Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 8, No. 1, pp. 105-114, February 2026

RESEARCH PAPER OPEN ACCES

Comparison Between K-Fold Cross Validation And Percentage Split In Decision Tree Algorithms For Anemia Classification

Nanda Putri Rahmawati[1], Irwan Budiman[2], Muhammad Itqan Mazdadi[3], Andi Farmadi[4], and Friska Abadi[5]

Department of Computer Science, Faculty of Mathematics and Natural Science, Lambung Mangkurat University, Banjarbaru, Indonesia

ABSTRACT

Anemia is a significant global health challenge characterized by a pathological deficit in hemoglobin concentration, often leading to physiological instability. Accurate clinical diagnosis typically relies on complete blood count (CBC) tests, which provide critical hematological parameters for classification. While machine learning models have demonstrated high efficacy in diagnosing anemia, existing research often relies on static data partitioning strategies that may overlook evaluation reliability and performance stability. This study addresses this gap by shifting the focus from architectural benchmarking to validation robustness, specifically evaluating the C4.5 algorithm's performance across different data-splitting techniques. The research uses a dataset comprising 1,281 clinical records with 14 numerical features and 9 anemia-type labels. To assess stability, two distinct partitioning strategies were implemented: a static Percentage Split (ranging from 60:40 to 90:10) and iterative K-Fold Cross Validation (with K values of 3, 5, 7, 10, and 15). Experimental results demonstrate that the C4.5 algorithm achieved its peak performance with the 90:10 Percentage Split, achieving an average accuracy of 99.46%, precision of 98.32%, and recall of 99.28%. In comparison, the K-Fold (K=10) approach yielded a slightly lower but more stable accuracy of 99.19% with a significantly reduced standard deviation (±0.09), highlighting its reliability for clinical applications. While the high-ratio percentage split maximizes training exposure and predictive potential, the K-Fold method provides a more objective, generalizable benchmark by accounting for the entire data distribution. The study further identifies challenges in classifying minority classes, such as Leukemia with thrombocytopenia, due to inherent data scarcity. Ultimately, this research confirms that the C4.5 algorithm, when paired with an optimal partitioning protocol, remains a robust and highly interpretable solution for clinical anemia screening, outperforming several complex modern architectures

Paper History

Received Dec 26, 2025
Revised Feb. 15, 2026
Accepted Feb. 20, 2026
Published Feb. 27, 2026

Keywords

Anemia;
Decision Tree;
K-Fold Cross Validation;
Percentage Split;
C4.5 Algorithm

Author Email

nandaputri15406@gmail.com
irwan.budiman@ulm.ac.id
mazdadi@ulm.ac.id
andifarmadi@ulm.ac.id
friska.abadi@ulm.ac.id

I. Introduction

Anemia is a global health problem clinically recognized as a pathological deficit in hemoglobin concentration, a condition that fundamentally compromises physiological stability [1]. The most common symptoms include fatigue, palpitations, headaches, shortness of breath, and conjunctival and palmar pallor. While these symptoms serve as a preliminary indication, they are not conclusive. They function as navigational instruments [2]. Anemia globally has a multifactorial origin; that is, there are numerous reasons for where the disease may arise, with iron deficiency being the most common cause in developing countries. It is estimated that 30% of women and 40% of children below five years of age globally are affected by this illness, which is linked to inadequate nutrition resulting from poverty [3]. Anemia can be categorized based on its origin. For instance, anemia may be microcytic due to iron deficiency, normocytic due to inflammation, or macrocytic due to vitamin B12 deficiency. Traditional techniques for identifying anemia demonstrate considerable obstacles for the least advantaged groups. The unavailability of quality healthcare access services, along with the expenses of laboratory tests, hinders numerous individuals from obtaining prompt diagnoses [4].

Many studies have been conducted on anemia diseases [5], [6], [7]. Anemia is generally diagnosed through a complete blood count (CBC) test that assesses red blood cell hemoglobin, hematocrit, and erythrocyte index levels [8]. The quantity of red blood cells must be sufficient to transport oxygen throughout the body. Any decrease in this cell count leads to a decline in oxygen levels, and occasionally, undiagnosed instances result in anemia [9]. The dataset from the CBC test contains 1281 observations, 14 predictors, and the diagnosis as a categorical target variable [10].

In previous studies, the decision tree, Naïve Bayes, and Support Vector Machine (SVM) approaches were used with two testing techniques, namely percentage split

Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315

(80% training data and 20% test data) and K-Fold Cross Validation (K=10) [6]. Another researcher using the same dataset with many classification model such as Ensemble Learning, Tree, SVM, Efficient Linear, Neural Network, KNN, and Naïve Bayes, only reaches the highest accuracy with Ensemble learning, achieving 99.22% accuracy, using a 90:10 split data [10]. Prior studies from [11], gaining 99.21% accuracy using Extreme Machine Learning with 190 data points and a 67:33 ratio of splitting data. The Latest studies from [12] Using the same dataset, achieved the highest accuracy of 99.00% with the bagging model as the classification model. The other new research with the same dataset using classification model Gradient Boosting, Decision Tree, CatBoost, XGBoost, and Random Forest, with the highest accuracy 99.19% achieved by Gradient Boosting using an 80:20 ratio data split [13].

However, it is crucial to note that the high accuracies reported in these previous studies are largely driven by the use of complex ensemble architectures (e.g., Boosting, Bagging) combined with specific, often aggressive, data splits (90:10) that maximize training exposure. While effective for benchmarking model capabilities, this reliance on static percentage splits often overlooks the reliability and variability of the partitioning strategy. Such static methods may fail to adequately capture the data variability encountered in real-world scenarios. Consequently, this research aims to address this gap by shifting the focus from architectural benchmarking to validation robustness. In this context, the C4.5 algorithm was specifically selected because, as noted by [14]It demonstrates outstanding performance on medical datasets and is highly interpretable. This high level of interpretability and transparent rule-based logic are vital for clinical anemia screening. By using a stable interpretable baseline, this study provides a clearer, systematic comparison of how various data partitioning strategies affect performance stability, free from the confounding effects of model complexity.

