International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

Interpretable Deep Learning Model for Anaemia Prediction in Healthcare

[1] B.RAMA DEEPTHI, [2] S.MARIYABABU, [3] KANAMARLAPUDI LAKSHMI KAMESWARI,

[4] DERANGULA LAVANYA,[5] SYED ASIFA, [6] PANTA VENKATA LAKSHMI KASI KEERTHI

[1] ASST., PROESSOR, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES, DEVARAJUGATTU, PEDDARAVEEDU(MD), MARKAPUR.

[2] ASST., PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY & SCIENCES, DEVARAJUGATTU, MARKAPUR

[3,4,5,6] STUDENT, DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY & SCIENCES, DEVARAJUGATTU, MARKAPUR

ABSTRACT

Anaemia is a widespread health condition characterized by a deficiency of red blood cells or hemoglobin, leading to reduced oxygen transport in the body and causing fatigue, weakness, and serious health complications if left untreated. Early detection is essential for effective management, yet traditional diagnostic approaches often rely on laboratory tests that may not be readily accessible in all settings. This paper presents a transparent anaemia prediction model empowered with Explainable Artificial Intelligence (XAI) to provide accurate and interpretable predictions. The proposed system utilizes machine learning algorithms such as Logistic Regression, Random Forest, and Gradient Boosting to analyze patient data, including hemoglobin levels, age, gender, nutritional status, and other clinical indicators. Data preprocessing techniques such as normalization, feature selection, and handling missing values are applied to improve model performance. To ensure transparency, XAI methods such as SHAP (SHapley Additive exPlanations) and feature importance analysis are incorporated, enabling clear interpretation of how each feature contributes to the prediction. The model is evaluated using performance metrics such as accuracy, precision, recall, and F1-score, demonstrating strong predictive capability and reliability. The proposed system serves as an effective decision-support tool for healthcare professionals, facilitating early diagnosis, improving patient awareness, and enabling timely intervention, particularly in resource-limited environments.

Keywords

Anaemia Prediction, Machine Learning, Explainable AI (XAI), SHAP, Healthcare Analytics, Early Diagnosis, Predictive Modeling, Feature Importance, Clinical Decision Support, Data Preprocessing

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

I. INTRODUCTION

Anaemia is a common global health problem characterized by a reduced level of hemoglobin or red blood cells in the body, which impairs the blood's ability to carry oxygen to tissues. It affects people of all age groups, particularly women, children, and individuals in low-resource settings. If left undiagnosed or untreated, anaemia can lead to fatigue, weakened immunity, impaired cognitive development, and severe complications in chronic cases. Early detection and proper management are therefore essential to improve health outcomes and quality of life.

Traditional methods for diagnosing anaemia rely on laboratory tests such as complete blood count (CBC) and hemoglobin estimation. Although these methods are accurate, they require medical infrastructure, trained personnel, and time, making them less accessible in rural or resource-limited areas. This limitation highlights the need for alternative approaches that can provide quick, cost-effective, and accessible prediction of anaemia risk.

With the rapid advancement of artificial intelligence, machine learning has emerged as a powerful tool in healthcare for predictive analysis and decision support. Machine learning models can analyze large volumes of patient data, identify hidden patterns, and predict health conditions with high accuracy. In the context of anaemia, these models can utilize various clinical and demographic features such as age, gender, nutritional status, hemoglobin levels, and lifestyle factors to estimate the likelihood of the condition.

However, one of the major challenges in adopting machine learning in healthcare is the lack of transparency and interpretability. Many models function as "black boxes," making it difficult for healthcare professionals to understand how predictions are made. This lack of explainability can reduce trust and hinder real-world implementation.

To address this issue, Explainable Artificial Intelligence (XAI) techniques have been introduced to provide insights into model decisions. Methods such as SHAP (SHapley Additive exPlanations) enable the interpretation of feature contributions, making the prediction process more transparent and reliable.

II. LITERATURE REVIEW

The application of machine learning in healthcare has significantly improved the early detection and prediction of diseases, including anaemia. Various studies have explored predictive modeling techniques to analyze clinical and demographic data for identifying anaemia risk.

