

Reviewer 2

After reviewing each of your comments in detail, we have responded to them. We appreciate the feedback that was provided.

Overall, your paper presents significant knowledge and considerable effort in the field. I have provided a few comments that I believe will help you improve your manuscript in terms of clarity and completeness and give greater impact to your study.

I hope you find these comments helpful in improving your manuscript.

I look forward to seeing the revised version and wish you the best in your research endeavors.

Best regards

The summary: The study entitled: Anemia classification system using machine learning. by Jorge Gómez Gómez, Camilo Parra Urueta, Daniel Salas Álvarez, Velssy Hernández Riaño and Gustavo Ramirez-Gonzalez.

The study developed a system using blood count data and supervised learning algorithms to predict the type of anemia. Anemia, characterized by low red blood cell levels, impairs oxygenation and often leads to symptoms like fatigue and shortness of breath. The system classifies anemia into three types: microcytic, normocytic, and macrocytic, based on metrics like mean corpuscular volume. The Random Forest model achieved an accuracy of 99.82%, surpassing previous studies.

Comments – Suggestions

The data definition should include the relevant link to the dataset you have chosen. Alternatively, you could include the dataset you used as a supplementary file.

R:/

This is the link to the dataset obtained from Kaggle:
<https://www.kaggle.com/datasets/biswaranjanrao/anemia-dataset?resource=download&select=anemia.csv>

This is the link to the file that was trained to solve the problem:
https://github.com/jeliecergomez/Machine_Learning/blob/main/Anemia_Confirmate_Type.csv

Also, the variables (features) of the Kaggle dataset should be described in detail in a supplementary file.

R:/

This is the file of the anemia dataset characteristics:
https://github.com/jeliecergomez/Machine_Learning/blob/main/About%20Dataset_Anemia.txt

The inclusion criteria for the dataset you are using should also be clearly stated.

R:/

The dataset consists of hematological parameters such as hemoglobin (HGB), mean cell volume (MCV) among others. Each value contributes differently to the diagnosis and classification of anemia as follows:

- HGB: Determines the presence of anemia based on threshold levels (e.g., <13.6 g/dL for men).
- MCV: Differentiates anemia types (microcytic, normocytic, macrocytic) based on red blood cell size.
- MCH and MCHC: Indicate the content and concentration of hemoglobin within cells, aiding in subclassification.

Once these metrics are understood, algorithms can set thresholds for classification.

You will make your study more comprehensive if you clearly describe the framework and workflow of your proposed model and accompany it with appropriate figures.

R:/

The framework and workflow of the proposed model is presented below, Figure 2 shows the main elements that provide workflow of the model.

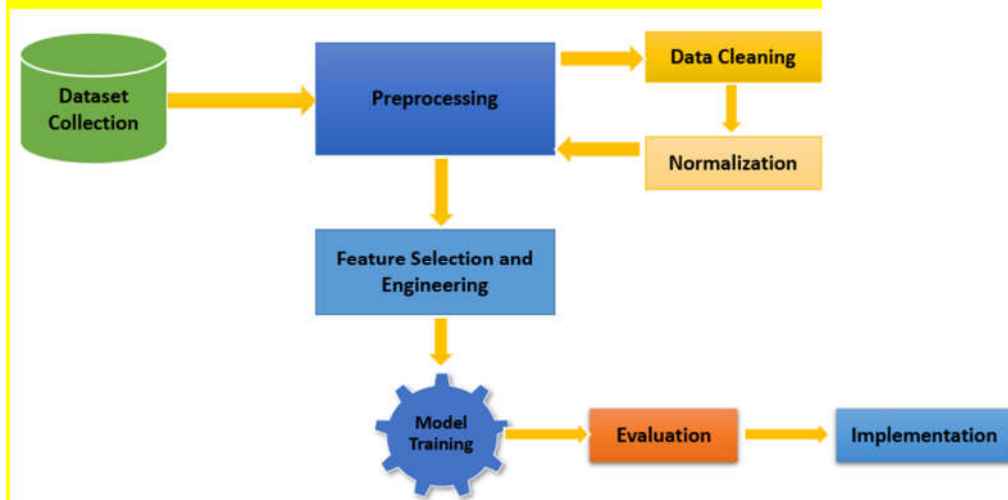


Figure 2. framework and workflow of the proposed model

An extended description of the framework is included in the following steps:

1. Dataset Collection and Preprocessing:

- Source: Kaggle dataset with 1,421 instances.
- Data Cleaning: Removal of incomplete or inconsistent records.
- Normalization: Ensuring values are scaled to eliminate bias due to variable magnitudes.

2. Feature Selection and Engineering:

- Parameters like HGB, MCV, MCHC, and RDW are selected based on their diagnostic significance.

- Thresholds are defined for anemia classification (e.g., MCV for microcytic: <80).

3. Model Training:

- Algorithms: Random Forest, Decision Trees, and Linear Discriminant Analysis.
- Dataset split into training (80%) and testing (20%) sets.

4. Evaluation:

- Metrics: Accuracy, precision, recall, F1 score, and confusion matrix.

5. Implementation:

- Integration into a diagnostic system for clinical use.
- Designed for regions with limited access to laboratory tests.