HemaExplainAI: A Hybrid Explainable Intelligence Framework for Transparent Anaemia Risk Prediction Using Interpretable Machine Learning

Swetha N CH, M.Tech Scholar, Department of CSE, Nimra College of Engineering and Technology, Jupudi, Ibrahimpatnam, Vijayawada, India, Email: nchswetha07@gmail.com
Mrs.G.Preeti Jyotsna, Assistant Professor, Department of CSE, Nimra College of Engineering and Technology, Jupudi, Ibrahimpatnam, Vijayawada, India, Email: preeti.gudapati@gmail.com

Abstract
Anaemia is one of the most common health disorders affecting millions of people worldwide, particularly women, children, and elderly individuals. Early diagnosis and prediction of anaemia are essential for preventing severe health complications and improving patient treatment outcomes. Traditional anaemia diagnosis methods mainly rely on laboratory blood testing and manual medical analysis, which can be time-consuming, expensive, and dependent on healthcare infrastructure availability. This paper presents the development of a transparent anaemia prediction model using machine learning techniques for intelligent healthcare analytics and disease prediction.

The proposed system utilizes machine learning algorithms including Random Forest, Decision Tree, and Logistic Regression to analyze medical parameters related to anaemia conditions. Data preprocessing techniques such as missing value handling, normalization, and feature optimization are applied to improve prediction accuracy and computational efficiency. The system is implemented using Python, Flask, HTML, CSS, and machine learning libraries to provide a user-friendly web-based healthcare prediction platform.

Experimental evaluation demonstrates that the proposed Random Forest model achieves superior prediction performance compared to other machine learning approaches in terms of Accuracy, Precision, Recall, and F1-Score. The transparent machine learning framework improves prediction interpretability and assists healthcare professionals in understanding disease prediction outcomes. The developed system provides a reliable, scalable, and cost-effective solution for intelligent anaemia prediction and healthcare decision support applications.

Keywords— Anaemia Prediction, Machine Learning, Random Forest, Healthcare Analytics, Disease Prediction, Flask Framework, Medical Data Analysis, Healthcare Monitoring.

I. INTRODUCTION

The rapid advancement of artificial intelligence and machine learning technologies has significantly transformed modern healthcare systems by enabling intelligent disease prediction, automated medical analysis, and efficient healthcare monitoring. Healthcare organizations increasingly adopt machine learning techniques to improve disease diagnosis, patient treatment planning, and clinical decision-making. Among various global health disorders, anaemia is considered one of the most common and serious medical conditions affecting millions of people worldwide.

Anaemia is a condition that occurs when the body does not have enough healthy red blood cells or sufficient hemoglobin levels to transport oxygen effectively to body tissues. Common symptoms of anaemia include fatigue, weakness, dizziness, pale skin, shortness of breath, and reduced physical performance. If not diagnosed and treated at an early stage, severe anaemia may lead to serious health complications including heart-related diseases, pregnancy complications, and reduced immune system functionality.

Traditional anaemia diagnosis methods mainly depend on laboratory blood tests and manual medical analysis performed by healthcare professionals. Although these methods provide reliable clinical evaluation, they can be time-consuming, expensive, and dependent on healthcare infrastructure availability. In rural and resource-limited environments, access to proper laboratory facilities and medical experts may be limited, resulting in delayed diagnosis and treatment. Therefore, there is a growing need for intelligent healthcare systems capable of providing fast, accurate, and cost-effective disease prediction support.

Machine learning techniques provide an effective solution for intelligent healthcare analytics by analyzing medical datasets and identifying hidden patterns associated with disease conditions. Algorithms such as Decision Tree, Logistic Regression, Random Forest, Support Vector Machine, and Neural Networks have demonstrated strong performance in medical prediction and classification applications. These approaches improve diagnostic efficiency and assist healthcare professionals in making accurate clinical decisions.

