Classification of Diabetic Retinopathy Using Feature Extraction and Statistical Learning

Ivan Echevarria Daniel Ruhnke Daniel Vasiliu Daniel McGibney

February 17, 2016

Acknowledgement: This research uses the HPC resources at W&M.
Many thanks to Eric Walter for his kind support.

William & Mary

Outline

Section 1

Medical Background and Technical Details

Diabetic Retinopathy

[Shafqat, 2011]

- Diabetic Retinopathy is a complication that can occur in people suffering from diabetes
- If allowed to progress can cause blindness.
- Can be treated effectively, especially if detected at an early stage before symptoms are present

[American Academy of Ophthalmology]

- Classified in five stages: Not present (0), Mild Non-Proliferative (1), Moderate Non-Proliferative (2), Severe Non-Proliferative (3), Proliferative (4)

Technical Details

- Data set from Kaggle.

Technical Details

- Data set from Kaggle.
- 35,000 training images with classifications.

Technical Details

- Data set from Kaggle.
- 35,000 training images with classifications.
- We are using Python, Julia, MATLAB, and R to implement our approach along with GPU computing libraries such as OpenCV and ArrayFire.

Characteristics of the Retina

Characteristics of the Retina

- Optic Nerve

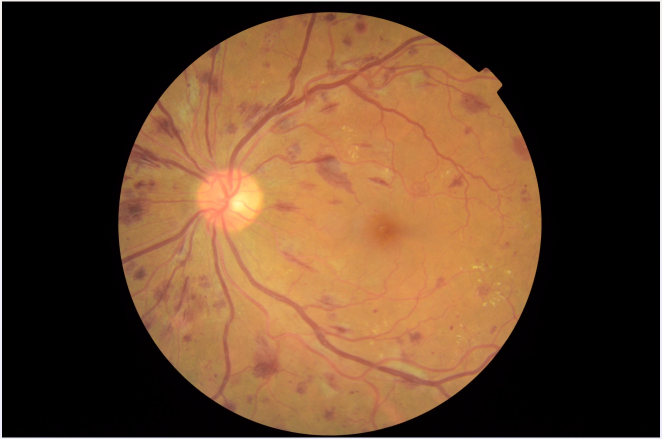
Characteristics of the Retina

- Optic Nerve
- Blood Vessels

Characteristics of the Retina

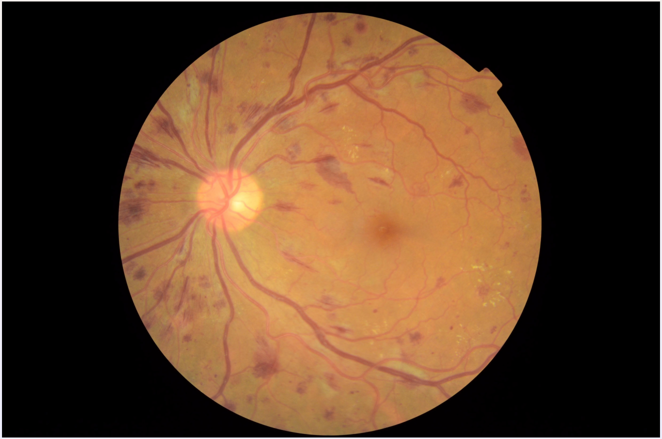
- Optic Nerve
- Blood Vessels
- Macula

Symptoms



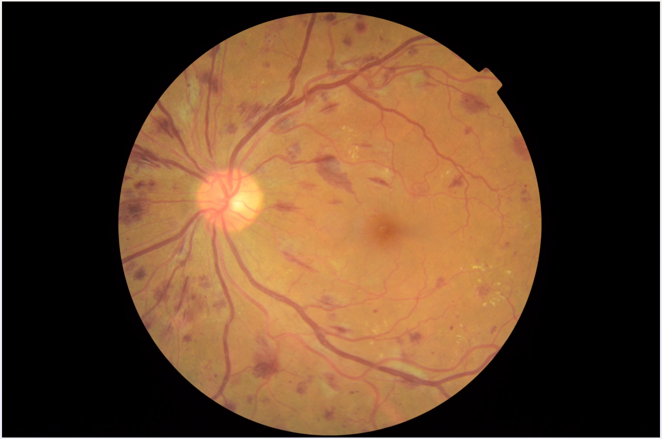
- Microaneurysms ("dots")

Symptoms



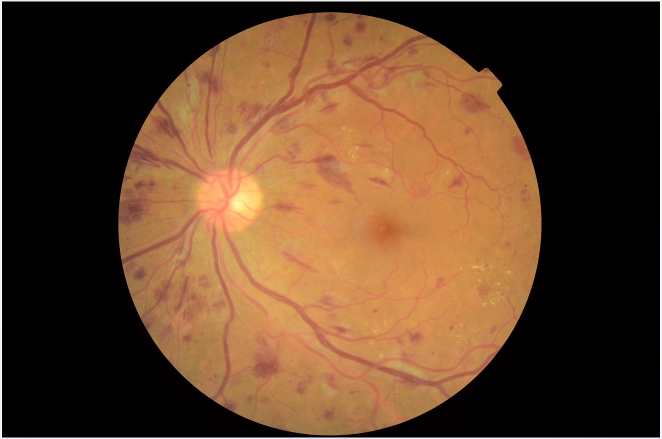
- Microaneurysms ("dots")
- Haemorrhages ("blots")

