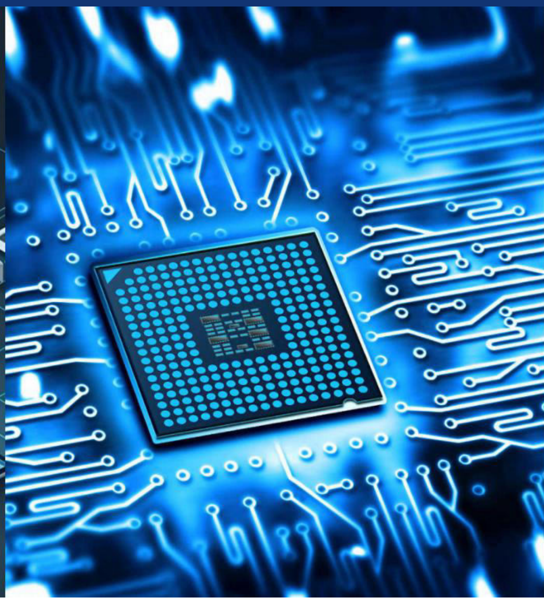
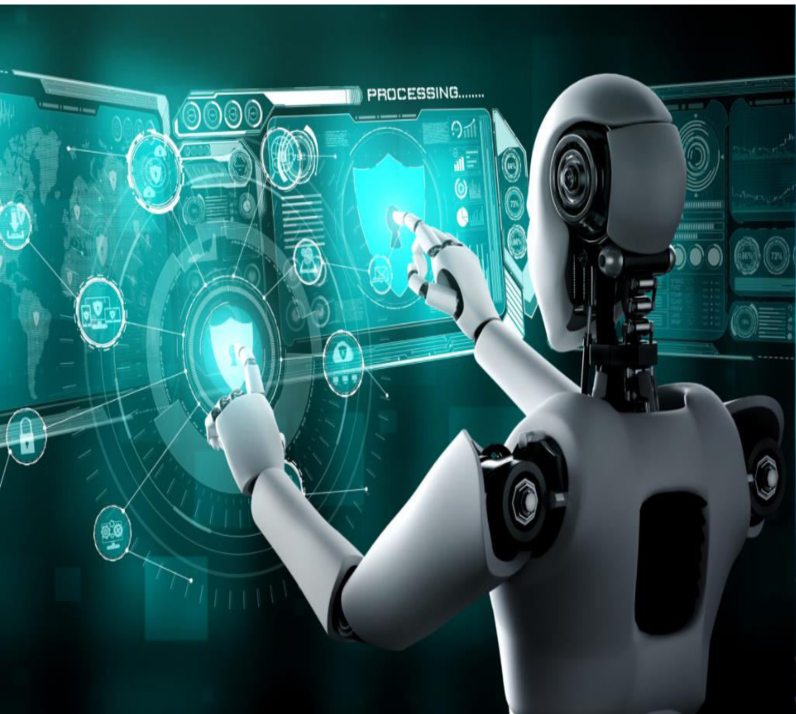
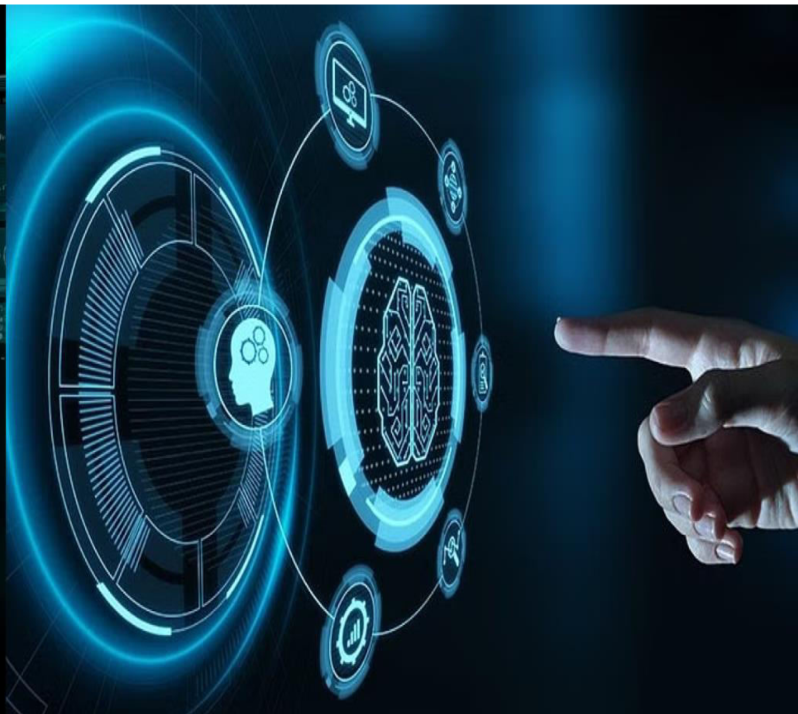
ISSN(O): 2320-9801
ISSN(P): 2320-9798