Among these algorithms, Random Forest has gained significant attention because of its high prediction accuracy, robustness, and ability to handle complex medical datasets. Random Forest combines multiple decision trees using ensemble learning techniques to improve classification performance and reduce overfitting problems. Logistic Regression provides interpretable prediction analysis, while Decision Tree algorithms offer simple and understandable classification structures.

This research focuses on developing a transparent anaemia prediction model using machine learning techniques for intelligent healthcare prediction and medical decision support. The proposed system integrates preprocessing methods, feature optimization techniques, and multiple machine learning algorithms to improve prediction performance and interpretability. The system is implemented using Python, Flask, HTML, CSS, and machine learning libraries to provide a user-friendly web-based healthcare prediction platform.

The developed framework allows users to enter medical parameters and receive instant prediction results through an interactive web application. The transparency of the machine learning model improves trust and understanding of prediction outcomes for healthcare professionals and users.

The major contributions of this work are summarized as follows:

1. Development of a transparent machine learning-based anaemia prediction system.
2. Implementation of multiple classification algorithms for healthcare analytics.
3. Integration of preprocessing and feature optimization techniques for improved prediction efficiency.
4. Development of a user-friendly Flask-based healthcare prediction platform.
5. Evaluation of model performance using Accuracy, Precision, Recall, and F1-Score metrics.
6. Improvement of healthcare prediction transparency and interpretability.

The proposed system provides a reliable, scalable, and cost-effective solution for intelligent anaemia prediction and healthcare decision support applications.

II. LITERATURE REVIEW

The rapid growth of machine learning and artificial intelligence technologies has significantly improved healthcare analytics and disease prediction systems. Intelligent healthcare applications are increasingly being used for early disease detection, patient monitoring, and medical decision support. Among various healthcare problems, anaemia prediction has gained considerable research attention because of its impact on global public health and the importance of early diagnosis.

Traditional anaemia diagnosis methods mainly depend on laboratory blood testing and manual clinical analysis performed by healthcare professionals. These methods are generally accurate but may require significant time, cost, and healthcare infrastructure support. In many rural and underdeveloped regions, limited medical resources and lack of specialized healthcare facilities can delay disease diagnosis and treatment. To overcome these limitations, researchers have explored machine learning techniques for automated healthcare prediction and intelligent disease analysis.

A. Traditional Healthcare Prediction Approaches

Early healthcare prediction systems primarily relied on statistical analysis and rule-based medical expert systems. These systems used predefined medical rules and threshold values for disease classification. Common statistical methods such as linear regression and probability-based analysis were applied to predict patient conditions based on medical attributes.

Although traditional approaches provided basic decision support, they faced several limitations including low scalability, poor adaptability to large datasets, and limited prediction accuracy. Manual feature selection and dependency on predefined rules reduced the efficiency of these systems in handling complex healthcare data.

B. Machine Learning in Healthcare Analytics

Machine learning algorithms have significantly improved disease prediction capability by automatically learning patterns from medical datasets. Several machine learning models including Decision Tree, Logistic Regression, Naive Bayes, Support Vector Machine, Random Forest, and Artificial Neural

Networks have been widely used in healthcare applications.

Decision Tree algorithms provide simple and interpretable classification structures, making them suitable for healthcare decision analysis. Logistic Regression is commonly used for binary classification problems and provides understandable prediction probabilities. Naive Bayes classifiers are efficient for probabilistic healthcare prediction tasks, while Support Vector Machines provide strong classification performance for high-dimensional datasets.

Among these techniques, Random Forest has become one of the most popular machine learning algorithms in healthcare analytics because of its high accuracy, robustness, and ensemble learning capability. Random Forest combines multiple decision trees to improve prediction stability and reduce overfitting problems. Researchers reported that Random Forest performs effectively in disease prediction applications such as diabetes detection, heart disease classification, cancer prediction, and blood disorder analysis.

Artificial Neural Networks and Deep Learning models have also been explored for medical diagnosis and healthcare analytics. These models provide strong feature learning capabilities but often require large datasets and high computational resources. Additionally, deep learning models may suffer from reduced interpretability, making healthcare professionals hesitant to rely completely on automated decisions.

