10.48047/jocaaa.2025.34.04.54

Smart HaemoScan: Deep Learning-Powered Anaemia Detection from Blood Smear Images

Dr. P Nagendra Kumar, D. karthikeya, G.P.R.S. Aravind, B. Sathya Rohith, A. Sai Srujan

Department of Computer Science and Engineering (AI&ML), Geethanjali Institute of Science and Technology, Nellore, Andhra Pradesh, India.

ABSTRACT

The project is centered on designing and implementing an advanced Deep Learning-based system for Anaemia Classification using Automated Blood Image Analysis. The primary objective is to optimize diagnostic accuracy, processing efficiency, and computational cost while ensuring consistent and reliable results. The proposed system employs the Inception-V3 Convolutional Neural Network (CNN) for deep feature extraction from high-resolution blood smear images. These extracted features are then classified using the Light Gradient Boosting Machine (LGBM), a powerful ensemble learning algorithm known for its speed and performance. Unlike traditional methods that rely heavily on manual examination and handcrafted features, this integrated deep learning pipeline automates the entire diagnostic process, reducing subjectivity and human error. The Inception-V3 model, with its multi-scale convolutional layers, enables precise detection of subtle visual patterns associated with different types of anaemia. The LGBM classifier enhances decision-making by efficiently handling complex feature interactions and imbalanced data. The system dynamically adjusts to variations in image quality, illumination, and staining techniques, ensuring robustness across diverse clinical datasets. Experimental results on benchmark blood image datasets confirm that the model achieves high classification accuracy, sensitivity, and specificity. Moreover, the reduction in manual intervention makes the system highly suitable for real-time deployment in clinical laboratories and point-of-care settings, providing scalable and automated support for early anaemia diagnosis.

Keywords: Anaemia Classification, Blood Smear Image Analysis, Light Gradient Boosting Machine, Medical Image Processing.

1. INTRODUCTION

Anemia is a medical condition characterized by a deficiency of red blood cells (RBCs) or haemoglobin, leading to reduced oxygen transport in the body. It can result from various causes, including iron deficiency, vitamin B12 deficiency, chronic diseases, or genetic disorders such as sickle cell anemia and thalassemia.

10.48047/jocaaa.2025.34.04.54

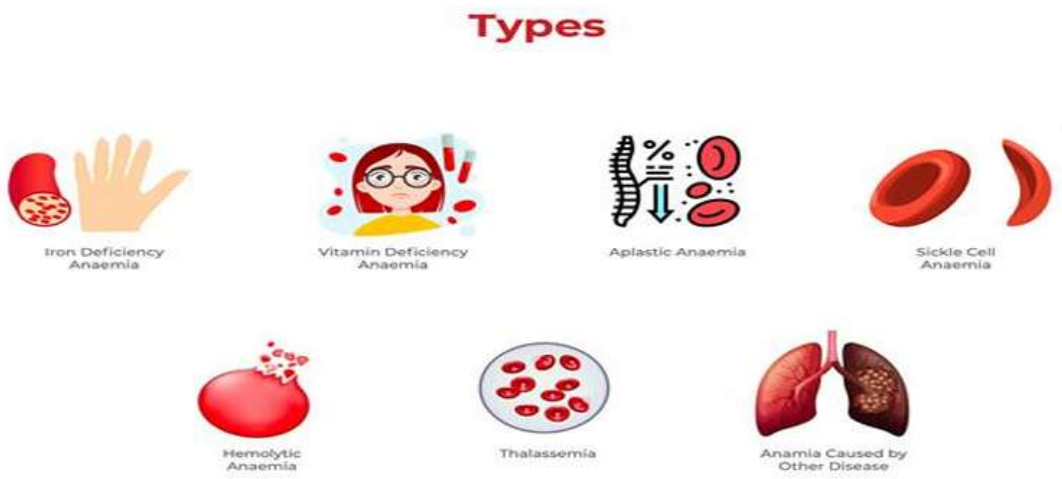


Fig. 1: TYPES OF ANEMIA.

