AUTOMATED CLASSIFICATION OF RETINAL DISEASES IN STARE DATABASE

Author
TOOBA
Registration Number
00000117334

Supervisor
DR. ADEEB SHEHZAD

DEPARTMENT OF BIOMEDICAL ENGINEERING AND SCIENCES

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD, PAKISTAN.

JUNE, 2018.

Automated Classification of Retinal Diseases in STARE Database

Author

Tooba

Registration Number

00000117334

A thesis submitted in partial fulfillment of the requirements for the degree of MS Biomedical Sciences

Thesis Supervisor

Dr. Adeeb Shehzad

Thesis Supervisor's Signature: _____

DEPARTMENT OF BIOMEDICAL ENGINEERING AND SCIENCES
SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,
ISLAMABAD, PAKISTAN.

JUNE, 2018.

Thesis Acceptance Certificate

It is certified that the final copy of MS Thesis written by Ms. Tooba (Registration No. 00000117334), of MS Biomedical Sciences, SMME, has been vetted by undersigned, found complete in all respects as per NUST statutes / regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/Phil Degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in this dissertation.

Signature:_____

Name of Supervisor: Dr. Adeeb Shehzad

Date:_____

Signature (HOD):_____

Date:_____

Signature (Principal):_____

Date:_____

Declaration

I, Tooba, certify that this research work titled "Automated Classification of Retinal Diseases in STARE Database" is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources has been properly acknowledged / referred.

Signature of Student
TOOBA
00000117334

Copyright Statement

Acknowledgements

In the name of ALLAH, the most Gracious and the most Merciful, all praise is for Him. He is the one who blessed man with Wisdom and Knowledge. I shower my thanks to Almighty Creator for flourishing my thoughts and helping me in my all efforts to complete this research work.

Many Darood on the Last Prophet Muhammad (peace be upon him) whose teachings have always been source of light in journey of life.

I would also like to express special thanks to my research supervisor Dr. Adeeb Shehzad and co-supervisor Dr. Omer Gilani for giving me opportunity, guidance, his kind support and encouragement from the beginning till the end period of completion of my thesis. I would like to pay my sincere thanks to all my teachers at SMME, NUST, for their cooperation during the course of my study of M.S.

I am also grateful to my research fellows and friends especially Naireen Zaheer and senior Zaid Ahsan for their support and cooperation. Without their help it wouldn't be possible for me to complete my thesis. I appreciate their patience and guidance throughout my thesis phase.

This note of acknowledgement will be incomplete without appreciating the efforts and boundless cooperation, love and prayers of my respectable parents and other family members in every phase of life.

May Allah bless the above mention personalities with honor and success, Ameen.

Tooba

Dedication

I dedicate this piece of work to my parents, sister and brother for their immense support and cooperation that led me to this wonderful accomplishment.

Abstract

People of all age groups are affected by a number of retinal diseases. These diseases are identified by conducting different medical examinations primary of which are visual examinations. One of the key issues in visual diagnosis of diseases is the human error due to poor decision making, for that a number of research projects are conducted which use the visual data and directly or the symptoms and generate decisions. An upcoming inter-disciplinary technology named Computer Aided Medical Diagnostic System provides precise detection and prediction of disease. Automated image analysis methods are far more helpful for early identification and evaluation of disease as compared to cryptic and time taking manual techniques of digital medical imaging. This thesis aims to develop an automated method for identification of eye disorders that affect the human retina which if left unidentified may result in blindness due to delayed detection and analysis. Image data was acquired by publically available STARE Database having fundus images and by implementation of exclusion and inclusion criterion it was pre-processed on MATLAB. Initial pre-processing increased the significance of the data to be analyzed. From 186 images, 16 diseases and 22 features were deduced. A support vector machine classifier was used for automated identification and classification, resulting in an accuracy of 94% and specificity of 98%. In the chosen technique sensitivity, specificity and accuracy of the results was affected by the problem of one-sided data. For the reduction of dimensionality of data (redundancy reduction) principal component analysis was employed. 5, 10 and 22 Principal components were obtained to reduce the amount of variables. PCA was performed prior to training of the SVM, results for different data dimensionality were compared for completeness.

Keywords: *Image processing, support vector machine, principal component analysis, MATLAB, fundus images, retinal diseases*

Table of Contents

List of Figures

List of Tables

Chapter 1: INTRODUCTION

1.1 Anatomy of the Human Eye

Retina of human is composed cells known as rods and cones that are responsible for the image and color identification. The neural layer connected to the retina through the optic nerve serves the most pivotal part in the human vision (Rath, 2017). In the structure of the human eye, blood vessels, optic disc, fovea and macula are fundamental elements which form the retina. To avoid confusion in the terminology the Latin terms for the Right and Left Eye (Oculus Dexter and Oculus Sinister respectively) are dropped in favor of English terms. The abbreviation for the Optic Disc is OD, in the figure below, the brightest parts shown as curved orange and pale pink form Optic Disc (OD).

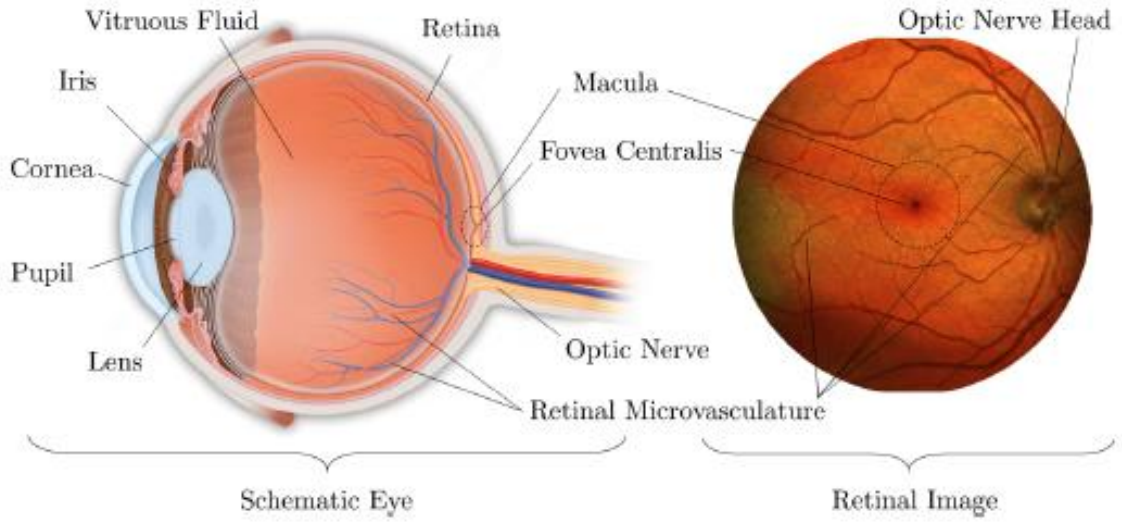


Figure 1.1 The Anatomy of the Human Eye

The focal point of the OD is also called the blind spot because it does not have any photoreceptor(Nirmala, Nath, & Dandapat, 2011). The OD is also the convergence point of the blood vessels in the eye. The macula lutea is an oval shaped dim spot around 0.4mm in width located parallel to blind spot in eye. The center of the macula, about the size of a pin head is called Fovea Centralis and it has only cones to assist in sharp vision(Wu et al., 2016). Most of the cones are present in macula and its thickness gradually reduces progressing towards the periphery. The blood circulation sustains the retina through the artery and the central vein vacates the eye through the focal point of OD, this vascular system can be observed directly through an ophthalmoscope.

1.2 Degeneration of the Human Retina

Human retina undergoes several retinal deformities due to unhealthy living style, which influences the overall health of the eye. Abnormalities such as hemorrhages, macular degeneration, glaucoma and diabetic retinopathy are a few of the many degenerative conditions of the retina (Sowmya, 2016).A wide range of characteristic changes occur in the retina due to different diseases, a few of the symptoms are hemorrhages, cotton wool spots, exudates and micro aneurisms of the retina. Evaluation of the retinal health can also predict events leading to visual impairment. Early and proper diagnosis of the diseases under various symptoms is crucial to predictive and preventive medical treatment of the patient.

1.3 Diagnosis of Retinal Diseases

A number of automated techniques are present in the latest research literature for identification of features of the human eye and their underlying abnormalities(Gharaibeh, 2016). Vision loss can be cured if it is diagnosed earlier, but for the treatment of advanced stages and intense conditions more complex procedures needs to be followed. As such the digital image processing techniques have seen remarkable development in the recent years,

however for the detection, analysis and treatment of the diseases there is need to further improve the existing methods and develop new ones (Kumar & P, 2017). The solution is presented by modern advancements in the digital imaging. The diagnosis and treatment related conclusions are easily and accurately drawn by the physician using automated medical image analysis. Essentially, these procedures reduce the repetitive and error prone part of physician during analysis by automating the recursive tasks. For the detection of retinal diseases in patients in Fundus photographs Kale et. al. have discussed a number of techniques in their study(Kale & Janwe, 2017). A number of automated techniques utilize machine learning and reinforcement learning as a primary solution for the decision and classification problems. Algorithms such as Support Vector Machines (SVM), Artificial Neural Networks, K-Nearest Neighbor and Naïve Bayes Classification are a few of the most popular techniques of diseases identification and classification.

1.4 STARE Database

The faculty of University of California, San Diego published the project "Structured Analysis of the Retina" under the acronym STARE for the detailed study of the human retina in 1975. Funded by various U.S. National Health Institutes it has contributed large databases for the research and development of diagnostic techniques of the diseases pertaining to the human eye ("The STARE Project," n.d.).The publically available data in the STARE Database was employed in this research to perform the automated classification of retinal diseases using support vector machine (SVM) classifier.

1.5 Support Vector Machines (SVM)

In numerically extensive automation applications such as data analysis, pattern recognition, classification and regression a very popular technique called the Support vector machine (SVM) is used to perform supervised learning with an associated learning algorithm (Shveta

&Kaur, 2015). This is a classification technique which was introduced by Boser et. al. in 1992(Boser, Guyon, & Vapnik, 1992). SVMs are widely used in bioinformatics and many other fields of engineering and machine learning because of its higher accuracies and tunable parameters and kernels. It can be readily employed to data with high-dimensionality such as that of the symptoms of retinal or any biological conditions. A few of the marked benefits of SVMs are their enhanced flexibility of design, conciseness of solution in large data sets, the ability to manage large feature spaces and reasonable control for the over-fitting problem (Karthikram, Kavya, Keerthika, & Veenmathi, 2016).

Chapter 2: LITRETURE REVIEW

Humans typically employ sight the most, among their five main senses, to perceive the incoming information from the world around them. A considerably large part of the human brain is dedicated exclusively to process visual information. The human eye can be considered a camera to the effect that it converts light into information that can be understood by the brain. Similar to a camera, the eye contains a lens that can concentrate and change the point of focus in the outside world. The aperture of the human eye is controlled by the iris which opens or closes to control the incoming luminance. A camera generates a picture by projecting the light from the lens onto a thin film whereas inside the human eye the image from the lens is focused onto the retina, a specialized layer of cells which convey image information to the brain.

As far as the operation is considered, this is as far as the similarity between the human eye and a camera goes, beyond this the human eye is far more superior to a camera. The human eye can resolve image orders of magnitude precisely in comparison to a camera. Also the rapid luminance control, brightness control and real time stereo focus adjust acts instantaneously.

The light approaching the eye initially passes through the cornea that focuses the image by passing through it first. To keep the front of eye firm and slightly covered with a layer of fluid, the inner part of the eye is filled with a fluid known as aqueous humor. For focusing the light on retina, the lens squeezes and stretches according to the image. The human brain actively controls all the muscles inside the eye and generates a seamless image, needless to say this is a sort of neural training every human being goes through in the early months after birth.The interior surface opposite to the lens is called the fundus which includes the multi-layered sensory tissue, retina (Cassin & Rubin, n.d.).The light rays falling on the retina are

converted into electrical impulses by photoreceptors present in retina. When these electrical impulses travel to brain through optic nerve the visual cortex generates appropriate visual perception inside the brain. Retina has two kinds of photoreceptors which, as their names suggest, have ‘rods’ and ‘cones’ shapes. Rod cells have high sensitivity and at low lights they change their contrast and recognize movements, but they are not-precise and sensitive to color. Generally, the rods are located at the retinal margin and they are utilized for night time vision. Contrary to this, cones are the highly precise cells that have the ability to recognize colors. The area of the retina responsible for day time vision is macula and it comprise mostly of cones. Fovea is the center of the macula, which helps human eyes to differentiate fine details in the image.

