

An Intelligent Leukemia Detection and Disease Classification System

L. Dharshana Deepthi¹, Preethi Shree J², Rhamya Sri Ishwarya Lakshmi³, Rohita Pandi R⁴ and Sabitha P⁵

¹Assistant Professor, Department of CSE, PSNA College of Engineering and Technology, Dindigul, India
Email: dharshnadeepthi@gmail.com

²⁻⁵PSNA College of Engineering and Technology, Dindigul, India,
Email: preethishreej21cs@psnacet.edu.in, rhamyasriiswaryalakshmit21cs@psnacet.edu.in, rohitapandir21cs@psnacet.edu.in, sabithap21cs@psnacet.edu.in

Abstract— The disease threatens life by blood cancer called leukemia because of clonal expansion of white blood cells, leading to abrogation of normal hematopoiesis as well as immune response. The disease originates in the bone marrow and further accelerates its pace into important organs through the blood vessels. Leukemia accounts for nearly 3% of all cancers on earth, with an estimated annual new incidence of 474,500 cases together with death of 311,500 individuals from the disease. Across all ages, leukemia affects people: Acute Lymphoblastic Leukemia (ALL) occurs more in children; Acute Myeloid Leukemia (AML) is more in adults. Some early features that manifest to give signs of leukemia are fatigue all the time, frequent infections, loss of weight without any reason, bruising without obvious cause, extended bleeding, and swelling in lymph nodes. Early diagnosis and accuracy in diagnosis are thus important to improve the prognosis. This research is an automated Leukemia Classification and Detection System using deep learning to improve diagnosis accuracy. This uses CNNs for feature extraction and VGG-19 for the ultimate classification into subtypes of leukemia from images of blood smears. The organized pipeline involves image preprocessing, feature extraction, and subtype classification for accurate detection. A web application has been created under Flask and Streamlit into which health care providers can upload patient images for diagnosis of leukemia in real time. Experimental results prove that the suggested model greatly enhances accuracy, making it an effective tool for early leukemia detection and clinical decision making.

Index Terms— Leukemia Detection, Deep Learning, Multi-Modal Data, VGG-19, Random Forest, Web Application, Medical Image Processing.

I. INTRODUCTION

Leukemia refers to a malignancy that affects blood and is produced in the bone marrow, leading to abnormal reproduction of white blood cells. This condition alters the process of production of blood cells as well as the immune function. Leukemia is among the most commonly reported forms of cancer as far as blood is concerned, with thousands of new cases being reported every year. Thus, early diagnosis and accurate identification of subtypes of leukemia greatly improve treatment efficacy and patient survival rates. Traditional diagnosis of leukemia usually includes microscopic viewing of blood smear images and clinical assessment of the patient symptoms done by senior hematologists.

Such processes of diagnosis are laborious, time-consuming, and are also prone to human errors. Progress made lately in artificial intelligence (AI) and deep learning has shown great promises in improving the efficiency and accuracy of processing and analyzing medical images and hence automated detection of leukemia promises to be an exciting area of exploration. The work pries open an already developed Automated Leukemia Detection and Classification System utilizing Convolutional Neural Networks (CNNs) for feature extraction and VGG-19 for the classification of subtypes of cancer. This system further enhances a complete systematic pipeline that is Image Preprocessing, Feature Extraction, Classification-VGG-19, the one to classify leukemia into different subtypes.

To enhance access, the model is web-based using Flask and Streamlit, which allows clinicians to upload leukemia images for diagnosis. The system performance will be evaluated according to parameters as accuracy, precision, recall, and F1 score so that it can be considered reliable enough to be used in the clinical environment. The system generates doctor recommendation, prevention, and treatment advice for patients.

