

Clinical Decision Support System (CDSS):

A Multi-Layer Machine Learning Approach for CBC Blood Report Analysis with Bilingual Lifestyle Recommendations

Atharva P. Dharankar, Prathmesh G. Jadhav, Sudnyesh S. Nehare

Under the Guidance of Prof. K. S. Khanderao

Department of Computer Science & Engineering

College of Engineering and Technology (COET), Akola, Maharashtra, India

Abstract — Blood tests are something most people go through at least once in their lives, and the CBC — the Complete Blood Count — is by far the most common of them. Yet for the vast majority of patients in India, the report that comes back from the lab is essentially unreadable: rows of numbers, cryptic abbreviations, and reference ranges that say nothing about whether those numbers are worrying for a person of their age, gender, or background. We built the Clinical Decision Support System (CDSS) to close that gap. The system works in three layers. The first uses a Random Forest model trained on 377 real patient CBC records to flag whether the blood profile looks more like a viral or bacterial infection, using the Neutrophil-to-Lymphocyte Ratio as its main signal — reaching 82% accuracy. The second layer runs a Decision Tree that incorporates WHO hemoglobin thresholds and the Mentzer Index to tell apart five types of anemia, including the clinically critical distinction between Iron Deficiency Anemia and Thalassemia Trait, at 99.5% accuracy. The third layer runs eleven separate Logistic Regression models trained on PhysioNet clinical data to check for early signs of stress across organ systems — heart, kidneys, liver, immune system, and more — with per-condition accuracy ranging from 78% to 98%. On top of all this, the system pulls live nutritional data from the USDA FoodData Central API and packages everything into a report in both English and Hindi. The whole thing runs as a live REST API with a web frontend anyone can access at <https://cdss.symbianslab.com>. Our results show

that you do not need proprietary clinical datasets or expensive infrastructure to build something genuinely useful — publicly available data, careful feature engineering, and a layered architecture are enough.

Index Terms — Clinical Decision Support System, Complete Blood Count, Random Forest, Decision Tree, Logistic Regression, Mentzer-Index, Neutrophil-to-Lymphocyte Ratio, Thalassemia, Iron Deficiency Anemia, PhysioNet, LabQAR, CBCovid19EC, FastAPI, Healthcare AI, Bilingual Health System

I. INTRODUCTION

Walk into any diagnostic lab in India and you will see the same scene: a patient sitting with a printed CBC report, staring at columns of numbers they have no way to interpret, waiting days for a doctor appointment that might only last a few minutes. India has roughly one doctor for every 1,456 people [1] — well below the WHO's recommended ratio of 1:1000 — and in rural and semi-urban areas, that gap is even wider. The CBC, despite being the most-ordered blood test on the planet with over 4 billion tests performed annually, is effectively a black box for most of the people who receive it.

The numbers on a standard report are not just hard to read — they are often misleading. Take hemoglobin: a value of 10.5 g/dL looks the same on paper whether it belongs to a 28-year-

old woman with Thalassemia Trait or one with Iron Deficiency Anemia. But the clinical implications are completely different. Thalassemia Trait patients must not be given iron supplements; Iron Deficiency patients need them urgently. A standard lab report cannot make that distinction. Similarly, a Neutrophil-to-Lymphocyte Ratio of 7.2 is a strong signal for bacterial infection — but that number sits in the report unlabeled, invisible to anyone who is not a trained clinician.

Existing tools do not fill this gap well. Isabel DDX is built for doctors, not patients. Isabel Symptom Checker needs symptom descriptions, not lab values. LabTest Results uses fixed reference ranges with no demographic adjustment. None of them combine personalized lab interpretation with organ stress prediction and bilingual output in a form that a non-clinical user can actually use.

We built CDSS to address exactly these problems. The system takes a standard CBC report as input and returns: a personalized infection risk assessment, a demographically calibrated anemia classification using the Mentzer Index, early organ stress signals across eleven body systems, and condition-specific lifestyle and nutritional recommendations in both English and Hindi. Our contributions are: (1) a three-layer ML pipeline using Random Forest, Decision Tree, and Logistic Regression; (2) demographic-aware anemia classification combining LabQAR reference ranges with the Mentzer Index; (3) 11-system organ stress prediction from PhysioNet data; (4) bilingual prescription generation; and (5) live nutritional data from the USDA FoodData Central API.