Symptoms



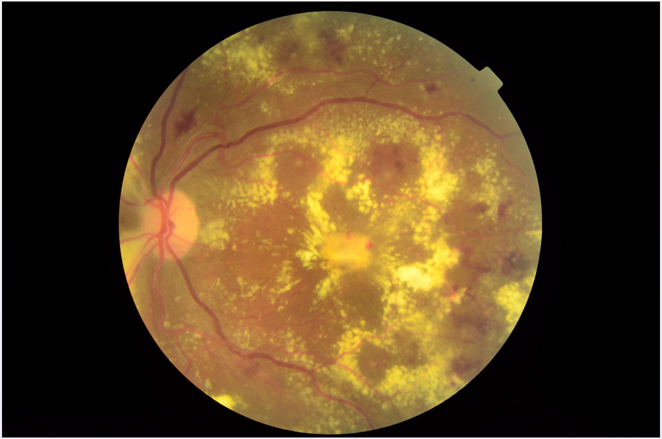
- Microaneurysms ("dots")
- Haemorrhages ("blots")
- Exudates

Symptoms



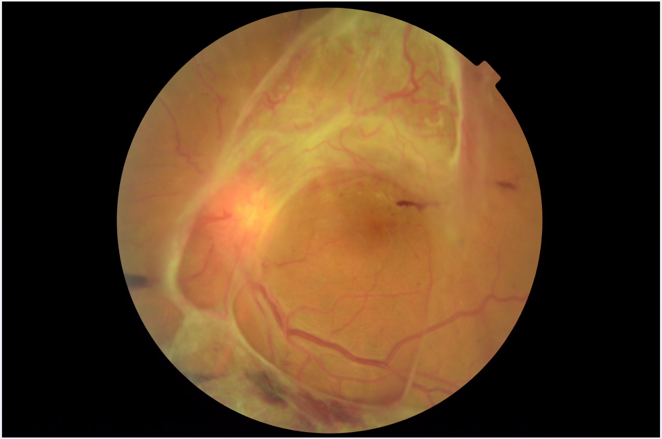
- Microaneurysms (”dots”)
- Haemorrhages (”blots”)
- Exudates
- Tortuosity of Blood Vessels

Proliferative DR



- Cotton Wool Spotting

Proliferative DR cont.

