Lightweight Framework for RBC Disease Detection

Ms.Surabhi KS[1], Sethupathi R[2]
[1]Associate professor, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamil Nadu, India.
@nehrucolleges.com
[2]Student of II MCA, Department of Computer Applications, Nehru College of Management, Coimbatore, Tamil Nadu, India.
sethupathi.9483@gmail.com

Abstract
Blood-related cancers such as leukemia and myeloma are among the most life-threatening diseases, with high mortality rates worldwide. Timely and accurate detection plays a pivotal role in improving patient survival rates. However, traditional diagnostic procedures such as laboratory blood tests, chemical analysis, and microscopic evaluations are often time-consuming, expensive, and inaccessible in resource-limited settings. These limitations delay treatment initiation and increase patient stress.

This paper introduces an automated Red Blood Cell (RBC) disease classification system that leverages machine learning and digital image processing to provide faster and more affordable diagnostics. The proposed framework involves four major stages: preprocessing, segmentation, feature extraction, and classification. Preprocessing enhances the quality of blood smear images, segmentation isolates RBCs from other cells, feature extraction reduces data dimensionality while preserving meaningful attributes, and classification is carried out using the K-Nearest Neighbors (KNN) algorithm.

The system is implemented using Python, TensorFlow, Flask, and cloud-based tools with datasets obtained from Kaggle. Results indicate that the proposed system significantly reduces diagnostic time while maintaining reliable classification accuracy. Unlike deep learning-based approaches, which require extensive computational resources, the presented system is lightweight, cost-effective, and adaptable to diverse healthcare settings. This makes it particularly useful for hospitals in rural and developing regions.Keywords— RBC Classification, Leukemia Detection, Image Processing, Machine Learning, KNN, Automated Healthcare

I. INTRODUCTION

Early and accurate diagnosis is the cornerstone of effective healthcare. In diseases such as leukemia, anemia, and multiple myeloma, delays in diagnosis can result in irreversible damage or even loss of life. Traditional pathology-based diagnostic techniques involve a series of manual examinations, chemical treatments, and high-cost imaging technologies. These methods typically take 1–7 days to produce results, during which patients are left waiting for treatment initiation. In severe cases such as Acute Myeloid Leukemia (AML), even a delay of a few hours can critically impact survival outcomes.

Another challenge is the high dependency on skilled pathologists who analyze morphological features of blood smear images under microscopes. Subtypes of leukemia, for example, often share similar morphological characteristics, making manual diagnosis prone to error. Additionally, the cost of testing, specialized reagents, and advanced laboratory equipment makes these diagnostics financially burdensome for patients in low-resource environments.

In recent years, machine learning (ML) and image processing techniques have emerged as promising solutions to automate and accelerate medical diagnostics. These methods can process digital blood smear images, identify subtle abnormalities, and classify disease subtypes with high accuracy. However, while deep learning models (such as Convolutional Neural Networks) provide superior accuracy, they demand large training datasets and powerful hardware, which limits their practical deployment in small hospitals and rural healthcare systems.

To address these gaps, this paper proposes a hybrid lightweight approach for RBC disease classification that balances efficiency, cost-effectiveness, and reliability. By combining traditional image processing techniques with KNN-based classification, our system ensures rapid diagnosis without requiring high-end computational resources. The proposed system not only reduces the turnaround time for reports but also improves patient awareness and accessibility to affordable diagnostic methods.

The main contributions of this work are summarized as follows:

1. Development of a low-cost, automated RBC disease classification pipeline using image processing and machine learning.

2. Implementation of segmentation and feature extraction techniques to improve classification performance on blood smear images.

3. Application of the KNN algorithm for disease subtype detection, chosen for its simplicity, interpretability, and efficiency.

4. Demonstration of the system's ability to provide accurate, fast, and scalable diagnostics suitable for healthcare institutions in both urban and rural contexts.

