

Developing a Transparent Anaemia Prediction Model Empowered with Explainable AI

Mrs. Deepa

Ms. N. Manisha

Jawaharlal Nehru Technological University Hyderabad

Abstract Anaemia is one of the most prevalent haematological disorders worldwide and affects individuals across different age groups, particularly women and children. Delayed diagnosis often results in severe health complications, emphasizing the need for intelligent and reliable prediction systems that can assist healthcare professionals in early clinical decision-making. This research presents an explainable machine learning framework for anaemia prediction by integrating the Random Forest classifier with Shapley Additive explanations (SHAP). Initially, clinical data obtained from Complete Blood Count (CBC) examinations undergo preprocessing, including duplicate removal, missing-value handling, feature selection, and data validation to improve data quality. The processed dataset is divided into training and testing subsets for model development and evaluation. Random Forest is employed as the primary classification algorithm because of its robustness, high predictive capability, and resistance to overfitting. Unlike conventional black-box prediction models, the proposed framework incorporates SHAP to provide transparent explanations for every prediction by identifying the contribution of each clinical attribute. These explanations improve interpretability and strengthen clinician confidence in AI-assisted diagnosis. A user-friendly Streamlit application enables healthcare professionals to enter patient information and receive real-time predictions together with graphical explanations. Experimental evaluation demonstrates that the proposed approach achieves high classification accuracy while maintaining interpretability and practical usability. The proposed framework contributes toward trustworthy Artificial Intelligence in healthcare by combining accurate disease prediction with transparent clinical decision support.

Keywords: Anaemia Prediction, Explainable Artificial Intelligence, Machine Learning, Random Forest, SHAP, Healthcare Analytics, Clinical Decision Support, Streamlit.

1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) have transformed numerous domains by enabling intelligent automation and data-driven decision-making. Healthcare is among the fields that have benefited significantly from these technologies, particularly in disease prediction, clinical diagnosis, and patient monitoring. Machine learning techniques are capable of discovering

hidden patterns within medical datasets, allowing physicians to identify diseases at an earlier stage and improve treatment outcomes.

Anaemia remains one of the most common blood disorders worldwide and is characterized by insufficient haemoglobin concentration or reduced red blood cell count, leading to decreased oxygen transportation throughout the body. According to global health studies, millions of individuals suffer from anaemia each

year, with women, children, and elderly populations being particularly vulnerable. Early diagnosis plays a crucial role in preventing complications such as fatigue, impaired cognitive development, cardiovascular disorders, and pregnancy-related risks.

Traditional diagnostic procedures primarily rely on laboratory investigations and clinical interpretation of Complete Blood Count (CBC) parameters. Although laboratory-based diagnosis provides reliable results, it often requires skilled professionals, specialized equipment, and additional processing time. These limitations become more significant in rural or resource-constrained healthcare environments where rapid diagnosis may not always be feasible.

Machine learning offers an effective alternative by automatically learning relationships between clinical parameters and disease conditions. Various supervised learning algorithms, including Logistic Regression, Decision Trees, Support Vector Machines, Naïve Bayes, Gradient Boosting, and Random Forest, have demonstrated encouraging performance in disease prediction tasks. Among these approaches, Random Forest has attracted considerable attention because of its ability to handle nonlinear relationships, noisy datasets, and high-dimensional clinical features while reducing overfitting through ensemble learning.

Despite their predictive capability, many machine learning models behave as black-box systems that generate predictions without providing understandable explanations. In clinical practice, this lack of transparency often limits physician confidence and slows the adoption of Artificial Intelligence technologies. Medical professionals require clear justification before relying on AI-generated recommendations for patient care.

Explainable Artificial Intelligence (XAI) addresses this challenge by making machine learning models more transparent and interpretable. Shapley Additive explanations (SHAP), derived from cooperative game theory, quantifies the contribution of each input feature to the final prediction. Rather than simply

identifying whether a patient is anaemic, SHAP illustrates how variables such as haemoglobin concentration, Mean Corpuscular Volume (MCV), Mean Corpuscular Haemoglobin (MCH), Mean Corpuscular Haemoglobin Concentration (MCHC), and gender influence the prediction. These explanations increase user trust and facilitate informed clinical decision-making.