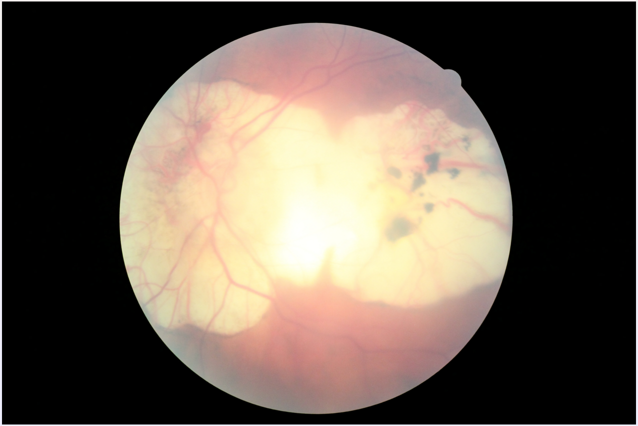


- Neovascularisation

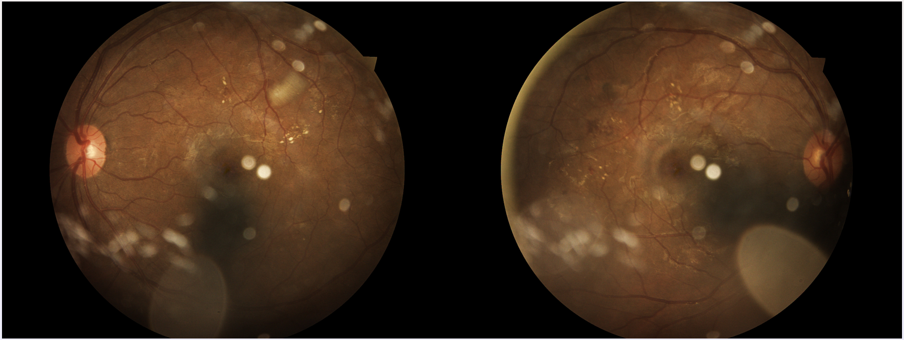
Section 2

Challenges with Feature Extraction

Misdiagnosed Images



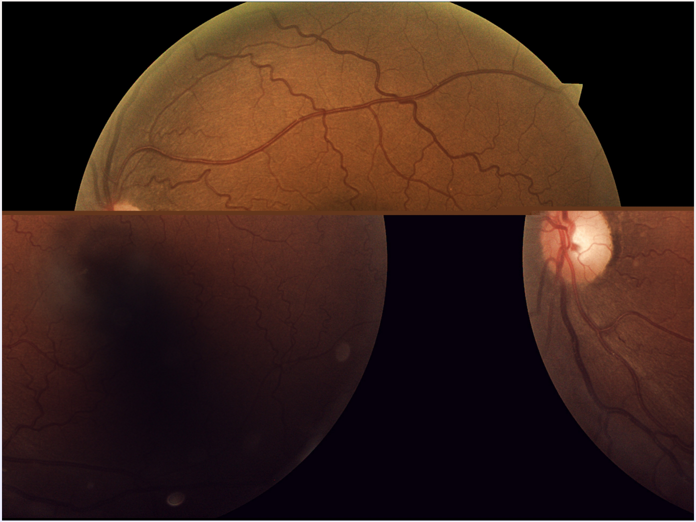
Deceptive Noise



Images with Different Transformations

- Inversion
- Scale

Bad Images

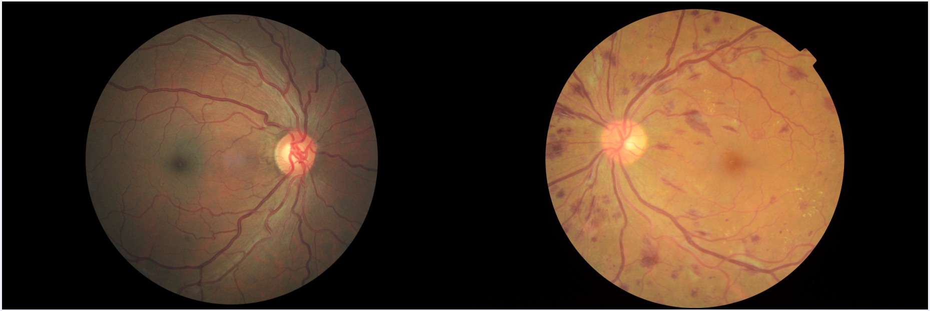


Section 3

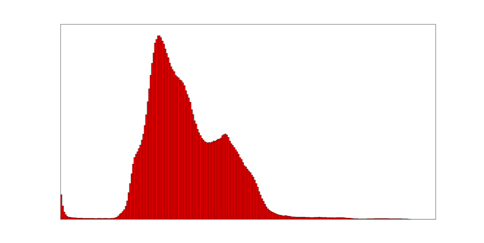
Image Analysis

Example Images

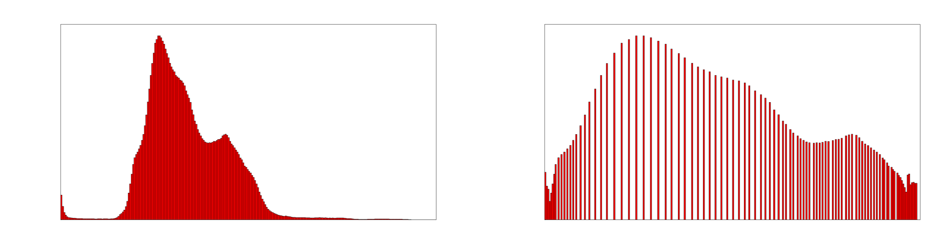
Example Images



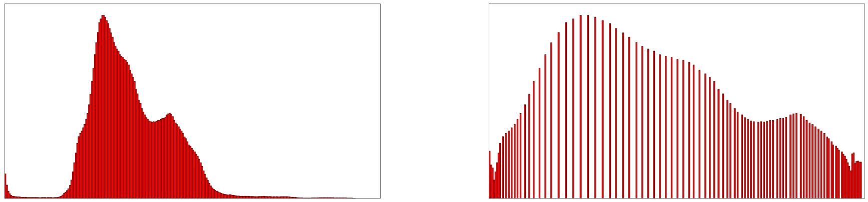
Histogram Equalization



Histogram Equalization



Histogram Equalization

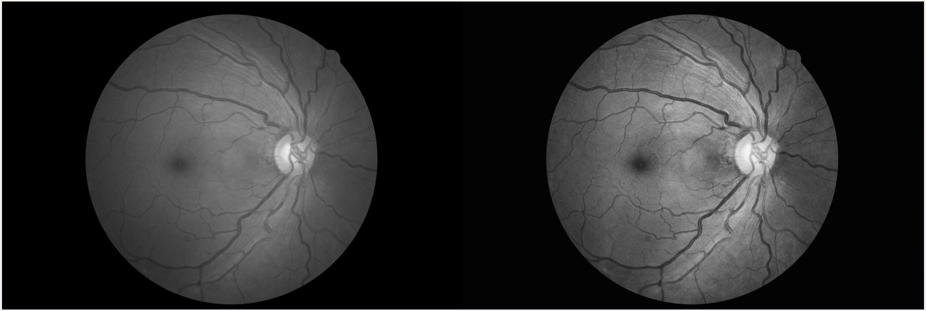


[Gonzalez and Woods, 2008]

