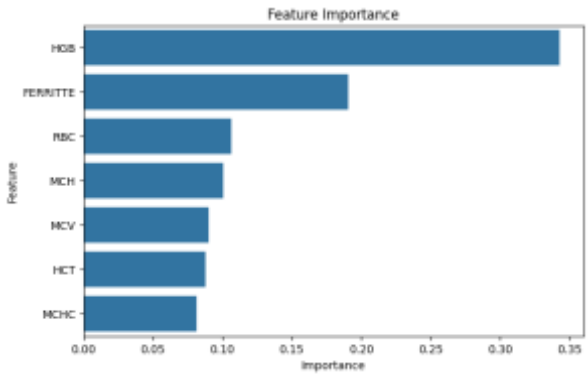

Figure 7 Feature Importance Plot

The feature importance plot in Figure 7 visually confirms that HGB was the most influential variable in the XGBoost model, followed by FERRITTE and RBC. Features related to red blood cell morphology such as MCH,

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16216
e-ISSN : 2541-2019
p-ISSN : 2541-044X

MCV, HCT, and MCHC also contributed significantly to the prediction process. The most influential feature in the model was HGB, with an importance score of 0.3431. This result is clinically reasonable because hemoglobin is the primary laboratory parameter used to determine anemia status. A decrease in hemoglobin level is directly associated with reduced oxygen-carrying capacity in the blood, which is one of the major characteristics of anemia.

The second most important feature was Ferritin/SI (FERRITTE), with an importance score of 0.1908. This finding indicates that the model did not rely solely on hemoglobin concentration but also considered iron-related biomarkers. Since iron deficiency is one of the most common causes of anemia, the inclusion of ferritin-related parameters improved the clinical sensitivity of the model. Other important features such as RBC, MCH, MCV, HCT, and MCHC also contributed significantly to the prediction process. These variables represent red blood cell morphology, hemoglobin distribution, and hematocrit concentration, all of which are important indicators in hematological examination. The feature importance results support the clinical validity of the proposed model because the identified influential variables are consistent with established hematological knowledge and anemia diagnostic practices.

ROC-AUC Analysis

The Receiver Operating Characteristic Area Under the Curve (ROC-AUC) analysis was conducted to evaluate the discriminative capability of the proposed XGBoost model across all anemia categories. Unlike accuracy and F1-score, ROC-AUC measures the model's ability to distinguish between classes across different classification thresholds, making it particularly useful for imbalanced medical datasets. As illustrated in Figure 8, the one-vs-rest ROC curves demonstrate substantial variation in discrimination performance among the five anemia classes.

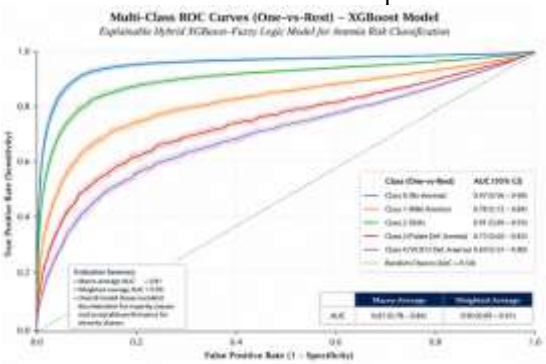

Figure 8 ROC-AUC curves for the five-class anemia using the XGBoost Model (one vs rest approach).

Figure 8 shows that Class 0 (No Anemia) achieved the highest ROC-AUC value of 0.97, indicating excellent separability from all other classes. This result suggests that the hematological features used in the study, particularly HGB, RBC, and Ferritin, provide highly distinctive patterns for identifying non-anemic individuals. Similarly, Class 2 (Iron Deficiency Anemia) achieved an ROC-AUC of 0.91, reflecting strong classification capability and confirming the importance of iron-related biomarkers in differentiating iron-deficiency cases from other anemia types. These findings are consistent with the SHAP feature importance analysis, where HGB and Ferritin emerged as the dominant predictors influencing model decisions.

