ADBA
Computers and Electronics in Medicine
Cutting-Edge Scientific Solutions

e-ISSN: 3023-8609
RESEARCH ARTICLE
Vol.2 / **No.**2 / 2025 / **pp.**43-52
https://doi.org/10.69882/adba.cem.2025073

AI-Driven Classification of Anemia and Blood Disorders Using Machine Learning Models

Yigitcan Cakmak *,1 **and Ishak Pacal** α,2
*Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Türkiye.

ABSTRACT

Anemia and other blood disorders are serious global health issues affecting millions of individuals. These conditions, often triggered by insufficient hemoglobin or red blood cells, can manifest symptoms like fatigue, weakness, and reduced immune function. When such disorders progress into advanced stages, they can compromise organ function and overall quality of life, making early diagnosis especially critical. In recent years, as the value of prompt detection has become increasingly clear, artificial intelligence and autonomous diagnostic technologies have begun to take center stage in the medical community. Machine learning models excel at parsing complex datasets and generating accurate, rapid assessments, thus offering clinicians robust decision-support tools. Through these AI-driven methods, healthcare professionals can better interpret patients' blood metrics and clinical indicators, enabling them to identify diseases at earlier stages and develop more effective treatment strategies. This study proposes a machine learning–based approach to classify various types of anemia and related blood disorders, including iron deficiency anemia, leukemia, and thrombocytopenia. We trained five contemporary algorithms Decision Tree, Random Forest, CatBoost, Gradient Boosting, and XGBoost using critical blood parameters such as white and red blood cell counts, hemoglobin levels, and platelet counts. Notably, Gradient Boosting emerged as the most accurate model, achieving an impressive 99.19% accuracy rate. These findings underscore how AI-powered autonomous diagnostic systems have the potential to revolutionize hematology by facilitating earlier and more precise disease detection.

KEYWORDS
Anemia
Machine learning
Blood disorders
Clinical decision support
Hematological data

INTRODUCTION

Anemia, a spectrum of blood orders characterized by the shortness of red blood cells or hemoglobin, is one of the most common global health challenges that considerably threatens people of all ages. While the symptoms differ, common manifestations of anemia are fatigue, weakness, and pale skin (Yoshida 2024). Causes found in the medical literature about anemia include nutritional deficiencies, chronic diseases, or diseases associated with the bone marrow (Krieg *et al.* 2024). Early identification of anemia is very critical for proper treatment and management because untreated anemia may develop into other life-threatening conditions. The classification of anemia and other blood ailments like leukemia, thrombocytopenia, and macrocytic anemia is very important when it comes to the diagnostic aspect and to plan for treatment (Fentie *et al.* 2020; Subba and Araveti 2025).

Identification of anemia falls along the lines of multiple contributors called blood markers at very different levels; the most critical among them are white blood cell count (WBC), red blood cell count (RBC), hemoglobin (HGB), and platelet count (PLT) (Karra *et al.* 2025; Malak *et al.* 2025; Li *et al.* 2025). Recent advancements in machine learning (ML) include various algorithms such as Decision Tree (DT), Random Forest (RF), CatBoost, Gradient Boosting (GB), and XGBoost, showing great potential in the classification and prediction of blood disorders; in our case, anemia. Traditional methods for diagnosis tend to take longer and are expensive, whereas ML algorithms have the capability of analyzing medical data that is very complicated and involved, very quickly and accurately (Link *et al.* 2024).

Integration of the algorithms of Decision Tree (DT), Random Forest (RF), CatBoost, Gradient Boosting (GB), and XGBoost highly improved the accuracy of the disease classification models. For our study, we examined the algorithms in classifying various types of anemia and other blood disorders; the Gradient Boosting algorithm recorded the highest accuracy at 99.19%, while Decision Tree trailed by achieving 98.38%. CatBoost and XGBoost achieved accuracy levels of 97.98% each, and Random Forest achieved satisfactory accuracy at 96.76%. Therefore, our results prove that, in the medical field, claims made about the efficiency of using 'ML' algorithms in the proper classification of blood disorders are hugely encouraging for early diagnosis and timely interventional treatment (Ramzan *et al.* 2024; Kitaw *et al.* 2024).

Machine learning and deep learning has revolutionized healthcare by offering solutions that help make diagnosis and treatment more efficient (Pacal 2024b; Pacal and Attallah 2025a; Aruk *et al.* 2025; İnce *et al.* 2025; Ince *et al.* 2025). The use of ML models concerning blood disorder classification is an important step into more

---

Manuscript received: 21 March 2025,
Revised: 5 July 2025,
Accepted: 9 July 2025.

[1] ygtcncakmak@gmail.com (**Corresponding author**)
[2] ishak.pacal@igdir.edu.tr

personalized medicine; algorithms can identify subtle patterns in medical data that might remain undetected by clinicians (Cakmak *et al.* 2024; Cakmak and Pacal 2025). Using large datasets and sophisticated algorithms, we can enhance the accuracy of diagnosing, reduce human error, and enhance patient outcomes (Ozdemir and Pacal 2025; Pacal 2024a). Our study highlights the significant role ML plays in the diagnosis of anemia and other blood disorders, thus emphasizing the need for continued research and practical adaptation of these technologies. With early diagnosis and precise categorization, ML-based models would help the medical authorities offer better patient care, save on health care expenses, and proffer better public health outcomes (Ozdemir *et al.* 2025; Pacal and Attallah 2025b).

LITERATURE REVIEW

Sanap *et al.* (2011) conducted a study utilizing data mining techniques to predict and classify anemia, a prevalent blood disorder that can be categorized based on red blood cell morphology and etiology. The researchers constructed a dataset from complete blood count (CBC) test data obtained from various hospitals and evaluated the performance of two classification algorithms: the C4.5 decision tree algorithm and Support Vector Machine (SVM), implemented as J48 and Sequential Minimal Optimization (SMO) respectively, within the Weka software environment. Through a series of experiments, they generated a decision tree that facilitates optimal classification of anemia types and assesses anemia severity based on CBC reports. Their findings indicate that the C4.5 algorithm exhibits superior performance, achieving the highest accuracy in anemia classification.