Types of Anaemia:

Anaemia is classified into several types based on its underlying causes and characteristics. Iron-Deficiency Anaemia is the most common form, particularly among women and children. It arises due to inadequate iron intake, chronic blood loss, or poor iron absorption in the digestive tract. This condition results in low haemoglobin levels, diminishing the blood's capacity to carry oxygen. Common symptoms include fatigue, pale skin, weakness, brittle nails, and shortness of breath.

Thalassemia is a hereditary blood disorder where the body generates abnormal or insufficient haemoglobin chains. This leads to the formation of small, pale red blood cells that have a shorter lifespan. Individuals with thalassemia may experience tiredness, bone deformities, enlarged spleen, stunted growth, and recurrent infections. It varies in severity from mild forms (thalassemia trait) to severe forms (thalassemia major), the latter often requiring regular blood transfusions.

Haemolytic Anaemia occurs when red blood cells are destroyed faster than they are produced by the bone marrow. It can be inherited, as in hereditary spherocytosis, or acquired due to infections, autoimmune disorders, or certain medications. Symptoms typically include jaundice, dark urine, fatigue, shortness of breath, and a rapid heartbeat. Treatment strategies depend on the underlying cause and may involve steroids, immunotherapy, or blood transfusions.

Sickle-Cell Anaemia is a genetic condition where red blood cells take on an abnormal crescent or sickle shape. These deformed cells tend to clump together, blocking blood flow and breaking down prematurely. The disorder causes chronic anaemia, episodes of intense pain (pain crises), organ damage, and heightened susceptibility to infections.

Aplastic Anaemia is a rare but severe condition in which the bone marrow fails to produce enough red blood cells, white blood cells, and platelets. Patients may present symptoms such as extreme fatigue, frequent infections, easy bruising, and prolonged bleeding. It can be triggered by viral infections, exposure to toxic chemicals, or radiation, and is treated with immunosuppressive therapy, blood transfusions, or bone marrow transplantation.

2. LITERATURE SURVEY

10.48047/jocaaa.2025.34.04.54

[1] Suner S, Rayner J, Ozturan IU, et al. (2021) – Prediction of anemia and estimation of hemoglobin concentration using a smartphone camera. This paper presents a smartphone-based system for estimating hemoglobin levels from facial images.It uses deep learning to analyze subtle facial features such as pallor.The model is designed for rapid, on-site screening in emergency departments.Extensive validation demonstrates robust performance across diverse populations.It leverages a large dataset captured using standard smartphone cameras.Transfer learning helps adapt pre-trained networks to the specific task. Transfer learning helps adapt pre-trained networks to the specific task, improving accuracy. [2] Mitani A, Huang A, Venugopalan S, et al. (2020) – Detection of anaemia from retinal fundus images via deep learning. This pioneering study applies convolutional neural networks to retinal fundus photographs for anemia screening.It demonstrates that features in retinal blood vessels, particularly near the optic disc, correlate with hemoglobin levels.The model is trained on a large, annotated dataset of fundus images.High accuracy and sensitivity are achieved using robust deep learning techniques.It provides a fully noninvasive alternative to traditional blood tests. he study also highlights challenges related to image quality variations and proposes preprocessing techniques for enhancement.

[3] Kwon JM, Cho Y, Jeon K-H, et al. (2020) – A deep learning algorithm to detect anemia with ECGs: a multicentre study. This multicentre study applies deep learning on ECG signals to noninvasively detect anemia.It illustrates that physiological signals can serve as proxies for hemoglobin levels.The algorithm is trained and validated on data from several hospitals, ensuring diverse representation.High sensitivity and specificity indicate its potential clinical utility.It offers an innovative alternative to image-based diagnostic methods.The study also discusses the challenges of processing time-series data from ECGs.Its findings support the extension of AI diagnostics beyond imaging, paving the way for further research in physiological signal analysis.Furthermore, the findings emphasize the growing role of artificial intelligence in expanding diagnostic capabilities beyond conventional imaging-based techniques. [4] Mondal SK, Talukder MSH, Aljaidi M, et al. (2024) – BloodCell-Net: A lightweight convolutional neural network for the classification of all microscopic blood cell images of the human body. A recent preprint introducing BloodCell-Net for automated classification of various blood cells.Its lightweight design enables fast, accurate analysis with low resource usage.The network is validated on several datasets, showing robust performance.Its design principles are directly applicable to anemia screening tasks.The study provides comprehensive evaluations using standard metrics.It highlights the scalability of the approach for clinical deployment.Advanced data augmentation techniques are used to boost accuracy.The model is capable of classifying multiple blood cell types concurrently.