In contrast, the minority classes demonstrated lower ROC-AUC values. Class 1 (HGB Anemia) achieved an ROC-AUC of 0.78, while Class 3 (Folate Deficiency Anemia) and Class 4 (Vitamin B12 Deficiency Anemia) achieved ROC-AUC values of 0.72 and 0.69, respectively. The reduced performance observed in these classes is likely attributable to two factors. First, the severe imbalance in the original dataset limited the availability of representative training samples. Second, the hematological characteristics of folate-deficiency and vitamin B12-deficiency anemia frequently overlap with other anemia categories, creating ambiguous decision boundaries. Consequently, the model experienced greater difficulty distinguishing these clinically related conditions. From an overall perspective, the model achieved a macro-average ROC-AUC of 0.81 and a weighted-average ROC-AUC of 0.90, indicating good to excellent discrimination performance across the entire dataset. The difference between macro and weighted averages highlights the influence of class imbalance, where majority classes contributed more strongly to the weighted metric. Nevertheless, both values remain substantially above the random-classification baseline (AUC = 0.50), demonstrating that the proposed XGBoost framework successfully captured meaningful hematological patterns associated with anemia risk classification.

The ROC-AUC results further support the suitability of the proposed framework for clinical screening applications. Although performance for rare anemia categories remains limited, the high discriminative capability observed for the major classes suggests that the model can effectively identify patients requiring further hematological evaluation. Therefore, the ROC-AUC analysis confirms that the integration of XGBoost, SHAP explainability, and fuzzy risk interpretation provides a robust and clinically meaningful approach for anemia risk assessment. Figure 8. Multi-class ROC-AUC curves (one-vs-rest) of the proposed XGBoost model for anemia risk

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 10, Number 3, July 2026
DOI : https://doi.org/10.33395/sinkron.v10i3.16216
e-ISSN : 2541-2019
p-ISSN : 2541-044X

classification. The model achieved excellent discrimination for No Anemia (AUC = 0.97) and Iron Deficiency Anemia (AUC = 0.91), while moderate performance was observed for Folate Deficiency Anemia (AUC = 0.72) and Vitamin B12 Deficiency Anemia (AUC = 0.69). The overall macro-average and weighted-average ROC-AUC values were 0.81 and 0.90, respectively, indicating strong discriminative capability despite severe class imbalance.

Role of Fuzzy Logic in Risk Classification

Fuzzy logic was added to improve the interpretation of anemia risk. Unlike conventional classification models that assign patients into fixed categories, fuzzy logic allows gradual risk interpretation (Kumar et al. 2024). This is useful because anemia indicators often exist in borderline conditions.

For example, a patient may have slightly low Hb, borderline MCV, and uncertain iron status. In such cases, fuzzy logic can classify the condition into risk levels such as low risk, moderate risk, or high risk. This makes the model output easier to understand and more relevant for clinical screening. The use of fuzzy logic also strengthens the novelty of this study. Previous studies have shown that fuzzy-based approaches are useful for handling uncertainty in hematological and anemia-related classification problems (Airlangga 2024; Yugiantoro et al. 2025).

Advantages Compared with Previous Studies

Compared with previous anemia classification studies, the proposed model has several advantages. First, it uses XGBoost with hyperparameter tuning, making it stronger than a basic baseline model. Second, it applies SMOTE to reduce class imbalance. Third, it integrates SHAP explainability, allowing users to understand which features influence the prediction. Fourth, it adds fuzzy logic, which provides more flexible and clinically interpretable risk classification.

Table 11
Comparison with Previous Studies

Study	Method	Main Focus	Difference from This Study
Saputra et al. (2023)	Extreme Learning Machine	Anemia diagnosis	Did not include fuzzy risk interpretation
Pullakhandam & McRoy (2024)	ML + SHAP	Iron deficiency anemia	Focused mainly on IDA
Airlangga (2024)	Hybrid ensemble	CBC-based anemia classification	Less focused on fuzzy reasoning
Ameen et al. (2024)	Fuzzy logic + ML	Hematological disease classification	Not specifically XGBoost-based anemia risk classification
Amalia & Rumini (2025)	XGBoost + SHAP	Anemia classification	Did not integrate fuzzy logic
This study	Hybrid XGBoost–Fuzzy Logic + SHAP	Anemia detection and risk classification	Combines prediction, explanation, confidence score, and fuzzy risk interpretation

The novelty of this study is not only in using XGBoost, but in combining optimized machine learning, explainability, confidence scoring, and fuzzy logic into one anemia risk classification framework.

