University of KwaZulu-Natal

Doctoral Thesis

Retinal Blood Vessel Segmentation using Random Forest Gabor Feature Selection and Automatic Thresholding

Author:
Mandlenkosi Victor Gwetu

Supervisors:
Prof. Jules Raymond Tapamo
Prof. Serestina Viriri

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

School of Mathematics, Statistics and Computer Science
University of KwaZulu-Natal
Durban, South Africa

March 2019

UNIVERSITY OF KWAZULU-NATAL

COLLEGE OF AGRICULTURE, ENGINEERING AND SCIENCE

All the research described in this thesis was performed at the University of KwaZulu-Natal under the supervision of Prof. Jules Raymond Tapamo and Prof. Serestina Viriri. I hereby declare that this thesis is my own original work and it has not been submitted for a degree in this or any other university. All referenced work in this thesis has been either acknowledged or appropriately cited.

Signed: _____

Mandlenkosi Victor Gwetu

Date: _____

As the candidate's supervisor, I have approved this dissertation for submission.

Signed: _____

Prof. Jules Raymond Tapamo

Date: _____

Signed: _____

Prof. Serestina Viriri

Date: _____

UNIVERSITY OF KWAZULU-NATAL

COLLEGE OF AGRICULTURE, ENGINEERING AND SCIENCE

Declaration of Plagiarism

I, Mandlenkosi Victor Gwetu, declare that:

- The thesis and the work presented therein is my original research work. Due credit and acknowledgement has been given to all the work of others that is mentioned or referred to in this thesis.
- This thesis or any part thereof has not been submitted for any degree, qualification or examination at any other university or institution.
- All the published work of others that has been consulted in this thesis is clearly credited.
- The source of all the external work that is quoted in this thesis is clearly stated. With the exception of such quotations, the text in this thesis was entirely written by me.
- This thesis does not contain any unacknowledged data, images, graphical illustrations or other information from other sources including the internet.
- The research work emanating from this thesis that is derived from collaborative work with others has been clearly stated along with the exact contributions from myself and all the collaborators.

Signed: _____

Mandlenkosi Victor Gwetu

Date: _____

UNIVERSITY OF KWAZULU-NATAL

COLLEGE OF AGRICULTURE, ENGINEERING AND SCIENCE

Declaration of Authorship

Some of the research work presented in this thesis has contributed towards research papers that are either under review or published. The details of such research papers are listed below along with the exact contributions of each collaborating author.

- Mandlenkosi Victor Gwetu[1], Jules Raymond Tapamo[2], and Serestina Viriri[3]. Segmentation of retinal blood vessels using normalized Gabor filters and automatic thresholding. *South African Computer Journal*. Volume 55, December 2014, pp 12-24.
- Mandlenkosi Victor Gwetu[1], Serestina Viriri[2], and Jules Raymond Tapamo[3]. Random Forests with a Steepend Gini-Index Split Function and Feature Coherence Injection. *South African Computer Journal*. Submitted August 2017. Under Review.
- Mandlenkosi Victor Gwetu[1], Serestina Viriri[2], and Jules Raymond Tapamo[3]. Purity and Out of Bag Confidence Metrics for Random Forest Weighting. *Computational Collective Intelligence, ICCCI 2018. Lecture Notes in Computer Science, Volume 11055. Springer.*

Signed: _____

Mandlenkosi Victor Gwetu

Date: _____

[1]Author1 was responsible for conceptualization, literature review, experimental design, implementation, write up and review editing.

[2]Author2 performed conceptual design, critical analysis of methodology and proof reading.

[3]Author3 assisted with conceptual design, verification of results and review analysis.

To my hero and father, Aaron Mfanyana Gwetu. Your memory and legacy of discipline, hard work, excellence and integrity have inspired me through this journey.

"If you can not do great things, do small things in a great way."

Napoleon Hill

Contents

List of Figures

List of Tables

Abbreviations

AI	**A**rtificial **I**ntelligence
ARM	**A**ge-**R**elated **M**aculopathy
AUC	**A**rea **U**nder the **C**urve
BoVW	**B**ag **o**f **V**isual **W**ords
BP	**B**lood **P**ressure
BRB	**B**lood **R**etinal **B**arrier
CAD	**C**omputer **A**ided **D**iagnosis
CBIR	**C**ontent **B**ased **I**mage **R**etrieval
CPGG	**C**ummulative **P**urity **G**ap **G**ain
DFP	**D**igital **F**undus **P**hotographs
DMO	**D**iabetic **M**acular **O**edema
DR	**D**iabetic **R**etinopathy
DRIVE	**D**igital **R**etinal **I**mages for **V**essel **E**xtraction
DVE	**D**eepened **V**alley **E**mphasis
EML	**E**xtreme **M**achine **L**earning
EPC	**E**ndothelial **P**rogenitor **C**ell
ETDRS	**E**arly **T**reatment of **D**iabetic **R**etinopathy **S**tudy
FCM	**F**uzzy **C**-**M**eans **C**lassifier
FNR	**F**alse **N**egative **R**ate
FPR	**F**alse **P**ositive **R**ate
FROC	**F**ree-response **R**eceiver **O**perating **C**haracteristic
GMM	**G**aussian **M**ixture **M**odel
GPU	**G**raphics **P**rocessing **U**nit
GVE	**G**aussian **V**alley **E**mphasis
ICDSSDR	**I**nternational **C**linical **D**isease **S**everity **S**cale for **D**iabetic **R**etinopathy