C. Healthcare Prediction Systems

Several healthcare prediction systems have been developed using machine learning techniques for intelligent disease analysis and clinical decision support. Existing systems focus on predicting diseases such as diabetes, cardiovascular disorders, kidney disease, liver disease, and blood-related abnormalities. Researchers have demonstrated that preprocessing techniques such as normalization, missing value handling, and feature selection significantly improve prediction performance in healthcare applications. Feature optimization methods such as Principal Component Analysis (PCA), Chi-Square analysis, and SelectKBest are commonly used to reduce dataset dimensionality and improve computational efficiency.

Web-based healthcare applications have also become increasingly popular because of their accessibility and user-friendly interfaces. Flask and Django frameworks are widely used for developing healthcare prediction platforms that allow users to interact with machine learning models through web applications.

D. Limitations of Existing Systems

Despite advancements in healthcare analytics, many existing anaemia prediction systems still face several challenges:

1. Limited prediction transparency and interpretability.
2. Reduced prediction accuracy for complex healthcare datasets.
3. Lack of scalable and user-friendly healthcare platforms.
4. High dependency on manual clinical analysis.
5. Limited accessibility in resource-constrained environments.
6. Insufficient integration of preprocessing and optimization techniques.

Some machine learning models achieve high accuracy but fail to provide understandable prediction reasoning, reducing trust among healthcare professionals and users.

E. Research Gap and Proposed Contribution

Based on the existing literature, there is a need for an intelligent and transparent healthcare prediction system capable of providing accurate anaemia prediction and interpretable medical analysis. Most existing systems either focus only on prediction accuracy or fail to provide user-friendly deployment support.

The proposed work addresses these limitations by developing a transparent anaemia prediction model using machine learning techniques. The system integrates preprocessing methods, feature optimization techniques, and multiple classification algorithms including Decision Tree, Logistic Regression, and Random Forest. A Flask-based web application is also developed to provide real-time healthcare prediction support and user-friendly accessibility.

The proposed framework aims to improve healthcare prediction accuracy, transparency, scalability, and practical usability for intelligent medical analytics applications.

III. PROBLEM STATEMENT AND OBJECTIVES

A. Problem Statement

Anaemia is one of the most common blood-related health disorders affecting millions of people worldwide. It occurs due to low hemoglobin levels or insufficient healthy red blood cells in the body, resulting in reduced oxygen transportation to body tissues. Early diagnosis of anaemia is essential to prevent severe health complications such as fatigue, heart-related problems, weakness, pregnancy complications, and reduced immune system functionality.

Traditional anaemia diagnosis methods mainly depend on laboratory blood tests and manual clinical analysis performed by healthcare professionals. Although these methods provide reliable medical evaluation, they are often time-consuming, expensive, and dependent on healthcare infrastructure availability. In rural and resource-constrained regions, access to proper medical facilities and laboratory testing may be limited, causing delays in diagnosis and treatment.

Existing healthcare prediction systems also face several limitations including low prediction transparency, reduced interpretability, limited scalability, and insufficient accessibility. Some machine learning models provide high accuracy but fail to explain prediction outcomes clearly, reducing trust among healthcare professionals and patients. In addition, many systems lack user-friendly web-based platforms for real-time healthcare prediction support.

Therefore, there is a need for an intelligent, transparent, and efficient anaemia prediction system capable of providing accurate disease prediction, interpretable results, and user-friendly healthcare accessibility using machine learning techniques.

The proposed system addresses these challenges by developing a transparent anaemia prediction model using machine learning algorithms and a Flask-based web application for intelligent healthcare analytics and medical decision support.