II. LITERATURE REVIEW

Faivdullah et al(2023) suggested a three-layered leukemia classification system using blood smear images. The system consisted of feature extraction by Dense Scale-Invariant Feature Transform (Dense-SIFT), followed by coding and classification with a multi-class SVM classifier. The method obtained a classification accuracy of 79.37% using a database of 400 samples. Likewise, Kumar et al.2 (2023) presented an automatic acute leukemia classification system that began with noise reduction of microscopic images. Different features were extracted, and classification was carried out by using k-nearest neighbor (K-NN) and Naïve Bayes classifiers. The K-NN classifier reached an accuracy level of 92.8% using a set of 60 blood samples. These studies exhibit progress in techniques of leukemia classification using both traditional machine learning as well as more advanced feature extraction. Supardi et al (2022) created a classification system to differentiate between Acute Myeloid Leukemia (AML) and Acute Lymphocytic Leukemia (ALL). They manually extracted twelve features from image samples and used a K-NN classifier, with an accuracy of 86% on a database of 1,500 photos. Qin et al. 4 (2022) tried to classify up to 32 types of white blood cells using a deep learning model. Their model, however, was hindered by limited training data and hence had fairly low classification performance. These papers emphasize the drawbacks and possible improvements in the detection of leukemia using image-based classification methods.

Setiawan et al.(2021) offered a classification model for AML subtypes M4, M5, and M7 through color k-means clustering-based cell segmentation. They obtained six statistical features and learned a multi-class SVM classifier with an 87% segmentation accuracy and 92.9% classification accuracy. In the same way, Madhukar et al.6(2021) proposed a classification method for AML by improving image contrast and extracting five significant features. With an SVM classifier, they obtained a classification accuracy of 93.5% using a database of 50 images. These research works highlight segmentation and feature extraction as the key areas for enhancing leukemia classification accuracy. Patel and Mishra (2020) proposed an automated detection approach of leukemia from microscopic images. Their method entailed noise removal, k-means clustering-based segmentation, and feature extraction, which was followed by SVM-based classification. They attained an accuracy of 93.57% on a dataset of 27 images.

III. PROPOSED SYSTEM

Leukemia Detection and Classification System is a web application with deep learning implementation to automatically classify and detect leukemia from microscopic blood smear images. The system implements CNN-based feature extraction and classification using VGG-19 for efficient and precise leukemia detection functionality.

A. Data Acquisition

The data used in this research is a publicly available dataset of peripheral blood smear (PBS) images for acute lymphoblastic leukemia (ALL) detection and classification. The dataset was developed in the Bone Marrow Laboratory of Taleqani Hospital, Tehran, Iran, and is composed of 3,256 high-resolution images from 89 suspected ALL patients. The laboratory personnel stained the blood samples, and photographs were taken with a Zeiss 100x magnification microscope. The dataset is organized into benign (normal) samples and malignant (ALL) conditions, and malignant samples are divided further into three leukemia types:

- Benign (Normal) – Contains hematogones (normal precursor B-cells) images.
- Early Pre-B ALL – Immature lymphoblasts are identified in early phases of leukemia.
- Pre-B ALL – Intermediate-stage malignant lymphoblasts.

- Pro-B ALL – Poorly differentiated blast cells aggressive subtype of ALL.

TABLE I. CLASSIFICATION OF BLOOD SMEAR IMAGES

Classes	Image Counts
Benign	504
Early	985
Pre	963
Pro	804

B. Preprocessing

- Load Dataset: The dataset is 3,256 images of blood smears classified as benign or malignant (Early Pre-B, Pre-B, and Pro-B ALL subtypes).
- Image Resizing: The images are all resized to 224×224 pixels for compatibility with the model.
- Grayscale Conversion: Images are transformed to grayscale through the formula $\text{imgGray} = 0.2989 \times R + 0.5870 \times G + 0.1140 \times B$
- Normalization: Pixel intensity is normalized to [0,1] to facilitate efficient training.

Data Augmentation: Augmentation techniques such as rotation, flipping, and contrast adjustment are used to enhance model robustness.