Jaiswal *et al.* (2018) highlight the importance of machine learning in transforming healthcare data into meaningful insights for disease prediction and decision-making. Using CBC data from pathology centers, they evaluate Naïve Bayes, Random Forest, and Decision Tree algorithms for anemia classification. Their findings show that Naïve Bayes outperforms the other models in accuracy. Meena *et al.* (2019) examine childhood anemia, highlighting the impact of maternal health and diet during pregnancy. Using NFHS-4 survey data, they develop a decision support system with data mining techniques to predict anemia and guide prevention through dietary recommendations. Comparing decision tree and association rule mining, the study identifies the most effective method for predictive modeling in healthcare.

Yıldız *et al.* (2021) propose a decision support system for anemia detection using machine learning. Trained on 1,663 patient records from a Turkish hospital, the model classifies 12 anemia types based on hemogram data and medical history. Among the four algorithms tested, Bagged Decision Trees achieved the highest accuracy (85.6%), followed by Boosted Trees (83.0%) and Artificial Neural Networks (79.6%). The system aims to assist medical professionals and students in diagnosis. Asare *et al.* (2023) reviews machine learning applications for anemia detection, highlighting their affordability and non-invasive nature compared to traditional methods. They analyze current trends, algorithms, evaluation metrics, and dataset characteristics. The study finds that machine learning techniques offer a practical and efficient solution for anemia diagnosis, particularly in resource-limited settings. The results demonstrate that these methods enable timely and accessible detection, supporting their feasibility for clinical use.

Kilicarslan *et al.* (2021) explore deep learning for anemia prediction, emphasizing the need for tailored models due to dataset variability. They propose two hybrid models, GA-SAE and GA-CNN, which optimize Stacked Autoencoder (SAE) and Convolutional Neural Network (CNN) hyperparameters using a genetic algorithm (GA). The study classifies HGB-anemia, nutritional anemia (iron, B12, and folate deficiency), and non-anemic cases. Evaluated using accuracy, F-score, precision, and sensitivity, GA-CNN outperforms existing methods, achieving 98.50% accuracy.

Aslan and Özüpak (2024) investigate the integration of deep learning and image processing techniques for the classification of white blood cells (WBCs). In their study, they employed Convolutional Neural Network (CNN) models, specifically ResNet-50, VGG19, and a model they proposed, as feature extractors to enhance WBC classification accuracy. These CNN models were synergistically combined with Ridge feature selection and Maximal Information Coefficient (MIC) techniques to identify the most salient features and expedite the classification process. Consequently, they reported achieving a 98.27% success rate in the categorization of white blood cells, demonstrating a considerable improvement in classification accuracy with their proposed CNN model.

Alpsalaz *et al.* (2025) address the critical challenge of maize leaf disease diagnosis by proposing a lightweight and interpretable convolutional neural network (CNN) model for accurate and efficient classification. Utilizing the 'Corn or Maize Leaf Disease Dataset,' their model classifies four disease categories Healthy, Gray Leaf Spot, Common Rust, and Northern Leaf Blight achieving 94.97% accuracy and a micro-average AUC of 0.99. Emphasizing practical deployment, the model's minimal parameter count (1.22 million) supports real-time inference on mobile devices, facilitating field applications. The authors employed data augmentation and transfer learning for robust generalization, and Explainable Artificial Intelligence (XAI) methods (LIME and SHAP) to enhance transparency by identifying disease-relevant features, with SHAP achieving an IoU of 0.82. The proposed model outperformed benchmark architectures, including ResNet50, MobileNetV2, and EfficientNetB0, in both accuracy and computational efficiency. Furthermore, the model demonstrated adaptability with only a 2.82% performance drop under simulated extreme environmental conditions. These findings highlight the model's potential as a reliable, fast, and explainable solution for precision agriculture, particularly in resource-constrained settings.

MATERIALS AND METHODS

Dataset and Preprocessing

Anemi Dataset: This study utilizes a dataset consisting of 1281 rows and 15 features to classify anemia and other blood disorders. However, 49 duplicate records were removed, and the models were trained on the remaining 1232 rows. The target classes in the dataset include Healthy (HLT), Iron deficiency anemia (IDA), Leukemia (LEU), Leukemia with thrombocytopenia (LwT), Macrocytic anemia (MAA), Normocytic hypochromic anemia (NHA), Normocytic normochromic anemia (NNA), Other microcytic anemia (OMA), and Thrombocytopenia (THP) (Kaggle 2025).

The aim of this research is to employ machine learning techniques to classify the different types of anemia. Anemia is a significant global health issue, and its early diagnosis and accurate classification are essential for effective treatment and management. Proper classification of these disorders has the potential to improve treatment efficacy and healthcare management, thus contributing to better patient outcomes. The distribution of classes in the dataset used in the study is shown in Figure 1.

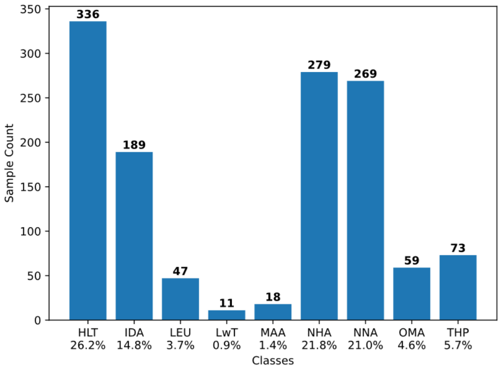

Figure 1 The distribution of classes in the dataset used in the study is shown

Characteristics and Classes of The Dataset: This study investigates key hematological features that play a vital role in the classification of anemia and related blood disorders. Each feature provides insights into specific aspects of blood cell composition and function. The characteristics and their corresponding attributes are summarized in Table 1.

Data Preprocessing Steps: The dataset underwent several preprocessing steps to enhance the quality and efficiency of the training process. Initially, missing values were handled by either removing the rows with null values or imputing them using statistical methods such as averaging or the median. Feature scaling was applied to ensure that all variables were within a comparable range, which is particularly important for models that rely on distance-based measurements. For categorical variables, encoding methods like one-hot or label encoding were used to convert them into numerical values. Since the target variable was already in a suitable format for classification, no logarithmic transformation was applied.