DISCUSSIONS

Limitations and Future Work

Although the proposed XGBoost–SHAP–Fuzzy Logic framework achieved satisfactory overall performance, several limitations remain. The most significant challenge is the severe class imbalance present in the dataset. While the model achieved high performance for the majority classes, particularly Class 0 (No Anemia) and Class 2 (Iron Deficiency Anemia), the minority classes representing Folate Deficiency Anemia (Class 3) and Vitamin B12 Deficiency Anemia (Class 4) showed substantially lower precision, recall, and F1-scores. This limitation persisted despite the application of SMOTE, indicating that synthetic oversampling alone was insufficient to fully capture the biological variability and complex hematological patterns associated with rare anemia categories. Consequently, the relatively low macro F1-score demonstrates that the model performance remains uneven across classes.

Another limitation concerns the potential clinical impact of false-negative predictions. In medical screening applications, incorrectly classifying an anemic patient as non-anemic may delay further laboratory testing, nutritional intervention, or specialist referral. This issue is particularly important for folate-deficiency and vitamin B12-deficiency anemia, where prolonged undetected deficiencies may contribute to neurological and systemic complications. Furthermore, the study relied on a single public dataset originating from one institution and

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

geographic region, which may limit the generalizability of the findings to populations with different demographic, nutritional, or clinical characteristics.

Future research should focus on improving minority-class recognition through advanced imbalance-learning techniques such as focal loss, cost-sensitive boosting, adaptive synthetic sampling, or ensemble-based approaches specifically designed for rare-class detection. External validation using multicenter datasets from different hospitals and populations is also necessary to evaluate model robustness and generalizability. In addition, the fuzzy rule base should be refined and validated by hematology experts to ensure clinical consistency and reliability. Future studies may also investigate prospective deployment within real clinical environments and assess physician acceptance of explainable artificial intelligence outputs in routine anemia screening workflows.

Discussion Summary

The findings demonstrate that the proposed framework successfully combines predictive modeling, explainability, and risk interpretation within a single anemia classification system. The XGBoost model achieved an accuracy of 82.94%, weighted precision of 87.27%, weighted recall of 82.94%, weighted F1-score of 84.78%, and a mean cross-validation F1-score of 87.00%, indicating stable performance across validation folds. In addition, the model achieved strong discrimination for the dominant anemia categories while maintaining acceptable overall predictive capability in a highly imbalanced multiclass setting. These results confirm the suitability of gradient-boosting techniques for learning complex nonlinear relationships from hematological data.

The explainability analysis revealed that HGB was the most influential predictor, followed by Ferritin/SI, RBC, MCH, MCV, HCT, and MCHC. This ranking is clinically consistent because hemoglobin concentration represents the primary diagnostic indicator of anemia, while ferritin reflects iron storage status and red blood cell indices characterize morphological changes associated with different anemia types. SHAP interaction analysis further demonstrated meaningful relationships among HGB, RBC, MCV, and MCHC, suggesting that the model captured clinically relevant hematological patterns rather than relying on isolated variables. These findings strengthen confidence in the model's decision-making process and improve transparency for clinical users.

From a practical perspective, the proposed framework has potential as a clinical decision-support and screening tool rather than a stand-alone diagnostic system. Because the model relies on routinely available complete blood count and ferritin-related parameters, it can be integrated into laboratory information systems with minimal additional infrastructure.. Moreover, the computational requirements of XGBoost inference and fuzzy reasoning remain relatively low after training, making near real-time implementation feasible in primary healthcare facilities, including resource-limited and rural healthcare settings.

