# CHAPTER 1

## INTRODUCTION

Acute Lymphoblastic Leukemia (ALL) is a rapidly growing blood cancer that mostly affects children and young adults. It occurs when immature white blood cells multiply uncontrollably in the bone marrow, reducing the body's ability to produce healthy blood cells. This leads to symptoms such as tiredness, fever and frequent infections. Early diagnosis is important for effective treatment.

Doctors usually diagnose ALL by examining blood smear images under a microscope. This process is slow, depends on expert experience and can lead to errors due to fatigue or subjective judgement.

Artificial intelligence, especially deep learning, is now helping improve accuracy in medical image analysis. Convolutional Neural Networks (CNNs) can automatically learn important patterns from images. Since medical datasets often lack detailed labels, Multiple Instance Learning (MIL) is used to train models using image groups instead of individual cells. When combined with attention mechanisms and Grad-CAM, the system becomes more understandable, allowing doctors to see which parts of the image influenced the prediction.

This automated approach reduces the workload on medical professionals. It provides faster and more consistent results than manual examination. AI models help identify early signs of leukemia that may be missed by the human eye. The system can analyze multiple images quickly and accurately. MIL makes training easier by using slide-level labels instead of cell-level labels. Attention maps highlight the most important patches in the image. Grad-CAM heatmaps increase trust by showing how the model makes decisions. Overall, AI-based screening supports doctors and improves the chances of early diagnosis.

## 1.1 Background

Leukemia diagnosis relies on blood tests, microscopic image analysis and bone marrow examination. Earlier computer-aided methods used shape-based or texture-based handcrafted features, but these struggled with variations in images.
Deep learning improved accuracy by learning visual patterns automatically. Techniques like MIL reduce the need for manual labeling. Attention mechanisms highlight important cell regions, and Grad-CAM provides heatmaps that explain model decisions.

## 1.2 Need for Automated Screening

**Automated screening is important because:**

- Manual diagnosis is time-consuming.
- Expert availability is limited.
- Interpretations can vary between doctors.
- Large numbers of patients require fast processing.
- Early detection improves survival chances.

AI-based systems offer quick, consistent and objective results that help doctors make better decisions.

## 1.3 Problem Statement

Manual detection of ALL is slow, subjective and dependent on expert availability. Limited labeled datasets and the black-box nature of deep learning create additional challenges. The project aims to develop an accurate, reliable and interpretable automated system using deep learning, Multiple Instance Learning and Grad-CAM explanations.

## 1.4 Scope of the Project

The system:

1.4.1 Accepts blood smear images as input

1.4.2 Uses MIL-based classification

1.4.3 Generates attention maps and Grad-CAM heatmaps

1.4.4 Provides predictions and visual explanations

1.4.5 Offers a simple user interface

The project supports clinicians by improving speed and reliability of screening.

## 1.5 Organization of the Report

This report is structured into several chapters to clearly present the development and analysis of the project. The introduction outlines the background, need, scope and challenges associated with leukemia diagnosis. The literature review examines existing approaches, highlighting gaps and motivating the proposed solution. The objectives of the study define the specific milestones achieved through the project.

Subsequent chapters detail the methodology, system architecture and implementation strategies. The results and analysis chapter presents the model performance, interpretability outputs and technical evaluation metrics. The conclusion summarizes the achievements and outlines possible future improvements. The appendix includes supplementary materials, including sample images, data descriptions and code explanations.

# CHAPTER 2

# SYSTEM ANALYSIS

System analysis involves understanding the complete workflow, requirements and operational behavior of the proposed leukemia detection system. It examines the existing challenges in manual leukemia screening, including variability in staining, limited expert availability and subjective interpretation. The analysis also evaluates current automated methods to identify gaps in accuracy, scalability and interpretability. By studying data characteristics, model requirements and clinical constraints, the system analysis defines a structured foundation for designing an efficient diagnostic solution. This ensures that the final system is robust, clinically aligned and capable of delivering reliable predictions.

## 2.1 Literature Survey

Research shows that early leukemia detection relied on handcrafted image features like shape, color and texture. Machine learning techniques such as SVMs and Random Forests improved accuracy but still struggled with variations in samples.

Deep learning, especially CNNs, significantly improved performance by learning features directly from images. Models like VGG, ResNet and Inception showed strong results in medical imaging. However, the lack of well-annotated datasets remains a limitation.

Multiple Instance Learning helped solve this issue by allowing training on groups of images. Attention mechanisms further improved interpretability. Grad-CAM techniques provide heatmaps that help understand the model's focus areas.

## 2.2 Findings of the Analysis

The analysis shows that:

- Traditional image processing methods are not robust.
- Classical machine learning depends heavily on manual features.
- Deep learning improves accuracy but needs better interpretability.
- MIL reduces dependency on detailed annotations.
- Grad-CAM improves trust by showing visual explanations.

These findings justify the need for an automated, interpretable and detection system.