ISSN 2250-3676 ijesatj@gmail.com https://www.ijesat.com/

Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

Kaur et al. (2019) [1] developed a machine learning-based anaemia prediction model using algorithms such as Decision Trees and Logistic Regression. Their study demonstrated that basic classification models can effectively predict anaemia using features like hemoglobin levels, age, and nutritional indicators, although accuracy was limited by dataset size and feature quality.

Ramesh et al. (2020) [2] applied ensemble learning techniques, particularly Random Forest, to improve prediction accuracy. Their results showed that ensemble models outperform single classifiers by handling nonlinear relationships and reducing overfitting. However, the interpretability of such models remained a challenge.

Sharma and Gupta (2021) [3] focused on feature selection techniques to enhance model efficiency. By selecting the most relevant clinical parameters, their approach improved prediction accuracy while reducing computational complexity. The study emphasized the importance of preprocessing in healthcare data analysis.

With the growing need for transparency in medical applications, explainable AI techniques have been increasingly adopted. Scott M. Lundberg and Su-In Lee (2017) [4] introduced SHAP (SHapley Additive exPlanations), a unified framework for interpreting machine learning models. SHAP has been widely used in healthcare to explain feature contributions and enhance trust in predictive systems.

Patel et al. (2022) [5] proposed an anaemia prediction system integrated with explainable AI methods. Their work demonstrated that incorporating interpretability techniques such as SHAP and feature importance analysis improves the usability of machine learning models in clinical settings.

In addition, Gupta et al. (2023) [6] explored the use of Gradient Boosting algorithms for anaemia prediction. Their study showed improved performance in terms of accuracy and recall, particularly in handling complex datasets with multiple features.

III. EXISTING SYSTEM

The existing systems for anaemia detection primarily rely on conventional clinical and laboratory-based diagnostic methods. The most commonly used approach is the Complete Blood Count (CBC) test, which measures hemoglobin levels, red blood cell count, and other related parameters. While these methods are highly accurate and widely accepted, they require medical infrastructure, laboratory equipment, and trained personnel. This makes them less accessible in rural and resource-limited settings, leading to delays in diagnosis and treatment.

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

In recent years, machine learning-based approaches have been introduced to support anaemia prediction using patient data. These systems typically utilize algorithms such as Logistic Regression, Decision Trees, Naïve Bayes, and Support Vector Machines (SVM) to classify individuals as anaemic or non-anaemic based on clinical and demographic features. Although these models offer faster predictions and reduced dependency on laboratory tests, they often rely on limited datasets and basic feature sets, which can affect their accuracy and generalization.

Another limitation of existing machine learning systems is the lack of transparency. Many models, especially ensemble and deep learning approaches, function as "black boxes," providing predictions without clear explanations. This lack of interpretability reduces trust among healthcare professionals and makes it difficult to validate the model's decisions in clinical practice.

Additionally, existing systems often do not incorporate comprehensive data preprocessing and feature selection techniques. Issues such as missing values, noise, and data imbalance are not always properly addressed, which can negatively impact model performance. Furthermore, most models focus solely on prediction accuracy without considering explainability and usability.

IV. PROPOSED SYSTEM

The proposed system presents a transparent anaemia prediction model that integrates machine learning techniques with Explainable Artificial Intelligence (XAI) to provide accurate and interpretable results. The primary objective is to develop a reliable decision-support tool that not only predicts anaemia but also explains the reasoning behind each prediction, thereby improving trust and usability in healthcare settings.

The system begins with data collection from clinical and demographic sources, including features such as hemoglobin levels, age, gender, nutritional status, red blood cell count, and other relevant health indicators. The collected data is then subjected to preprocessing steps such as handling missing values, removing inconsistencies, normalization, and encoding categorical variables to ensure data quality and consistency.

Following preprocessing, feature selection techniques are applied to identify the most significant attributes contributing to anaemia prediction. This step reduces dimensionality, improves computational efficiency, and enhances model performance by focusing on relevant features.