The loss of peripheral vision can go unnoticed for years, even decades, in conditions such as Glaucoma, whereas the loss of central vision is evident immediately and can occur when the macula is damaged (Wyszecki & Stiles, 2000). Approximately 1.2 million nerve fiber endings connects the photoreceptors to the brain (Jonas, Schneider, & Naumann, 1992). These nerve fibers exit the eye in a highly unusual array in the optic nerve. Punctum caecum is the opening in the retina from where the optic nerve originates and leaves the eye. The sustaining nutrients, salts and oxygen are provided to the internal and external layers of retina by many retinal blood vessels. The inner layers contain 35% of the blood vessels and become noticeable through transparent humor in normal images of fundus. Other 65% blood vessels are for the external layers of retina and infrequently manifest in images of fundus camera because they are located in choroid(Abràmoff, Garvin, & Sonka, 2010).

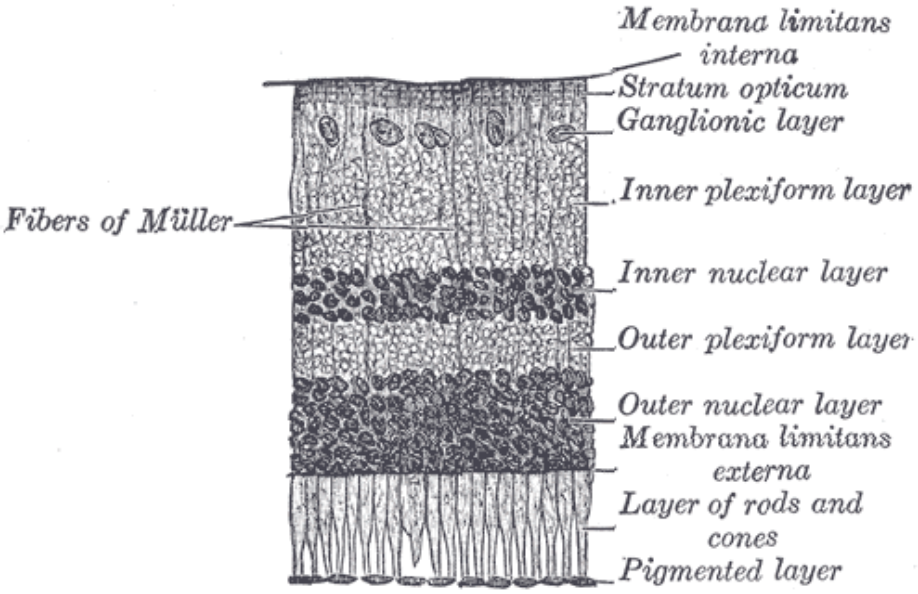
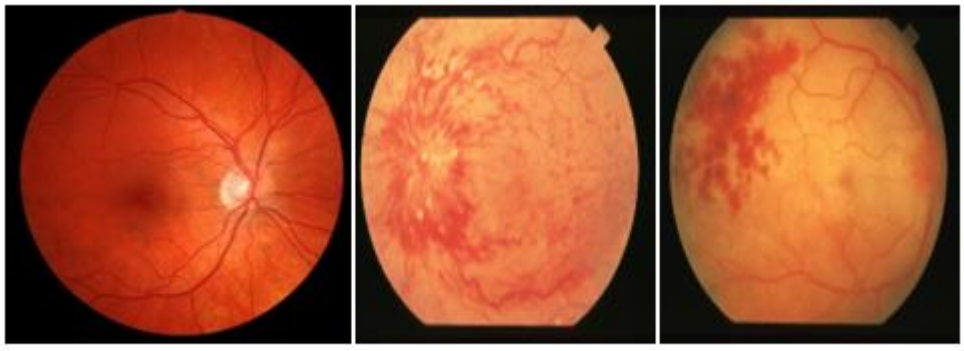


Figure 2.1 Layers of Retina (Williams et al., Gray's Anatomy, 1989)

With the help of bipolar and amacrine cells, the type of neuron cells named ganglion cells receives signals from photoreceptors. To combine the signals received from photoreceptors, interconnecting neurons known as horizontal cells are responsible (Masland, 2001).

2.1 Techniques of Retinal Imaging

With the use of funduscopic techniques, the eye is illustrated by optometrists. The ophthalmoscope is used to determine the health of the retina. Earlier, an instrument with several lenses having the ability to amplify up to 15 times, was used to provide the wide perspective view of the eye fundus. Its size was around that of a small flashlight, then later it was modified to the size of headband and a wearable alternate was developed to facilitate the physician (Chernecky & Berger, 2004). Nowadays a camera is available which directly captures the fundus images (Hutchinson et al., 2000; Lin, Blumenkranz, Brothers, & Grosvenor, 2002). Internal exposure of eye including posterior pole, retina, macula and optic disc are viewed upright and magnified by the modern fundus camera.



(a) (b) (c)

Figure 2.2 (a) Healthy retinal image; (b) retinal image having micro aneurism's signs; (c) retinal image with hemorrhage's signs (STARE Database)

A typical camera with an amplification of 2.5 views around 30-50 degrees area of the retina but with slight adjustments through additional lenses about 5 times the amplification can be achieved for a viewing angle of around 15 degrees (Saine & Tyler, 2002). Stereo fundus photography, hyper spectral imaging, Fluorescein Angiography (FA), scanning laser ophthalmoscope (SLO) are also uses for image acquisition instead of color digital fundus cameras. Optical coherence tomography (OCT) is another important imaging technique. It is a non-invasive imaging technique utilizing an interferometer to quantify the light's traveling time that is backscattered by the retina. In vivo illustrations of the retinal anatomical layers are obtained through this technique. Thus it can be utilized to analyze diseases, for example, AMD, Glaucoma and DME, with a considerably higher accuracy than with a basic image of fundus(Walsh, Wildey, Lara, Ouyang, & Sadda, 2010).

2.2 Retinal Abnormalities

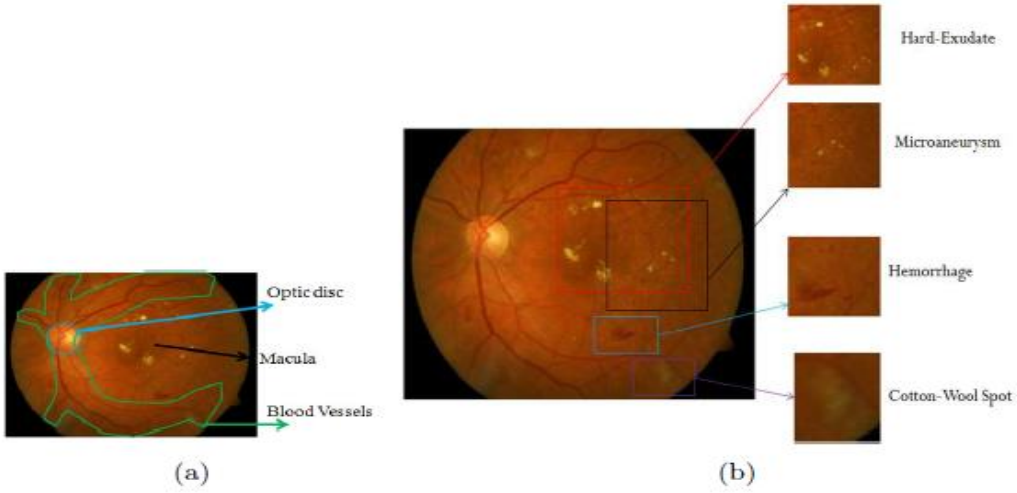


Figure 2.3 Fundus image; (a) fundus image's anatomy (b) abnormalities in fundus images (Sohini Roy Chowdhury, 2014)

2.2.1 Micro Aneurism

Micro aneurisms (MA's) are responsible lesions that may leak blood and fluid into retina, resulting in macular edema, hemorrhages and exudates whish are dangerous for vision. For the cure of macular edema and diabetic retinopathy, by laser, these MA lesions are the basic focus.

2.2.2 Hemorrhage

The loss of blood from the vasculature is called Hemorrhage. They come into view in the fundus as a red structure of various shapes and sizes. In retina their shape may be deeply correlated to the symptoms of the disease causing it. A flame shaped hemorrhages near to ganglion layer have a tendency to vanish inside a brief timeframe. Round shaped hemorrhages located in inner nuclear of retina are referred to as blot and dot hemorrhages(Niki, Muraoka, & Shimizu, 1984).

2.2.3 Exudate

Yellowish fluid accumulation composed of proteins and fats inside the sensory retina are called exudates. MAs due to increased pressure, having thinner walls than normal capillaries are typically responsible for the leakage of lipid and proteins resulting in exudates.

2.2.4 Cotton Wool Spots

A reduced blood supply to the nerve fiber layers causes cotton wool spots, which are micro obstructions which appear as yellowish or white composition or smudges with blurred edges. They are exceedingly correlated with conditions that influence the retinal flow, for example, diabetic retinopathy but they do not cause visual difficulties themselves.

2.2.5 Drusen

Yellow accumulations, located outside the cells, beneath the retina in the Bruch's layer are called drusen. These are the indications of macular degeneration and sometimes, they are not clearly visible in a fundus image because they are beneath the retina.

2.2.6 Diabetic Retinopathy

If Diabetic retinopathy is not treated at earlier stage it can damage the retina which leads to permanent loss of vision. As compared to non-diabetic subjects, the diabetic patients are 25 times more likely to have blindness. It is a vascular difficulty of diabetes mellitus(Baker, Hand, Wang, & Wong, 2008).

2.2.7 Diabetic Macular Edema

The most typical reason of sight loss is diabetic macular edema (DME) that known as a drawback or DR (Singer, Nathan, Fogel, & Schachat, 1992). This results in retinal swelling in patients of diabetes by fluid leakage from micro aneurisms, inside the center of macula, which is a result of incurable damage caused by high blood sugar levels.

2.2.8 Age-Related Macular Degeneration

Age-Related Macular Degeneration (AMD) has a similar impact as diabetic retinopathy, which influences the macula of retina. Main attribute of initial stage of AMD is existence of soft drusen. It not always the case that macular degeneration is present in a person that has drusen, however both conditions have strong correlation to build up macular degeneration. Dry or wet AMD are the two forms of late AMD. Leakage of blood/serum is not caused by dry AMD but vision loss can happen. These patients may have good focal vision yet numerous functional constraints that include: reading hindrance due to the central vision limitation, fluctuating vision and night time restricted vision. Wet AMD causes the growth of abnormal blood vessels beneath the retina and macula, this directly causes the loss of patient's central vision(Gottlieb, 2002).

2.2.9 Glaucoma

Glaucoma damages the ganglion cells and their axons that eventually affect the retina. Due to increasing pressure on optic nerves it causes peripheral visual field loss initially and progresses gradually inward. It usually goes undetected for many years and it has turned into the second most prominent reason for visual impairment in whole world (Resnikoff et al., 2004).

2.3 Publically Available Databases

To design, implement and test algorithms that are capable of analyzing retinal morphology or diagnosis in the field of medical imaging, there are some annotated and publicly available collections, STARE, DRIVE, DIARETDB, MESSIDOR, HEI-MED, etc have retinal images for with conditions and characteristics. Their purpose comprises DR diagnosis, vessel segmentation and localization of micro aneurisms. However, the photographs separately are not sufficient to make dataset valuable for algorithm expansion. The main feature is the ground truth information that gives the golden level that can be achieved by training algorithms and testing them. The pros and cons of proposed approaches were computable and comparable, whenever a single common data set (with GT) was used by different research groups(Niemeijer et al., 2010).

2.4 Prior Work

Keith P. Thompson (1992) promoted the rapid development of laser applications in ophthalmology, including access to the human eye, its transparency and the properties of its internal tissue absorption. In the eye, with photo disruptive, photo-thermal and photochemical processes, laser achieves its effect.

Spencer et al. (1996) median filteredthe image on green channel to evaluate the retinal background and then from that real image they subtract the estimated background.

