**Introduction**

The objective of this project is to leverage image processing techniques to assist in the classification of ex-vivo brain images as either containing or being free of amyloid-beta (Aβ) plaque. The presence of Aβ is characteristic of the development of Alzheimer’s disease, and with a categorization approach only utilizing image processing, the use of dyes or staining to distinguish between samples that have such plaques and samples without plaques could be avoided or reduced in degree.

The image dataset consisted of 52 greyscale images taken by confocal microscope of samples of the mouse hippocampus and cortex. Specifically, the imaged samples were taken from either 5xFAD mice -- transgenic mice used in Alzheimer’s disease research that develop Aβ plaques -- or C57 strain mice which were used as a control. Cloudy, circular areas are visible as distinct against the background noise in some of the images; these are the areas where Aβ is present in the given sample.

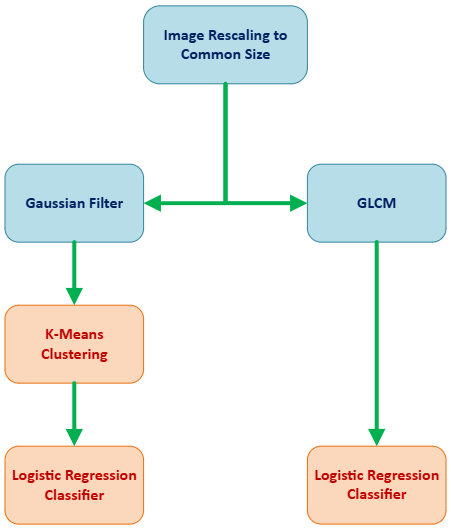
**Image Processing Approach**

To determine whether an image contained Aβ plaques, two image processing approaches were attempted as listed below. Figure 1 illustrates the overall steps in the process.

1. The use of K-means clustering (K = 2) to generate two-color images of each image roughly distinguishing areas of plaque from areas of background noise. These generated images were then fed into a logistic regression model to categorize the images.
2. The generation of Gray-Level Co-occurrence Matrices (GLCMs) for each image to distinguish each image by the texture of the plaque areas as compared to the background noise. Statistical parameters calculated from these matrices were then fed into a logistic regression model to categorize the images.

In the first case, an initial filtering step was performed on all images. This step consisted of applying a simple Gaussian filter with a constant σ of 1 in an attempt at reducing the influence of background noise on subsequent clustering. Visually, this filtering approach resulted in images that appeared to have clearer distinction between the plaque-containing regions and other areas, since the effects of background noise present within the plaque regions and the presence of small brightly lit areas (a consequence of the autofluorescence of stray proteins in the field of view) was diminished by the blurring. In the process of determining the appropriate pre-processing steps to use, a median filter was applied on one image arbitrarily selected from the image set – this filter considerably increased the influence of background noise on the resulting clustering, so the Gaussian filter was selected instead.

For the GLCM generation, the algorithm being a form of texture analysis meant that any blurring of the images to reduce noise would obscure the differences in pixel intensity changes between areas with Aβ and background areas. For this reason, Gaussian filtering was not performed on these images after resizing.



**Figure 1: Two Classification Approaches Applied to the Image Set**

Blue indicates image processing steps utilizing the Python Sci-kit Image library, and red indicates classification steps utilizing the Sci-kit Learn library.

Other image processing techniques were also attempted. Applying a Fast-Fourier Transform algorithm on one image arbitrarily selected from the image set revealed no significant patterns in the frequency spectrum, so no frequency filtering was applied as a pre-processing step. The use of histogram equalization, in a similar manner to the aforementioned median filter, emphasized the noise in the images to such a degree that subsequent K-Means clustering resulted in qualitatively less differentiation between areas with and without plaque.

Additionally, edge detection was explored as a potential means of first delineating relatively large plaque containing regions from background noise. Specifically, the Canny edge detection and Multi-Otsu thresholding algorithms were attempted on a representative image from the dataset; the former was chosen for its general effectiveness as an edge-detection algorithm, especially since it combines Gaussian filtering with hysteresis for edge classification, and the latter was chosen …

**[insert image of Canny edge detection not working well – Figure 2]**

**Results**

**[insert diagram of an image going through each step – Figure 3]**

With the K-Means clustering of each image, the resulting accuracy score of the logistic regression model was **0.45**. With the use of GLCM parameters (specifically angular second moment, contrast, correlation, and dissimilarity), the resulting accuracy score of the logistic regression model was **0.73.** In both cases, the training size was 80% of the image data set.

[…]

**Limitations**

The most notable limitation of both image classification approaches was in the relative paucity of image data. Since only 52 images were present in total to be divided into training and testing subsets, the resulting accuracy scores with a typical split of training and testing data of 80%/20% could be easily swayed by the difference between a single image’s classification.

Additionally, the noise inherent to all the images – including the presence of fluoresced protein “specks” – posed a challenge to determining effective image processing steps. Therefore, there was no case where a clearly superior method to delineate between plaque-containing and plaque-free areas became apparent.