

PREDICTION OF CUP-TO-DISC RATIO FROM OPTIC FUNDUS IMAGE

By

Jermaine Ramdass, BEng, Ryerson University, June 2010

A Project

presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Engineering

in the program of

Electrical and Computer Engineering

Toronto, Ontario, Canada, 2016

©Jermaine Ramdass 2016

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A PROJECT

I hereby declare that I am the sole author of this project. This is a true copy of the project, including any required final revisions.

I authorize Ryerson University to lend this project to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this project by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my project may be made electronically available to the public.

Prediction of Cup-to-Disc Ratio From Optic Fundus Image

Master of Engineering, 2016

Jermaine Ramdass

Electrical and Computer Engineering

Ryerson University

ABSTRACT

A technique is proposed that can be used to predict the cup-to-disc ratio from a single optic fundus image and determine which image features have the highest contribution to a specific ophthalmologist's measured cup-to-disc ratio. The procedure starts with image pre-processing. The main step of the procedure is feature extraction where image features related to pixel intensities are found. These features are used to train three different classifiers: neural networks, support vector machines, and sparse representation classifiers. The classifiers are tested and evaluated to see how accurately they can predict the cup-to-disc ratio. The best obtained results are in the 70-75% success range. Finally, feature ranking is performed using the methods of chi square and information gain on a combined feature vector using measured cup-to-disc ratios from each ophthalmologist to determine the importance and contribution of each feature to that ophthalmologist.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Kaamran Raahemifar, and my co-supervisor, Dr. Vengu Lakshminarayanan, for giving me the opportunity to work on this project and for all their guidance and support throughout the entire process, my family for providing the stability needed for pursuing my master's degree, and my friends for believing in me.

TABLE OF CONTENTS

Abstract	iii
Acknowledgements	iv
List of Tables	vi
List of Figures	vii
Chapter 1 – Introduction	1
1.1 – Background	1
1.2 – Motivation	2
1.3 – Challenges and Assumptions.....	4
1.4 – Chapter Overview.....	5
Chapter 2 – Related Works	6
Chapter 3 – Theory	8
3.1 – Sparse Representation Classifier.....	8
3.2 – Chi Square Feature Ranking.....	9
3.3 – Information Gain Feature Ranking.....	9
Chapter 4 – Methodology	11
4.1 – Pre-processing.....	12
4.2 – Feature Extraction.....	17
4.3 – Classifier Training.....	25
4.4 – Classifier Testing.....	28
4.5 – Feature Ranking.....	30
Chapter 5 – Results and Discussion	32
5.1 – CDR Prediction.....	32
5.2 – Comparison of Ophthalmologists using Feature Ranking.....	43
5.3 – Comparison with Similar Papers in the Literature.....	50
Chapter 6 – Conclusion	53
6.1 – Accomplishments.....	53
6.2 – Future Work.....	53
References	56

LIST OF TABLES

TABLE I - COMPARISON OF LOW AND HIGH CDR IMAGES BY AVERAGE PIXEL INTENSITY

TABLE II - COMPARISON OF LOW AND HIGH CDR IMAGES BY ENERGY

TABLE III- COMPARISON OF LOW AND HIGH CDR IMAGES BY ENTROPY

TABLE IV - COMPARISON OF LOW AND HIGH CDR IMAGES BY STANDARD DEVIATION

TABLE V- COMPARISON OF LOW AND HIGH CDR IMAGES BY VARIANCE

TABLE VI- CLASSIFIER SUCCESS RATES FOR FEATURE: ALL PIXELS

TABLE VII - CLASSIFIER SUCCESS RATES FOR FEATURE: PCA COEFFICIENTS

TABLE VIII - CLASSIFIER SUCCESS RATES FOR FEATURE: IMAGE HISTOGRAM

TABLE IX - CLASSIFIER SUCCESS RATES FOR FEATURE: PCA COEFFICIENTS USING HSV IMAGES

TABLE X - CLASSIFIER SUCCESS RATES FOR FEATURE: COMBINED FEATURE VECTOR

TABLE XI - CLASSIFIER SUCCESS RATES ACROSS ALL OPHTHALMOLOGISTS: HISTOGRAM

TABLE XII - CLASSIFIER SUCCESS RATES ACROSS ALL OPHTHALMOLOGISTS: COMBINED
FEATURE VECTOR

TABLE XIII - 10-FOLD CROSS-VALIDATION: HISTOGRAM

TABLE XIV - 10-FOLD CROSS-VALIDATION: COMBINED FEATURE VECTOR

TABLE XV - TOP 100 FEATURES USING CHI SQUARE FEATURE RANKING

TABLE XVI - REDUCED FEATURE RANK LIST

TABLE XVII - CLASSIFICATION USING REDUCED SETS OF FEATURES

TABLE XVIII - COMPARISON WITH SIMILAR PAPERS

LIST OF FIGURES

Figure 1 - Cup and Disc Regions

Figure 2 - Comparison of cup and disc regions by different ophthalmologists

Figure 3 - Histogram of CDR Values

Figure 4 - The SRC Algorithm

Figure 5 - High-level Flowchart for Proposed Technique

Figure 6 - Pre-processing Stage of Proposed Technique

Figure 7 - Full-size fundus image

Figure 8 - Localized image with cup and disc region

Figure 9 - Individual RGB Channels

Figure 10 - HSV Model of Localized Image

Figure 11 - Individual HSV Channels

Figure 12 - Feature Extraction Stage of Proposed Technique

Figure 13 - Comparison of Histograms Between Low and High CDR Fundus Images

Figure 14 - Eigenvectors Found Using PCA for 100 Image Data Set

Figure 15 - Structure of Combined Feature Vector

Figure 16 - Classifier Training Stage of Proposed Technique

Figure 17 - Classifier Testing Stage of Proposed Technique

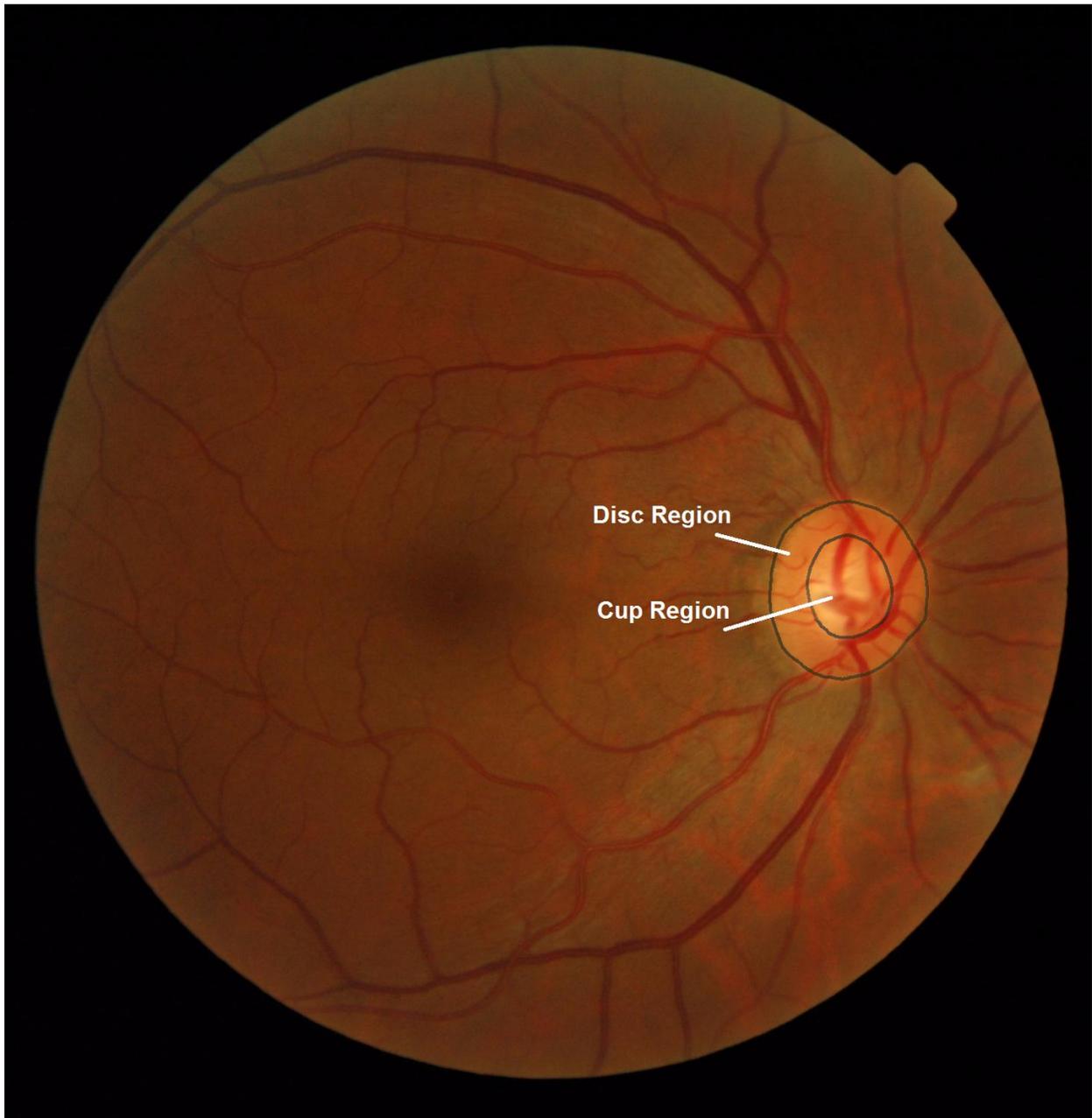
Figure 18 - Feature Ranking Stage of Proposed Technique

CHAPTER 1 – INTRODUCTION

1.1 Background

The cup-to-disc ratio (CDR) found from optic fundus images is a popular and simple measure of determining whether a patient may have the degenerative optic nerve condition known as glaucoma. Glaucoma, if left untreated, can lead to partial or complete vision loss and, as it has few noticeable symptoms at an early stage, can be difficult to detect. This makes quick, affordable screening methods such as fundus image CDR measurements very important. There are two regions located around the optic nerve that are used in measuring the CDR. The first is the disc region which is where the optic nerve and blood vessels enter the retina. The second is the cup region which is the center region of the disc. Fig 1 below shows the entire fundus image with the cup and disc regions labelled as marked by an ophthalmologist. The outer circle region is the disc and the inner circle is the cup. The cup region grows if there is an increased intraocular pressure and this leads to a higher CDR, a good indicator that a patient has glaucoma. However, it does not guarantee the presence of glaucoma and other factors such as the intraocular pressure and condition of the optic nerve need to be considered for a full diagnosis.

Fig. 1. Cup and Disc Regions

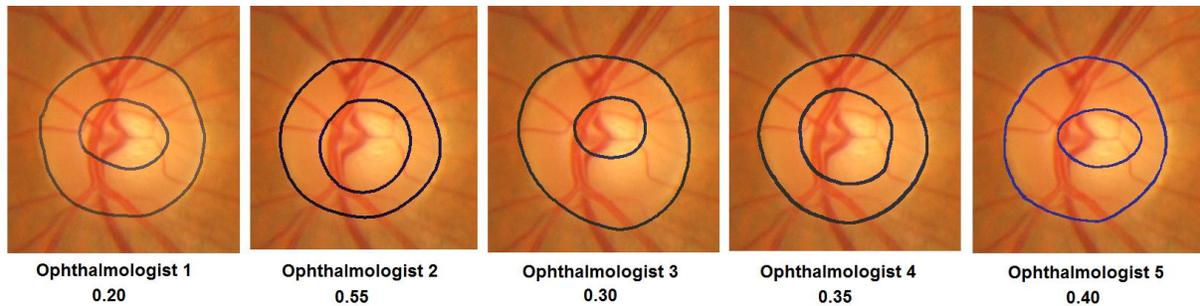


1.2 Motivation

A lot of research has already been done in accurately segmenting the cup and disc regions from fundus images in order to calculate the CDR. This paper attempts something different. In the examined data set there are 100 images, each of which have been marked by six different

ophthalmologists showing an outline around both the cup and disc regions. Also included is the unmarked version of each image. It has been noticed that the six CDRs for each image can sometimes vary by up to 50% between the six ophthalmologists which is a significant difference. Fig 2 shows one such example where each ophthalmologist assigned a different CDR value for the same fundus image. In this example, the CDRs vary by up to 35%. Looking closely at this comparison, it seems that the disc region is roughly similar across each manual marking. Therefore, the variance might mostly be attributed to the different cup markings.

Fig. 2. Comparison of cup and disc regions by different ophthalmologists

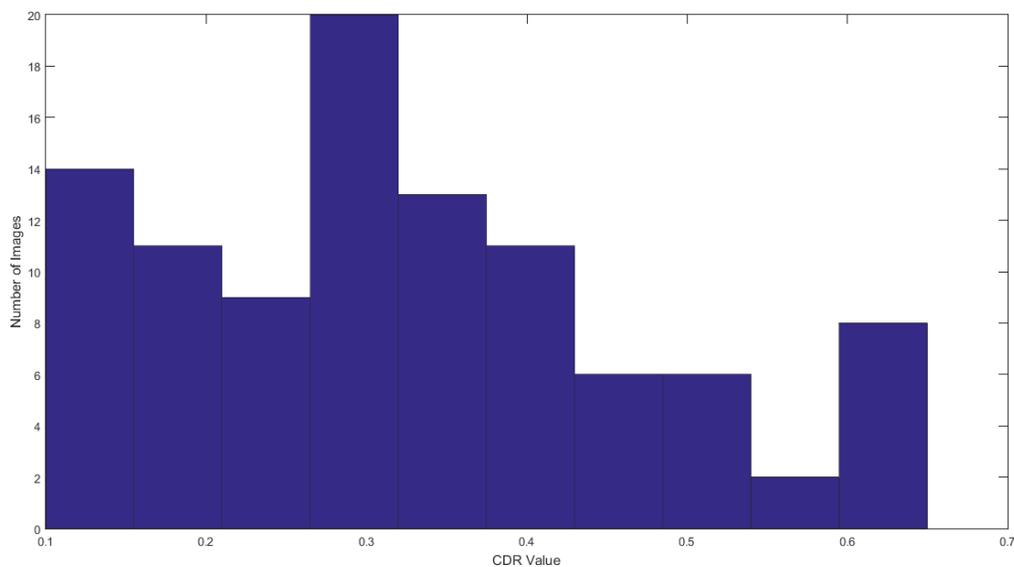


The purpose of this project is two-fold. First, classification is performed based on fundus image features in order to accurately and consistently predict what CDR value a given ophthalmologist will measure. The goal is to be able to predict the CDR value any given ophthalmologist will measure for a fundus image based on all the information about that ophthalmologist from fundus images they have already marked. The second purpose of this project is to try and determine why there is such a large variation between some of the CDR measurements between different ophthalmologists. This is done by examining which features are more correlated to each CDR measurement through the use of feature ranking, thus determining which features a given ophthalmologist prefers when deciding a CDR value.

