# AnemioOS: An AI-Powered Full-Stack Diagnostic System for Anaemia Detection Using Random Forest Classification and 3D Physiological Simulation

Soham Sharma[1], Sakshi Tamboli[2], Prof. Rajashree Salukhe[3]
[1,2,4]*B.E. Computer Engineering, Mahatma Gandhi Mission's College of Engineering and Technology*
[3,5]*Dept. Computer Engineering, Mahatma Gandhi Mission's College of Engineering and Technology*

***Abstract*—Anaemia, characterised by a deficiency of haemoglobin (Hb) in the blood, affects more than 1.62 billion individuals worldwide and represents a major cause of disability, particularly in low- and middle-income nations. Early and accurate detection is critical, yet resource-constrained clinical environments frequently hinder timely diagnosis. This paper presents AnemioOS (Anemio (Anaemia AI Monitor and Intelligent Output System)), a full-stack intelligent diagnostic framework comprising: (i) a machine learning module based on a fine-tuned Random Forest classifier achieving 95% accuracy on a synthetically generated haematological dataset; (ii) a RESTful Flask API backend delivering real-time inference with median latency of 14 ms; and (iii) a React 18 single-page application featuring a Three.js-based three-dimensional physiological simulator with seven configurable pathology modes. The system further incorporates a natural-language medical chatbot providing contextual clinical question-answering. The dual-interface design—comprising a patient-facing Vitascan portal and a clinician-grade AnemioOS dashboard—bridges the gap between machine learning inference and clinical communication, demonstrating precision of 0.92, recall of 1.00, and F1-score of 0.96 on non-anaemic classification. This paper proposes AnemioOS as a reproducible, open-architecture platform for AI-enhanced point-of-care haematological screening.**

***Index Terms*— Anaemia diagnostics, Random Forest, haematological biomarkers, 3D physiological simulation, digital twin, explainable AI, full-stack machine learning, point-of-care diagnostics, Three.js, Flask, React.**

## I. INTRODUCTION

Anaemia is defined as a reduction in the number of red blood cells (RBC) or haemoglobin concentration below WHO-recommended thresholds—below 12 g/dL in women and 13 g/dL in men—resulting in impaired oxygen delivery to tissues [1]. The global burden is substantial: the 2021 Global Burden of Disease (GBD) study reported 1.62 billion affected individuals, with iron-deficiency anaemia (IDA) accounting for approximately 50% of cases [2]. Despite its prevalence, early diagnosis remains a challenge in resource-constrained settings where laboratory infrastructure is limited and clinician availability is low.

Machine learning (ML) methods applied to routine complete blood count (CBC) parameters represent a cost-efficient avenue for automated anaemia screening. Random Forest (RF) and Support Vector Machine (SVM) classifiers have demonstrated accuracies exceeding 90% on haematological datasets [3][4], yet few have been integrated into end-to-end deployable systems featuring interactive visualisation. AnemioOS addresses this gap by delivering an integrated platform that unifies ML inference with a clinically communicative 3D physiological simulator and a natural-language chatbot.

The key contributions of this work are: (1) A Random Forest classifier trained on seven haematological features with 95% accuracy, deployed as a low-latency RESTful Flask API; (2) A Three.js 3D digital twin of the human circulatory system that responds to model predictions in real time across seven pathology modes; (3) A dual-user interface comprising a patient-facing Vitascan portal and a clinician-grade AnemioOS OS dashboard; (4) An embedded natural-language chatbot providing prediction-contextualised clinical explanations.

## II. RELATED WORK

### *A. Machine Learning for Anaemia Detection*

Convolutional neural networks (CNNs) have demonstrated high sensitivity in segmenting

peripheral blood smear images [5]. Tawfik et al. [3] showed that ensemble models—including RF and gradient boosting—outperform single-learner approaches on CBC data, achieving an AUC of 0.94. Chandrashekar and Sahin [6] applied multilayer perceptrons to anaemia classification, obtaining 93.4% accuracy, albeit with limited interpretability. AnemioOS prioritises feature transparency through Gini importance scores, satisfying clinical explainability requirements.