Lee and Wang (1999) approached the problem of automatically detecting the quality of the fundus image. Their approach starts from the perspective of addressing a pure signal by applying the global image intensity chart roughly by the Gaussian distribution.

Lee and Wang's approach was extended by Lalonde et al. (2001) using two distinct sets of attributes: edge distribution in the image and pixel intensity’s local distribution. Instead of

accurately associating image quality with noise, its quality concept depends upon capability of experts for making the right diagnosis.

Jim Beach (2002) provides spectral reflection curves obtained using a spectral camera from structures that produce hemoglobin signatures. These include retinal artery, macular area, veins and optic disk. For the diagnosis of oxygen-dependent changes in hemoglobin signature surrounding vessels and tissues were used.

In 2003 Ushar et al. applied a vessel fragmentation algorithm that evaluates image deformation. Through solid threshold, monitored vessel area is directly related to image quality.

Li and Chutatape (2004)used edge detection to reduce the region size for computational issues. Then to identify all structures with expected shapes, morphological operators were used in grayscale (e.gvesicles).

In 2005 Foracchia et al. specified pixels by assessing the standard deviations and mean on the local window, these pixels belonged to the background and a normalized image was obtained through the background pixels.Feng et al. (2007) adopted a contour-based approach that was a multi-scale image evaluation process depending upon Wavelets theory.

Kevin Noronha (2006) elaborated the ways to reveal key attributes of the fundus images like the optic disc, blood vessels, fovea and exudates. Afterwards they applied the Hough transform to detect the brightest part of the fundus.

Sanjayol Lee (2007) suggested a variety of ways to record retinal images, but evaluating such methods objectively was difficult because there was no benchmark for real alignment of the individual images that make up the montage. The author also provides a validation tool for

any method of recording an image of retina by tracking the distortion path and recording miss-alignment from the reference benchmark.

E. Ricci and R. Perfetti (2007) used the line operators and the SVM with a set of 3 features per pixel. This method was very sensitive to training data and due to SVM classifier this was computationally intensive.

S. Sekhar (2008) used a retinal image to diagnose and treat several diseases of eye like glaucoma and diabetic retinopathy. In initial stage, a round area of significance was found by disconnecting the brightest locale of the picture by morphological procedures. In the next stage, to detect a circular feature the Hough transform was used.

Riries Rulaningtyas and Khusnul Ain (2009) used edge detection for glaucoma detection. In this work, glaucoma is diagnosed by doctors. Doctors typically tend to disagree on the detection of Glaucoma thus in order to facilitate the physicians in the diagnosis, a software research was developed with edge detection methods which could give the glaucoma and retinal edge pattern itself. The accurate detection of glaucoma in this study is the first step in the study of glaucoma classification. This research found the best way to detect the glaucoma between Sobel, Prewitt and Robert edge filters. Out of those three techniques, the Sobel edge filter was appropriate for glaucoma detection. The Sobel filter was one standard deviation value smaller than the other edge detection methods.

Ehab F. Badran, Esraa Galal Mahmoud and Nader Hamdy (2010) elaborate the new algorithm for detecting brain infection. They introduced a computer-based method to identify the retinal infection area using magnetic resonance imaging (MRI) images. The retina was first classified as either healthy or infectious and further classified into a benign or malignant infection.

N. Nandha Gopal, Dr. M. Karnan (2010) presented the discovery of glaucoma using the C-Mean clustering algorithm. In this research, an intelligent system for the diagnosis of glaucoma was designed through magnetic resonance imaging by utilizing image processing aggregation algorithms, which was Fuzzy C Means accompanied by intelligent optimization tools.

Yanqing Xue, Shuicai Wu and Hongjian Gao (2011) created a three-dimensional image of the retina. In their work, the support vector machine (SVM) was used for automated detection of glaucoma and the classic mobile cube algorithm was used to obtain 3D visualization model of glaucoma. In addition, three-dimensional model coefficients based on morphological analysis were extracted, providing reliable information on the quantization of radio-glaucoma eradication.

Zafer Yavuz (2011) suggested a method for automatically categorizing the blood vessels of the retina. In this study, the author applied a top-hat transform after the Gabor filter to improve blood vessels. Later, the output of the transform was converted to a binary image with a p-tile threshold.

K.Sangeetha (2012) revealed irregularities in diabetic patient's eye for early identification of DR. The system was introduced, utilizing digital image processing (DIP), for the programmed identification of veins and irregularities in the eye of diabetic patients.

Geeta Ramani (2012) introduced a new approach to automate disease detection, which was proposed to be used to analyze retinal images along with data extraction techniques to accurately classify diseases in retinal images such as diabetic retinopathy.

Kumar Parasuraman (2012) has proposed another procedure that gathers data about all veins in the retina and recognizes the genuine vessel in image of retina. In the proposed technique, the input image was chosen first then the blood vessels go under fragmentation. From this,

the intersection point identification was used to distinguish the vessels that cross each other utilizing the window with adjoining pixels.

G.T. Pavai and S.T. Selvi (2013) have fragmented the optic disc using intrusive texture descriptors, inertia, entropy and energy.

R.A. Welikala, et al. (2014) utilized an adjusted line administrator and double categorization in retinal image investigation to distinguish new vessels while lessening false reactions caused by other basic retinal attributes.

Nasr Al-Gharaibeh, Al-Huson (2015) suggested an approach that consists of pretreatment, vascular segmentation (FPCM), localization of necrosis, fovea removal, extraction and classification of malignant neural features. Neuro-Fuzzy states as a hybrid of fuzzy logical and neural networks were employed in the design. Evaluations were performed utilizing MATLAB and the MESSIDOR database for their analysis that provides productive conclusions in vulnerability, quality, categorization, precision and reliability.

Jing Wu et al. (2016) introduced a fully documented approach to detect fovea SD-OCT scans in a healthy and diseased macula. This allows the use of the fovea as a key parameter in building a population reference frame to identify and exact key spatial temporal features from a large group of patients consisting of different time points, devices and imaging modalities.

Priyanka B. Kale and Nitin Janwe (2017) proposed using an automated system to identify patients with diabetic retinopathy using fundus images. To enhance intensity of images and noise removal images were pre-processed. After blood vessels detection they delete them and focused on hemorrhages. Finally, the textured area was used to classify images into severe, moderate and normal DR. STARE database was used to test this approach.

2.5 Machine Learning

2.5.1 Naive Bayes

In the medical data the Naïve-Bayes NB classification has been widely applied(Arora & Sharma, 2015). The NB classifier is one of the most efficient classification algorithms, in comparison to other techniques such as nearest neighbor, logistic regression, decision tree, neural network and rule based on medical datasets. Classifiers are compared for the region under the ROC curve(Mangai, Nayak, & Kumar, n.d.). Kononenko (2001) considered NB as the reference algorithm in any medical field to be tested before any other advanced method. Naive Bayes are simple, computationally efficient and require relatively little training data and require many parameters and are naturally robust to non-available data and noise in contrast to other classifiers.

2.5.2 KNN

The K Nearest Neighbor is a type of instance-based learning, where the function is rounded locally only and each computation is converted up to the classification. Because it does not require any extensive training or simple training phase this technique is called lazy learning. All training data is required only during the testing phase so that if we have a large set of data, we need a special way to work on a piece of data that is known as the algorithmic approach. This can also be used to estimate the density inside a population. From all automated learning algorithms the KNN algorithm is one of the simplest algorithms. The KNN classification has been formulated from the requirement for multiple analyses.

2.5.3 Support Vector Machine (SVM)

SVM is a supervised learning framework with a linked learning algorithm that can analyze data and identify patterns that are then used for regression analysis and classification(Shveta

& Kaur, 2015). Given a set of training examples, each of which has been set as a reference for one category in one of two categories, the SVM training algorithm creates a model that divides new examples into one class or the other and distributes them in a non-probability linear binary classifier. The SVM model is representative of the example where the points are divided into the space assigned to those examples from different categories(Arora & Sharma, 2015). In addition to linear classification performance, SVM can quickly execute a non-linear classification using a trick called the kernel trick, meaning implicitly assigning it in distinct spaces of high dimensions.

2.5.4 Hidden Markov Model (HMM)

A HMM-based approach for recognition and detection uses an effective set of observational vectors acquired from binary-DCT coefficients. HMM embedded dimension sculpture can have two dimensional data of one-dimensional HMM which is less difficult than two-dimensional HMM. This model is quite suitable for facial images because it exploits the important facial properties and structure of the "cases" within each of these super states.

2.5.5 DCT

In image processing and recognition, DCT and linear discrimination are widely used technologies. It can dramatically improve face recognition and palm pattern recognition rates and effectively reduce the dimensions of feature space (Arora & Sharma, 2015).

2.5.6 PCA

A new technique is formulated in which a two-dimensional component analysis is performed to represent images. Instead, image contrast measurements are created directly using the original image matrices and their self-extracting content is derived to extract the image

features. The experimental result shows that drawing image features is a very effective way of using 2DPCA instead of simple PCA.

2.5.7 AUC

The AUC is part of the regression matrix. It is an evaluation matrix widely used for binary classification problems, such as predicting an existing disease.

Chapter 3: METHODOLOGY

Publically accessible database of retinal image at Structured Analysis of the Retina (STARE) project was used to evaluate automated classification of retinal diseases. The initial idea was an image interpretation approach that differentiates the diseases of retina using fundus images. This contain colored images of retina obtained using 50 fundus camera that is of TRV (Topcon Corp., Tokyo, Japan) with 605x700 pixels resolution at 35 degree field. 397 images in 14 disease categories are present in database including cilio retinal artery occlusion (CRAO), emboli, branch retinal artery occlusion (BRAO), hemi-CRVO, proliferative diabetic retinopathy (PDR), arteriosclerotic retinopathy, central background diabetic retinopathy (BDR), branch retinal vein occlusion (BRVO), choroidal neo-vascularization (CNV), hypertensive retinopathy, Coat's disease, retinal vein occlusion (CRVO), macroaneurism, and other retinal status. These other retinal statuses were further separately classified with new

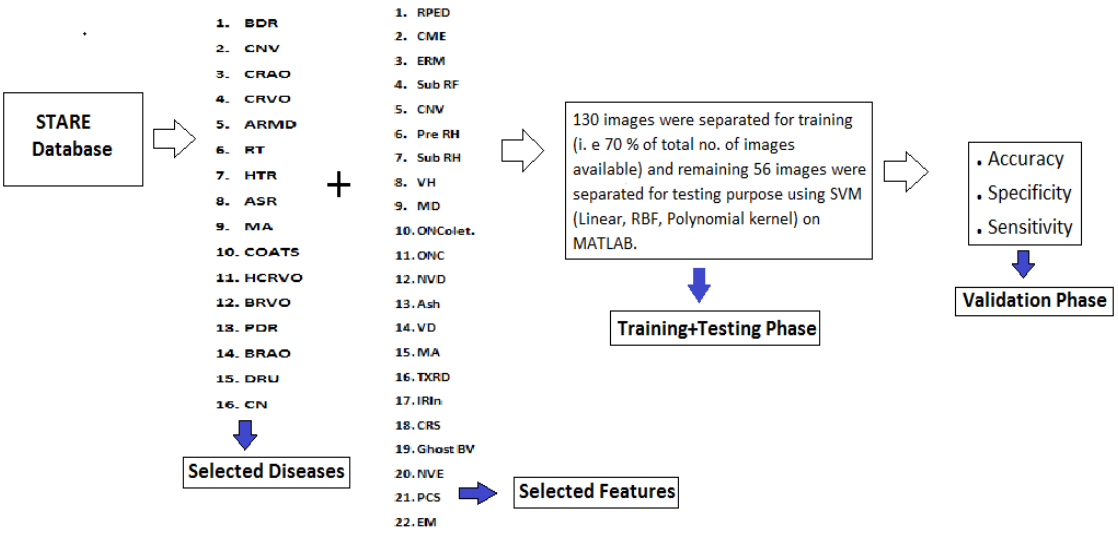


Figure 3.1 Proposed Procedure

codes resulting in 41 diseases.

