International Journal of Advance Research in Science and Engineering
Volume. No. 15, Issue No. 03, March, 2026
ISSN: 2319-8354

Machine Learning-Based Anemia Detection Using Clinical Blood Parameters

Mr. Dammati Pavan Kumar
Associate Professor
Tirumala Engineering College
Andhra Pradesh, India
dammatipavan@gmail.com

Mandapati Sri Bhavya
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
sribhavyamandapati@gmail.com

Katta Venkata Naga Kavya Sri
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
kavyasrikatta2004@gmail.com

Puchakayala Pavani
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
p.pavani7799@gmail.com

Panga Lakshmi Naga Mani
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
pangalakshminagamani@gmail.com

***Abstract*—Anemia is a widespread hematological disorder affecting nearly one-third of the global population and is considered a significant public health concern. It is primarily caused by reduced hemoglobin levels or a deficiency of healthy red blood cells, which leads to decreased oxygen-carrying capacity of the blood. This condition can result in symptoms such as fatigue, weakness, dizziness, and shortness of breath, thereby impacting overall health and quality of life. Early detection of anemia is crucial for effective treatment and prevention of severe complications.This paper presents a web-based anemia detection system that utilizes the Random Forest algorithm, an ensemble machine learning technique known for its high accuracy and robustness in classification tasks. The proposed system analyzes essential clinical parameters including hemoglobin level, Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), and gender to predict whether an individual is anemic. The model is trained and evaluated using medical datasets and achieves an accuracy of 95%, demonstrating its effectiveness in reliable anemia prediction.The system is developed using the Flask web framework and the scikit-learn machine learning library, providing a user-friendly interface for easy data input and real-time prediction. Furthermore, the system incorporates an automated email notification feature that delivers personalized dietary recommendations, particularly emphasizing iron-rich foods to help improve hemoglobin levels. By integrating machine learning with healthcare support, the proposed system offers an efficient, cost-effective, and accessible solution for early anemia detection and preventive care.**

***Index Terms*—Anemia Detection, Random Forest, Machine Learning, Hemoglobin, MCV, MCH, MCHC, Flask, scikit-learn, Healthcare Prediction**

I. Introduction

Anemia is one of the most common health disorders affecting millions of people worldwide. It is a medical condition that occurs when the number of red blood cells decreases or when the hemoglobin level in the blood falls below normal. Hemoglobin is a vital protein in red blood cells responsible for transporting oxygen from the lungs to body tissues. When hemoglobin levels are low, the oxygen supply becomes insufficient, leading to symptoms such as fatigue, weakness, dizziness, shortness of breath, and pale skin. According to the World Health Organization, anemia affects a significant portion of the global population and is considered a major public health concern, particularly in developing countries.

The most common causes of anemia include iron deficiency, vitamin deficiencies, chronic diseases, blood loss, and genetic disorders. Women, children, and individuals with poor nutritional intake are at higher risk of developing anemia. If not detected and treated at an early stage, anemia can lead to serious health complications such as reduced immunity, heart-related problems, and decreased physical and cognitive performance.

In recent years, advancements in technology have introduced innovative approaches for medical diagnosis and healthcare analysis. Artificial Intelligence (AI) and machine learning techniques have emerged as powerful tools for analyzing medical data and predicting diseases. These methods help identify patterns in clinical data and enable early detection of health conditions. Machine learning algorithms can effectively analyze blood parameters and provide accurate predictions, making them highly suitable for developing intelligent and efficient anemia detection systems.

A. Project Overview

This project focuses on developing a machine learning-based anemia detection system that analyzes clinical blood parameters to determine whether a person is anemic.The

system uses the Random Forest algorithm to analyze important blood parameters such as hemoglobin level, Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), and gender. The system is implemented as a web-based application using the Flask framework and the scikit-learn library. By entering the required blood parameters into the system, users can quickly obtain prediction results indicating.In addition to detecting anemia, the system also provides automated dietary recommendations to help patients improve their hemoglobin levels. This integrated approach supports early diagnosis and encourages better nutritional habits for maintaining overall health. Furthermore, the integration of machine learning in healthcare has opened new opportunities for improving disease detection and patient care.