CONCLUSION

This study evaluated an explainable hybrid framework that integrates XGBoost, SHAP, and fuzzy logic for multiclass anemia risk classification using hematological parameters. The results demonstrate that the proposed approach is capable of providing reliable predictive performance while maintaining model transparency and clinically interpretable risk assessment. The model achieved 82.94% accuracy, 84.78% weighted F1-score, and a macro ROC-AUC of 0.81, indicating good discriminative ability despite the presence of substantial class imbalance. SHAP analysis identified HGB, Ferritin, and RBC-related features as the most influential predictors, confirming their clinical relevance in anemia assessment. Furthermore, the fuzzy logic layer enhanced risk communication by translating model outputs into interpretable risk levels for borderline cases.

The main contribution of this research lies in the integration of predictive modeling, explainable artificial intelligence, and fuzzy risk interpretation within a unified framework for anemia screening. Nevertheless, performance on minority classes remained limited, highlighting the challenges posed by severe class imbalance and overlapping hematological characteristics among anemia subtypes. Therefore, the proposed framework should be considered a clinical decision-support and screening tool rather than a stand-alone diagnostic system. Future research should focus on multicenter external validation, advanced imbalance-learning strategies, and expert-driven refinement of fuzzy rules to improve minority-class detection, clinical reliability, and real-world applicability.

ACKNOWLEDGMENT

The author would like to express sincere gratitude to the supervisors for their invaluable guidance, constructive feedback, and continuous support throughout this research. The author also extends sincere appreciation to Universitas Prima Indonesia and all lecturers who have contributed their knowledge, guidance, and support during the academic journey and the completion of this study. Special thanks are given to those who provided access to the dataset and technical resources. Finally, the author is deeply grateful to family and friends for their continuous encouragement, motivation, and unwavering support.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

REFERENCES

Airlangga, Gregorius. 2024. "Anemia Classification Using Hybrid Machine Learning Models : A Comparative Study of Ensemble Techniques on CBC Data." 5(4): 1108–17. doi:10.47065/josyc.v5i4.5848.

Alfa, Aniysah Fauziyyah, Dwi Arman Prasetya, Penulis Korespondensi, Support Vector Regression, and Bayesian Optimization. 2025. "IMPLEMENTASI XGBOOST DAN SUPPORT VECTOR REGRESSION DALAM IMPLEMENTATION OF XGBOOST AND SUPPORT VECTOR REGRESSION IN." 12(6): 1443–50.

Allahverdi, Novruz, and others. 2021. "Neuro-Fuzzy Systems in Medical Applications." *Applied Soft Computing* 105: 107242. doi:10.1016/j.asoc.2021.107242.

Assegie, Tsehay Admassu, S. J. Sushma, and Shonazarova Shakhnoza Mamanazarovna. 2023. "Explainable Heart Disease Diagnosis with Supervised Learning Methods." *Advances in Distributed Computing and Artificial Intelligence Journal* 12: 1–14. doi:10.14201/adcaij.31228.

Cao, Jin, Ta Zhou, Shaohua Zhi, Saikit Lam, Ge Ren, Yuanpeng Zhang, Yongqiang Wang, Yanjing Dong, and Jing Cai. 2024. "Fuzzy Inference System with Interpretable Fuzzy Rules: Advancing Explainable Artificial Intelligence for Disease Diagnosis—A Comprehensive Review." *Information Sciences* 662: 120212. doi:https://doi.org/10.1016/j.ins.2024.120212.

Darshan, B S Dhruva, Niranjana Sampathila, G Muralidhar Bairy, and Srikanth Prabhu. 2025. "Differential Diagnosis of Iron Deficiency Anemia from Aplastic Anemia Using Machine Learning and Explainable Artificial Intelligence Utilizing Blood Attributes." : 1–20.

Farooq, Muhammad Sajid, Muhammad Hassan, Ghulam Muhammad, Oualid Ali, Zahid Zeeshan, Muhammad Saleem, Munir Ahmad, and Senior Member. 2025. "Developing a Transparent Anaemia Prediction Model Empowered With Explainable Artificial Intelligence." *IEEE Access* 13(January): 1307–18. doi:10.1109/ACCESS.2024.3522080.