## 2.3 System Requirement Specification

The development of an automated Acute Lymphoblastic Leukemia detection and prognosis system requires careful consideration of both hardware and software requirements. Since the system involves deep learning model training, image preprocessing, visualization modules and a graphical interface, the specifications must support computationally intensive processes. The system must also ensure smooth execution during inference, where predictions must be generated efficiently and accurately for clinical applicability.

The design includes components for data preprocessing, CNN-based feature extraction, MIL- based classification, attention visualization and Grad-CAM heatmap generation. In addition, the system incorporates a user-friendly interface that enables clinicians or laboratory personnel to upload images, view predictions and analyze visual explanations.

### 2.2.1 Software Requirements

Table 2.1:Software Requirements

| **Component** | **Description** |
|---|---|
| Operating System | Windows / Linux / macOS |
| Programming Language | Python |
| Deep Learning Libraries | PyTorch or TensorFlow |
| Image Processing | OpenCV, NumPy, Pandas |
| Visualization | Matplotlib |
| User Interface | Streamlit |
| IDE | Jupyter Notebook / VS Code |

### 2.2.2 Hardware Requirements

Table 2.2:Hardware Requirements

| **Component** | **Minimum Requirement** |
|---|---|
| Processor | Intel i5 or above |
| RAM | 16 GB |
| GPU | NVIDIA GTX 1660 / RTX 3060 or higher |
| Storage | 512 GB SSD |
| Inference System | Any mid-range laptop |

# CHAPTER 3

# SYSTEM DESIGN

This chapter presents the design of the proposed automated Acute Lymphoblastic Leukemia (ALL) screening and prognosis system. It explains the dataset used, preprocessing steps, system architecture, Multiple Instance Learning (MIL) approach, attention mechanism, workflow, user interface, and visualization modules. Standardized preprocessing is applied to correct variations in staining, illumination, and resolution, making the model reliable across different clinical environments.

The architecture is structured to mimic the diagnostic process followed by hematologists, enabling the system to capture subtle morphological features related to malignancy. MIL reduces the need for detailed cell-level annotations, making the model scalable for large datasets. Attention mechanisms further enhance interpretability by highlighting image regions that influence the classification. An explainability module using Grad-CAM ensures transparency in decision-making.

Finally, a user-friendly interface and automated visual reports help clinicians, researchers, and students easily interact with the system, making the design suitable for both academic and clinical applications.

## 3.1 Data Understanding & Preprocessing

The dataset consists of peripheral blood smear images labeled at slide/bag level as ALL or NORMAL. Images vary by staining, illumination and magnification. Preprocessing steps standardize images: resizing to a fixed dimension, color/stain normalization to reduce inter-lab variability, denoising when necessary, and pixel normalization. Patches are extracted from slides to form instances for MIL training. Patch extraction preserves morphological detail required for discriminating lymphoblasts from normal lymphocytes.

After standardization, images are segmented into smaller patches that represent individual or grouped cell regions. These patches form the basis for MIL training. Finally, pixel values are normalized to stabilize the learning process and improve convergence during CNN-based feature extraction.

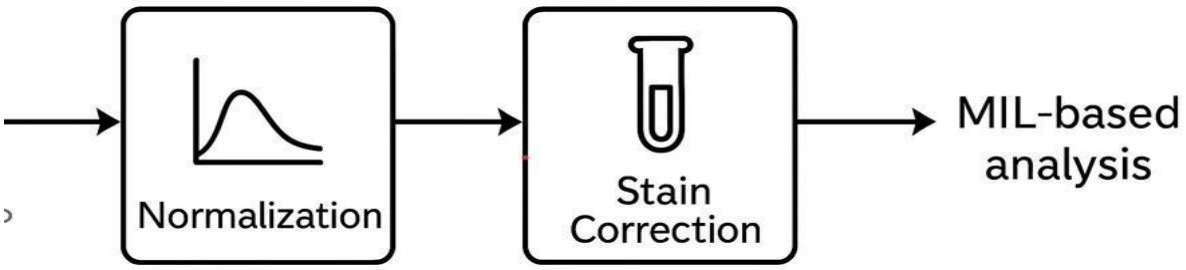


Figure 3.1:Cell Image Preprocessing Workflow

***Description:*** The figure illustrates the preprocessing pipeline: input raw smear image → resize to standard dimensions → stain/color normalization → noise reduction → segmentation
/ patch extraction → normalization → patch-ready tensors for CNN. This workflow ensures consistent model inputs and improves convergence during feature learning.

## 3.2 System Architecture

The architecture comprises these layers: data acquisition, preprocessing, CNN backbone for patch-level feature extraction, MIL attention module for bag aggregation, classifier for bag-level prediction, Grad-CAM explainability module and Streamlit UI for interaction. The attention module computes instance weights that highlight diagnostically significant patches.

## 3.3 Functional Design

At runtime, users upload single images or folders. Preprocessing extracts patches which the CNN encodes into feature vectors. The MIL attention mechanism assigns relevance scores and aggregates features to a bag-level representation that the classifier uses to predict the slide label. Grad-CAM overlays and attention heatmaps provide visual explanations alongside prediction confidence.