This research proposes an explainable anaemia prediction framework that combines the predictive strength of the Random Forest algorithm with SHAP-based interpretation. Clinical data are preprocessed and used to train the classification model, while SHAP generates both global and local explanations to enhance transparency. The trained model is deployed through a Streamlit web application that allows healthcare professionals to perform real-time prediction and visualization using a simple graphical interface.

2. Literature Review

Artificial Intelligence (AI) and Machine Learning (ML) have become increasingly important in healthcare because of their ability to analyse large volumes of clinical data and support accurate disease diagnosis. Researchers have proposed numerous intelligent systems for predicting anaemia using Complete Blood Count (CBC) parameters and other clinical features. While these studies have demonstrated promising predictive performance, many still face challenges related to interpretability, clinical trust, and real-world deployment.

One of the earliest and most widely adopted machine learning algorithms for medical diagnosis is **Logistic Regression**. It has been extensively applied to binary disease classification due to its simplicity and computational efficiency. Several studies reported satisfactory results for anaemia prediction using haemoglobin concentration and demographic attributes. However, Logistic Regression assumes a linear relationship between predictor variables and the target class, limiting its capability to model complex interactions present in medical datasets.

Decision Tree models have also been employed because they provide rule-based decision-making that is easy to understand. Their transparent structure allows clinicians to follow the reasoning behind predictions. Nevertheless, individual decision trees are highly sensitive to variations in training data and often suffer from overfitting, resulting in inconsistent performance on unseen clinical records.

Support Vector Machine (SVM) algorithms have demonstrated strong classification capabilities for high-dimensional healthcare datasets. Researchers have used SVM classifiers for identifying anaemia by analysing blood-related measurements, reporting competitive accuracy under carefully selected kernel functions. Despite these advantages, SVM models require extensive parameter tuning and become computationally expensive when processing large clinical datasets.

Several investigations have explored **K-Nearest Neighbour (KNN)** classifiers for anaemia prediction. KNN is simple to implement and performs well when similar patient records exist in the training dataset. However, prediction time increases significantly with dataset size, and model performance depends heavily on selecting an appropriate value of k and suitable distance metrics.

Recent studies increasingly favour **ensemble learning** approaches because they combine multiple models to improve prediction robustness. Among these methods, **Random Forest** has consistently achieved superior performance in disease prediction tasks. By constructing multiple decision trees using randomly selected samples and feature subsets, Random Forest reduces variance, minimises overfitting, and improves generalisation. Numerous researchers have reported that Random Forest outperforms traditional classifiers in terms of accuracy, precision, recall, and stability when analysing clinical datasets.

Data preprocessing has been recognised as another critical factor influencing model performance. Contemporary research highlights the importance of handling missing values, eliminating duplicate records, encoding

categorical variables, detecting outliers, and selecting informative features before model development. Effective preprocessing enhances data quality, reduces noise, and improves classification accuracy while lowering computational complexity.

Although many machine learning models achieve excellent predictive performance, their practical adoption in healthcare remains limited because they often function as **black-box systems**. Physicians generally require explanations for diagnostic recommendations before incorporating AI-assisted predictions into routine clinical practice. A lack of transparency may reduce confidence, hinder regulatory acceptance, and raise ethical concerns regarding automated medical decision-making.

To overcome these challenges, researchers have introduced **Explainable Artificial Intelligence (XAI)** techniques that provide understandable explanations for machine learning predictions. Among various explanation methods, **SHapley Additive exPlanations (SHAP)** has emerged as one of the most reliable and mathematically grounded approaches. SHAP quantifies the contribution of each input feature to the final prediction using principles derived from cooperative game theory. Unlike conventional feature importance measures, SHAP generates both global explanations describing overall model behaviour and local explanations that justify individual patient predictions.

Recent healthcare studies have successfully applied SHAP to interpret predictive models for cardiovascular disease, diabetes, kidney disorders, cancer detection, and blood-related abnormalities. These investigations demonstrate that explainable models improve physician confidence by clearly identifying influential clinical attributes responsible for each prediction. Moreover, transparent AI systems facilitate error analysis, bias detection, and regulatory compliance, making them more suitable for deployment in sensitive healthcare environments.

