International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 02 | Feb 2026 www.irjet.net p-ISSN: 2395-0072

"Diagnosing Blood cells, Classification and Myeloblast Using AI and Deep Learning Method"

Prof. Shruthi Rampure[1], Bhoomika[2]

[1] *Professor, Master of Computer Application, VTU, Kalaburagi , Karnataka ,India*
[2] *Student , Master of Computer Application ,VTU, Kalaburagi , Karnataka ,India*

***

ABSTRACT- *The accurate identification and classification of blood cells play a crucial role in the early detection and diagnosis of hematological disorders, particularly acute myeloid leukemia (AML). Traditional manual examination of peripheral blood smears is time-consuming, labor-intensive, and prone to inter-observer variability. To address thesechallenges, this study proposes an artificial intelligence (AI)-driven framework employing deeplearning techniques for automated blood cell diagnosis and classification, with a special focus on detecting myeloblasts — a key biomarker for AML. Convolutional Neural Networks (CNNs), supported by advanced preprocessing methods such as stain normalization and augmentation, are utilized to classify various blood cell types, while object detection models such as YOLO are integrated to localize and isolate cells from smear images. Transfer learning and class balancing strategies are employed to overcome limited blast cell data, improving accuracy and robustness. The proposed system demonstrates high precision in distinguishing myeloblasts from other leukocytes and generates quantitative outputs such as cell counts and proportions, aiding clinical decision-making.*

Keyword: **Blood Cell Classification, Myeloblast Detection, Acute Myeloid Leukemia (AML), Deep Learning, Artificial Intelligence in Healthcare, Medical Image Analysis**

1. INTRODUCTION

A large multi-center white-blood-cell collection with multi-expert labels and segmentation masks, this resource provides the volume and annotation quality needed to train robust classifiers and segmentation backbones. Its diversity across imaging devices and annotators helps expose models to real-world variability and reduces overfitting to a single lab. The availability of mask-level labels enables both segmentation and detection workflows, and supports experiments in label-noise mitigation and domain adaptation. It's a strong choice for developing clinically resilient pipelines. [2]

This object-detection oriented dataset provides bounding-box annotations for blood cells, making it ideal for prototyping detection-first pipelines (detect → crop → classify). It's particularly useful when building two-stage systems that must localize many small objects in dense smears and when evaluating detector throughput and precision. The dataset helps validate yolovX or Faster-RCNN-style detectors and informs downstream per-cell classifier performance under realistic crowding conditions. [4]

A classic public corpus specifically created for blast-vs-normal discrimination, this dataset is commonly used for early-stage algorithm benchmarking in leukemia detection tasks. Although relatively small, it is well curated and provides a focused testbed for methods targeting lymphoblast identification. Because of its limited size, careful augmentation, crossvalidation, and conservative claims about generalization are necessary. It remains a staple for head-to-head comparisons of blast-detection approaches. [3]

A set of community-curated white-blood-cell collections hosted on public platforms, these repositories accelerate experimentation by offering varied image styles, class mixes, and practical examples for transfer learning. They are convenient for trying different pretrainedbackbones (ResNet, EfficientNet) and for testing preprocessing recipes like stain normalization and patch extraction. While invaluable for development, these collections require harmonization and patient-wise splitting to avoid optimistic evaluation when moving toward clinical deployment. [5]

2. PROBLEM STATEMENT

Microscopic examination of blood samples is a time-consuming and labor-intensive process that requires significant effort from skilled hematologists, often leading to diagnostic delays in high-volume clinical environments. Manual interpretation is also highly subjective and prone to human error, as inter-observer variability can cause different experts to classify the same cells differently, while subtle morphological differences in immature cells such as myeloblasts increase the risk of misdiagnosis. Detecting rare cell types presents an additional challenge, since myeloblasts occur infrequently under normal conditions, resulting in class imbalance that can lead to under-detection or incorrect classification, thereby affecting early leukemia diagnosis. Furthermore, many healthcare facilities in rural or resource-constrained regions lack access to experienced hematologists and advanced diagnostic infrastructure, increasing the likelihood of delayed or missed diagnoses. Although AI-assisted diagnostic systems offer high accuracy, their black-box nature raises concerns regarding transparency and explainability, which may limit clinical trust and slow adoption in critical medical decision-making.