Images were tallied for every disease in MATLAB and a reduction criteria was employed for

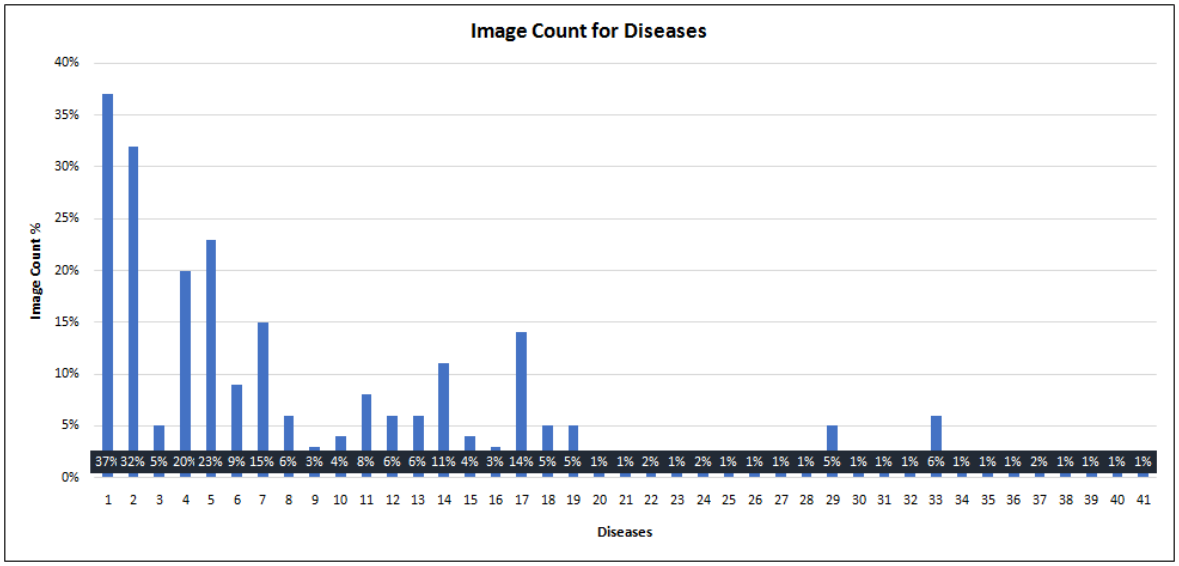


Figure 3.2 Total image count (%) for each disease

diseases, features and images. Database contains 402 images, 41 diseases and 37 features.

Among 402 images only 311 images have valid information about features and diseases. First criteria was the removal of diseases with less than 5 images. 91 images were removed due to improper diagnosis (according to the database), thus resulting in 21 diseases afirst reduction, 2 categories (unknown diagnosis and choroidal haemangioma) were not used as r recommended by the STARE project, leaving only 19 diseases, 37 features and 301 images in second reduction case.

The following listings show the MATLAB code for the reduction of the images in the database to satisfy the selection criterion.

```
% IMAGE VS FEATURE ANALYSIS FOR EACH DISEASE

finalFea = xlsread('finalFea.xls');
finalDis = xlsread('finalDis.xls');

% Diseases are remapped as following
% Old Disease Number        -       New Disease Number
%   1-17                    -           1-17
%   18                      -       Insufficient Cases
%   19,20                   -           18,19
%   21-clc
%   23                      -       Insufficient Cases
% Features are mapped as following
% Old Feature Number        -       New Feature Number
%   1-37                    -           1-37
%   38-44   `               -           Not Used

redDis = zeros([402 19], 'logical');
redDis(:,1:17) = finalDis(:,1:17);
redDis(:,18:19) = finalDis(:,19:20);
redFea = zeros([402 37], 'logical');
redFea = finalFea(:,1:37);
[nImg, nFeat] = size(finalFea);
[nImg, nDis] = size(finalDis);
nIperD = sum(finalDis);
nIperF = sum(finalFea);
DFchart = zeros([length(finalDis) nFeat nDis], 'logical');

for k = 1:1:nDis
    for m = 1:1:length(finalDis)
        if(finalDis(m,k))
            DFchart(m,:,k) = finalFea(m,:);
        end
    end
end

scd = sum(sum(DFchart(:,:,testDis)));
dScat = zeros([2 scd]);
[rd, cd] = size(DFchart(:,:,testDis));
counter = 1;

for m = 1:1:rd
    for n = 1:1:cd
        if(DFchart(m,n,testDis))
            dScat(:,counter) = [m n]';
            counter = counter + 1;
        end
    end
end
```

```
% IMAGE VS DISEASE ANALYSIS FOR EACH FEATURE

nIperF = sum(finalFea);

FDchart = zeros([length(finalFea) nDis nFeat], 'logical');


for k = 1:1:nFeat
    for m = 1:1:length(finalFea)
        if(finalFea(m,k))
            FDchart(m,:,k) = finalDis(m,:);
        end
    end

end


scf = sum(sum(FDchart(:,:,testFeat)));

fScat = zeros([2 scf]);

[rf, cf] = size(FDchart(:,:,testFeat));

counter = 1;
for m = 1:1:rf
    for n = 1:1:cf
        if(FDchart(m,n,testFeat))
            fScat(:,counter) = [m n]';
            counter = counter + 1;
        end
    end
end
```

Listing 3.1 Matlab® Code for First Reduction less than 5 images

```
% DISEASE VERSUS FEATURE TABLE FOR REDUCED COUNTS
% FvD - Features / Diseases Table

FvD = zeros([37 19]);

for i = 1:1:402
    for j = 1:1:37
        if(redFea(i,j))
            FvD(j,:) = FvD(j,:) + redDis(i,:);
        end
    end
end

FvDcheck = zeros([37 19], 'logical');

for i = 1:1:37
    for j = 1:1:19
        if(FvD(i,j) > 0)
            FvDcheck(i,j) = 1;
        end
    end
end

% FvDcheck shows if a feature and disease are present together
% Less than 50% check for reduction of features

redFea2 = zeros([402 22], 'logical');

% Features Mapping Sequence
% Old Features      -       New Features
% 1-5               -       1-5
% 7,8               -       6,7
% 10                -       8
% 15,16             -       9,10
% 19,20             -       11,12
% 22                -       13
% 25                -       14
% 28                -       15
% 30                -       16
% 32-37             -       17-22

% Features removed by 50% disease occurance check are;
% 6, 9, 11, 12, 13, 14, 17, 18, 21, 23, 24, 26, 27, 29, 31
```

```
redFea2(:,1:5)      = redFea(:,1:5);
redFea2(:,6:7)      = redFea(:,7:8);
redFea2(:,8)        = redFea(:,10);
redFea2(:,9:10)     = redFea(:,15:16);
redFea2(:,11:12)    = redFea(:,19:20);
redFea2(:,13)       = redFea(:,22);
redFea2(:,14)       = redFea(:,25);
redFea2(:,15)       = redFea(:,28);
redFea2(:,16)       = redFea(:,30);
redFea2(:,17:22)    = redFea(:,32:37);

% Feature Suppression mask
fsm = ones([1 37] , 'logical');
fsm(6) = 0;
fsm(9) = 0;
fsm(11:14) = 0;
fsm(17:18) = 0;
fsm(21) = 0;
fsm(23:24) = 0;
fsm(26:27) = 0;
fsm(29) = 0;

fsm(31) = 0;

frm = ~fsm;

for i = 1:1:402
    maskFea(i,:) = redFea(i,:).*fsm;
end

F = (sum(redFea'))';
M = (sum(maskFea'))';

redDis2 = redDis;

for i = 1:1:402
    if(M(i) == 0)
        redFea2(i,:) = zeros([1 22], 'logical');
        redDis2(i,:) = zeros([1 19], 'logical');
    end
end
```

Listing 3.2 Matlab Code for the second reduction for 50% features

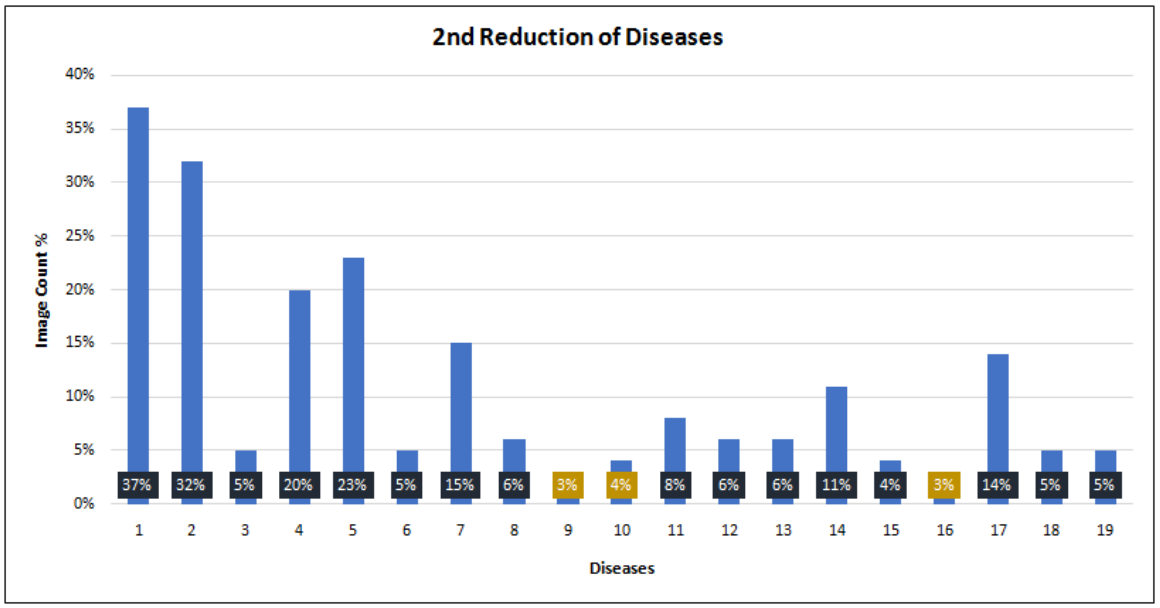


Figure 3.3 Reduced image count for each disease

Afterwards the dataset was further reduced by applying a 50% criteria to 19 diseases eliminating 15 features and 112 images as a result we get 189 valid images for 22 features with 19 diseases.

The images which were removed in the above step also had some images of valid 19 diseases that causes skewed results for instance none of the 22 selected features were present however the disease was present in the image diagnosis. So a final reduction to fix the diseases was performed and images of 3 diseases were removed leaving valid 186 images and 16 diseases.

```
% Removing diseases with less than 5 valid images
% These are Diseases 8, 16 and 19

redDis3 = zeros([402 16], 'logical');
redDis3(:,1:7)       = redDis2(:,1:7);
redDis3(:,8:14)      = redDis2(:,9:15);
redDis3(:,15:16)     = redDis2(:,17:18);

% Final Datasets

validImg = zeros([402 1],'logical');

temp = sum(redDis3');

for i=1:1:402
    if(temp(i) > 0)
        validImg(i) = true;
    end
end
```

Listing 3.3MATLAB Code for the third reduction (16 Diseases with 22 Features in 186 images)