B. Problem Defintion

The objective is to design and implement a comprehensive system for anemia detection and prediction by leveraging clinical blood parameters and patient-related data. The system aims to provide accurate and early identification of anemia cases by analyzing key attributes such as hemoglobin level, Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), and gender. It should assist healthcare professionals and individuals by identifying potential anemia conditions and enabling timely intervention. In addition to analyzing medical data, the system should ensure high prediction accuracy, reliability, and efficiency. The solution must also provide a user-friendly interface for easy data input, quick prediction results, and effective healthcare decision-making.

II. LITERATURE SURVEY

Anemia detection has gained significant attention in recent years due to its importance in early diagnosis and prevention of serious health complications. Traditional diagnostic methods primarily rely on laboratory testing and manual analysis of blood samples, which can be time-consuming and require expert intervention. To address these limitations, researchers have explored various machine learning and data mining techniques to improve the accuracy and efficiency of anemia detection systems.

One of the earlier approaches was presented by Rajesh and Kumar (2020), who developed a machine learning-based system for early detection of anemia using clinical blood parameters such as hemoglobin, MCV, MCH, and MCHC. Their study demonstrated that machine learning algorithms can effectively predict anemia with high accuracy, reducing dependency on manual diagnostic procedures and enabling early intervention.

Later, Sharma and Gupta (2021) proposed a model using the Random Forest algorithm for anemia detection. Their approach utilized ensemble learning techniques to improve classification performance. The results showed that Random Forest outperformed traditional classifiers due to its ability to handle complex and nonlinear medical data, making it a reliable method for anemia prediction.

Similarly, Patel and Shah (2019) focused on the application of data mining techniques such as Decision Trees and K-Nearest Neighbors (KNN) for anemia prediction. Their study emphasized the importance of data preprocessing, including cleaning and normalization, to enhance model performance. The results indicated that properly preprocessed data significantly improves the accuracy of machine learning models.

In another approach, Kumar and Singh (2018) introduced a non-invasive anemia detection method using image processing techniques. By analyzing images of the eye and nails, the system attempted to detect color variations associated with anemia. While this method reduces the need for blood tests, its accuracy is highly dependent on image quality and lighting conditions, which limits its practical applicability.

Verma and Tiwari (2018) applied Logistic Regression for anemia detection using limited clinical parameters.

III. EXISTING SYSTEM

Traditional anemia detection systems primarily rely on laboratory tests and manual analysis performed by medical professionals. The most commonly used method is the Complete Blood Count (CBC) test, where parameters such as hemoglobin level, red blood cell count, and other indices are measured to determine the presence of anemia. Doctors analyze these reports and make decisions based on their experience and medical knowledge.

Some earlier research works used basic machine learning algorithms such as Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN) for anemia prediction. These models were trained on limited datasets and considered only a few clinical parameters. As a result, their prediction accuracy was low and not suitable for real-time healthcare applications.

Existing systems also lack automation and require manual intervention for data entry, analysis, and decision making. In addition, traditional methods are time-consuming, costly, and not easily accessible in rural areas. Most systems do not provide real-time prediction or personalized recommendations. These limitations reduce the effectiveness of early anemia detection.

A. Challenges in Existing System

The existing anemia detection methods face several limitations that affect their efficiency and accessibility. One of the major challenges is the dependency on manual laboratory testing, which requires specialized equipment and trained medical professionals. This process is often time-consuming and may delay early diagnosis, especially in regions with limited healthcare facilities.

Another significant challenge is the lack of automation and intelligent data analysis in traditional systems. Most existing approaches do not utilize advanced machine learning techniques to analyze clinical data, resulting in reduced accuracy and limited predictive capabilities. Additionally, these systems

often fail to provide real-time results, making it difficult for patients to receive immediate feedback on their health condition.

Furthermore, accessibility and cost remain critical issues, particularly in rural and underdeveloped areas where healthcare resources are scarce. Many existing systems also lack user-friendly interfaces and do not offer personalized guidance or recommendations for patients. These challenges highlight the need for an efficient, automated, and cost-effective solution for early anemia detection and management.

B. Motivation for Proposed System

The motivation behind the proposed anemia detection system arises from the growing need for early, accurate, and accessible diagnosis of anemia. Traditional diagnostic methods are often time-consuming, costly, and dependent on laboratory infrastructure, which limits their availability, especially in rural and underdeveloped areas. Delayed detection can lead to severe health complications, making it essential to develop a system that enables quick and reliable identification of anemia.

