Automated Diagnosis of Acute Lymphoblastic Leukemia

Computer Vision Course Project Report

Department of Artificial Intelligence

FAST National University of Computer and Emerging Sciences

Peshawar Campus

Submitted By

Tazmeen Afroz

Roll No: 22P-9252

Instructor:

Dr. M Tahir

2 December 2025

ABSTRACT

Acute Lymphoblastic Leukemia (ALL) is a malignant cancer characterized by the overproduction of immature lymphoblasts, accounting for approximately 25% of all pediatric cancers. Traditional diagnosis through manual examination of blood smears is time-consuming, subjective, and prone to human error. This report presents a comprehensive review of state-of-the-art deep learning approaches for automated ALL diagnosis. We examine three major benchmark datasets (C-NMC 2019, ALL Image Dataset, and ALL-IDB) and analyze recent advancements in the field, including lightweight architectures (YOLOv8/v11), explainable AI techniques (Grad-CAM, SHAP), and synthetic data generation using GANs. Recent studies demonstrate that optimized YOLO-based models achieve up to 98.6% precision with real-time inference capabilities (20–30 ms per image), making them suitable for clinical deployment.

1 Introduction

Acute Lymphoblastic Leukemia (ALL) is a malignant white blood cell cancer characterized by the overproduction of immature lymphoblasts. The disease presents a significant diagnostic challenge due to the visual similarity between malignant lymphoblasts and normal lymphocytes. Both cell types exhibit morphological similarities including irregular nucleus structure, high nucleus-to-cytoplasm ratios, and prominent nucleoli, making manual differentiation extremely difficult even for experienced pathologists.

1.1 The Clinical Challenge

ALL accounts for approximately 25% of all pediatric cancers, making it one of the most common childhood malignancies. Traditional diagnosis relies on manual examination of blood smears by pathologists—a process that is time-consuming, subjective, and prone to errors related to human fatigue and inter-observer variability.

1.2 Motivation for Automation

Automated AI-based diagnostic systems offer several critical advantages over traditional manual methods. First, they provide rapid screening capabilities, significantly reducing the time from sample collection to diagnosis. Second, these systems deliver standardized results that are not subject to human fatigue or subjective interpretation.

2 Methodology

2.1 Benchmark Datasets

This review examines three major benchmark datasets that have driven research in automated ALL diagnosis:

- **C-NMC 2019 (ISBI Challenge):** The largest dataset from The Cancer Imaging Archive (TCIA) containing 15,135 pre-segmented single-cell images.
- **ALL Image Dataset (Kaggle):** Contributed by Mehrad Aria et al. (2023), containing 3,256 images categorized into four classes: Benign, Early Pre-B, Pre-B, and Pro-B.
- **ALL-IDB1 & ALL-IDB2:** Includes 108 whole-slide high-resolution images and 260 expert-cropped single-cell images.

2.2 Deep Learning Approaches

Recent research has converged on several key methodological trends:

Transfer Learning with YOLO: YOLOv8 and YOLOv11 models have been adapted for blast cell detection, offering real-time inference capabilities.

GAN-Based Synthetic Data Generation: GANs are employed to address class imbalance by generating high-fidelity synthetic healthy or blast cells.

3 Results and Discussion

3.1 Dataset Characteristics

The transition from single-cell to multi-cell imaging represents a crucial advancement in realistic clinical application.

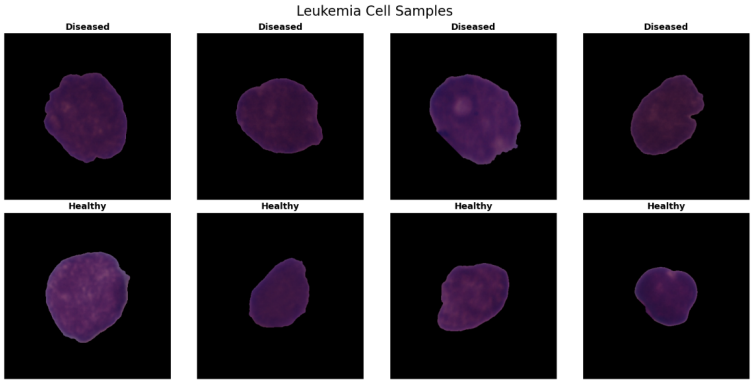

Figure 1: Pre-segmented single-cell images from C-NMC 2019 dataset.

