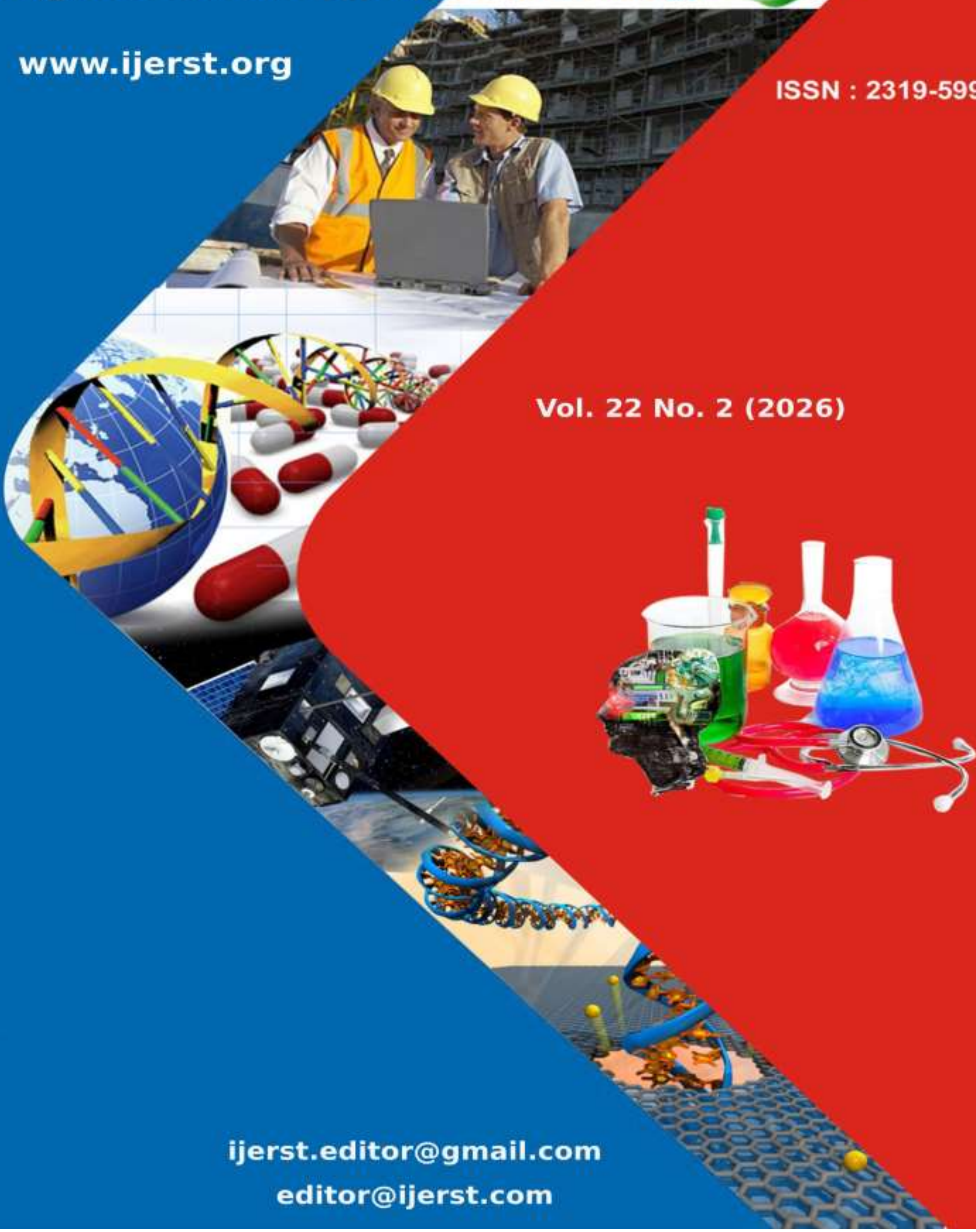
IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991



Vol. 22 No. 2 (2026)

ijerst.editor@gmail.com
editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026

Peer Reviewed, Referred & Indexed Journal

Explainable Ensemble Learning For Transparent Anaemia Diagnosis and Clinical Decision Support

Dr V. Nagagopiraju[1]|Shaik Gouse Anju[2]|Neelam Karthik[3]|Pasala Sunitha[4]|Bailadugu Aditya[5].