## 3.4 User Interface & Reporting

The Streamlit UI supports batch uploads and displays: bag prediction, image-wise attention scores, Grad-CAM heatmaps and downloadable PDF/HTML reports summarizing predictions and metrics.

## 3.5 System Architecture (Layered)

The system architecture follows a layered structure where each layer performs a specific function while supporting the next. The process begins with the data acquisition layer, which receives microscopic blood smear images from the user. These images are then passed to the preprocessing layer, where they are resized, normalized, and converted into tensors to ensure input consistency for the neural network.

At the core of the system is the Multiple Instance Learning (MIL) module. Instead of using single-image labels, MIL processes groups of image patches as bags, enabling the model to evaluate multiple cell regions collectively. Features extracted by the convolutional backbone are weighted using an attention mechanism, which identifies the most informative instances for distinguishing leukemic from normal cells.

The classifier layer outputs a binary prediction indicating whether the sample corresponds to ALL. In parallel, explainability modules such as Grad-CAM generate visual heatmaps that highlight critical regions. Finally, the user interface layer, built using Streamlit, displays both the prediction and the interpretability maps in a clear and clinically meaningful format. This architecture provides a complete end-to-end diagnostic pipeline.

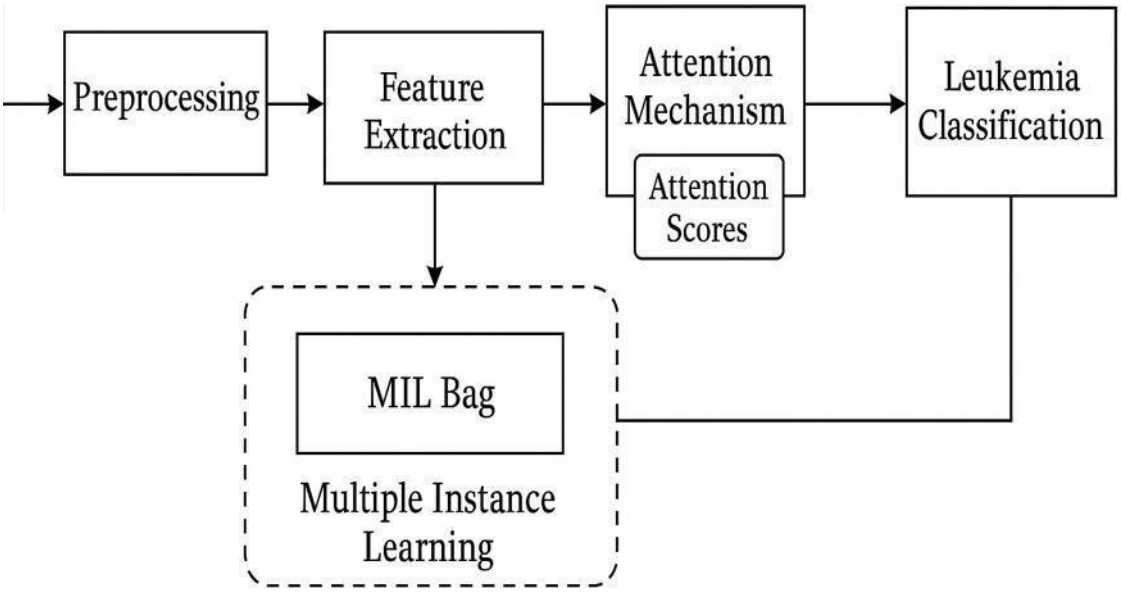


Figure 3.2:System Architecture (MIL + Attention Model)

***Description:*** The diagram depicts the end-to-end flow from slide input to prediction and visualization. Patches from each slide are passed through the CNN backbone. Feature vectors flow into the attention-based MIL aggregator which computes bag-level representation. The classifier outputs the prediction and the Grad-CAM module provides heatmaps for interpretability

## 3.6 Data Flow Diagram — Level 0

Level 0 provides an overview: User → Preprocessing → Model → Results (prediction + confidence + heatmaps).