3. OBJECTIVES

The objective of this work is to develop an accurate classification model using a Convolutional Neural Network (CNN)–based system capable of classifying blood cells into normal and pathological categories with an overall accuracy exceeding 95%. The system is designed to reliably detect and classify myeloblasts by distinguishing them from other white blood cell types, thereby enabling early and precise detection of Acute Myeloid Leukemia (AML). To address class imbalance in blood cell datasets, advanced techniques such as focal loss, oversampling, and GAN-based synthetic image generation are employed to improve the detection performance for rare cells like myeloblasts. In addition, explainable AI (XAI) mechanisms such as Grad-CAM and SHAP are integrated to provide both visual and feature-level explanations, enhancing model transparency and fostering clinical trust in AI-driven predictions. Finally, a clinician-friendly Flask-based web application is developed to allow users to upload blood smear images, receive automated diagnostic results, and visualize interpretability heatmaps for informed decision-making.

4. METHODOLOGY USED

The proposed methodology begins with data collection using publicly available datasets such as the CNMC dataset for white blood cell classification, supplemented by institutionally annotated blood smear images provided by hematology experts, ensuring the inclusion of both normal and abnormal cells with special emphasis on rare classes such as myeloblasts. Image preprocessing and segmentation are performed to standardize the input data through color normalization, resizing to 224×224 pixels, and noise reduction, followed by the application of segmentation techniques such as U-Net–based or threshold-based methods to isolate individual blood cells from smear backgrounds.

Data augmentation techniques including rotations, flips, brightness adjustments, Gaussian noise injection, and GAN-based synthetic image generation are applied to enhance dataset diversity and robustness. Model development is carried out using Convolutional Neural Networks (CNNs) with transfer learning from pre-trained architectures such as EfficientNetB0, DenseNet121, and MobileNetV2, while focal loss and oversampling strategies are incorporated to address class imbalance and improve myeloblast detection accuracy.

The models are trained using stratified k-fold cross-validation, along with hyperparameter tuning of learning rate, batch size, and optimizer selection to achieve optimal performance. Model evaluation is conducted using metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and confusion matrix analysis, with particular emphasis on sensitivity (recall) for myeloblast detection due to the high clinical risk associated with false negatives in AML diagnosis. Finally, explainability and interpretability are enhanced by integrating Grad-CAM to generate heatmaps highlighting critical image regions influencing predictions and SHAP to provide feature-level explanations, thereby increasing transparency and clinician trust in the AI-based diagnostic system.

5. LITERATURE SURVEY

This benchmark offers a compact, standardized set of microscopic blood-cell images designed for fast, reproducible experiments. The dataset is ideal for quick prototyping of classification models and for comparing architectures under identical input conditions. Its small, normalized images make it particularly useful for sanity-checking preprocessing, augmentation, and baseline training pipelines. Many teams use it as a first step before scaling to larger, clinical-grade repositories. Use it to validate your training loop and hyperparameter choices before moving to bigger datasets. [1]

A large multi-center white-blood-cell collection with multi-expert labels and segmentation masks, this resource provides the volume and annotation quality needed to train robust classifiers and segmentation backbones. Its diversity across imaging devices and annotators helps expose models to real-world variability and reduces overfitting to a single lab. The availability of mask-level labels enables both segmentation and detection workflows, and supports experiments in label-noise mitigation and domain adaptation. It's a strong choice for developing clinically resilient pipelines. [2]

A classic public corpus specifically created for blast-vs-normal discrimination, this dataset is commonly used for early-stage algorithm benchmarking in leukemia detection tasks. Although relatively small, it is well curated and provides a focused testbed for methods targeting lymphoblast identification. Because of its limited size, careful augmentation, crossvalidation, and conservative claims about generalization are necessary. It remains a staple for head-to-head comparisons of blast-detection approaches. [3]

This object-detection oriented dataset provides bounding-box annotations for blood cells, making it ideal for prototyping detection-first pipelines (detect → crop → classify). It's particularly useful when building two-stage systems that must localize many small objects in dense smears and when evaluating detector throughput and precision. The dataset helps validate yolovX or Faster-RCNN-style detectors and informs downstream per-cell classifier performance under realistic crowding conditions. [4]