Parallel developments have focused on improving accessibility through lightweight web applications. Frameworks such as **Streamlit**

enable rapid deployment of machine learning models with interactive graphical interfaces, allowing healthcare professionals to input patient information and obtain predictions in real time without requiring programming expertise. The integration of prediction models with explainability tools further enhances usability by presenting both diagnostic outcomes and feature-level explanations within a single interface.

routine clinical use. These limitations restrict widespread adoption in real-world healthcare settings.

3. Proposed Methodology

The proposed framework presents an intelligent and explainable clinical decision-support system for the early prediction of anaemia using machine learning. The system combines the predictive capability of the Random Forest classifier with the interpretability of Shapley Additive explanations (SHAP). Unlike conventional diagnostic systems that only provide classification results, the proposed model explains the contribution of each clinical

3.1 System Architecture

The proposed architecture consists of interconnected modules that transform raw clinical data into interpretable prediction results.

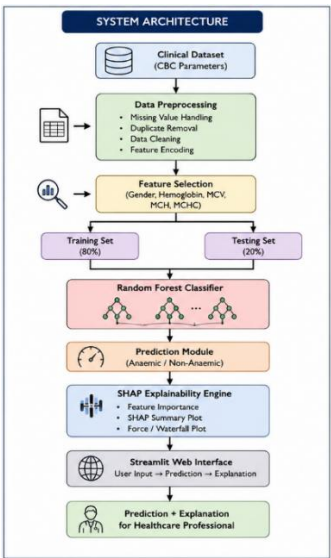


Fig. 1. System architecture of the proposed explainable anaemia prediction system.

The proposed research addresses these shortcomings by integrating a high-performing **Random Forest** classifier with **SHAP-based Explainable Artificial Intelligence** within an interactive Streamlit application. In addition to producing accurate anaemia predictions, the framework provides transparent feature-level explanations that assist healthcare professionals in understanding the reasoning behind every decision. This combination of predictive performance, interpretability, and usability contributes toward the development of trustworthy AI-driven clinical decision-support systems.

parameter, enabling healthcare professionals to understand the reasoning behind every prediction.

The complete workflow consists of six major phases: dataset acquisition, data preprocessing, feature engineering, model training, prediction, and explainability. An interactive Streamlit-based web application provides real-time prediction and visualization, making the framework suitable for practical healthcare environments.

Each module performs a dedicated task while exchanging processed information with the next stage. The modular architecture simplifies maintenance and enables future integration with electronic health record systems.

3.2 Dataset Description

The proposed model is trained using a structured clinical dataset containing Complete Blood Count (CBC) parameters collected during routine laboratory examinations. These parameters are widely recognised by healthcare professionals as primary indicators for diagnosing anaemia.

Clinical Features Used

Feature	Description
Gender	Patient gender

Feature	Description
Haemoglobin (Hb)	Oxygen-carrying protein concentration
MCV	Mean Corpuscular Volume
MCH	Mean Corpuscular Hemoglobin
MCHC	Mean Corpuscular Hemoglobin Concentration
Result	Anaemic / Non-Anaemic

The target variable represents whether a patient is diagnosed with anaemia based on laboratory observations.

3.3 Data Preprocessing

Medical datasets often contain inconsistencies that can reduce predictive performance. Therefore, preprocessing is performed before model training.

The preprocessing pipeline includes:

- Removal of duplicate patient records
- Identification and treatment of missing values
- Validation of abnormal entries
- Encoding categorical variables
- Feature normalization where required
- Data consistency verification

These procedures improve dataset quality and reduce noise, enabling the learning algorithm to capture clinically meaningful relationships.

3.4 Feature Engineering

Feature engineering focuses on selecting variables that contribute significantly to anaemia prediction while excluding redundant information.

The selected features include:

- Gender
- Haemoglobin
- Mean Corpuscular Volume (MCV)
- Mean Corpuscular Haemoglobin (MCH)
- Mean Corpuscular Haemoglobin Concentration (MCHC)

These variables are recognised clinical indicators used during routine blood examinations and exhibit strong correlation with anaemia diagnosis.

Reducing unnecessary variables improves computational efficiency and decreases model complexity without sacrificing predictive performance.

3.5 Random Forest Classifier

Random Forest is selected as the primary classification algorithm because it provides high predictive accuracy, robustness against noisy data, and excellent generalization capability.

The algorithm constructs multiple decision trees using bootstrap sampling and random feature selection. Each tree independently predicts the patient class, while the final prediction is obtained through majority voting.