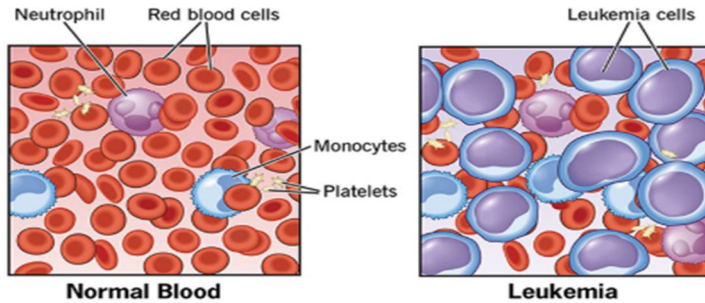


Figure 1. Normal blood vs leukemia blood cell

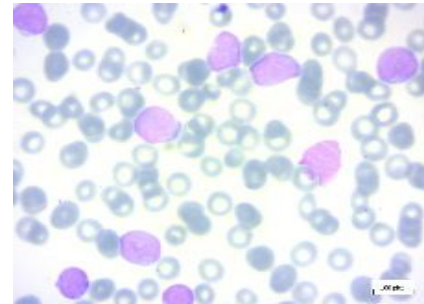


Figure 2. Affected Blood Smear Image

C. Feature Extraction

To improve leukemia classification, key features were learned from peripheral blood smear (PBS) images with the help of Convolutional Neural Networks (CNN). Feature extraction was done through the following steps: Texture Analysis & Edge Detection, Processing in Gray Level, Feature Extraction via CNN

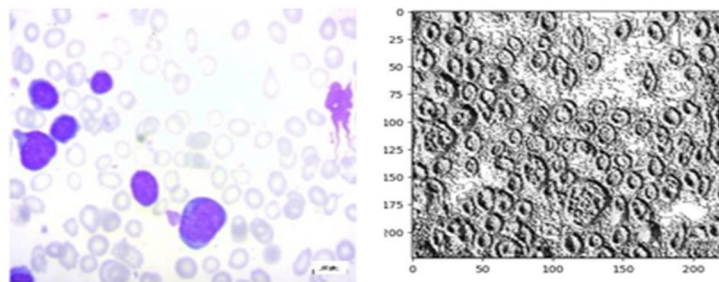


Figure.3. Original image vs Featured image

D. Data Splitting

The dataset is split into 80% for training and 20% for testing to facilitate effective model assessment. The stratified split ensures class distribution and avoids bias during training.

E. Model Training and Classification

For leukemia classification, a deep convolutional neural network (CNN) is utilized, based on VGG-19 for feature extraction and classification. The model structure includes Convolutional Layers, Pooling Layers, Fully Connected Layers & Softmax Activation. The proposed approach enhances leukemia diagnosis by leveraging deep learning-based feature extraction and classification, improving accuracy and reliability in medical imaging applications.

Block Diagram

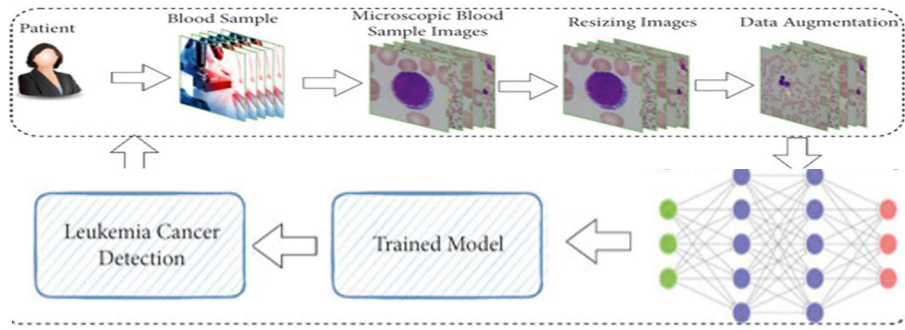


Figure 4: Architecture of the proposed system

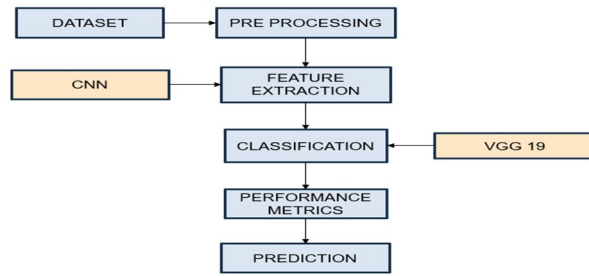


Figure 5. Block diagram

F. Performance Evaluation

To assess the effectiveness of the Leukemia Detection and Classification System, the model is evaluated using standard performance metrics: Accuracy, Precision, Recall, F1-Score, and Confusion Matrix.

