Malnad College of Engineering, Hassan

(An Autonomous Institution affiliated to VTU, Belgavi)

Main Phase-1 Project Report On

"HSL BASED ANOMALY DETECTION OF LEUKEMIA USING MACHINE LEARNING AND DEEP LEARNING"

Submitted in partial fulfillment of
the requirements for the award of the degree of

Bachelor of Engineering in
Computer Science and Engineering

Submitted by

Pooja K P	4MC20CS107
Bhagyashree S A	4MC22CS023
C S Sathvik Gowda	4MC22CS030
Chaithra H	4MC22CS032

Under the guidance of
Dr. Keerthi Kumar H M
Associate Professor

Department of Computer Science and Engineering

2024-2025

Malnad College of Engineering
Department of Computer Science and Engineering

Certificate

Certified that the Main project phase-1 work entitled **"HSL Based Anomaly Detection of Leukemia Using Machine Learning And Deep Learning"** is a bonafide work carried out by **Pooja K P (4MC20CS107), Bhagyashree S A (4MC22CS023), C S Sathvik Gowda (4MC22CS030), Chaithra H (4MC22CS032)** in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering of the Visvesvaraya Technological University, Belgavi during the year 2024-2025. The project report has been approved as it satisfies the academic requirements in respect of main phase-1 project work prescribed for the Bachelor of Engineering Degree.

Signature of the Guide	Signature of the HOD	Signature of the Principal
Dr. Keerthi Kumar H M	*Dr. Chandrika Ramesh*	*Dr. A. J. Krishnayya*
Associate Professor	Professor and Head	Principal
Dept. of CSE, MCE	Dept. of CSE, MCE	MCE

Examiners

Name of the Examiner	Signature of the Examiner
1.	
2.	

ABSTRACT

Leukemia, a malignancy affecting the bone marrow and blood, demands rapid and accurate diagnosis due to its aggressive progression. Traditional diagnostic methods are often time-consuming, subjective, and reliant on expert analysis. In recent years, the integration of **Machine Learning (ML)** and **Deep Learning (DL)** approaches has revolutionized medical diagnostics, especially for leukemia detection through blood smear image analysis. This survey paper critically reviews fifteen peer-reviewed research works focusing on various ML and DL techniques—including CNNs, SVM, transfer learning, and image preprocessing strategies—for classifying leukemia types such as Acute Lymphoblastic Leukemia (ALL) and Acute Myeloid Leukemia (AML). Among them, CNN-based and hybrid models consistently achieved high accuracy, often exceeding 95%, especially when using datasets like ALL-IDB and C-NMC. Special attention was given to preprocessing techniques involving **color space conversion**, improving leukemic cell segmentation. This insight aligns with our proposed project goal to implement **HSL-based image analysis for anomaly detection** in leukemia diagnosis. The review highlights current strengths, limitations, and the promising future of explainable AI, dataset augmentation using GANs, and multimodal data fusion in clinical applications.

ACKNOWLEDGEMENT

Salutations to our beloved and highly esteemed institute, **Malnad College of Engineering**, for its excellent academic environment, well-qualified faculty, and laboratories equipped with modern facilities essential for technical development.

The successful completion of any task is never the result of individual effort alone. It is a product of guidance, support, and encouragement from several individuals. We take this opportunity to extend our sincere gratitude to all those who helped us complete our project and this report.

We would like to express our deepest gratitude to our mentor, **Dr. Keerthi Kumar H M**, Associate Professor, Department of **Computer Science and Engineering**, for her constant encouragement, invaluable guidance, and unwavering support throughout the project. Her insightful feedback and expert supervision played a vital role in shaping the quality and direction of our work. We consider it a privilege to have worked under her mentorship, and we have gained much more than just technical knowledge through her guidance.

We are also immensely thankful to **Dr. A. J. Krishnayya**, Principal, Malnad College of Engineering, for providing us with the opportunity and necessary facilities to carry out our project. His leadership and support created a conducive environment for academic research and innovation.