This research evaluates the reliability of anemia classification using the C4.5 algorithm by comparing two distinct data partitioning strategies: a static Percentage Split and an iterative K-Fold Cross Validation. While static partitioning is a common practical approach, it often carries a risk of selection bias. In contrast, K-Fold Cross Validation is implemented to provide a more stable and representative performance estimate by minimizing variance across the dataset. By systematically contrasting these methods, this study aims to establish a robust evaluation protocol that ensures both accuracy and stability for clinical anemia diagnosis, shifting the focus from mere algorithm benchmarking to the reliability of model evaluationConsequently, the main aim of this research is to assess the effectiveness of C4.5 using various data splitting methods. This study will contrast the data splitting technique of percentage split with k-fold cross-validation.

This research makes several key contributions to the anemia classification. Firstly, to investigates the specific impact of data partitioning strategies (K-Fold vs. Percentage Split) on Anemia Classification, shifting the focus from algorithm comparison to evaluation reliability. Secondly, provides empirical evidence on whether K-Fold Cross Validation minimizes the systematic errors and bias often found in static train-test splits within medical datasets. Third contribution, extends the comparative validation framework previously successful in medical research to the anemia domain to test the robustness of Decision Tree algorithms, Lastly, determines the most reliable evaluation protocol to maximize diagnostic accuracy and stability for anemia prediction models.

The structure of this paper is outlined as follows: Section 2 examines the dataset utilized, the suggested training and testing split strategies, and the proposed approach. Section 3 shows the outcomes of C4.5 using the training testing split schemes. Section 4 explores the interpretation of the results, compares them with other research, and addresses the limitations of the method used. Lastly, Section 5 wraps up the study by reiterating the aim, encapsulating the key discovery, and outlining prospects for future research.

II. Materials And Method

A. Dataset

The dataset used in this study consisted of 1,281 rows of data with 15 attributes, comprising 14 numerical features from complete blood count (CBC) tests and 1 target column containing anemia type labels. Each feature represents a hematological parameter, such as hemoglobin level, hematocrit, erythrocyte count, erythrocyte index, and other blood components that play an important for diagnosing anemia. The target column contains classifications of 9 anemia types, including normocytic hypochromic anemia, iron deficiency anemia, other microcytic anemia, leukemia, healthy, thrombocytopenia, normocytic normochromic anemia, leukemia with thrombocytopenia, and lastly macrocytic anemia [15]. With these characteristics, this dataset can be used to train and test classification algorithms to identify patterns and distinguish types of anemia based on patient data.

Table 1. The 9 anemia types in the dataset are categorized by diagnosis

Anemia Types	Data Records
Healthy	336
Iron deficiency anemia	189
Leukemia	47
Leukemia with thrombocytopenia	11
Macrocytic anemia	18
Normocytic hypochromic anemia	279
Normocytic normochromic anemia	269
Other microcytic anemia	59
Thrombocytopenia	73

Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315

Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 8, No. 1, pp. 105-114, February 2026

Table 2. 14 numerical features from complete blood count (CBC) test in the dataset this research use

No	WBC	LYMp	NEUTp	LYMn	NEUTn	RBC	HGB	HCT	MCV	MCH	MCHC	PLT	PDW	PCT
1	10	43.2	50.1	4.3	5	2.77	7.3	24.2	87.7	26.3	30.1	189	12.5	0.17
2	10	42.4	52.3	4.2	5.3	2.84	7.3	25	88.2	25.7	20.2	180	12.5	0.16
3	7.2	30.7	60.7	2.2	4.4	3.97	9	30.5	77	22.6	29.5	148	14.3	0.14
4	6	30.2	63.5	1.8	3.8	4.22	3.8	32.8	77.9	23.2	29.8	143	11.3	0.12
5	4.2	39.1	53.7	1.6	2.3	3.93	0.4	316	80.6	23.9	29.7	236	12.8	0.22
…..	…..	…..	…..	…..	…..	…..	…..	…..	…..	…..	…..	…..	…..	…..
1277	4.4	25.845	77.511	1.881	5.141	4.86	13.5	46.153	80.7	27.7	34.4	180	14.312	4.4
1278	5.6	25.845	77.511	1.881	5.141	4.85	15	46.153	91.7	31	33.8	215	14.312	5.6
1279	9.2	25.845	77.511	1.881	5.141	4.47	13.1	46.153	88.7	29.3	33	329	14.312	9.2
1280	6.48	25.845	77.511	1.881	5.141	4.75	13.2	46.153	86.7	27.9	32.1	174	14.312	6.48
1281	8.8	25.845	77.511	1.881	5.141	4.95	15.2	46.153	89.7	30.6	34.2	279	14.312	8.8

Table 3. 14 numerical features description which used in this research

No	Features
1	WBC (White Blood Cells)
2	LYMp (Lymphocyte Percentage)
3	NEUTp (Neutrophil Percentage)
4	LYMn (Lymphocyte Number)
5	NEUTn (Neutrophil Number)
6	RBC (Red Blood Cells)
7	HGB (Hemoglobin)
8	HCT (Hematocrit)
9	MCV (Mean Corpuscular Volume)
10	MCH (Mean Corpuscular Hemoglobin)
11	MCHC (Mean Corpuscular Hemoglobin Concentration)
12	PLT (Platelets)
13	PDW (Platelet Distribution Width)
14	PCT (Plateletcrit)

Table 1. displays the allocation of data points based on the target variable. The dataset shows more observations in certain categories and fewer in others. The Healthy category has 336 data, the Normocytic hypochromic anemia category has 279 data, and the Normocytic normochromic anemia category has 269 data; however, the Macrocytic anemia category has only 18 data, while the Leukemia with thrombocytopenia category includes 11 data. Table 3. describes the 14 numerical features in the dataset, while Table 2. shows the dataset, which contains 14 numerical features from the complete blood count (CBC) test.