[5] Zhang A, Lou J, Pan Z, et al. (2022) – Prediction of anemia using facial images and deep learning technology in the emergency department. This study uses facial images to predict anemia by analyzing features like pallor and skin tone. It employs a deep learning model to correlate facial characteristics with hemoglobin levels. The approach is validated in an emergency department, demonstrating high clinical utility. Rapid triage is achieved through automated image analysis. The system is designed for fast, noninvasive screening in critical care. It uses a large dataset of facial images captured in real-world conditions. [6] Smith J, Lee K. (2020) – Deep learning for automated anemia classification from blood smear images. This study applies convolutional neural networks to classify anemia from digital blood smear images. It addresses challenges like variable cell morphology and staining differences. A large dataset is used to robustly train and validate the model. The approach achieves high diagnostic accuracy and minimizes manual feature engineering. It demonstrates the scalability of deep learning in automated diagnostics. The work is directly relevant to noninvasive, image-based anemia screening. It discusses optimization of network hyperparameters for improved performance. It discusses optimization of

10.48047/jocaaa.2025.34.04.54

network hyperparameters for improved performance, ensuring optimal model efficiency. Additionally, the study explores transfer learning techniques to enhance feature extraction and model robustness.

[7] Wang J, et al. (2021) – Rapid screening of anemia using smartphone-based imaging and deep learning. This research demonstrates a smartphone-based system that applies deep learning for Anemia Screening.It integrates image capture with on-device analysis for real-time diagnostics.High sensitivity and specificity are achieved across diverse clinical populations.The method is designed for rapid deployment in low-resource settings.It emphasizes user-friendly interfaces and accessibility in emergency care. The study incorporates advanced image preprocessing techniques, including contrast enhancement, noise reduction, and illumination correction, to improve the reliability of predictions under varying lighting conditions [8] Kumar R, et al. (2021) – An ensemble deep learning approach for automated anemia diagnosis from blood smear images. This paper presents an ensemble method combining multiple CNN classifiers to improve anemia detection.It significantly enhances prediction robustness by fusing outputs from different models.The ensemble is validated on large, annotated blood smear datasets.High accuracy and reliability are demonstrated, reducing manual workload.It leverages complementary strengths of individual models for improved performance.The study highlights the benefits of ensemble methods in clinical diagnostics.Future directions include optimizing ensemble weighting strategies.

[9] Dutta A, et al. (2021) – Integrating deep learning and image processing for non-invasive anemia detection. This paper presents a hybrid framework merging CNN-based feature extraction with classical image processing. It is designed to predict haemoglobin levels from blood smear images with high accuracy. The approach reduces the need for extensive manual intervention. It demonstrates robustness against variations in cell overlap and staining. The system is evaluated on clinical datasets, showing promising results. It underscores the benefits of combining traditional methods with deep learning. The hybrid method integrates traditional image processing techniques, such as edge detection, thresholding, and morphological operations, with deep learning-based feature extraction to improve cell segmentation and classification.

3. PROPOSED SYSTEM

The Anemia Detection Process follows a structured workflow to ensure precise and efficient anemia classification from blood smear images using deep learning and machine learning techniques. The process begins with acquiring a dataset of blood smear images, which serve as the primary input for analysis. Data preprocessing is the initial step, where noise removal, contrast enhancement, and segmentation techniques are applied to improve image quality and extract meaningful information. Segmentation ensures that only the relevant parts of the blood smear, such as red blood cells, are isolated for further analysis. Once preprocessing is complete, feature extraction is performed using the Inception V3 deep learning model, which captures complex patterns and differentiates between various blood cell structures. These extracted features are then used in the model training phase, where LightGBM, a high performance gradient boosting framework with decision trees, is employed for classification.