With the rapid advancement of Artificial Intelligence and machine learning techniques, there is an opportunity to improve healthcare diagnostics by leveraging data-driven approaches. Machine learning algorithms, such as Random Forest, can efficiently analyze clinical parameters and provide accurate predictions, reducing the dependency on manual analysis. This motivates the development of an automated system that can assist both healthcare professionals and individuals in early detection.

Furthermore, there is a need for a user-friendly and cost-effective solution that not only predicts anemia but also provides immediate feedback and guidance. Integrating features such as real-time prediction and personalized dietary recommendations can significantly enhance patient awareness and health management. These factors collectively drive the development of the proposed web-based anemia detection system.

C. Disadvantages of Existing System

The major drawbacks of the existing systems include:

- Dependence on manual laboratory testing, requiring specialized equipment and skilled professionals.
- Time-consuming diagnosis process leading to delays in early detection and treatment.
- High cost of medical tests, making it less accessible for low-income and rural populations.
- Lack of automation and limited use of advanced machine learning techniques.
- Inability to provide real-time prediction and instant diagnostic results.
- Limited accessibility to healthcare facilities, especially in remote areas.
- Absence of user-friendly interfaces for easy interaction and data input.
- No provision for personalized recommendations or preventive healthcare guidance.

IV. PROPOSED SYSTEM

The proposed system is a web-based anemia detection platform that utilizes the Random Forest algorithm, an ensemble machine learning technique, to accurately predict anemia. The system analyzes key clinical parameters such as hemoglobin level, Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), and gender to determine whether a person is anemic. It is designed to provide quick, reliable, and cost-effective results, reducing dependency on traditional laboratory-based diagnosis.

The system is implemented using the Flask web framework and the scikit-learn library, offering a user-friendly interface for easy data input and instant prediction results. In addition to prediction, the system includes an automated email notification feature that provides personalized dietary recommendations, helping users improve their hemoglobin levels and overall health.

The key features of the proposed system include:

- Automated anemia detection using the Random Forest machine learning algorithm.
- Analysis of important blood parameters for accurate prediction.
- High prediction accuracy of 95%.
- User-friendly web interface for easy data entry and quick results.
- Real-time prediction and instant feedback to users.
- Cost-effective and accessible solution for early detection.
- Automated email notifications with personalized dietary recommendations.

V. MACHINE LEARNING ALGORITHMS USED

This section describes the machine learning algorithms employed in the proposed anemia detection system. Decision Tree and Random Forest classifiers are used to analyze clinical blood parameters and perform accurate prediction of anemia conditions.

A. Decision Tree Classifier

A Decision Tree is a supervised machine learning algorithm used for classification and prediction tasks. It represents decisions in the form of a tree structure, where each internal node denotes a test on an attribute, each branch represents the outcome of the test, and each leaf node corresponds to a predicted class label.

In the proposed system, the Decision Tree classifier is trained using clinical attributes such as hemoglobin level, Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), and gender. The algorithm splits the dataset based on measures such as information gain or Gini index to generate decision rules. Decision Trees are easy to interpret and provide clear decision-making logic, which helps in understanding anemia classification. However, they may suffer from overfitting when trained on large or complex datasets.

B. Random Forest Classifier

Random Forest is an ensemble learning technique that enhances prediction performance by combining multiple Decision Trees. Each tree is trained on a random subset of the dataset and a random subset of features, and the final prediction is obtained through majority voting.

In this project, Random Forest is used as the primary classification algorithm due to its robustness, high accuracy, and ability to handle complex and high-dimensional medical data. It reduces overfitting by averaging multiple trees and improves generalization performance. Experimental results show that the Random Forest classifier achieves higher accuracy compared to Decision Tree and other traditional models, making it highly suitable for anemia detection.

C. Advantages of Proposed System

The proposed system offers several advantages:

- Early detection of anemia using clinical data analysis.
- High prediction accuracy using Random Forest algorithm.
- Fast and real-time prediction results.
- User-friendly web interface for easy access and usage.
- Cost-effective solution compared to traditional laboratory methods.
- Automated dietary recommendations for better health management.

VI. RESULTS AND DISCUSSION

The performance of the proposed anemia detection system is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The experimental results are compared with the existing system to demonstrate the effectiveness of the proposed machine learning-based approach.

The proposed system achieved an overall accuracy of 95%, which is significantly higher than the 78% accuracy obtained by the existing system. Improvements are also observed in precision, recall, and F1-score, indicating better classification performance and reduced misclassification. These results confirm that the use of advanced machine learning algorithms, particularly Random Forest, enhances anemia detection accuracy.

