Assignment # 1
(Problem Based Learning)
(CLO4 -> PLO5)
Digital Image Processing
Hematological Image Segmentation using Connected Component Labeling
Submission Deadline: 16 Feb 2025

Note: Students should score 50% in OBE specific questions to ensure their accumulated scores towards respective PLOs are above 50%

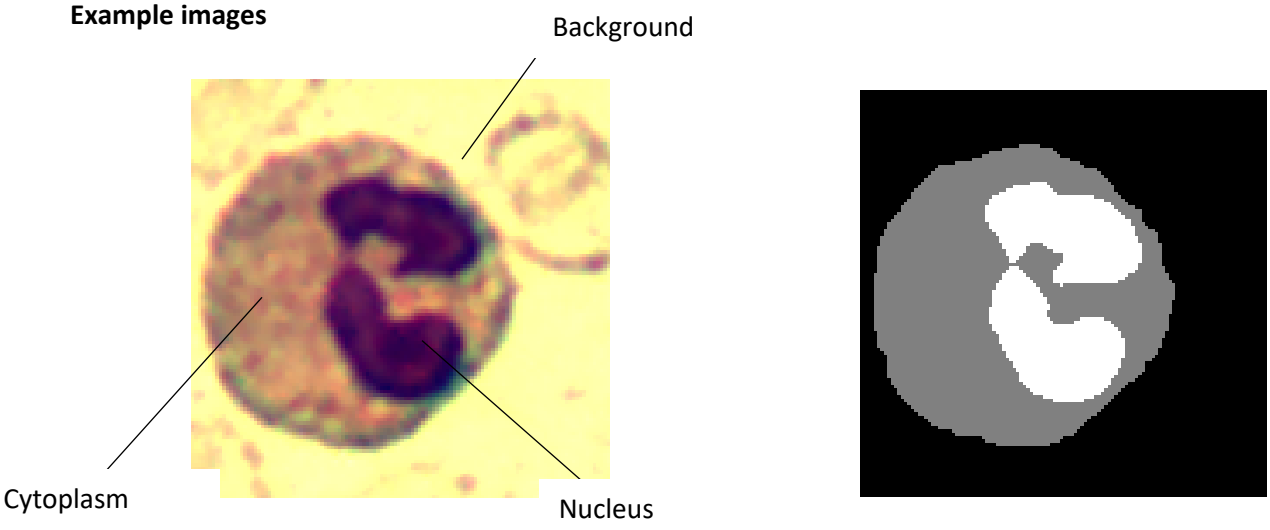
Introduction

Blood smear films are thin layers of blood spread on a microscope slide and stained to allow microscopic examination of blood cells. They are crucial in diagnosing various hematological disorders by analyzing the morphology and count of blood cells. White blood cell (WBC) disorders include leukemias, lymphomas, and conditions such as neutropenia and leukocytosis, which indicate infections, immune system abnormalities, or malignancies. Accurate segmentation of blood smear images aids in the diagnosis and classification of these disorders.

Automated segmentation of hematological images is essential for blood cell analysis and disease diagnosis. In this assignment, we will implement **Connected Component Labeling (CCL)** for segmenting different components of blood cell images. The dataset consists of microscopic images of blood smears with corresponding masks that classify each pixel into different categories:

- **Nucleus (White Mask Region)**
- **Cytoplasm (Gray Mask Region)**
- **Background (Black Mask Region)**

Example images



Using this dataset, we will develop a segmentation pipeline that applies pre-processing techniques, connected component labeling, and post-processing refinement.

Dataset Details

The dataset consists of paired microscopic images and manually annotated ground truth masks. Each image has a corresponding labeled mask where:

- **White region** represents the nucleus of the WBC.
- **Gray region** represents the cytoplasm of the WBC.
- **Black region** represents the background.

Each image is labeled pixel-wise, enabling precise segmentation.

Dataset Link:
https://drive.google.com/drive/folders/1DUDnYXZQF6zSZDl0RJIsdo8lqiZOJQoV?usp=sharing

Tasks

1. **Define a V set**:

   In first phase, Separate white blood cell (WBC) from the background using 8-connectivity based Connected Component Analysis (CCA). For this intelligently design V set using images from train folder and see how good they are working. This should give you 2 labels i.e. one for background and other for complete WBC

2. **Differentiate WBC components**:

   Define new groups of V sets again from train data to further divide the cell into nucleus and cytoplasm. Now you should have 3 labels in total i.e. background, nucleus and cytoplasm.

3. **Compute Dice Coefficient**:

   Once you are done with designing complete pipeline, use data from test folder to compute Dice Coefficient for each class i.e. background, nucleus and cytoplasm.

   Count the number of **True** and **False** pixels.

   True pixels: Where your segmented labels overlap with the ground truth masks.

   False pixels: Where they do not overlap.

   Normalize by dividing with the total number of true pixels given in the original mask.

Submission Requirements

You need to submit a report (Word or PDF) against this assignment containing:

1. **Explanation of your approach**:

   Clearly define the **V set** used.

Outline the steps followed for all tasks.

Include a **flow diagram** for better explanation. You can complete CCA with other logic to improve your results but you cannot utilize any image processing technique which we haven't studied so far.

2. **Complete editable code**:

   Copy the entire code in an editable format within the Word document.

3. **Results and Comparisons**:

   Add some sample test images along with:

   Their **original microscopic images**.

   Ground truth masks (provided dataset labels).

   Your segmented outputs based on the implemented algorithm.

   Present results in a **tabular format** against different experiments which you have conducted.