1.3 Challenges and Assumptions

There are a few challenges associated with this project, mainly in terms of the machine learning aspect of it. The first challenge is the size of the sample space. There are only 100 images for each ophthalmologist. This is a relatively small amount of data for training an accurate and general classifier. This is made even more challenging due to the higher number of classes when performing multi-class classification. Multi-class classification is highly dependent on each class having a good amount of representation in the data set. In Fig 3, a histogram of the CDR values for one ophthalmologist is shown. This describes how many images of the 100 image data set fall into each class. As can be seen, four classes have eight or fewer images to represent them. Therefore, training a classifier that can accurately predict these underrepresented classes is much more difficult than predicting the better represented classes. An explanation on how these challenges are mitigated will be made throughout the procedure.

Fig. 3. Histogram of CDR Values



For the purpose of this project, a few assumptions are made in order to approach the problem in a preferred way. The first assumption made is that the method should avoid using segmentation. Even though segmentation can result in accurate CDR values, the aim in this project is to use the entire fundus image itself in order to predict the CDR. Taking too many steps that lead to segmenting the cup and disc regions is also avoided as there is less value in predicting the CDR when only a step or two away from having a completely segmented image which can be used to directly measure the CDR. Furthermore, while lengths, widths, and areas of cup or disc regions would result in accurate CDR predictions, it is assumed that the value of each of these features are dependent on the ophthalmologist and are not prior knowledge or information that can be obtained directly from an image without segmentation.

1.4 Chapter Overview

The remainder of this paper is organized into the following chapters:

- In Chapter 2, a brief overview of various research papers that have been examined in this research domain is given.
- In Chapter 3, the basic theory on some of the methods used throughout the procedure is explained.
- Chapter 4 presents the step-by-step procedure used to reach the goals that have been outlined for this project.
- Chapter 5 presents the results along with a discussion on them.
- Finally, in Chapter 6, overall conclusions and ideas for future work are given.

CHAPTER 2 – RELATED WORKS

After performing a literature review of papers in this research area, only two have been found that have previously done something similar to what is proposed here. In [1], 2-class classification (presence of glaucoma or not) is performed purely with the image data. Image features of pixel intensities, textures, fast Fourier transform (FFT) coefficients, and histograms are used to test naive Bayes, K-nearest neighbours (KNN), and support vector machine (SVM) classifiers. Success rates ranging from 73-83% are achieved. In [2], image texture and higher order spectra features are used to perform 2-class classification to detect the presence of glaucoma. Success rates ranging from 85-91.7% are achieved. A very similar paper, [3], performs 2-class and 3-class classification to determine the presence of diabetic macular edema (DME) which is also an abnormal eye condition. Global fundus image features obtained from a motion pattern method are used.

Most papers calculate the CDR after performing segmentation. There has been extensive research done in segmenting the cup and disc regions and the methods have become more advanced and accurate over the years. An extensive list of glaucoma detection techniques is given in [4]. A very recent method which has become popular is an approach using polar maps [5]. In this method, both supervised and unsupervised segmentation is performed to extract the cup from the disc region. The cup is then represented as a sector based polar map and features are taken from each of these sectors, resulting in 192 features. Histogram features are also added to this. A random forest classifier is then trained and tested using these features. An area under the curve (AUC) of 0.8964 is achieved. As stated in the assumptions in Chapter 1.3, segmentation methods are not used in this paper. The focus in this paper is predicting CDR values based on the ophthalmologist and also to compare the importance of features between

them. Therefore, even though segmentation methods are highly accurate in finding the cup and disc region, as well as directly measuring the CDR, the purposes of this paper require a different approach.

Although the aim is to determine what features might have more contribution to a given ophthalmologist's measurement, papers that give methods for segmentation have also been examined to see what kind of pre-processing they perform on fundus images. They also help in gaining some prior knowledge about fundus images and CDRs in general. For example, in [6], [7], and [8] it is stated that the red colour channel is better for segmenting the disc region while the green colour channel is better for segmenting the cup region. Therefore, not only are RGB or grayscale images used for feature extraction but the individual colour channels as well. Different feature ranking methods that are used in machine learning problems were also found. In [9], feature ranking was performed using linear SVM and in [10] it was performed using AdaBoost. However, the methods from [2] which include chi square and information gain are used in this proposed technique mainly due to their simplicity and how commonly they are used.

CHAPTER 3 – THEORY

In this chapter, the basic theory for some of the methods used is given so that a reader can understand how they work and why they are used in this proposed technique.

3.1 Sparse Representation Classifier

While the other classifiers used are very common, a Sparse Representation Classifier (SRC) which is less common is also used. A colleague suggested using this classifier since it is good at using the entire sample space. This would help with the challenge mentioned about having underrepresented classes. Some background information about SRC is given here as it might not be as well known. Fig 4 shows the algorithm for SRC. As shown, it is a four step process. In step 1, the columns of the data set, \mathbf{X} , which correspond to samples of each class are l_2 normalized. In step 2, the test image, \mathbf{y} , is coded as $\mathbf{y} \approx \mathbf{X}\boldsymbol{\alpha}$ where $\boldsymbol{\alpha}$ is known as the coding vector. In step 3, the residuals of \mathbf{y} are calculated using (2). In step 4, the coding vector, $\boldsymbol{\alpha}_i$, with the most significant values for \mathbf{y} is found as the one resulting in the minimum residual. This means that \mathbf{y} most likely belongs to class i .

Fig. 4. The SRC Algorithm [11]

-
1. Normalize the columns of \mathbf{X} to have unit l_2 -norm.
 2. Code \mathbf{y} over \mathbf{X} via l_1 -minimization
$$(\hat{\boldsymbol{\alpha}}) = \arg \min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ s.t. } \|\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}\|_2 < \varepsilon \quad (1)$$
where constant ε is to account for the dense small noise in \mathbf{y} , or to balance the coding error of \mathbf{y} and the sparsity of $\boldsymbol{\alpha}$.
 3. Compute the residuals
$$e_i(\mathbf{y}) = \|\mathbf{y} - \mathbf{X}_i \hat{\boldsymbol{\alpha}}_i\|_2 \quad (2)$$
where $\hat{\boldsymbol{\alpha}}_i$ is the coding coefficient vector associated with class i .
 4. Output the identity of \mathbf{y} as
$$\text{identity}(\mathbf{y}) = \arg \min_i \{e_i\} \quad (3)$$
-

For a more detailed description of SRC, please see [11]. For this project, a MATLAB implementation of SRC was used which can be found at [12]. In order to use this implementation, a MATLAB-based software for convex programming known as CVX is also needed and it can be found at [13].

3.2 *Chi Square Feature Ranking*

The first feature ranking method used is Chi Square which is basically a statistical method for determining if there is a relationship between different categories. It is a popular method for performing feature ranking or selection. This is done by seeing how significant the relationship is between each feature in a feature vector and the target class. The results are then sorted to form a ranking based on significance. In order to calculate chi square, (4) is used where χ is the chi square value, n is to total number of samples, O is the observed value, and E is the expected value. The value, χ , is higher if there is more correlation between the two values and therefore, features with high chi square values will be ranked higher. For more information on how to calculate chi square, please see [14].

$$\chi = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \quad (4)$$

3.3 *Information Gain Feature Ranking*

The second feature ranking method used is called information gain, also known as mutual information, which uses entropy information. To calculate information gain, (5) is used where $IG(Y/X)$ is the information gain, $H(Y)$ is the total entropy of the class Y , and $H(Y/X)$ is the conditional entropy of Y given the feature X . This basically means that each feature is checked to

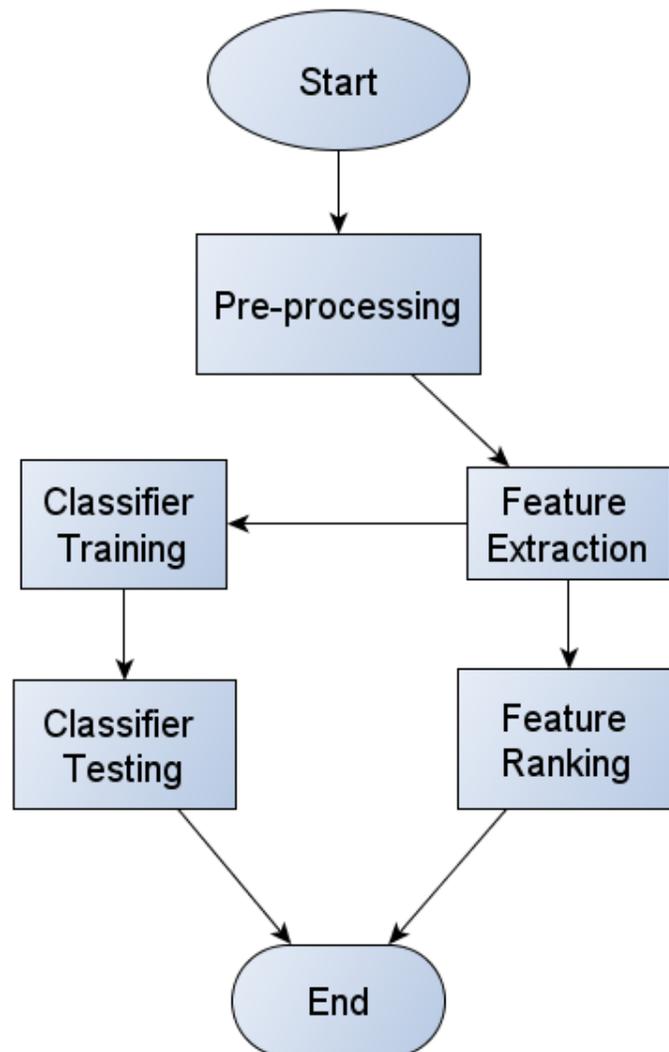
see how often it results in a specific classification. The more times a feature results in a specific classification, the lower the expected entropy will be, thus resulting in a higher information gain. Features that have higher information gain values are ranked higher. For more information on how to calculate information gain, please see [15].

$$IG(Y, X) = H(Y) - H(Y|X) \quad (5)$$

CHAPTER 4 – PROPOSED TECHNIQUE

The overall flow of the project can be split into several main steps. Fig 5 shows the high-level flow of the entire project. The remainder of this chapter will examine each of these steps in greater detail. Once feature extraction is performed, CDR prediction can take place by proceeding to the classifier training and testing steps. However, feature ranking can also be performed directly after the feature extraction step as the classifier training and testing results are unnecessary for it.

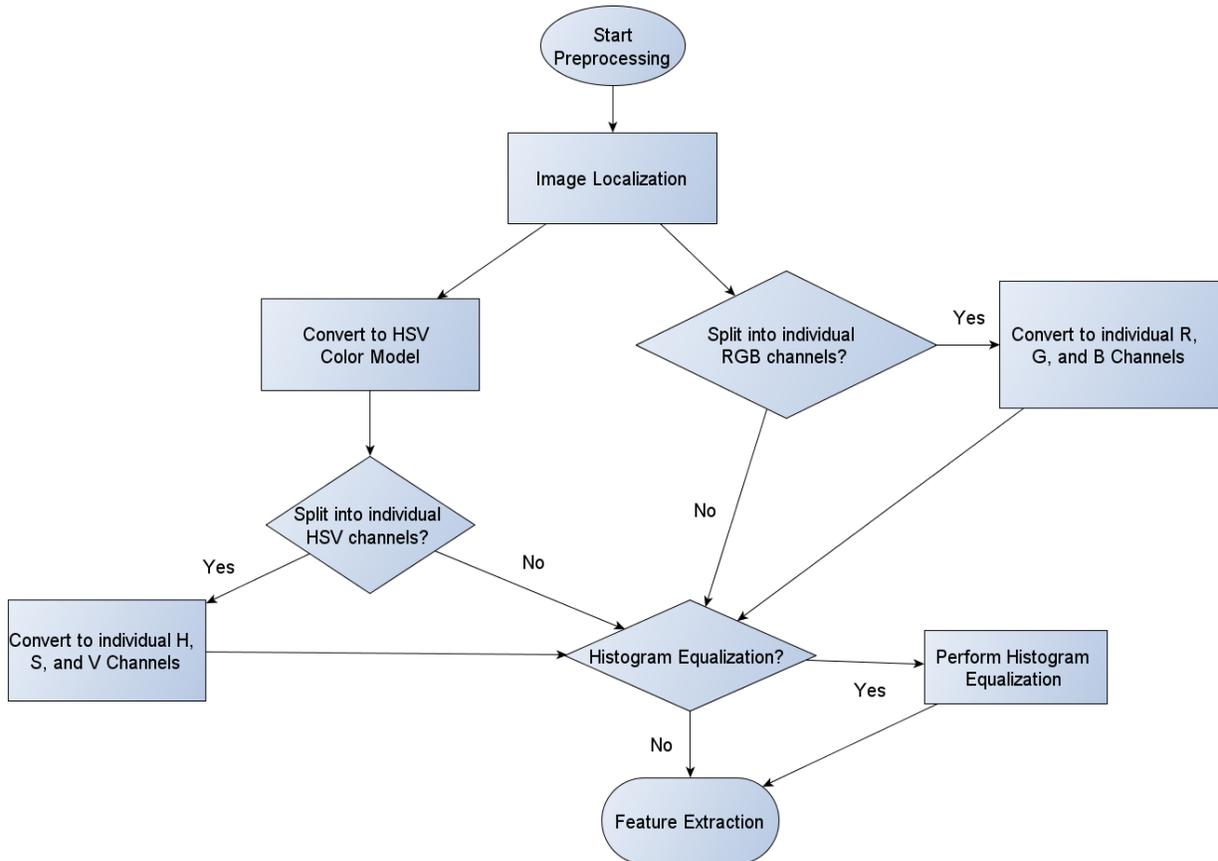
Fig. 5. High-level Flowchart for Proposed Technique



4.1 Pre-processing

A flowchart of the pre-processing step can be seen in Fig 6.