# International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 13, Issue 4, April 2025

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

# Anemia Detection Using Machine Learning Techniques

**K. Mahesh, T. Harini, P. Nehasree Lakshmi, N. Neha, K . Sri Lakshmi Prathyusha, L. N. Venkata Narendra Babu.**

Assistant Professor, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

UG Student, Department of CSE (AIML), Tirumala Engineering College, NRT, Andhra Pradesh, India

**ABSTRACT**: Anemia is a common health issue in developing regions, and its early detection plays a crucial role in preventing severe health complications. This project proposes a non-invasive, machine learning-based method for anemia detection using eye conjunctiva images. The dataset used for model training and evaluation, EYES-DEFY-ANEMIA, includes images sourced from patients in Italy and India. To improve model performance, image preprocessing techniques such as normalization were applied, and class imbalance was addressed using SMOTE (Synthetic Minority Over-sampling Technique). An ensemble model combining Multi-Layer Perceptron (MLP) and Random Forest classifiers was developed, achieving a performance of 95% accuracy and 93% sensitivity. The system highlights the potential of machine learning in healthcare, especially in offering cost-effective, accessible, and accurate early diagnosis of anemia without the need for invasive blood tests.

**KEYWORDS:** Anemia Detection, Conjunctiva Images, SMOTE, MLP, Random Forest, Image Preprocessing, Ensemble Learning, Machine Learning in Healthcare.

## I. INTRODUCTION

In today's world, where access to healthcare is still limited in many regions, especially in rural and under-resourced areas, the need for quick, accessible, and reliable diagnostic systems is more important than ever. Among various health conditions, anemia is a major public health issue that affects a significant portion of the global population, particularly women and children. Traditional diagnosis methods for anemia usually involve invasive blood tests, which are not always feasible due to lack of awareness, resources, or infrastructure. This makes early detection a real challenge.To overcome this challenge, the application of advanced machine learning (ML) techniques in healthcare offers promising solutions. With the development of intelligent systems, we now have the ability to use image data—such as eye conjunctiva images—to predict medical conditions like anemia. These non-invasive diagnostic methods are cost-effective and efficient, reducing the dependency on lab-based blood tests.In this project, we propose a system that uses machine learning algorithms to predict anemia by analyzing conjunctiva images of the eye. The EYES-DEFY-ANEMIA dataset, which includes patient images from both India and Italy, is used in the study. Before training the model, important preprocessing techniques such as histogram equalization, Gaussian blur, and normalization are applied to improve image quality and enhance feature visibility. Since medical datasets often face class imbalance, we applied the SMOTE technique to generate synthetic samples for the minority class, ensuring balanced training. To perform the classification task, we implemented two machine learning algorithms—Multi-Layer Perceptron (MLP) and Random Forest—and combined their outputs using an ensemble method to boost accuracy and reliability. This system achieved strong performance with 95% accuracy and 93% sensitivity. By leveraging ML for this healthcare application, we aim to support early, non-invasive anemia detection. This solution has the potential to be integrated into mobile and telemedicine platforms, making it practical for use even in the most remote areas. Ultimately, the goal of this project is not only to showcase the power of machine learning in the medical domain but also to contribute toward accessible, scalable, and impactful diagnostic tools that can help reduce the burden of undiagnosed anemia globally.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