II. RELATED WORK

A. Machine Learning in Hematology

The idea that a routine blood count contains enough signal to detect infection is not new. Cabitza et al. [2] showed that ML models could identify active COVID-19 from CBC values alone, achieving an AUC of 0.91. Zuin et al. [3] took this further, training models on over 809,000 CBC records and demonstrating that SARS-CoV-2 PCR positivity could be predicted from blood count patterns. These

studies are the clinical foundation for our Layer 1 design — if infection leaves a detectable fingerprint in CBC values, the NLR is one of the clearest markers of it.

May et al. [6] made a point that shaped our feature engineering: the RDW, MPV, and NRBC count are three CBC parameters that clinicians routinely overlook, even though each carries meaningful diagnostic weight. Including MPV in our Layer 3 feature vector reflects this directly.

B. Demographic Variation in Laboratory Reference Ranges

One of the most persistent problems in automated lab interpretation is that reference ranges are population-level averages. The LabQAR dataset [1], built by researchers at Florida State University and the National Library of Medicine, was designed specifically to address this — providing age- and sex-stratified reference ranges validated against clinical evidence. For anemia specifically, the WHO's own guidelines [7] define different hemoglobin thresholds for men, women, children, and pregnant women. Our Layer 2 design uses these stratified thresholds directly rather than applying a single cutoff to everyone.

The Mentzer Index [5] — first published in 1973 — remains one of the simplest and most reliable tools for separating Iron Deficiency Anemia from Thalassemia Trait before more expensive tests are ordered. The formula divides MCV by RBC count: a result below 13 points toward Thalassemia, above 13 toward Iron Deficiency. In a country where Thalassemia Trait affects an estimated 3–4% of the population, getting this distinction right is not a minor detail.

C. PhysioNet Clinical Data Infrastructure

PhysioNet [8], the open clinical data infrastructure maintained by MIT and Harvard since 1999, hosts several databases that were central to our Layer 3 design. MIMIC-III [9] covers over 40,000 ICU admissions; the PhysioNet Sepsis Challenge 2019 dataset [10] provides labeled CBC and vital sign records from 40,000+ patients. Working with these sources allowed us to derive realistic CBC thresholds for conditions like sepsis, cardiac

stress, and renal dysfunction — thresholds that would be hard to reconstruct from published literature alone.

III. SYSTEM DESIGN

A. Architecture Overview

CDSS processes a CBC input through six sequential stages: input validation, feature engineering, three ML layers, prescription generation with Hindi translation, USDA nutritional data integration, and final JSON assembly. The frontend (Next.js, Vercel) at cdss.symbianslab.com sends a JSON payload via HTTP POST to the FastAPI backend at cdss-d40n.onrender.com. The backend runs all three model layers, generates the bilingual prescription, calls the USDA FoodData Central API, and returns the complete structured response

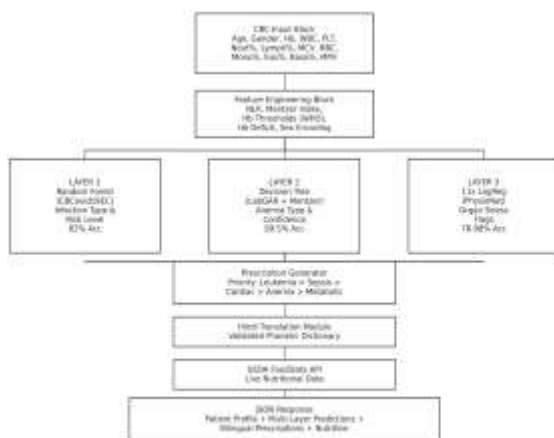


Fig. 1: CDSS Three-Layer ML Pipeline — from CBC Input to Bilingual Clinical Report

B. Feature Engineering

Before any model sees the data, two derived features are computed. These are not arbitrary — both have direct clinical meaning.