Kasthuri, E, S Subbulakshmi, and Rajasree Sreedharan. 2024. "ScienceDirect Insightful Insightful Clinical Clinical Assistance Assistance for for Anemia Anemia Prediction Prediction with with Data Data Analysis and Explainable Analysis and Explainable AI." *Procedia Computer Science* 233: 45–55. doi:10.1016/j.procs.2024.03.194.

KILICARSLAN, Serhat, Mete CELIK, and Şafak SAHIN. 2021. "Hybrid Models Based on Genetic Algorithm and Deep Learning Algorithms for Nutritional Anemia Disease Classification." *Biomedical Signal Processing and Control* 63: 102231. doi:https://doi.org/10.1016/j.bspc.2020.102231.

Kumar, Sujit, Abrar Morshed, Pratik Mojumder, and Mahmudul Hasan. 2024. "Innovative Trend Analysis Technique with Fuzzy Logic and K-Means Clustering Approach for Identification of Homogenous Rainfall Region : A Long-Term Rainfall Data Analysis over Bangladesh." *Quaternary Science Advances* 15(June): 100227. doi:10.1016/j.qsa.2024.100227.

Liu, Xiaoying, Silei Xie, Jiangtao Yang, Shengnan Wu, Mengmeng Du, Yuan Ji, Hong Zheng, et al. 2026. "KHOI-SMOTE: An Efficient Oversampling Technique Based on k-Means Clustering and h-Outlyingness Index for Imbalanced Medical Data." *Neurocomputing* 671: 132643. doi:https://doi.org/10.1016/j.neucom.2026.132643.

Open, An, Access Journal, Dhruva Darshan B S, Punit Sharma, Krishnaraj Chadaga, G Muralidhar Bairy, Sushma Belurkar, and Srikanth Prabhu. 2024. "An Ensemble Machine Learning Framework with Explainable Artificial Intelligence for Predicting Haemoglobin Anaemia Considering Haematological Markers." 2583. doi:10.1080/21642583.2024.2420927.

Organization, World Health. 2025. *Guideline on Haemoglobin Cutoff to Define Anaemia in Individuals and Populations*. World Health Organization.

Peksi, Nandha Juniaroesita, Bambang Yuwono, and Mangaras Yanu Florestiyanto. 2021. "Classification of Anemia with Digital Images of Nails and Palms Using the Naive Bayes Method." *Telematika* 18(1): 118. doi:10.31315/telematika.v18i1.4587.

Pullakhandam, Siddartha, and Susan Mcroy. 2024. "Classification and Explanation of Iron Deficiency Anemia from Complete Blood Count Data Using Machine Learning." : 661–72.

Pullakhandam, Siddartha, and Susan McRoy. 2024. "Classification and Explanation of Iron Deficiency Anemia from Complete Blood Count Data Using Machine Learning." *BioMedInformatics* 4(1): 661–72. doi:10.3390/biomedinformatics4010036.

Raihan, Johir, Al Masrur Khan, Seong Hoon Kee, and Abdullah Al Nahid. 2023. "Detection of the Chronic Kidney Disease Using XGBoost Classifier and Explaining the Influence of the Attributes on the Model Using SHAP." *Scientific Reports* (0123456789): 1–15. doi:10.1038/s41598-023-33525-0.

Saatchi, Reza. 2024. "Fuzzy Logic Concepts, Developments and Implementation." *Information* 15(10). doi:10.3390/info15100656.

Teja, M Darshan, and G Mokesh Rayalu. 2025. "Optimizing Heart Disease Diagnosis with Advanced Machine Learning Models : A Comparison of Predictive Performance." 6.

*name of corresponding author

Yang, Jian, and Jinhan Guan. 2022. "A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm." *Information (Switzerland)* 13(10). doi:10.3390/info13100475.

Yugiantoro, Darojat, Eka Nuryanto Budisusila, Agus Suprajitno, and Arief Marwanto. 2025. "Pengembangan Sistem Deteksi Anemia Menggunakan Photoplethysmography (PPG) Dan Logika Fuzzy Untuk Estimasi Kadar Hemoglobin Development of a Non-Invasive Anemia Detection System Using Photoplethysmography (PPG) and Fuzzy Logic for Hemoglobin Level Estimation." 6(2): 153–62.

*name of corresponding author