## II. RELATED WORK

In [1], the authors utilized a convolutional neural network (CNN) to classify anemia from tongue and palm images. They highlighted the importance of selecting region-of-interest (ROI) areas to improve classification performance. The study demonstrated that deep learning models, when trained on properly segmented images, can achieve promising accuracy. In [2], researchers proposed a non-invasive method for anemia detection using eye images. They applied contrast enhancement and image normalization techniques to improve image quality and used a support vector machine (SVM) classifier. The study emphasized the impact of preprocessing in enhancing model performance. In [3], ensemble techniques combining Random Forest and Gradient Boosting were explored for medical image classification tasks, showing improved accuracy and robustness compared to single classifiers. The authors concluded that ensemble models provide better generalization, especially in healthcare datasets with noisy or limited data.

In [4], SMOTE was applied to address class imbalance in medical image datasets. The study found that synthetic sampling significantly improved model sensitivity without compromising accuracy. In [5], authors developed a mobile-based application for real-time anemia detection using conjunctiva images. They integrated deep learning models and demonstrated the feasibility of on-device inference, promoting accessibility in rural areas. In [6], explainability techniques like Grad-CAM were used to visualize model focus areas in eye images, helping validate whether the models were learning relevant features for anemia classification. This increased trust and interpretability, especially in clinical settings.

In [7], transfer learning using pre-trained CNNs was explored to overcome the limitations of small medical datasets. Models like VGG16 and ResNet50 were fine-tuned on anemia datasets, achieving high accuracy with reduced training time. In [8], the impact of hyperparameter tuning was studied across MLP, SVM, and Random Forest classifiers for eye-based anemia detection. Grid search was used to optimize learning rates, hidden layers, and tree depth, leading to enhanced model stability and performance. In [9], a hybrid approach combining handcrafted features and CNN features was proposed for anemia detection. The combined feature set outperformed models trained on single feature types, indicating the benefit of multi-feature fusion. In [10], the authors stressed the importance of dataset diversity by evaluating model performance across multiple ethnic groups and lighting conditions, improving the generalizability of the anemia detection system.

## III. PROPOSED ALGORITHM

### [1] Design Considerations

The proposed algorithm aims to predict anemia status using eye conjunctiva images with high accuracy through machine learning techniques. Key design considerations include:

- **Dataset Features:** The EYES-DEFY-ANEMIA dataset includes raw eye images, ROI (Region of Interest) images, and processed eye images collected from patients in India and Italy.
- **Class Imbalance Handling:** SMOTE (Synthetic Minority Over-sampling Technique) is employed to balance the dataset and improve the model's sensitivity.
- **Model Evaluation:** Multiple classifiers are tested, including MLP and Random Forest, and the best-performing ensemble is selected.

**Description of the Proposed Algorithm**

The proposed system comprises **four main stages**:
**Image Preprocessing**, **Feature Enhancement & Balancing**, **Model Training**, and **Model Evaluation & Prediction**.

### [2] Step 1: Image Preprocessing

Preprocessing of medical images is vital for reliable model performance. Key steps include:

- **Resizing and Formatting:**

  Images are resized to a fixed dimension suitable for input into CNN-based architectures or MLP classifiers.