B. K-Fold Cross Validation

In this research, the K-Fold Cross Validation method is used as a model evaluation technique to ensure more stable and reliable prediction results. Previous studies indicate that cross-validation performs better than other accuracy estimation techniques and improves performance by lowering bias and variance [16]. K-Fold Cross Validation is a method used to assess MLM. In this method, the data is split into k folds. In each iteration, one fold serves as the testing data, while the other folds act as training data [17]. Thus, the process is repeated until the entire dataset is evaluated. The value of K=3,5,7,10,15 was selected due to the high percentage of machine learning practitioners who use this value, as supported by [18], [19].

C. Percentage Split

Percentage Split, often referred to as train-test split, is used as a model evaluation technique by dividing the dataset into two main parts, namely training data and testing data, in certain proportions, for example, 70:30 or 80:20 [16]. The ratio used in this study is systematically conducted by dividing the dataset into several proportions of training and testing data, specifically 60:40, 67:33, 70:30, 80:20, and 90:10. This approach is implemented to identify the most optimal splitting ratio for the specific characteristics of the dataset, as dataset splitting is considered an indispensable practice to reduce bias and prevent overfitting. As explained in [20], the determination of the ideal ratio should be based on the total number of instances within the dataset to ensure an effective mapping of inputs that produce good and effective outputs. Furthermore, the literature suggests that although certain ratios like 70:30 are established as common practice, there is no single definitive rule for selecting the training/test splitting ratio. Therefore, using the various test set structures proposed in [21] enables a more comprehensive evaluation of classification performance stability across different data availability scenarios. By systematically testing ratios from 60:40 to

Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315

90:10, this research aims to provide a robust evaluation of how different partitioning strategies affect classification stability, thereby addressing the limitations of static split protocols commonly used in practical applications [22].

D. Decision Tree and C4.5 Classification

Classification is among the most widely applied techniques in data mining. The process works by analyzing input predictors (independent variables) to assign a specific class label (dependent variable) to each data instance [23]. The decision tree and C4.5 algorithms essentially have the same objective: to build a classification model that groups patients based on hematological parameters to detect types of anemia. A decision tree is a graphical flowchart that illustrates the decision-making process, with its structure resembling a tree. A person can utilize a decision tree to navigate tough decisions by breaking them down into simpler options. Every decision tree contains a node along with a branch linking each node [24].

This research specifically implements the C4.5 algorithm, developed by J.R Quinlan. C4.5 is a landmark classification algorithm in machine learning, particularly valued for its interpretability and its robustness when handling unbalanced datasets [25]. The primary advantage of C4.5 over its predecessor, ID3, is its ability to handle both categorical and numeric features. While ID3 is limited to nominal or ordinal data, C4.5 handles numeric data by establishing threshold values and partitioning the data into intervals. Furthermore, to address the bias toward attributes with numerous variations, C4.5 utilizes Gain Ratio instead of simple Information Gain as the primary splitting criterion [26]. Previous research has classified anemia using various models, including DecisionTree, ExtraTree, RandomForest, and ExtraTrees. Decision Tree exhibiting the highest performance among the tree algorithms, with an accuracy of 98.96% [7].

E. Confusion Matrix

Among the commonly used methods for evaluating classification-oriented machine learning models, confusion matrices are prominent. They provide a thorough overview of the effectiveness of machine learning models by evaluating their performance using numerical metrics, which makes them among the simplest performance indicators [27].

Table 4. Confusion matrix table to evaluate classification performance by providing accuracy, precision, recall and f1-score.

	Predicted Class	
True Class	True Positive (TP)	False Negative (FN)
	False Positive (FP)	True Negative (TN)

True Positive (TP) mean a classification method is correctly at classifying positive cases as positive. False Positive (FP) it is an outcome when incorrectly predicts the negative class as positive [28].

$$Accuracy = \frac{TP+TN}{TP+FP+FN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

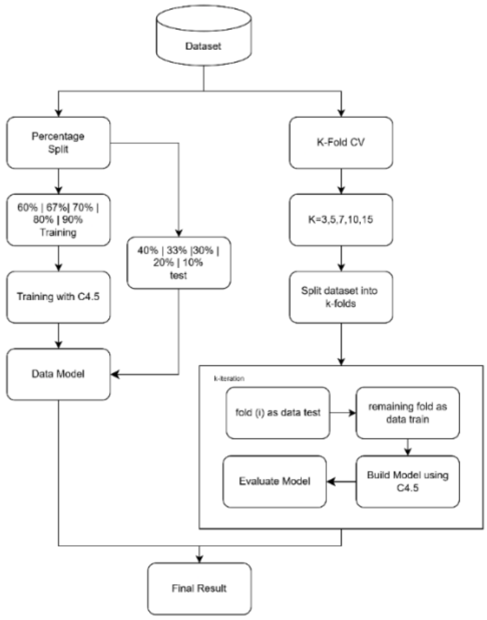
$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = 2 \times \left(\frac{Precision\ x\ Recall}{Precision+Recall}\right) \quad (4)$$

Accuracy Eq. 1. indicate the accuracy rate of correct predictions in the classifier model. Precision Eq. 2. represents the result of the precision of a positively predicted category. Recall Eq. 3. calculates the proportion of true positives accurately recognized for a specific category. Finally, F1-Score Eq. 4. Measures the overall performance of a classifier by integrating precision and recall.