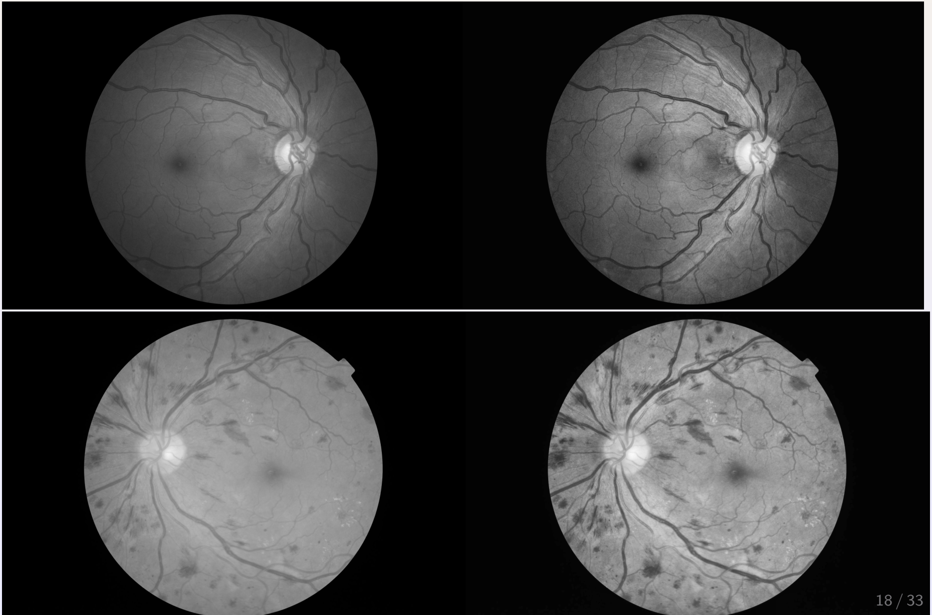
- Pixel $g_{i,j} = floor((L-1)\sum_{n=0}^{f_{i,j}} p_n)$
- Where L is the number of intensity levels in the input image, $f_{i,j}$ is the original intensity of the pixel, and $p_n = \frac{number\ of\ pixels\ with\ intensity\ n}{total\ number\ of\ pixels}$ for $n = 0, 1, ..., L-1$

Results

Results



Results



Adaptive Thresholding

[Fisher, Perkins, Walker, and Wolfart, 2003]

- $T = mean$
- T is the threshold value and *mean* is the mean value of the pixels of the image

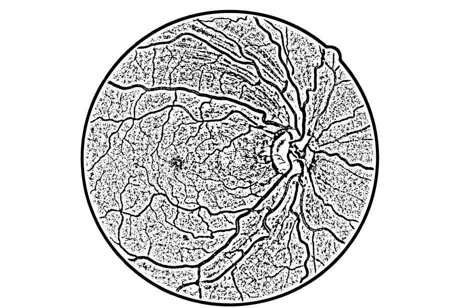
Adaptive Thresholding

[Fisher, Perkins, Walker, and Wolfart, 2003]

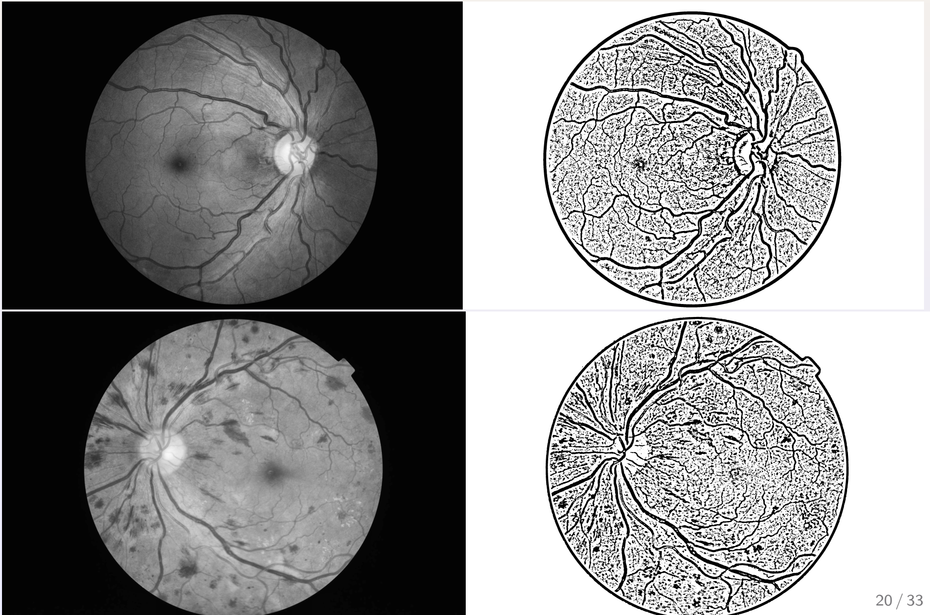
- $T = mean$
- T is the threshold value and *mean* is the mean value of the pixels of the image
- This threshold is computed for a neighborhood of specified size

Results

Results



Results



Connected Components

- Once an image is in binary it can be separated into connected components

Connected Components

- Once an image is in binary it can be separated into connected components
- Difference between 4-connected and 8-connected:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Connected Components