Fig. 6. Pre-processing Stage of Proposed Technique



Pre-processing begins with image localization. Each fundus image in the data set is in TIF format and has a resolution of 2240 x 1488. This equals to over three hundred thousand pixels in each image. However, only a small fraction of about 3% of the total pixels contains the disc and cup region. An example of one of these data images can be seen in Fig 7.

Fig. 7. Full-size fundus image



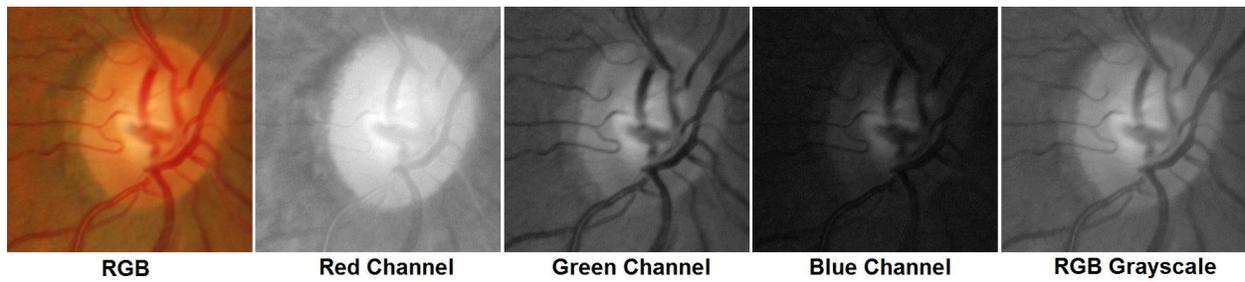
Since only a small portion of the image is actually needed, pre-processing is carried out in order to create a localized image containing only the region of interest: the cup and disc region. This is done by finding high intensity regions in the image from the image histograms. A population of high intensity regions is found through a genetic algorithm and then the largest of these regions is cropped from the image as the region of interest. This resulting region is significantly smaller than the original image having a resolution of only 351x351. The cup and disc region is accurately found in 95/100 of the test images. A sample of a localized image is shown in Fig 8. The code for this was provided by a colleague.

Fig. 8. Localized image with cup and disc region



During pre-processing, histogram equalization of each image is also performed and the red, green, and blue channels of each image are taken separately. Different combinations of these steps lead to different results in later stages. For example, extracting certain features from the blue colour channel leads to better classification results than performing that same feature extraction on the green colour channel. However, for a different feature, the green colour channel might give better results than the blue channel. In Fig 9, a side-by-side comparison of each RGB channel is shown so that a reader can get a better idea of what each looks like and how different information might be better obtained in different colour channels.

Fig. 9. Individual RGB Channels



The HSV (Hue, Saturation, Value) colour model of each image is also found as it contains additional information that can help with the image classification. The HSV version of Fig 8 can be seen in Fig 10 below. Similar to the RGB image, each channel of the HSV image is also used separately. In Fig 11, a side-by-side comparison of each individual HSV channel is shown.

Fig. 10. HSV Model of Localized Image

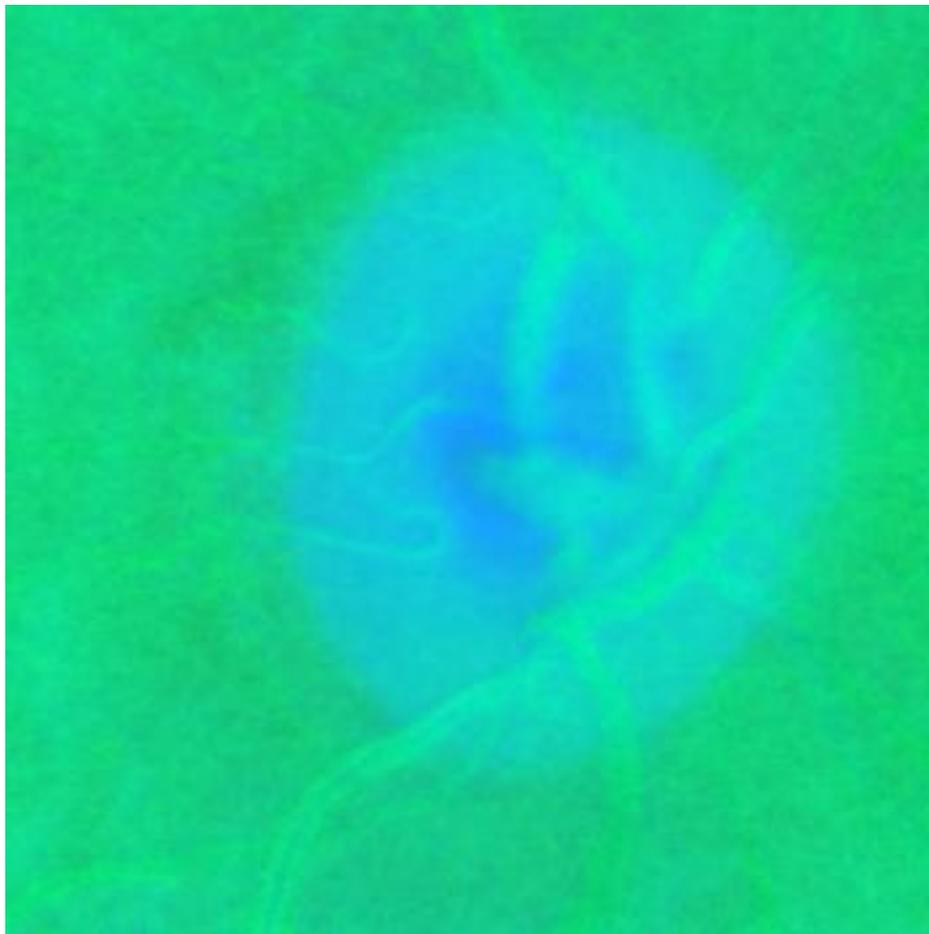
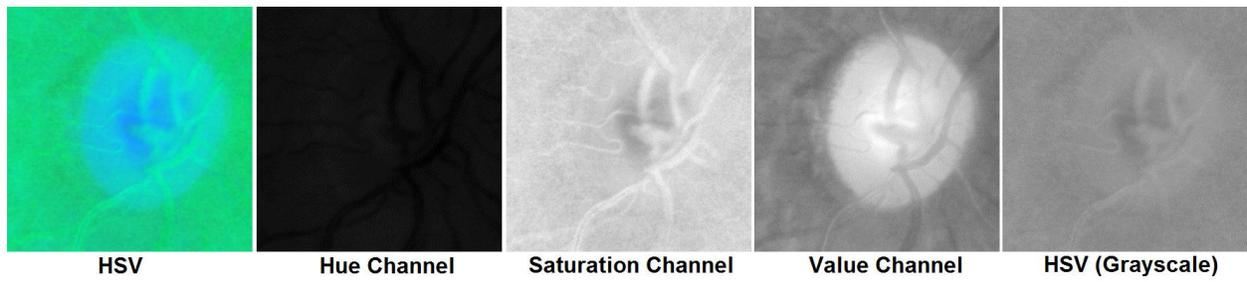


Fig. 11. Individual HSV Channels

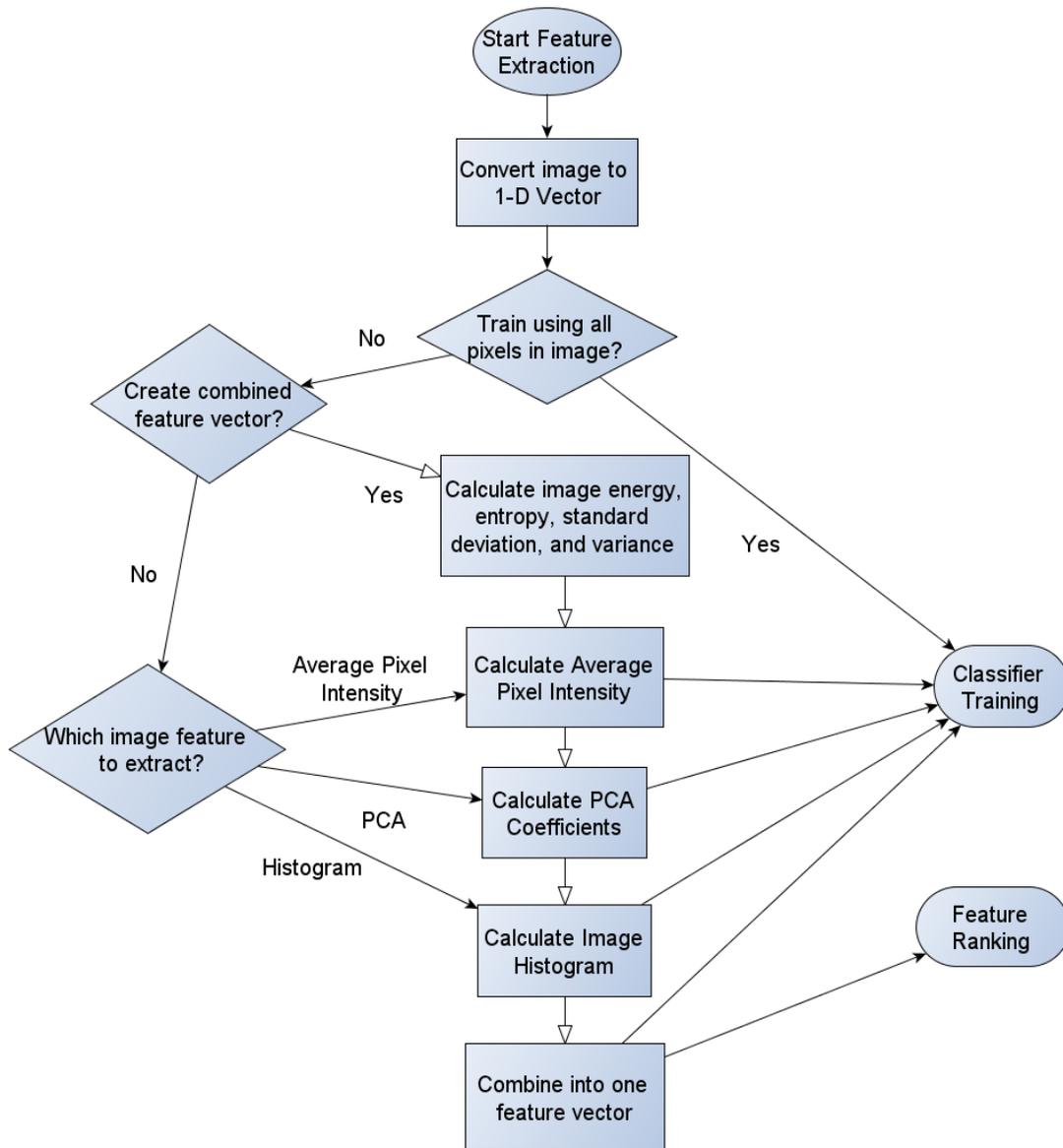


After the image region of interest is localized, the various colour channels are obtained and histogram equalization is performed if necessary, the images are ready to have the desired features extracted from them.

4.2 Feature Extraction

A flowchart of the feature extraction stage is shown in Fig 12.

Fig. 12. Feature Extraction Stage of Proposed Technique



The most important step of the procedure is feature extraction. With images there is a wide variety of features that can be extracted. In order to determine which features have a higher contribution towards predicting the CDR, several need to be tried. The main goals of this project

are to find features that can be used to accurately predict the CDR and also to determine how the importance of each feature varies between different ophthalmologists to explain how they can assign highly varied CDRs to the same fundus image. These goals are mainly reflected in this stage of feature extraction.

The simpler features that are used include all of the pixel intensities in an image and the average pixel intensity of an image. As stated before, this can be done on each of the colour channels or a histogram equalized image to lead to different results. These are common features used for training classifiers related to images, but there might be some correlation between pixel intensities and the CDR. This is because the disc region has a higher intensity than the remainder of the image while the cup region has an even higher intensity than the disc region. This can be seen in Fig 8 by observing how the pixel intensities change throughout the image. It is one of the main features used when being manually marked by an ophthalmologist. In Table I below the average pixel intensities between low CDR and high CDR images as marked by a single ophthalmologist are compared. As can be seen from the table, there is no clear distinction between the average pixel intensities whether the CDR is low or high so it may not be a useful feature for classification based just on this comparison.

TABLE I. COMPARISON OF LOW AND HIGH CDR IMAGES BY AVERAGE PIXEL INTENSITY

CDR (Low Values)	Average Pixel Intensity	CDR (High Values)	Average Pixel Intensity
0.1	95.1657	0.6	93.8719
0.1	111.8675	0.6	119.6729
0.1	131.1622	0.65	117.3108
0.1	141.4763	0.6	108.1754
0.1	133.9871	0.65	109.9459

The next simple feature used is image energy which is just the sum of all pixels in an image. In Table II below a similar comparison is performed as before with energy instead of average pixel intensity. Similar to average pixel intensities, there is no clear distinction between the energy of low CDR images and high CDR images.

TABLE II. COMPARISON OF LOW AND HIGH CDR IMAGES BY ENERGY

CDR (Low Values)	Image Energy	CDR (High Values)	Image Energy
0.1	11724506	0.6	11565106
0.1	13782184	0.6	14743826
0.1	16159315	0.65	14452810
0.1	11181550	0.6	13327316
0.1	17430016	0.65	13545450

Next, the entropy of each image is compared in Table III below. Here it can be seen that entropy is slightly higher in high CDR images, which means that it might be a decent feature for classification.

