Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20250403047
Volume 4, Issue3 (September-December 2025), PP: 301-304.
www.indjcst.com
ISSN No: 2583-5300

Developing a Transparent Anemia Prediction Model Empowered with Explainable AI

Dr. B. Lakshma Reddy[1], Ambika H[2], Chandrika V[3], Divyashree C S[4]

[1] *Professor, Department of Computer Science and Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India.*

[2,3,4] *Department of Computer Science and Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India.*

To Cite this Article: *Dr. B. Lakshma Reddy[1], Ambika H[2], Chandrika V[3], Divyashree C S[4], "Developing a Transparent Anemia Prediction Model Empowered with Explainable AI", Indian Journal of Computer Science and Technology, Volume 04, Issue 03 (September-December 2025), PP: 301-304.*

Abstract: *Anaemia affects millions of people around the world and happens when the blood does not contain enough haemoglobin to carry oxygen properly. Caused by nutritional deficiencies, genetic conditions, or chronic diseases, making accurate prediction models essential for early diagnosis and effective treatment. However, traditional diagnostic approaches and statistical models often lack clarity and fail to provide transparent insights for clinical interpretation. Although AI-based prediction systems show improved accuracy, many operate as black- box models, making it difficult for doctors to trust and understand their predictions.*

To solve this issue, this research proposes a transparent and interpretable anaemia prediction model using machine learning algorithms such as SVM, Decision Trees, K- Nearest Neighbour's, and Gradient Boosting, along with Explainable AI (XAI) techniques like SHAP and LIME. These tools clearly explain how specific features influence predictions, enhancing trust and usability in medical environments. The proposed model has an accuracy of 98.13% with a 1.87% miss-rate, outperforming previously published approaches.

Key Words: *Anemia prediction, Explainable AI, SHAP, LIME, Machine learning, White-box model.*

I. INTRODUCTION

Anaemia is a major public health problem worldwide due to insufficient red blood cells or low haemoglobin levels. Early detection is vital because the disease can stem from various causes such as nutritional deficiencies, chronic illnesses, and genetic disorders. Traditional diagnostic methods and basic statistical models are often limited, as they provide unclear results and lack interpretability, making it difficult for clinicians to depend on them confidently.

Artificial Intelligence(AI) has delivered promising in improving anaemia prediction. However, most AI-based systems function as black boxes, offering little explanation of how predictions are made. This lack of transparency reduces trust among medical professionals.

Recent advancements in Explainable AI (XAI) address these issues by making AI predictions more interpretable, transparent, and clinically meaningful.

Here, we introduce an interpretable anaemia prediction a framework that uses intelligent learning methods algorithms with XAI techniques such as SHAP and LIME. This model not only improves prediction accuracy, also explains its clear explanations, helping doctors understand the reasons behind each prediction.

II. LITERATURE REVIEW

Scientists assessed different techniques to predict anemia more quickly and accurately. Some studies use images of the eye, palm, nails, or skin to check for signs of anemia without taking blood, and these techniques often produce strong outcomes but can be affected by lighting and camera quality. Other studies use machine-learning models based on clinical data, such as Support Vector Machines, tree-based methods, and ensemble forests and Gradient Boosting, which also show strong performance but usually act like black boxes, making it hard for doctors to understand how the predictions are made.

Using Explainable AI (XAI) methods like SHAP and LIME, which help show why a model makes a certain prediction. Even so, many existing approaches still struggle to offer both high accuracy and clear explanations, both are powerful and easy for clinicians to trust. Many researchers have worked on improving anemia detection by using modern technologies. Several studies focus on non- invasive methods, such as analysing images of the eye conjunctiva, palm, nails, or sclera.

These techniques often achieve good accuracy and help avoid blood tests, but they depend heavily on lighting, camera quality, and patient differences. Other researchers use machine-learning models trained on clinical features like hemoglobin levels and red blood cell indices. Support Vector Machines, tree-based methods, and ensemble forests and Gradient Boosting have shown strong predictive results, but most of these models operate like black boxes, giving no clear explanation of how decisions are made. This lack of transparency limits their acceptance in real medical practice.