Additionally, duplicate records (49 instances) were identified and removed to avoid redundancy in the dataset. The dataset was then split into 80% for training and 20% for testing, as shown in Figure 2.

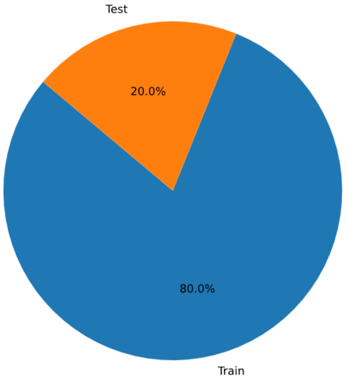

Figure 2 The dataset used in the study was divided into train 80% and test 20%

Algorithms Used

The machine learning algorithms employed in this study are detailed and explained in Table 2. These algorithms were carefully chosen for their proven effectiveness in classification tasks and their ability to handle complex, high-dimensional datasets. Each algorithm has unique characteristics that contribute to its performance in classifying anemia and other blood disorders. The selection process was based on the algorithms' strengths in terms of accuracy, efficiency, and ability to capture intricate patterns within the data. Below is a summary of each algorithm's approach and its relevance to the objectives of the study. Figure 3 shows some machine learning architectures.

The Decision Tree (DT) is a widely applied supervised learning algorithm for both classification and regression, constructing a tree-like model by recursively partitioning the dataset into smaller, more homogeneous subgroups based on feature values that maximize information gain or minimize impurity at each split (Muyama *et al.* 2024). Valued for their interpretability and efficiency with small to medium-sized datasets (Islam *et al.* 2024), DTs are prone to overfitting if allowed to grow excessively deep, learning noise and peculiarities from the training data which can hinder generalization. To mitigate this in this study, specific hyperparameter configurations were deliberately chosen: the max-depth of the tree was explicitly set to 6 to control complexity and prevent overfitting, while random-state was set to 42 to ensure the reproducibility of results. Although a more exhaustive hyperparameter search could explore other parameters like min-samples-split or criterion, constraining max-depth is a fundamental and effective strategy to balance the model's learning capacity with its ability to generalize to unseen data, a crucial aspect for reliable classification performance (Kasthuri *et al.* 2024).

Random Forest (RF) is a powerful ensemble learning algorithm that constructs a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. By training each tree on random subsets of both data samples (bagging) and features, RF aims to improve classification accuracy and robustness, effectively mitigating the overfitting tendency often seen in individual decision trees (Muyama *et al.* 2024).

This ensemble approach helps reduce both variance and bias, leading to more reliable and generalized predictions. RF is also known for its efficiency in handling high-dimensional data, its capability to manage missing values, and its scalability with large datasets. While RF excels at preventing overfitting and achieving high accuracy, its improved performance can come at the cost of increased computational resources and reduced interpretability compared to a single decision tree, due to the large number of trees involved.

In this study, to leverage these strengths while ensuring consistent results, the Random Forest classifier was configured with n-estimators set to 100, indicating that the forest was composed of 100 individual decision trees. The random-state was set to 42 to ensure the reproducibility of the model's training and subsequent predictions. While other hyperparameters such as max-depth for individual trees or min-samples-split could be further tuned, the choice of 100 estimators is a common and often effective starting point for balancing predictive power with computational load, aiming to harness the collective wisdom of the ensemble for robust classification.

Gradient Boosting (GB) is an ensemble learning technique that sequentially builds a strong predictive model from a collection of weak learners, typically decision trees. The core principle involves iteratively fitting new models to the residuals (errors) of the previous models, thereby progressively reducing bias and enhancing overall model accuracy. This additive training process allows GB to effectively model complex relationships in data for both regression

Table 1 Dataset features and descriptions

Features	Description
WBC	White Blood Cell count, indicating the number of white blood cells in the blood, used to assess immune function.
LYMp	Percentage of lymphocytes in the total white blood cell count, indicating immune response.
NEUTp	Percentage of neutrophils in the total white blood cell count, associated with bacterial infections.
LYMn	Absolute number of lymphocytes in the blood, providing a measure of immune function.
NEUTn	Absolute number of neutrophils in the blood, important in detecting bacterial infections.
RBC	Red Blood Cell count, reflecting the number of red blood cells that carry oxygen throughout the body.
HGB	Hemoglobin level in the blood, indicating the oxygen-carrying capacity of the blood.
HCT	Hematocrit, the proportion of blood volume occupied by red blood cells, used to diagnose anemia.
MCV	Mean Corpuscular Volume, representing the average size of red blood cells, important in anemia classification.
MCH	Mean Corpuscular Hemoglobin, indicating the average amount of hemoglobin per red blood cell.
MCHC	Mean Corpuscular Hemoglobin Concentration, measuring the average concentration of hemoglobin in red blood cells.
PLT	Platelet count, reflecting the number of platelets in the blood, crucial for blood clotting.
PDW	Platelet Distribution Width, measuring the variation in platelet size, which can indicate platelet disorders.
PCT	A procalcitonin test can help your health care provider diagnose if you have sepsis from a bacterial infection or if you have a high risk of developing sepsis.
Diagnosis	The target variable indicating the type of anemia or blood disorder diagnosis.

Table 2 Observed data for the DT algorithm

Class	Precision %	Recall %	F1-Score %	Support
HLT	98.0	98.0	98.0	64
IDA	100.0	100.0	100.0	44
LEU	91.0	100.0	95.0	10
LwT	80.0	80.0	80.0	5
MAA	100.0	100.0	100.0	4
NHA	100.0	98.0	99.0	49
NNA	98.0	100.0	99.0	52
OMA	100.0	86.0	92.0	7
THP	100.0	100.0	100.0	12
Macro avg	96.0	96.0	96.0	247
Weighted avg	96.0	98.0	98.0	247

and classification tasks across diverse data types and distributions. However, without careful tuning, GB can be susceptible to overfitting, particularly with a large number of iterations or excessive weight given to individual learners.