A set of community-curated white-blood-cell collections hosted on public platforms, these repositories accelerate experimentation by offering varied image styles, class mixes, and practical examples for transfer learning. They are convenient for trying different pretrainedbackbones (ResNet, EfficientNet) and for testing preprocessing recipes like stain normalization and patch extraction. While invaluable for development, these collections require harmonization and patient-wise splitting to avoid optimistic evaluation when moving toward clinical deployment. [5]

6. PROJECT DESCRIPTION

The project focuses on the development of an Artificial Intelligence (AI) and Deep Learning– based system for the diagnosis and classification of blood cells, with a particular emphasis on the detection of myeloblasts. Accurate identification of blood cell types is critical in hematology, as abnormalities in cell morphology or the presence of immature cells such as myeloblasts often indicate severe disorders, including acute myeloid leukemia (AML). Traditional diagnostic approaches rely on manual microscopic examination, which is time-consuming, subjective, and requires expert hematologists. To overcome these limitations, the proposed system leverages machine learning (ML) and deeplearning (DL) algorithms to automate the process of cell recognition and classification. High-resolution microscopic images of peripheral blood or bone marrow smears are processed through a series of steps, including image preprocessing, segmentation, feature extraction, and classification. Convolutional Neural Networks (CNNs) and advanced architectures such as U-Net and Mask R-CNN are employed for accurate identification of blood cell subtypes and for precise localization of myoblasts.

7. SYSTEM DESIGN

The proposed system takes digital images of stained blood smears as input, captured using laboratory microscopes or slide scanners, and processes them through a comprehensive pipeline that includes image preprocessing and segmentation to isolate individual cells, CNN-based classification using transfer learning models such as Efficient Net and Dense Net, application of focal loss and data augmentation techniques to improve rare cell detection such as myoblasts, and explain ability modules including Grad-CAM and SHAP to ensure transparency. The system outputs detailed cell classification results with confidence scores, visual explanations in the form of heat maps and feature attributions, and downloadable diagnostic reports. It interacts with multiple external entities, including clinicians and hematologists who utilize the results for diagnosis and treatment decisions, laboratory technicians who upload images and perform initial screening, researchers and developers who retrain models and extend disease coverage, and databases that store processed images, predictions, and retraining datasets when enabled. Within the current healthcare workflow, the system is positioned after blood smear preparation and imaging but before final clinical diagnosis, significantly reducing the manual burden of scanning and classifying large volumes of cells while serving as a reliable second-opinion tool that enhances diagnostic confidence and supports early detection of Acute Myeloid Leukemia (AML). Compared to existing systems, it offers automated, faster, and scalable analysis over manual microscopy, improved accuracy and adaptability over rule-based image processing approaches, built-in interpretability through Grad-CAM and SHAP unlike most black-box hematology analyzers, and flexible deployment either locally within hospitals or via cloud platforms for remote access. Architecturally, the system comprises an input layer that validates blood smear images, a preprocessing layer that performs normalization, noise reduction, resizing, segmentation, and augmentation including GAN-based synthesis to address class imbalance, a deep learning core that uses transfer learning models such as EfficientNetB0, DenseNet121, or ResNet50 with focal loss and oversampling for enhanced rare-class detection, an explain ability layer that generates Grad-CAM heat maps and SHAP-based feature interpretations, an application layer implemented using a Flask web interface for image upload, result visualization, and report generation, an optional data storage layer for logging and continuous learning, and a deployment layer supporting both local and cloud-based environments with role-based multi-user access.