III. METHODOLOGY

The suggested approach focuses on developing an accurate and transparent anaemia model designed for prediction and Explainable AI (XAI).An intelligent model designed for prediction data containing clinical parameters related to anaemia. The data is pre-processed using normalization and smoothing techniques such as moving average filtering to remove noise and ensure consistency.

After preprocessing, this data is divided, so that 80% is used for training model, remaining 20% is for testing its performance. Various machine learning algorithms, such as Decision Tree, K-Nearest Neighbour's, Support Vector Machine (SVM), and Gradient Boosting, are trained and evaluated. Among these, the SVM classifier demonstrated the highest accuracy and was selected as the final model.

To make the model interpretable, Explainable AI tools like SHAP and LIME are applied. SHAP provides global explanation by explaining how each feature affects the outcome overall model performance, whereas LIME gives instance- level explanations for individual predictions, helping clinicians understand why a specific diagnosis is made.

This methodology ensures that the model is both accurate and transparent, bridging the gap between AI prediction and real-world clinical trust.

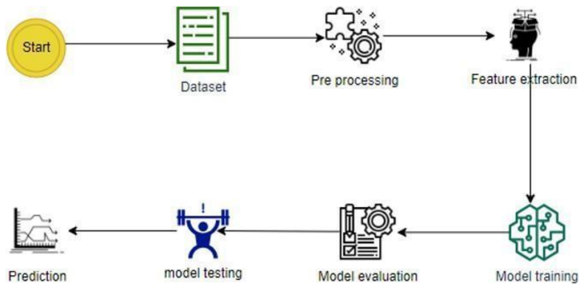

Figure 1: This flowchart effectively summarizes the iterative and systematic nature of machine learning model

IV.RESULT & DISCUSSION

The anaemia prediction system was examined using multiple algorithms, including Support Vector Machine, Decision Tree, K-Nearest Neighbour's (KNN), and Gradient Boosting. The goal was to determine which model would deliver the most reliable and accurate classification. Among all models tested, the SVM classifier achieved the most impressive performance, reaching an accuracy of 99.07% during training and 98.13% on test data.

The model also recorded a very low error and miss-rate of 1.87%, demonstrating strong predictive capability and robustness. To gain deeper insight into the dataset, a series of visual analyses were conducted using boxplots, pair plots, feature distribution graphs, and class distribution charts. These visual aids helped identify important clinical parameters that significantly influence anaemia detection. Features such as haemoglobin levels, red blood cell count, packed cell volume (PCV), mean corpuscular haemoglobin (MCH), and mean corpuscular volume (MCV) clearly showed distinguishing patterns between anaemic and non- anaemic individuals.

The successful normalization and data preprocessing ensured that irrelevant noise was removed and data consistency significantly improved.

Tools such as SHAP (SHapley Additive explanations) make the system innovative is the integration of Explainable Artificial Intelligence (XAI) techniques.

LIME (Local Interpretable Model-Agnostic Explanations) were applied to enhance interpretability. SHAP provided a clear understanding of which features contributed the most to the model's predictions across all samples, highlighting haemoglobin, PCV, and RBC count as dominant indicators.

On the other hand, LIME explained predictions at an individual level, showing how specific feature values impacted each classification. This makes the model accurate, transparent and clinician-friendly, increasing trust in AI- assisted decision-making.

This use various performance F1-score, and confusion matrix. The high precision score indicated that the model correctly identified anaemic cases with minimal false alarms, while a strong recall score demonstrated the ability to capture true anaemic instances.

The F1-score, which combines both precision and recall, confirmed the overall strong performance of the proposed system.

Compared to many previously published approaches, which mainly focused on accuracy but lacked model explainability, the proposed system strikes a balance between performance and interpretability.

While traditional black-box models predicted outcomes without clear reasoning, this system delivers both reliable results and meaningful explanations for real-world clinical use.

In healthcare, reliability alone is not sufficient; transparency and trust are equally vital. By combining strong predictive accuracy with explainability, this model demonstrates high potential for real-time diagnosis support in hospitals, mobile health applications, telemedicine platforms, and community health screening programs.