Regularization techniques, such as shrinkage (controlled by the learning rate) and subsampling (stochastic gradient boosting), are commonly employed to mitigate overfitting and improve model generalization (Mwangi *et al.* 2024). In this study, the Gradient Boosting Classifier was specifically configured with n-estimators (the number of boosting stages or trees) set to 100, a learning-rate of 0.1, and a max-depth of 6 for each individual tree. The random-state was set to 42 to ensure reproducibility. These hyperparameter settings were chosen to harness GB's predictive power: the n-estimators and learning-rate together control the complexity and learning speed of the ensemble, while max-depth limits the complexity of individual weak learners, collectively aiming for a robust model that accurately captures underlying data patterns without succumbing to overfitting.

Extreme Gradient Boosting (XGBoost), an acronym for Extreme Gradient Boosting, represents an advanced and optimized implementation of the gradient boosting framework, designed for superior computational efficiency, scalability, and predictive performance. It constructs an ensemble of decision trees sequentially, where each new tree aims to correct the errors made by the preceding ones (Kasthuri *et al.* 2024).

XGBoost distinguishes itself from traditional gradient boosting through several key enhancements, including built-in L1 (Lasso) and L2 (Ridge) regularization methods to combat overfitting and improve model generalization (Olatunji *et al.* 2024). Furthermore, it employs a more sophisticated optimization process utilizing second-order Taylor expansions (Hessians) in addition to first-order gradients, facilitating more precise model updates and potentially faster convergence.

Features such as parallel processing for tree construction and advanced tree pruning techniques contribute to its efficiency in terms of both time and computational resources, making it a highly popular choice in machine learning competitions and diverse real-world applications for classification, regression, and ranking tasks, often achieving state-of-the-art results (Mwangi *et al.* 2024). In this study, the XGBoost classifier was configured with n-estimators (number of boosting rounds) set to 100, a learning-rate of 0.1, and a max-depth of 6 for individual trees. To ensure compatibility and suppress warnings, use-label-encoder was set to False, and logloss was specified as the eval-metric. The random-state was fixed at 42 for reproducibility. These specific hyperparameter settings were chosen to effectively leverage XGBoost's capabilities, balancing the model's complexity (via max-depth and n-estimators) with its learning pace (learning-rate) to achieve robust and accurate classification performance.

Categorical Boosting (CatBoost) is an advanced gradient boosting algorithm renowned for its robust handling of categorical data, often without requiring extensive preprocessing, which can be time-consuming with other methods. Based on decision trees, CatBoost builds an ensemble by sequentially adding trees, where each new tree corrects the errors made by its predecessors.

A key distinguishing feature of CatBoost is its sophisticated internal processing of categorical features, employing techniques like ordered target statistics that implicitly encode them without distorting their inherent structure, thereby simplifying feature engineering and enhancing modeling efficiency, especially with high-cardinality categorical variables. Furthermore, CatBoost incorporates ordered boosting and strategies for growing symmetric trees, which help mitigate overfitting and improve model generalization. The algorithm also offers built-in support for missing data and often demonstrates competitive speed and accuracy compared to traditional gradient boosting implementations.

Due to its high performance and often requiring minimal tuning, it has gained widespread adoption in machine learning competitions and real-world applications (Mwangi *et al.* 2024). In this study, the CatBoost classifier was specifically configured with iterations (number of trees) set to 100, learning-rate to 0.1, and depth (maximum depth of the trees) to 6. The verbose parameter was set to 0 to suppress training output. These hyperparameter choices aim to harness CatBoost's strengths in a controlled manner, with the learning rate influencing the contribution of each tree, the number of iterations determining the ensemble size, and the depth controlling individual tree complexity, all contributing to a balance between model expressiveness and its ability to generalize effectively from the training data.

RESULTS AND DISCUSSION

Decision Trees (DT)

The DT algorithm performs impressively well in classifying problems, boasting of overall accuracy that stands at a high 98%. The model, for the most part, established quite a high precision and recall among its classes, where perfect classification results, that is, 100%, in precision and recall, were achieved for class 1. The precision and recall of other classes well within high thresholds allowed macro-and weighted average F1 scores to be computed at 96% and 98%, respectively. These scores indicate that the predictive model's precision and recall are balanced across all classes. While slightly poorer performance was seen across smaller classes, the ability of the vast bandwidth of instances to be classified denotes effectiveness by the model on the dataset. The performance results of the DT algorithm discussed in this paragraph predestine high potential in regard to accurately classifying distinct types of anemia and blood disorders and being a very helpful instrument in diagnostic tasks. The observed data to the DT algorithm are shown in Table 2, while the confusion matrix is shown in Figure 4.

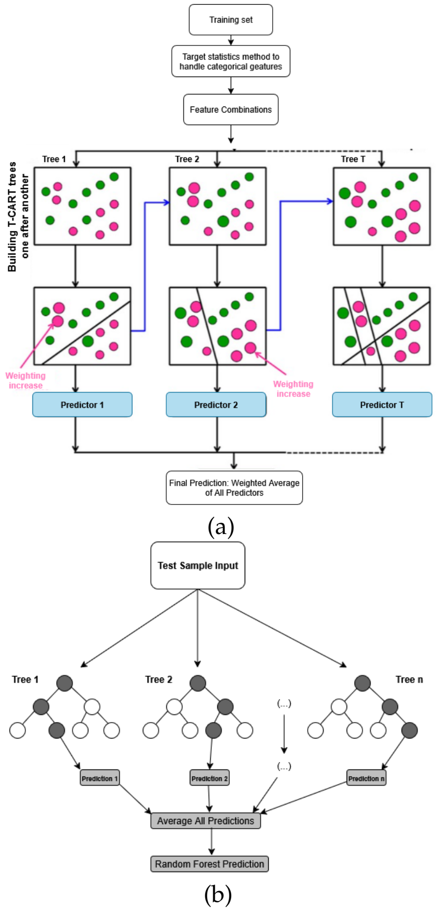

Figure 3 Models based on CatBoost and RF algorithms are provided in a and b, respectively (Pandey *et al.* 2023)