III. Results

The experimental framework utilized in this study is illustrated in the flowchart in Fig 1. The process initiates with a dataset containing anemia clinical records. To evaluate the stability and reliability of the classification model, the data are split into two distinct partitions. The first strategy employs a static Percentage Split with ratios of 60:40, 67:33, 70:30, 80:20, and 90:10. This partitioning is essential for reducing bias and preventing overfitting by allocating the larger portion to training and the remaining to testing. The second strategy uses K-Fold Cross-Validation with K=3, 5, 7, 10, and 15. This method partitions the data into k-disjoint subsets to ensure that each sample is used for both training and validation, thereby reducing random errors. Subsequently, the C4.5 Classification algorithm is applied to the data prepared by both data splitting methods. Finally, the process concludes with the Evaluation phase, where the performance of the C4.5 algorithm under both partitioning schemes is analyzed and compared to determine the optimal evaluation protocol.



Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315

Fig 1. Flowchart illustrating the research methodology structure from the dataset to the evaluation.

Table 5. shows the C4.5 evaluation using a confusion matrix with 2 schemes of method. The first scheme is C4.5 with a percentage split using a 60:40 to 90:10 ratio for the training and testing data. The next scheme is C4.5 with K-Fold Cross Validation using K =3, 5, 7, 10, and 15. The C4.5 with a 90:10 split had the highest accuracy (99.46), whereas K-Fold Cross Validation with K=10 had the highest accuracy (99.19).

Table 5. The C4.5 performance between Percentage Split using 60:40, 67:33, 70:30, 80:20, 90:10, and K-Fold Cross Validation (K=3, 5, 7, 10, 15) in percent(%) based on 10-run aggregation

Method used	Accuracy	Precision	Recall	F1-Score
Percentage Split (60:40)	98.67 ±0.85	96.28 ±3.06	96.37 ±2.37	95.87 ±2.80
Percentage Split (67:33)	99.34 ±0.35	98.56 ±1.10	98.21 ±1.46	98.29 ±1.03
Percentage Split (70:30)	99.04 ±0.39	96.38 ±2.30	96.80 ±2.27	96.30 ±2.17
Percentage Split (80:20)	99.34 ±0.39	97.69 ±2.12	98.31 ±1.86	97.83 ±1.85
Percentage Split (90:10)	**99.46 ±0.61**	**98.32 ±2.53**	**99.28 ±0.87**	**98.48 ±2.25**
K-Fold CV (K=3)	99.11 ±0.12	97.03 ±0.49	97.53 ±0.69	97.00 ±0.48
K-Fold CV (K=5)	99.14 ±0.14	**97.65** ±0.44	97.50 ±0.80	97.27 ±0.71
K-Fold CV (K=7)	99.12 ±0.10	97.48 ±0.99	97.58 ±0.73	97.07 ±0.83
K-Fold CV (K=10)	**99.19** ±0.09	97.44 ±0.91	**97.84** ±0.71	**97.32** ±0.75
K-Fold CV (K=15)	99.15 ±0.05	97.20 ±0.64	97.40 ±0.62	97.05 ±0.59

The experimental results demonstrate that the decision tree model achieves exceptionally high classification performance on the CBC dataset, with the 90:10 split outperforming other ratios, achieving an average accuracy of 99.46%. This superior performance in high-ratio splits is primarily due to the increased volume of training data, which enables the model to develop more precise splitting criteria and robust rules across the major diagnostic categories. While the percentage split method yielded higher peak metrics, it is prone to lower variance and potential over optimality due to the limited size of the test set. In contrast, the K-Fold Cross-Validation approach provided a more conservative yet objective estimation of model performance. The findings indicate that K=10 tends to stabilize accuracy, recall, and f1-score by balancing the learning curve, whereas K=5 occasionally yields higher precision. Primarily due to the larger validation volume in the K=5 configuration. K=5 provides a more representative sample, reducing the model's sensitivity to outliers and noise, which might otherwise inflate false-positive counts in the smaller test sets in K=10.

The disparity between the two evaluation methods highlights the trade-off between model optimization and generalization integrity. The model's reliance on information gain and gain ratio makes it sensitive to the availability of representative training samples. In the 90:10 split, the tree effectively minimizes bias by learning from nearly the entire data distribution, resulting in near-perfect classification across the diagnostic categories. However, the K-Fold results, specifically at K=10 and K=15, offer a more reliable benchmark for clinical applications as they account for the entire dataset through iterative validation. It is within these more rigorous testing environments that the challenge of class imbalance becomes evident. The performance drop in minority classes where samples are extremely limited in each fold underscores the difficulty of maintaining high sensitivity for rare conditions in medical diagnostics. These results suggest that while this tree-based approach is highly effective for general CBC-based diagnosis, the choice of evaluation strategy is crucial in determining the model's reliability for both common and rare clinical cases.

Table 6. Performance metric of the C4.5 algorithm with the Percentage Split (90:10) method over 10 test runs.