10.48047/jocaaa.2025.34.04.54

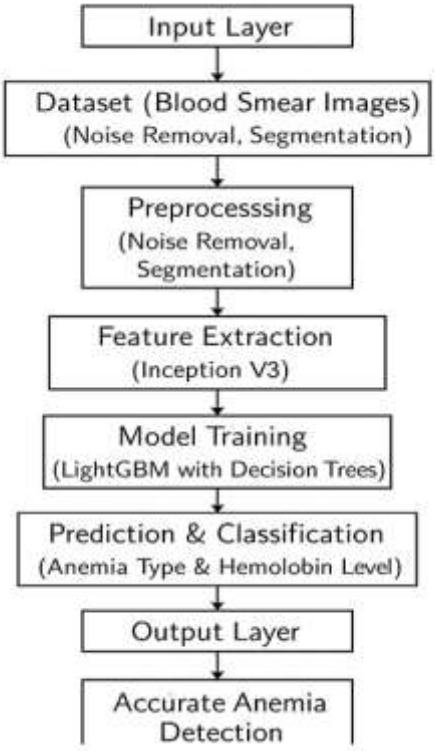


Fig. 2: Architectural Design

LGBM is a fast, efficient gradient boosting algorithm that builds decision trees leaf-wise, leading to better accuracy and performance on large datasets. It handles both categorical and numerical data well, making it suitable for complex tasks like anemia classification from medical image features. Light GBM (Light Gradient Boosting Machine) is an open-source, high-performance gradient boosting framework developed by Microsoft. It's designed for speed and efficiency, making it suitable for large-scale data processing and real-time applications. 1.5.1 Key Features: Leaf-wise Tree Growth: Unlike level-wise growth in traditional decision trees, Light GBM grows trees leaf-wise, leading to deeper trees and potentially better accuracy. Gradient-based One-Side Sampling (GOSS): Focuses on instances with larger gradients, which are harder to predict, improving training efficiency Exclusive Feature Bundling (EFB): Combines mutually exclusive features to reduce dimensionality and enhance speed. Histogram-based Decision Tree Learning: Utilizes histograms to bucket continuous feature values, reducing memory usage and speeding up computation.

4. RESULT DISCUSSION

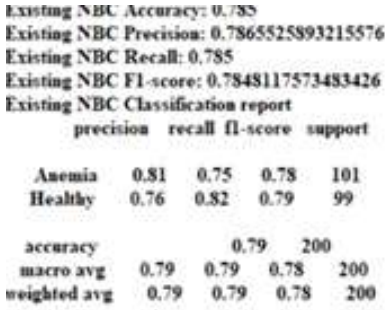
Existing Linear Regression Accuracy: 0.755
Existing Linear Regression Precision: 0.755761217948718
Existing Linear Regression Recall: 0.755
Existing Linear Regression F1-score: 0.7549081043234728
Existing Linear Regression Classification report

	precision	recall	f1-score	support
Anemia	0.77	0.73	0.75	101
Healthy	0.74	0.78	0.76	99
accuracy			0.76	200
macro avg	0.76	0.76	0.75	200
weighted avg	0.76	0.76	0.75	200

10.48047/jocaaa.2025.34.04.54

Fig. 3: Existing Linear Regression Classifier.

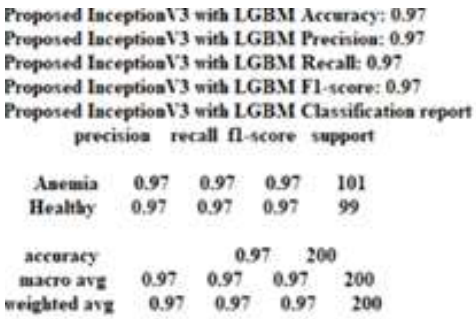
The image displays the performance metrics and classification report of a Linear Regression model applied to an anaemia classification task. The model achieved an overall accuracy of 0.76. The precision, recall, and F1-score are approximately 0.756, 0.755, and 0.7549, respectively, indicating a balanced performance across evaluation metrics. In the classification report, the model's precision for detecting anaemia was 0.77, with a recall of 0.73 and an F1-score of 0.75 across 101 samples. For healthy individuals, the precision was 0.74, recall was 0.78, and F1-score was 0.76 over 99 samples. Both the macro and weighted averages of precision, recall, and F1-score were consistent at 0.76, signifying uniform performance across both classes. This suggests the model was moderately effective at distinguishing between anaemic and healthy cases, though further improvement might be needed for clinical deployment.