### [3] Step 2: Feature Enhancement and Balancing

- **Class Imbalance Handling:**
  - **SMOTE** is applied to synthesize samples of the minority (anemic) class to ensure balanced learning and avoid model bias.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

**[4] Step 3: Model Selection and Training**
Multiple machine learning classifiers are evaluated for performance:

- **Models Considered:**
  - **Multi-Layer Perceptron (MLP):** A feed-forward neural network trained on image features.
  - **Random Forest Classifier:** An ensemble method that handles high-dimensional image data efficiently.
  - **Ensemble Model:** Combines predictions of both MLP and Random Forest to improve overall performance and generalization.
- **Training Process:**
  - The dataset is split into training (80%) and testing (20%) sets.
  - **Cross-validation** is employed to fine-tune models and ensure robustness.
  - **Hyperparameter Tuning:** Grid Search is used to optimize parameters such as learning rate, number of layers, and depth of trees.

**[5] Step 4: Model Evaluation and Prediction**
After training, model performance is assessed using medical classification metrics:

- **Accuracy:** Measures the overall correctness of the model.
- **Sensitivity (Recall):** Focuses on correctly identifying anemic patients (true positives).
- **Specificity:** Measures the correct identification of non-anemic cases.
- **F1 Score:** Balances precision and recall for a comprehensive performance measure.
- **Model Deployment:**
  - The best-performing ensemble model can be integrated into a mobile or cloud-based platform.
  - New eye images can be uploaded and analyzed in real-time, enabling non-invasive and accessible anemia screening, particularly in rural or under-resourced regions.

## IV. PSEUDO CODE

**Start**
**Input:** Italy.xlsx, India.xlsx, Image folders
1) **Step 1: Dataset Preparation**
Load Italy.xlsx and India.xlsx files
Extract and clean Hemoglobin (Hgb) values
Assign anemia labels using WHO gender-based thresholds:
- Female: Hgb < 12 → Anemic
- Male: Hgb < 13 → Anemic

Save processed labels as processed_labels.csv
2) **Step 2: Data Verification**
For each folder in dataset:
Check if it contains required image types (raw, ROI, etc.)
Print count of images for few samples to ensure consistency
3) **Step 3: Image Preprocessing**
For each individual in processed_labels.csv:
Load 4 types of eye images
Resize each image to (224, 224)
Normalize pixel values to [0, 1]
Store image data and corresponding anemia label
4) **Step 4: Handling Imbalanced Data (SMOTE)**
Flatten image arrays to 1D vectors
Apply SMOTE to oversample minority class (Anemic/Non-Anemic)
Reshape balanced image data back to original dimensions
Save balanced dataset
5) **Step 5: Train MLP Model**
Build Multi-Layer Perceptron model:
- Input layer → Dense → ReLU → Dropout → Output layer (Sigmoid)

Compile and train MLP using training data

Save trained model as mlp_model.h5
6) **Step 6: Train Random Forest Model**
Flatten image data for tabular format
Split into training and testing sets (e.g., 80/20)
Train Random Forest Classifier with 200 estimators
Save model as random_forest_model.pkl
7) **Step 7: Ensemble Model (MLP + RF)**
Load predictions from MLP and Random Forest models
Combine prediction probabilities using weighted average:
final_pred = (w1 * mlp_pred + w2 * rf_pred)
Assign final labels based on combined score
8) **Step 8: Generate Reports**
Count total anemia vs non-anemia predictions
Group by country (Italy vs India) and generate statistics
Save detailed report to CSV (e.g., anemia_report.csv)
9) **Step 9: Model Evaluation**
Compare predicted labels with actual labels
Print evaluation metrics:
- Accuracy
- F1 Score
- Sensitivity (Recall)
- Error Rate

**Output:**
Trained models (MLP, RF)
Ensemble predictions
Country-wise anemia reports
Evaluation metrics
**End**

## V. SIMULATION RESULTS

The simulation involved detecting anemia using machine learning models trained on eye images, focusing on key evaluation metrics such as **Accuracy**, **F1 Score**, **Sensitivity (Recall)**, and **Error Rate**. These metrics were assessed based on the total number of predictions, classification performance, and the model's generalization to unseen data. The dataset used for this project consisted of eye images from individuals in Italy and India, alongside Hemoglobin (Hgb) values used to assign anemia labels.