C4.5 with Percentage Split	Precision	Recall	F1-Score
Healthy	99.71%	100%	99.86%
Iron Deficiency Anemia	99.50%	100%	99.74%
Leukemia	100%	100%	100%
Leukemia with thrombocytopenia	88.33%	100%	91.67%
Macrocytic anemia	100%	100%	100%
Normocytic hypochromic anemia	100%	99.29%	99.64%
Normocytic normochromic anemia	100%	99.26%	99.62%
Other microcytic anemia	98.57%	95.00%	96.50%
thrombocytopenia	98.75%	100%	99.33%
Accuracy		99,46%	
Macro Avg	98.32%	99.28%	98.48%
Weighted Avg	99.63%	99.46%	99.50%

Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315

Based on the comparative performance of Table 6. and Table 7. of the C4.5 algorithm, the Percentage Split (90:10) method demonstrates higher stability and performance compared to K-Fold Cross Validation across all evaluation metrics. Percentage Split achieved the highest accuracy of 99.46%, exceeding the 99.19% accuracy obtained by the best performer on K-Fold Cross Validation (K=10).

Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315

Table 7. Performance metric of the C4.5 algorithm with K-Fold Cross Validation (K=10) method over 10 test runs.

C4.5 with Percentage Split	Precision	Recall	F1-Score
Healthy	99.54%	99.70%	99.62%
Iron Deficiency Anemia	99.30%	99.47%	99.37%
Leukemia	98.23%	100%	99.03%
Leukemia with thrombocytopenia	83.33%	92.50%	86.23%
Macrocytic anemia	100%	96.50%	97.67%
Normocytic hypochromic anemia	100%	99.64%	99.82%
Normocytic normochromic anemia	99.26%	99.29%	99.26%
Other microcytic anemia	98.67%	94.27%	96.09%
thrombocytopenia	98.58%	99.20%	98.83%
Accuracy		99.19%	
Macro Avg	97.44%	97.84%	97.32%
Weighted Avg	99.27%	99.19%	99.18%

IV. Discussion

Based on the experimental results, the application of the C4.5 algorithm using the Percentage Split method (90:10) gives higher performance compared to K-Fold Cross Validation across all evaluation metrics. The Percentage Split method achieved the highest accuracy of 99.46%, exceeding the 99.19% accuracy obtained by K-Fold Cross Validation (K=10). This trend is consistent across other metrics, with percentage split recording a precision of 98.32%, recall of 99.28%, and F1-score of 98.48%, whereas K-Fold Cross Validation (K=10) recorded values of 97.44%, 97.58%, and 97.84%, respectively. While these differences might appear numerically marginal, they are practically meaningful in a clinical diagnostic context. A higher recall and F1-score, even by a small margin, are critical for minimizing life-threatening false negatives, particularly in rare conditions like leukemia. These findings suggest that for this specific anemia classification dataset, the 90:10 partitioning scenario yields a more precise and sensitive prediction model, as the larger volume of training data effectively optimizes the mapping of input features.

The experimental results indicate a performance variance between the majority and minority classes, particularly within the Leukemia with thrombocytopenia category. In the 90:10 percentage split and K-Fold (K=10) scenarios, this class exhibited lower metric values compared to other prevalent conditions. This trend is primarily attributed to the inherent scarcity of data, as the category comprises only 11 samples within the entire dataset. In a 90:10 split, the validation set may contain only a single representative of this class, where even a single misclassification can lead to a significant fluctuation in the F1-score. From an algorithmic standpoint, the C4.5 decision tree uses gain ratio to determine optimal splits; in cases of extreme class imbalance, the gain-ratio normalization can still be influenced by entropy reduction driven by the majority classes, which may limit the granularity of rules generated for rarer conditions. Furthermore, in the K=10 cross-validation setting, the limited sample size is fragmented across multiple folds, making it challenging for the model to consistently establish a robust decision boundary during each training iteration.

The implementation of different evaluation strategies reveals distinct advantages for the C4.5 model's application on the CBC dataset. The percentage split approach, particularly at the 90:10 ratio, excels in achieving peak classification performance (99.46% ± 0.61%) by providing a high volume of training data. This allows the model to construct highly granular decision rules and minimize learning bias, making it ideal for demonstrating the model's maximum predictive potential. Conversely, K-Fold Cross Validation, especially at K=10, offers superior stability and reliability. As evidenced by its significantly lower standard deviation (99.19% ±0.09%), K-Fold Cross Validation ensures that the evaluation is not biased by a single, potentially favorable data split. By iteratively validating the model across the entire dataset, K-Fold Cross Validation provides a more objective and generalizable benchmark, confirming the model's consistent performance across diverse data distributions, which is critical for real-world clinical implementation.

The proposed C4.5 model achieves highly competitive results when benchmarked against existing anemia classification models. While recent studies, according to Table 8. such as [11], utilized more modern architectures like Extreme Learning Machines (ELM) to achieve an accuracy of 99.21%, their evaluation was limited to a single 67:33 percentage split ratio on a different dataset. In contrast, this study shows that by employing a systematic data-splitting strategy, a traditional algorithm such as C4.5 can achieve an accuracy of 99.46%. Furthermore, compared with studies using the same dataset, such as the ensemble approach in [10], which achieved 99.22% using only a 90:10 ratio, our C4.5 model yields more promising results by maintaining stability across multiple iterations. Similarly, the proposed model outperforms the bagging techniques in [12] and remains highly competitive against modern classifiers in [13], which recorded 99.19% accuracy using an 80:20 split. Notably, even our K-Fold Cross-Validation results (e.g., K=10 at 99.19%) effectively match the performance of these more complex modern models, reinforcing that the C4.5 algorithm, when paired with an optimal partitioning protocol, remains a robust and interpretable solution for clinical anemia screening.

Despite the high accuracy achieved, this study has notable limitations, particularly regarding class imbalance in categories such as Leukemia with thrombocytopenia

Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315

and Macrocytic anemia. This imbalance results in lower F1-score, recall, and precision for minority classes compared to the majority ones, which subsequently affects the overall model performance. In a medical context, these three metrics are critical, as misclassifying rare conditions can have fatal diagnostic consequences. Another limitation is reliance on a static Percentage Split, which introduces a risk of selection bias; however, this study mitigated this risk by employing a 10-run scheme to ensure stable results. The findings indicate that providing a larger, more optimal proportion of training data, such as the 90:10 ratio and K=10, yields better values for these critical metrics, enabling the model to capture the characteristics of minority classes more effectively.