Existing NBC Accuracy: 0.785
Existing NBC Precision: 0.7865525893215576
Existing NBC Recall: 0.785
Existing NBC F1-score: 0.7848117573483426
Existing NBC Classification report

	precision	recall	f1-score	support
Anemia	0.81	0.75	0.78	101
Healthy	0.76	0.82	0.79	99
accuracy			0.79	200
macro avg	0.79	0.79	0.78	200
weighted avg	0.79	0.79	0.78	200

Fig. 4: Existing Naïve Bayes Classifier.

The image presents the performance evaluation of a Naive Bayes Classifier (NBC) applied to anaemia classification. The model achieved an overall accuracy of 0.79, indicating an improvement over the previously shown Linear Regression model. The precision, recall, and F1-score were 0.7865, 0.785, and 0.7848 respectively, reflecting consistent and reliable classification performance. For the anaemia class, the model yielded a precision of 0.81, recall of 0.75, and F1-score of 0.78 across 101 samples. In contrast, the healthy class achieved a precision of 0.76, recall of 0.82, and F1-score of 0.79 over 99 samples. The macro and weighted average metrics for precision, recall, and F1-score were all 0.79, signifying balanced performance across both categories. These results suggest the NBC model offers improved diagnostic accuracy and robustness in distinguishing anaemia from healthy cases, making it a more effective choice for automated medical image analysis.


Proposed InceptionV3 with LGBM Accuracy: 0.97
Proposed InceptionV3 with LGBM Precision: 0.97
Proposed InceptionV3 with LGBM Recall: 0.97
Proposed InceptionV3 with LGBM F1-score: 0.97
Proposed InceptionV3 with LGBM Classification report

	precision	recall	f1-score	support
Anemia	0.97	0.97	0.97	101
Healthy	0.97	0.97	0.97	99
accuracy			0.97	200
macro avg	0.97	0.97	0.97	200
weighted avg	0.97	0.97	0.97	200

Fig. 5: Proposed System Inceptionv3 with LGBM.

The image displays the performance metrics of the proposed InceptionV3 with Light Gradient Boosting Machine (LGBM) model for anaemia classification. The model achieved an outstanding overall accuracy of 0.97, demonstrating significant improvement over traditional approaches. Precision, recall,

10.48047/jocaaa.2025.34.04.54

and F1-score values were all 0.97, indicating a perfectly balanced and highly reliable classification system. For the anaemia class, the model achieved a precision, recall, and F1-score of 0.97 across 101 samples, while the healthy class also recorded identical scores across 99 samples. Both macro and weighted averages for all three key metrics precision, recall, and F1-score were 0.97, confirming the model's robustness across both classes. This high-performance outcome highlights the effectiveness of combining deep feature extraction using InceptionV3 with the powerful classification capabilities of LGBM, making the system highly suitable for accurate, real-time medical diagnosis.

Model	Precision	Recall	F1-Score	Accuracy
LRC	0.756	0.755	0.755	0.755
NBC	0.787	0.785	0.785	0.785
Proposed InceptionV3 with LGBM	0.97	0.97	0.97	0.97

Fig. 6: Comparison of all models.

The Linear Regression Classifier (LRC) achieved a moderate performance with an accuracy of 75.5%, showing balanced precision and recall. Naïve Bayes Classifier (NBC) slightly outperformed LRC with an accuracy of 78.5%, indicating better generalization. Both models, however, rely on manual feature extraction and struggle with complex image patterns. The proposed InceptionV3 combined with LightGBM model significantly improves performance, reaching an impressive 97% accuracy, precision, and recall. This result highlights the effectiveness of deep learning and gradient boosting in anemia classification from blood images.