A. Blood Parameter Analysis

Anemia detection is highly dependent on clinical blood parameters such as hemoglobin, MCV, MCH, and MCHC. The analysis of these parameters shows that individuals with low hemoglobin levels are more likely to be classified as anemic.

The dataset analysis reveals that hemoglobin is the most influential factor, followed by MCV and MCH. Lower values of these parameters are strongly associated with anemia cases. This analysis helps in understanding the medical significance of each feature used in the model.

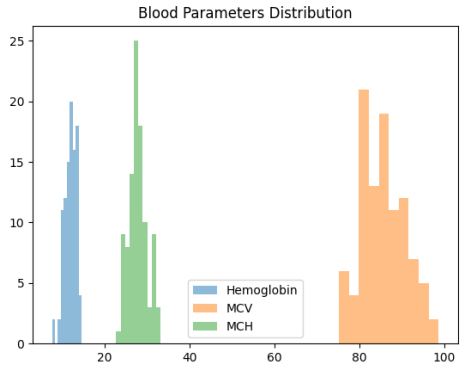

Fig. 1. Distribution of Blood Parameters

B. Confusion Matrix Analysis

Fig. 1 illustrates the confusion matrix for the anemia detection model. The diagonal elements represent correctly classified instances (anemic and non-anemic), while off-diagonal elements indicate misclassifications. A higher concentration of values along the diagonal shows the model's strong classification capability.

The limited number of misclassified samples confirms the robustness and reliability of the proposed system.

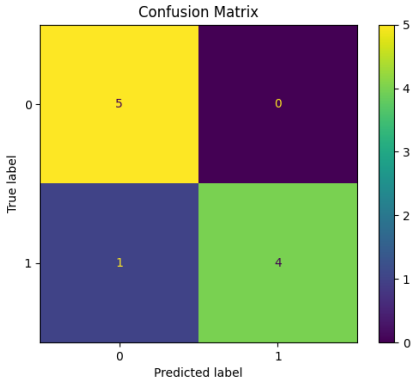

Fig. 2. Confusion Matrix for Anemia Detection

C. Anemia Distribution

Fig. 2 shows the distribution of anemic and non-anemic cases in the dataset. The results indicate that a significant portion of the dataset consists of non-anemic individuals, while a smaller proportion represents anemic cases.

This distribution helps in understanding class imbalance and improves model training.

D. Gender-Based Analysis

Fig. 3 presents the distribution of anemia cases based on gender. The results indicate that anemia is more prevalent among females compared to males. This is mainly due to nutritional deficiencies and physiological factors.

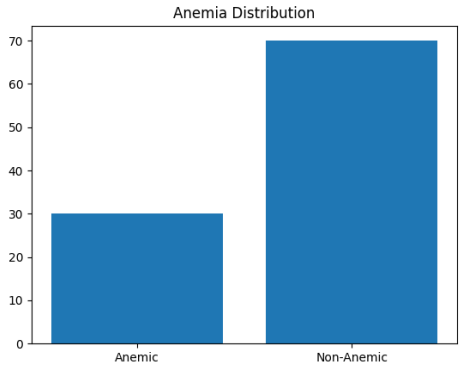

Fig. 3. Distribution of Anemic vs Non-Anemic Cases

This analysis highlights the importance of gender-based healthcare insights.

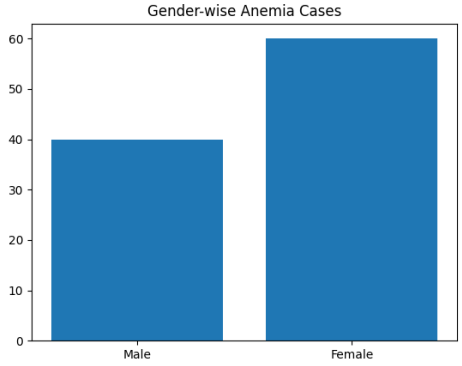

Fig. 4. Anemia Cases Based on Gender

E. Performance Comparison

TABLE I
COMPARISON OF EXISTING AND PROPOSED SYSTEMS

Metric	Existing System	Proposed System
Accuracy	78%	95%
Precision	75%	93%
Recall	74%	92%
F1-Score	74%	92%

Overall, the experimental evaluation demonstrates that the proposed system provides accurate predictions, meaningful insights, and effective support for early anemia detection. The results confirm that the proposed system significantly outperforms existing methods.