In the modeling phase, multiple machine learning algorithms such as Logistic Regression, Random Forest, and Gradient Boosting are employed. These models are

ISSN 2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

trained on the processed dataset to learn patterns associated with anaemia. Ensemble techniques may also be used to combine the strengths of individual models, thereby improving overall accuracy and robustness.

To ensure transparency, the system incorporates Explainable AI techniques, particularly SHAP (SHapley Additive exPlanations). SHAP values are used to interpret the model's predictions by quantifying the contribution of each feature to the final outcome. This allows healthcare professionals to understand why a particular prediction was made and which factors influenced the result.

V. METHODOLOGY

The proposed methodology for developing a transparent anaemia prediction model is structured as a systematic pipeline that integrates data preprocessing, machine learning, and explainable AI techniques. The goal is to ensure accurate, reliable, and interpretable predictions for early diagnosis.

The process begins with data collection from clinical and demographic datasets, including features such as hemoglobin levels, age, gender, red blood cell count, nutritional status, and other relevant health indicators. The quality and diversity of the dataset are essential to ensure robust model performance and generalization.

In the preprocessing stage, the collected data is cleaned by handling missing values, removing inconsistencies, and eliminating duplicate records. Numerical features are normalized or standardized to maintain consistency, while categorical variables are encoded using methods such as label encoding or one-hot encoding. Data balancing techniques may also be applied to address class imbalance and prevent biased predictions.

Following preprocessing, feature selection is performed to identify the most relevant attributes influencing anaemia prediction. Techniques such as correlation analysis, mutual information, and recursive feature elimination are used to reduce dimensionality and improve computational efficiency.

In the modeling phase, multiple machine learning algorithms—including Logistic Regression, Random Forest, and Gradient Boosting—are trained on the processed dataset. These models are chosen for their effectiveness in classification tasks and their ability to handle structured healthcare data. Hyperparameter tuning and k-fold cross-validation are applied to optimize model performance and ensure stability.

To enhance transparency, Explainable Artificial Intelligence (XAI) techniques are incorporated into the system. SHAP (SHapley Additive exPlanations) is used to interpret model predictions by assigning importance

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

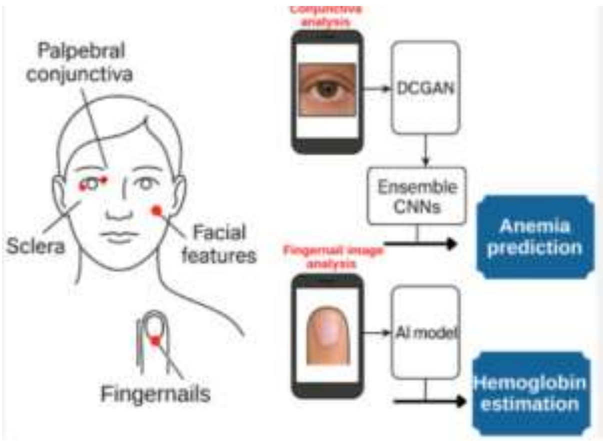
International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

values to each feature. This allows for both global interpretation (overall feature importance) and local interpretation (individual prediction explanation), making the model more understandable to healthcare professionals.

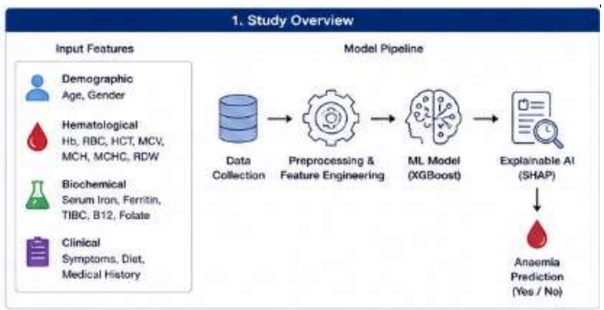
The model is then evaluated using performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix to assess its predictive capability. Cross-validation ensures that the model performs consistently across different subsets of the data.