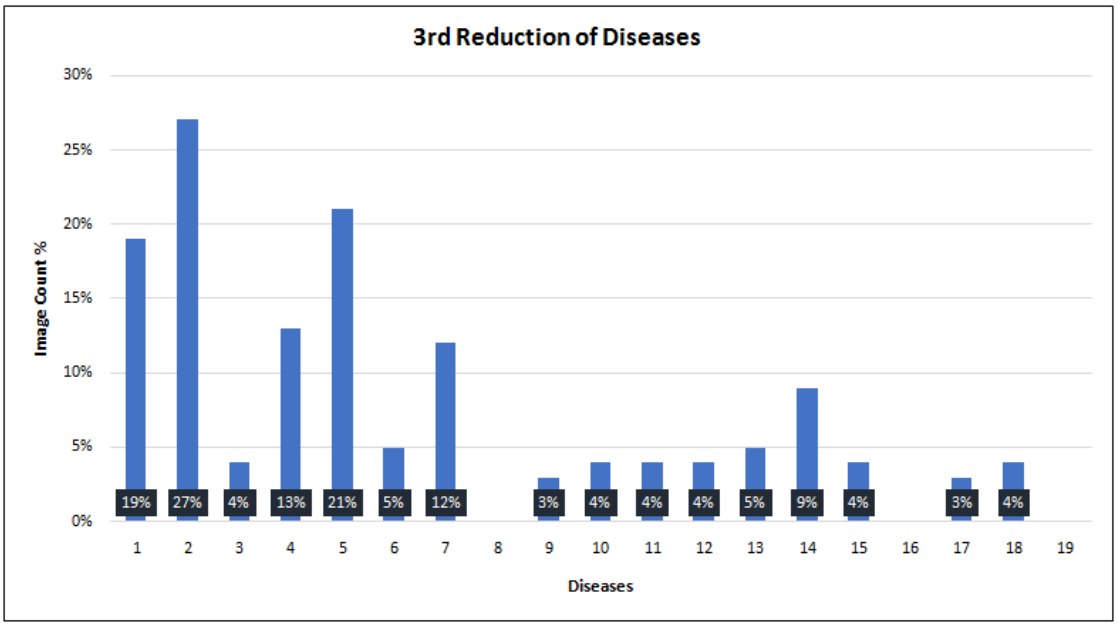


Figure 3.4 Disease reduction containing 5 images

In order to utilize a Support Vector Machine the Data is separated into training and testing samples. Biasedrandom sampling was performed on the reduced data of 186 images to split into 70% training and 30% testing data. The listing 3.4 shows the MATLAB Code for the biased random sampling of the data.

```
nImg = sum(validImg);
IvF = zeros([nImg 22], 'logical');
IvD = zeros([nImg 16], 'logical');

sweep = 1;
for i = 1:1:402
    if(validImg(i))
        IvF(sweep,:) = redFea2(i,:);
        IvD(sweep,:) = redDis3(i,:);
        sweep = sweep + 1;
    end
end

% The Matrices IvF and IvD are the training / testing Matrices
% IvF -> Input Matrix for training / testing Features for Valid
Images
% IvD -> Output Matrix for training  / testing Diseases for
Valid Images

DiseaseCount = 16;
totalSamplesPerDisease = 186;
positiveSamplesPerDisease = sum(IvD);
pIndexPerDisease = zeros([max(positiveSamplesPerDisease)
DiseaseCount]);
% A value of 0 in the pIndexPerDisease corresponds to no image

sweepCounter = ones([1 16]);
for i = 1:1:totalSamplesPerDisease
    for j = 1:1:DiseaseCount
        if(IvD(i,j))
            pIndexPerDisease(sweepCounter(j),j) = i;
            sweepCounter(j) = sweepCounter(j) + 1;
        end
    end
end
```

```
mxLine = positiveSamplesPerDisease;
mxRepVal = max(mxLine) + 1;
minIndex = 0;
minVal = 0;

totalTrainingSamples = floor(0.7*totalSamplesPerDisease);
totalTestingSamples = totalSamplesPerDisease -
totalTrainingSamples;

trainingSamples = zeros([totalTrainingSamples 1]);
trainingIndex = 1;

testingSamples = zeros([totalTestingSamples 1]);
testingIndex = 1;

trainingPositiveSamplesCountPerDisease =
floor(0.7*positiveSamplesPerDisease);
testingPositiveSamplesCountPerDisease =
positiveSamplesPerDisease -
trainingPositiveSamplesCountPerDisease;

for i = 1:1:DiseaseCount
    if(testingIndex > totalTestingSamples)
        break;
    end
    [minVal, minIndex] = min(mxLine);
    rTrainSample =
datasample(pIndexPerDisease(1:minVal,minIndex),trainingPositive
SamplesCountPerDisease(minIndex),'Replace',false);
    rTestSample =
setdiff(pIndexPerDisease(1:minVal,minIndex),rTrainSample);

    crossCheckTrain = ismember(rTrainSample, trainingSamples);
    crossCheckTest = ismember(rTrainSample, testingSamples);

    for j = 1:1:length(crossCheckTrain)
        if(~(crossCheckTrain(j)||crossCheckTest(j)))
            trainingSamples(trainingIndex) = rTrainSample(j);
            trainingIndex = trainingIndex + 1;
        end
    end

    crossCheckTrain = ismember(rTestSample, trainingSamples);
    crossCheckTest = ismember(rTestSample, testingSamples);
```

```
    for k = 1:1:length(crossCheckTest)
        if(~(crossCheckTrain(k)||crossCheckTest(k)))
            testingSamples(testingIndex) = rTestSample(k);
            testingIndex = testingIndex + 1;
        end
        if(testingIndex > totalTestingSamples)
            break;
        end
    end
    mxLine(minIndex) = mxRepVal;
end

allSamples = 1:1:totalSamplesPerDisease;
remainingSamples = setdiff(allSamples,trainingSamples);
remainingSamples = setdiff(remainingSamples,testingSamples);

trainingSamples(trainingIndex:end) = remainingSamples;

testingSamples = sort(testingSamples);
trainingSamples = sort(trainingSamples);

%clearvars -except IvF IvD trainingSamples testingSamples

% Partitioning Training and Testing Data

TrainingIvF = zeros([0 0],'logical');
TestingIvF = zeros([0 0],'logical');

TrainingIvD = zeros([0 0],'logical');
TestingIvD = zeros([0 0],'logical');

for i=1:1:186
    if(ismember(i,trainingSamples))
        TrainingIvF(end+1,:) = IvF(i,:);
        TrainingIvD(end+1,:) = IvD(i,:);
    else
        TestingIvF(end+1,:) = IvF(i,:);
        TestingIvD(end+1,:) = IvD(i,:);
    end
end

clearvars -except TestingIvD TestingIvF TrainingIvD TrainingIvF
```

Listing 3.4 MATLAB Code for Biased Random Sampling

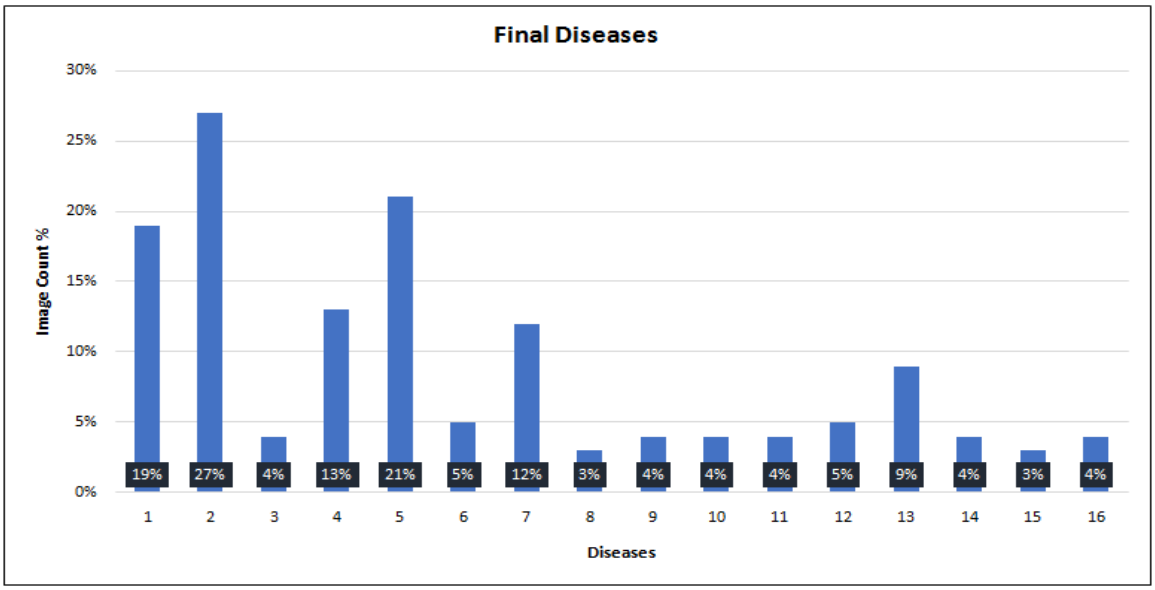


Figure 3.5 Final disease count (16 Diseases, 22 Features and 186 Images)

This data was split into training and testing samples through biased random sampling in the ascending order of positive samples of the diseases, min number of occurrences were for disease no. 15 (i.e. 5 positive samples). This ensured that the uneven distribution of the data was catered for in the sampling process. Around 70% of the samples were sampled through biased random sampling for training and the remaining 30% were separated for testing.

Total 130 images were separated for training (i. e 70 % of total no. of images available) and remaining 56 images were separated for testing purpose.

The Table 3.1 below shows the split positive samples of data for all diseases in the available reduced data. The diseases with very less positive samples were prone to over-fitting in the case of true random sampling, hence biased random sampling was performed to minimize over-fitting.

Disease	Total Positive Samples	Total Negative Samples	Training Positive Samples	Testing Positive Samples
1	35	145	25	10
2	51	129	41	10
3	7	173	4	3
4	25	155	16	9
5	39	141	28	11
6	10	170	7	3
7	22	158	14	8
8	6	174	4	2
9	7	173	4	3
10	8	172	5	3
11	8	172	5	3
12	9	171	6	3
13	17	164	10	7
14	7	173	5	2
15	5	175	3	2
16	8	172	6	2

Table 3.1 Count of total, training and testing positive samples for each disease

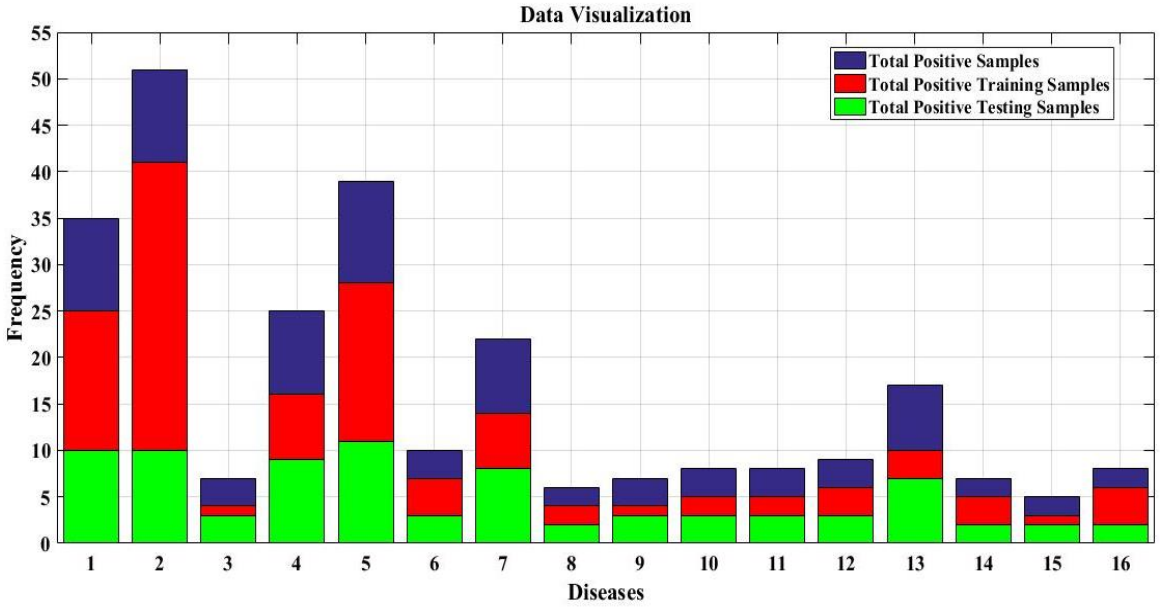


Figure 3.6 Representation of positive samples for training and testing (Blue: total positive samples. Red: total Positive training samples, Green: Total positive testing samples)

This biased random sampling was used because total number of positive samples per disease varies from 5 to 51 with normal random sampling all positive samples of some diseases were selected for training; while down to 50% of the positive samples of some diseases were selected for training. Biasing the random sampling ensures the diseases with least number of positive samples were treated first while ensuring random sampling for training. Figure 2.17 proves that biased random sampling (green line trend) is more closely related to 70% training set (blue reference line for 70%) as compared to normal random sampling (red line trend).