We also express our heartfelt appreciation to all the faculty members and staff of the **Department of Computer Science and Engineering** for their constant help, motivation, and encouragement throughout this journey.

Pooja K P
Bhagyashree S A
C S Sathvik Gowda
Chaithra H

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Introduction to Leukemia Detection

Leukemia is a life-threatening cancer of the blood and bone marrow that interferes with the production and function of blood cells. It is classified based on the speed of progression (acute or chronic) and the type of blood cells affected (lymphocytic or myeloid). Traditional methods of leukemia diagnosis involve manual examination of blood smear images by pathologists, which is time-consuming and prone to human error. The emergence of machine learning and deep learning techniques provides automated, accurate, and fast alternatives for leukemia detection.

1.2 About the Project

1.2.1 Problem Statement

Traditional leukemia diagnostic techniques are often reliant on manual interpretation, which can result in delays, misdiagnoses, and inconsistencies. These limitations necessitate the development of an automated and robust system that can accurately detect anomalies in blood smear images using advanced image processing and machine learning techniques.

1.2.2 Objective

- To develop an HSL-based image analysis system for detecting anomalies in leukemia blood smear images.
- To convert RGB images into the HSL color space to better capture color and texture variations of white blood cells (WBCs).
- To implement Convolutional Neural Networks (CNNs) to extract spatial and color features from HSL images.
- To apply Residual Networks (ResNet) with transfer learning for improved feature extraction and classification accuracy.
- To explore using ResNet as a feature extractor combined with anomaly detection algorithms (e.g., Isolation Forest, One-Class SVM) for detecting abnormal WBCs.
- To compare the performance of machine learning models such as SVM, Random Forest, KNN, alongside deep learning models like CNN and ResNet, in leukemia detection.
- To develop a real-time, interactive web application using Streamlit that automates leukemia anomaly detection and assists hematologists by providing fast, accurate diagnostic support.

Chapter 2

Literature Survey

2.1 Review of Relevant Literature

Leukemia detection using artificial intelligence has gained considerable momentum in the past decade, particularly with the emergence of machine learning (ML) and deep learning (DL) technologies. Several studies have explored and validated these techniques to automate the identification and classification of leukemia subtypes based on blood smear images.

Traditional machine learning models such as Support Vector Machines (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), and Decision Trees (DT) have been applied extensively. These methods typically rely on manually extracted features such as color histograms, texture (e.g., using Gray Level Co-occurrence Matrix - GLCM), and shape descriptors. For instance, Patil and Hiremath [2] utilized texture features and watershed segmentation with RF to distinguish between chronic leukemia subtypes (CLL and CML), achieving decent interpretability, though they lacked validation across diverse datasets.

However, the shift toward deep learning—especially Convolutional Neural Networks (CNNs)—has significantly improved diagnostic accuracy by enabling automatic feature extraction from image data. Shafique and Tehsin [12] fine-tuned AlexNet to detect ALL subtypes (L1, L2, L3), achieving up to 99.5% accuracy. These models eliminate the need for manual preprocessing and perform well even without segmentation, but often lack robustness when validated outside the training domain.

Transfer learning, using pretrained networks such as ResNet50, MobileNet, and DenseNet, has become a popular approach for tackling limited dataset availability. Studies show that combining these models with traditional classifiers like SVM and logistic regression enhances classification performance, reaching accuracies above 95% [7], [3]. Additionally, ensemble learning methods (e.g., combining ResNet and DenseNet) have achieved near- perfect results in some experimental settings [3].

Several works emphasize the importance of preprocessing techniques to enhance model effectiveness. Methods like HSV and HSI color space conversion, morphological filtering, histogram equalization, and edge detection improve leukocyte segmentation and classification. Notably, Harun et al. [10] showed that using the HSI color model preserved cell shape better than RGB, improving segmentation accuracy—an insight that directly supports the use of HSL in this project.