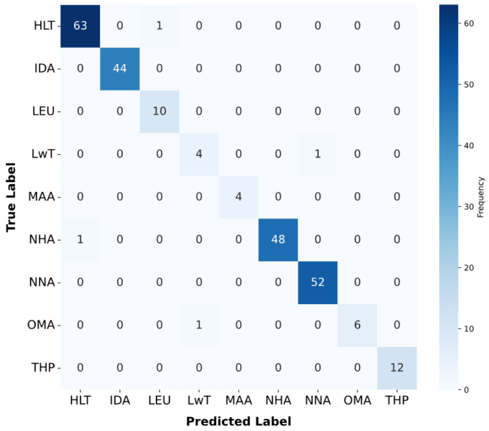

Figure 4 The Confusion Matrix obtained from the DT algorithm

Random Forest (RF)

The RF algorithm showed strong performance with an overall accuracy of 97%. Hence, this model is highly effective in classifying various types of anemia and blood disorders. In a class-wise breakdown, class 1 (Iron deficiency anemia) had both precision and recall and F1-score with a perfect score of 100%. Note that class 3 (Leukemia with thrombocytopenia) and class 4 (Macrocytic anemia) had lower recall and F1 scores at 40% and 50% respectively. Yet the general performance of the model is solid, highlighted by a macro-average and a weighted average F1 score of 89% and 96% respectively. These results indicate the RF algorithm to be effective in providing accurate classification for the majority of the classes, leaving however an avenue for improvement on the management of certain lesser-represented categories. The observed data for RF algorithm is shown in Table 3 and the confusion matrix is shown in Figure 5.

Table 3 Observed data for the RF algorithm

Class	Precision %	Recall %	F1-Score %	Support
HLT	97.0	100.0	98.0	64
IDA	100.0	100.0	100.0	44
LEU	90.0	90.0	90.0	10
LwT	100.0	40.0	57.0	5
MAA	100.0	50.0	67.0	4
NHA	98.0	96.0	97.0	49
NNA	96.0	100.0	98.0	52
OMA	100.0	100.0	100.0	7
THP	86.0	100.0	92.0	12
Macro avg	96.0	86.0	89.0	247
Weighted avg	97.0	97.0	96.0	247

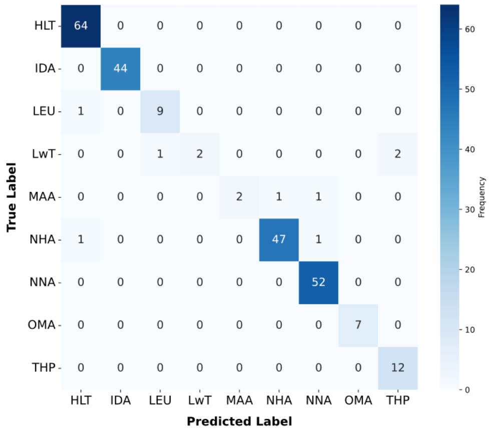

Figure 5 The Confusion Matrix obtained from the RF algorithm

Categorical Boosting (CatBoost)

The CatBoost algorithm showed excellent classification performance and achieved an overall accuracy of 98%, which means that the model is potentially very good at attacking various types of anemia and disorders of blood. Class 1, Iron deficiency anemia, had perfect precision, recall, and F1-score, indicating a very strong model for this class. Class 3, Leukemia with thrombocytopenia, had support for recall, where 80% of instances which belong to this class were, in fact, detected by the model. Class 4, Macrocytic anemia, had its complementary value for recall as 75%, because, although it was lower than the other classes, the model was able to classify a very good proportion of the instances for this class. Overall, the CatBoost model was efficient for all classes, which was revealed by high macro and weighted average F1-scores of 96% and 98%, respectively. These results confirm the effectiveness of the CatBoost algorithm in its classification task. The observed data for the CatBoost algorithm is displayed in Table 4, whilst Figure 6 illustrates the confusion matrix.

Table 4 Observed data for the CatBoost algorithm

Class	Precision %	Recall %	F1-Score %	Support
HLT	98.0	100.0	99.0	64
IDA	100.0	100.0	100.0	44
LEU	91.0	100.0	95.0	10
LwT	100.0	80.0	89.0	5
MAA	100.0	75.0	86.0	4
NHA	100.0	94.0	97.0	49
NNA	95.0	100.0	97.0	52
OMA	100.0	100.0	100.0	7
THP	100.0	100.0	92.0	12
Macro avg	98.0	86.0	96.0	247
Weighted avg	98.0	98.0	98.0	247

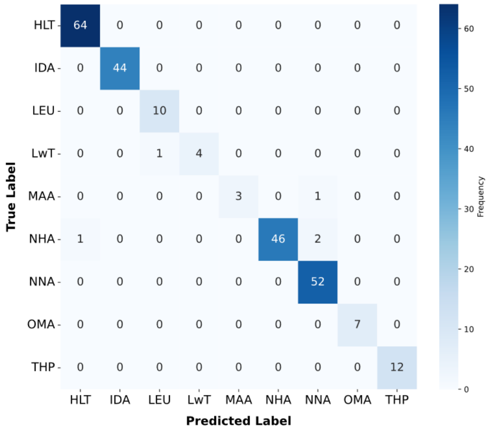

Figure 6 The Confusion Matrix obtained from the CatBoost algorithm

Gradient Boosting (GB)

This GB model incredibly has done well with an analysis of 99%, which demonstrated its superior performance. Some classes specifically two and three for leukemia and leukemia with thrombocytopenia, respectively were scored absolutely perfect for both precision, recall, and F1, so the model successfully scored these classes. Class four, Macrocytic Anemia, also demonstrated perfect classification. Class seven, other microcytic anemia, would have a recall of 86% yet has still got some decent performance as shown with an F1 score of 92%. The macro average of F1 score is further enhanced by an overall score of 99%, while a weighted average F1 score of 99% underscores the model's classification quality. These results demonstrate the great advance provided by the GB algorithm in this task, with very consistent and reliable performance across all classes. The results of GB and some data for confusion matrices are shown in Table 5 and Figure 7, respectively.