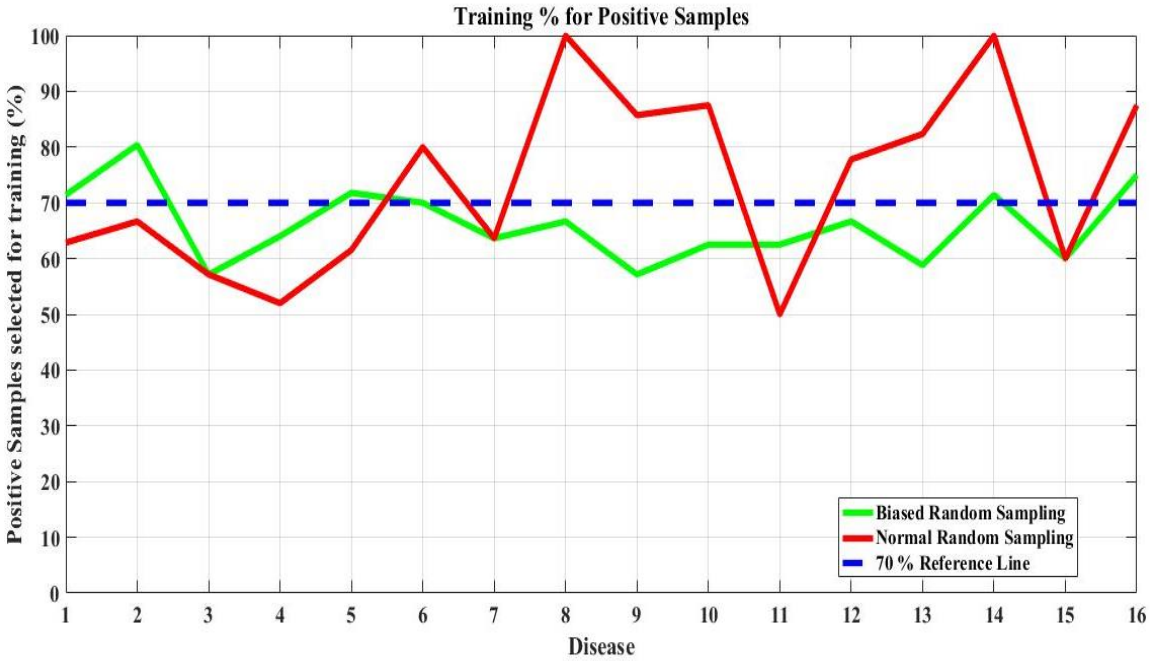


Figure 3.7 Graph showing trend of biased random sampling

Each sample point has 22 features and labels for presence (or absence) of 16 diseases. A total of 130 samples were used for training and 56 samples were used for testing the SVM algorithm.

```
% Support Vector Machine Training Algorithm

% Parameters:
% Matrix          Description          Size         Data Type
% TrainingIvF     Training Feature     130 x 22     binary (1-0)
% TrainingIvD     Training Labels      130 x 16     binary (1-0)
% TestingIvF      Testing Features     56 x 22      binary (1-0)
% TestingIvD      Testing Labels       56 x 16      binary (1-0)

% Since the SVM requires the data to be numerical instead of
logical
% (binary), it must first be converted into an equivalent
representation
```

Listing 3.5 The parameters of the training and testing data for SVM

A Support Vector Machine (SVM) performs classification by finding the hyper plane that maximizes the margin between the two classes. The vectors (cases) that define the hyper

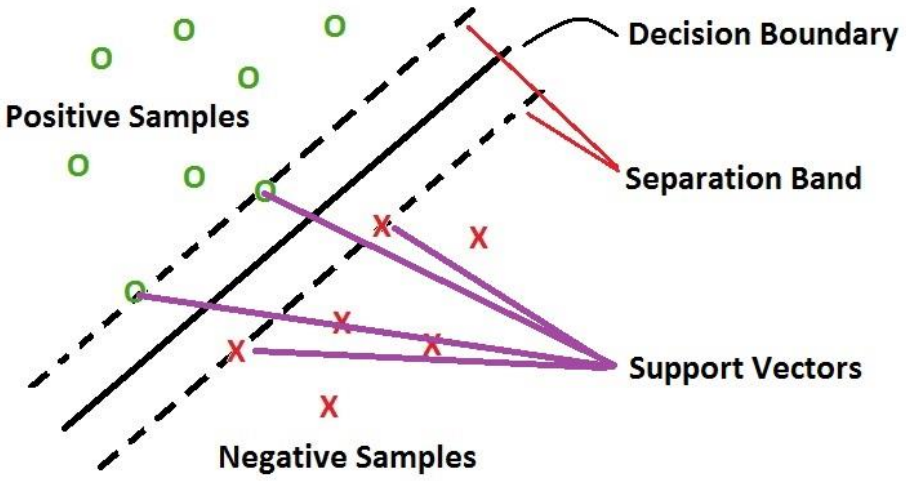


Figure 3.8 Basic Scheme of an SVM Classifier

plane are termed as the support vectors (Umesh, Mrunalini, & Shinde, 2016).

The kernel can be changed to incorporate data that is not linearly separable by a hyperplane boundary. In machine learning, kernel methods are a class of algorithms for pattern analysis, any linear model can be turned into a non-linear model by applying a non-linear kernel.

Three different kernel functions were tested for classification, Linear Kernel, Gaussian Kernel (radial basis function) and a Polynomial Kernel to analyze which performed better. Kernel is shape of separation boundary- linear/nonlinear boundary represents the similarity of vectors (training samples) in a feature space.

Four types of data sets were analyzed DDC (Direct Data Conversion) in which logical data (0 and 1) converted to numerical data (0 and 1), PCA Full (22 feature were reoriented into 22 Principal components), PCA10 / PCARED (22 feature were reoriented into 22 Principal components and 10 were used) and PCA05 (22 feature were reoriented into 22 Principal

```
% For this 3 types of data matrices are created;
%
% 1- Direct Data Conversion:
%       logical 1 is converted to numerical 1
%       logical 0 is converted to numerical 0
% 2- PCA Full:
%       All 22 features are reoriented into 22 Principle
Components
% 3- PCA Reduced:
%       All 22 features are reoriented and 10 Principle
Components are used

% *** DIRECT DATA CONVERSION ***
TrainingIvF_DDC = double(TrainingIvF);
TestingIvF_DDC = double(TestingIvF);
% *** PRINCIPLE COMPONENT ANALYSIS
LoadingsIvF = pca(TrainingIvF_DDC);
% *** PCA FULL ***
TrainingIvF_PCA_FULL = TrainingIvF_DDC * LoadingsIvF;
TestingIvF_PCA_FULL = TestingIvF_DDC * LoadingsIvF;
% *** PCA REDUCED ***
TrainingIvF_PCA_RED = TrainingIvF_PCA_FULL(:,1:10);
TestingIvF_PCA_RED = TestingIvF_PCA_FULL(:,1:10);
% *** PCA REDUCED 5 PC ***
TrainingIvF_PCA_RED_5 = TrainingIvF_PCA_FULL(:,1:5);
TestingIvF_PCA_RED_5 = TestingIvF_PCA_FULL(:,1:5);
```

Listing 3.6 MATLAB Code for generating different datasets for SVM

components and only 05 were used).

The different datasets were used to initially train and afterwards test the SVM algorithm for three different kernels, the following listing 3.7 represents the MATLAB code for training

```
% USING SVM ALGORITHM FOR SEPARATION
Kern = 'linear';
% For 16 Diseases and 56 Testing cases we have;
% 56 x 16 predicted scores
Dis_Pred_DDC = zeros([56 16]);
Dis_Scor_DDC = zeros([56 16 2]);
Dis_Pred_PCA_FULL = zeros([56 16]);
Dis_Scor_PCA_FULL = zeros([56 16 2]);
Dis_Pred_PCA_RED = zeros([56 16]);
Dis_Scor_PCA_RED = zeros([56 16 2]);
Dis_Pred_PCA_RED_5 = zeros([56 16]);
Dis_Scor_PCA_RED_5 = zeros([56 16 2]);
for i = 1 : 1 : 16
    SVMModel =
fitcsvm(TrainingIvF_DDC,TrainingIvD(:,i),'KernelFunction',Kern)
;
    [label score] = predict(SVMModel, TestingIvF_DDC);
    Dis_Pred_DDC(:,i) = label;
    Dis_Scor_DDC(:,i,:) = score;
end
for i = 1 : 1 : 16
    SVMModel =
fitcsvm(TrainingIvF_PCA_FULL,TrainingIvD(:,i),'KernelFunction',
Kern);
    [label score] = predict(SVMModel, TestingIvF_PCA_FULL);
    Dis_Pred_PCA_FULL(:,i) = label;
    Dis_Scor_PCA_FULL(:,i,:) = score;
end
for i = 1 : 1 : 16
    SVMModel =
fitcsvm(TrainingIvF_PCA_RED,TrainingIvD(:,i),'KernelFunction',K
ern);
    [label score] = predict(SVMModel, TestingIvF_PCA_RED);
    Dis_Pred_PCA_RED(:,i) = label;
    Dis_Scor_PCA_RED(:,i,:) = score;
end
for i = 1 : 1 : 16
    SVMModel =
fitcsvm(TrainingIvF_PCA_RED_5,TrainingIvD(:,i),'KernelFunction'
,Kern);
    [label score] = predict(SVMModel, TestingIvF_PCA_RED_5);
    Dis_Pred_PCA_RED_5(:,i) = label;
```

Listing 3.7 MATLAB Code for training and testing SVM

and testing the data for ‘Linear’ Kernel.

For the assessment of the performance of classifiers prediction values were calculated for all 3 kernels using all 4 datasets.

```
% Calculating Correct Predictions
Match_DDC = Dis_Pred_DDC == TestingIvD;
Match_PCA = Dis_Pred_PCA_FULL == TestingIvD;
Match_RED = Dis_Pred_PCA_RED == TestingIvD;
Match_5PC = Dis_Pred_PCA_RED_5 == TestingIvD;

% Initializing Wrong Predictions
Miss_DDC = 1 - Match_DDC;
Miss_PCA = 1 - Match_PCA;
Miss_RED = 1 - Match_RED;
Miss_5PC = 1 - Match_5PC;

% Initializing True Positives
TP_DDC = zeros([1 16]);
TP_PCA = zeros([1 16]);
TP_RED = zeros([1 16]);
TP_5PC = zeros([1 16]);

% Initializing True Negatives
TN_DDC = zeros([1 16]);
TN_PCA = zeros([1 16]);
TN_RED = zeros([1 16]);
TN_5PC = zeros([1 16]);

% InitializingFalse Positives
FP_DDC = zeros([1 16]);
FP_PCA = zeros([1 16]);
FP_RED = zeros([1 16]);
FP_5PC = zeros([1 16]);

% InitializingFalse Negatives
FN_DDC = zeros([1 16]);
FN_PCA = zeros([1 16]);
FN_RED = zeros([1 16]);
FN_5PC = zeros([1 16]);
```

```
for i = 1:1:56
    for j = 1:1:16
        if( Miss_DDC(i,j) )
            if( TestingIvD(i,j) )
                FN_DDC(j) = FN_DDC(j) + 1;
            else
                FP_DDC(j) = FP_DDC(j) + 1;
            end

        else
            if( TestingIvD(i,j) )
                TP_DDC(j) = TP_DDC(j) + 1;
            else
                TN_DDC(j) = TN_DDC(j) + 1;
            end
        end
        if( Miss_PCA(i,j) )
            if( TestingIvD(i,j) )
                FN_PCA(j) = FN_PCA(j) + 1;
            else
                FP_PCA(j) = FP_PCA(j) + 1;
            end
        else
            if( TestingIvD(i,j) )
                TP_PCA(j) = TP_PCA(j) + 1;
            else
                TN_PCA(j) = TN_PCA(j) + 1;
            end
        end

        if( Miss_RED(i,j) )
            if( TestingIvD(i,j) )
                FN_RED(j) = FN_RED(j) + 1;
            else
                FP_RED(j) = FP_RED(j) + 1;
            end
        else
            if( TestingIvD(i,j) )
                TP_RED(j) = TP_RED(j) + 1;
            else
                TN_RED(j) = TN_RED(j) + 1;
            end
        end
```

```
if( Miss_5PC(i,j) )
            if( TestingIvD(i,j) )
                FN_5PC(j) = FN_5PC(j) + 1;
            else
                FP_5PC(j) = FP_5PC(j) + 1;
            end
        else
            if( TestingIvD(i,j) )
                TP_5PC(j) = TP_5PC(j) + 1;
            else
                TN_5PC(j) = TN_5PC(j) + 1;
            end
        end
    end
end
```

Listing 3.8 MATLAB Code for checking predictions

PCA Reduced (10 and 5 Principal components) represent the reduction of dimensionality that can be made in two distinctive methods: by just keeping the most applicable factors from the first dataset or by exploiting the redundancy present inherently in the data.For the reduction of data dimensionality principal component analysis (PCA) is used, eitherusing the factors from the first set or using redundancy in the reoriented features (Bro & Smilde, 2014). Linear combinations are used to distinguish the variance-covariance structure of an arrangement of factors.