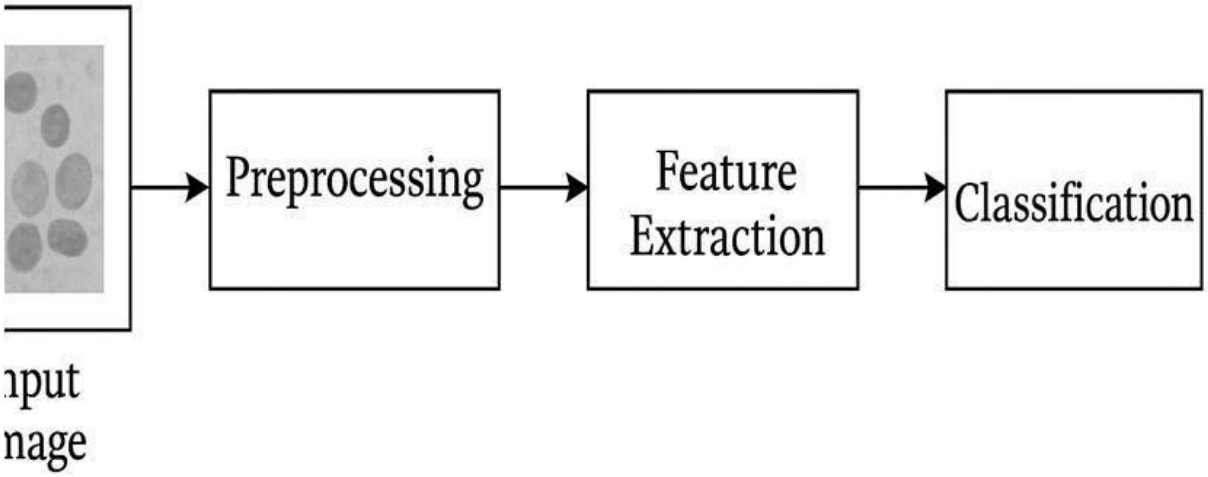


Figure 3.3: Level 0 Data Flow Diagram (Overall Pipeline)

***Description:*** High-level pipeline showing user upload, preprocessing, model inference and UI output. It summarizes the major data transitions without module internals.

## 3.7 Data Flow Diagram — Level 1

Level 1 expands internal operations: preprocessing substeps, CNN extraction, MIL attention scoring, Grad-CAM generation and report export.

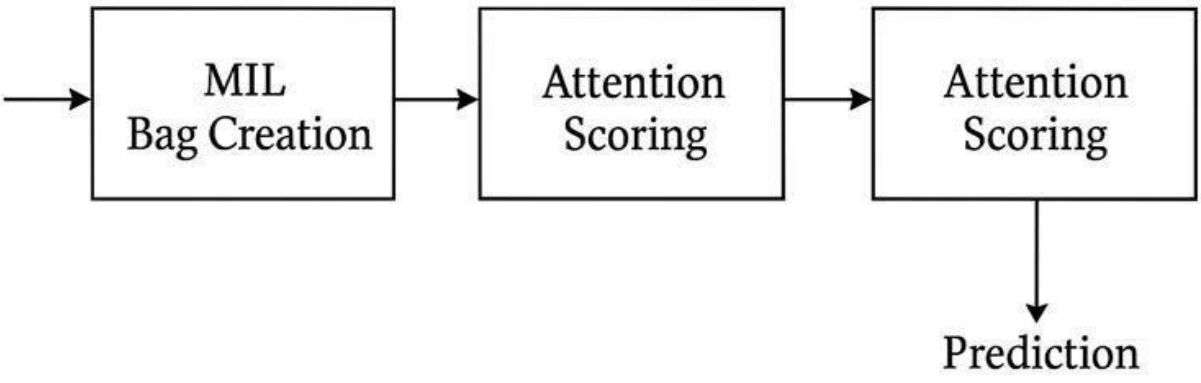


Figure 3.4:Level 1 Data Flow Diagram (Image Classification Process)

***Description:*** Detailed flowchart that shows how patches are created, features extracted, attention weights computed and Grad-CAM maps produced for each patch prior to aggregation and UI display

## 3.8 UML Use Case & Class Descriptions

### Use Case Description

- The primary actor is the **medical professional or technician**.
- The actor uploads microscopic blood smear images into the system.
- The actor requests **predictions** from the diagnostic model.
- The actor views **explainable outputs** such as Grad-CAM heatmaps.
- The system processes the image, generates predictions, and returns interpretable results.

### Core Class Descriptions

- **DataLoader**
    - Handles loading of microscopic images.
    - Organizes batches and ensures compatibility with the model.
- **Preprocessing**
    - Performs resizing, normalization, and tensor conversion.
    - Ensures consistent input format for the CNN and MIL model.
- **MILModel**
    - Implements Multiple Instance Learning architecture.
    - Processes image patches (bags) and generates feature representations.
    - Integrates the attention mechanism to compute instance importance.
- **GradCAM**
    - Generates class-discriminative heatmaps.
    - Highlights nuclear and chromatin regions contributing to predictions.
- **StreamlitApp**
    - Manages user interaction with the system.
    - Displays predictions, heatmaps, and relevant interpretation results.

