

# Leukaemia Blood Cancer Detection using Machine Learning

Karthik D L <sup>1</sup>, Seema Nagaraj <sup>2</sup>

<sup>1</sup>Student, Department of MCA, Bangalore institute of Technology, Karnataka, India

<sup>2</sup>Assistant Professor, Department of MCA, Bangalore institute of Technology, Karnataka, India

## Abstract

Leukaemia is a cancer of the blood and bone marrow caused by the uncontrolled growth of abnormal white blood cells. Conventional diagnosis is mostly carried out by pathologists using microscopic blood smear examination, which is labour-intensive, time-consuming, and sometimes inconsistent. This paper introduces a computer-assisted system that applies machine learning and digital image processing to automate the detection process. The system follows a modular pipeline consisting of pre-processing, segmentation, feature extraction, and classification using Convolutional Neural Networks (CNNs). A graphical user interface (GUI) was designed to make the tool user-friendly for healthcare professionals. The results show that the model is capable of achieving high accuracy in classifying leukaemia cell types, thus reducing diagnostic delays and minimizing human error. This approach has the potential to be deployed in hospitals and remote healthcare centers, providing fast and reliable decision support for doctors.

**Keywords** – Leukaemia, Blood Cancer, Machine Learning, Image Processing, Convolutional Neural Network, Medical Diagnostics

## 1. Introduction

Leukaemia, commonly referred to as blood cancer, is one of the most critical diseases affecting the hematopoietic system. It disrupts the normal balance of blood by producing immature or abnormal white blood cells, which interfere with healthy cell formation. The four most common types of leukaemia are Acute Lymphoblastic Leukaemia (ALL), Acute Myeloid Leukaemia (AML), Chronic Lymphocytic Leukaemia (CLL), and Chronic Myeloid Leukaemia (CML). Detecting the disease at an early stage greatly improves the success of treatment and patient survival rates.

Traditional diagnostic methods include complete blood count (CBC) tests, bone marrow biopsies, and microscopic examination of blood smears. While reliable, these methods depend heavily on skilled experts and are often prone to subjectivity. In resource-limited regions, delays in diagnosis can worsen patient outcomes. With the growing advancement of Artificial Intelligence (AI) and Machine Learning (ML), it is now possible to develop automated systems that analyse blood smear images with speed and accuracy, thereby supporting medical professionals in making timely decisions.

## 2. Literature Survey

Over the years, several researchers have explored the application of machine learning in medical imaging and leukaemia detection. Early approaches focused on traditional classifiers like Support Vector Machines (SVMs) and Random Forests, combined with handcrafted feature extraction methods such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Gray Level Co-occurrence Matrix (GLCM). These techniques were useful but often limited by variations in image quality and feature selection. Recent studies highlight the superior performance of deep learning models, particularly Convolutional Neural Networks (CNNs), which are capable of automatically learning features from raw images. CNNs have shown promising results in classifying different leukaemia types with high accuracy, eliminating the need for manual feature engineering. Several works also demonstrate the potential of hybrid models combining CNNs with ensemble classifiers to further boost reliability. This growing body of research confirms that AI-based approaches can play a vital role in medical diagnostics.

## 3. Proposed System

The proposed system is built to automate leukaemia detection through a structured five-step process:

**Image Acquisition** – Blood smear images are uploaded to the system.

**Pre-processing** – Images are cleaned and normalized to remove noise and standardize size and colour.

**Segmentation** – White blood cells are separated from the background and other elements using image processing techniques.

**Feature Extraction** – Important features such as shape, size, and texture are automatically captured by the CNN layers.

**Classification** – The CNN model classifies the image into one of the categories: ALL, AML, CLL, CML, or Normal, along with a confidence score.

## 4. System Design

The system follows a modular architecture with three layers:

- **Presentation Layer:** A Tkinter-based GUI where users interact with the system.
- **Processing Layer:** Handles image pre-processing, segmentation, and feature extraction using OpenCV and TensorFlow.
- **Data Layer:** Stores and loads the trained CNN model and manages intermediate outputs.