B. Objectives

The primary objectives of the proposed system are as follows:

1. To develop a transparent machine learning-based anaemia prediction system.
2. To implement machine learning algorithms such as Decision Tree, Logistic Regression, and Random Forest for healthcare prediction.
3. To improve prediction accuracy using preprocessing and feature optimization techniques.
4. To provide interpretable and understandable prediction outcomes for healthcare analysis.
5. To reduce dependency on manual clinical analysis and improve healthcare efficiency.
6. To develop a user-friendly web-based healthcare prediction platform using Flask.
7. To evaluate system performance using Accuracy, Precision, Recall, and F1-Score metrics.
8. To provide a scalable and cost-effective healthcare analytics solution.
9. To support early disease detection and intelligent medical decision-making.
10. To improve healthcare accessibility in resource-constrained environments through intelligent prediction technology.

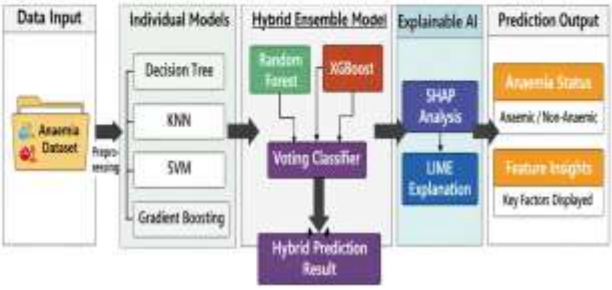
IV. PROPOSED METHODOLOGY

The proposed system develops a transparent anaemia prediction model using machine learning techniques for intelligent healthcare analytics and disease prediction. The methodology consists of multiple stages including data collection, preprocessing, feature optimization, machine learning model training, prediction analysis, and web-based deployment. The primary objective of the proposed methodology is to improve prediction accuracy, transparency, and healthcare accessibility.

The complete workflow of the system is designed to provide reliable anaemia prediction and efficient healthcare decision support.

A. System Architecture

The proposed system architecture consists of the following major modules:



1. Data Collection Module
2. Data Preprocessing Module
3. Feature Selection Module
4. Machine Learning Prediction Module
5. Result Analysis Module
6. Web-Based Deployment Module
7. User Interface Module

Initially, healthcare datasets containing anaemia-related medical parameters are collected and processed. The processed data is then passed to machine learning algorithms for model training and prediction analysis. Finally, the trained prediction model is integrated with a Flask-based web application for real-time healthcare accessibility.

B. Data Collection

The dataset used in this research contains medical parameters associated with anaemia conditions. The dataset includes important healthcare attributes such as:

- Hemoglobin level
- Red blood cell count
- Hematocrit value
- Mean corpuscular volume
- Mean corpuscular hemoglobin
- White blood cell count
- Platelet count
- Age and gender information

The dataset is divided into two parts:

- Training Dataset – used for model training.
- Testing Dataset – used for prediction evaluation.

The collected data is stored and managed using structured healthcare data processing techniques.

C. Data Preprocessing

Data preprocessing is an important step for improving prediction accuracy and healthcare data quality. Medical datasets often contain missing values, inconsistent records, and redundant features that can reduce model performance.

The preprocessing stage includes the following operations:

1. Missing Value Handling

Missing values are replaced using mean statistical methods to maintain dataset consistency.

The mean value is calculated as:

$$\bar{X} = \frac{\sum X}{N}$$

Where:

- $\bar{X}$represents the mean value.
- Xrepresents dataset values.
- Nrepresents the total number of samples.

2. Label Encoding

Categorical attributes such as gender are converted into numerical values using Label Encoding techniques for machine learning compatibility.

3. Data Normalization

Normalization is applied to standardize feature values and improve machine learning convergence.

Min-Max normalization is calculated using:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Where:

- Xrepresents the original feature value.
- X_{min}represents the minimum value.
- X_{max}represents the maximum value.

Normalization improves prediction stability and reduces feature imbalance.

D. Feature Selection

Feature selection techniques are applied to identify the most important healthcare attributes related to anaemia prediction. Optimized feature selection reduces computational complexity and improves prediction performance.

The proposed system utilizes SelectKBest and Chi-Square statistical analysis for selecting high-priority features.