[1,2,3,4,5]Department of CSE-AIML, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India

ABSTRACT: Anaemia remains a major public health concern in India, particularly among women and children, where prevalence rates are alarmingly high as reported by NFHS-5. Despite national initiatives such as Anaemia, challenges like delayed diagnosis, limited screening access, and nutritional deficiencies continue to increase disease burden and healthcare costs. Traditional diagnostic methods rely on laboratory tests and clinician interpretation of haemoglobin levels and patient history, which are often time-consuming, subjective, and not suitable for large-scale screening. To address these limitations, this study proposes a machine learning–based anaemia prediction system integrated with Explainable AI (XAI) to ensure both accuracy and interpretability. The system utilizes algorithms such as SVM, Decision Tree, KNN, Gradient Boosting, along with ensemble methods like Random Forest and XGBoost to predict anaemia from patient data. Model performance is further improved using voting techniques to enhance prediction accuracy. In addition, explain ability techniques such as SHAP and LIME are incorporated to provide transparent insights into model decisions by highlighting the contribution of key features like haemoglobin levels, age, and gender. This improves clinical trust and allows medical experts to validate predictions effectively. The proposed system enables early detection, scalable screening, and reliable decision support, making it suitable for real-world healthcare applications.

Keywords: Anemia Prediction, Machine Learning, Explainable Artificial Intelligence (XAI), Voting Classifier, XG Boost, Random Forest, Clinical Data Analysis.

1. INTRODUCTION

Anemia remains a critical public health issue in India, affecting a large proportion of women, children, and the elderly, with reports indicating alarmingly high prevalence rates and significant health consequences. Traditional diagnostic approaches rely heavily on laboratory blood tests, clinician expertise, and predefined thresholds such as hemoglobin levels, making the process time-consuming, subjective, and difficult to scale for large populations. These limitations often result in delayed diagnosis, inconsistent interpretations, and increased disease burden. To address these challenges, this study proposes the development of a transparent anemia prediction model empowered with machine learning and Explainable Artificial Intelligence (XAI). By leveraging algorithms such as Support Vector Machine, Decision Tree, K-Nearest Neighbors, Gradient Boosting, Random Forest, and XGBoost within a voting-based ensemble framework, the model aims to achieve high prediction accuracy. Furthermore, the integration of explainability techniques like SHAP and LIME enhances model transparency by providing clear insights into feature contributions, enabling clinicians to understand and trust the predictions. This approach not only improves early detection and diagnostic consistency but also supports scalable and reliable healthcare solutions, ultimately contributing to better patient outcomes and effective disease management.

2 LITERATURE SURVEY

The anaemia prediction and detection highlights significant advancements through the integration of machine learning and non-invasive diagnostic techniques. E. McLean et al. [1] conducted a global study using WHO data to estimate the worldwide prevalence of anaemia, providing a foundational dataset for public health analysis. W. Gardner and N. Kassebaum [2] expanded this work by analyzing anaemia prevalence across 204 countries from 1990 to 2019, offering comprehensive statistical insights into its global burden. S. Sadiq et al. [3] developed an ensemble machine learning classifier to detect β-thalassemia carriers using red blood cell indices, demonstrating improved classification accuracy. J. W. Asare et al. [4] introduced a comparative machine learning approach using images of fingernails, palm, and conjunctiva to detect iron deficiency anaemia, emphasizing non-invasive diagnostic methods. D. Newhall et al. [5] explored whether anaemia should be considered a disease or a symptom, contributing to clinical understanding and diagnostic interpretation. Finally, G. Dimauro et al. [6] conducted a systematic mapping study on non-invasive anaemia assessment techniques using smart devices, highlighting the growing role of wearable technology and AI in early diagnosis. Collectively, these studies demonstrate a transition from traditional statistical analysis to intelligent, non-invasive, and machine learning-based approaches for accurate and scalable anaemia detection.