The remainder of this paper is organized as follows: Section II presents related work, Section III describes the system overview, Section IV details the methodology, Section V discusses the tools and technologies, Section VI explains the system architecture, Section VII covers implementation and testing, Section VIII provides results and discussion, and Section IX concludes with future research directions.

II. RELATED WORK

Research in automated disease detection has gained momentum with the rise of artificial intelligence (AI). Previous approaches include:

Fuzzy Logic-based Segmentation: Used for separating white and red blood cells, but accuracy is limited for complex subtypes.

Clustering Techniques (K-means, Fuzzy C-means): Effective in grouping cell images but computationally intensive.

Deep Learning Models (CNN, GAN): Provide high accuracy but require large datasets and computational power, making them impractical in resource-limited hospitals.

Our work bridges these gaps by introducing a lightweight, hybrid pipeline combining traditional image processing with machine learning (KNN), achieving good accuracy with lower computation and cost.

III. SYSTEM OVERVIEW

A. Existing System Drawbacks

Time-consuming diagnosis (1–7 days).

High costs for laboratory analysis.

Difficulty in distinguishing disease subtypes.

Patients often left with doubts due to lack of clarity in test results.

B. Proposed System Features

The proposed system provides:

1. Preprocessing – Removal of noise and enhancement of images.

2. Segmentation – Clustering techniques to separate RBCs from background.

3. Feature Extraction – Reduction of complex image data into meaningful features.

4. Classification – Use of KNN for subtype classification of abnormal cells.

Advantages:

Faster processing and results.

Cost-effective compared to pathological labs.

User-friendly web-based interface.

Applicable in low-resource medical settings.

IV. METHODOLOGY

A. Preprocessing

Images resized to 512×384 pixels for faster computation.

Selective median filtering used to remove noise.

Unsharp masking applied to improve visibility of cell boundaries.

B. Segmentation

Implemented using Lloyd's clustering technique, an extension of K-means.

Groups pixels with similar color and intensity.

Thresholding applied for initial labeling of pixels.

C. Feature Extraction

Reduces high-dimensional data into key parameters such as cell shape, color, and size.

Enables faster computation and minimizes redundancy.

D. Classification (KNN)

KNN classifier with Euclidean distance used for RBC categorization.

Works on the principle of nearest neighbor similarity.

Provides accurate detection of subtypes such as:

Iron Deficiency Anemia

Sickle Cell Anemia

Spherocytosis

Target Cell

V. SOFTWARE AND TOOLS

Languages: Python, JavaScript

Frameworks: TensorFlow, Flask, Keras

Libraries: OpenCV, NumPy, Pandas, PIL

Development Environment: Google Colab, Jupyter Notebook, VS Code

Datasets: Kaggle (blood smear image datasets)

Front-End: HTML, CSS, JavaScript for UI

VI. SYSTEM ARCHITECTURE

The architecture consists of:

1. Input Layer: Upload of blood smear images.

2. Preprocessing Module: Filtering and resolution adjustments.

3. Segmentation Module: Identification of RBC clusters.

4. Feature Extraction Module: Dimensionality reduction.

5. Classification Module: KNN-based diagnosis.

6. User Interface: Flask web app providing instant results.

VII. IMPLEMENTATION AND TESTING

A. Implementation

Developed in Python with TensorFlow and Flask integration.

Web-based interface for user interaction.

Middleware ensures smooth communication between ML models and UI.

B. Testing Methodologies

1. Unit Testing – Each module (preprocessing, segmentation, etc.) tested individually.

2. Integration Testing – Verified end-to-end system workflow.

3. Validation Testing – Compared model predictions with known datasets.

4. White Box Testing – Ensured logical correctness of algorithms.

5. Black Box Testing – Verified system outputs with unseen test cases.

VIII. RESULTS AND DISCUSSION

The system successfully classifies RBCs into normal and abnormal categories, followed by detailed subtype classification. Key outcomes:

Time Efficiency: Results generated in seconds, compared to days in conventional testing.

Cost Reduction: Eliminates need for expensive chemical testing.