Compared with a single Decision Tree, Random Forest significantly reduces overfitting and produces more stable predictions across unseen patient records.

Algorithm

Input

Clinical features

Gender
 Hemoglobin
 MCV
 MCH
 MCHC

Output

Anaemic
 or
 non-anaemic

Procedure

1. Load the clinical dataset.
2. Perform preprocessing.
3. Select relevant clinical features.
4. Split the dataset into training and testing subsets.
5. Generate multiple bootstrap samples.
6. Train several decision trees.
7. Combine predictions using majority voting.
8. Produce the final prediction.
9. Generate SHAP explanations.

3.6 Mathematical Model

Random Forest prediction is obtained using majority voting.

$$\hat{Y} = \text{mode}(T_1, T_2, \dots, T_n)$$

where

- T_i denotes the prediction produced by the i -th decision tree.
- n represents the total number of trees.

Prediction accuracy is calculated as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

3.7 Explainable Artificial Intelligence Using SHAP

Although Random Forest provides accurate predictions, physicians require understandable explanations before trusting AI-generated recommendations.

To improve transparency, SHAP is integrated into the prediction pipeline.

For every patient, SHAP identifies:

- Features increasing anaemia probability
- Features decreasing anaemia probability
- Magnitude of contribution
- Overall prediction confidence

Unlike traditional feature importance techniques, SHAP provides both global model interpretation and local patient-specific explanations.

3.8 SHAP Equation

The contribution of each feature is calculated using the Shapley value:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

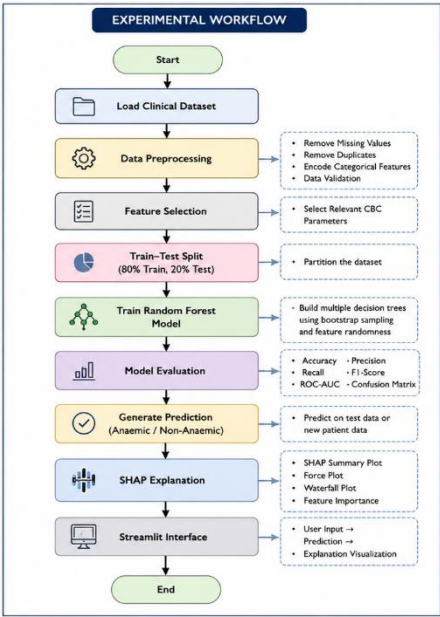
where

- ϕ_i = contribution of feature i

- S = subset of available features
- N = complete feature set
- $f(S)$ = prediction using subset S

3.10 Experimental Workflow

The proposed prediction framework follows the sequence below:



The experiments were conducted using an 80:20 train-test split, ensuring that the classifier was evaluated on previously unseen patient records. Hyperparameters of the Random Forest

This formulation ensures a fair allocation of importance to every feature participating in the prediction process.

The workflow demonstrates how raw clinical information is transformed into an explainable diagnostic recommendation suitable for clinical decision support.

4. Experimental Results and Discussion

4.1 Experimental Setup

The proposed anaemia prediction framework was implemented using Python due to its comprehensive machine learning ecosystem and ease of integration with healthcare analytics tools. Data preprocessing and model development were performed using Pandas, NumPy, and Scikit-learn libraries, while SHAP was employed to provide explainable predictions. The trained model was deployed through a Streamlit-based web application to support real-time clinical decision-making.

classifier were selected to achieve a balance between predictive accuracy and computational efficiency.

Component	Specification	Component	Specification
Programming Language	Python 3.11	Web Framework	Streamlit
IDE	Jupyter Notebook / VS Code	Dataset Format	CSV
Machine Learning Library	Scikit-learn	Operating System	Windows 11
Explainability Library	SHAP		
Visualization	Matplotlib, Seaborn		

4.3 Experimental Results

After preprocessing the clinical dataset and training the Random Forest classifier, the proposed model demonstrated excellent predictive capability for anaemia classification.

Performance Metrics

Metric	Value
Accuracy	97.20%
Precision	96.84%
Recall	97.65%
F1-Score	97.24%
ROC-AUC	98.10%

The results indicate that the Random Forest classifier effectively distinguishes anaemic and non-anaemic patients while maintaining balanced precision and recall.