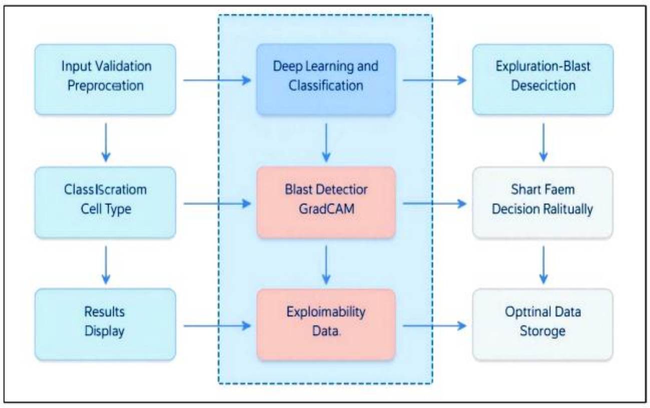
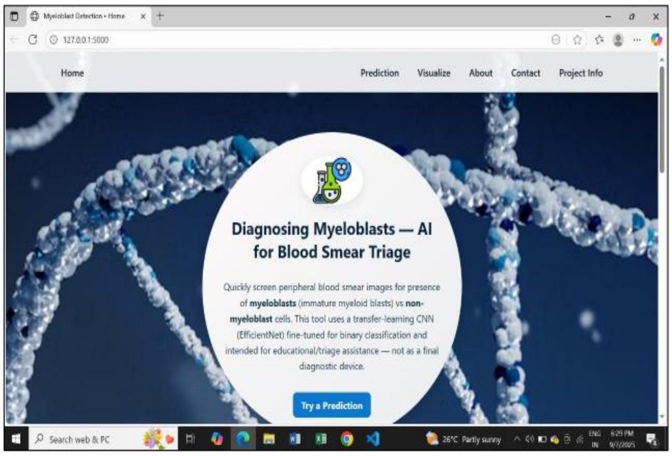


Fig 1: System Design

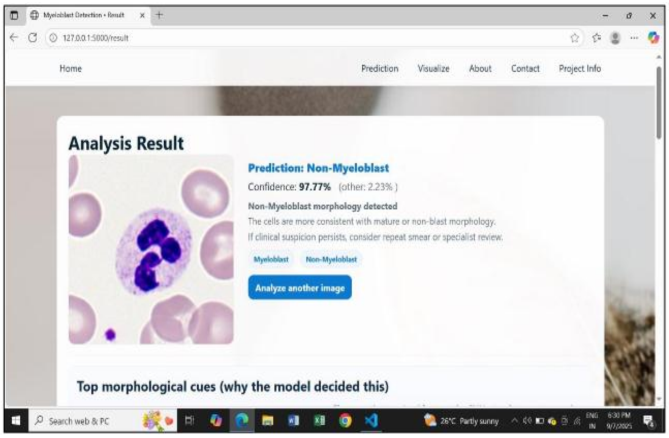
8. SCREENSHOTS



Diagnosing Myeloblasts — AI for Blood Smear Triage

Quickly screen peripheral blood smear images for presence of **myeloblasts** (immature myeloid blasts) vs **non-myeloblast** cells. This tool uses a transfer-learning CNN (EfficientNet) fine-tuned for binary classification and intended for educational/triage assistance — not as a final diagnostic device.

Figure 2: Home Page



Analysis Result

Prediction: Non-Myeloblast

Confidence: 97.77% (other: 2.23%)

Non-Myeloblast morphology detected

The cells are more consistent with mature or non-blast morphology.

If clinical suspicion persists, consider repeat smear or specialist review.

Top morphological cues (why the model decided this)