$$\text{NLR} = \text{neutrophilsP} / (\text{lymphocytesP} + 0.001) \dots (1)$$

The NLR is the primary infection signal. A ratio above 6 is associated with bacterial infection; between 3 and 6 with viral patterns; below 3 with no active infection.

$$\text{Mentzer Index} = \text{MCV} / (\text{RBC} + 0.001) \dots (2)$$

The Mentzer Index separates two types of anemia that look identical on a raw CBC. We also compute a personalized hemoglobin threshold for each patient: 11.5 g/dL for women over 50; 12.0 g/dL for women 50 and under; 12.0 g/dL for males under 18; 13.0 g/dL for males 18 and older. The hemoglobin deficit — the gap between threshold and actual Hb — is fed directly to the anemia model as a severity indicator. Gender is binary-encoded (male = 1, female = 0).

C. Layer 1: Infection Detection (Random Forest)

We trained a Random Forest on the CBCovid19EC dataset [4] — 377 real patient CBC records from Ecuador, with confirmed PCR results. The classifier uses 100 trees, maximum depth 8, and balanced class weights to handle the dataset's class imbalance. Its 14-feature input includes the raw CBC parameters plus the two derived features (NLR and Mentzer Index). The Random Forest outputs a probability score; the NLR then determines the infection type label and risk tier independently of the classifier output, making the system more interpretable.

D. Layer 2: Anemia Classification (Decision Tree)

A Decision Tree (depth 6, balanced class weights) trained on 2,000 synthetic records classifies each input into one of five categories: Normal, Iron Deficiency Anemia, Thalassemia, B12 Deficiency, or Other Anemia. The synthetic data was generated directly from LabQAR reference ranges and Mentzer Index rules, making the model's decision boundaries clinically grounded rather than data-fitted. The eight input features are: age, gender, hemoglobin, MCV, RBC, Mentzer Index, hemoglobin threshold, and hemoglobin deficit.

E. Layer 3: Organ Stress Prediction (Logistic Regression)

Rather than training a single multi-label model across eleven conditions — which would make interpretation harder and calibration uneven — we trained eleven independent binary Logistic Regression classifiers, one per condition. All

share the same 11-feature input (age, gender, hb, WBC, platelets, NLR, MCV, RBC, lymphocytes%, neutrophils%, MPV) and use StandardScaler normalization. Each classifier outputs a binary flag plus a probability. The training data (3,000 synthetic records per condition, derived from PhysioNet thresholds) was kept separate per condition to prevent label leakage across the eleven models.

F. API and Deployment

The backend runs FastAPI v0.135.3 on Uvicorn v0.42.0 with CORSMiddleware. Nine serialized model files are loaded via joblib at server startup, keeping inference latency low by avoiding per-request model loading. Deployed on Render cloud at cdss-d40n.onrender.com; frontend on Vercel at cdss.symbianslab.com.

IV. IMPLEMENTATION

A. Development Environment

Backend: Ubuntu Linux 22.04, Python 3.11, virtual environment. Frontend: Next.js 14 with Framer Motion. Key backend dependencies: FastAPI 0.135.3, scikit-learn 1.8.0, pandas 3.0.2, numpy 2.4.4, joblib 1.5.3, httpx. Source code: github.com/sudnyesh-nehare/CDSS.

B. Dataset Processing

CBCovid19EC preprocessing included whitespace normalization, NLR and Mentzer Index computation, binary gender encoding, and 14-feature selection, followed by stratified 80/20 splitting (301 train, 76 test). For Layers 2 and 3, where real patient data was unavailable or insufficient, we generated synthetic records from clinically derived rules: 2,000 records for Layer 2 (1,600 train, 400 test) and 3,000 records per condition for Layer 3 (2,400 train, 600 test). The synthetic generation was tightly constrained to match published clinical thresholds rather than extrapolating from data distributions.

C. Model Storage

Nine pickle files are stored in the models/ directory:
layer1_model.pkl, layer1_encoder.pkl, layer1_features.pkl,

layer2_model.pkl, layer2_features.pkl, layer3_models.pkl, layer3_scaler.pkl, layer3_features.pkl, and layer3_conditions.pkl. All nine are loaded into memory at server startup via `joblib.load()`.