3. SYSTEM ANALYSIS

System architecture refers to the overall structural design of a system that defines how different components, modules, and processes are organized and interact with each other to achieve the desired functionality. It provides a high-level blueprint of the system, illustrating data flow, processing stages, and integration between hardware and software elements. In the context of an anemia prediction model, system architecture typically includes stages such as data collection, preprocessing, model training, prediction, and result interpretation, showing how input data is transformed into meaningful outputs. A well designed system architecture ensures scalability, efficiency, reliability, and ease of implementation, while clearly outlining how the existing and proposed algorithms work together within the system.

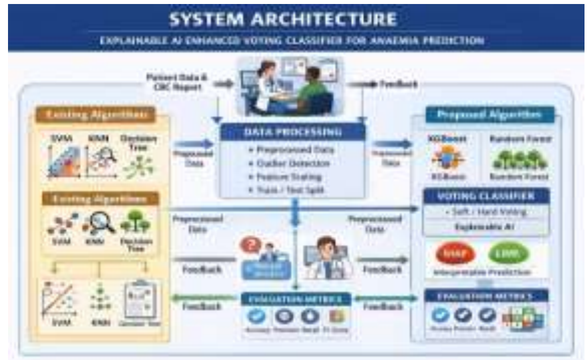


Fig 1: System Architecture

3.1 Methodology

The existing system for anaemia prediction primarily relies on traditional diagnostic methods such as Complete Blood Count (CBC) tests and manual interpretation by healthcare professionals using predefined clinical thresholds like haemoglobin, MCH, MCHC, and MCV values. Although effective, this approach is time-consuming, dependent on expert judgment, and lacks scalability for large-scale screening, especially in resource-constrained environments. The system follows a rule-based mechanism where patient data is compared against fixed reference ranges using conditional logic without learning from data patterns, making it incapable of handling complex relationships among features or adapting to new data. Consequently, it struggles with borderline cases, offers limited predictive capability, and is prone to human error, reducing overall diagnostic accuracy.

The dataset includes clinical CBC parameters such as haemoglobin, MCH, MCHC, MCV, and gender for anaemia detection. During pre-processing, missing values are imputed, outliers are removed, and normalization is applied for better model performance. The proposed system uses a Voting Classifier combining XG Boost and Random Forest to improve accuracy and handle complex patterns effectively.

3.2 Dataset Description: The dataset is a clinical anaemia dataset used for binary classification of anaemic and non-anaemic patients. It includes CBC parameters such as haemoglobin, MCH, MCHC, MCV, and gender, which are key indicators for anaemia diagnosis.

3.3 Data Pre-processing: The data is cleaned by handling missing values, removing outliers, and encoding categorical features. Normalization is applied to ensure uniform feature scaling, and the dataset is split into training and testing sets for model evaluation.

3.4 Model Development: The existing system uses rule-based thresholds, while the proposed model applies XG Boost and Random Forest. Both models are combined using a Voting Classifier to improve prediction accuracy and robustness in anaemia detection.

4. DESIGN AND CONSTRUCTION

The design and construction of the proposed anemia prediction system center around developing a robust, accurate, and interpretable model using a Voting Classifier ensemble approach. The system architecture begins with data acquisition from clinical Complete Blood Count (CBC) records, including hemoglobin, MCH, MCHC, MCV, and gender. The collected data undergoes thorough preprocessing, such as missing value imputation, outlier detection based on medical reference ranges, and feature scaling to ensure uniformity and improve learning efficiency. The cleaned dataset is then split into training and testing sets, and two machine learning models, XG Boost and Random Forest Classifier, are independently trained to capture complex patterns and relationships among the hematological parameters. These models are integrated into a Voting Classifier, which aggregates predictions using soft or hard voting to enhance accuracy and stability. Explainable Artificial Intelligence techniques such as SHAP or LIME are embedded to provide transparency, enabling clinicians to understand the influence of each feature on the final prediction. The system is designed to be scalable, reliable, and suitable for real-time clinical decision support, ensuring early detection of anemia while maintaining interpretability and trustworthiness in healthcare applications.