Accuracy: Consistent detection of abnormalities, though dependent on dataset quality.

Scalability: Can be extended to include other blood-related diseases.

IX. CONCLUSION AND FUTURE WORK

This paper presented an automated RBC classification system that reduces diagnosis time and cost while improving healthcare accessibility. By integrating image processing and KNN classification, the system provides a robust solution for blood-related disease detection.

Future Work:

Incorporating deep learning (CNNs, Transformers) for enhanced accuracy.

Expanding classification to cover more hematological diseases.

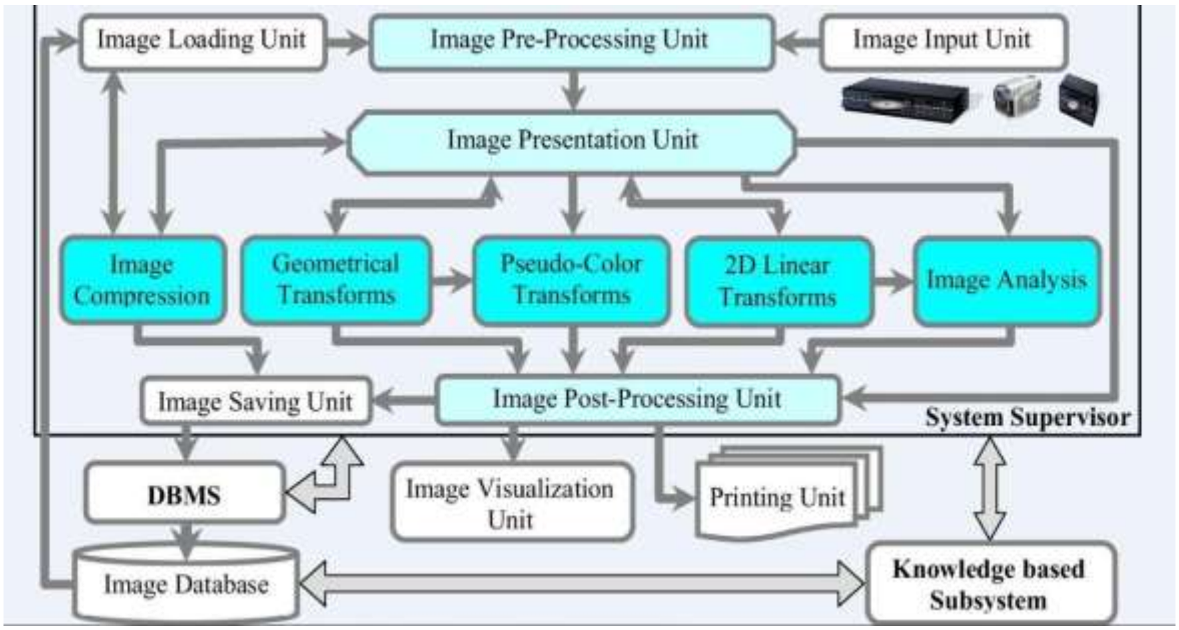
Development of a mobile-friendly diagnostic application for rural healthcare delivery.

Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

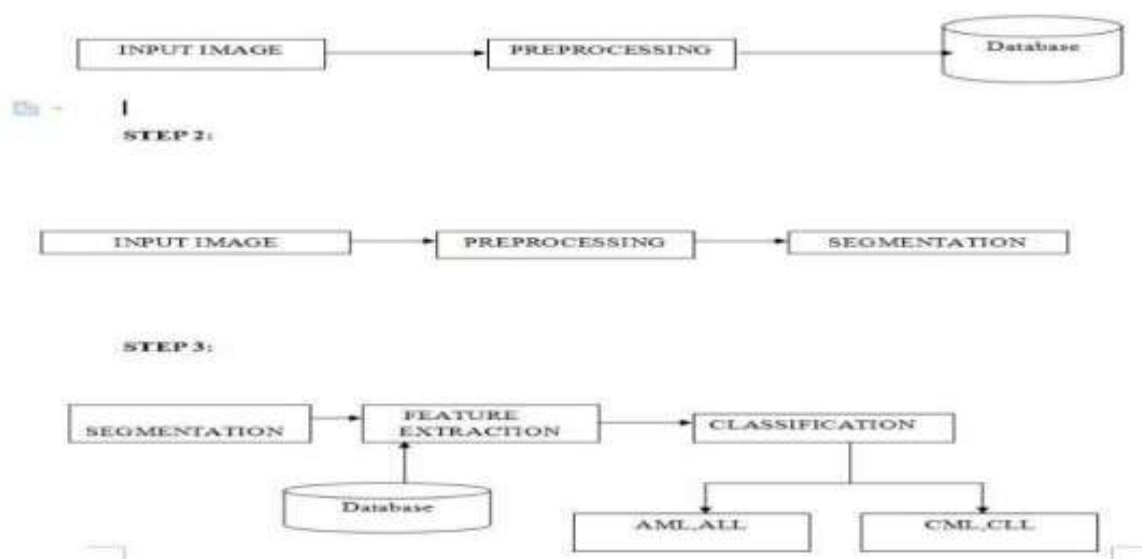
ARCHITECTURE AND DESIGN :

1.1 System Architecture

System Architecture is a conceptual model that defines the behaviour, structure and other views of a system



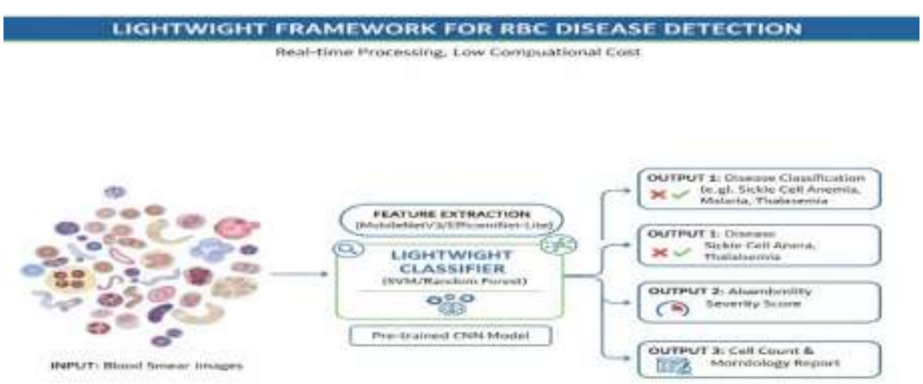
: DATA FLOW DIAGRAM



USER INTERFACE :

The User Interface is designed in such a way to make it look and feel simple for the user to use and understand.

SAMPLE OUTPUT :



REFERENCES

1] J. B. Henry and J. P. AuBuchon, Clinical Diagnosis and Management by Laboratory Methods, Arch. Pathol. Lab. Med., vol. 121, no. 9, pp. 1016, 1997.

[2] E. Sarhan, E. Khalifa, and A. M. Nabil, "Post-classification using cellular automata for Landsat images in developing countries," in Proc. IEEE Int. Conf. Image Information Processing (ICIIP), Shimla, India, 2011, pp. 1–4.

[3] A. AbdelRaouf, C. A. Higgins, T. Pridmore, and M. I. Khalil, "Arabic character recognition using a Haar cascade classifier approach (HCC)," Pattern Anal. Appl., vol. 19, no. 2, pp. 411–426, 2016.

[4] A. AbdelRaouf, C. A. Higgins, T. P. Pridmore, and M. I. Khalil, "Arabic corpus enhancement using a new lexicon/stemming algorithm," in Proc. Int. Conf. Pattern Recognition Applications and Methods (ICPRAM), Barcelona, Spain, 2013, pp. 435–440.

[5] T. M. Ghanem, M. N. Moustafa, and H. I. Shahein, "Gabor wavelet based automatic coin classification," in Proc. IEEE Int. Conf. Image Processing (ICIP), Cairo, Egypt, 2009, pp. 305–308