

DS-3002 Data Mining Spring 2022

Project Statement and Guidelines

Due Date: 11:59 PM 1st June 2022

Must read instructions before starting

- It is **not** a group project
- You **cannot** use any advance and pre-trained model in your project (Like VGG-16, YOLO, and so on)
- Zero marks will be awarded to the students involved in plagiarism and late submission.
- All the submissions will be done on google classroom
- The name of the python file should be like this **19iXXXX.ipynb** and the name of the word report should be **19iXXXX_ProjectReport.docx**
- You must Zip your python file and report (in the word form) before zip
- Naming convention must be followed strictly

Segmentation and Classification of Diabetic Retinopathy

GOAL:

- Image sampling using k-mean clustering
- Image segmentation
- Image classification using ML, ANN, and CNN models

Description About the Dataset:

You are provided 500 retina images taken using fundus photography under a variety of imaging conditions. A clinician has rated each image for the severity of diabetic retinopathy on a scale of 0 to 4:

0 - No DR

1 - Mild

2 - Moderate

3 - Severe

4 - Proliferative DR

And a .csv file contains the label information against each image.

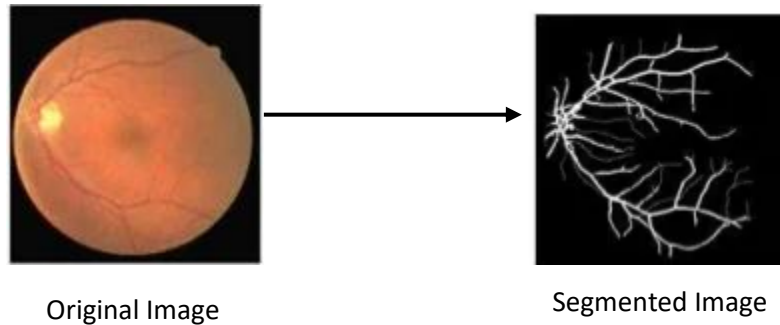
Step 01: Sampling

Your 1st task is to select the sample from your dataset. For this purpose, you are required to cluster all your images. This will be done using k-mean clustering with an appropriate number of k. To find the optimal number of k, you may use the elbow method or any other possible approach (as per choice). After clustering your images, select 70% of images against each cluster as your final dataset.

HINT: Save the label of selected images in this step for further use.

Step 02: Segmentation

After the above step, Convert your images into segments (for better classification). You can use K mean clustering to find the segments in each image and save the segmented images for the next step.



Step 03: Image classification and analysis

Classify your images before the segmentation and after segmentation to see the result. Split your data into training and testing. Apply any two ML models (from the lectures), a simple Neural Network (using fully connected layers), and a CNN model to classify the segmented and original images from the training sample.

Report the accuracy (for ML model), and error rate (for ANN and CNN) on testing data. Also, perform time analysis using proper visualization.

Note: Use Sklearn for ML and Pytorch for ANN and CNN

Deliverables:

1. Code file (19iXXXX.ipynb) which include,

- Clear and optimize code without any ambiguity
- Detail comments and headings
- Visualization plot for the analysis

2. A word report which includes,

- A detailed conclusion of clustering and sample selection method (BONUS: Visualization of image clustering)
- Analyze any two segmented images from the step 2 to show the perfect segmentation.
- A detailed comparison between each model against segmented and non-segmented images. Report all necessary steps

Marking Rubrics

Description		Marks (130)
Code	Comments and Heading	10
	Clear and optimize steps	30 ML + 15 ANN + 15 CNN = 60
	Visualization and proper output	20
Report	Analysis of steps 1 and 2	5
	Comparison table and detailed discussion for step 3	40
BONUS	Segment the image using advanced algorithms (Step 02)	10