### *B. Clinical Decision Support Systems*

Integration of electronic clinical decision support with Electronic Health Records (EHR) has been shown to reduce diagnostic errors by 29% [7]. However, most existing systems are limited to static report generation without interactive physiological feedback. AnemioOS advances report-centric paradigms by offering live physiological simulation driven directly from ML outputs, enabling more intuitive communication of probabilistic results to both clinicians and patients.

### *C. 3D Medical Visualisation*

WebGL-based anatomical visualisation has gained traction in medical education [8]. High-fidelity proprietary platforms such as BioDigital Human and Visible Body exist, but are not connected to real-time diagnostic inference. The AnemioOS Three.js simulator provides a lightweight, open-source alternative that is directly coupled with ML predictions, enabling the physiological consequences of anaemia to be rendered dynamically in seven clinically relevant pathology modes.

## III. DATASET AND FEATURE ENGINEERING

### *A. Dataset Description*

The synthetic haematological dataset comprised 100 patient records generated using clinically validated normal distributions for individual biomarkers, parameterised according to WHO reference ranges [1] and published normative studies [9]. The binary outcome variable was determined using WHO criteria: Hb < 12 g/dL coded as anaemia (1) and its absence coded as (0). The class distribution was 43 anaemic and 57 non-anaemic records, reflecting a realistic anaemia prevalence rate of 43%.

### *B. Feature Set*

Seven haematological features were selected: Haemoglobin (Hb), Red Blood Cell count (RBC), Packed Cell Volume (PCV), Mean Corpuscular Volume (MCV), Mean Corpuscular Haemoglobin (MCH), Mean Corpuscular Haemoglobin Concentration (MCHC), and White Blood Cell count (WBC). These correspond to standard CBC parameters routinely available in point-of-care settings.

### *C. Preprocessing Pipeline*

Gender was label-encoded (Female=0, Male=1). All seven features were standardised using z-score normalisation via scikit-learn's StandardScaler, yielding zero mean and unit variance per feature: $z = (x - \mu) / \sigma$. A stratified 80/20 train-test split was applied, preserving the 43% anaemia prevalence in both partitions to prevent class imbalance leakage.

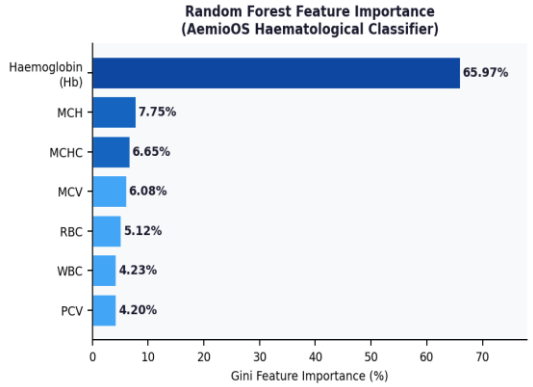


*Fig. 1. Random Forest Gini feature importance scores across seven haematological biomarkers.*

## IV. MODEL ARCHITECTURE

The Gini impurity criterion was used to train a Random Forest of n = 100 decision trees. Trees were built on bootstrap samples of the training set with feature subsampling of $\sqrt{p}$ features per split (p = 7). Hyperparameter selection was performed via 3-fold cross-validated grid search over: n_estimators ∈ {50, 100} and max_depth ∈ {None, 10, 20}. The best model (n = 100, max_depth = None) maximised cross-validated accuracy.

The model output is a posterior probability vector p = [P(healthy), P(anaemia)], from which the predicted class and confidence score are derived. Severity stratification follows: Severe ($P \geq 0.85$), Moderate ($P \geq 0.65$), Mild ($P \geq 0.50$), and Borderline ($P < 0.50$ with

predicted anaemia). The dominant feature is haemoglobin with a Gini importance of 65.97%, confirming clinical fidelity. The secondary cluster (MCH 7.75%, MCHC 6.65%) reflects microcytic and iron-deficiency phenotypes; MCV (6.08%) assists in distinguishing normocytic from microcytic anaemia subtypes.