TABLE III. COMPARISON OF LOW AND HIGH CDR IMAGES BY ENTROPY

CDR (Low Values)	Image Entropy	CDR (High Values)	Image Entropy
0.1	6.0185	0.6	6.3487
0.1	6.1895	0.6	6.4614
0.1	6.2683	0.65	6.5784
0.1	6.0693	0.6	6.5844
0.1	6.2765	0.65	6.2290

The next simple feature examined is standard deviation and the comparison is shown in Table IV. For this set of images, it can be seen that the standard deviation is usually around twice as large in the high CDR images compared to the low CDR images. This might be an indication that standard deviation is a good feature for classifying between low and high CDR images.

TABLE IV. COMPARISON OF LOW AND HIGH CDR IMAGES BY STANDARD DEVIATION

CDR (Low Values)	Image Standard Deviation	CDR (High Values)	Image Standard Deviation
0.1	0.0355	0.6	0.0484
0.1	0.0346	0.6	0.0600
0.1	0.0204	0.65	0.0596
0.1	0.0310	0.6	0.0643
0.1	0.0367	0.65	0.0611

The final simple feature used is image variance and the comparison is shown in Table V below. The values here show that variance is about an order of magnitude larger in high CDR

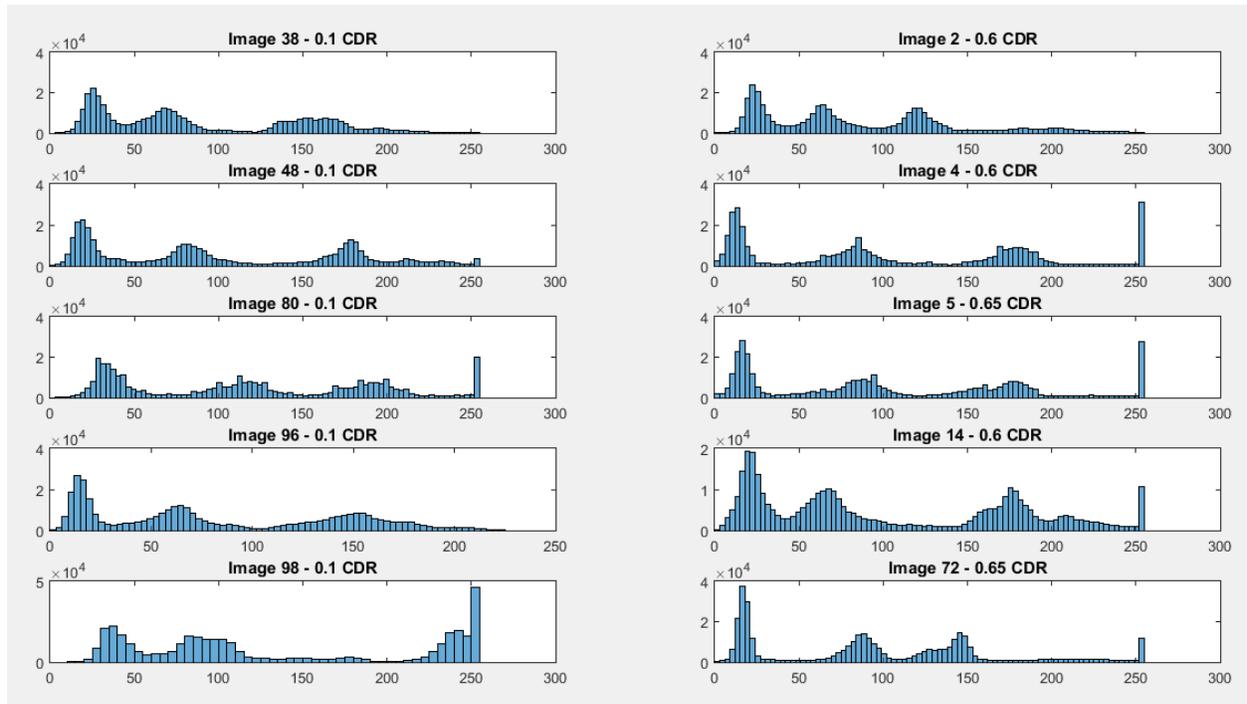
images compared to low CDR images. This is also a good indication that variance might be a strong feature for classification.

TABLE V. COMPARISON OF LOW AND HIGH CDR IMAGES BY VARIANCE

CDR (Low Values)	Image Variance	CDR (High Values)	Image Variance
0.1	2.4621e-05	0.6	6.4474e-05
0.1	2.1445e-05	0.6	1.5620e-04
0.1	6.4762e-06	0.65	1.5294e-04
0.1	1.6085e-05	0.6	2.0927e-04
0.1	2.2883e-05	0.65	1.3919e-04

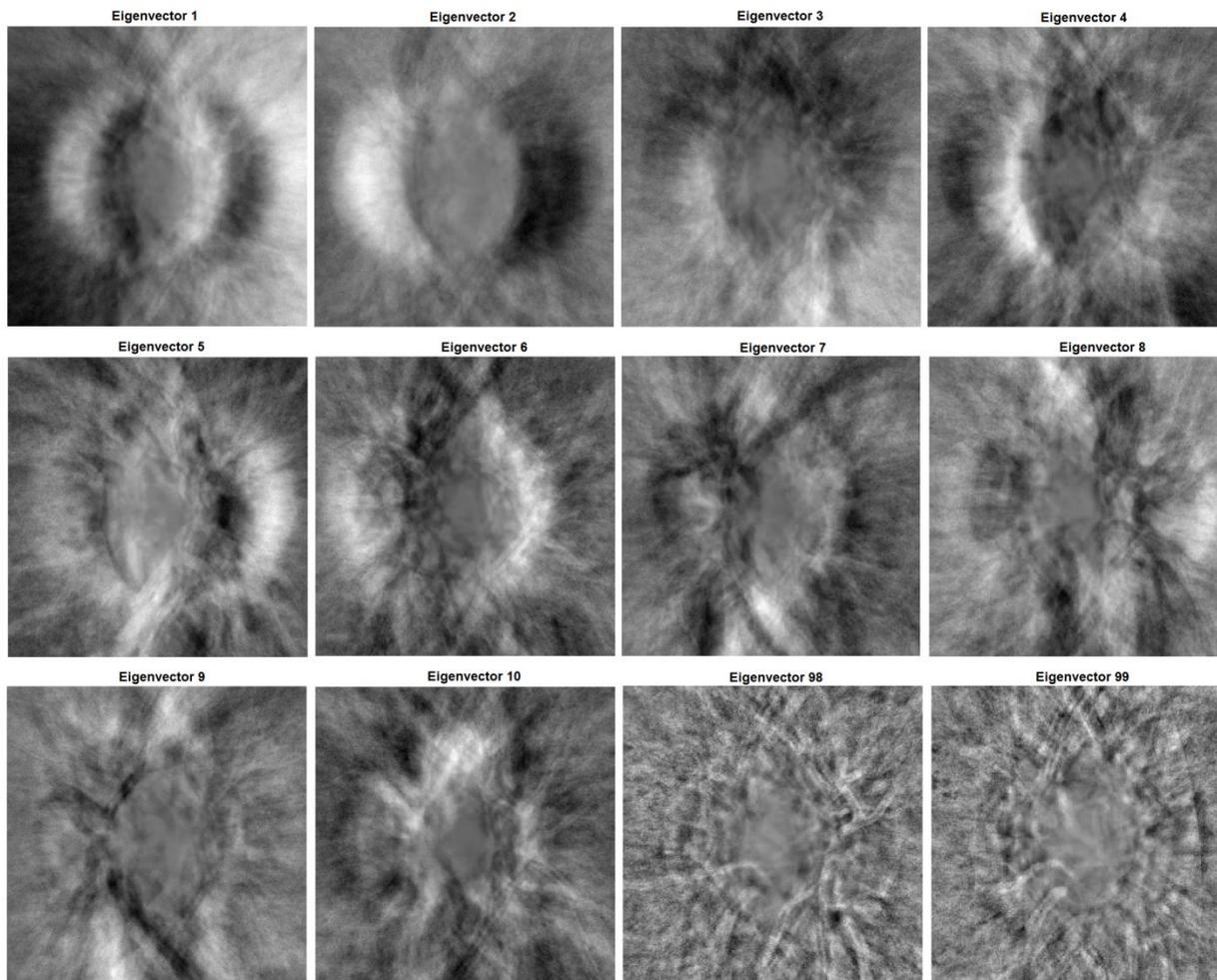
For more advanced features, image histograms are used. A histogram shows the distribution of pixel values in bins ranging from low intensity to high intensity. This also follows the idea that higher intensities are attributed to cup and disc regions. Fig 13 below shows the histogram plots of the same images used in the tables above. The left side of the figure shows the histograms for the low CDR images and the right side shows the histograms for the high CDR images. It was expected that high CDR images would have more pixels grouped on the right side of the histogram because of the higher quantity of high intensity pixels due to the larger cup and disc regions. Based on Fig 13, there does appear to be a spike at the highest intensity bin of the histogram for most of the high CDR images. However, some of the low CDR images also have this spike.

Fig. 13. Comparison of Histograms Between Low (Left) and High (Right) CDR Fundus Images



Principal component analysis (PCA) was also used. PCA finds the component images of a set of images along with weights/coefficients that can be used to linearly add these component images in order to reproduce the original image. PCA is primarily used to reduce the dimensionality of a data set. Although built in MATLAB functions were used to calculate the eigenvectors for PCA, a tutorial, [16], was helpful in understanding how PCA works and to calculate the PCA coefficients. In Fig 14 below, the top 10 eigenvectors that result from using PCA on the data set of 100 images are shown. When using the PCA function in MATLAB, the eigenvectors are automatically sorted to have the highest variance eigenvectors listed first. Since there are 100 images in the data set, 99 eigenvectors are generated. Also included in Fig 14 are the last two eigenvectors to show what the low variance eigenvectors look like. The majority of the 99 eigenvectors are visually similar to eigenvectors 98 and 99 shown below. The high variance eigenvectors contribute the most to the overall image.

Fig. 14. Eigenvectors Found Using PCA for 100 Image Data Set



Although each feature is tested separately, a combined feature vector which is composed of several of the features mentioned above is also created. The exact features are described in the next paragraph. This procedure was used so that it could be immediately seen if any single feature was a strong decider in the final CDR value. The combined feature vector is important since it is used later for feature ranking. Therefore, more testing was done with this feature vector than with the individual features, with the exception of features that performed well. By looking at the rank assigned to each feature in the vector, it would be possible to see which features contributed more to predicting the output CDR. These ranks could also be compared between different ophthalmologists to see how the importance of each feature varies between them.

The feature vector was assembled using the simple image features of average pixel intensity, energy (which is simply a sum of all pixel intensities in the image), entropy, standard deviation and variance. For each image, these values were found for the three colour channels separately and also for the RGB grayscale image which resulted in a feature vector with a length of 20. Furthermore, the top 10 PCA coefficients and a 10 bin histogram for each of the three colour channels and the RGB grayscale image were also taken. This results in another 40 PCA features and 40 histogram features, for a total of 100 features. Finally, these same measurements were taken for the HSV version of the image, resulting in an overall feature vector of 200 features. More PCA coefficients and larger histograms would have been taken but classifier training becomes much slower around 200 features so the feature vector was kept at this size. A diagram of the feature vector components can be seen in Fig 15. Each colour channel has this same set of features with a length of 25 and there are eight in total (red, green, blue, RGB grayscale, hue, saturation, value, HSV grayscale) resulting in a total length of 200. The histogram bins are arranged from low intensity to high intensity going from left to right in the feature vector. The PCA coefficients are arranged from high variance to low variance from left to right.

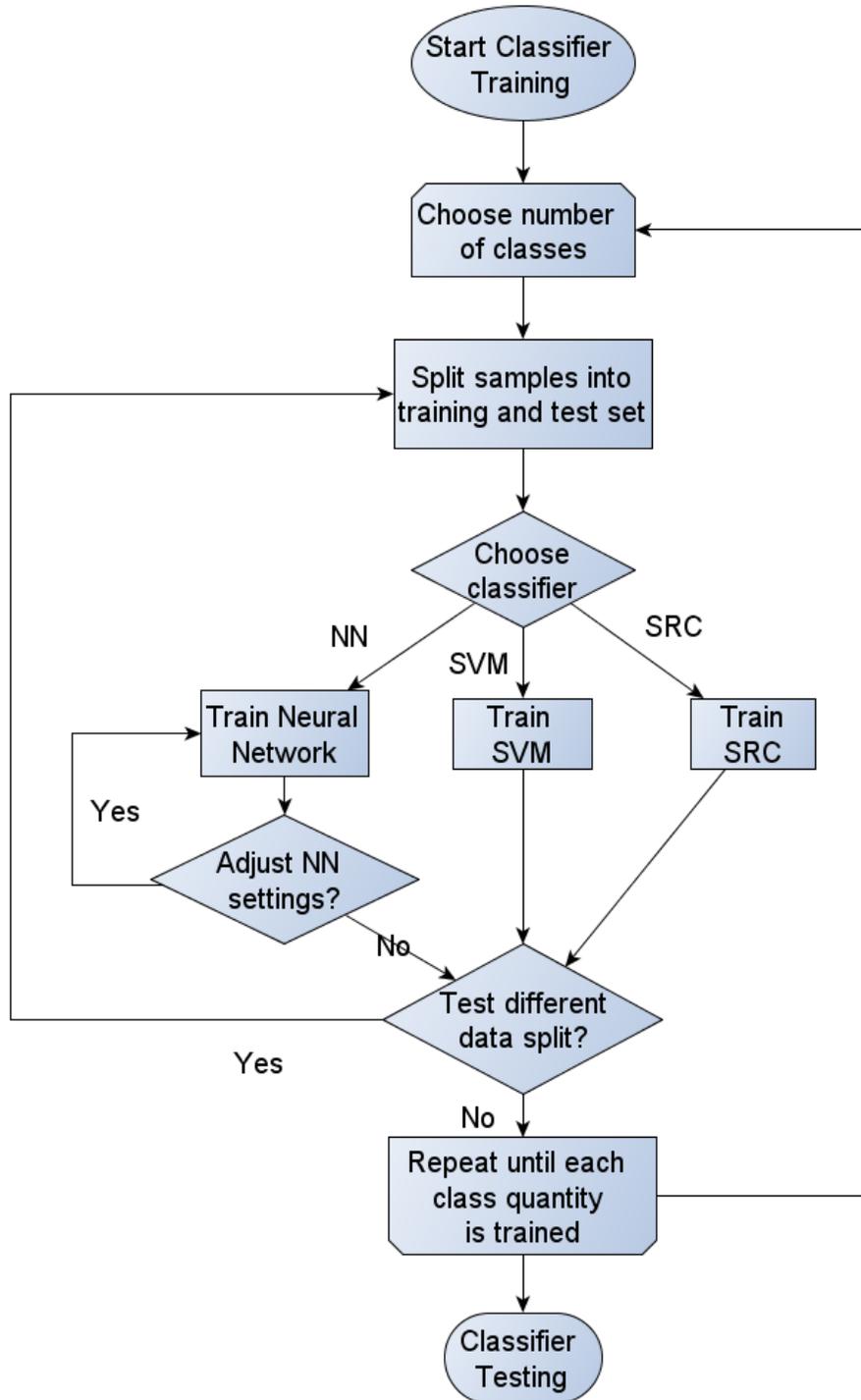
Fig. 15. Structure of Combined Feature Vector

	Colour Channel						
Feature	Average Pixel Intensity	Energy	Entropy	Standard Deviation	Variance	Top 10 PCA Coefficients	Histogram (10 Bins)
Feature Length	5					10	10

4.3 Classifier Training

A flowchart of the classifier training stage is shown in Fig 16.