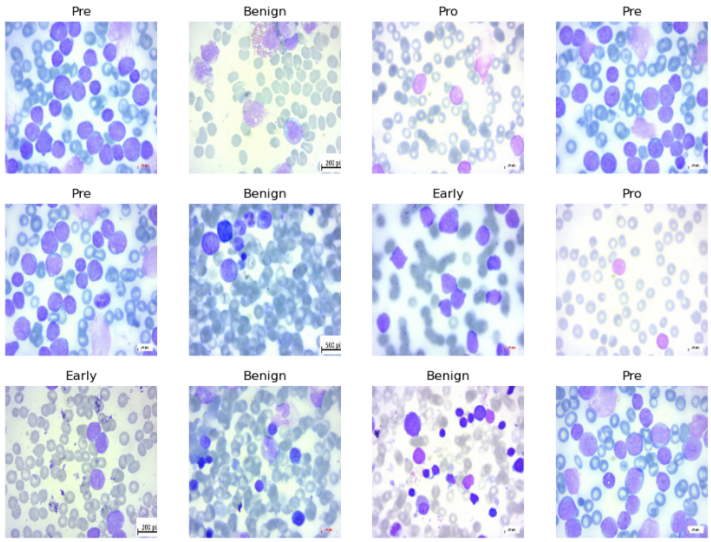

Figure 2: Multi-cell crowded peripheral blood smear images from ALL Image Dataset.

3.2 State-of-the-Art Performance

Table 1: Recent Published Articles on Automated ALL Diagnosis

Authors	Title & Approach	Dataset	Acc/Perf
A. Awad et al.	**YOLOv8 & YOLOv11 Models:** Real-time detection with lightweight architecture (IEEE ICSPIS 2024)	Blood smears	20ms/img
G. N. Prashanth	**Custom GAN + ResNet:** Addressing class imbalance via synthetic generation (IEEE ICD-CECE 2024)	C-NMC 2019	98% Acc
J. Peng et al.	**Enhanced YOLOv11:** Using DWSCNN + RFCBAM for precise detection (IEEE ICCNSE 2025)	Kaggle ALL	98.6% Prec

3.3 Key Trends and Achievements

Over 75% of publications from 2024–2025 have transitioned from single-cell to multi-cell imaging. Lightweight architectures enable smartphone deployment, while explainable AI techniques (Grad-CAM, SHAP) build clinical trust.

4 Conclusion

This comprehensive review demonstrates that automated ALL diagnosis using deep learning has reached clinical viability. Current state-of-the-art approaches achieve precision exceeding 98% while maintaining real-time processing speeds suitable for point-of-care applications.

References

[1] A. Awad and S. A. Aly, "Early Diagnosis of Acute Lymphoblastic Leukemia Using YOLOv8 and YOLOv11 Deep Learning Models," in *Proc. 2024 2nd Int. Conf. Signal Processing and Intelligent Systems (ICSPIS)*, Dec. 2024.

[2] G. Naveen Prashanth et al., "Integrating Custom GAN Segmentation with Advanced Deep Learning Classifiers," in *Proc. 2024 Int. Conf. Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, Apr. 2024.

[3] J. Peng and F. Li, "An Enhanced YOLOv11 Model for Lightweight and Efficient Precise Leukemia Detection," in *Proc. 2025 IEEE Int. Conf. Communication Networks and Smart Systems Engineering (ICCNSE)*, Aug. 2025.