Accuracy

Accuracy measures the overall correctness of the model and is given by:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where TP = True Positives (correct leukemia detections), TN = True Negatives (correct normal case detections), FP = False Positives (normal cases misclassified as leukemia), FN = False Negatives (leukemia cases misclassified as normal)

Precision

Precision measures how many of the predicted leukemia cases are actually correct:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Higher precision indicates fewer false positives.

Recall (Sensitivity or True Positive Rate)

Recall evaluates how well the model identifies actual leukemia cases:

$$\text{Recall} = \frac{TP}{TP + FN}$$

A higher recall means fewer false negatives, ensuring fewer missed leukemia cases.

F1-Score

The F1-score is the harmonic mean of Precision and Recall, balancing false positives and false negatives:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

A high F1-score indicates a well-balanced model.

Confusion Matrix

The confusion matrix provides a detailed breakdown of predictions:

TABLE II. CONFUSION MATRIX

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

The evaluation results indicate that the Leukemia Detection and Classification System achieves high accuracy and reliability in identifying leukemia cases. The use of deep learning for image classification enables robust detection, ensuring an effective and comprehensive diagnostic approach.

G. Web Application Deployment

The VGG-19 model is trained and deployed as a web application with Flask used as the backend for processing and Streamlit for the interactive user interface, allowing users to upload blood smear images for automated leukemia classification.

Some key features of the web application include Image Upload, Leukemia Classification, Result Visualization, Report Download

H. Summary of Implementation

The proposed Leukemia Detection and Classification System uses deep learning to automate the leukemia diagnosis from blood smear images. The proposed system includes the following characteristics:

- Deep Learning (VGG-19): For image-based leukemia classification using blood smear images.
- Flask & Streamlit: Create an interactive web application for real-time classification and result visualization.

By giving impetus to CNN-based feature extraction, this system accentuates early detection of leukemia, thus bringing improved accuracy in diagnosis, and hence is fast becoming an AI-oriented approach concerning medical image classification.

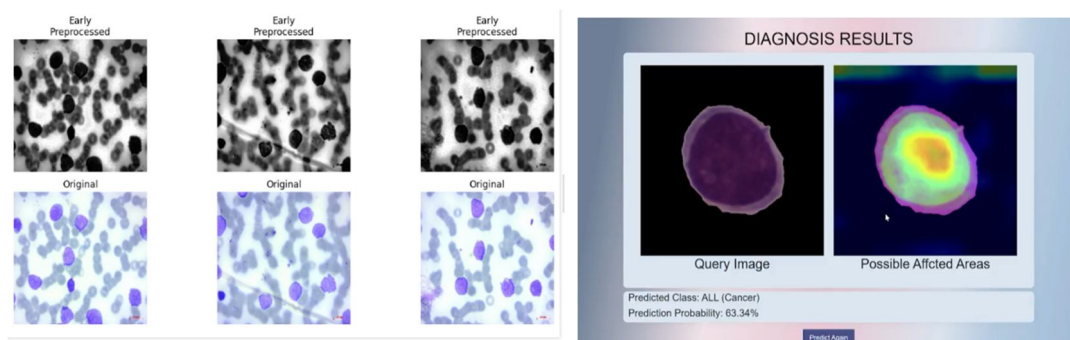


Figure 6. Possible output

Advantages

- Early Detection
- High Accuracy
- Automated and Fast Processing
- Reduced Dependency on Experts

IV. CONCLUSION

The Leukemia Detection and Classification System implements deep learning techniques to automate the diagnosis of leukemia from blood smear images. The VGG-19 model is employed for the extraction of features and classification concerning the detection of leukemia with very high accuracy. Furthermore, the application integrating Flask and Streamlit allows seamless interaction with users, making it a useful and readily accessible solution for health professionals and researchers. The evaluated performances of the system, i.e., accuracy, precision, recall, F1-score, and confusion matrix, suggest that the system is reliable in identifying different subtypes of leukemia.