Table 8. Benchmarking results for various study of anemia disease classification model

Researcher	Method	Accuracy(%)
[11]	**ELM**	99.21
[6]	DT, **LR,** MLP, NB, RF and SVM	94.44
[7]	**DT**, ExtraTree, RF and ExtraTrees	98.96
[29]	RF, **NB** and C4.5	96.09
[30]	LogitBoost, RF, DT, NB, **XGB**, GB, SVM, KNN, LR, MLP, LR, MLP and AdaBoost	98.05
[16]	LR(66:33) and **LR(K=16)**	76.30
[10]	**Ensemble**, DT, SVM, Efficient Linear, NN, Kernel, KNN, NB and Discriminant	99.22
[12]	**Bagging**, RF, GB, DT, ExtractTrees, SVM, KNN, LR and AdaBoost	99.00
[13]	**GB**, DT, CatBoost, XGB and RF	99.19
[31]	ELM, **RELM**, SVM and KNN	95.59
[32]	**Multilayer Perceptron**	93.77
[33]	LR, RF, NB, MLP, DT and **KNN**	92.00
[34]	**Support Vector Machine**	76.00
[35]	RF, DA, NB, **SVM**, KNN and Ensemble Learning	84.30
[36]	Decision Tree, Random Forest, XGB and **CatBoost**	97.56
[37]	LR, LDA, KNN, SVM, QDA, NN, CART and **RF**	81.29
Proposed Model	**C4.5 (90:10)**	**99.46**

V. Conclusion

In conclusion, this study demonstrates that the Percentage Split method (90:10) provides high metrics for anemia classification using the C4.5 algorithm on this specific dataset. Achieving a peak accuracy of 99.46% and a Weighted Average F1-Score of 99.50%. Simultaneously, the K-Fold Cross Validation (K=10) also showed strong and reliable results, with an accuracy of 99.19% and a Weighted Average F1-Score of 99.18%. Notably, the K-Fold approach offers a distinct advantage in terms of iterative stability across runs, as evidenced by its low standard deviation, which ensures consistent model performance and reduced dependence on a single data partition. While the overall results are high, the study highlights that for high-risk conditions like Leukemia with thrombocytopenia, the 90:10 ratio is practically meaningful as it provides a larger training volume that yields better critical metricsprecision, recall, and F1-score, which are paramount in clinical diagnostics. Therefore, future research is recommended to incorporate data-level techniques, such as oversampling (e.g., SMOTE), in the training data to further enhance these critical metrics without sacrificing the test data proportion, and to validate the model on external datasets to ensure broader generalizability.

References

[1] Z. Faradila, A. Homaidi, and J. D. Prasetyo, “Classification of Anaemia Status Using The K-Nearest Neighbor Algorithm,” *G-Tech: Jurnal Teknologi Terapan*, vol. 9, no. 1, pp. 436–444, Jan. 2025, doi: 10.70609/gtech.v9i1.6377.

[2] M. N. Garcia-Casal, O. Dary, M. E. Jefferds, and S. R. Pasricha, “Diagnosing anemia: Challenges selecting methods, addressing underlying causes, and implementing actions at the public health level,” Jun. 01, 2023, *John Wiley and Sons Inc*. doi: 10.1111/nyas.14996.

[3] L. Del Castillo *et al.*, “Prevalence and risk factors of anemia in the mother–child population from a region of the Colombian Caribbean,” *BMC Public Health*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12889-023-16475-0.

[4] J. G. Gómez, C. Parra Urueta, D. S. Álvarez, V. Hernández Riaño, and G. Ramirez-Gonzalez, “Anemia Classification System Using Machine Learning,” *Informatics*, vol. 12, no. 1, Mar. 2025, doi: 10.3390/informatics12010019.

[5] M. Mert Usta, M. Çakmak, and D. Ekmekçi, “Anemia Types Prediction Using Ensemble Learning.” [Online]. Available: https://www.icensos.com/

[6] R. Vohra, A. Hussain, A. K. Dudyala, J. Pahareeya, and W. Khan, “Multi-class classification algorithms for the diagnosis of anemia in an outpatient clinical setting,” *PLoS One*, vol. 17, no. 7 July, Jul. 2022, doi: 10.1371/journal.pone.0269685.

[7] G. Airlangga, “Anemia Classification Using Hybrid Machine Learning Models: A Comparative Study of Ensemble Techniques on CBC Data,” *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 4, pp. 1108–1117, Aug. 2024, doi: 10.47065/josyc.v5i4.5848.

Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315

[8] M. K. Hirok, S. Rahman, and M. Parvin, "Anemia prediction and classification of all classes with and without anemia patients using a machine learning model," in *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)*, IEEE, Sep. 2024, pp. 1–6. doi: 10.1109/COMPAS60761.2024.10796730.

[9] S. J. M. Mohammed, A. A. Ahmed, A. A. Ahmad, and M. S. Mohammed, "Anemia Prediction Based on Rule Classification," in *Proceedings - International Conference on Developments in eSystems Engineering, DeSE*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 427–431. doi: 10.1109/DeSE51703.2020.9450234.