Lastly the predictions were analyzed to assess classifier performance in all cases, the reporting scores of accuracy, precision, sensitivity, specificity and F1-score were estimated through the MATLAB code shown in Listing 3.9 below.

```
Acc = zeros([16 4]);
Pre = zeros([16 4]);
Rec = zeros([16 4]);
Spe = zeros([16 4]);
F1 = zeros([16 4]);

Acc(:,1) = (TP_DDC + TN_DDC)/56;
Acc(:,2) = (TP_PCA + TN_PCA)/56;
Acc(:,3) = (TP_RED + TN_RED)/56;
Acc(:,4) = (TP_5PC + TN_5PC)/56;

Pre(:,1) = TP_DDC ./ (TP_DDC + FP_DDC);
Pre(:,2) = TP_PCA ./ (TP_PCA + FP_PCA);
Pre(:,3) = TP_RED ./ (TP_RED + FP_RED);
Pre(:,4) = TP_5PC ./ (TP_5PC + FP_5PC);

Rec(:,1) = TP_DDC ./ (TP_DDC + FN_DDC);
Rec(:,2) = TP_PCA ./ (TP_PCA + FN_PCA);
Rec(:,3) = TP_RED ./ (TP_RED + FN_RED);
Rec(:,4) = TP_5PC ./ (TP_5PC + FN_5PC);

Spe(:,1) = TN_DDC ./ (TN_DDC + FP_DDC);
Spe(:,2) = TN_PCA ./ (TN_PCA + FP_PCA);
Spe(:,3) = TN_RED ./ (TN_RED + FP_RED);
Spe(:,4) = TN_5PC ./ (TN_5PC + FP_5PC);

F1 = 2 * Pre .* Rec ./ (Pre + Rec);
```

Listing 3.9 Classifier Performance Estimation

The following chapter 4 discusses the results of the analysis performed in this chapter.

Chapter 4: RESULTS

In total four types of data sets were analyzed for 16 diseases using Support Vector Machine algorithm in MATLAB to check the effect of dimensionality reduction in the original data. The total training labels of all 16 diseases are shown in the table below.

Disease	Training	
	Positive	***Negative***
1	25	105
2	41	89
3	4	126
4	16	114
5	28	102
6	7	123
7	14	116
8	4	126
9	4	126
10	5	125
11	5	125
12	6	124
13	10	120
14	5	125
15	3	127
16	6	124

Table 4.1 Training Samples of all Diseases

The 4 datasetsanalyzed were labeled DDC (Direct Data Conversion), PCA22 (All 22 Principal Components used as new features), PCA10 (10 Principal Components were used as new features) and PCA5 (5 Principal Components were used as new features). In DDC, binary data converted to numerical because the SVM Classifier cannot operate on logical data. After this all three kernels Linear, Polynomial (with order 3) and Gaussian were applied to all four types of data separately, kernel is a shape of separation boundary each kernel has different physical/dimensional shape. Accuracy, precision, recall, specificity and f1 score were calculated for performance analysis and results are shown in the following tables;

Disease	Accuracy				Precision			
	DDC	PCA 22	PCA10	PCA5	DDC	PCA 22	PCA 10	PCA5
1	0.9107	0.9107	0.8393	0.7679	0.7778	0.7778	0.5385	0.4118
2	0.9464	0.9286	0.9464	0.9643	0.7692	0.7143	0.7692	0.9
3	1	1	0.9643	0.9464	1	1	1	---
4	0.8929	0.8929	0.8929	0.8393	0.6667	0.6667	0.6667	---
5	0.9643	0.9643	0.9643	0.9643	1	1	1	1
6	0.9821	0.9821	0.9286	0.9464	1	1	0	---
7	0.8571	0.8571	0.8571	0.8571	---	---	---	---
8	0.9643	0.9643	0.9643	0.9643	---	---	---	---
9	0.9286	0.9286	0.9643	0.9464	0.3333	0.3333	0.6667	0.5
10	0.9464	0.9464	0.9464	0.9464	---	---	---	---
11	0.9464	0.9464	0.9464	0.9464	---	---	---	---
12	0.9107	0.9107	0.9464	0.9464	0.25	0.25	---	---
13	0.9464	0.9464	0.9643	0.875	1	1	1	---
14	0.9643	0.9643	0.9643	0.9643	0.5	0.5	---	---
15	0.9643	0.9643	0.9643	0.9643	---	---	---	---
16	0.9643	0.9643	0.9643	0.9643	---	---	---	---

Table 4.2 Accuracy and Precision for the Linear Kernel

Disease	Recall				Specificity			
	DDC	PCA 22	PCA10	PCA5	DDC	PCA22	PCA10	PCA 5
1	0.7	0.7	0.7	0.7	0.9565	0.9565	0.8696	0.7826
2	1	1	1	0.9	0.9348	0.913	0.9348	0.9783
3	1	1	0.3333	0	1	1	1	1
4	0.6667	0.6667	0.6667	0	0.9362	0.9362	0.9362	1
5	0.8182	0.8182	0.8182	0.8182	1	1	1	1
6	0.6667	0.6667	0	0	1	1	0.9811	1
7	0	0	0	0	1	1	1	1
8	0	0	0	0	1	1	1	1
9	0.3333	0.3333	0.6667	0.3333	0.9623	0.9623	0.9811	0.9811
10	0	0	0	0	1	1	1	1
11	0	0	0	0	1	1	1	1
12	0.3333	0.3333	0	0	0.9434	0.9434	1	1
13	0.5714	0.5714	0.7143	0	1	1	1	1
14	1	1	0	0	0.963	0.963	1	1
15	0	0	0	0	1	1	1	1
16	0	0	0	0	1	1	1	1

Table 4.3 Recall and Specificty for Linear Kernel

Disease	F1 Score			
	DDC	PCA 22	PCA 10	PCA 5
1	0.7368	0.7368	0.6087	0.5185
2	0.8696	0.8333	0.8696	0.9
3	1	1	0.5	----
4	0.6667	0.6667	0.6667	----
5	0.9	0.9	0.9	0.9
6	0.8	0.8	----	----
7	----	----	----	----
8	----	----	----	----
9	0.3333	0.3333	0.6667	0.4
10	----	----	----	----
11	----	----	----	----
12	0.2857	0.2857	----	----
13	0.7273	0.7273	0.8333	----
14	0.6667	0.6667	----	----
15	----	----	----	----
16	----	----	----	----

Table 4.4 F1 Score values for the Linear Kernel

In the case of linear kernel the classifier had different performance for every disease and every condition, some of the diseases showed higher score in DDC data but some has better

results with other data types. Each data type shows significant results with overall accuracy of 94%, 98% specificity 69% F1-score. As expected the dimensionality reduction (PCA22, PCA10 and PCA5) tends towards reduced accuracy however this decrease in performance is really insignificant.

The table 4.5 below shows the Accuracy and Precision for a Polynomial Kernel with an order of 3.

Disease	Accuracy				Precision			
	DDC	PCA 22	PCA 10	PCA 5	DDC	PCA 22	PCA 10	PCA 5
1	0.9286	0.9286	0.8929	0.8571	0.75	0.75	0.7	0.5833
2	0.9286	0.9286	0.9464	0.9464	0.7143	0.7143	0.7692	0.7692
3	1	1	1	1	1	1	1	1
4	0.8214	0.8214	0.8214	0.8571	0.4	0.4	0.4	1
5	0.9286	0.9286	0.9643	0.9643	0.7692	0.7692	0.8462	0.9091
6	0.9464	0.9464	0.9464	0.9464	----	----	----	----
7	0.8036	0.8036	0.7857	0.8571	0.2857	0.2857	0	----
8	0.9643	0.9643	0.9643	0.9643	----	----	----	----
9	0.9643	0.9643	0.9643	0.9464	1	1	1	0.5
10	0.9464	0.9464	0.9464	0.9464	----	----	----	----
11	0.9643	0.9643	0.9286	0.9464	1	1	0.3333	----
12	0.9107	0.9107	0.9464	0.9286	0.3333	0.3333	0.5	0
13	0.9464	0.9464	0.9286	0.875	1	1	1	----
14	0.9643	0.9643	0.9643	0.9464	0.5	0.5	0.5	0
15	0.9643	0.9643	0.9643	0.9643	----	----	----	----
16	0.9107	0.9107	0.9286	0.9643	0.2	0.2	0.25	----

Table 4.5 Accuracy and Precision for the Polynomial Kernel

Disease	Recall				Specificity			
	DDC	PCA 22	PCA 10	PCA 5	DDC	PCA 22	PCA 10	PCA 5
1	0.9	0.9	0.7	0.7	0.9348	0.9348	0.9348	0.8913
2	1	1	1	1	0.913	0.913	0.9348	0.9348
3	1	1	1	1	1	1	1	1
4	0.2222	0.2222	0.2222	0.1111	0.9362	0.9362	0.9362	1
5	0.9091	0.9091	1	0.9091	0.9333	0.9333	0.9556	0.9778
6	0	0	0	0	1	1	1	1
7	0.25	0.25	0	0	0.8958	0.8958	0.9167	1
8	0	0	0	0	1	1	1	1
9	0.3333	0.3333	0.3333	0.3333	1	1	1	0.9811
10	0	0	0	0	1	1	1	1
11	0.3333	0.3333	0.3333	0	1	1	0.9623	1
12	0.6667	0.6667	0.3333	0	0.9245	0.9245	0.9811	0.9811
13	0.5714	0.5714	0.4286	0	1	1	1	1
14	1	1	1	0	0.963	0.963	0.963	0.9815
15	0	0	0	0	1	1	1	1
16	0.5	0.5	0.5	0	0.9259	0.9259	0.9444	1

Table 4.6 Recall and Specificity Values for the Polynomial Kernel

Disease	F1 Score			
	DDC	PCA 22	PCA 10	PCA 5
1	0.8182	0.8182	0.7	0.6364
2	0.8333	0.8333	0.8696	0.8696
3	1	1	1	1
4	0.2857	0.2857	0.2857	0.2
5	0.8333	0.8333	0.9167	0.9091
6	----	----	----	----
7	0.2667	0.2667	----	----
8	----	----	----	----
9	0.5	0.5	0.5	0.4
10	----	----	----	----
11	0.5	0.5	0.3333	----
12	0.4444	0.4444	0.4	----
13	0.7273	0.7273	0.6	----
14	0.6667	0.6667	0.6667	----
15	----	----	----	----
16	0.2857	0.2857	0.3333	----

Table 4.7 F1Scores for the Polynomial Kernel

Lastly the tables 4.8 through 4.10 show the performance of the Gaussian Kernel based SVM.