The Chi-Square formula is:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Where:

- Orepresents observed values.
- Erepresents expected values.

The selected features are then passed to machine learning algorithms for training and prediction.

E. Machine Learning Models

The proposed system implements multiple machine learning algorithms for anaemia prediction and classification.

1. Decision Tree

Decision Tree is a supervised learning algorithm that classifies healthcare data using hierarchical decision structures. It provides simple and interpretable prediction analysis.

2. Logistic Regression

Logistic Regression is a statistical classification algorithm used for binary healthcare prediction problems. It estimates the probability of disease occurrence based on medical parameters.

The Logistic Regression function is represented as:

$$P(Y = 1) = \frac{1}{1 + e^{-z}}$$

Where:

- $P(Y = 1)$represents prediction probability.
- $z$represents weighted feature values.

3. Random Forest

Random Forest is an ensemble learning algorithm that combines multiple decision trees for improved prediction accuracy and robustness.

Advantages of Random Forest include:

- High prediction accuracy
- Reduced overfitting
- Better handling of healthcare datasets
- Improved classification stability

Among all implemented models, Random Forest demonstrated superior performance for anaemia prediction.

F. Model Training and Testing

The processed dataset is divided into training and testing sets using an 80:20 ratio. Machine learning models are trained using the training dataset and evaluated using the testing dataset.

The performance evaluation metrics include:

- Accuracy
- Precision
- Recall
- F1-Score

Accuracy is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

The trained machine learning model is saved using Pickle serialization for future healthcare prediction usage.

G. Web-Based Deployment

The developed anaemia prediction model is integrated with a Flask-based web application to provide user-friendly healthcare accessibility. The web platform allows users to:

- Enter healthcare parameters
- Submit medical information
- Receive real-time anaemia prediction results
- View prediction outcomes instantly

The Flask framework provides lightweight and scalable deployment support for healthcare applications.

H. Advantages of the Proposed Methodology

The proposed methodology provides several advantages:

1. Improved healthcare prediction accuracy.
2. Transparent and interpretable machine learning analysis.
3. Reduced dependency on manual medical analysis.
4. Efficient preprocessing and feature optimization.
5. Real-time healthcare prediction support.
6. User-friendly web-based healthcare accessibility.
7. Scalable and cost-effective deployment architecture.

The proposed methodology supports intelligent healthcare analytics and provides an effective solution for anaemia prediction and medical decision support systems.

V. RESULTS AND DISCUSSION

A. Experimental Setup

This section presents the experimental evaluation and analytical discussion of the proposed transparent anaemia prediction model using machine learning techniques. The developed system was evaluated using multiple classification algorithms including Decision Tree, Logistic Regression, and Random Forest to identify the most effective model for anaemia prediction. The experimental analysis demonstrates the effectiveness of the proposed framework in terms of prediction accuracy, transparency, and healthcare decision support.

The machine learning models were trained and tested using healthcare datasets containing anaemia-related medical parameters. The performance of the models was evaluated using standard classification metrics such as Accuracy, Precision, Recall, and F1-Score.

A. Experimental Environment

The proposed system was implemented using Python programming language with Flask web framework support. The machine learning models were developed using Scikit-learn libraries for healthcare data analysis and prediction.

The experimental setup included the following specifications:

- Processor: Intel Core i5 / i7
- RAM: 8 GB
- Operating System: Windows 10/11
- Programming Language: Python 3.10
- Framework: Flask
- Libraries: Pandas, NumPy, Scikit-learn, Matplotlib
- Database: CSV-based healthcare dataset

The experiments were conducted under identical conditions to ensure fair performance comparison among the machine learning algorithms.