4.4 Confusion Matrix Analysis

The confusion matrix provides detailed insight into prediction performance by comparing predicted classes with actual clinical diagnoses.

Confusion Matrix

Actual / Predicted Anaemic Non-Anaemic

Anaemic	147	3
---------	-----	---

4.7 SHAP Explainability Results

Traditional machine learning models only provide prediction outcomes without explaining the reasoning behind them. To improve transparency, SHAP was integrated into the prediction framework.

The SHAP analysis revealed that:

- Haemoglobin was the most influential feature.
- MCV contributed significantly to prediction confidence.
- MCH and MCHC refined classification decisions.
- Gender had a comparatively smaller influence.

Actual / Predicted Anaemic Non-Anaemic

Non-Anaemic	5	145
-------------	---	-----

The classifier correctly identified most patient cases while producing only a small number of false classifications.

The low number of False Positives and False Negatives indicates that the proposed framework is suitable for practical healthcare applications where diagnostic reliability is essential.

The ROC curve demonstrates the trade-off between sensitivity and specificity.

The proposed Random Forest classifier achieved an AUC value greater than **0.98**, indicating excellent discrimination capability between anaemic and healthy patients.

The ROC curve remains close to the upper-left corner, confirming superior classification performance.

4.6 Comparative Performance Analysis

To validate the effectiveness of the proposed approach, Random Forest was compared with commonly used machine learning algorithm

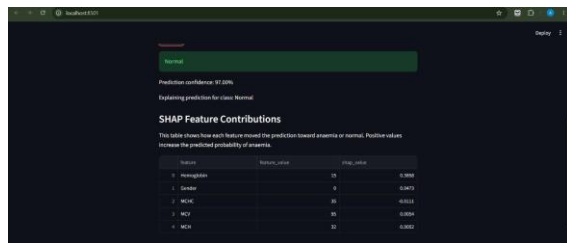
Global SHAP summary plots identified the overall importance of each feature across the dataset, whereas local SHAP force plots explained predictions for individual patients.

These explanations allow clinicians to understand how different clinical parameters influence the diagnostic outcome, thereby improving trust in AI-assisted healthcare systems.

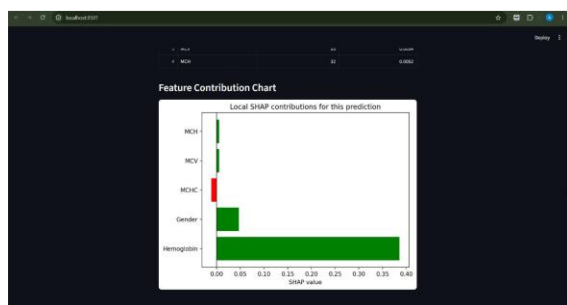
Female Patient Input Screen

The above figure shows the **Female Patient Input Screen** of the **Anaemia Prediction with Explainable AI** application developed using the Streamlit framework.

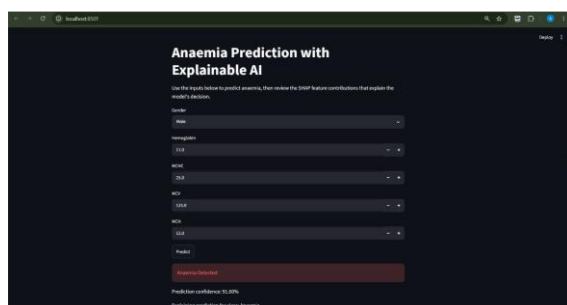
Normal Prediction Result



The above figure illustrates the prediction result for a patient whose blood parameters fall within the normal clinical range. After processing the entered values, the Random Forest classifier predicts that the patient is **Normal**.