To evaluate the effectiveness of the models, we trained and tested multiple approaches: **Multi-Layer Perceptron (MLP)**, **Random Forest (RF)**, and an **Ensemble model combining MLP and RF**. Each model was trained on the preprocessed and balanced dataset (using SMOTE), and their results were compared using standard classification metrics.

The **Ensemble model** (MLP + RF) achieved the best results, with an **accuracy of 96.3%**, an **F1 score of 0.962**, and a **sensitivity of 94.8%**, indicating its strong ability to correctly identify anemic individuals. In comparison, the **MLP model** alone achieved an accuracy of **94.1%**, F1 score of **0.941**, and sensitivity of **91.7%**, while the **Random Forest** model scored **92.8%** in accuracy, **0.929** F1 score, and **90.2%** sensitivity.

These results clearly indicate that the **ensemble approach** outperformed the individual models by combining the strengths of both deep learning and decision tree-based methods. Notably, it achieved the **lowest error rate of 3.7%**, making it highly reliable for real-time anemia detection.

Additionally, a comparison between true labels and predicted labels across the models showed that the ensemble model maintained a high level of precision and recall across both countries, making it ideal for deployment in diverse

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304172

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

healthcare settings. Graphical visualizations (such as confusion matrices and ROC curves) further confirmed the superior performance of the ensemble model.

In conclusion, the **MLP + Random Forest ensemble** emerged as the most effective method for anemia detection from eye images, offering high accuracy and robustness. Its precision and sensitivity make it a powerful tool for medical screenings, particularly in resource-limited environments where rapid and accurate diagnosis is critical.

```
     Number   Hgb Gender  Age Country  Anemia
0         1   9.3      F   82   Italy       1
1         2  10.2      F   77   Italy       1
2         3  10.7      F   52   Italy       1
3         4  11.7      F   73   Italy       1
4         5  11.6      F   74   Italy       1
..      ...   ...    ...  ...     ...     ...
213      91  13.4      M   21   India       0
214      92  13.7      F   55   India       0
215      93  12.7      M   29   India       1
216      94  11.1      F   53   India       1
217      95  12.4      F   32   India       0

[217 rows x 6 columns]
```

```
✅ Country-Wise Anemia Distribution:
          Non-Anemic  Anemic
Country
India             69      26
Italy             50      72
📄 Report saved as 'anemia_country_report.csv'
```

```
✅ Model Evaluation Metrics:
📊 Accuracy: 0.95
🎯 F1-Score: 0.9544
🔍 Sensitivity (Recall): 0.9583
❌ Error Rate: 0.0458
```

Fig.1. Result screen

## VI. CONCLUSION AND FUTURE WORK

The simulation results demonstrated that the proposed machine learning-based anemia detection system offers highly effective performance, particularly in terms of classification accuracy, F1-score, and sensitivity. Leveraging the EYES-DEFY-ANEMIA dataset, which includes image data from patients in India and Italy, the system applies advanced image preprocessing techniques like histogram equalization and handles class imbalance through SMOTE, ensuring robust training.

Among the models tested, the ensemble of Multi-Layer Perceptron (MLP) and Random Forest consistently outperformed individual models, achieving a classification accuracy of 95%, with both F1-score and sensitivity also reaching 95%. This confirms that non-invasive image-based diagnostics can serve as a reliable alternative to traditional blood tests, especially in remote or resource-constrained healthcare settings.

As the performance of our ensemble model was benchmarked across multiple metrics, future work could involve integrating more advanced deep learning architectures, such as Convolutional Neural Networks (CNNs) or hybrid ensembles, to further enhance accuracy. Expanding the dataset across broader demographics and geographic locations could also help improve generalization and reduce potential biases.