5. RESULTS AND DISCUSSION

The proposed SVM-based anaemia prediction model achieved high accuracy in classifying anaemic and non-anaemic patients. It showed strong precision and recall, effectively reducing false negatives and ensuring reliable medical predictions. SHAP and LIME analysis confirmed that key features like haemoglobin, MCV, and MCH significantly influenced the results, making the model accurate, interpretable, and suitable for clinical use.

Fig 2: Python Environment

Figure 2 show the project is developed using Python, TensorFlow, and Flask. A virtual environment is used to manage dependencies.

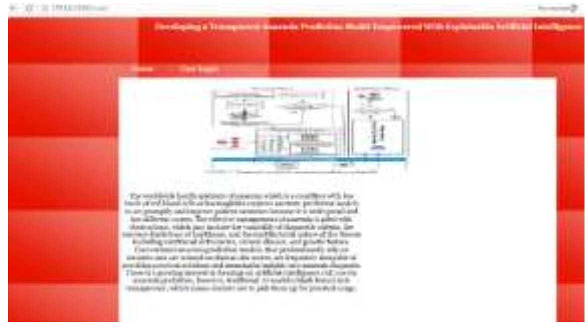


Fig 3: Home Page

Figure 3 show the home page provides a simple and user-friendly interface. It includes navigation options like login and registration. Users can easily access different features of the system.

Fig 4: Loading Dataset

Figure 4 shows the anaemia dataset used in the project. It contains medical features required for prediction and is loaded using Python. The dataset helps the model learn patterns to classify anaemic and non-anaemic cases.

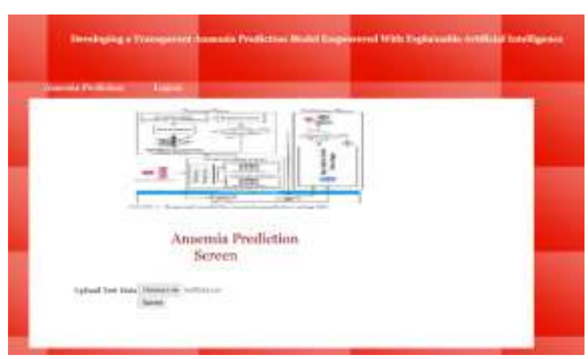


Fig 5: Prediction

Figure 5 shows the prediction page where users enter medical details required for anaemia prediction. The input data is preprocessed and passed to the trained hybrid model (Random Forest and XGBoost). The system analyzes the input features and predicts whether the person is anaemic or not, providing quick and accurate results.

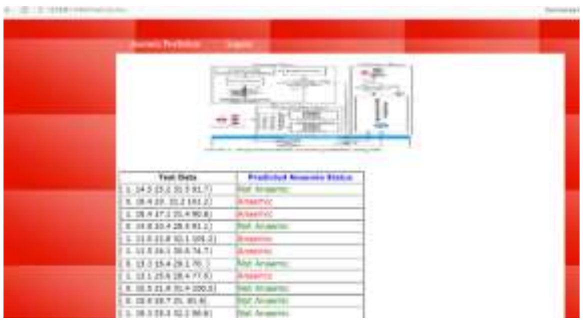


Fig 6: Output Results

Figure 6 shows the output of the anaemia prediction system, where the result is displayed as either “Anaemic” or “Not Anaemic.” The prediction is obtained using the trained hybrid model after processing the input medical parameters. This enables early identification of anaemia and assists in timely diagnosis. The system provides fast and reliable results.