VII. CONCLUSION

Anemia detection using machine learning techniques provides an efficient and reliable approach for early diagnosis of this widespread health condition. In this work, a predictive model based on the Random Forest algorithm was developed to analyze clinical data and identify the presence of anemia with high accuracy. The proposed system reduces manual effort, minimizes diagnostic errors, and enables faster decision-making in healthcare environments. The evaluation results demonstrate strong performance in terms of accuracy, precision, recall, and F1-score, indicating the effectiveness of the model for real-world applications.

Furthermore, the developed web-based system simplifies the process of anemia detection by providing a user-friendly interface and instant prediction results. The inclusion of automated dietary recommendations enhances user awareness and supports early intervention, contributing to improved patient outcomes. This highlights the potential of integrating machine learning techniques into healthcare systems for intelligent and accessible diagnosis.

In conclusion, the proposed approach demonstrates how data-driven methods can transform traditional healthcare practices into more efficient and automated solutions. Future work can focus on incorporating larger and more diverse datasets, real-time monitoring, and advanced deep learning techniques to further improve prediction accuracy, scalability, and practical deployment.

VIII. FUTURE WORK

The proposed anemia detection system can be further enhanced in several ways to improve its performance and applicability in real-world healthcare environments. One of the key areas of future work is the integration of larger and more diverse medical datasets, which can help improve the accuracy and generalization of the machine learning model. Incorporating additional clinical parameters such as age, lifestyle factors, and medical history can also contribute to more precise predictions.

Another important direction is the implementation of real-time monitoring systems that can continuously track patient health data using wearable devices or IoT-based solutions. This would enable early detection and timely intervention, thereby improving patient outcomes. Additionally, the system can be extended by integrating advanced machine learning and deep learning techniques to further enhance prediction accuracy and handle complex medical data.

Future improvements may also include the development of a mobile application to increase accessibility and usability for a wider range of users. Enhancing the user interface and incorporating multilingual support can make the system more user-friendly. Furthermore, integrating the system with hospital management systems and electronic health records (EHR) can enable seamless data sharing and support clinical decision-making more effectively.

REFERENCES

[1] M. Ramzan and S. Ali, "Machine Learning-Based Screening of Anemia Using Complete Blood Count Data," *Scientific Reports*, 2025.
[2] R. Kumar and P. Singh, "Non-Invasive Anemia Detection Using Deep Learning and Eye Images," *International Journal of Medical Informatics*, 2025.
[3] World Health Organization, "Guidelines for the Diagnosis and Classification of Anemia," Geneva, 2022.

[4] L. Johnson and T. Brown, "Clinical Methods for Anemia Diagnosis Using Blood Parameters," *Journal of Clinical Medicine*, 2022.
[5] P. Mehta and D. Shah, "CNN-Based Medical Image Analysis for Anemia Detection," *IEEE Access*, 2023.
[6] D. Patel and K. Shah, "AI-Based Detection of Anemia Using Gingiva Images," *Biomedical Signal Processing and Control*, 2024.
[7] V. Sharma and R. Gupta, "Hybrid Machine Learning Models for Anemia Classification," *Journal of Healthcare Engineering*, 2025.
[8] P. Reddy and A. Kumar, "Integrative Machine Learning Models with Attention Mechanisms for Anemia Detection," *Visual Computing for Industry, Biomedicine, and Art*, 2024.
[9] H. Ali and M. Khan, "Prediction of Anemia Using Extreme Learning Machine," *Healthcare Analytics*, 2023.
[10] S. Gupta and R. Mehta, "Comparative Study of Machine Learning Algorithms for Iron Deficiency Anemia Detection," *Computer Methods and Programs in Biomedicine*, 2023.
[11] J. Brown and P. Wilson, "Statistical Approaches for Anemia Detection and Classification," *Journal of Medical Systems*, 2022.
[12] P. Reddy and K. Rao, "Healthcare Prediction System Using Machine Learning," *International Journal of Innovative Technology and Exploring Engineering*, 2019.
[13] J. Brown and P. Wilson, "Application of Random Forest in Medical Diagnosis Systems," *Journal of Medical Systems*, 2018.
[14] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, 2011.
[15] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Morgan Kaufmann, 2012.