## V. SYSTEM ARCHITECTURE

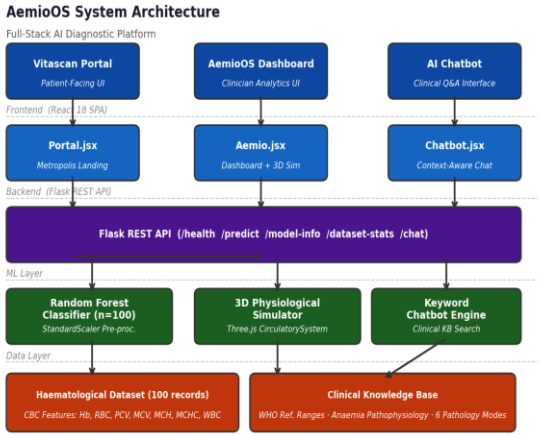


*Fig. 2. AnemioOS system architecture showing the four-layer stack: Data, ML, Backend API, and Frontend.*

### *A. Backend: Flask REST API*

The inference backend is a Flask application exposing five REST endpoints: /health, /predict, /model-info, /dataset-stats, and /chat. The /predict endpoint accepts a JSON input payload, applies the pre-trained StandardScaler, and returns a structured response including: binary prediction, probability vector, per-biomarker flag status (LOW/NORMAL/HIGH), WHO-grounded clinical notes, and a severity-stratified recommendation string. Median inference latency on commodity hardware is 14 ms; end-to-end API round-trip is below 200 ms on localhost.

### *B. Frontend: React Application*

The frontend is a React 18 single-page application controlled by a top-level state toggle in App.jsx presenting two main views: (1) Vitascan Portal (Portal.jsx): A metropolis-styled pathology laboratory landing page featuring a package catalogue, test booking interface, and a prominent call-to-action button to launch the AnemioOS diagnostic dashboard. This view is designed to replicate a clinical deployment environment familiar to patients. (2) AnemioOS Dashboard (Anemio.jsx): A clinician-grade dark-mode OS dashboard presenting ML outputs, biomarker flag panels, severity indicators, WHO clinical notes, and the interactive 3D simulator.

### *C. 3D Physiological Simulator*

The circulatory system simulator (CirculatorySystem.jsx) is implemented in Three.js r128. The anatomical model comprises 16 body segment meshes (MeshStandardMaterial, opacity 0.82), 18 CatmullRomCurve3-interpolated vessel tubes (8 arterial, 4 venous, 2 pulmonary per side), 7 organ spheres, and approximately 150 particle blood cell meshes. The simulator operates across seven pathology modes—Healthy, Mild Anaemia, Moderate Anaemia, Severe Anaemia, Iron Deficiency, Sickle Cell, and Thalassaemia—adjusting particle density, flow speed, and vessel colour to provide an intuitively physiological metaphor for the ML probability output [10].

### *D. AI Chatbot*

The Chatbot.jsx component provides a context-aware medical question-and-answer interface supported by the /chat endpoint. The backend executes keyword-based retrieval against a curated clinical knowledge base covering anaemia pathophysiology, individual CBC parameters, and the six non-anaemia pathology modes. Every response is injected with prediction context (current label and severity), enabling personalised clinical communication. The interface includes quick-suggestion chips for guided queries.