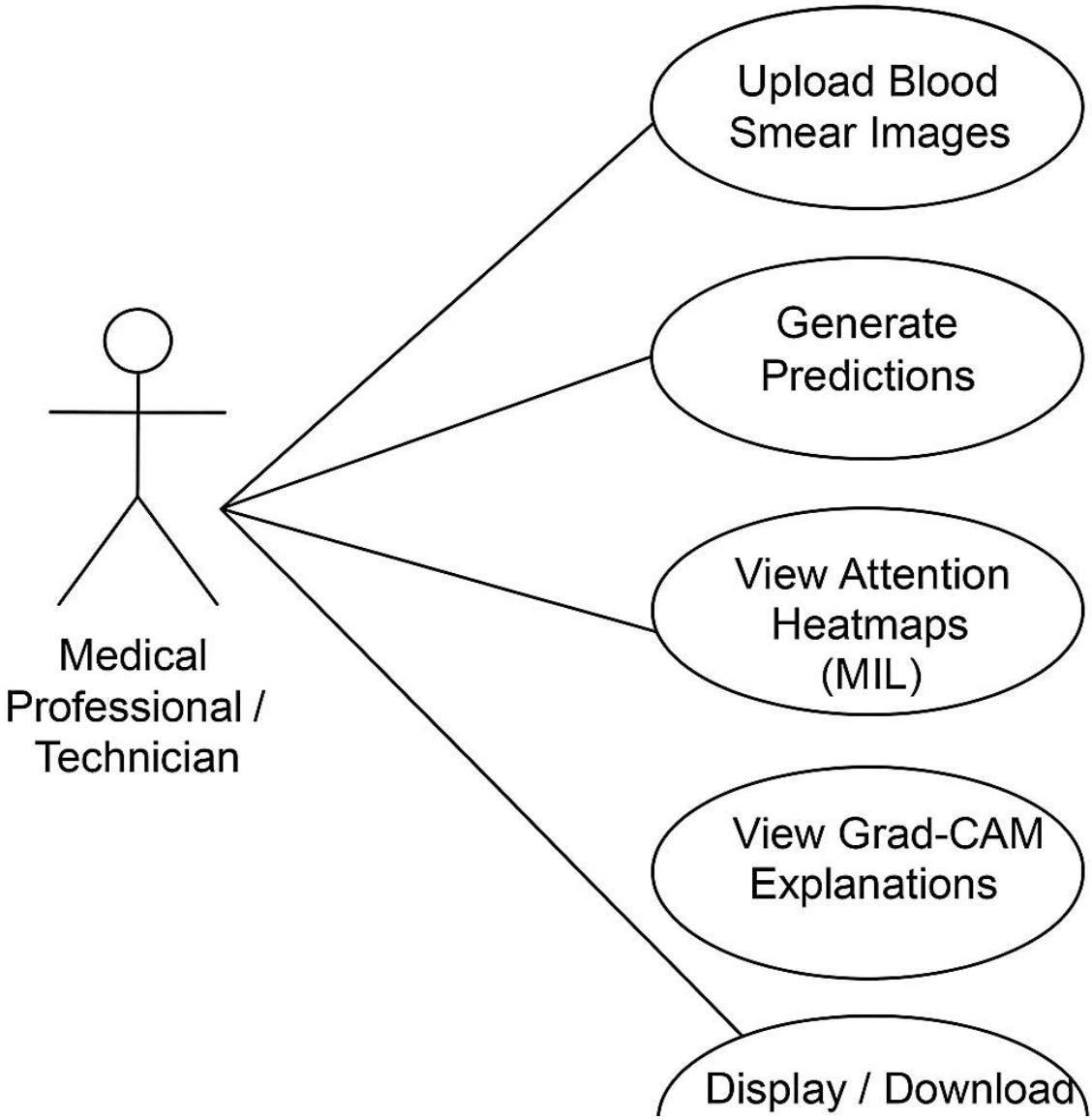


**Figure 3.5: UML Use Case Diagram Description**

*This diagram illustrates the interaction between the medical professional and the automated leukemia screening system. The actor uploads blood smear images, requests predictions, and views explainable outputs such as attention heatmaps and Grad-CAM visualizations. The system processes the images using the AI model and returns interpretable diagnostic results to support clinical decision-making.*

# CHAPTER 4

# SYSTEM IMPLEMENTATION

The implementation of the automated Acute Lymphoblastic Leukemia (ALL) screening and prognosis system involves translating the system design into a functional software solution capable of processing microscopic blood smear images, training deep learning models, generating predictions and producing interpretability visualizations. This chapter explains how each component of the architecture was implemented, including preprocessing modules, CNN- based feature extraction, the Multiple Instance Learning framework, attention mechanism integration and Grad-CAM visualization generation. The implementation also covers the Streamlit-based interface that enables real-time interaction with the system. The objective of this chapter is to describe how the theoretical design is transformed into executable modules, detailing the workflow from input acquisition to final diagnostic output.

## 4.1 Modules Implemented

The system includes:

- Preprocessing module
- CNN backbone
- MIL + attention module
- Grad-CAM visualization
- Streamlit interface
- Model inference module

Major modules: preprocessing (resizing, normalization, patching), CNN backbone (pretrained ImageNet nets fine-tuned), MIL module with attention, Grad-CAM visualization module, Streamlit front-end, and the main inference/training scripts connecting modules.

## 4.2 Model Training and Evaluation

The model is trained with:

- Bag-level datasets
- Cross-entropy loss
- Accuracy and loss monitoring
- Validation during training

## 4.3 Grad-CAM Visualization Output

Grad-CAM computes gradients of the target class flowing into the last convolutional layer, combines them with activation maps and upsamples the result to the original image size. The overlay highlights regions that most influenced the model's prediction.