Future work can enhance the full realization of the system by:

- Expanding Accessibility – Web platform optimization for maximal uptake within the clinical settings.
- Leveraging Advanced Technologies – Cloud-deployable solutions aimed toward scalability and real-time processing.

Through the development of these future advancements, the system can be transformed into a real-time, scalable, and worldwide accessible platform for the detection of leukemia, establishing further milestones in AI-driven medical diagnostics development.

REFERENCES

- [1] Md. Abu Ahnaf Mollick, Md. Mahfujur Rahman, D.M. Asadujjaman, Abdullah Tamim, Nosin Anjum Dristi, Md. Takbir Hossen. "Detection and Classification of Acute Lymphoblastic Leukemia Utilizing Deep Transfer Learning" in arXiv preprint arXiv:2501.14228,2025.
- [2] Rakhmonalieva Farangis Oybek Kizi, Tagne Poupi Theodore Armand, Hee-Cheol Kim. "A Review of Deep Learning Techniques for Leukemia Cancer Classification Based on Blood Smear Images" in Applied Biosciences, Vol. 4, No. 1, 2025.
- [3] Hakar Hasan Rasheed, Adnan Mohsin Abdulazeez. "Leukemia Detection and Classification Based on Machine Learning and CNN: A Review" in The Indonesian Journal of Computer Science, Vol. 13, No. 3, 2024.
- [4] Chandravaradhan Singh Raghaw, Arnav Sharma, Shubhi Bansal, Mohammad Zia Ur Rehman, Nagendra Kumar. "CoTCoNet: An Optimized Coupled Transformer-Convolutional Network with an Adaptive Graph Reconstruction for Leukemia Detection" in arXiv preprint arXiv:2410.08797, 2024.
- [5] Maksuda Akter, Rabea Khatun, Md Manowarul Islam. "A Hybrid Feature Fusion Deep Learning Framework for Leukemia Cancer Detection in Microscopic Blood Samples Using Gated Recurrent Unit and Uncertainty Quantification" in arXiv preprint arXiv:2410.14536, 2024.
- [6] Wahidul Hasan Abir, Md. Fahim Uddin, Faria Rahman Khanam, Mohammad Monirujjaman Khan. "Explainable AI in Diagnosing and Anticipating Leukemia Using Transfer Learning Method" in arXiv preprint arXiv:2312.00487, 2023.
- [7] Abdelmageed Ahmed, Alaa Nagy, Ahmed Kamal, Daila Farghl. "Leukemia Detection Based on Microscopic Blood Smear Images Using Deep Learning" in arXiv preprint arXiv:2301.03367, 2022.
- [8] Md. Shafiqul Islam, Md. Kamrul Hasan, Md. Nazmul Haque, Md. Rabiul Islam. "Automatic Detection of Leukemia from Microscopic Blood Images Using Convolutional Neural Networks" in Journal of Healthcare Engineering, hindawi.com, Vol. 2021, Article ID 6641079, 2021.
- [9] Muhammad Zakir Ullah, Yuanjie Zheng, Jingqi Song, Sehrish Aslam, Chenxi Xu, Gogo Dauda Kiazolu, Liping Wang. "An Attention-Based Convolutional Neural Network for Acute Lymphoblastic Leukemia Classification" in Applied Sciences, Vol. 11, No. 22, November 2021, pp. 1-16.
- [10] Shakir M. Abas, Adnan M. Abdulazeez. "Detection and Classification of Leukocytes in Leukemia using YOLOv2 with CNN" in Asian Journal of Research in Computer Science, journalajrcos.com, Vol. 8, No. 3, May 2021, pp. 64-75.