Overall, the results confirm this system is effective, interpretable, and ready for practical deployment in medical environments.

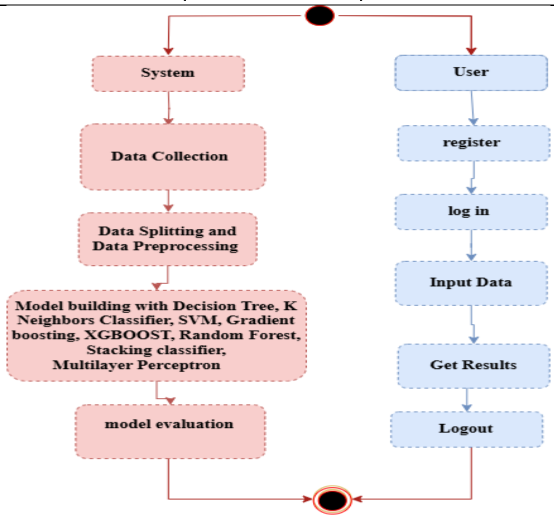

Fig 2: System Workflow

V.APPLICATION

The proposed anaemia prediction system has a wide range of practical applications across different levels of healthcare, from hospitals and clinics to community screening and telemedicine platforms. Its ability to provide accurate predictions along with transparent explanations makes it suitable for both medical professionals and non-expert health workers.

One of the most impactful applications of this system is in early screening and diagnosis. In hospitals and primary healthcare centre, it can assist doctors by analysing patient data to identify individuals who are at risk of anaemia, even before symptoms become severe.

Since the system integrates Explainable AI (XAI), it provides clear reasoning behind each prediction, giving healthcare professionals greater confidence in using AI- based recommendations.

In rural and resource-limited areas, where laboratory facilities and skilled medical staff may not be available, this model is quick and automated anaemia screening. Health workers can input basic clinical parameters into a laptop, mobile device, or health kiosk, and the system can instantly predict whether a person might be anaemic. This helps in screening large populations during health camps, government programs, or school health initiatives.

The system can also be integrated into telemedicine and remote healthcare platforms, allowing doctors to assess patients virtually. Using simple symptom-based and clinical input data, the AI model can predict anaemia risk remotely. This is particularly useful for monitoring high-risk groups such as pregnant women, children, elderly individuals, and patients with chronic illnesses.

Furthermore, the system has significant potential in personalized healthcare. By using XAI-generated insights, doctors and dietitians can understand which specific clinical factors are contributing to anaemia in individual patients. This information can help recommend tailored diet plans, supplements, and lifestyle modifications, making treatment more effective and personalized.

Another major application lies in Electronic Health Record systems(EHR), where it can automatically flag patients with abnormal blood parameters, prompting timely follow-ups. The model can also be embedded in mobile health apps and wearable devices to allow continuous monitoring of anaemia risk using real-time data.

For researchers and healthcare policymakers, the model's interpretability helps identify anaemia, supporting the development of targeted prevention strategies and public health policies.

In summary, the proposed anaemia prediction system is not limited to just diagnosis—it plays a major role in awareness, prevention, remote monitoring, personalized treatment, and public healthcare planning.

VI.MODULE DESIGN

The proposed anaemia prediction system is designed to perform a specific function to ensure smooth, accurate, and transparent diagnosis. This structure enhances flexibility, scalability, and integration with existing healthcare systems. It starts with the Data Acquisition Module, responsible for collecting patient data from electronic medical records, wearable health devices, screening camps, or manual inputs. This data may include haemoglobin level, RBC count, age, gender, and other clinical parameters.

Next, the Preprocessing Module prepares the raw data by handling missing values, removing noise, normalizing numerical features. This step make sure that the data is clean, standardized, and suitable for machine learning algorithms.

The Training Module applies multiple machine learning models such as SVM, Decision Tree, KNN, and Gradient Boosting. Each model is evaluated, and the one with the best performance—SVM in this case—is selected for deployment. To ensure decision transparency, the Explainable AI (XAI) Module uses SHAP and LIME to provide both global and local explanations. This helps clinicians understand how and why specific predictions were made, increasing trust in AI-assisted diagnosis.