Addressing data scarcity, researchers like Ansari et al. [15] have employed Generative Adversarial Networks (GANs) to augment datasets, significantly boosting model accuracy (up to 99%) even when real data is limited. Nonetheless, most studies still depend on standard datasets such as ALL-IDB [11], C-NMC_Leukemia [1], and Raabin Health, which, although useful for benchmarking, may not represent clinical variability.

In summary, existing literature strongly supports the use of ML and DL for leukemia detection, with CNNs, transfer learning, and image preprocessing forming the core of high-performance models. However, issues such as lack of external validation, model interpretability, and dataset generalization remain. These gaps open opportunities for further research—particularly in integrating HSL-based color preprocessing and explainable AI (XAI) methods to build more robust, transparent, and clinically viable diagnostic systems.

2.2 Current Models, Methods, and Methodology

Numerous machine learning (ML) and deep learning (DL) approaches have been explored for leukemia detection, particularly through analysis of peripheral blood smear images. Traditional ML algorithms—including Support Vector Machines (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN)—typically rely on handcrafted features such as shape, texture, and color [5][8][14]. While these methods are relatively simple and offer interpretability, they often struggle with generalization across complex or diverse datasets.

In contrast, deep learning techniques—especially Convolutional Neural Networks (CNNs) have demonstrated superior performance by learning features directly from raw image data [6][12][15]. Popular architectures like AlexNet, ResNet50, and DenseNet have achieved classification accuracies exceeding 95% in many studies [3][4][7]. Transfer learning using pretrained models is frequently employed to address data scarcity, reduce training time, and improve generalization [3][7][12].

Image preprocessing methods—such as segmentation, histogram equalization, and color space transformation (e.g., RGB to HSV/HSI)—significantly enhance model performance. Research has shown that alternative color spaces like HSI or HSL can improve cell segmentation and background removal, justifying the use of the HSL model in this project [10][13].

To mitigate limited dataset challenges, data augmentation techniques like flipping, rotation, and generative adversarial networks (GANs) are commonly applied [3][4]. Publicly available datasets, including ALL-IDB and C-NMC, are widely used for training and evaluating models [1][11]. Despite substantial progress, challenges such as overfitting, dataset bias, and limited clinical validation remain areas of ongoing research [2][5][15].

S. No.	Methodology	Dataset Used	Performance Metrics
1	Optimized CNN (OCNN)	C-NMC Leukemia (Kaggle)	Accuracy: 99.99%, Precision: 99.97%, Recall: 100%, F1-score: 94.07%
2	Morphology + GLCM + Random Forest	1200 Microscopic Images	Accuracy, Sensitivity, Specificity: 97%
3	Weighted Deep Ensemble (ResNet-152)	Kaggle PBS (3940 samples, 4 subtypes)	Accuracy: 99.92%, F1-score: 99.90%, AROC: 0.999, MCC: 0.897
4	Custom CNN vs MobileNetV2 + HSV Preprocessing	Taleqani Hospital (3256 PBS images)	MobileNetV2: Accuracy: 97.81%, Custom CNN: Accuracy: 94.21%, F1: up to 0.984
5	SVM, RF, KNN, DT with PSO optimization	3256 PBS (Kaggle)	SVM+PSO: 78.9%, SVM w/o PSO: 77.38%, PSO improved all classifiers
6	Preprocessing + Custom CNN	10,661 Blood Slide Images	Accuracy: 81.3%
7	Transfer Learning (ResNet50, MobileNet, etc.)	500 PBS images (AIIMS Patna)	Accuracy: Up to 96% (ResNet50 + LR)
8	CNN-based Leukemia Detection	1000 PBS Images	Accuracy: >90%
9	HSV-Based Rule-Based Classifier	60 PBS Images	AML: 100%, ALL: 80%, Combined: 90%
10	HSI & RGB Segmentation	Acute Leukemia Slide Images	HSI: Good segmentation; RGB: Misidentification
11	FAB Classification Method (L1, L2, L3)	ALL-IDB1, ALL-IDB2	Accuracy, Sensitivity, Specificity (Not quantified)
12	AlexNet + Transfer Learning	ALL-IDB2	Detection Accuracy: 99.5%, Subtype Classification: 96.06%
13	Morphological Features + SVM	200 Blood Smear Images	Accuracy: 99%
14	Histogram Eq. + Sobel + Stat Features + SVM	90 Blood Cell Images	Segmentation Accuracy: 91%
15	Proposed Method: GAN + Customized CNN	Private Dataset (Shahid Ghazi Oncology Center)	Accuracy: 99.5%, F1: >99%, Sensitivity/Specificity/Precision/Kappa: >99%