Fig. 16. Classifier Training Stage of Proposed Technique



Each feature mentioned in Chapter 4.2 was then used as the training input to a classifier. The 100 images in the overall data set were divided between a training set and test set. A size of 70 images for the training set and 30 for the test set was often used. However, different train/test data splits were also tested to see if they would give better results. This means that each classifier was trained with a number of inputs equal to the size of the training set in the form of vectors with varying lengths depending on the feature. For example, with pixel intensities the length of the feature vector was 123 201 pixel intensities (351x351), for PCA the length was one less than the training set size of coefficients, and for average pixel intensity the length was 1 pixel intensity.

The main classifier type used was neural networks (NN). This is because MATLAB allowed NNs to be easily created and highly customized. Several parameters could be varied including the number of hidden nodes, training algorithm, number of epochs, target performance, and minimum gradient. The main training algorithms provided by MATLAB were Scaled Conjugate Gradient (SCG), Gradient Descent (GD), Levenberg-Marquardt (LM), and Bayesian regularization (BR). These parameters also added variations to the results and so they were all tested for each NN in order to obtain the best performance. While different parameters were tested, generally 10 hidden nodes and the BR training algorithm were used which seemed to give the best results most often. SVMs were used as an alternative classifier and they were trained with the same training data as the NNs to see if they would give better results. Finally, upon recommendation from a colleague, SRC was used as it is a good classifier for making use of training samples from each class.

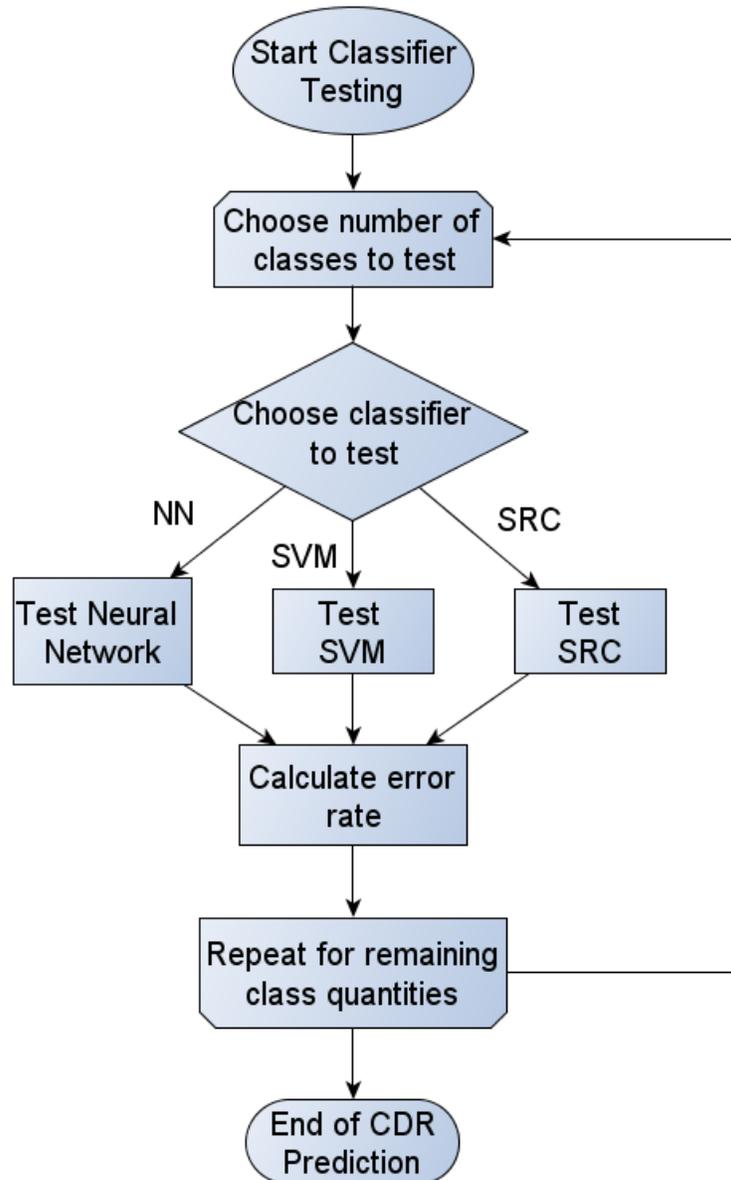
Furthermore, different quantities of classes were used including: full (which was between 12-14 classes, depending on the ophthalmologist, and resulted from taking the CDRs at 0.05 intervals), half (which was 6-7 classes and resulted from taking the CDRs at 0.10 intervals), 3 (classification ranges of 0.0 – 0.24, 0.25 – 0.49, 0.50+), and 2 classes (0.0 – 0.49, 0.50+). The

reduced numbers of classes were used to see if the results would be significantly improved. Also, 2 classes is the most common number of classes when classifying CDR images as shown in papers [1] and [2]. This is because the main goal is to diagnose glaucoma and thus the most important factor is whether the CDR is above or below a certain threshold, usually 0.50. By testing with different numbers of classes, the challenge of having a fairly low amount of samples was partially mitigated. Testing with reduced numbers of classes also helped give better class representation.

4.4 Classifier Testing

A flowchart of the classifier testing stage is shown in Fig 17.

Fig. 17. Classifier Testing Stage of Proposed Technique



As with training, each of the three classifier types needed to be tested. Furthermore, each multi-class amount also needed to be tested resulting in 12 tests for each feature. The process shown in Fig 17 was followed to help keep the large amount of testing and resulting data

organized. To test a classifier, the test set is inputted into the classifier. The NN and SVM classifiers used output a set of probabilities equal in size to the number of classes. The highest probability was then taken as the class that the input sample most likely belonged to. The SRC classifier output a single class label value directly. Each class label corresponded directly to a distinct CDR value.

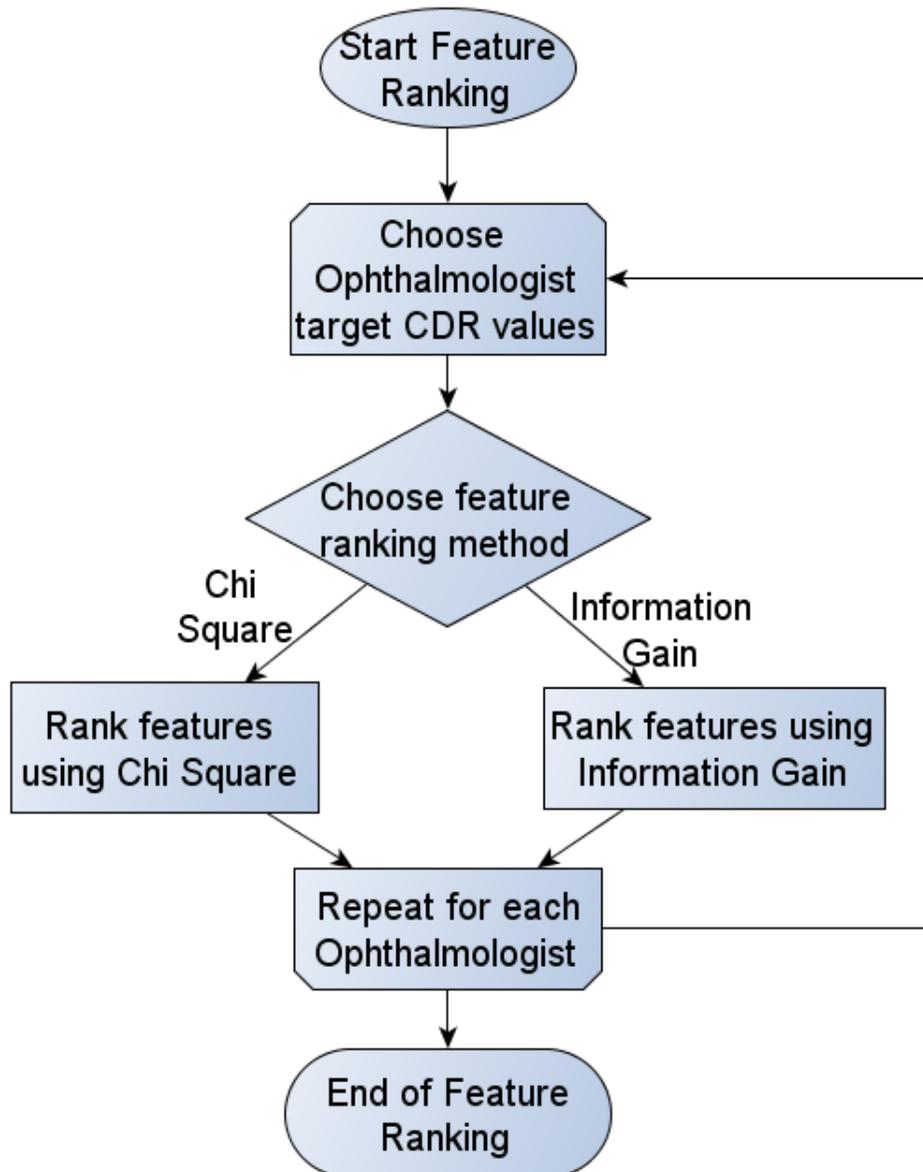
In order to evaluate a classifier, a binary 1 or 0 corresponding to a successful prediction or a failure was used. This was calculated by (6) where y is the predicted CDR and T is the target CDR. A value of 0.10 was chosen so that a predicted CDR needed to be within $\pm 10\%$ of the target CDR to be considered successfully predicted. Originally, a success criterion of within $\pm 5\%$ was used but due to poor results, it had to be raised. Note that for the 3-class and 2-class classification cases, this criterion does not apply and the success of a prediction is measured based on whether or not the predicted CDR matches the target CDR exactly. By having the full and half-class cases use a less strict success criterion, the challenge of doing multi-class classification with some underrepresented classes was partially mitigated.

$$\begin{cases} 1, & |y - T| \leq 0.10 \\ 0, & |y - T| > 0.10 \end{cases} \quad (6)$$

4.5 Feature Ranking

A flowchart of the feature ranking stage is shown in Fig 18.

Fig. 18. Feature Ranking Stage of Proposed Technique



The final step of the proposed technique was to perform feature ranking. The goal of this step was to determine the contribution of each feature in the combined feature vector towards successfully predicting the CDR. However, the main use of feature ranking was to compare the

feature contributions between each ophthalmologist. This was done to gain an idea of why for some fundus images each ophthalmologist assigned a different CDR value even though it was the exact same image. By seeing how the feature ranks change for each ophthalmologist, it can be determined if different features are more important for different ophthalmologists.

Two different feature ranking methods were used: Chi Square and Information Gain. These are commonly used methods for feature ranking and their basic theory can be found in Chapter 3. The program, Weka, which is data mining software, was used and it has both of these feature ranking algorithms built in. The process started with choosing an ophthalmologist to rank the features on. The features are run with the specific ophthalmologists given CDR values. Since they each assigned different CDR values to the dataset of images, the feature ranking would be performed using different targets for each ophthalmologist. The result was a sorted list of features by rank which could then be examined to see feature contributions for the ophthalmologist. Furthermore, each list of ranked features could then be compared to the resulting lists of every other ophthalmologist to see how they differ.

CHAPTER 5 – RESULTS AND DISCUSSION

The results presented here are split into two sections corresponding to the two main goals of this project which are CDR prediction and examining why different ophthalmologists assign different CDR values to the same fundus image. A comparison of these results with other similar papers in the literature is also performed. Every step of the process was performed using MATLAB R2015a, except for feature ranking which used Weka 3.6.13 [17], on a laptop with an Intel Core i3 Processor at 2.13 GHz and 4.0 Gb of RAM.

5.1 *CDR Prediction*

The results for CDR prediction are obtained at the proposed technique stage described in Chapter 4.4, feature testing. The results for some of the features mentioned in the feature extraction stage can be seen in the tables below. This is only a subset of the more successful features and configurations that were tried. Most of the results are based on target CDRs from a single ophthalmologist. A few of the highest achieving classifiers are tested with targets from each ophthalmologist and the results are also included below. Most of the NNs are using the BR algorithm and have 10 hidden nodes unless otherwise stated. The aim is to have a high success rate, particularly for the test set as this shows whether the classifier will be good for unseen data.