[10] A. Végh, L. Takáč, O. Czakóová, K. Dansca, and D. Nagy, "Evaluating Optimizable Machine Learning Models for Anemia Type Prediction from Complete Blood Count Data," *International Journal of Advanced Natural Sciences and Engineering Researches*, vol. 7, no. 7, pp. 108–119, 2024, [Online]. Available: https://as-proceeding.com/index.php/ijanser

[11] D. C. E. Saputra, K. Sunat, and T. Ratnaningsih, "A New Artificial Intelligence Approach Using Extreme Learning Machine as the Potentially Effective Model to Predict and Analyze the Diagnosis of Anemia," *Healthcare (Switzerland)*, vol. 11, no. 5, Mar. 2023, doi: 10.3390/healthcare11050697.

[12] O. O. Okundalaye, N. Özdemir, and F. Evirgen, "Leveraging Machine Learning for Early and Accurate Anaemia Diagnosis: A Comparative Study of Classification Algorithms," in *Advances in Mathematical Modelling, Applied Analysis and Computation*, J. Singh, G. A. Anastassiou, D. Baleanu, and D. Kumar, Eds., Cham: Springer Nature Switzerland, 2025, pp. 42–52.

[13] Y. Cakmak and I. Pacal, "AI-Driven Classification of Anemia and Blood Disorders Using Machine Learning Models," *Computers and Electronics in Medicine*, vol. 2, no. 2, pp. 43–52, Jul. 2025, doi: 10.69882/adba.cem.2025073.

[14] Y. Zhang, Y. Xin, and Q. Li, "Research on parameter selection and optimization of C4.5 algorithm based on algorithm applicability knowledge base," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-11901-2.

[15] M. Teke, T. Etem, and M. Karhan, "Enhancing anemia diagnosis using ensemble machine learning and feature selection techniques on CBC data," *European Physical Journal: Special Topics*, Oct. 2025, doi: 10.1140/epjs/s11734-025-01838-y.

[16] M. Bhagat and B. Bakariya, "Implementation of Logistic Regression on Diabetic Dataset using Train-Test-Split, K-Fold and Stratified K-Fold Approach," *National Academy Science Letters*, vol. 45, no. 5, pp. 401–404, Oct. 2022, doi: 10.1007/s40009-022-01131-9.

[17] M. Rhifky Wayahdi, D. Syahputra, S. Hafiz, and N. Ginting, "EVALUATION OF THE K-NEAREST NEIGHBOR MODEL WITH K-FOLD CROSS VALIDATION ON IMAGE CLASSIFICATION," *JURNAL INFOKUM*, vol. 9, no. 1, pp. 1–6, Dec. 2020, [Online]. Available: http://infor.seaninstitute.org/index.php/infokum/index

[18] I. K. Nti, O. Nyarko-Boateng, and J. Aning, "Performance of Machine Learning Algorithms with Different K Values in K-fold CrossValidation," *International Journal of Information Technology and Computer Science*, vol. 13, no. 6, pp. 61–71, Dec. 2021, doi: 10.5815/ijitcs.2021.06.05.

[19] K. Jung, D. H. Bae, M. J. Um, S. Kim, S. Jeon, and D. Park, "Evaluation of nitrate load estimations using neural networks and canonical correlation analysis with K-fold cross-validation," *Sustainability (Switzerland)*, vol. 12, no. 1, 2020, doi: 10.3390/SU12010400.

[20] I. O. Muraina, "IDEAL DATASET SPLITTING RATIOS IN MACHINE LEARNING ALGORITHMS: GENERAL CONCERNS FOR DATA SCIENTISTS AND DATA ANALYSTS." [Online]. Available: www.artuklukongresi.org

[21] B. Vrigazova, "The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems," *Business Systems Research*, vol. 12, no. 1, pp. 228–242, May 2021, doi: 10.2478/bsrj-2021-0015.

[22] J. Tan, J. Yang, S. Wu, G. Chen, and J. Zhao, "A critical look at the current train/test split in machine learning," Jun. 2021, [Online]. Available: http://arxiv.org/abs/2106.04525

[23] A. Z. Abdullah, B. Winarno, and D. R. S. Saputro, "The decision tree classification with C4.5 and C5.0 algorithm based on R to detect case fatality rate of dengue hemorrhagic fever in Indonesia," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1742-6596/1776/1/012040.

[24] Sumiati, V. V. R. Repi, P. Hendriyati, Anharudin, A. Yusta, and A. Triayudi, "Classification of cardiac disorders based on electrocardiogram data using a decision tree classification approach with the C45 algorithm," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 3, pp. 1128–1138, Sep. 2023, doi: 10.11591/ijai.v12.i3.pp1128-1138.

[25] M. M. Mijwil and R. A. Abttan, "Utilizing the Genetic Algorithm to Pruning the C4.5 Decision Tree

Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315

Algorithm," 2021. [Online]. Available: www.ajouronline.com

[26] M. Yunus, M. K. Biddinika, and A. Fadlil, "Classification of Stunting in Children Using the C4.5 Algorithm," *Jurnal Online Informatika*, vol. 8, no. 1, pp. 99–106, Jun. 2023, doi: 10.15575/join.v8i1.1062.

[27] A. Sharma, M. Grover, J. Malhotra, and S. Sharma, "Predicting Maternal Health Risk Using Machine Learning Models And Comparing The Performance Of Percentage Split And K-Fold Cross Validation," 2024. [Online]. Available: www.ijnrd.org

[28] L. Pawar, J. Malhotra, A. Sharma, D. Arora, and D. Vaidya, "A Robust Machine Learning Predictive Model for Maternal Health Risk," in *3rd International Conference on Electronics and Sustainable Communication Systems, ICESC 2022 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 882–888. doi: 10.1109/ICESC54411.2022.9885515.

[29] P. Verma and V. Chopra, "A Review on Machine Learning Algorithms for Anemia disease Prediction," 2022.