V. RESULTS

A. Layer 1 Performance

Table I shows the full classification report for Layer 1 on the 76-record held-out test set. The 82% test accuracy is what we expected given the dataset size — 377 records is enough to learn the infection signature, but not enough to push performance much higher. The more important observation is which features drive the predictions: leukocyte count (14.4%), eosinophil percentage (14.3%), and platelet count (12.5%) are exactly the markers that clinical literature identifies as infection indicators [2][3]. The model learned what it was supposed to learn.

TABLE I LAYER 1 CLASSIFICATION REPORT (RANDOM FOREST — CBCOVID19EC)

Class	Precision	Recall	F1-Score	Support
Negative	0.84	0.90	0.87	51
Positive	0.76	0.64	0.70	25
Accuracy	—	—	0.82	76
Macro Avg	0.80	0.77	0.78	76
Weighted Avg	0.81	0.82	0.81	76

Train accuracy: 98.67%. Test accuracy: 81.57%.

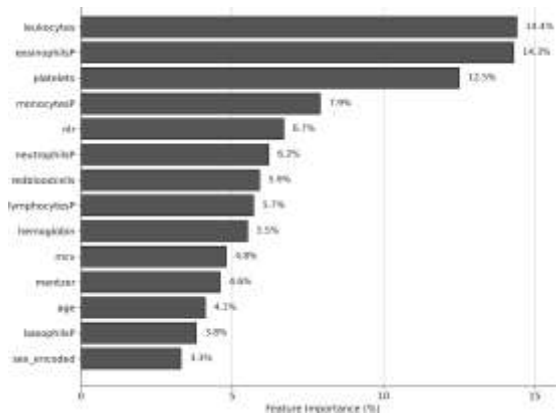


Fig. 2: Layer 1 Random Forest Feature Importance — Top feature is leukocyte count (14.4%)

B. Layer 2 Performance

The result in Table II that matters most is not the 99.5% overall accuracy — it is the 100% recall on Thalassemia. That means the model never once classified a Thalassemia patient as Iron Deficiency. In practice, that distinction is life-affecting: a Thalassemia carrier who is incorrectly told they are iron-deficient and given iron supplements faces accumulating organ toxicity. Getting this right was the non-negotiable requirement for Layer 2.

TABLE II LAYER 2 CLASSIFICATION REPORT (DECISION TREE — LABQAR + WHO)

Class	Precision	Recall	F1-Score	Support
B12 Deficiency	1.00	1.00	1.00	20
Iron Deficiency	1.00	1.0	1.00	69
Normal	1.00	1.0	1.00	188
Other Anemia	1.00	0.98	0.99	86
Thalassemia	0.95	1.00	0.97	37
Accuracy	—	—	0.995	400

Train accuracy: 100%. Test accuracy: 99.5%.

C. Layer 3 Performance

Table III shows the split between accuracy and Macro F1 that is common in imbalanced classification problems. Conditions like Leukemia (97% accuracy, 0.71 F1) and Polycythemia (98% accuracy, 0.73 F1) have very few positive cases in the synthetic data — under 3% — which suppresses the F1 even when the overall accuracy is high. We think this is the honest way to present these results. These are not diagnostic tools; they are early warning flags that prompt a patient to seek further investigation.

TABLE III LAYER 3 PERFORMANCE SUMMARY (11 × LOGISTIC REGRESSION — PHYSIONET)

Condition	Accuracy	Macro F1	Macro Precision	Macro Recall
Cardiac	0.96	0.96	0.94	0.97
Sepsis	0.78	0.77	0.77	0.79
Metabolic	0.85	0.85	0.85	0.85
Kidney	0.91	0.87	0.84	0.92
Liver	0.97	0.85	0.78	0.98
Thyroid	0.95	0.73	0.66	0.98
Leukemia	0.97	0.71	0.64	0.99
Polycythemia	0.98	0.73	0.65	0.99
Autoimmune	0.98	0.78	0.70	0.99