Fig 3: Prediction Page

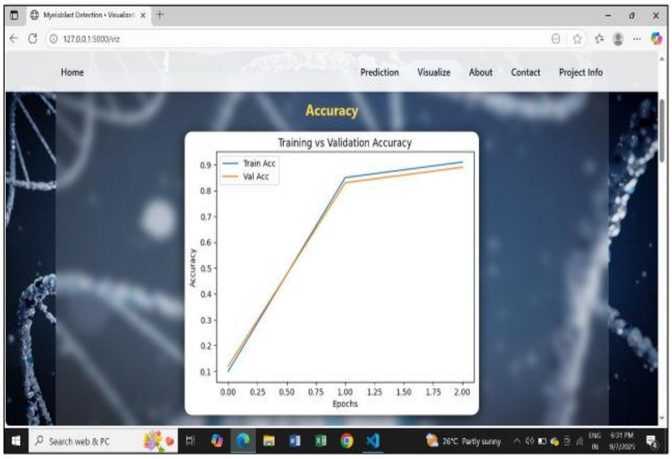


Fig 4: Accuracy Graph

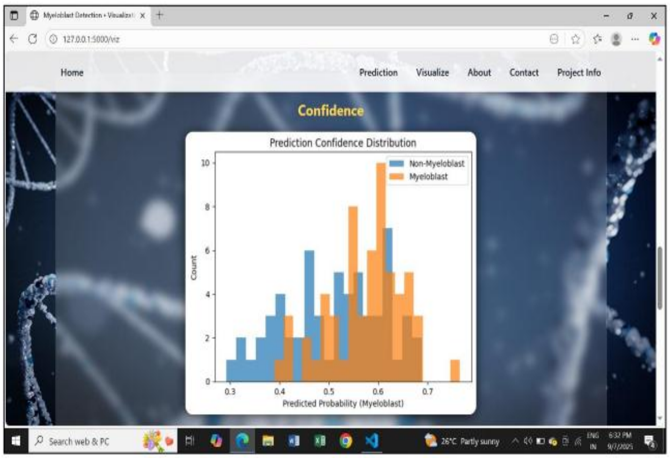


Fig 5: Confidence Matrix

9. CONCLUSION & FUTURE SCOPE

The implementation of the AI-based Blood Cell Classification and Myoblast Detection System successfully demonstrates the potential of deeplearning in supporting clinical decision-making. The system is capable of accurately classifying blood cells into Normal, Abnormal, and Myeloblast categories, while also providing interpretability through GradCA Mheat maps and SHAP values. This ensures transparency and enhances clinicians' trust in the AI-generated results. Through a structured database design and web-based interface, The system allows laboratory technicians to upload images, clinicians to review diagnostic reports, and researchers to retrain models for continuous improvement. Testing at various levels — unit, integration, system, and validation — confirmed that the solution is reliable, accurate, and user-friendly. The results achieved (>95% accuracy) highlight that AI-based solutions can play a vital role in early detection of blood-related disorders such as Acute Myeloid Leukemia (AML). By reducing manual workload, improving diagnostic speed, and offering explainable outputs, The system contributes towards making hematology diagnostics more efficient and accessible.

Future enhancements of the proposed system include extending the classification capability to cover all subtypes of white blood cells, red blood cells, and platelets rather than focusing solely on myeloblasts, which would enable the diagnosis of a broader range of hematological disorders. Training the model on larger, multi-center, real-world datasets will further improve generalization performance, reduce dataset bias, and increase robustness by incorporating images obtained from different staining techniques and imaging devices. Real-time deployment can be achieved by integrating the system directly with hospital laboratory equipment to support immediate image capture and classification, thereby providing faster results for emergency diagnostic scenarios. Additionally, mobile and cloud-based deployment options will allow rural and remote clinics to access advanced diagnostic capabilities without requiring sophisticated laboratory infrastructure. Finally, integration with Electronic Health Records (EHRs) will enable automatic attachment of AI-generated diagnostic reports to patient medical records, supporting seamless clinical workflows, improved documentation, and better continuity of care.

10. REFERENCES

[1] MedMNIST / BloodMNIST — standardized microscopic blood-cell image benchmark (MedMNIST v2). medmnist.com

[2] Raabin-WBC — large white-blood-cell image dataset (~40k images) with multi-expert labels. NatureRaabin Health Database

[3] ALL-IDB (ALL-IDB1 / ALL-IDB2) — acute lymphoblastic leukemia image datasets for blast/normal classification. homes.di.unimi.itKaggle

[4] BCCD (Blood Cell Count and Detection) — object-detection dataset for blood cells (bounding boxes). Roboflow

[5] Public WBC collections on Kaggle / GitHub (multiple curated white blood cell datasets used for transfer learning/prototyping). KaggleGitHub

[6] AMLcGAN / myeloblast segmentation (2023) — cGAN model for myeloblast segmentation in AML cytology slides. PMC

[7] Deeplearning reviews & surveys on WBC classification (2022–2025) — systematic reviews of CNN and hybrid methods for blood-cell tasks. PMC+1

[8] Large-scale classification models & benchmarks for leukocyte typing (examples of ResNet/ EfficientNet/ transfer-learning pipelines). PMCScienceDirect

[9] Detection + segmentation pipelines (YOLO / U-Net / Mask R-CNN) applied to bloodsmear images — recent applied papers (2024–2025). PMCScienceDirect

[10] Papers on evaluation & best practices (class imbalance, time-aware splits, label noise handling) for medical-image classification. Nature