[30] S. S. Abdul-Jabbar, A. K. Farhan, and A. S. Luchinin, "A Comparative Study of Anemia Classification Algorithms for International and Newly CBC Datasets," *International journal of online and biomedical engineering*, vol. 19, no. 6, pp. 141–157, 2023, doi: 10.3991/ijoe.v19i06.38157.

[31] B. Çil, H. Ayyıldız, and T. Tuncer, "Discrimination of β-thalassemia and iron deficiency anemia through extreme learning machine and regularized extreme learning machine based decision support system," *Med. Hypotheses*, vol. 138, p. 109611, 2020, doi: https://doi.org/10.1016/j.mehy.2020.109611.

[32] D. A. Tyas, S. Hartati, A. Harjoko, and T. Ratnaningsih, "Morphological, Texture, and Color Feature Analysis for Erythrocyte Classification in Thalassemia Cases," *IEEE Access*, vol. 8, pp. 69849–69860, 2020, doi: 10.1109/ACCESS.2020.2983155.

[33] S. De and B. Chakraborty, "Case-Based Reasoning (CBR)-Based Anemia Severity Detection System (ASDS) Using Machine Learning Algorithm," in *Advanced Machine Learning Technologies and Applications*, A. E. Hassanien, R. Bhatnagar, and A. Darwish, Eds., Singapore: Springer Singapore, 2021, pp. 621–632.

[34] Y. K. Fu *et al.*, "The tvgh-nycu thal-classifier: Development of a machine-learning classifier for differentiating thalassemia and non-thalassemia patients," *Diagnostics*, vol. 11, no. 9, Sep. 2021, doi: 10.3390/diagnostics11091725.

[35] P. Memmolo *et al.*, "Differential diagnosis of hereditary anemias from a fraction of blood drop by digital holography and hierarchical machine learning," *Biosens. Bioelectron.*, vol. 201, p. 113945, 2022, doi: https://doi.org/10.1016/j.bios.2021.113945.

[36] B. E. Dejene, T. M. Abuhay, and D. S. Bogale, "Predicting the level of anemia among Ethiopian pregnant women using homogeneous ensemble machine learning algorithm," *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, p. 247, 2022, doi: 10.1186/s12911-022-01992-6.

[37] Md. M. Islam *et al.*, "Risk Factors Identification and Prediction of Anemia among Women in Bangladesh using Machine Learning Techniques," *Curr. Womens Health Rev.*, vol. 17, Feb. 2021, doi: 10.2174/1573404817666210215161108.

Author Biography

Nanda Putri Rahmawati is a student in the Computer Science program at the Faculty of Mathematics and Natural Sciences, Universitas Lambung Mangkurat, having commenced her academic journey in 2020. Throughout her studies, she has demonstrated a profound interest in data analysis, leading her to choose Data Science as her major field of specialization. To complete her bachelor's degree, Nanda is conducting a final project specifically focused on medical data classification. Her research, titled "Comparison Between K-Fold Cross Validation and Percentage Split in Decision Tree Algorithms for Anemia Classification". In this study, she analyzes how selecting validation techniques can optimize the accuracy of anemia detection.

Irwan Budiman is an esteemed academic and lecturer at Lambung Mangkurat University. He laid the foundation for his education by earning a Bachelor's Degree in Informatics Engineering from Universitas Islam Indonesia in Yogyakarta. He further advanced his expertise by completing a Master's degree in Information Systems at Diponegoro University, Semarang. His comprehensive research portfolio spans several critical domains, including data mining, human-computer interaction, applied business intelligence, and e-government system development. Through his scholarly work, he actively contributes to the evolution of information technology. He can be reached for academic collaboration via email at: irwan.budiman@ulm.ac.id. Orcid ID: 0000-0002-0514-7429

Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315

Muhammad Itqan Mazdadi is a distinguished lecturer in the Department of Computer Science at Lambung Mangkurat University. His academic foundation was established at the same university, where he completed his undergraduate degree in 2013, followed by a Master's degree in Informatics from Islamic University of Indonesia, Yogyakarta. His primary research expertise and interests are in Data Science and Computer Networking. With a strong background in both theoretical and applied sciences, he actively contributes to advancing technological innovation and academic excellence within his department. He is reachable for academic inquiries at mazdadi@ulm.ac.id. Orcid ID: 0000-0002-8710-4616.

Andi Farmadi, a senior lecturer in the Computer Science program at Lambung Mangkurat University. He has been teaching since 2008 and has beenthe Head of the Data Science Lab since 2018. He completed his undergraduate studies at Hasanuddin University and his graduate studies at Bandung Institute of Technology. His research area, to date, focuses on Data Science. One of his research projects, along with other researchers, published in the International Conference of Computer and Informatics Engineering (IC2IE), is titled "Hyperparameter tuning using GridsearchCV on the comparison of the activation function of the ELM method to the classification of pneumonia in toddlers," and this research was published in 2021. Email: andifarmadi@ulm.ac.id. Orcid ID: 0009-0009-0926-8082

Friska Abadi finished his bachelor's degree in Computer Science from Lambung Mangkurat University in 2011. Subsequently, in 2016, he obtained his master's degree from the Department of Informatics at STMIK Amikom, Yogyakarta. Following that, he joined Lambung Mangkurat University as a lecturer in Computer Science. As a lecturer, he teaches programming. Apart from that, he also carries out research and community service. Other activities as an application developer, whether using a web or mobile platform. Currently, he holds the position of head of the software engineering laboratory. His current area of research revolves around software engineering, and he is also interested in machine learning. He can be contacted at Email: friska.abadi@ulm.ac.id. Orcid ID: 0000-0002-9449-8000

Corresponding author: Irwan Budiman, irwan.budiman@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia
Digital Object Identifier (DOI): https://doi.org/10.35882/ijeeemi.v8i1.315