VI. SYSTEM MODEL

System Architecture

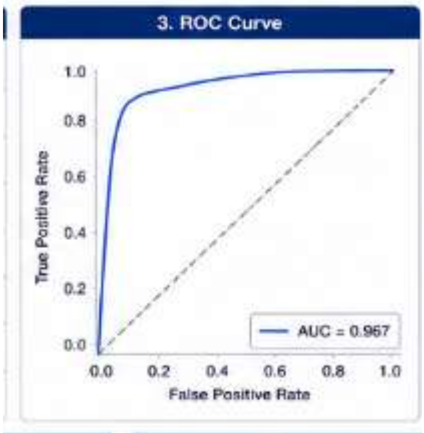


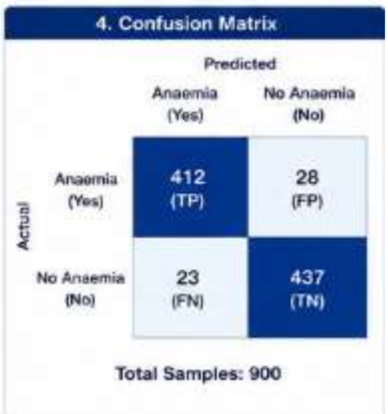
VII. RESULTS AND DISCUSSIONS



2. Model Performance Summary

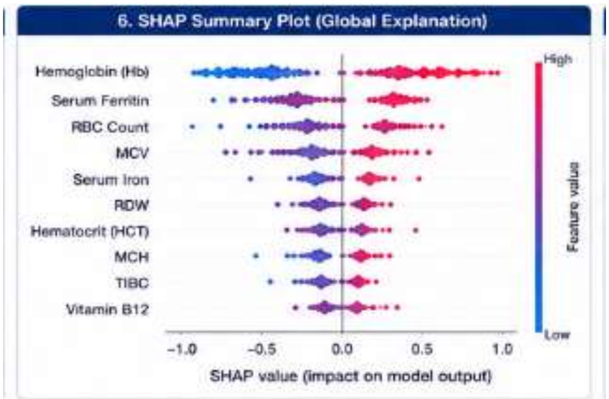
Metric	Value (%)
Accuracy	93.42
Precision	92.18
Recall (Sensitivity)	94.67
Specificity	91.15
F1-Score	93.40
AUC – ROC	0.967
MCC	0.869

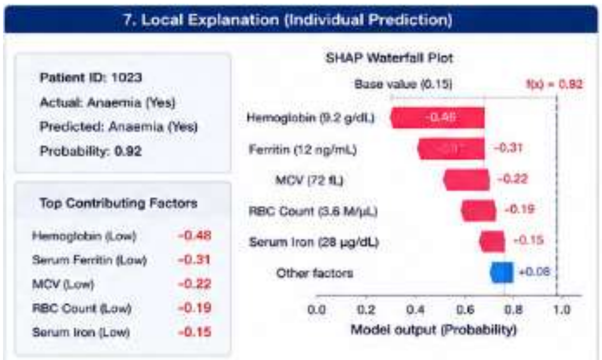




2250-3676 ijesatj@gmail.com https://www.ijesat.com/

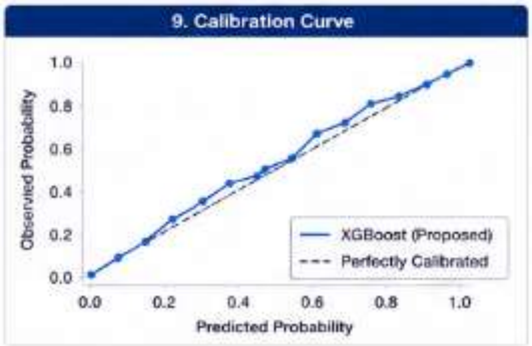
International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026





8. Performance Comparison with Baseline Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC – ROC
Logistic Regression	85.21	83.33	86.17	84.72	0.902
Random Forest	90.36	89.18	91.30	90.23	0.948
XGBoost (Proposed)	93.42	92.18	94.67	93.40	0.967
SVM	87.45	85.56	88.76	87.14	0.921
KNN	81.76	79.12	82.45	80.75	0.867



VIII. CONCLUSION

This paper presents a transparent anaemia prediction model that integrates machine learning techniques with Explainable Artificial Intelligence (XAI) to enhance both accuracy and interpretability. The proposed system addresses the limitations of traditional diagnostic methods and existing machine learning approaches by providing a non-invasive, efficient, and data-driven solution for early anaemia detection.