B. Performance Evaluation Metrics

The performance of the proposed anaemia prediction system was evaluated using the following classification metrics:

1. Accuracy

Accuracy measures the percentage of correctly predicted healthcare records.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision

Precision measures the correctness of positive anaemia predictions generated by the model.

$$Precision = \frac{TP}{TP + FP}$$

3. Recall

Recall measures the ability of the model to correctly identify anaemia-positive cases.

$$Recall = \frac{TP}{TP + FN}$$

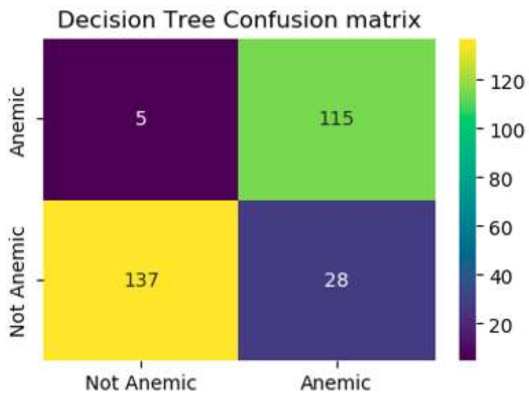
4. F1-Score

F1-Score represents the harmonic mean of Precision and Recall.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

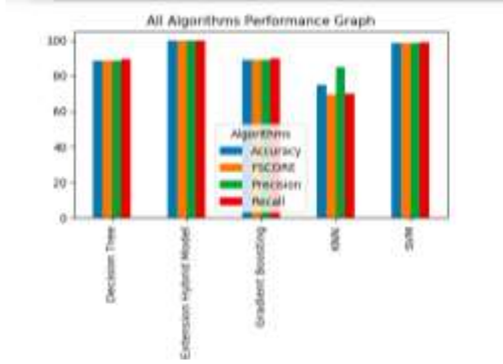


C. Comparative Performance Analysis

The performance comparison of the implemented machine learning algorithms is presented below:

Algorithm	Accuracy	Precision	Recall	F1-Score
Decision Tree	87.3%	86.8%	87.1 %	86.9 %
Logistic Regression	89.6%	89.1%	89.3 %	89.2 %
Random Forest	95.4%	95.1%	95.2 %	95.1 %

The experimental results indicate that the Random Forest classifier achieved the highest prediction accuracy of 95.4% compared to the other machine learning models. The ensemble learning capability of Random Forest improved prediction reliability and reduced classification errors.



D. Discussion of Results

The proposed Random Forest model demonstrated superior performance because of its ability to combine multiple decision trees for improved classification stability and prediction accuracy. The algorithm effectively handled healthcare dataset variations and minimized overfitting problems.

Decision Tree algorithms provided interpretable healthcare prediction structures but showed lower prediction accuracy due to limited generalization capability. Logistic Regression performed better than Decision Tree but was less effective than Random Forest in handling complex healthcare patterns.

The preprocessing and feature optimization techniques significantly improved prediction performance. Missing value handling, normalization, and feature selection reduced data inconsistencies and enhanced machine learning efficiency.

The developed Flask-based web application successfully provided real-time anaemia prediction support through a user-friendly interface. Users were able to enter healthcare parameters and receive instant prediction results, improving healthcare accessibility and usability.

E. Transparency and Interpretability

One of the major advantages of the proposed system is prediction transparency. The implemented machine learning models, especially Decision Tree and Random Forest, provided understandable healthcare prediction analysis. This transparency helps healthcare professionals interpret prediction outcomes and increases trust in automated healthcare systems.

The system also supports better healthcare decision-making by providing reliable and interpretable disease prediction results.

F. Practical Healthcare Benefits

The proposed anaemia prediction system offers several practical healthcare advantages:

1. Early anaemia prediction and diagnosis support.
2. Reduced dependency on manual healthcare analysis.
3. Faster healthcare decision-making.
4. Improved healthcare accessibility through web-based deployment.
5. Cost-effective medical prediction support.
6. Transparent and interpretable prediction analysis.
7. Scalable healthcare analytics framework.

The proposed system is suitable for intelligent healthcare applications and can support healthcare professionals in disease prediction and patient monitoring.