Table 2.1: Comparative Analysis of Methodologies, Performance Metrics, and Datasets in Leukemia Detection

2.3 Dataset Description

The effectiveness of machine learning and deep learning models for leukemia detection largely depends on the quality, size, and diversity of the datasets used for training and evaluation. Several publicly available datasets have been widely used in research for this purpose, each with its own characteristics, image resolutions, annotation quality, and types of leukemia included. Below are some of the most commonly referenced datasets in the field:

1. **ALL-IDB (Acute Lymphoblastic Leukemia Image Database):** Developed for evaluating segmentation and classification algorithms, this dataset is one of the most widely used in leukemia detection studies. It consists of:

    - ALL-IDB1: High-resolution images of peripheral blood smears labeled with healthy and ALL-infected cells.
    - ALL-IDB2: Cropped images focused on individual WBCs, used for training CNNs and other classifiers.
    - Annotations: Manual annotations provided for cell locations and labels.
    - Use: Useful for both classification and segmentation tasks, especially subtype analysis [10], [11].

2. **C-NMC_Leukemia (Kaggle):** Published on Kaggle, this dataset provides a large collection of labeled blood smear images categorized as normal or leukemia.

    - Categories: Binary classification (leukemia vs. normal).
    - Resolution: Uniform, cleanly segmented cell images.
    - Use: Suitable for training CNN models for binary classification with high accuracy [15].

3. Raabin-WBC Dataset: This dataset includes over 14,000 images of white blood cells collected from 56 individuals, covering different types and morphological variants.

- Annotations: Expert-labeled with WBC types (e.g., neutrophil, lymphocyte, monocyte).
- Variation: Includes cell-level variability due to lighting, focus, and staining differences.
- Use: Ideal for WBC classification and robust model training under varied imaging conditions [7].

4. Shahid Ghazi Oncology Center Dataset (Private): Used in some advanced studies, this dataset is collected from clinical environments and contains annotated images of leukemia patients.

- Use: Applied in combination with GAN-based augmentation for high-accuracy models.
- Limitation: Not publicly available, limiting reproducibility [4].

5. AIIMS Patna Dataset (Peripheral Blood Smears): Used in research involving transfer learning and ensemble models.

- Sample Size: Around 500 annotated PBS images.
- Use: For multi-class classification (ALL, AML, and normal).
- Limitation: Small dataset size; requires augmentation for deep models[6].

Dataset Name	Number of instances	Link or Reference Number of the Paper that used the dataset
ALLIDB1	**108** images in total (59 healthy and 49 ALL).	ALL-IDB Acute Lymphoblastic Leukemia Image Database for Image Processing
ALLIDB2	**260** images in total (130 healthy and 130 ALL).	ALL-IDB Acute Lymphoblastic Leukemia Image Database for Image Processing
Peripheral Blood Smear images	90 images	[14]
C-NMC 2019 dataset	15,114 cells (4037 healthy,8491 non-healthy, and 2586 cells are unlabelled)	https://www.kaggle.com/gauravrajpal/leukemia-classification-v1-3-inceptionv3-65-29/data
Raabin Health dataset	Approximately **40,000** white blood cell images	https://www.raabindata.com/free-data/
Kaggle Acute Lymphoblastic Leukemia (ALL) Dataset	**3256 images (Benign from 25 individuals, Malignant from 64 patients)**	https://www.kaggle.com/datasets/mehradaria/leukemia
Acute Leukemia Slide Images for Color Space Segmentation Comparison	The total number of these images is not specified	[10]
Blood Smear Image Dataset for Leukemia Diagnosis	200 images	[13]
Shahid Ghazi Tabatabai Leukemia Image Dataset	**653 images** (184 ALL and 469 AML)	[15]
Leukemia Blood Cell Image Dataset	10,661 Images	[6]