Note that the overall average is not obtained by directly taking the average of the training set and test set. The overall average is a weighted average that depends on the size of the training set and test set. This is why equivalent results for the training and testing set can have different overall averages. Also note that, while four different amounts of classes were tested for each classification, the focus is on the full-class and 2-class classification cases. The half-class and 3-

class cases usually do not perform as well as the others and would not be used as commonly either. As stated at the end of chapter 4.4, the success rates obtained for the 3-class and 2-class cases are based on the predicted CDR exactly matching the target CDR, while the success rates for the full and half-class cases are based on the predicted CDR being within 10% of the target CDR as in (6). All success rates are rounded to two decimal places. Finally, when training NNs in MATLAB, there is an element of randomness with the initial values. Therefore, multiple runs of training the same NN can give varying results. The results presented here are the best results that were obtained after three runs. This does not occur with the SVM or SRC algorithms used which give the same results every run. The most significant results are bolded.

TABLE VI. CLASSIFIER SUCCESS RATES FOR FEATURE: ALL PIXELS

	NN – GD Algorithm – 10 hidden nodes				SVM				SRC			
	12	6	3	2	12	6	3	2	12	6	3	2
Number of Classes	12	6	3	2	12	6	3	2	12	6	3	2
Average Success Rate (Training Set)	0.77	0.52	0.70	0.82	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Average Success Rate (Test Set)	0.47	0.58	0.56	0.86	0.62	0.66	0.60	0.90	0.53	0.62	0.62	0.88
Overall Average Success Rate	0.68	0.55	0.63	0.84	0.81	0.83	0.80	0.95	0.72	0.77	0.77	0.93

Table VI shows the results when using all the pixel intensities in an image as the feature to train and test with. This results in a feature vector with a length of 123 201 (351x351). As can be seen from the table, a very simple feature vector of all pixel intensities actually gives decent results, particularly for the SVM and SRC classifications. The best test set success rate obtained was 62% for the full amount of classes and 90% for the 2-class case. SRC classification performed slightly worse than SVM classification. NN classification was not effective for this feature, even for the training set. NN also took significantly longer to train, approximately 8 minutes, due to the high length of the feature vector. The memory requirements were over the maximum limit required by the BR and LM training algorithms.

TABLE VII. CLASSIFIER SUCCESS RATES FOR FEATURE: PCA COEFFICIENTS

	NN – BR Algorithm – 7 hidden nodes				SVM				SRC			
	12	6	3	2	12	6	3	2	12	6	3	2
Number of Classes	12	6	3	2	12	6	3	2	12	6	3	2
Average Success Rate (Training Set)	0.88	0.85	0.93	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Average Success Rate (Test Set)	0.60	0.70	0.65	0.92	0.62	0.66	0.60	0.90	0.67	0.68	0.63	0.90
Overall Average Success Rate	0.71	0.76	0.76	0.94	0.81	0.83	0.80	0.95	0.80	0.81	0.78	0.94

Table VII shows the results when using the PCA coefficients of an image as the feature to train and test with. This results in a feature vector with a length of one less than the number of images in the training set. This is a feature that was expected to obtain the highest success rates. However, it gave only slightly better results compared to simply using all of the pixel intensities as shown above. The best test set success rate was obtained using SRC and was 67% for the full amount of classes. Using NN, 92% was obtained for the 2-class case. NN performed better with this feature, but that is most likely because the BR training algorithm was used instead of the previously used GD algorithm.

TABLE VIII. CLASSIFIER SUCCESS RATES FOR FEATURE: IMAGE HISTOGRAM

	NN – BR Algorithm – 10 hidden nodes – 10 bins				SVM – 200 bins				SRC – 100 bins			
	12	6	3	2	12	6	3	2	12	6	3	2
Number of Classes	12	6	3	2	12	6	3	2	12	6	3	2
Average Success Rate (Training Set)	0.93	0.97	0.95	0.97	0.88	0.96	0.92	1.00	1.00	1.00	1.00	1.00
Average Success Rate (Test Set)	0.73	0.65	0.65	0.85	0.52	0.42	0.50	0.64	0.56	0.50	0.58	0.92
Overall Average Success Rate	0.85	0.84	0.83	0.92	0.70	0.69	0.71	0.82	0.78	0.75	0.79	0.96

Table VIII shows the results when using the image histogram of an image as the feature to train and test with. The length of this feature vector depends on how many bins the histogram is

divided into. Each classifier was tested with several different quantities of bins (10, 20, 50, 100, 200, and 500) and the best results were recorded. As shown in the table, NN worked best with 10 bins, SVM with 200 bins, and SRC with 100 bins. The best test set success rate that was obtained was 73% from NN classification. This is actually the overall highest test set success rate for the full-class case that is achieved across all the features that were tested. Also, with SRC a success rate of 92% was achieved for the 2-class classification. For this feature, SVM classification had the worst performance.

TABLE IX. CLASSIFIER SUCCESS RATES FOR FEATURE: PCA COEFFICIENTS USING HSV IMAGES

	NN – BR Algorithm – 10 hidden nodes				SVM				SRC			
Number of Classes	12	6	3	2	12	6	3	2	12	6	3	2
Average Success Rate (Training Set)	0.90	0.90	0.98	0.93	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Average Success Rate (Test Set)	0.50	0.32	0.62	0.92	0.58	0.54	0.66	0.84	0.50	0.45	0.60	0.90
Overall Average Success Rate	0.66	0.55	0.76	0.92	0.79	0.77	0.83	0.92	0.80	0.78	0.84	0.96

Table IX shows the results when using the PCA coefficients of the HSV colour model image as the feature to train and test with. The length of this feature vector is equal to one less than the number of training images. This test was done to see how well PCA coefficients for the HSV

version of the image would perform compared to PCA coefficients of the normal RGB image. As shown in the table, it performed worse with success rates around 10% lower for the full-class cases. The best test set success rate that was obtained was 58% from SVM classification for the full-class case and 92% from NN classification in the 2-class case.

TABLE X. CLASSIFIER SUCCESS RATES FOR FEATURE: COMBINED FEATURE VECTOR

	NN – BR Algorithm – 10 hidden nodes				SVM				SRC			
	12	6	3	2	12	6	3	2	12	6	3	2
Number of Classes	12	6	3	2	12	6	3	2	12	6	3	2
Average Success Rate (Training Set)	0.97	0.90	0.90	0.97	0.65	0.62	0.68	0.85	1.00	1.00	1.00	1.00
Average Success Rate (Test Set)	0.60	0.63	0.57	0.83	0.50	0.53	0.60	0.85	0.50	0.43	0.40	0.63
Overall Average Success Rate	0.86	0.82	0.80	0.93	0.59	0.58	0.65	0.85	0.85	0.83	0.82	0.89

The results for the final feature tested, the combined feature vector, are shown in Table X. As stated before, the length of this feature vector is 200. The hope was to obtain the best results by using this feature vector, since this would be used later for feature ranking. Better results would show that these features were strongly correlated to the predicted CDR. Unfortunately, the results were around the same as the other features, and slightly worse in some cases. As shown in the

table, the best test set success rate that was obtained was 60% from NN classification for the full-class case. This was actually the highest overall success rate for the full-class case with a performance of 86%, but the test set success rate is more important. It did not perform as well with the 2-class case however, achieving a maximum of only 85% using SVM classification. To obtain the results in Table X, histogram equalization was used on the 100 HSV features, which slightly improved the success rates over using no histogram equalization.

Another observation that is general to all of the above results is the high success of the training set classification compared to the test set. As shown in Tables VI to X, it is often in the 90+% range and even 100% in several cases. This could mean that the classifiers are being overfit and not generalized enough. Early stopping was tried when training the NNs but it did not help to improve the results significantly.

Only two of the best or more significant classifiers above were tested with target CDRs from all six ophthalmologists as testing every feature with every ophthalmologist would have taken a very long time. The first was the histogram feature vector shown in Table VIII. The results are shown in Table XI. This is for 10-bin histograms and using a BR NN classifier with 10 hidden nodes.

As can be seen, the results are fairly consistent between all of the ophthalmologists. An even higher success rate of 75% was obtained with ophthalmologist 4 for the full-class case. The highest success rate for the 2-class case remained at 85%. However, ophthalmologist 5 and 6 seemed to be more difficult to predict based on these results for both the full-class and 2-class classifications.

TABLE XI. CLASSIFIER SUCCESS RATES ACROSS ALL OPHTHALMOLOGISTS: HISTOGRAM

Number of Classes	Full Classes						2 Classes					
Ophthalmologist	1	2	3	4	5	6	1	2	3	4	5	6
Average Success Rate (Training Set)	0.93	0.93	0.90	0.95	0.98	0.97	0.97	0.98	0.95	0.95	0.97	0.93
Average Success Rate (Test Set)	0.73	0.65	0.65	0.75	0.50	0.65	0.85	0.80	0.83	0.75	0.70	0.55
Overall Average Success Rate	0.85	0.82	0.80	0.87	0.79	0.84	0.92	0.91	0.90	0.87	0.86	0.78

The second was with the combined feature vector. The results are shown in Table XII. This is also using a BR NN classifier with 10 hidden nodes.

TABLE XII. CLASSIFIER SUCCESS RATES ACROSS ALL OPHTHALMOLOGISTS: COMBINED FEATURE VECTOR

Number of Classes	Full Classes						2 Classes					
Ophthalmologist	1	2	3	4	5	6	1	2	3	4	5	6
Average Success Rate (Training Set)	0.97	0.93	0.94	0.99	0.96	0.90	0.97	0.96	0.96	0.99	0.97	0.97
Average Success Rate (Test Set)	0.60	0.50	0.50	0.57	0.43	0.60	0.83	0.83	0.73	0.77	0.60	0.73
Overall Average Success Rate	0.86	0.80	0.81	0.86	0.80	0.81	0.93	0.92	0.89	0.92	0.86	0.90

Again, with the combined feature vector, the results were generally consistent between all of the ophthalmologists which was expected since there were many more features. Therefore, even if some features that were good for one ophthalmologist were bad for another, there were more

options available. The best full-class success rate remained at 60% and the best 2-class success rate was 83%. Ophthalmologist 5 was once again the most difficult to predict.

The final method employed to evaluate these classifiers further was 10-fold cross-validation. Cross-validation is basically carried out by dividing the entire data set into 10 equal folds (or whatever amount of folds is desired). For this data set, each fold was comprised of 10 images. Then the classifier was trained with 9 of the 10 folds and tested with the remaining fold. This is then repeated so that each fold would be used as the test set exactly once. The purpose of using cross-validation is to see how well the classifier responds to unseen data. The results for the histogram feature and BR NN classifier with 10 hidden nodes are shown in Table XIII.

While some folds achieved high test set success rates of up to 80%, the overall average across all the folds was 68%. This is slightly lower than the best success rate achieved of 73% shown in Table VIII meaning that with unseen data the results might be slightly lower.

TABLE XIII. 10-FOLD CROSS-VALIDATION: HISTOGRAM

Fold	Average Success Rate (Training Set)	Average Success Rate (Test Set)	Overall Average Success Rate
1	0.96	0.70	0.93
2	0.90	0.70	0.88
3	0.97	0.60	0.93
4	0.88	0.80	0.87
5	0.91	0.60	0.88
6	0.93	0.80	0.92
7	0.97	0.80	0.95
8	0.92	0.60	0.89
9	0.93	0.60	0.90
10	0.94	0.60	0.91
Overall Average	0.93	0.68	0.91

The 10-fold cross-validation for the combined feature vector and a BR NN classifier with 10 hidden nodes is shown in Table XIV.

TABLE XIV. 10-FOLD CROSS-VALIDATION: COMBINED FEATURE VECTOR

Fold	Average Success Rate (Training Set)	Average Success Rate (Test Set)	Overall Average Success Rate
1	0.93	0.60	0.90
2	0.93	0.80	0.92
3	0.90	0.40	0.85
4	0.93	0.70	0.91
5	0.93	0.60	0.90
6	0.96	0.70	0.93
7	0.90	0.60	0.87
8	0.93	0.40	0.88
9	0.91	0.30	0.85
10	0.91	0.50	0.87
Overall Average	0.92	0.56	0.89

As with cross-validation on the histogram feature, the overall average across all folds for the combined feature vector is slightly lower than the maximum success rate achieved of 60% shown in Table X. This is mainly due to about four of the folds performing badly with rates of 30-40%. As stated earlier, multiple runs of training the same NN can give varying results. If the process was repeated several times and the best classifier results taken, the overall success may have been improved.

5.2 *Comparison of Ophthalmologists Using Feature Ranking*

The next set of results are concerned with the feature ranking stage described in detail in Chapter 4.5. Once the combined feature vector is created during the feature extraction stage, it is ready to be ranked. In order to perform the chi square and information gain methods, Weka 3.6.13 was used. This machine learning software has functions built in for a variety of feature ranking algorithms including chi square and information gain. The feature vector is loaded and the target is set as the CDR values for a specific ophthalmologist. Weka allows the feature ranking to be performed with the normal training set or to be cross-validated over the entire data set with a desired number of folds. For these results, the option for 10-fold cross-validation was used as this helps generalize the results better.

In Table XV below, the top 100 features by rank for each ophthalmologist obtained through the use of the chi square algorithm are listed. Only the top 100 features are listed as they are the most significant. Only the results for chi square are shown as information gain gave the exact same feature ranking. The name of each feature is broken into four parts: location in feature vector, feature, colour model, and number for multipart features. For example, “200 histHSV10” means that it was the 200th feature in the 200 length feature vector as shown in Fig 15 above, it was the histogram feature for the HSV image, and it was the 10th bin of the histogram.