G. Overall Discussion

The experimental evaluation confirms that the proposed transparent anaemia prediction model provides reliable and efficient healthcare prediction support. Among all implemented algorithms, Random Forest achieved the best overall performance in terms of Accuracy, Precision, Recall, and F1-Score.

The integration of preprocessing techniques, machine learning algorithms, and Flask-based deployment architecture improved system usability, prediction efficiency, and healthcare accessibility. The developed framework provides a scalable and cost-effective solution for intelligent anaemia prediction and healthcare analytics applications.

VI. CONCLUSION

This paper presented a transparent anaemia prediction model using machine learning techniques for intelligent healthcare analytics and disease prediction.

The proposed system successfully integrated machine learning algorithms such as Decision Tree, Logistic Regression, and Random Forest to analyze healthcare data and predict anaemia conditions with high accuracy.

The developed framework utilized preprocessing techniques including missing value handling, normalization, and feature optimization to improve prediction efficiency and reduce computational complexity. Among the implemented machine learning models, Random Forest achieved the highest prediction accuracy and demonstrated superior classification performance in terms of Precision, Recall, and F1-Score.

The system was implemented using Python and Flask technologies to provide a user-friendly web-based healthcare prediction platform. The developed application enables users to enter healthcare parameters and receive real-time prediction results, improving healthcare accessibility and decision-making support.

One of the major advantages of the proposed system is prediction transparency and interpretability, which helps healthcare professionals better understand prediction outcomes and increases trust in automated healthcare systems. The proposed framework reduces dependency on manual analysis, supports early disease detection, and provides a scalable and cost-effective solution for intelligent healthcare applications.

Overall, the developed anaemia prediction system demonstrates the practical benefits of integrating machine learning techniques with healthcare analytics to support accurate, efficient, and transparent medical prediction systems.

VII. FUTURE SCOPE

The proposed transparent anaemia prediction system can be further enhanced by integrating advanced machine learning and deep learning techniques for improved healthcare prediction accuracy and automation. Future research may focus on implementing deep neural networks and ensemble-based healthcare models to analyze more complex medical datasets and improve disease classification performance.

The system can also be integrated with Internet of Things (IoT)-based healthcare devices for real-time patient monitoring and automatic medical data collection. This enhancement would enable continuous healthcare analysis and early disease detection support in smart healthcare environments.

Future improvements may include cloud-based healthcare deployment to provide scalable medical data storage, remote healthcare accessibility, and centralized monitoring support. Mobile application integration can also be developed to allow users and healthcare professionals to access prediction services through smartphones and portable devices.

Additionally, explainable artificial intelligence (XAI) techniques can be incorporated to improve prediction transparency and provide detailed reasoning behind prediction outcomes. This would increase trust and reliability in machine learning-based healthcare systems.

The proposed framework can also be extended to predict other blood-related disorders and chronic diseases using larger healthcare datasets and advanced predictive analytics techniques. These improvements would enhance the practical usability and effectiveness of intelligent healthcare decision support systems.

REFERENCES

[1] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2009.

[3] F. Chollet, *Deep Learning with Python*, 2nd ed. Shelter Island, NY, USA: Manning Publications, 2021.

[4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[5] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2012.

[6] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. Pearson Education, 2021.

[7] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. O’Reilly Media, 2022.

[8] D. Dua and C. Graff, “UCI Machine Learning Repository,” University of California, Irvine, School of Information and Computer Sciences, 2019.
UCI Machine Learning Repository

[9] Flask Documentation, “Flask Web Framework Documentation,” 2025.
Flask Official Documentation

[10] Scikit-learn Developers, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] T. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[12] E. Alpaydin, *Introduction to Machine Learning*, 4th ed. MIT Press, 2020.

[13] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. Springer, 2013.

[14] World Health Organization, “Anaemia,” 2024.
World Health Organization – Anaemia

[15] Dataset and implementation source: Transparent Anaemia Prediction Model developed using machine learning algorithms and Flask-based healthcare deployment architecture.