Disease	Accuracy				Precision			
	DDC	PCA22	PCA 10	PCA 5	DDC	PCA 22	PCA 10	PCA 5
1	0.9643	0.9643	0.9107	0.8036	1	1	0.7778	0.4667
2	0.9643	0.9643	0.9643	0.9643	0.8333	0.8333	0.8333	0.8333
3	0.9464	0.9464	0.9464	0.9464	----	----	----	----
4	0.8214	0.8214	0.8393	0.8393	0.3333	0.3333	0.5	----
5	0.9643	0.9643	0.9821	0.9821	0.9091	0.9091	0.9167	1
6	0.9464	0.9464	0.9464	0.9464	----	----	----	----
7	0.875	0.875	0.8571	0.8571	1	1	----	----
8	0.9643	0.9643	0.9643	0.9643	----	----	----	----
9	0.9643	0.9643	0.9821	0.9821	1	1	1	1
10	0.9464	0.9464	0.9464	0.9464	----	----	----	----
11	0.9464	0.9464	0.9464	0.9464	----	----	----	----
12	0.9464	0.9464	0.9464	0.9464	0.5	0.5	----	----
13	0.9286	0.9286	0.9464	0.875	1	1	1	----
14	0.9821	0.9821	0.9643	0.9643	0.6667	0.6667	----	----
15	0.9643	0.9643	0.9643	0.9643	----	----	----	----
16	0.9643	0.9643	0.9643	0.9643	0.5	0.5	----	----

Table 4.8 Accuracy and Precision of the SVM based on Gaussian Kernel

Disease	Recall				Specificity			
	DDC	PCA22	PCA10	PCA5	DDC	PCA22	PCA10	PCA5
1	0.8	0.8	0.7	0.7	1	1	0.9565	0.8261
2	1	1	1	1	0.9565	0.9565	0.9565	0.9565
3	0	0	0	0	1	1	1	1
4	0.1111	0.1111	0.2222	0	0.9574	0.9574	0.9574	1
5	0.9091	0.9091	1	0.9091	0.9778	0.9778	0.9778	1
6	0	0	0	0	1	1	1	1
7	0.125	0.125	0	0	1	1	1	1
8	0	0	0	0	1	1	1	1
9	0.3333	0.3333	0.6667	0.6667	1	1	1	1
10	0	0	0	0	1	1	1	1
11	0	0	0	0	1	1	1	1
12	0.3333	0.3333	0	0	0.9811	0.9811	1	1
13	0.4286	0.4286	0.5714	0	1	1	1	1
14	1	1	0	0	0.9815	0.9815	1	1
15	0	0	0	0	1	1	1	1
16	0.5	0.5	0	0	0.9815	0.9815	1	1

Table 4.9 Recall and Specificity of the SVM based on Gaussian Kernel

Disease		F1 Score			
		DDC	PCA22	PCA10	PCA5
1		0.8889	0.8889	0.7368	0.56
2		0.9091	0.9091	0.9091	0.9091
3		----	----	----	----
4		0.1667	0.1667	0.3077	----
5		0.9091	0.9091	0.9565	0.9524
6		----	----	----	----
7		0.2222	0.2222	----	----
8		----	----	----	----
9		0.5	0.5	0.8	0.8
10		----	----	----	----
11		----	----	----	----
12		0.4	0.4	----	----
13		0.6	0.6	0.7273	----
14		0.8	0.8	----	----
15		----	----	----	----
16		0.5	0.5	----	----

Table 4.10 F1 Score of an SVM Classifier based on a Guassian Kernel

The results were tabulated for both Polynomial and Gaussian Kernels. The Polynomial Kernel based SVM shows an overall average accuracy of across all diseases 93% and the Gaussian Kernel based SVM shows an overall average accuracy across all diseases of 94%. Similar to the Linear Kernel, the difference between the different datasets was insignificant for both Polynomial and Gaussian Kernels.

All diseases have significant result but some of the diseases show extraordinary results because the training data was skewed because the number of positive sample images was far less than the number of negative samples for the disease. This was catered for slightly by the biased random sampling for the training set in the data, however the overall effect of over-fitting is prominent in the data.

This also meant that for some diseases after splitting the data randomly for training and testing the positive samples were significantly reduced which resulted in lack of information.

Hence some of the diseases (e.g. 5 images for disease 15) resulted in heavily biased training. A considerably low number of positive samples in data is the major contributing factor for the low recall scores of all classifiers in all datasets for almost all diseases.

In dimensionality reduction it was noted that as move towards 5 Principal Components instead of 10 or 22, the accuracy showed a slight decreasing trend but number of false positive increased due to which number of true positive also increased.

Chapter 5: DISCUSSION & CONCLUSION

Using digital retinal imaging various clinical disorders of the human eye can be examined noninvasively. Automated image analysis can assist in the early detection of a disease based on its earlier symptoms. In order to facilitate the physicians in making better diagnosis, different measurements can be done to detect the retinal diseases quickly and make accurate suggestions based on machine learning algorithms. The aim of this study was to develop an automated tool for classification of retinal diseases in STARE database using SVM classifier in MATLAB, this approach resulted in an average accuracy of 94%.

The proposed algorithm results in significant accuracies forthe classification of disease with problems like over-fitting because of biased random sampling for the training data. This also tries to adjust the disproportionate number of positive sample images for some of the diseases. A few diseases were not detected at all due to a heavily biasing training, resulting in wrong classification, it is also expected since the training set consists of 3 positive samples and 127 negative samples for some of the diseases.

The effect of dimensionality reduction on performance was studied, as the system moves towards the lower dimensionality, accuracies generally reduced however this reduction in accuracy was insignificant. Among all three kernels linear, polynomial and gaussian there were insignificant differences between results as gaussian perform just slightly better for our dataset.

On accounts of heavily biased training, resulting in false negative, a problem arose for the diseases with less number of positive samples in the database. This can be improved by applying slightly adjusting the cutoff criterion such as, increasing the number of positive samples (up to 15 instead of 5). This can also be corrected by changing the classifier type or using more features.

REFERENCES

Abràmoff, M. D., Garvin, M. K., & Sonka, M. (2010). Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering*, *3*, 169–208. https://doi.org/10.1109/RBME.2010.2084567

Arora, P., & Sharma, N. (2015). A Study on Object Classification and Filteration Approaches. *Journal of Computer Science and Mobile Computing*, *4*(5), 92–98.

Baker, M. L., Hand, P. J., Wang, J. J., & Wong, T. Y. (2008). Retinal Signs and Stroke: Revisiting the Link Between the Eye and Brain. *Stroke*, *39*(4), 1371–1379. https://doi.org/10.1161/STROKEAHA.107.496091

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A Training Algorithm for Optimal Margin Classiiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 144–152. https://doi.org/10.1.1.21.3818

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Anal. Methods*, *6*(9), 2812–2831. https://doi.org/10.1039/C3AY41907J

Cassin, B., & Rubin, M. L. (n.d.). *Dictionary of eye terminology*. Retrieved from https://www.triadpublishing.com/eyedictionary.shtml

Chernecky, C. C., & Berger, B. J. (2004). *Laboratory tests & diagnostic procedures*. Saunders.

Gharaibeh, N. Y. (2016). A Novel Approach for Detection of Microaneurysms in Diabetic Retinopathy Disease from Retinal Fundus Images. *Computer and Information Science*, *10*(1), 1. https://doi.org/10.5539/cis.v10n1p1

Gottlieb, J. L. (2002). Age-Related Macular Degeneration. *JAMA*, *288*(18), 2233. https://doi.org/10.1001/jama.288.18.2233

Hutchinson, A., McIntosh, A., Peters, J., O'Keeffe, C., Khunti, K., Baker, R., & Booth, A. (2000). Effectiveness of screening and monitoring tests for diabetic retinopathy--a systematic review. *Diabetic Medicine : A Journal of the British Diabetic Association*, *17*(7), 495–506. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10972578

Jonas, J. B., Schneider, U., & Naumann, G. O. (1992). Count and density of human retinal

photoreceptors. *Graefe's Archive for Clinical and Experimental Ophthalmology = Albrecht von Graefes Archiv Fur Klinische Und Experimentelle Ophthalmologie*, *230*(6), 505–510. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1427131

Kale, P. B., & Janwe, N. (2017). Detection and Classification of Diabetic Retinopathy in Color Fundus Image. *International Journal for Research in Advanced Computer Science and Engineering*, (7), 1–4.

Karthikram, A., Kavya, P., Keerthika, P., & Veenmathi, T. (2016). Detection of Retinal Diseases Based on SVM Classifier. *International Journal of Innovative Research in Computer and Communication Engineering*, *3*(2), 1548–1554. https://doi.org/10.15680/IJIRCCE.2016.

Kumar, B. R., & P, S. M. (2017). Automatic Detection of Eye Cataracts and Disease Classification Using Hybrid Techniques. *Nternational Journal on Future Revolution in Computer Science & Communication Engineering*, *3*(12), 304–308. Retrieved from http://www.ijfrcsce.org/download/browse/Volume_3/December_17_Volume_3_Issue_12/1514555629_29-12-2017.pdf

Lin, D. Y., Blumenkranz, M. S., Brothers, R. J., & Grosvenor, D. M. (2002). The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *American Journal of Ophthalmology*, *134*(2), 204–213. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12140027

Mangai, J. A., Nayak, J., & Kumar, V. S. (n.d.). A Novel Approach for Classifying Medical Images Using Data Mining Techniques. Retrieved from https://www.semanticscholar.org/paper/A-Novel-Approach-for-Classifying-Medical-Images-Mangai-Nayak/f7dd3f123078b124815bf168e69099f2104cde93

Masland, R. H. (2001). The fundamental plan of the retina. *Nature Neuroscience*, *4*(9), 877–886. https://doi.org/10.1038/nn0901-877

Niemeijer, M., van Ginneken, B., Cree, M. J., Mizutani, A., Quellec, G., Sanchez, C. I., … Abramoff, M. D. (2010). Retinopathy Online Challenge: Automatic Detection of Microaneurysms in Digital Color Fundus Photographs. *IEEE Transactions on Medical*

Imaging, *29*(1), 185–195. https://doi.org/10.1109/TMI.2009.2033909

Niki, T., Muraoka, K., & Shimizu, K. (1984). Distribution of capillary nonperfusion in early-stage diabetic retinopathy. *Ophthalmology*, *91*(12), 1431–1439. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6084212

Nirmala, S. R., Nath, M. K., & Dandapat, S. (2011). Retinal Image Analysis : A Review. *International Journal of Computer & Communication Technology*, *2*(Vi), 11–15.

Rath, P. (2017). CONTRIBUTION OF IMAGE PROCESSING AND MACHINE LEARNING FOR AUTOMATED ANALYSIS OF RETINAL VESSELS : A REVIEW. *International Journal of Recent Innovation in Engineering and Research*, *2*(2), 1–7.

Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G. P., & Mariotti, S. P. (2004). Global data on visual impairment in the year 2002. *Bulletin of the World Health Organization*, *82*(11), 844–851. https://doi.org//S0042-96862004001100009

Saine, P. J., & Tyler, M. E. (2002). *Ophthalmic Photography: Retinal Photography, Angiography, and Electronic Imaging*. Butterworth-Heinemann.

Shveta, S., & Kaur, G. (2015). Review on: Detection of Diabetic Retinopathy using SVM and MDA. *International Journal of Computer Applications*, *117*(20), 1–3. https://doi.org/10.5120/20667-2485

Singer, D. E., Nathan, D. M., Fogel, H. A., & Schachat, A. P. (1992). Screening for diabetic retinopathy. *Annals of Internal Medicine*, *116*(8), 660–671. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1546868

Sowmya, R. (2016). A survey on automatic detection of retinal disorders from fundus images. *International Journal of Research in Computer Applications and Robotics*, *4*(1), 9–15.

The STARE Project. (n.d.). Retrieved August 26, 2018, from http://cecas.clemson.edu/~ahoover/stare/

Umesh, L., Mrunalini, M., & Shinde, D. S. (2016). Review of image processing and machine learning techniques for eye disease detection and classification. *International Research Journal of Engineering and Technology*, *3*, 547–551.

Walsh, A. C., Wildey, R. C., Lara, C., Ouyang, Y., & Sadda, S. R. (2010). Detection of Fundus Abnormalities Using 3D-OCT versus Mydriatic Color Fundus Imaging. In *Investigative Ophthalmology & Visual Science* (Vol. 51, pp. 3863–3863). C.V. Mosby Co. Retrieved from https://iovs.arvojournals.org/article.aspx?articleid=2372489

Wu, J., Waldstein, S. M., Montuoro, A., Gerendas, B. S., Langs, G., & Schmidt-Erfurth, U. (2016). Automated Fovea Detection in Spectral Domain Optical Coherence Tomography Scans of Exudative Macular Disease. *International Journal of Biomedical Imaging, 2016*, 1–9. https://doi.org/10.1155/2016/7468953

Wyszecki, G., & Stiles, W. S. (Walter S. (2000). *Color science : concepts and methods, quantitative data, and formulae*. John Wiley & Sons. Retrieved from https://www.wiley.com/en-us/Color+Science%3A+Concepts+and+Methods%2C+Quantitative+Data+and+Formulae%2C+2nd+Edition-p-9780471399186