TABLE XV. TOP 100 FEATURES USING CHI SQUARE FEATURE RANKING

Rank	Ophth1	Ophth2	Ophth3	Ophth4	Ophth5	Ophth6
1	200 histHSV10	200 histHSV10	200 histHSV10	33 pcaG3	68 histR8	18 entropyRGB
2	63 histR3	63 histR3	66 histR6	9 stdG	67 histR7	8 entropyG
3	67 histR7	67 histR7	67 histR7	68 histR8	69 histR9	66 histR6
4	66 histR6	66 histR6	63 histR3	67 histR7	65 histR5	67 histR7
5	65 histR5	65 histR5	65 histR5	69 histR9	66 histR6	68 histR8
6	68 histR8	68 histR8	68 histR8	66 histR6	70 histR10	65 histR5
7	69 histR9	69 histR9	69 histR9	70 histR10	72 histG2	69 histR9
8	70 histR10	70 histR10	70 histR10	73 histG3	71 histG1	70 histR10
9	72 histG2	72 histG2	72 histG2	64 histR4	63 histR3	73 histG3
10	73 histG3	73 histG3	64 histR4	72 histG2	73 histG3	71 histG1
11	71 histG1	71 histG1	71 histG1	65 histR5	64 histR4	200 histHSV10
12	64 histR4	64 histR4	62 histR2	71 histG1	54 pcaRGB4	72 histG2
13	62 histR2	62 histR2	73 histG3	200 histHSV10	56 pcaRGB6	64 histR4
14	54 pcaRGB4	61 histR1	61 histR1	63 histR3	62 histR2	90 histB10
15	61 histR1	199 histHSV9	54 pcaRGB4	55 pcaRGB5	74 histG4	63 histR3
16	53 pcaRGB3	54 pcaRGB4	53 pcaRGB3	75 histG5	57 pcaRGB7	75 histG5
17	55 pcaRGB5	53 pcaRGB3	199 histHSV9	56 pcaRGB6	61 histR1	55 pcaRGB5
18	56 pcaRGB6	52 pcaRGB2	52 pcaRGB2	57 pcaRGB7	55 pcaRGB5	56 pcaRGB6
19	57 pcaRGB7	55 pcaRGB5	55 pcaRGB5	62 histR2	75 histG5	54 pcaRGB4
20	60 pcaRGB10	56 pcaRGB6	56 pcaRGB6	58 pcaRGB8	58 pcaRGB8	14 stdB
21	199 histHSV9	57 pcaRGB7	57 pcaRGB7	10 varG	59 pcaRGB9	57 pcaRGB7
22	59 pcaRGB9	60 pcaRGB10	74 histG4	61 histR1	200 histHSV10	58 pcaRGB8
23	75 histG5	59 pcaRGB9	60 pcaRGB10	59 pcaRGB9	60 pcaRGB10	62 histR2
24	74 histG4	74 histG4	59 pcaRGB9	60 pcaRGB10	76 histG6	59 pcaRGB9
25	58 pcaRGB8	58 pcaRGB8	58 pcaRGB8	54 pcaRGB4	80 histG10	74 histG4
26	52 pcaRGB2	75 histG5	75 histG5	74 histG4	10 varG	61 histR1
27	76 histG6	76 histG6	76 histG6	76 histG6	92 histRGB2	60 pcaRGB10
28	92 histRGB2	92 histRGB2	92 histRGB2	77 histG7	91 histRGB1	76 histG6
29	77 histG7	77 histG7	91 histRGB1	92 histRGB2	77 histG7	92 histRGB2
30	91 histRGB1	91 histRGB1	90 histB10	91 histRGB1	94 histRGB4	78 histG8
31	90 histB10	90 histB10	93 histRGB3	90 histB10	93 histRGB3	91 histRGB1
32	93 histRGB3	93 histRGB3	77 histG7	93 histRGB3	95 histRGB5	15 varB
33	94 histRGB4	94 histRGB4	87 histB7	52 pcaRGB2	79 histG9	94 histRGB4
34	95 histRGB5	95 histRGB5	94 histRGB4	43 pcaB3	53 pcaRGB3	93 histRGB3
35	98 histRGB8	98 histRGB8	78 histG8	94 histRGB4	78 histG8	95 histRGB5
36	97 histRGB7	97 histRGB7	95 histRGB5	95 histRGB5	98 histRGB8	77 histG7
37	96 histRGB6	96 histRGB6	98 histRGB8	98 histRGB8	90 histB10	98 histRGB8
38	78 histG8	89 histB9	97 histRGB7	97 histRGB7	9 stdG	97 histRGB7

39	89 histB9	88 histB8	96 histRGB6	96 histRGB6	97 histRGB7	88 histB8
40	88 histB8	87 histB7	89 histB9	89 histB9	89 histB9	96 histRGB6
41	87 histB7	79 histG9	88 histB8	88 histB8	20 varRGB	89 histB9
42	79 histG9	86 histB6	79 histG9	19 stdRGB	96 histRGB6	80 histG10
43	81 histB1	78 histG8	81 histB1	87 histB7	52 pcaRGB2	87 histB7
44	80 histG10	81 histB1	80 histG10	78 histG8	88 histB8	79 histG9
45	82 histB2	80 histG10	86 histB6	79 histG9	81 histB1	52 pcaRGB2
46	51 pcaRGB1	82 histB2	82 histB2	81 histB1	87 histB7	81 histB1
47	86 histB6	85 histB5	85 histB5	80 histG10	82 histB2	53 pcaRGB3
48	85 histB5	84 histB4	84 histB4	82 histB2	18 entropyRGB	82 histB2
49	84 histB4	83 histB3	83 histB3	86 histB6	199 histHSV9	86 histB6
50	83 histB3	51 pcaRGB1	51 pcaRGB1	53 pcaRGB3	83 histB3	85 histB5
51	50 pcaB10	50 pcaB10	50 pcaB10	85 histB5	86 histB6	83 histB3
52	16 meanRGB	49 pcaB9	49 pcaB9	84 histB4	85 histB5	84 histB4
53	15 varB	24 pcaR4	24 pcaR4	83 histB3	84 histB4	51 pcaRGB1
54	17 energyRGB	16 meanRGB	16 meanRGB	50 pcaB10	51 pcaRGB1	50 pcaB10
55	49 pcaB9	15 varB	15 varB	20 varRGB	16 meanRGB	19 stdRGB
56	14 stdB	14 stdB	14 stdB	17 energyRGB	17 energyRGB	17 energyRGB
57	24 pcaR4	17 energyRGB	17 energyRGB	23 pcaR3	19 stdRGB	21 pcaR1
58	18 entropyRGB	18 entropyRGB	18 entropyRGB	16 meanRGB	2 energyR	20 varRGB
59	9 stdG	19 stdRGB	19 stdRGB	51 pcaRGB1	21 pcaR1	10 varG
60	19 stdRGB	22 pcaR2	22 pcaR2	18 entropyRGB	22 pcaR2	22 pcaR2
61	22 pcaR2	21 pcaR1	21 pcaR1	21 pcaR1	24 pcaR4	16 meanRGB
62	21 pcaR1	20 varRGB	20 varRGB	15 varB	3 entropyR	24 pcaR4
63	20 varRGB	13 entropyB	13 entropyB	22 pcaR2	50 pcaB10	13 entropyB
64	13 entropyB	12 energyB	12 energyB	14 stdB	15 varB	23 pcaR3
65	12 energyB	11 meanB	11 meanB	13 entropyB	14 stdB	25 pcaR5
66	3 entropyR	3 entropyR	3 entropyR	3 entropyR	13 entropyB	12 energyB
67	11 meanB	2 energyR	2 energyR	24 pcaR4	23 pcaR3	3 entropyR
68	2 energyR	5 varR	5 varR	199 histHSV9	25 pcaR5	2 energyR
69	5 varR	4 stdR	4 stdR	2 energyR	7 energyG	199 histHSV9
70	4 stdR	6 meanG	6 meanG	5 varR	5 varR	5 varR
71	6 meanG	10 varG	10 varG	4 stdR	12 energyB	4 stdR
72	10 varG	9 stdG	9 stdG	12 energyB	4 stdR	27 pcaR7
73	25 pcaR5	8 entropyG	8 entropyG	6 meanG	11 meanB	6 meanG
74	23 pcaR3	7 energyG	7 energyG	25 pcaR5	6 meanG	11 meanB
75	8 entropyG	23 pcaR3	23 pcaR3	11 meanB	8 entropyG	9 stdG
76	7 energyG	25 pcaR5	25 pcaR5	8 entropyG	49 pcaB9	7 energyG
77	26 pcaR6	48 pcaB8	48 pcaB8	7 energyG	26 pcaR6	26 pcaR6

78	48 pcaB8	26 pcaR6	26 pcaR6	26 pcaR6	27 pcaR7	49 pcaB9
79	41 pcaB1	41 pcaB1	41 pcaB1	42 pcaB2	99 histRGB9	43 pcaB3
80	40 pcaG10	40 pcaG10	40 pcaG10	45 pcaB5	100 histRGB10	42 pcaB2
81	42 pcaB2	39 pcaG9	39 pcaG9	44 pcaB4	42 pcaB2	28 pcaR8
82	39 pcaG9	42 pcaB2	42 pcaB2	27 pcaR7	43 pcaB3	45 pcaB5
83	43 pcaB3	43 pcaB3	43 pcaB3	46 pcaB6	44 pcaB4	44 pcaB4
84	44 pcaB4	44 pcaB4	44 pcaB4	41 pcaB1	28 pcaR8	46 pcaB6
85	47 pcaB7	47 pcaB7	47 pcaB7	49 pcaB9	48 pcaB8	40 pcaG10
86	46 pcaB6	46 pcaB6	46 pcaB6	40 pcaG10	45 pcaB5	48 pcaB8
87	45 pcaB5	45 pcaB5	45 pcaB5	47 pcaB7	40 pcaG10	41 pcaB1
88	38 pcaG8	38 pcaG8	38 pcaG8	48 pcaB8	41 pcaB1	47 pcaB7
89	27 pcaR7	37 pcaG7	37 pcaG7	39 pcaG9	46 pcaB6	39 pcaG9
90	37 pcaG7	36 pcaG6	36 pcaG6	28 pcaR8	47 pcaB7	29 pcaR9
91	28 pcaR8	28 pcaR8	28 pcaR8	29 pcaR9	39 pcaG9	31 pcaG1
92	36 pcaG6	27 pcaR7	27 pcaR7	100 histRGB10	29 pcaR9	30 pcaR10
93	30 pcaR10	30 pcaR10	30 pcaR10	30 pcaR10	30 pcaR10	33 pcaG3
94	29 pcaR9	29 pcaR9	29 pcaR9	38 pcaG8	31 pcaG1	32 pcaG2
95	31 pcaG1	31 pcaG1	31 pcaG1	31 pcaG1	38 pcaG8	38 pcaG8
96	35 pcaG5	35 pcaG5	35 pcaG5	32 pcaG2	37 pcaG7	34 pcaG4
97	34 pcaG4	34 pcaG4	34 pcaG4	34 pcaG4	32 pcaG2	37 pcaG7
98	32 pcaG2	33 pcaG3	33 pcaG3	37 pcaG7	36 pcaG6	36 pcaG6
99	33 pcaG3	32 pcaG2	32 pcaG2	36 pcaG6	33 pcaG3	35 pcaG5
100	99 histRGB9	99 histRGB9	99 histRGB9	35 pcaG5	34 pcaG4	99 histRGB9

Based on the results from the above table, it can be seen that the feature ranking is similar between the ophthalmologists. Ophthalmologists 4, 5, and 6 have the most variance in feature ranks, followed by ophthalmologist 3, and then by Ophthalmologists 1 and 2 which are the most similar. This may be why ophthalmologist 5 and sometimes ophthalmologist 6 had lower success rates than the others in Tables XI and XII above. Overall the structure is similar between all of the ophthalmologists. In order to make it easier to see ranks, a simplified version of the results is presented in Table XVI. The main difference is that multi-part features are grouped as one. Therefore, if 5 out of 10 red PCA features are close together, all 10 of the red PCA features are grouped as one around the rank that the group appears in. Some of the multipart HSV features appear alone so they are mentioned specifically.

TABLE XVI. REDUCED FEATURE RANK LIST

Rank	Ophth1	Ophth2	Ophth3	Ophth4	Ophth5	Ophth6
1	histHSV10	histHSV10	histHSV10	stdG	histR	entropyRGB
2	histR	histR	histR	histR	histG	entropyG
3	histG	pcaRGB	pcaRGB	histHSV10	pcaRGB	histR
4	pcaRGB	histG	histHSV9	histG	histHSV10	histHSV10
5	histHSV9	histHSV9	histG	pcaRGB	varG	histG
6	histRGB	histRGB	histRGB	varG	histRGB	stdB
7	histB	histB	histB	histRGB	stdG	pcaRGB
8	meanRGB	meanRGB	meanRGB	histB	varRGB	varB
9	varB	varB	varB	stdRGB	histB	histRGB
10	energyRGB	stdB	stdB	varRGB	entropyRGB	histB
11	stdB	energyRGB	energyRGB	energyRGB	histHSV9	stdRGB
12	entropyRGB	entropyRGB	entropyRGB	meanRGB	meanRGB	energyRGB
13	stdG	stdRGB	stdRGB	entropyRGB	energyRGB	varRGB
14	stdRGB	varRGB	varRGB	varB	stdRGB	varG
15	pcaR	entropyB	entropyB	stdB	energyR	meanRGB
16	varRGB	energyB	energy	entropyB	entropyR	entropyB
17	entropyB	meanB	meanB	entropyR	varB	pcaR
18	energyB	entropyR	entropyR	pcaR	stdB	energyB
19	entropyR	energyR	energy	histHSV9	entropyB	entropyR
20	meanB	varR	varR	energyR	pcaR	energyR
21	energyR	stdR	stdR	varR	energyG	histHSV9
22	varR	meanG	meanG	stdR	varR	varR
23	stdR	varG	varG	energyB	energyB	stdR
24	meanG	stdG	stdG	meanG	stdR	meanG
25	varG	entropyG	entropyG	meanB	meanB	meanB
26	entropyG	energyG	energy	entropyG	meanG	stdG
27	energyG	pcaR	pcaR	energyG	entropyG	energyG
28	pcaB	pcaB	pcaB	pcaB	pcaB	pcaB
29	pcaG	pcaG	pcaG	pcaG	pcaG	pcaG

Table XVI shows that the red channel histogram is commonly the most important feature as it is always ranked in one of the top 3 spots for every ophthalmologist. The 10th bin of the HSV histogram is also always in one of the top 4 spots, and in the top spot for three of the ophthalmologists. Some other consistently important features are the green channel histogram and the RGB PCA coefficients. Some of the lowest ranked features include the green and blue channel PCA coefficients, and the green channel energy, entropy, mean and variance, except with ophthalmologist 6 where the green channel entropy is one of the top ranked features. Based on the

table, ophthalmologist 4, 5, and 6 have the most differences compared to the others. Some examples are the standard deviation and variance of the green channel being ranked high for ophthalmologist 4 and the RGB and green channel entropy features ranked high for ophthalmologist 6. These same features are ranked low for all of the others. Note that most of the HSV features were not ranked in the top 100, but this is because the feature vector was already ordered with the 100 RGB features first and then the 100 HSV features. Upon ranking the feature vector when the RGB and HSV features were switched, the opposite was true and most of the RGB features were not ranked in the top 100. This is probably because the RGB and HSV features carry a lot of redundant and similar information.