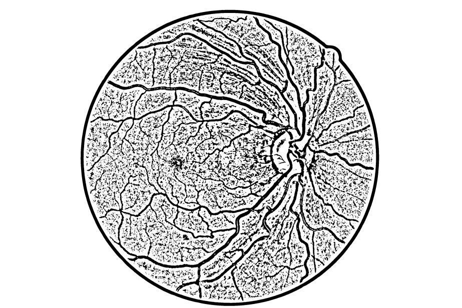
- Once an image is in binary it can be separated into connected components
- Difference between 4-connected and 8-connected:

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

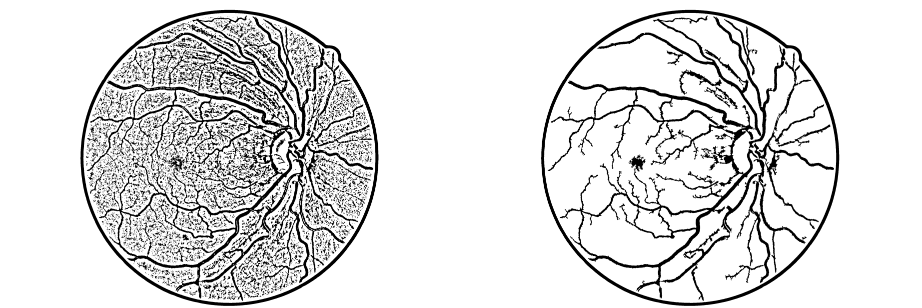
$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

- Eliminate noise by erasing the small connected components

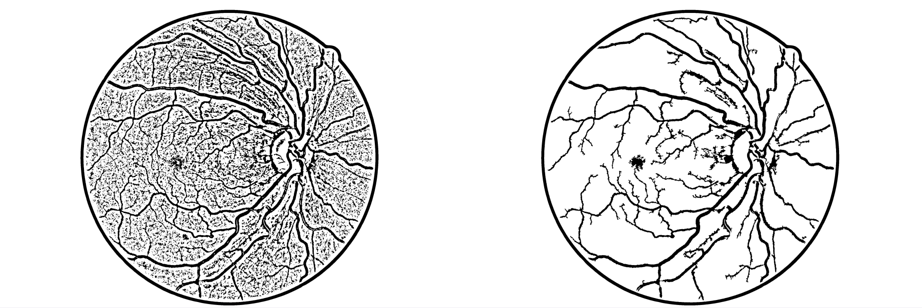
Results

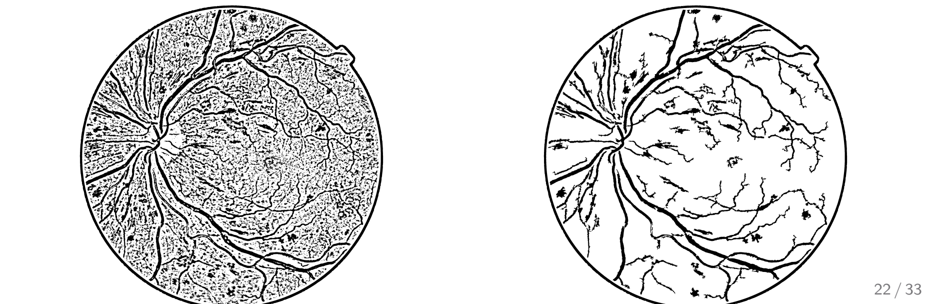


Results



Results





Skeletonization

[Fisher, Perkins, Walker, and Wolfart, 2003]

- Skeletonization reduces a line to a single pixel in width

Skeletonization

[Fisher, Perkins, Walker, and Wolfart, 2003]

- Skeletonization reduces a line to a single pixel in width
- $\textit{thin}(I, J) = I - \textit{hit–and–miss}(I, J)$
- I is an image and J is a structuring element

Skeletonization

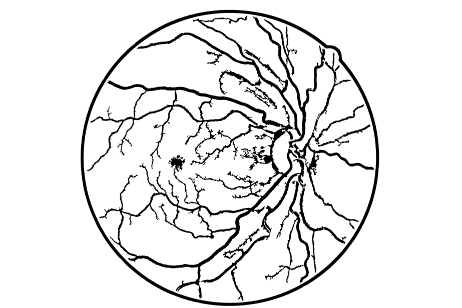
[Fisher, Perkins, Walker, and Wolfart, 2003]

- Skeletonization reduces a line to a single pixel in width
- $thin(I, J) = I - hit\text{–}and\text{–}miss(I, J)$
- I is an image and J is a structuring element
- Structuring elements:

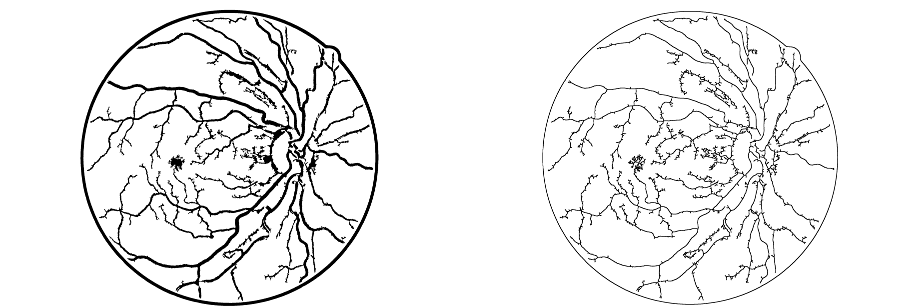
$$\begin{bmatrix} 0 & 0 & 0 \\ & 1 & \\ 1 & 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} & 0 & 0 \\ 1 & 1 & 0 \\ & 1 & \end{bmatrix}$$

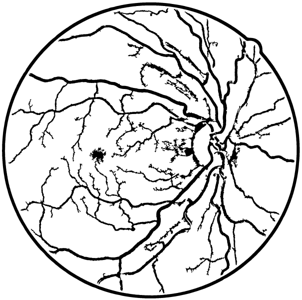
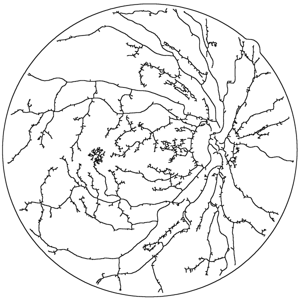
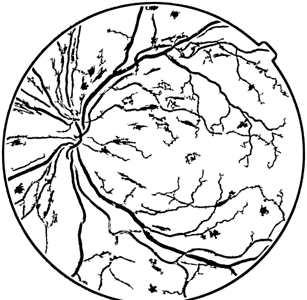
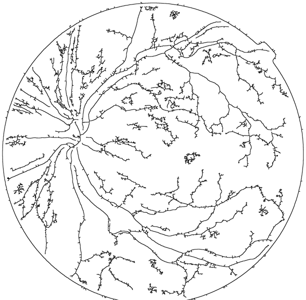
Results



Results



Results

Calculating Tortuosity

- Find branchpoints in the skeleton and extract discrete segments

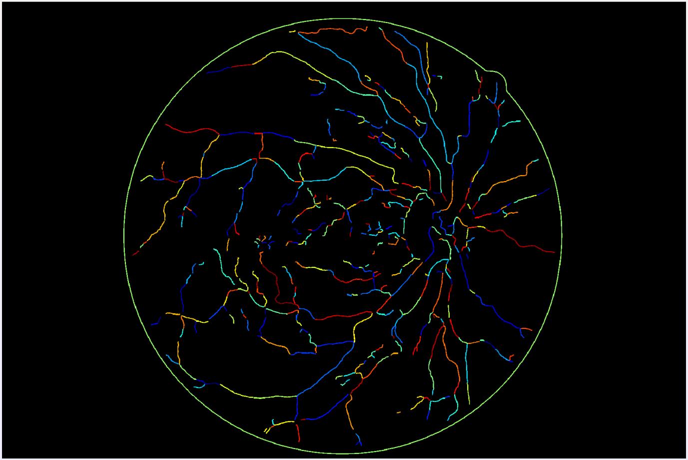
Calculating Tortuosity

- Find branchpoints in the skeleton and extract discrete segments
- $Tortuosity = \frac{L}{C}$

Calculating Tortuosity

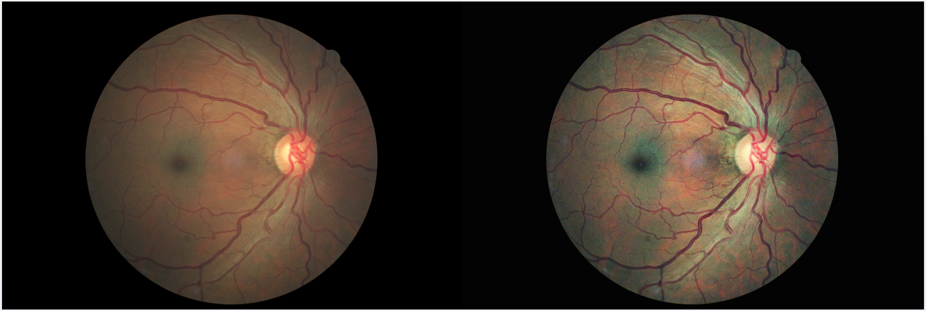
- Find branchpoints in the skeleton and extract discrete segments
- $Tortuosity = \frac{L}{C}$
- L is the length of the segment and C is the Euclidean distance between the endpoints