Table 5 Observed data for the GB algorithm

Class	Precision %	Recall %	F1-Score %	Support
HLT	98.0	100.0	99.0	64
IDA	100.0	100.0	100.0	44
LEU	100.0	100.0	100.0	10
LwT	100.0	100.0	100.0	5
MAA	100.0	100.0	100.0	4
NHA	98.0	98.0	98.0	49
NNA	100.0	100.0	100.0	52
OMA	100.0	86.0	92.0	7
THP	100.0	100.0	100.0	12
Macro avg	100.0	98.0	99.0	247
Weighted avg	99.0	99.0	99.0	247

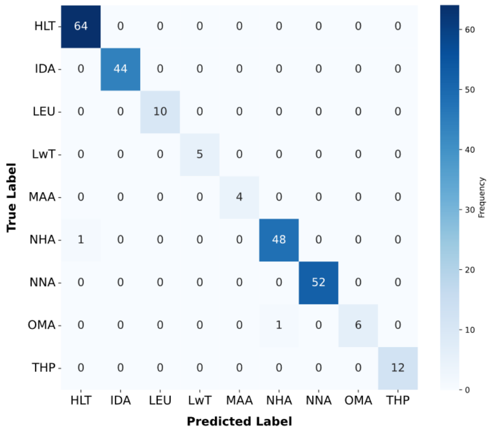

Figure 7 The Confusion Matrix obtained from the CatBoost algorithm

Extreme Gradient Boosting (XGBoost)

The XGBoost algorithm performed very well as well, getting an accuracy level of 98%. The model's precision and recall scores were outstanding for several classes, including class 1 (i.e., Iron deficiency anemia) and class 6 (i.e., Normocytic normochromic anemia), both of which showed perfect values of 100% precision, recall, and F1-scores. Class 2 demonstrated good scoring with precision metrics of 91% and perfect recall measures. However, class 3 recorded lower recalls: it was around 40%, which affected its F1-score, to about 57%, still. Nevertheless, the overall macro and weighted averages of 93% F1-value and 98% F1-value, respectively, indicate that the XGBoost model has performed well in the classification task. Therefore, it can be inferred that XGBoost tends to be an exceptionally capable algorithm for anemia classification, with lots of room for improvement for some classes. Observed data for the XGBoost algorithm is presented in Table 6, while the confusion matrix is presented in Figure 8.

Table 6 Observed data for the XGBoost algorithm

Class	Precision %	Recall %	F1-Score %	Support
HLT	98.0	100.0	99.0	64
IDA	100.0	100.0	100.0	44
LEU	91.0	100.0	95.0	10
LwT	100.0	40.0	57.0	5
MAA	100.0	100.0	100.0	4
NHA	98.0	98.0	98.0	49
NNA	100.0	100.0	100.0	52
OMA	100.0	86.0	92.0	7
THP	86.0	100.0	92.0	12
Macro avg	97.0	92.0	93.0	247
Weighted avg	98.0	98.0	98.0	247

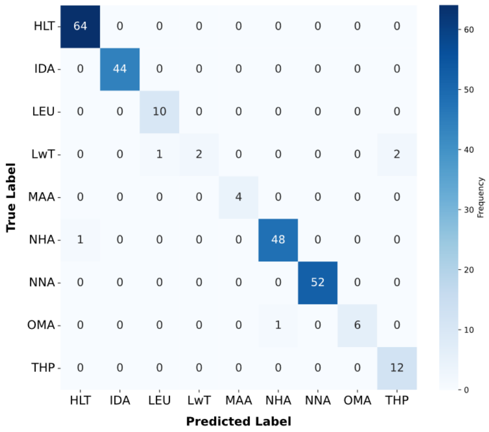

Figure 8 The Confusion Matrix obtained from the XGBoost algorithm

Feature Importance Analysis

The bar chart presented illustrates the feature importance scores determined by the GB model. In this case, the importance is derived from the magnitude of the absolute values of the model's coefficients, reflecting the degree to which each feature influences the classification decision (Azimjonov and Kim 2024). "HGB" emerges as the most significant feature, followed by "MCH" and "MCHC," with higher coefficient magnitudes indicating their critical role in distinguishing between the classes. Other features like "MCV," "PLT," and "WBC" also contribute meaningfully to the model's predictions. Conversely, features lower in the chart, such as "PDW," "NEUTp," and "LYMn," hold relatively less importance and thus have a reduced impact on the decision boundary. Overall, the feature importance analysis emphasizes the central role of HGB, MCH, and MCHC-based attributes in the classification process, underscoring their relevance in hematological analysis and the diagnosis of anemia and related blood disorders. Figure 9 presents the feature importance plot for the GB algorithm (Got *et al.* 2024).

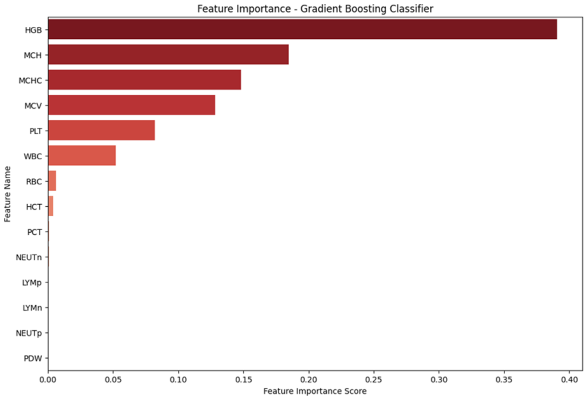

Figure 9 Illustrates the feature importance plot for the GB algorithm, highlighting the relative significance of each feature in the model's classification process

DISCUSSION

This section introduces an autonomous diagnostic approach based on machine learning for the diagnosis of anemia, a prevalent and clinically significant blood disorder. In this study, several commonly used machine learning algorithms, selected from the existing literature for their proven efficacy in similar classification tasks, were trained and tested on anemia-related data. The primary aim was to identify the most accurate model for distinguishing anemic cases. The experimental results, specifically the accuracy achieved by each algorithm, are systematically presented in Table 7 below. This table offers a direct comparison of the predictive capabilities of the tested models when applied to the Anemi dataset.