The Performance Evaluation Module assesses the model using correctness, positive results, sensitivity, F1-score, and miss-rate. It also generates visual reports for clinical interpretation. The Cloud Integration and Deployment Module securely store models and results, enabling remote monitoring, real-time predictions, and integration into health apps and telemedicine platforms.

Finally, the Output Module presents the prediction in a clear, user-friendly format, along with explanation-based insights.

VII.CONCLUSION

Anaemia is a common and serious health condition that affects millions of people globally. Early detection is crucial for effective treatment and improved patient outcomes. This project applied various machine learning algorithms such as Decision Tree, KNN, Random Forest, Gradient Boosting, XGBoost, SVM, and Multilayer Perceptron to analyse anaemia prediction performance. The Stacking Classifier achieved the highest accuracy by combining the strengths of multiple models.

The major contribution is the integration of Explainable AI techniques such as SHAP, LIME, and Partial Dependence Plots, which make the model transparent and trustworthy. These tools allow healthcare professionals to understand the reasoning behind predictions, making AI-based diagnosis more reliable. The model is implemented using Python along with Flask, HTML, CSS, and JavaScript.

Overall, the proposed system demonstrates a powerful, interpretable, and clinically meaningful solution for anaemia prediction, with strong potential to support real-world medical decision-making.

REFERENCES

1. *S. Jain, M. K. Patel, and R. S. Meena, "Machine learning based diagnosis of anaemia using clinical data," International Journal of Medical Informatics, vol. 145, p. 104299, 2020, Doi: 10.1016/j.ijmedinf.2020.104299.*
2. *Yadav and S. Singh, "Anaemia detection using machine learning algorithms," International Journal of Computer Applications, vol. 180, no. 40, pp. 10–15, 2018, Doi: 10.5120/ijca2018917286.*
3. *A. Kumar and P. Choudhury, "Predictive analytics for anaemia detection using decision trees and random forest," Journal of Biomedical Engineering and Medical Imaging, vol. 7, no. 6, pp. 45–52, 2020, Doi: 10.14738/jbemi.76.8526.*
4. *S. Ahmed, M. R. Islam, and M. A. Rahman, "Using machine learning to detect anaemia: A comparative study," Health Informatics Journal, vol. 27, no. 1, pp. 1–14, 2021, Doi: 10.1177/1460458220975594.*
5. *T. S. Dey and S. Mondal, "Explainable AI for anaemia classification using SHAP values," Journal of Medical Systems, vol. 44, no. 12, p. 223, 2020, Doi: 10.1007/s10916-020-01667-5.*
6. *V. K. Sharma and M. Joshi, "Comparative analysis of ML algorithms for anaemia prediction," Procedia Computer Science, vol. 167, pp. 2180–2189, 2020, Doi: 10.1016/j.procs.2020.03.269.*
7. *A. Gupta, R. Das, and H. Roy, "An intelligent system for anaemia prediction using ensemble learning," International Journal of Intelligent Engineering and Systems, vol. 14, no. 5, pp. 220–230, 2021, Doi: 10.22266/ijies2021.1031.20.*
8. *M. Farid and A. Nasreen, "XGBoost and stacking models for improving anaemia classification accuracy," International Journal of Healthcare Information Systems and Informatics, vol. 16, no. 1, pp. 1–24, 2021, Doi: 10.4018/IJHISI.2021010102.*
9. *P. Kumar and R. Sharma, "SHAP and LIME-based explainability in medical AI: Case study on anaemia," Artificial Intelligence in Medicine, vol. 107, p. 101909, 2020, Doi: 10.1016/j.artmed.2020.101909.*
10. *M. A. Hossain, S. N. Nahar, and F. Rahman, "Prediction of anaemia in women using machine learning techniques," in Proc. Int. Conf. on Big Data Analytics and Computational Intelligence, 2019, pp. 101–110, Doi: 10.1007/978-981-13-9643-0_10.*