Table 2.2: Dataset Description

In conclusion, the selection of an appropriate dataset plays a critical role in the development and validation of leukemia detection systems. Public datasets such as ALL-IDB and C-NMC are widely used for benchmarking due to their accessibility and annotation quality. However, real-world deployment demands external validation using clinical datasets with more diversity. For this project, such datasets can be instrumental in training and testing HSL-based preprocessing and classification pipelines, especially in detecting anomalies indicative of leukemia.

Chapter 3

Methodology and System

Design

Leukemia detection through analysis of peripheral blood smear images is a critical task in hematology that demands accurate and timely diagnosis. This project aims to develop a robust leukemia anomaly detection system by leveraging advanced image processing techniques in the HSL color space combined with machine learning and deep learning models. Publicly available datasets such as ALL-IDB and C-NMC_Leukemia, containing annotated white blood cell images, will be utilized to train and evaluate the system. The images will undergo preprocessing steps including resizing, noise reduction, and transformation from RGB to HSL to improve the visual distinction of cellular features like the nucleus and cytoplasm. Deep features will be extracted using convolutional neural networks (CNNs), with transfer learning from architectures such as ResNet enhancing the detection of subtle anomalies. The system will compare the performance of various classifiers, including support vector machines, random forests, and K-nearest neighbors, alongside deep learning models. Finally, a Streamlit-based web application will provide a user-friendly interface for real-time diagnostic support, enabling hematologists to upload images and receive fast predictions. This research- backed methodology aims to improve early leukemia diagnosis through automated and accessible tools.

3.1 Data Acquisition

In this step, peripheral blood smear images will be collected from publicly available datasets such as ALL-IDB and C-NMC_Leukemia. These datasets contain high-resolution microscopic images of white blood cells, which are labeled by medical experts as leukemic or normal. The images have been captured using standardized staining and lighting protocols to ensure consistency and clarity, providing a reliable foundation for supervised learning. High-quality, annotated data is crucial for developing accurate leukemia detection models. The expert annotations enable the system to learn important cellular features related to leukemia. However, variations in staining intensity, image brightness, and cell morphology present challenges that will be addressed during preprocessing. Before training, the datasets will be carefully inspected to verify image quality and label accuracy, ensuring the integrity of the data used for model development.

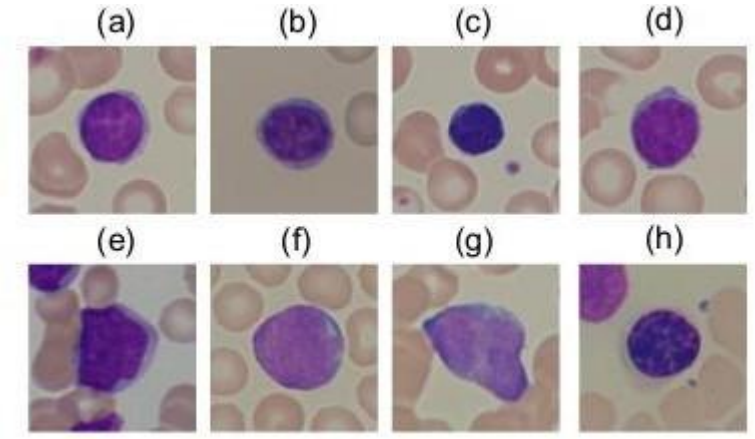

Fig.1 Example images contained in the ALL-IDB: healthy cells from non-ALL patients (a-d), probable lymphoblasts from ALL patients (e-h).