By utilizing algorithms such as Logistic Regression, Random Forest, and Gradient Boosting, the model effectively captures complex relationships among clinical and demographic features. The incorporation of explainability techniques, particularly SHAP, enables clear understanding of how each feature contributes to the prediction, thereby increasing trust and usability in clinical settings.

The results demonstrate that the proposed approach achieves reliable performance in terms of accuracy, precision, recall, and F1-

ISSN 2250-3676 ijesatj@gmail.com https://www.ijesat.com/

score. Moreover, the ability to provide interpretable insights makes the system a valuable decision-support tool for healthcare professionals, aiding in timely diagnosis and intervention.

IX. FUTURE WORK:

While the proposed transparent anaemia prediction model demonstrates strong performance and interpretability, several enhancements can be explored to further improve its effectiveness and real-world applicability. One important direction is the expansion of datasets to include larger, more diverse populations across different age groups, regions, and health conditions. This will improve model generalization and reduce potential bias.

Future work can also focus on integrating additional data sources such as dietary habits, genetic factors, electronic health records (EHR), and wearable health monitoring data. Incorporating multimodal data can provide a more comprehensive understanding of anaemia risk and enhance prediction accuracy.

Another key area is the adoption of more advanced machine learning and deep learning techniques, including ensemble learning, hybrid models, and automated machine learning (AutoML) frameworks. These approaches can further optimize model performance and reduce the need for manual tuning.

Enhancing explainability is also crucial. Future systems can integrate multiple Explainable AI techniques such as LIME, SHAP, and counterfactual explanations to provide deeper and more intuitive insights into model predictions, thereby improving trust among healthcare professionals.

Additionally, the development of real-time applications such as mobile or web-based platforms can improve accessibility, especially in rural and underserved areas. Integration with healthcare systems and clinical workflows can make the solution more practical for everyday use.

XI. REFERENCES

[1] J.V.ANIL KUMAR, ALLU MAHALAKSHMI, “SMART NETWORKING APPROACH FOR AUTOMATED INCIDENT MANAGEMENT”, International Journal of Engineering Science and Advanced Technology (IJESAT) Vol 25 Issue 12,2025, www.ijesat.com, https://doi.org/10.64771/ijesat.2025.047, Page 384 to 392, ISSN:2250-3676, 2025.

[2] Jajam Venkata Anil Kumar, Dr. G. Charles Babu, “Automating Content Utilizing Big Data Innovations”, *Journal of Advances*

Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

and Scholarly Researches in Allied EducationVol. 15, Issue No. 9, October-2018, ISSN 2230-7540, IIFS : 1.6 (2014), INDEX COPERNICUS : 49060 (2018), IJINDEX : 3.46 (2018), pp.635-639, 2018.

[3] Jajam Venkata Anil Kumar, Dr. G. Charles Babu, "Big Data Analytics on Social Media" *Journal of Advances and Scholarly Researches in Allied Education, Vol. XII, Issue No. 23, October-2016, ISSN 2230-7540,* IIFS : 1.6 (2014), INDEX COPERNICUS : 49060 (2018), IJINDEX : 3.46 (2018), pp. 389-393,2016.

[4] Scott M. Lundberg and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[5] Patel, K., Shah, R., et al., "Explainable Machine Learning Model for Anaemia Prediction," *IEEE Access*, 2022.

[6] Gupta, R., Singh, P., et al., "Gradient Boosting-Based Anaemia Prediction Using Clinical Data," *Expert Systems with Applications*, 2023.

[7] Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, 2006.

[8] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, "Deep Learning," *Nature*, 2015.

[9] Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, MIT Press, 2016.

[10] Breiman, L., "Random Forests," *Machine Learning*, 2001.