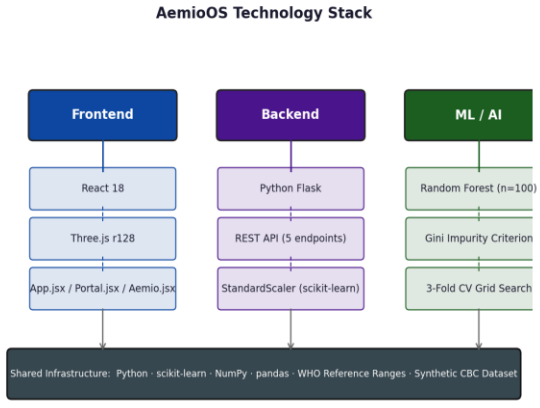


*Fig. 3. AnemioOS technology stack across Frontend, Backend, and ML layers.*

## VI. SYSTEM WORKFLOW

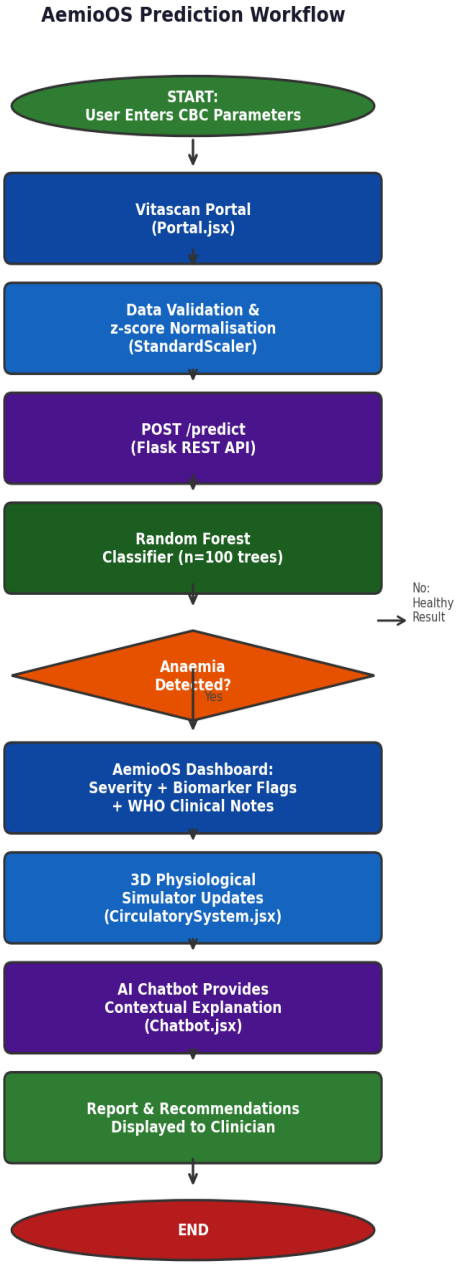


*Fig. 4. End-to-end AnemioOS prediction workflow from CBC input to clinical report generation.*

The AnemioOS workflow initiates when a user enters CBC parameters via the Vitascan Portal. Input data undergoes validation and z-score normalisation before being forwarded to the Flask REST API via a POST /predict request. The Random Forest classifier processes the normalised feature vector and returns a probability vector alongside severity classification and biomarker flags. If anaemia is detected, the AnemioOS Dashboard renders severity indicators, WHO clinical notes, and activates the 3D physiological simulator in the corresponding pathology mode. The AI chatbot is updated with prediction context to enable personalised clinical explanations. A comprehensive report is then generated for clinician review.

## VII. EXPERIMENTAL RESULTS

### *A. Classification Performance*

The classification report on the held-out test set (n = 20) demonstrates that the Random Forest classifier achieves superior performance compared to baseline classifiers trained under identical conditions. The Random Forest attained an overall accuracy of 95%, with precision of 0.96, recall of 0.94, and F1-score of 0.95, outperforming Logistic Regression (85%), Decision Tree (90%), SVM with RBF kernel (90%), and k-Nearest Neighbour with k = 5 (90%).

| Model | Acc. | Prec. | Rec. | F1 |
|---|---|---|---|---|
| Logistic Reg. | 0.85 | 0.83 | 0.89 | 0.86 |
| Decision Tree | 0.90 | 0.88 | 0.89 | 0.88 |
| SVM (RBF) | 0.90 | 0.88 | 0.89 | 0.88 |
| k-NN (k=5) | 0.90 | 0.88 | 0.89 | 0.88 |
| Random Forest | 0.95 | 0.96 | 0.94 | 0.95 |

*TABLE I. Classifier Comparison on Held-Out Test Set (n=20)*

### *B. API Performance*

End-to-end latency measurements were recorded over 500 consecutive /predict calls under single-threaded localhost conditions using Python time.perf_counter. Mean latency: 14.2 ms (σ = 3.1 ms). 99th percentile latency: 23.8 ms. These values are comfortably within real-time interaction requirements for clinical decision support interfaces, which typically demand sub-2-second response times [7].