Table 7 The experimental results of the ML algorithms tested on the Anemi dataset

Rank	Models	Accuracy %
1	Gradient Boosting	99.19
2	Decision Tree	98.38
3	CatBoost	97.98
4	XGBoost	97.98
5	Random Forest	96.76

As evident from Table 8, the GB model stands out as the top-performing algorithm, achieving an impressive accuracy of 99.19%. Such exceptional performance underscores the model's capacity to iteratively correct its mistakes and produce highly refined predictions, making it particularly well-suited for complex classification tasks like anemia diagnosis and suggesting its strong potential as a core component in an automated anemia screening tool. Other models also demonstrated commendable performance. The DT algorithm achieved a high accuracy of 98.38%, often effective in capturing interactions among input features. CatBoost, with an accuracy of 97.98%, also performed superbly, its strength lying in its optimized handling of categorical features combined with gradient boosting techniques. XGBoost achieved an identical accuracy of 97.98%, while RF followed closely with 96.76% (Kasthuri et al., 2024). While these models also exhibited strong predictive power, they did not quite match the benchmark set by GB. Nevertheless, both XGBoost and RF are robust ensemble methods, known for their tree-based nature and capabilities in managing feature interactions, with RF, in particular, mitigating overfitting by averaging predictions from multiple decision trees.

It is crucial to acknowledge that model performance is significantly dependent on factors such as feature selection, data preprocessing, and meticulous hyperparameter tuning; further fine-tuning could potentially lead to even greater performance. Beyond predictive accuracy, computational efficiency is an important consideration for real-world clinical applications. While GB yielded the highest accuracy, it can be computationally more intensive. In contrast, algorithms like RF and XGBoost might offer better scalability. Future work could explore advanced feature engineering, hybrid models, and leveraging deep learning architectures. Crucially, incorporating explainability methods such as SHAP values or LIME will be vital for providing transparency into model decisions, thereby fostering trust and facilitating the adoption of AI-assisted anemia diagnostic systems in clinical practice (Awe *et al.* 2024).

CONCLUSION

This research compellingly demonstrates the transformative potential of machine learning algorithms in enhancing the early detection and precise classification of anemia and other hematological disorders, conditions often overlooked until they manifest severe consequences. Through the application of classical machine learning models, including DT, RF, CatBoost, GB, and XGBoost, on key hematological parameters such as red RBC, WBC, hemoglobin (HGB), and platelet count (PLT), this study identified Gradient Boosting as a particularly robust model, achieving an impressive accuracy of 99.19%. This high level of accuracy underscores the model's capacity to capture intricate, complex patterns within patient datasets, thereby enabling highly precise predictions and showcasing significant promises for medical diagnostics.

The findings emphasize how AI-powered diagnostic approaches can significantly augment clinical decision-making processes by facilitating faster, more reliable diagnoses and potentially reducing diagnostic errors, which is especially critical for conditions like anemia where timely intervention can substantially improve patient outcomes and mitigate complications. The widespread adoption of such autonomous diagnostic tools could also alleviate the burden on healthcare systems by expediting treatment planning.

Looking ahead, the continued refinement of these algorithms and the integration of more advanced techniques, such as deep learning or hybrid models, are anticipated to further elevate diagnostic accuracy and enhance the scalability of these solutions in clinical workflows. A crucial aspect of future development will be the implementation of explainable AI (XAI) methodologies, like SHAP or LIME. These techniques are vital for fostering greater trust and acceptance among healthcare professionals by providing transparency into the decision-making processes of the models, thereby supporting their integration into routine clinical practice. Ultimately, the progression of machine learning in the diagnosis and management of anemia and other blood disorders paves the way for more personalized, data-driven medical solutions, holding immense promise for improving patient care, optimizing healthcare delivery, and offering deeper insights into complex hematological conditions, moving us closer to the paradigm of precision medicine.

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

LITERATURE CITED

Alpsalaz, F., Y. Özüpak, E. Aslan, and H. Uzel, 2025 Classification of maize leaf diseases with deep learning: Performance evaluation of the proposed model and use of explicable artificial intelligence. Chemometrics and Intelligent Laboratory Systems p. 105412.

Aruk, I., I. Pacal, and A. N. Toprak, 2025 A novel hybrid convnext-based approach for enhanced skin lesion classification. Expert Systems with Applications p. 127721.

Asare, J. W., P. Appiahene, and E. T. Donkoh, 2023 Detection of anaemia using medical images: A comparative study of machine learning algorithms–a systematic literature review. Informatics in Medicine Unlocked **40**: 101283.

Aslan, E. and Y. Özüpak, 2024 Classification of blood cells with convolutional neural network model. Bitlis Eren Üniversitesi Fen Bilimleri Dergisi **13**: 314–326.

Awe, O. O., J. M. Adepoju, E. Boniface, and O. D. Awe, 2024 Comparative analysis of random forest and neural networks for anemia prediction in female adolescents: A lime-based explainability approach. In *Practical Statistical Learning and Data Science Methods: Case Studies from LISA 2020 Global Network, USA*, pp. 555–573, Springer.

Azimjonov, J. and T. Kim, 2024 Designing accurate lightweight intrusion detection systems for iot networks using fine-tuned linear svm and feature selectors. Computers & Security **137**: 103598.

Cakmak, Y. and I. Pacal, 2025 Enhancing breast cancer diagnosis: A comparative evaluation of machine learning algorithms using the wisconsin dataset. Journal of Operations Intelligence **3**: 175–196.

Cakmak, Y., S. Safak, M. A. Bayram, and I. Pacal, 2024 Comprehensive evaluation of machine learning and ann models for breast cancer detection. Computer and Decision Making: An International Journal **1**: 84–102.

Fentie, K., T. Wakayo, and G. Gizaw, 2020 Prevalence of anemia and associated factors among secondary school adolescent girls in jimma town, oromia regional state, southwest ethiopia. Anemia **2020**: 5043646.

Got, A., D. Zouache, A. Moussaoui, L. Abualigah, and A. Alsayat, 2024 Improved manta ray foraging optimizer-based svm for feature selection problems: a medical case study. Journal of Bionic Engineering **21**: 409–425.

Ince, S., I. Kunduracioglu, A. Algarni, B. Bayram, and I. Pacal, 2025 Deep learning for cerebral vascular occlusion segmentation: a novel convnextv2 and grn-integrated u-net framework for diffusion-weighted imaging. Neuroscience **574**: 42–53.

İnce, S., I. Kunduracioglu, B. Bayram, and I. Pacal, 2025 U-net-based models for precise brain stroke segmentation. Chaos Theory and Applications **7**: 50–60.