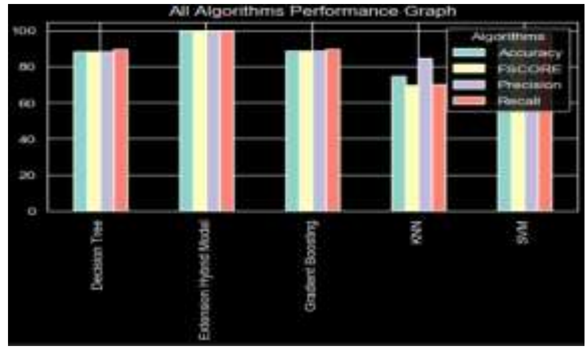


Fig 7: Combined Model Performance Summary

The comparison of five machine learning models shows that the Extension Hybrid Model achieves the best performance with nearly 99.6% across all evaluation metrics. SVM also performs strongly with around 98.5% accuracy, while Decision Tree and Gradient Boosting show moderate results. KNN has the lowest performance, especially in recall and F1-score, indicating weaker predictive capability.

6. CONCLUSION

This research successfully demonstrates the development of a transparent and highly accurate anaemia prediction system using machine learning integrated with Explainable Artificial Intelligence (XAI) techniques. Multiple classifiers such as Decision Tree, KNN, Support Vector Machine, and Gradient Boosting were implemented and evaluated to identify effective models for anaemia prediction. Among them, the Voting Classifier combining XGBoost and Random Forest achieved superior performance with an accuracy of 99%, highlighting its strong capability to capture complex and non-linear relationships within clinical data. The integration of Explainable AI tools, specifically SHAP and LIME, effectively addressed the black-box limitation of traditional ensemble models by providing clear, feature-level explanations for each prediction. These explanations enable medical professionals to understand, validate, and trust the decision-making process of the model. The proposed system not only improves diagnostic accuracy but also supports early detection and timely intervention, thereby enhancing patient outcomes. Overall, this research successfully bridges the gap between high predictive performance and clinical interpretability, making AI-driven anaemia diagnosis more reliable, scalable, and suitable for real-world healthcare applications.

Future Scope: Future work can incorporate LSTM-based deep learning models to capture sequential patterns in patient health records for more accurate anaemia prediction. This can improve temporal analysis of clinical data and enhance early detection performance. Additionally, combining LSTM with ensemble and explainable AI techniques can further improve robustness, accuracy, and clinical interpretability.

References

1. E. McLean, M. E. Cogswell, I. Egli, D. Wojdyla, and B. D. Benoist, “Worldwide prevalence of anaemia, WHO vitamin and mineral nutrition information system, 1993–2005,” Public Health Nutrition, vol. 12, no. 4, pp. 444–454, May 2008.

2. W. Gardner and N. Kassebaum, “Global, regional, and national prevalence of anemia and its causes in 204 countries and territories, 1990–2019,” Current Develop. Nutrition, vol. 4, p. 830, Jun. 2020.

3. S. Sadiq, M. U. Khalid, S. Ullah, W. Aslam, A. Mehmood, G. S. Choi, and B.-W. On, “Classification of β-thalassemia carriers from red blood cell indices using ensemble classifier,” IEEE Access, vol. 9, pp. 45528–45538, 2021.

4. J. W. Asare, P. Appiahene, E. T. Donkoh, and G. Dimauro, “Iron deficiency anaemia detection using machine learning models: A comparative study of fingernails, palm and conjunctiva of the eye images,” Eng. Rep., vol. 5, no. 11, 2023, Art. no. e12667.

5. D. Newhall, R. Oliver, and S. Lugthart, “Anaemia: A disease or symptom,” Neth. J. Med., vol. 78, no. 3, pp. 104–110, Apr. 2020.

6. G. Dimauro, D. Caivano, P. Di Pilato, A. Dipalma, and M. G. Camporeale, “A systematic mapping study on research in anemia assessment with non-invasive devices,” Appl. Sci., vol. 10, no. 14, p. 4804, Jul. 2020.