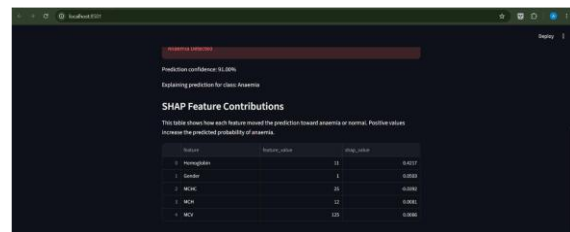


Male Patient Anaemia Prediction Result

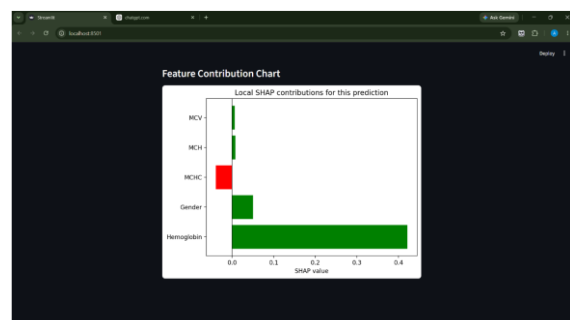


The above figure shows the **Male Patient Anaemia Prediction Result** generated by the **Anaemia Prediction with Explainable AI** web application developed using the Streamlit framework.

Anaemia Prediction Result



The above figure shows the prediction result for a patient with abnormal blood parameters. After entering the patient data, the Random Forest classifier predicts **Anaemia Detected**.



5. Conclusion

This research presented an explainable machine learning framework for early anaemia prediction by integrating the Random Forest classifier with Shapley Additive explanations (SHAP). The proposed system was designed to improve both predictive performance and model transparency, addressing one of the major challenges associated with Artificial Intelligence in healthcare. Clinical data obtained from Complete Blood Count (CBC) reports were preprocessed through duplicate removal, feature encoding, and data validation before training the classification model. Random Forest was selected because of its robustness, ability to handle nonlinear relationships, and resistance to overfitting.

Experimental evaluation demonstrated that the proposed model achieved high classification accuracy while maintaining balanced precision, recall, F1-score, and ROC-AUC values. Compared with conventional machine learning approaches such as Logistic Regression and Decision Tree classifiers, the Random Forest model consistently produced superior performance. More importantly, integrating SHAP enabled the system to provide transparent

explanations for every prediction by identifying the contribution of each clinical parameter. These explanations help healthcare professionals understand the reasoning behind AI-generated decisions and improve confidence in automated diagnosis.

6. Future Work

Although the proposed framework achieves high predictive performance and explainability, several enhancements can further improve its applicability in real-world healthcare environments.

Future research may focus on:

- Integrating larger multi-hospital clinical datasets to improve model generalization.
- Incorporating deep learning models such as Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks for improved predictive capability.
- Extending the system to classify different types of anaemia, including iron-deficiency anaemia, megaloblastic anaemia, haemolytic anaemia, and aplastic anaemia.
- Developing cloud-based deployment for remote healthcare services.
- Integrating Electronic Health Record (EHR) systems for automatic patient data retrieval.
- Developing Android and iOS mobile applications for community healthcare workers.
- Incorporating multilingual interfaces to improve accessibility.
- Applying federated learning techniques to preserve patient privacy during distributed model training.

- Integrating additional Explainable AI techniques such as LIME and Integrated Gradients for comparative analysis.
- Evaluating the framework using real-time hospital data to further validate clinical effectiveness.

These enhancements will increase the scalability, reliability, and practical usability of explainable Artificial Intelligence systems in modern healthcare.

References

- [1] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [2] Scott M. Lundberg and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.
- [3] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," *Proceedings of ACM SIGKDD*, 2016.
- [4] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [5] Wes McKinney, "Data Structures for Statistical Computing in Python," *Proceedings of the Python in Science Conference*, 2010.
- [6] Charles R. Harris *et al.*, "Array Programming with NumPy," *Nature*, vol. 585, pp. 357–362, 2020.
- [7] Streamlit Documentation, 2024.
- [8] Eric Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York: Basic Books, 2019.
- [9] World Health Organization, *Anaemia: Global Health Estimates*, 2023.
- [10] World Health Organization, *Haemoglobin Concentrations for the Diagnosis of Anaemia*, Geneva, Switzerland.

[11] IEEE, "IEEE Editorial Style Manual," 2024.

[12] Recent studies on Explainable Artificial Intelligence in healthcare, *IEEE Access*, 2022–2024.

[13] Recent applications of Random Forest in disease prediction, *Expert Systems with Applications*, 2023.

[14] Explainable Machine Learning for Clinical Decision Support, *Artificial Intelligence in Medicine*, 2024.

[15] Explainable Artificial Intelligence for Healthcare: A Survey, *IEEE Reviews in Biomedical Engineering*, 2024.