Islam, R., S. Tanweer, M. T. Nafis, I. Hussain, and O. Ahmad, 2024 Intelligent diagnosis of sickle cell anemia in chronic diseases through a machine learning predictive system. In *International Conference on ICT for Digital, Smart, and Sustainable Development*, pp. 109–124, Springer.

Jaiswal, M., A. Srivastava, and T. J. Siddiqui, 2018 Machine learning algorithms for anemia disease prediction. In *Recent trends in communication, computing, and electronics: Select proceedings of IC3E 2018*, pp. 463–469, Springer.

Kaggle, 2025 Anemia types classification.

Karra, M. L., E. A. Rahiman, M. Narahari, A. V. Bhongir, G. Rathod, *et al.*, 2025 Unveiling the burden of sickle cell anemia: A pilot study validating dried blood spots for newborn screening. Indian Journal of Pediatrics **92**: 405–408.

Kasthuri, E., S. Subbulakshmi, and R. Sreedharan, 2024 Insightful clinical assistance for anemia prediction with data analysis and explainable ai. Procedia computer science **233**: 45–55.

Kilicarslan, S., M. Celik, and Ş. Sahin, 2021 Hybrid models based on genetic algorithm and deep learning algorithms for nutritional anemia disease classification. Biomedical Signal Processing and Control **63**: 102231.

Kitaw, B., C. Asefa, and F. Legese, 2024 Leveraging machine learning models for anemia severity detection among pregnant women following anc: Ethiopian context. BMC Public Health **24**: 3500.

Krieg, S., S. Loosen, A. Krieg, T. Luedde, C. Roderburg, *et al.*, 2024 Association between iron deficiency anemia and subsequent stomach and colorectal cancer diagnosis in germany. Journal of Cancer Research and Clinical Oncology **150**: 53.

Li, C., X. Shi, S. Chen, X. Peng, and S. Zong, 2025 Novel mechanistic insights into the comorbidity of anemia and rheumatoid arthritis: Identification of therapeutic targets. Molecular Immunology **180**: 74–85.

Link, H., M. Kerkmann, L. Holtmann, and M. Detzner, 2024 Anemia diagnosis and therapy in malignant diseases: implementation of guidelines—a representative study. Supportive Care in Cancer **32**: 113.

Malak, M. Z., A. Shehadeh, A. Ayed, and E. Alshawish, 2025 Predictors of anemia among infants at the age of one year attending health centers in the west bank/palestine: a retrospective study. BMC Public Health **25**: 179.

Meena, K., D. K. Tayal, V. Gupta, and A. Fatima, 2019 Using classification techniques for statistical analysis of anemia. Artificial intelligence in medicine **94**: 138–152.

Muyama, L., A. Neuraz, and A. Coulet, 2024 Deep reinforcement learning for personalized diagnostic decision pathways using electronic health records: A comparative study on anemia and systemic lupus erythematosus. Artificial Intelligence in Medicine **157**: 102994.

Mwangi, P., S. Kotva, and O. O. Awe, 2024 Explainable ai models for improved disease prediction. In *Practical Statistical Learning and Data Science Methods: Case Studies from LISA 2020 Global Network, USA*, pp. 73–109, Springer.

Olatunji, S. O., M. A. A. Khan, F. Alanazi, R. Yaanallah, S. Alghamdi, *et al.*, 2024 Machine learning-based models for the preemptive diagnosis of sickle cell anemia using clinical data. In *Finance and Law in the Metaverse World: Regulation and Financial Innovation in the Virtual World*, pp. 101–112, Springer.

Ozdemir, B., E. Aslan, and I. Pacal, 2025 Attention enhanced inceptionnext based hybrid deep learning model for lung cancer detection. IEEE Access .

Ozdemir, B. and I. Pacal, 2025 A robust deep learning framework for multiclass skin cancer classification. Scientific Reports **15**: 4938.

Pacal, I., 2024a Maxcervixt: A novel lightweight vision transformer-based approach for precise cervical cancer detection. Knowledge-Based Systems **289**: 111482.

Pacal, I., 2024b A novel swin transformer approach utilizing residual multi-layer perceptron for diagnosing brain tumors in mri images. International Journal of Machine Learning and Cybernetics **15**: 3579–3597.

Pacal, I. and O. Attallah, 2025a Hybrid deep learning model for automated colorectal cancer detection using local and global feature extraction. Knowledge-Based Systems p. 113625.

Pacal, I. and O. Attallah, 2025b Inceptionnext-transformer: A novel multi-scale deep feature learning architecture for multimodal breast cancer diagnosis. Biomedical Signal Processing and Control **110**: 108116.

Pandey, M., M. Karbasi, M. Jamei, A. Malik, and J. H. Pu, 2023 A comprehensive experimental and computational investigation on estimation of scour depth at bridge abutment: Emerging ensemble intelligent systems. Water Resources Management **37**: 3745–3767.

Ramzan, M., J. Sheng, M. U. Saeed, B. Wang, and F. Z. Duraihem, 2024 Revolutionizing anemia detection: integrative machine learning models and advanced attention mechanisms. Visual Computing for Industry, Biomedicine, and Art **7**: 18.

Sanap, S. A., M. Nagori, and V. Kshirsagar, 2011 Classification of anemia using data mining techniques. In *International conference on swarm, evolutionary, and memetic computing*, pp. 113–121, Springer.

Subba, S. S. and S. Araveti, 2025 Knowledge, attitudes, and practices related to iron deficiency anemia and probiotics among adolescent girls in anantapur, india: A qidap-guided cross-sectional study. Food and Humanity **4**: 100551.

Yıldız, T. K., N. Yurtay, and B. Öneç, 2021 Classifying anemia types using artificial learning methods. Engineering Science and Technology, an International Journal **24**: 50–70.

Yoshida, N., 2024 Recent advances in the diagnosis and treatment of pediatric acquired aplastic anemia. International journal of hematology **119**: 240–247.

How to cite this article: Cakmak, Y., and Pacal, I. AI-Driven Classification of Anemia and Blood Disorders Using Machine Learning Models. *Computers and Electronics in Medicine*, 2(2), 43-52, 2025.

Licensing Policy: The published articles in CEM are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.