Example Image

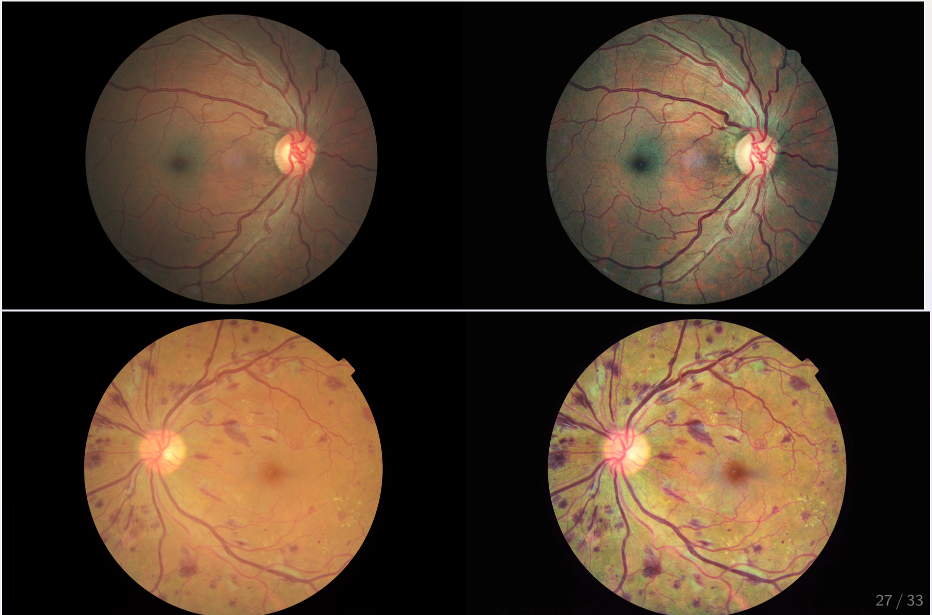


Separating RGB Channels and Equalizing the Histograms

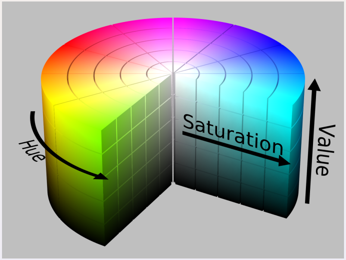
Separating RGB Channels and Equalizing the Histograms



Separating RGB Channels and Equalizing the Histograms

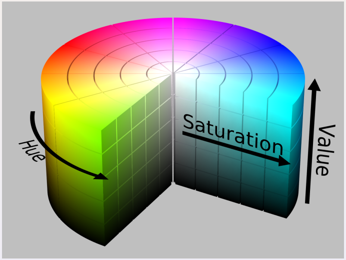


Detecting Blots and Microaneurysms



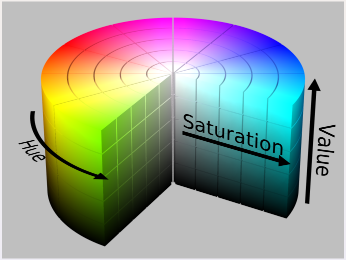
- Looked at image on the HSV color model and defined slice of cylinder which would contain blood vessels and microaneurysms

Detecting Blots and Microaneurysms



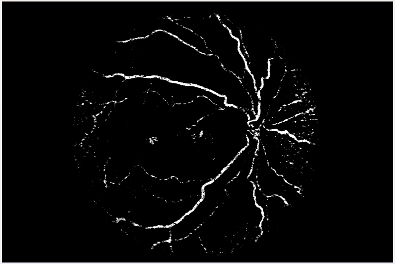
- Looked at image on the HSV color model and defined slice of cylinder which would contain blood vessels and microaneurysms
- Connected components which are in this confidence interval are identified as blood vessels or microaneurysms

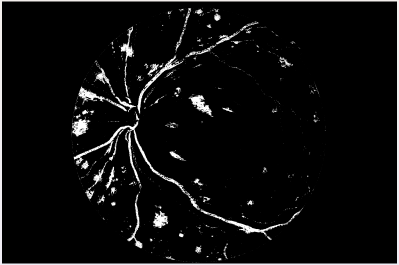
Detecting Blots and Microaneurysms



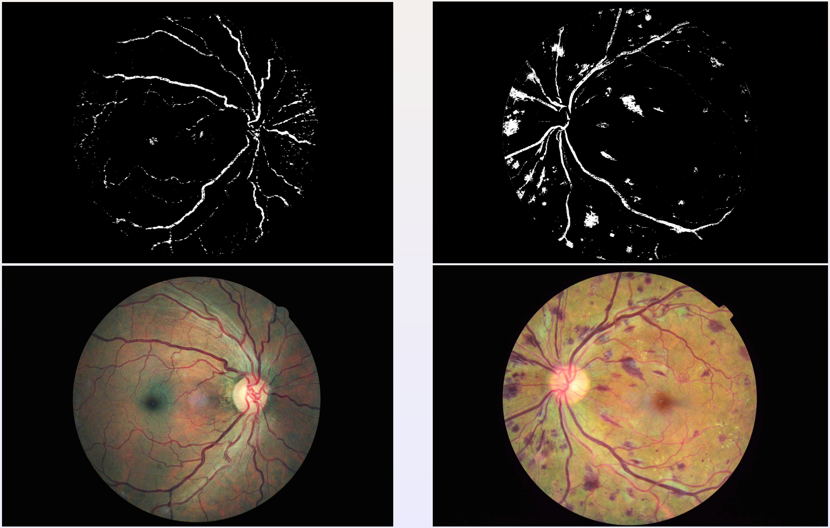
- Looked at image on the HSV color model and defined slice of cylinder which would contain blood vessels and microaneurysms
- Connected components which are in this confidence interval are identified as blood vessels or microaneurysms
- Calculated correlation between x and y coordinates of each connected component. Low correlation considered to be a microaneurysm