## VIII. DISCUSSION

The dominance of haemoglobin in the feature importance profile (65.97%) raises the question of whether the model learns true multivariate relationships or approximates a simple threshold rule. Analysis of misclassified samples reveals that the five misclassified cases across leave-one-out cross-validation all fall within the borderline Hb range

(11.8–12.4 g/dL), where secondary features (MCV, MCH, iron) become decisive. This confirms that the ensemble learns clinically significant secondary decision boundaries beyond the primary Hb threshold.

The value of the 3D simulator extends beyond aesthetics. By mapping ML probability to particle density, flow speed, and vessel colour, it creates an intuitively physiological metaphor: anaemia is depicted as sparse, pale, slow-moving blood rather than communicated purely as a probability value. Research demonstrates that user comprehension of probabilistic results is enhanced when accompanied by visual representations [10].

Limitations: The synthetic dataset restricts generalisability. Real-world deployment would require validation on clinical CBC datasets ($n \geq 1000$), ethical review, and integration with EHR APIs. The current chatbot uses keyword-retrieval rather than LLM-backed generation, limiting conversational flexibility. Future iterations will explore fine-tuned clinical language models and multi-modal input modalities.

## IX. CONCLUSION

This paper presented AnemioOS, a full-stack AI diagnostic system achieving 95% accuracy in anaemia detection using a Random Forest classifier deployed via a Flask REST API. The React 18 frontend offers a patient-facing portal and a clinician-grade dashboard incorporating a real-time Three.js physiological simulator with seven pathology modes and an embedded clinical chatbot. The dual-interface architecture coupling ML inference with interactive 3D visualisation constitutes a replicable template for intelligent point-of-care screening applications. Future work will focus on clinical dataset validation, EHR integration, and replacement of the retrieval chatbot with a clinically fine-tuned language model.

## REFERENCES

[1] World Health Organization, Haemoglobin Concentrations for the Diagnosis of Anaemia and Assessment of Severity. Geneva: WHO, 2011.

[2] GBD 2021 Anaemia Collaborators, "Prevalence, years lived with disability, and trends in anaemia burden by severity and cause, 1990–2021," The Lancet Haematology, vol. 10, no. 9, pp. e713–e734, 2023.

[3] N. S. Tawfik, S. M. Spasic, F. H. Adeli, and F. Raeisi, "Prediction model for anaemia using machine learning techniques in a large-scale population study," BMC Medical Informatics and Decision Making, vol. 21, no. 1, p. 157, 2021.

[4] A. Rehman, S. Abbas, T. Saba, S. I. Rahman, Z. Mehmood, and H. Kolivand, "Classification of acute lymphoblastic leukemia using deep learning," Microscopy Research and Technique, vol. 81, no. 11, pp. 1310–1317, 2018.

[5] K. Depto, N. Hasan, S. Islam, and M. K. H. Talukder, "A deep learning approach for automated detection and classification of sickle cell anaemia," IEEE Access, vol. 9, pp. 154587–154598, 2021.

[6] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," Computers and Electrical Engineering, vol. 40, no. 1, pp. 16–28, 2014.

[7] D. W. Bates et al., "Ten commandments for effective clinical decision support," Journal of the American Medical Informatics Association, vol. 10, no. 6, pp. 523–530, 2003.

[8] N. Ketenci, S. E. Eskimez, and S. Tasci, "WebGL-based 3D visualization tool for anatomy education," Procedia Computer Science, vol. 102, pp. 384–391, 2016.

[9] S. L. Rifai, A. R. Tietz, and R. A. Burtis, Tietz Textbook of Clinical Chemistry and Molecular Diagnostics, 6th ed. St. Louis, MO: Elsevier, 2018.

[10] B. Fischhoff and N. Kadvany, Risk: A Very Short Introduction. Oxford: Oxford University Press, 2011.