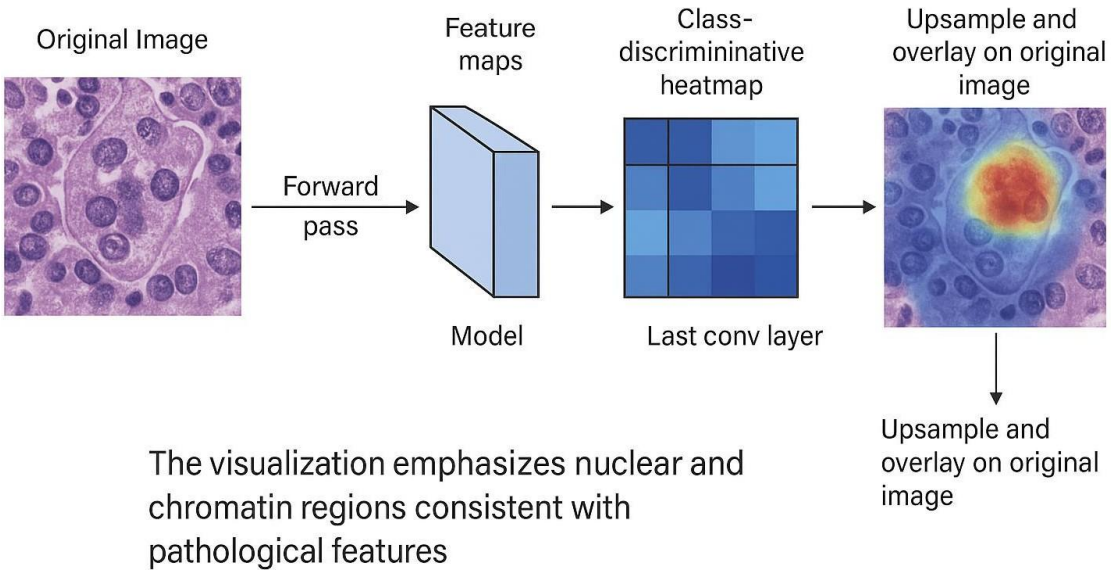


Figure 4.3 — Grad-CAM Heatmap Generation

***Description:*** The figure shows the Grad-CAM pipeline: forward pass → gradient extraction at last conv layer → weighted combination of feature maps → production of class-discriminative heatmap → upsample and overlay on original image. The visualization emphasizes nuclear and chromatin regions consistent with pathological feature

# CHAPTER 5

# TESTING AND EVALUATION

Testing and evaluation form one of the most critical stages of this project because they determine the reliability, accuracy and robustness of the automated screening and prognosis system for Acute Lymphoblastic Leukemia. The entire system, ranging from data preprocessing, model training, attention-based feature extraction and Grad-CAM visualization, is rigorously tested to ensure that it performs consistently on unseen microscopic blood smear images. The testing process aims to validate whether the system can generalize effectively, produce clinically meaningful predictions and maintain high diagnostic quality under various conditions. Evaluation metrics such as accuracy, precision, recall, F1-score and ROC-AUC are used to measure the performance of the deep learning model. In addition, visual interpretability is assessed through Grad-CAM overlays and attention heatmaps to confirm that the model focuses on biologically relevant cellular regions during prediction. This chapter describes the detailed procedures used to test the system and presents the results obtained from extensive experimentation.

## 5.1 Training and Validation Strategy

Data was split into:

- Training
- Validation
- Testing

MIL structure was preserved during splitting.

## 5.2 $R^2$ Score, Loss and Confusion Matrix

Metrics used:

- Accuracy
- Loss curves
- $R^2$ score
- Confusion matrix

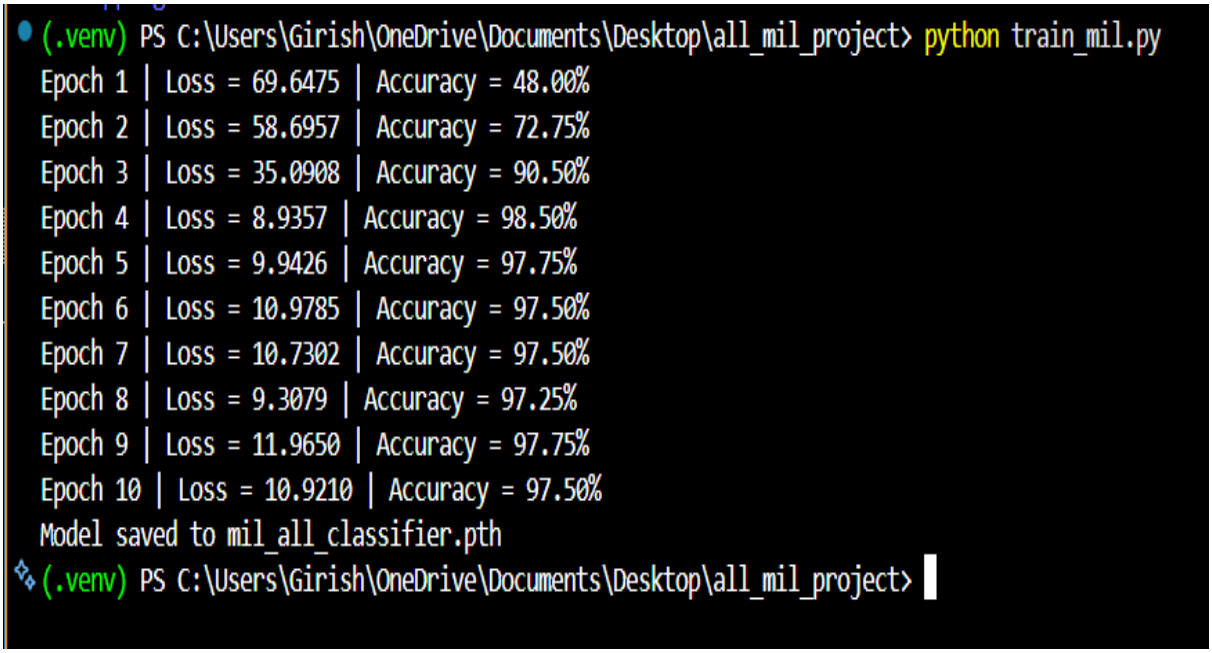


Figure 7.2 Model Accuracy and Loss

***Description:*** The figure shows training and validation accuracy curves and loss curves across epochs. The plot demonstrates steady training improvement and validation stability, indicating reliable convergence.

## 5.3 Visual Interpretation using Grad-CAM & Attention Maps

Grad-CAM heatmaps are generated on test images. In ALL-positive samples, heatmaps typically light up regions consistent with blast cell nuclei, increased nucleus-to-cytoplasm ratio and irregular chromatin. Attention maps show which individual patches in a bag contributed most to the classification.

## 5.4 System Testing & Real-world Validation

Testing included images with varied staining and magnification, stress tests with large files and validation on external datasets not used during training. Stability testing confirmed consistent model outputs across repeated runs.

## 5.5 Test Cases & Results Summary

Test cases covered preprocessing correctness, MIL bag formation, classifier outputs and UI error handling for corrupted files. Results showed consistent classification performance, meaningful heatmaps and robust error handling.

## 5.6 Final Outcome

The system demonstrates that a hybrid approach combining CNN-based feature extraction, attention-based MIL and Grad-CAM explainability can deliver accurate and interpretable ALL screening on peripheral blood smears. The Streamlit UI provides an accessible interface for clinicians to upload slides and view model explanations.