Detecting Blots and Microaneurysms

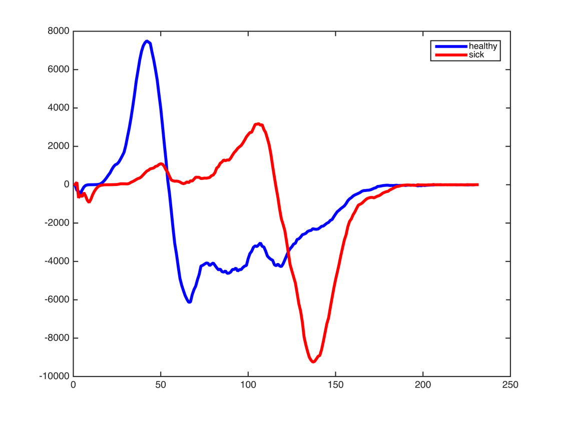




Detecting Blots and Microaneurysms



Incorporating Computational Homology



- Look at b_0 and b_1 over a range of thresholds
- Calculate Euler Characteristic $X = b_0 - b_1$ and plot it.
- Count peaks as an additional variable

Statistical Learning - Logistic Regression

For logistic regression the main idea comes from population biology. If we solve the following differential equation for some initial data we get a CDF:

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right). \tag{1}$$

Let's write $p(x) = Pr(y = 1|x)$ for the response $y = \begin{cases} 1 & \text{if Class I} \\ 2 & \text{if Class II.} \end{cases}$

If we fit the log-odds to a linear univariate model we have

$$\log\left(\frac{p(x)}{1 - p(x)}\right) = \beta_0 + \beta_1 x. \tag{2}$$

From the log-odds, we can calculate the corresponding probability

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}. \tag{3}$$

Maximum Likelihood

We use the maximum likelihood concept to estimate the parameters, where the likelihood equation is given by

$$L(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i:y_i=0} (1 - p(x_i)). \quad (4)$$

This likelihood gives the probability of the observed zeros and ones in the data. We choose β_0 and β_1 to maximize the likelihood of the observed data. Since the logarithmic function is increasing we can maximize instead

$$\frac{1}{n}\sum_{i=1}^{n}\{I(y_i = 1)log(p(x_i)) + I(y_i = 2)log(1 - p(x_i))\} \quad (5)$$

GLMNET

[Friedman, Hastie, and Tibshirani, 2009]

- When working with a binary classification have response variable $y = \{1, 2\}$ represent probabilities with predictors $Pr(y = 1|x) = \frac{1}{1+e^{-(\beta_0 + x^T\beta)}}$ and $Pr(y = 2|x) = \frac{1}{1+e^{(\beta_0 + x^T\beta)}} = 1 - Pr(y = 1|x)$

GLMNET

[Friedman, Hastie, and Tibshirani, 2009]

- When working with a binary classification have response variable $y = \{1, 2\}$ represent probabilities with predictors $Pr(y = 1|x) = \frac{1}{1+e^{-(\beta_0 + x^T\beta)}}$ and $Pr(y = 2|x) = \frac{1}{1+e^{(\beta_0 + x^T\beta)}} = 1 - Pr(y = 1|x)$
- The idea is to search for $(\beta_0, \beta) \in \mathbb{R}^{p+1}$ that maximize penalized log likelihood

$$\frac{1}{n}\sum_{i=1}^{n}\{I(y_i = 1)log(p(x_i)) + I(y_i = 2)log(1 - p(x_i))\} - \lambda P_\alpha(\beta) \quad (6)$$

- Where $P_\alpha(\beta) = \alpha * ||\beta||_1 + (1 - \alpha) * ||\beta||_2^2$

GLMNET

[Friedman, Hastie, and Tibshirani, 2009]

- When working with a binary classification have response variable $y = \{1, 2\}$ represent probabilities with predictors $Pr(y = 1|x) = \frac{1}{1+e^{-(\beta_0 + x^T\beta)}}$ and $Pr(y = 2|x) = \frac{1}{1+e^{(\beta_0 + x^T\beta)}} = 1 - Pr(y = 1|x)$
- The idea is to search for $(\beta_0, \beta) \in \mathbb{R}^{p+1}$ that maximize penalized log likelihood

$$\frac{1}{n}\sum_{i=1}^{n}\{I(y_i = 1)log(p(x_i)) + I(y_i = 2)log(1 - p(x_i))\} - \lambda P_\alpha(\beta) \quad (6)$$

- Where $P_\alpha(\beta) = \alpha * ||\beta||_1 + (1 - \alpha) * ||\beta||_2^2$
- When response variable y has $K > 2$ classifications, then for class ℓ $Pr(y = \ell|x) = \frac{e^{\beta_{0\ell} + x^T\beta_\ell}}{\sum_{k=1}^{K} e^{\beta_{0k} + x^T\beta_k}}$