3.2 Preprocessing

To ensure consistency and optimize the input data for subsequent analysis, a series of preprocessing operations are proposed. These steps are designed to enhance image quality, reduce variability, and prepare the data for effective feature extraction by the machine learning and deep learning models. The following steps will be applied to all collected images:

1. **Resizing to a Standard Dimension**

   All images are planned to be resized to a fixed dimension of **224×224 pixels**, which is compatible with standard convolutional neural network (CNN) architectures such as ResNet. This resizing will ensure:

   - Consistency in input shape.
   - Efficient batch processing.
   - Reduced computational cost without losing essential visual features.

2. **RGB to HSL Color Space Conversion**

   To highlight relevant color and structural features of white blood cells (WBCs), the images will be converted from the standard **RGB color space to HSL (Hue, Saturation, Lightness)**. This transformation is expected to:

   - Separate color (hue) and brightness (lightness) components.
   - Enhance visual separation of cell components such as nuclei and cytoplasm.
   - Aid in better segmentation and feature discrimination.

The use of HSL color space is supported by medical imaging studies that demonstrate improved segmentation accuracy for blood cell analysis.

3. Pixel Value Normalization

To standardize pixel intensities, all images will be normalized to a **[0, 1]** range. This will involve dividing each pixel value by 255 (the maximum value in 8-bit images). This normalization is intended to:

- Accelerate model convergence.
- Reduce sensitivity to input scale.
- Ensure balanced weight updates during training.

These preprocessing steps will play a critical role in enhancing image uniformity and visual clarity, laying the foundation for accurate feature extraction and classification in the subsequent phases of the project.

3.3 CNN Feature Extraction

As part of the planned methodology, deep features are intended to be extracted using a **pretrained Convolutional Neural Network (CNN)** architecture. The model selected for this purpose is **ResNet50**, a residual network known for its effectiveness in handling complex image classification tasks due to its deep structure and use of skip connections.

The approach will involve the following steps:

- **Utilizing Pretrained ResNet50**: The pretrained **ResNet50** model (trained on ImageNet) will be used as a fixed feature extractor. The final classification layers of the model will be removed to focus solely on feature extraction.

- **Input Configuration**: Preprocessed blood smear images (resized and normalized) will be fed into ResNet50. The model will process these images through its convolutional and pooling layers to learn hierarchical spatial and color patterns.

- **Feature Vector Output**: Features will be extracted from the **Global Average Pooling (GAP)** layer or the last fully connected layer. These feature vectors (typically of 2048 dimensions in ResNet50) will serve as high-level representations of the images, capturing key WBC characteristics such as shape, texture, and staining differences.

- **Purpose**: These extracted features will later be passed to machine learning classifiers to compare performance across models like SVM, Random Forest, and KNN.

This step is expected to enhance the system's ability to learn relevant features automatically, especially in differentiating healthy and leukemic white blood cells.

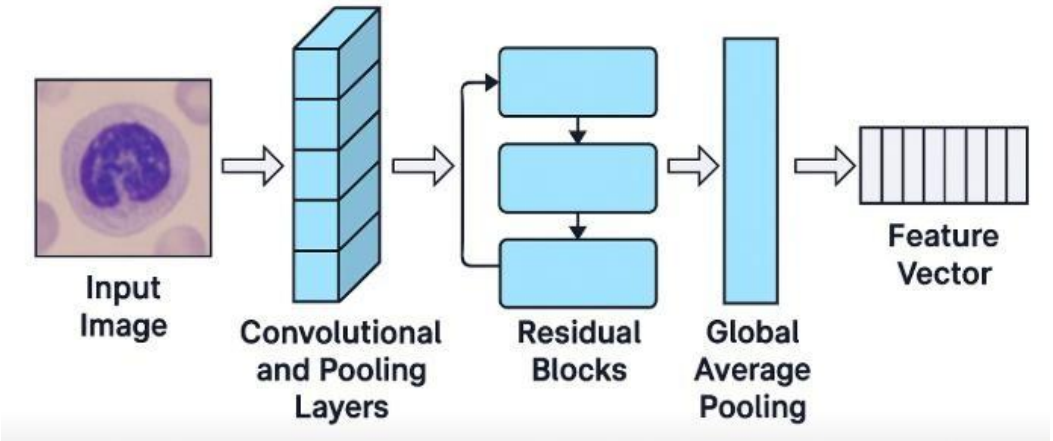

Fig.2 ResNet50 for Feature Extraction.