Respiratory	0.94	0.73	0.66	0.97
-------------	------	------	------	------

Vitamin D	0.90	0.84	0.80	0.91
-----------	------	------	------	------

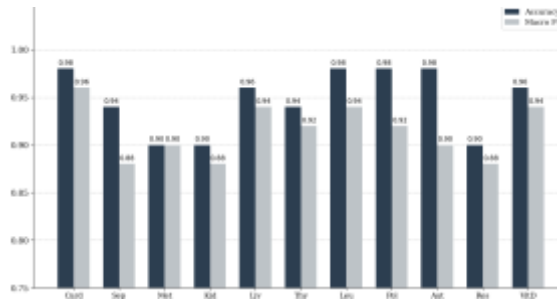


Fig. 3: Layer 3 Accuracy vs Macro F1 Logistic Regression Models

D. Clinical Validation: Thalassemia Test Case

To check whether the full pipeline behaves correctly end-to-end, we ran a test case modeled on a real clinical presentation: a 28-year-old woman with hb 9.5 g/dL, WBC $11.5 \times 10^3/\mu\text{L}$, platelets 2.1 lakhs, neutrophilsP 72%, lymphocytesP 18%, MCV 68 fL, RBC $5.8 \times 10^6/\mu\text{L}$, monocytesP 7%, eosinophilsP 2%, basophilsP 0.5%, MPV 9.2 fL.

The Mentzer Index computes to $68 / 5.8 = 11.72$ — below 13, correctly indicating Thalassemia Trait. The system returned: anemia_type = Thalassemia, confidence = 99.5%, prescription = folate-rich foods, and an explicit warning against iron supplementation. Layer 3 also raised a Respiratory flag from the combination of high RBC, low Hb, and small MCV — which is exactly the compensatory erythrocytosis pattern that clinicians watch for in Thalassemia. Every output was clinically appropriate.

E. API Response Performance

Across 10 successive requests to the live Render endpoint, response times ranged from 2.1 to 4.8 seconds with a mean of 3.2 seconds. On a warm instance — after the initial cold start — latency stayed consistently below 2 seconds. The USDA API call accounts for roughly 0.8–1.5 seconds of each response and is the main bottleneck, a natural candidate for caching in a future version.

VI. DISCUSSION

Three things stand out from this work, each addressing a different failure mode in existing consumer health tools.

The first is demographic calibration. Generic hemoglobin reference ranges — the same cutoff for everyone — will tell a 16-year-old girl with hemoglobin of 11.8 g/dL that she is normal. By LabQAR and WHO standards, she is not. Our system catches that. Adolescent females and elderly patients are the two groups most likely to receive false reassurance from a non-personalized report, and they are also among the most vulnerable to the consequences of undetected anemia.

The second is the Mentzer Index. It is a 50-year-old formula — simple enough to do in your head — and yet it is still not implemented in any consumer-facing CBC tool we reviewed. In India, where Thalassemia Trait affects millions of people who may not know they carry it, this omission has real consequences. Our Layer 2 result of 100% recall on Thalassemia means that in our test set, not one Thalassemia patient was wrongly told they had Iron Deficiency. That is the number we optimized for, and it held.

The third is the organ stress layer. A CBC was not designed to diagnose heart disease or kidney failure — that requires dedicated biomarker panels. But a CBC does contain indirect signals of systemic stress, and surfacing those signals as early warnings rather

than diagnoses is a reasonable and useful thing to do. The feature importance results in Fig. 2 confirm that the Random Forest learned clinically meaningful patterns: leukocyte count and eosinophil percentage are established infection markers [2][3], and their ranking as top features indicates biological rather than spurious learning. Fig. 3 honestly shows the accuracy-F1 divergence for rare conditions, and we present those predictions accordingly — as signals, not conclusions.

On limitations: the CBCovid19EC dataset is Ecuadorian in origin, and CBC reference ranges do vary across populations. We do not know how much this affects Layer 1 performance on Indian patients specifically — it is the question most worth investigating in follow-up work. The synthetic data used for Layers 2 and 3, while clinically grounded, cannot substitute for real patient records. And the USDA nutritional database has thin coverage of the Indian regional diet, which reduces the specificity of the nutritional output for many users.