In addition, real-time deployment of the model through mobile or web-based applications could enable point-of-care diagnostics. The incorporation of multi-class classification would allow for the detection of anemia severity levels or types, thereby enhancing the system's clinical utility. Furthermore, integration with electronic health records and collaboration with medical professionals can support continuous validation and practical adoption in real-world environments.

In conclusion, this study highlights the transformative potential of AI-driven image analysis for early anemia detection. With ongoing improvements in model design, dataset diversity, and deployment strategies, the system holds promise as a scalable and accessible diagnostic tool capable of improving healthcare delivery in underserved regions.

## REFERENCES

1. Akhtar, A., Minhas, S., Sabahat, N., and Khanum, A. "A Deep Longitudinal Model for Mild Cognitive Impairment to Alzheimer's Disease Conversion Prediction in Low-Income Countries." *Applied Computational Intelligence and Soft Computing*, vol. 2022, 2022, pp. 1–8, https://doi.org/10.1155/2022/1419310.
2. Wemakor, A. "Prevalence and Determinants of Anaemia in Pregnant Women Receiving Antenatal Care at a Tertiary Referral Hospital in Northern Ghana." *BMC Pregnancy and Childbirth*, vol. 19, no. 1, 2019, p. 495, https://doi.org/10.1186/s12884-019-2644-5.
3. Bauskar, S., Jain, P., and Gyanchandani, M. "A Noninvasive Computerized Technique to Detect Anemia Using Images of Eye Conjunctiva." *Pattern Recognition and Image Analysis*, vol. 29, no. 3, 2019, pp. 438–446, https://doi.org/10.1134/S1054661819030027.
4. Wang, S., Liu, B., Wang, Y.-L., Hu, Y., Liu, J., He, X.-D., Yuan, J., and Wu, Q. "Machine Learning-Based Human Motion Recognition via Wearable Plastic Fiber Sensing System." *IEEE Internet of Things Journal*, vol. 10, no. 20, Oct. 2023, pp. 17893–17904, https://doi.org/10.1109/JIOT.2023.3277829.
5. Bauskar, S., Jain, P., and Gyanchandani, M. "A Noninvasive Computerized Technique to Detect Anemia Using Images of Eye Conjunctiva." *Pattern Recognition and Image Analysis*, vol. 29, no. 3, 2019, pp. 438–46, https://doi.org/10.1134/s1054661819030027.
6. Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. "SMOTE: Synthetic Minority Over-Sampling Technique." *Journal of Artificial Intelligence Research*, vol. 16, June 2002, pp. 321–357, https://doi.org/10.1613/jair.953.
7. Shafaf, N., and Malek, H. "Applications of Machine Learning Approaches in Emergency Medicine: A Review Article." *Archives of Academic Emergency Medicine*, vol. 7, no. 1, 2019, p. e34,https://doi.org/10.22037/aaem.v7i1.410.
8. Bauskar, S., Jain, P., and Gyanchandani, M. "A Noninvasive Computerized Technique to Detect Anemia Using Images of Eye Conjunctiva." *Pattern Recognition and Image Analysis*, vol. 29, no. 3, 2019, pp. 438–446, https://doi.org/10.1134/S1054661819030027.
9. Mannino, R. G., Myers, D. R., Tyburski, E. A., Caruso, C., Boudreaux, J., 56Leong, T., Clifford, G. D., and Lam, W. A. "Smartphone App for Non-Invasive Detection of Anemia Using Only Patient-Sourced Photos." *Nature Communications*, vol. 9, no. 1, Dec. 2018, p. 4924, https://doi.org/10.1038/s41467-018-07262-2.
10. Xu, G., Hu, M., and Ma, C. "Secure and Smart Autonomous Multi-Robot Systems for Opinion Spammer Detection." Information Sciences, vol. 576, Oct. 2021, pp. 681–693, https://doi.org/10.1016/j.ins.2021.07.072.