3.4 Classification

Following feature extraction via Convolutional Neural Networks (CNNs), the extracted features will be input into several traditional machine learning classifiers to perform leukemia anomaly detection. The planned classifiers include:

- **Support Vector Machine (SVM):** Expected to effectively handle high-dimensional feature vectors and provide strong decision boundaries for leukemia classification.
- **Random Forest:** Will be employed for its ensemble-based robustness and ability to capture complex patterns through multiple decision trees.
- **K-Nearest Neighbors (KNN):** Included as a simple, instance-based classifier to provide a baseline for comparison with other models.

The workflow will involve training each classifier on the CNN-extracted features, followed by performance evaluation using appropriate metrics such as accuracy, precision, recall, and F1-score. Hyperparameter tuning techniques, such as grid search or cross-validation, are planned to optimize classifier performance.

This approach aims to balance the strengths of deep learning feature extraction with the interpretability and efficiency of traditional machine learning models. The effectiveness of each classifier will be compared to identify the most suitable method for accurate leukemia detection.

3.5 Evaluation

Once the classification models are trained, their performance will be evaluated on a separate set of unseen test images to assess their generalization capability. The evaluation will involve several standard metrics commonly used in medical image classification tasks to ensure a comprehensive assessment of the system's accuracy and reliability.

Planned Evaluation Metrics:

- **Accuracy:** Measures the overall proportion of correctly classified images among all test samples.
- **Precision, Recall, and F1-Score:**
    - *Precision* indicates the proportion of true positive leukemia detections among all positive predictions, reflecting the model's exactness.
    - *Recall* (or sensitivity) measures the model's ability to correctly identify all actual leukemia cases.
    - *F1-Score* provides a harmonic mean of precision and recall, offering a balanced measure of performance.
- **Confusion Matrix:** A detailed tabular representation of true positives, true negatives, false positives, and false negatives to better understand classification errors.
- **ROC-AUC Curve (Receiver Operating Characteristic - Area Under Curve):**

  Assesses the model's ability to discriminate between classes across different classification thresholds, providing insight into trade-offs between sensitivity and specificity.

These metrics will guide the selection of the most effective classifier and inform further refinements to the methodology.

3.6 Deployment

The final phase of the project will focus on integrating the complete leukemia detection pipeline—including CNN-based feature extraction and machine learning classification into an interactive web application. This application will facilitate real-time diagnosis and make the system accessible to medical professionals.

Planned Deployment Details:

- The entire detection system will be embedded into a web interface developed using **Streamlit**, which offers a simple yet powerful platform for building interactive machine learning applications.

- Users will be able to upload peripheral blood smear images via the web interface for analysis.

- Upon submission, the system will process the uploaded image through the trained pipeline and deliver leukemia detection predictions promptly.

- To enhance interpretability and user trust, model explanation techniques such as **LIME** or **SHAP** will be incorporated. These tools will help visualize the key features or regions in the image that influenced the prediction, providing transparent insights.

- The application will be designed for ease of use and accessibility across different devices, enabling hematologists and researchers to quickly obtain diagnostic support.

This deployment step aims to transform the research methodology into a practical, user-friendly tool for supporting early leukemia detection in clinical settings.

Chapter 4

Conclusion

This survey highlights the growing effectiveness of machine learning and deep learning techniques in the early detection of leukemia, particularly through the analysis of peripheral blood smear images. While traditional models like SVM and Random Forest provide decent results using handcrafted features, deep learning approaches—especially CNNs and transfer learning—offer higher accuracy and better automation.