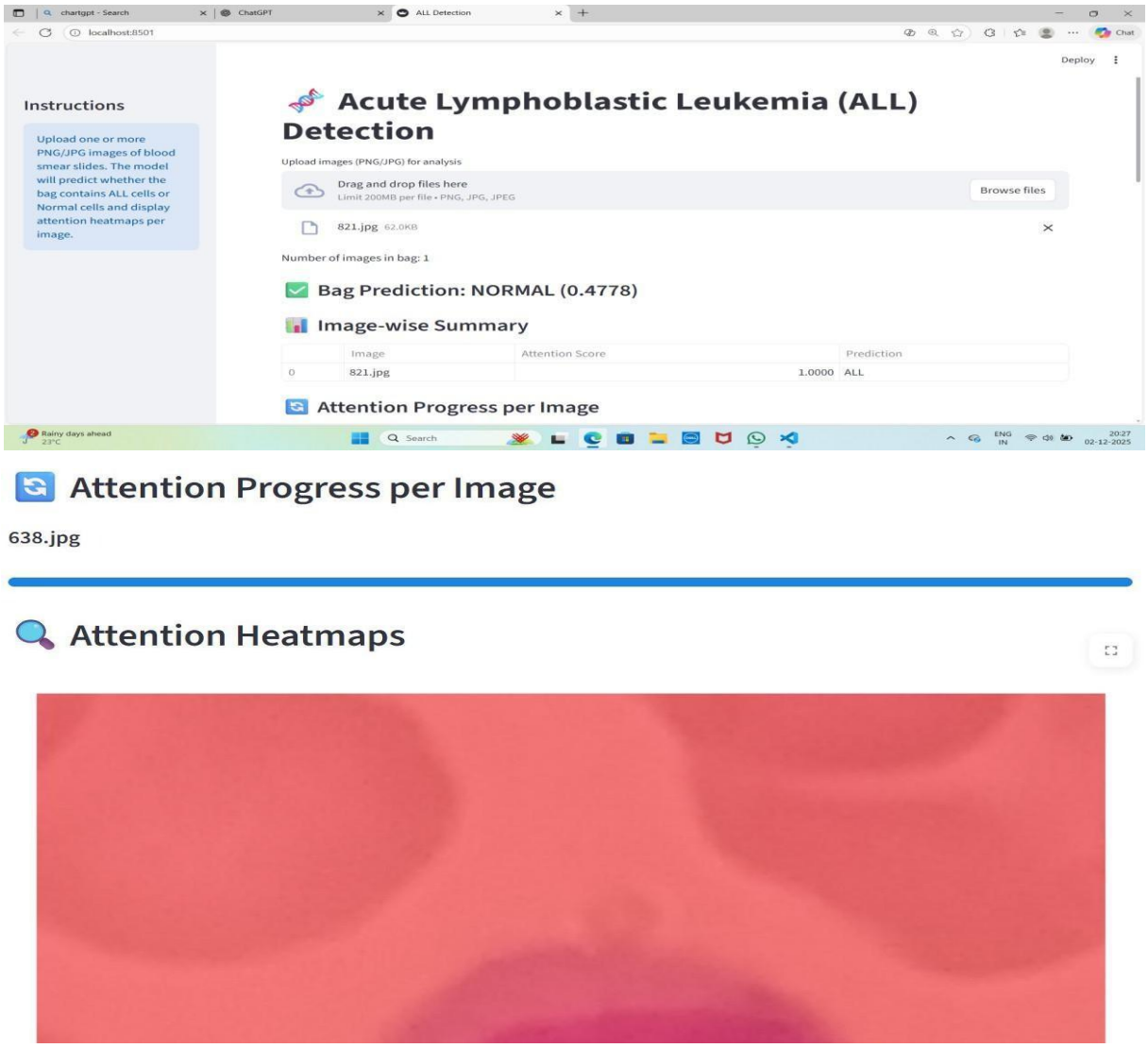


Figure 8.1 Final Output

***Description:*** The figure shows the Streamlit application displaying bag-level prediction (e.g., NORMAL with a probability), image-wise attention progress bar, attention heatmaps and a Grad-CAM overlay for a selected image. The screenshot demonstrates user-facing outputs that combine prediction, confidence and visual explanations in a single view

# CHAPTER 6

# CONCLUSION AND FUTURE WORKS

The project titled **"Automated Screening and Prognosis of Acute Lymphoblastic Leukemia"** presents a comprehensive approach to applying deep learning, Multiple Instance Learning and Grad-CAM–based interpretability techniques for early diagnosis of leukemia from microscopic blood smear images. The system addresses critical challenges faced in medical diagnostics, such as subjectivity, delays in interpretation and the need for expert pathologists. The integration of machine learning into the diagnostic workflow enhances the speed and accuracy of disease detection and provides decision support that is both reliable and clinically interpretable. This chapter summarizes the key findings, contributions and overall outcomes of the work while highlighting potential areas for future research and improvement

In this project, an automated model was developed to identify Acute Lymphoblastic Leukemia (ALL) from peripheral blood smear images. The system uses MIL to analyze cell patches, attention mechanisms to highlight important regions, and Grad-CAM to generate visual explanations. These features make the diagnostic process more transparent and help medical professionals understand why the model arrived at a particular decision. The model reduces manual effort, supports early detection, and enhances the reliability of the diagnostic process. Overall, the project demonstrates that AI-based systems can effectively assist clinicians and improve diagnostic quality in hematology.

## 6.2 Future Works

Although the proposed system performs well, several improvements can be made in future research:

1. **Larger and more diverse datasets:**
   Training the model on bigger and more varied medical datasets can further increase accuracy and robustness.
2. **Multi-class leukemia classification:**
   The system can be extended to identify different leukemia types, not only ALL.
3. **Whole slide image (WSI) analysis:**
   Future models can process entire slide images automatically instead of smaller patches.

# REFERENCE

[1] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," **IEEE International Conference on Computer Vision (ICCV)**, pp. 618–626, 2017. Available: https://arxiv.org/abs/1610.02391

[2] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh and D. Batra, "Grad-CAM: Gradient-weighted Class Activation Mapping," **International Journal of Computer Vision**, vol. 128, pp. 336–359, 2020. Available: https://link.springer.com/article/10.1007/s11263-019-01228-7

[3] S. S. Shafique and S. Tehsin, "Acute Lymphoblastic Leukemia Detection and Classification Using Deep Learning," **Biomedical Research International**, vol. 2018, pp. 1–12, 2018. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC6161200/

[4] P. Wang, X. Hu, Y. Li, "Machine Learning Techniques for Leukemia Detection: A Comprehensive Survey," **Cognitive Computation**, vol. 14, no. 3, pp. 789–806, 2022.

[5] M. Matek, S. Schwarz, E. Spiekermann and C. Marr, "Human-level recognition of blast cells in acute myeloid leukemia using deep neural networks," **Nature Machine Intelligence**, vol. 1, pp. 538–544, 2019.

[6] A. M. Sarraf and C. Pul, "Deep Convolutional Neural Network for Acute Lymphoblastic Leukemia Classification," **arXiv Preprint**, 2021. Available: https://arxiv.org/abs/2105.03995

[7] T. Dietterich, "Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms," **Neural Computation**, vol. 10, no. 7, pp. 1895–1923, 1998.

[8] M. T. Ribeiro, S. Singh and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," **Proceedings of the 22nd ACM SIGKDD**, pp. 1135–1144, 2016.