Overall, the feature ranking across the ophthalmologists was similar with only a few clear differences throughout. As mentioned, the top 2 or 3 features are very different for some of the ophthalmologists. This list of features does not seem to be enough to truly distinguish between each of the ophthalmologists and determine why they sometimes assign very different CDR values to the same fundus image. Most of these features are not entirely visible. It is possible to say that some ophthalmologists might be able to see the overall entropy or intensity changes in an image clearer than others, but it would be helpful to test with more physically apparent features in the future such as lengths and areas of the cup and disc regions. As stated earlier, it was assumed that these were highly dependent on an ophthalmologist's manual marking of the cup and disc regions. However, the physical features from an automatically segmented image could be used as absolute standard values and the ophthalmologist dependent values could be compared to them for the sake of feature ranking.

The feature ranking did show that different ophthalmologists had different overall rankings. It also showed that a mix of different features is useful since, as shown in Table XVI, they are

ranked in a mixed order. For example, the top features are not all from one colour channel or they are not all histogram features. To see if the feature ranking was effective, a BR NN classifier with 10 hidden nodes was tested to see how the results would change. The results are shown in Table XVII. For the full-class and 2-class cases, testing was done with reduced versions of the combined feature vector which were rearranged according to the feature ranks shown in Table XV. For example, when testing with 100 features, the feature vector would be the top 100 features as shown in Table XV. For a number of features equal to 50, the feature vector is the top 50 features as shown and so on.

TABLE XVII. CLASSIFICATION USING REDUCED SETS OF FEATURES

Number of Classes	Full Classes				2 Classes					
	200	100	50	25	200	100	50	25	10	5
Average Success Rate (Training Set)	0.97	0.90	0.93	0.96	0.97	1.00	0.99	0.97	0.99	0.99
Average Success Rate (Test Set)	0.60	0.63	0.67	0.50	0.83	0.80	0.80	0.80	0.73	0.63
Overall Average Success Rate	0.86	0.82	0.85	0.82	0.93	0.94	0.93	0.92	0.91	0.88

As can be seen, having a reduced set of ranked features in the full-class case, even up to 25% of the total amount of features, gives similar results. The test set success rate is actually higher for 100 and 50 features. This could probably be attributed to noise that the extra features create and also an increased focus on the most important features. Around 25 features, or 12.5% of the total, the success rates start to decrease showing that too few features does not have enough information. In the 2-class case, the results are also similar. However, none of the reduced feature amounts manage to increase the test set classification success rate. Since 2-class classification is simpler and depends on fewer features, the success rate actually does not start to decrease until around 10 features, or 5% of the total number of features, remain. This shows that a fairly accurate decision can still be made with just a few features.

5.3 Comparison With Similar Papers in the Literature

As mentioned in Chapter 2, only two papers were found that did something similar to what is presented here and they were written 4-8 years ago. Table XVIII below shows a comparison of some the 2-class classification results from this proposed technique with those papers. Note that the results obtained in this paper are based on predicting whether or not a fundus image contains a CDR above or below 0.50 while their results are based on classifying whether or not a fundus image is glaucomatous or not. As stated earlier, a CDR above 0.50, while having an increased chance of being glaucomatous, is not entirely indicative of the presence of glaucoma. Other factors such as intraocular pressure and optic nerve condition must also be considered. Therefore, the results cannot be directly compared but they are presented since similar methodologies are followed and since the CDR is still a major factor in determining the presence of glaucoma from fundus images. Unfortunately, the data set used in this paper did not come with the images

labelled as glaucomatous or not. Otherwise, classification could have been performed based on that and a more direct comparison could be provided. Note that histograms are compared side by side between [1] and the proposed technique given here. However, in [1] PCA is used on the image histograms while in this paper, histogram bin values are used directly. Also, the combined feature vector is different between each paper and the proposed technique. In [1], the feature vector is comprised of pixel intensities, texture features, FFT coefficients, and histograms and in [2] it is comprised of texture and higher order spectra features. The combined feature vector for the proposed technique is described as in Fig 15 above.

TABLE XVIII. COMPARISON WITH SIMILAR PAPERS

	Success Rate (%)		
Method	Paper [1]	Paper [2]	Proposed Technique
PCA on Intensities	81.0	-	92.0
Histogram	61.0	-	92.0
Combined Feature Vector	80.0	91.7	85.0

As shown, the proposed technique performs better in most cases where the same or similar features were used between papers, except for the combined feature vector comparison. Although higher success rates were achieved with 2-class classification for some of the individual features tested, the combined feature vector did not perform as well. While the 2-class classification results are similar, the main advantage of this research over [1] and [2] was to attempt multi-class

classification with more than two classes. In future work, it would be useful to try some texture and higher order spectra features, mainly for the CDR prediction part of the project.

CHAPTER 6 – CONCLUSION

6.1 Accomplishments

There were two main goals to accomplish during this project. The first was to see how accurately predictions could be made for a given ophthalmologist's assigned CDR value to a fundus image. The second was to determine the contributions of different features to a given ophthalmologist and how these features contributed differently between ophthalmologists. This was to get an idea about why different ophthalmologists sometimes assigned very different CDR values to the exact same fundus image. Overall, the results were satisfactory. Prediction success rates in the 70-75% range were achieved for a full number of CDR classes and the 85-92% range for 2-class classification. While better results were expected for the full amount of classes, the methodology laid out here is a good start and there are many more ideas to try in the future. The method used to compare the feature ranking between different ophthalmologists also worked as intended. However, more apparent features need to be tested. Finally, this project has opened the way to interesting biological and psychological understanding of the ophthalmologists and CDRs on top of the normal machine learning and image processing perspective as will be discussed in the next section.

6.2 Future Work

For future work on this project, there are many ideas to try. The first is simply to try more image features. In this project, most of the common and simple pixel intensity features were tested. However, as shown earlier, there are many advanced feature extraction methods, both

related to optic fundus images and to images in general. Some that might be useful to try are the motion patterns from [3], superpixels which is shown to be a very popular approach for segmentation in [4] and used in several research papers, and quaternion features described in [18]. There is a large amount of image features that can be tested. The challenge would be determining which would be most relevant to this problem.

As stated as a project challenge in Chapter 1, only a fairly small data set of 100 images for each ophthalmologist was used. In the future, it would be great to test with a much larger data set. This would help train a much more general classifier and gain more feature information for each class. Some classes during this project had only one or two images to represent them so a larger data set would result in better representation for each class. There would also be more image variation in a larger data set so any image anomalies and quality variation would be better addressed and result in classifiers that are better for a wide variety of real world data. Recent research papers also need to be examined to see methods of improving multi-class classification where the number of classes is fairly high.

The next idea is to do more pre-processing on the fundus images. As stated in the assumptions shown in Chapter 1, the aim was to not perform too many steps that would lead into segmentation or to perform segmentation itself. However, it would be interesting to see how more physical features such as disc/cup length, width, area and distance would be ranked by each ophthalmologist. These features could only be obtained through segmentation or object recognition of the cup and disc regions. Another pre-processing step that might improve the results is artery removal. As shown in [1] and [19], there are various methods for removing the blood vessels from the fundus image and doing this helps to improve the segmentation results. It

would also help in this project since the vessels often cause too much variance in the image pixel intensities.

The final idea for future testing is to have actual features related to both the patient and ophthalmologist in question. For example, the patient's age, gender, country of residence, and ethnicity might have an effect on their CDR. Furthermore, when the ophthalmologist has this information, the CDR value they assign to the patient's fundus image may be affected. A recent study has done something similar to this. In [20], a study is done to see how various features such as age, gender, and even physiological features such as blood pressure, smoking status, alcohol intake, height, and weight affect the CDRs of a population of people. They found that higher age, males, and high blood pressure are more correlated to higher CDRs. Therefore, features such as these would be helpful to take into account. Finally, some personal features about the ophthalmologist such as their age, gender, years of work experience, ethnicity, country of residence, and even personality type might also have an effect on how they determine the CDR they assign to a particular fundus image. It would be very interesting to test and rank these features to see what effect, if any, they have on the assigned CDR. Currently, this project has only taken into account the fundus images, which approaches the problem from a machine learning, image processing, and possibly biological perspective. Using these additional features would give a valuable psychological and social perspective on the determination of CDRs, which should be helpful since the variance in CDRs is due to the ophthalmologists themselves.

REFERENCES

- [1] R. Bock, J. Meier, G. Michelson, L. G. Nyul, and J. Hornegger. (2007). *Classifying Glaucoma with Image-based Features from Fundus Photographs*. [Online]. Available: <http://www5.informatik.uni-erlangen.de/Forschung/Publikationen/2007/Bock07-CGW.pdf>
- [2] U. R. Acharya, S. Dua, X. Du, V. Sree S, and C. K. Chua. (2011, May). *Automated Diagnosis of Glaucoma Using Texture and Higher Order Spectra Features*. [Online]. Available: <http://ieeexplore.ieee.org.ezproxy.lib.ryerson.ca/stamp/stamp.jsp?tp=&arnumber=5720314>
- [3] K. S. Deepak, and J. Sivaswamy. (2012, March). *Automatic Assessment of Macular Edema From Colour Retinal Images*. [Online]: Available: <http://ieeexplore.ieee.org.ezproxy.lib.ryerson.ca/stamp/stamp.jsp?tp=&arnumber=6097060>
- [4] P. S. J. Kumar, and S. Banerjee. (2014, December). *A Survey on Image Processing Techniques for Glaucoma Detection*. [Online]. Available: http://www.researchgate.net/publication/277714514_A_Survey_on_Image_Processing_Techniques_for_Glaucoma_Detection
- [5] A. Ramaswamy, K. Ram, N. Joshi, and M. Sivaprakasam. *A Polar Map Based Approach Using Retinal Fundus Images for Glaucoma Detection*. [Online]. Available: <http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1038&context=omia>
- [6] N. Dey, A. B. Roy, A. Das, and S. S. Chaudhuri. (2012, July 26-28). *Optical Cup to Disc Ratio Measurement for Glaucoma Diagnosis Using Harris Corner*. [Online]. Available: <http://ieeexplore.ieee.org.ezproxy.lib.ryerson.ca/stamp/stamp.jsp?tp=&arnumber=6395971>

- [7] N. Halder, D. Roy, A. Chattaraj, and T. Chowdhury. (2015, March). *Identification of Cup-Disk Ratio for Glaucoma Prone Eyes*. [Online]. Available: http://www.ijircce.com/upload/2015/march/47_Identification.pdf
- [8] Mandeep Singh, Mooninder Singh, and J. K. Virk. (2015). *A simple approach to Cup-to-Disk Ratio determination for Glaucoma Screening*. [Online]. Available: <http://csjournals.com/IJCSC/PDF6-2/18.%20Jiwan.pdf>
- [9] Y. Chang, and C. Lin. (2008). *Feature Ranking Using Linear SVM*. [Online]. Available: <http://www.jmlr.org/proceedings/papers/v3/chang08a/chang08a.pdf>
- [10] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi. (2014, September). *DREAM: Diabetic Retinopathy Analysis Using Machine Learning*. [Online]. Available: <http://ieeexplore.ieee.org.ezproxy.lib.ryerson.ca/stamp/stamp.jsp?tp=&arnumber=6680633>
- [11] L. Zhang, M. Yang, and X. Feng. (2011). *Sparse Representation or Collaborative Representation: Which Helps Face Recognition?*. [Online]. Available: <http://ieeexplore.ieee.org.ezproxy.lib.ryerson.ca/stamp/stamp.jsp?tp=&arnumber=6126277>
- [12] B. Gaonkar. (2012, March 22). *Sparse representations classifier*. [Online]. Available: <http://www.mathworks.com/matlabcentral/fileexchange/35813-sparse-representations-classifier>
- [13] CVX Research. (2012). *CVX: Matlab Software for Disciplined Convex Programming*. [Online]. Available: <http://cvxr.com/cvx/>
- [14] Andale. (2013, July 31). *Chi-Square Statistic: How to calculate it*. [Online]. Available: <http://www.statisticshowto.com/what-is-a-chi-square-statistic/>
- [15] A. W. Moore. (2003). *Information Gain*. [Online]. Available: <http://www.autonlab.org/tutorials/infogain11.pdf>

- [16] D. F. Gillies. (2011, September 21). *Lecture 15: Principal Component Analysis*. [Online]. Available: <http://www.doc.ic.ac.uk/~dfg/ProbabilisticInference/IDAPILecture15.pdf>
- [17] Machine Learning Group at the University of Waikato. (2015). *Downloading and Installing Weka*. [Online]. Available: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>
- [18] R. Zeng, J. Wu, Z. Shao, L. Senhadji, and H. Shu.. (2014, December 4). *Quaternion softmax classifier*. [Online]. Available: <http://ieeexplore.ieee.org.ezproxy.lib.ryerson.ca/stamp/stamp.jsp?tp=&arnumber=6975762>
- [19] A. Osareh, and B. Shadgar. (2009). *Automatic Blood Vessel Segmentation In Colour Images Of Retina*. [Online]. Available: http://www.sid.ir/en/vewssid/j_pdf/8542009b205.pdf
- [20] Y. J. Kim, J. M. Kim, S. H. Shim, J. H. Bae, and K. H. Park. (2015, April 15). *Associations between Optic Cup-to-disc Ratio and Systematic Factors in the Healthy Korean Population*. [Online]. Available: <http://ekjo.org/Synapse/Data/PDFData/0065KJO/kjo-29-336.pdf>