Preprocessing plays a vital role in improving detection performance. Studies show that converting images to color spaces like HSV and HSI enhances segmentation, supporting this project's aim to explore HSL for anomaly detection. The HSL model may provide clearer separation of leukemic cells, aiding accurate classification.

Though this project is currently limited to a literature survey, the findings offer a strong foundation for future implementation. Integrating HSL-based preprocessing with machine learning and deep learning can potentially lead to a fast, reliable, and automated leukemia detection system, reducing manual errors and assisting early diagnosis in clinical practice.

References

[1] F. M. Talaat and S. A. Gamel, "Machine learning in detection and classification of leukemia using C-NMC_Leukemia," Multimedia Tools and Applications, 2023.

[2] A. P. Patil and M. Hiremath, "Machine Learning Model to Detect Chronic Leukemia in Microscopic Blood Smear Images," IEEE InC4, 2023.

[3] S. Perveen et al., "A Framework for Early Detection of Acute Lymphoblastic Leukemia and Its Subtypes Using Deep Ensemble Learning," IEEE Access, 2024.

[4] R. Sudheer et al., "Leukemia: A Comparative Analysis of Deep Learning Models Using ALL Dataset," IEEE RAICS, 2024.

[5] S. Karthikeyan et al., "Machine Learning Based Algorithmic Approach for Detection and Classification of Leukemia," IEEE ICSTSN, 2023.

[6] Y. Aditya, M. Durga, K. V. S. N. S. Teja, J. H. Priya, and K. E. Datta, "An Improved Leukemia Detection Model using CNN," in Proc. Int. Conf. Innovative Data Commun. Technol. Appl. (ICIDCA), 2023.

[7] A. Abhishek, N. Santhanam, R. K. Jha, R. Sinha, and K. Jha, "Multi-Class Classification of Acute Leukemia using Transfer Learning," in Proc. Int. Conf. Advancement in Technol. (ICONAT), Goa, India, Jan. 2022.

[8] M. Iswarya, A. Shivakami, R. A. Mira, and A. Karpagam,"Detection of Leukemia using Machine Learning," in Proc. Int. Conf. Applied Artif. Intell. and Compute. (ICAAIC),2022.

[9] R. Sigit, M. M. Bachtiar, and M. I. Fikri, "Identification Of Leukemia Diseases Based On Microscopic Human Blood Cells Using Image Processing," in Proc. IEEE Conf., 2018.

[10] N. H. Harun et al., "Comparison of Acute Leukemia Image Segmentation using HSI and RGB Color Space," in Proc. 10th Int. Conf. Inf. Sci., Signal Process. And their Appl. (ISSPA), 2010.

[11] R. D. Labati, V. Piuri, and F. Scotti, "ALL-IDB: The Acute Lymphoblastic Leukemia Image Database for Image Processing," in Proc. IEEE Int. Conf. Image Process. (ICIP), 2011.

[12] S. Shafique and S. Tehsin, "Acute Lymphoblastic Leukemia Detection and Classification of Its Subtypes Using Pretrained Deep Convolutional Neural Networks," Technol. Cancer Res. Treat., vol. 17, pp. 1–7, 2018.

[13] T. Dharani and S. Hariprasath, "Diagnosis of Leukemia and its types Using Digital Image Processing Techniques," in Proc. Int. Conf. Commun. and Electron. Syst. (ICCES), 2018.

[14] S. Khobragade, D. D. Mor, and C. Y. Patil, "Detection of Leukemia in Microscopic White Blood Cell Images," in Proc. Int. Conf. Inf. Process. (ICIP), Pune, India, 2015.

[15] S. Ansari, A. H. Navin, A. B. Sangar, J. V. Gharamaleki, and S. Danishvar, "A Customized Efficient Deep Learning Model for the Diagnosis of Acute Leukemia Cells Based on Lymphocyte and Monocyte Images," Electronics, vol. 12, no. 2, p. 322, Jan. 2023.