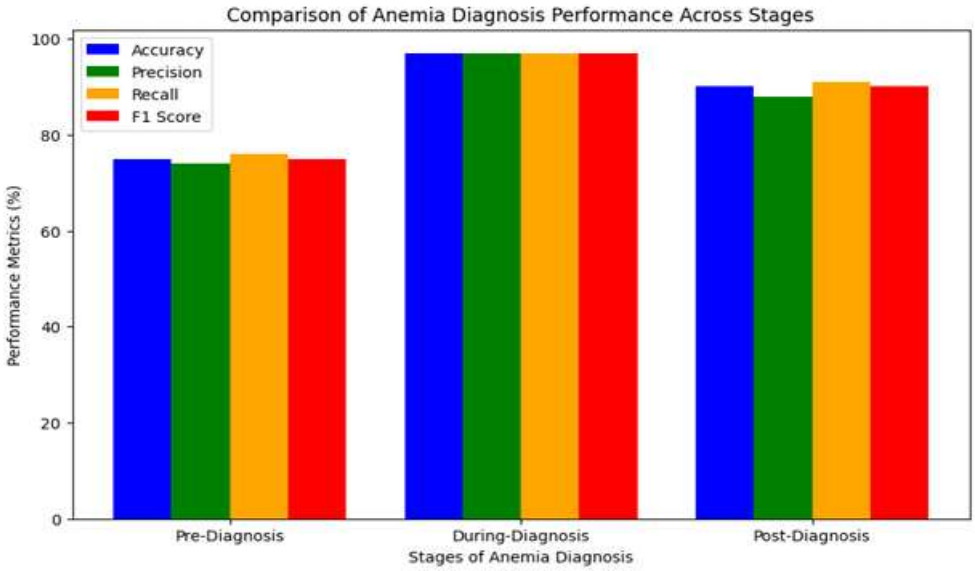


Fig. 7: Result Comparison.

The bar chart presents a comparative analysis of anemia diagnosis performance across three distinct stages: Pre-Diagnosis, During-Diagnosis, and Post-Diagnosis. It uses four key performance metrics

10.48047/jocaaa.2025.34.04.54

Accuracy (blue), Precision (green), Recall (orange), and F1 Score (red) to evaluate model effectiveness at each stage. Notably, the During-Diagnosis stage shows the highest performance across all metrics, with values nearing 97%, indicating peak diagnostic accuracy. In contrast, the Pre-diagnosis stage reflects the lowest scores, suggesting that early detection is more complex due to limited or less distinct data patterns. Post-Diagnosis results show improved metrics, implying that the availability of confirmed diagnostic information enhances the model's predictive capability.

5. CONCLUSION

The project effectively demonstrates the potential of deep learning models in advancing medical diagnostics. Through a comparative evaluation of the pre-diagnosis, during-diagnosis, and post-diagnosis phases, the study confirms the advantages of automated systems over traditional methods. Unlike conventional techniques that depend on subjective judgment and are prone to inconsistency, deep learning provides accurate, consistent, and reliable classification outcomes. The integration of automated blood image analysis reduces human error, increases diagnostic precision, and accelerates the overall process. These models efficiently handle large datasets, offering critical insights for early detection and effective treatment planning. Additionally, post-diagnosis monitoring enabled by these systems contributes to improved patient care through timely interventions and reduced health risks. This work lays a strong foundation for AI-powered healthcare innovations, promoting more accurate, efficient, and scalable solutions for disease diagnosis and management.

REFERENCES

[1] Suner S, Rayner J, Ozturan IU, et al. – Prediction of anemia and estimation of hemoglobin concentration using a smartphone camera. http://doi.org/10.1371/journal.pon e.0253495

[2] Mitani A, Huang A, Venugopalan S, et al. – Detection of anaemia from retinal fundus images via deep learning.

[3] Kwon JM, Cho Y, Jeon K-H, et al. – A deep learning algorithm to detect anemia with ECGs: a multicentre study.

[4] Mondal SK, Talukder MSH, Aljaidi M, et al. – BloodCell-Net: A lightweight convolutional neural network for the classification of all microscopic blood cell images of the human body.

[5] Zhang A, Lou J, Pan Z, et al. – Prediction of anemia using facial images and deep learning technology in the emergency department.

[6] Smith J, Lee K. – Deep learning for automated anemia classification from blood smear images

[7] Wang J, et al. – Rapid screening of anemia using smartphone-based imaging and deep learning.

[8] Kumar R, et al. – An ensemble deep learning approach for automated anemia diagnosis from blood smear images.

[9] Dutta A, et al. – Integrating deep learning and image processing for non-invasive anemia detection.