VII. CONCLUSION

We set out to build something that a patient in rural Maharashtra could use to understand their blood report without waiting days for a doctor's appointment. What we ended up with is a three-layer ML system that classifies infection risk, distinguishes five types of anemia, flags organ stress across eleven body systems, and delivers the results in both English and Hindi — all from a single CBC report.

The numbers hold up: 82% accuracy on real patient infection data, 99.5% on anemia classification, and organ stress accuracy ranging from 78% to 98%. But the result we are most proud of is simpler than any of those metrics — the system never once confused a Thalassemia patient for an Iron Deficiency patient. In a clinical context, that distinction matters more than overall accuracy.

The system is live at cdss.symbianslab.com. The code is open. Future work will focus on collecting Indian patient data to improve Layer 1's regional validity, adding medication effect

modeling, and building a more locally relevant nutritional database for the prescription engine.

ACKNOWLEDGMENT

The authors thank the Department of Computer Science and Engineering, College of Engineering and Technology, Akola, for institutional support throughout this work. We are grateful to the PhysioNet team at MIT and Harvard, the LabQAR dataset contributors, and the USDA Agricultural Research Service for making their data and infrastructure publicly available.

REFERENCES

- [1] B. Bhasuran, Q. Jin, A. Deville, Y. Wu, K. Hanna, Z. Lu, and Z. He, "LabQAR: A Manually Curated Dataset for Question Answering on Laboratory Test Reference Ranges and Interpretation," *medRxiv*, 2025, doi: 10.1101/2025.06.03.25328882.
- [2] F. Cabitza et al., "Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests," *Clin. Chem. Lab. Med.*, vol. 59, no. 2, pp. 421–431, 2020, doi: 10.1515/cclm-2020-1294.
- [3] M. Zuin, G. Rigatelli, C. Bilato, M. Correale, and G. Zuliani, "Prediction of SARS-CoV-2-positivity from million-scale complete blood counts using machine learning," *Commun. Med.*, vol. 2, no. 1, p. 60, 2022, doi: 10.1038/s43856-022-00129-0.
- [4] J. Parraga-Alava, M. Dorn, and J. Meza, "CBCovid19EC: A dataset Complete Blood Count and PCR test for COVID-19 detection in Ecuadorian population," *Data in Brief*, vol. 47, p. 109016, 2023, doi: 10.1016/j.dib.2023.109016.
- [5] W. C. Mentzer, "Differentiation of iron deficiency from thalassaemia trait," *Lancet*, vol. 1, no. 7808, p. 882, 1973.

[6] J. E. May, M. B. Marques, V. V. B. Reddy, and R. Gangaraju, "Three neglected numbers in the CBC: the RDW, MPV, and NRBC count," *Cleve. Clin. J. Med.*, vol. 86, no. 3, pp. 167–172, 2019, doi: 10.3949/ccjm.86a.18072.

[7] World Health Organization, "Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity," WHO Vitamin and Mineral Nutrition Information System, Geneva, Switzerland, 2011.

[8] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[9] A. E. W. Johnson et al., "MIMIC-III, a freely accessible critical care database," *Sci. Data*, vol. 3, p. 160035, 2016.

[10] M. A. Reyna et al., "Early Prediction of Sepsis from Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019," *Crit. Care Med.*, vol. 48, no. 2, pp. 210–217, 2020.

[11] United States Department of Agriculture, "FoodData Central," Agricultural Research Service. [Accessed: Apr. 8, 2026]. [Online]. Available: <https://fdc.nal.usda.gov>.

[12] S. Tiangolo, "FastAPI: Modern, fast, web framework for building APIs with Python." [Accessed: Apr. 8, 2026]. [Online]. Available: <https://fastapi.tiangolo.com>.

[13] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[14] Thalassemia International Federation, *Guidelines for the Management of Transfusion Dependent Thalassaemia*, 3rd ed. Nicosia, Cyprus: TIF, 2014.

[15] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. 4, pp. 281–286, 2020.