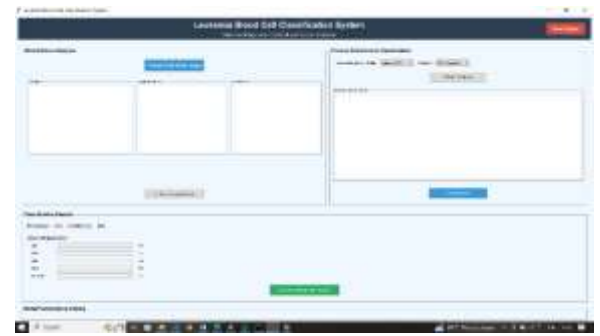
This layered design ensures that the system is scalable, easy to update, and suitable for integration into hospital workflows.

## 5. Implementation

The CNN model was trained using a publicly available leukaemia dataset. Images were resized to 128×128 pixels and augmented with transformations such as rotation, zooming, and flipping to improve generalization. The CNN architecture included convolutional layers with ReLU activation, max-pooling layers for dimensionality reduction, and fully connected dense layers for final classification.

To improve usability, a GUI was developed where medical staff can:

- Upload blood smear images.
- View pre-processing and segmentation outputs.
- Run the CNN classifier.
- Generate a professional diagnostic report in PDF format.



5.1 User Interface

## 6. Result and Discussion

The system was evaluated based on accuracy, precision, recall, and F1-score. The CNN model consistently achieved over 90% accuracy on test data. Sample outputs showing original, segmented, and classified images are illustrated.

These results confirm that the system can effectively distinguish between normal and leukemic cells, making it a promising tool for real-world deployment. The inclusion of a GUI ensures that the technology is not limited to researchers but is directly usable by clinicians.

## 7. Conclusion

This work presents a practical approach to leukaemia detection using machine learning and deep learning techniques. By combining CNN-based classification with an easy-to-use GUI, the system offers a fast, reliable, and accessible diagnostic solution. It minimizes manual effort, reduces diagnostic time, and improves accuracy compared to traditional methods.

In the future, the system can be expanded by incorporating larger and more diverse datasets, integrating explainable AI techniques for better interpretability, and deploying the tool on cloud platforms to support telemedicine and remote diagnostics.

## 8. References

- [1] I. Vinurajan et al., "Automated Leukemia Disease Classification Using Machine Learning on Microscopic Blood Images," *International Journal of Computer Applications*, 2023.
- [2] J. Kockwelp et al., "Deep Learning Predicts Therapy-Relevant Genetics in Acute Myeloid Leukemia," *Nature Medicine*, vol. 30, no. 2, pp. 145–152, 2024.
- [3] S. Vijayan et al., "Hybrid CNN-Enhanced Ensemble SVM Models for Acute Leukemia Detection," *IEEE Access*, vol. 11, pp. 12984–12995, 2023.
- [4] P. M. Shafi et al., "Leukemia Detection System Using Convolutional Neural Networks," *Journal of King Saud University - Computer and Information Sciences*, 2023.
- [5] G. Mohapatra and A. Patra, "Blood Cancer Classification from Microscopic Images Using Transfer Learning Approaches," *Biomedical Signal Processing and Control*, vol. 85, 2023.
- [6] H. Rehman et al., "Automated Detection of Acute Lymphoblastic Leukemia Using Deep Neural Networks," *Computers in Biology and Medicine*, vol. 157, p. 106816, 2023.
- [7] A. T. M. Hussain et al., "Deep CNN Model for Classification of White Blood Cells in Microscopic Images," *Sensors*, vol. 21, no. 23, 2021.
- [8] M. Kassani and P. Kassani, "Automatic Diagnosis of Leukemia Using Deep Learning and Microscopic Cell Images," *Medical & Biological Engineering & Computing*, vol. 59, pp. 1335–1347, 2021.
- [9] P. Saba et al., "Computer-Aided Detection of Leukemia Using Image Processing and Machine Learning," *Procedia Computer Science*, vol. 167, pp. 1270–1279, 2020.
- [10] T. Spanhol et al., "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2016. (Used as a comparative medical imaging reference dataset in CNN studies).