ICER	**I**ncremental **C**ost-**E**ffectiveness **R**atio
IDF	**I**nternational **D**iabetics **F**ederation
IRMA	**I**ntra-**R**etinal **M**icrovascular **A**nomaly
kNN	**k**-**N**earest **N**eighbour
LBP	**L**ocal **B**inary **P**attern
LDP	**L**ocal **D**irectional **P**attern
LSI	**L**atent **S**emantic **I**ndexing
MDA	**M**ean **D**ecrease **A**ccuracy
MDD	**M**inimum **D**istance **D**iscriminant
MDI	**M**ean **D**ecrease **I**mpurity
MFR	**M**atched **F**ilter **R**esponse
NCD	**N**on **C**ommunicable **D**isease
NPDR	**N**on-**P**roliferative **D**iabetic **R**etinopathy
OCT	**O**ptical **C**oherence **T**omography
OOB	**O**ut **O**f **B**ag
PDF	**P**robability **D**ensity **F**unction
PDR	**P**roliferative **D**iabetic **R**etinopathy
PGG	**P**urity **G**ap **G**ain
RF	**R**andom **F**orest
RGB	**R**ed **G**reen **B**lue
RMDI	**R**elative **M**ean **D**ecrease **I**mpurity
ROC	**R**eceiver **O**perating **C**haracteristic
ROI	**R**egion **O**f **I**nterest
ROP	**R**etinopathy **O**f **P**rematurity
SGNVE	**S**elective **G**aussian **N**eighbourhood **V**alley **E**mphasis
SOM	**S**elf **O**rganizing **M**ap
STARE	**S**tructured **A**nalysis of the **RE**tina
SURF	**S**peeded-**U**p **R**obust **F**eature
SVE	**S**elective **V**alley **E**mphasis
SVM	**S**upport **V**ector **M**achine
TNR	**T**rue **N**egative **R**ate
TPR	**T**rue **P**ositive **R**ate
UCI	**U**niversity of **C**alifornia, **I**rvine

VE	**V**alley **E**mphasis
WESDR	**W**isconsin **E**pidemiologic **S**tudy of **D**iabetic **R**etinopathy
WHO	**W**orld **H**ealth **O**rganisation
YLD	**Y**ears **L**ost due to **D**isability

Abstract

Successful computer aided diagnosis of ocular diseases is normally dependent on the accurate detection of components such as blood vessels, optic disk, fovea and microaneurysms. The properties of these components can be indicative of the presence and/or severity of pathology. Since most prevalent forms of ocular diseases emanate from vascular disorders, it is expected that accurate detection of blood vessels is essential for ocular diagnosis.

In this research work, we investigate several opportunities for improvement of retinal blood vessel segmentation with the hope that they will ultimately lead to improvement in the diagnosis of vascular related ocular diseases. We complement existing work in this domain by introducing new Gabor filter features and selecting the most effective of these using Random Forests feature selection. The actual segmentation of blood vessels is then done using an improved automatic thresholding scheme based on the preferred Gabor feature.

We propose Random Forest (RF) feature ranking algorithms that demonstrate reliable feature set partitions over several University of California, Irvine (UCI) datasets. To circumvent instances of unreliable rankings, we also propose feature rank and RF strength correlation as an alternative indicator. Of the four proposed Gabor features, the maximum magnitude response is confirmed as the most effective, as is the general trend in previous literature. The proposed Selective Valley Emphasis thresholding technique achieves identical segmentation results to the legacy approach while improving on computational efficiency. Sensitivity and specificity outcomes of up to 76.8% and 97.9% as well as 78.8% and 97.8% are achieved on the DRIVE and STARE datasets, respectively.

Acknowledgements

In writing this thesis and conducting its research work, the author has received support from several fronts and would like to acknowledge the following sources for all the support rendered throughout the duration of this study:

- My Creator, Lord and Saviour, **Jesus Christ**; for the gift of life, my abilities and the strength to overcome the numerous challenges faced on this academic journey.
- My wife, Thando; children and family at large; for believing in me. I am grateful for their patience and unwavering support throughout my many hours of research.
- My supervisors, Prof. Jules Raymond Tapamo and Prof. Serestina Viriri; for their reliable advice, encouragement and support throughout this study. I have been inspired by their passion for academic excellence and their commitment to supervising this thesis.
- The University of KwaZulu-Natal and the School of Mathematics, Statistics and Computer Science for the opportunity to conduct this study and the computational resources to run all its related experiments.
- All the reviewers that have assessed/critiqued the publications and presentations emanating from this study. I am a lot wiser as a result of the scientific advice and recommendations made in the peer review process.
- All the researchers and institutions that have availed their scientific data sets to the public for experimentation. Some of those used in this study include the STARE, DRIVE and UCI data sets.
- All the open source and proprietary institutions that have made community editions of their high quality software available for public download and use. Some of those used in this study include: the Cimg library, Opencv, Textlive, Texmaker, Dia, Octave, Netbeans, Mingw, Dlib and CMake.

Chapter 1

General Introduction

1.1 Introduction

One of the long term effects of the physiological processes of diabetes is the deterioration of retinal blood vessels, leading to possible loss of eye sight [1]. This condition is known as diabetic retinopathy and is currently the leading cause of blindness in the working class globally [2]. Although early laser treatment of damaged ocular blood vessels can be effective, delayed therapy may have side effects such as abnormal vascular tissue sprouting and scar tissue formation on the retina [3]. The presence of abnormal components on the retina may cause it to malfunction or detach from its membrane leading to compromised vision [4]. It is essential to have an effective means of early diagnosis of diabetic retinopathy, since the early stages of the disease are asymptomatic and its long term effects are potentially irreversible [5].

Although retinal structures are minute, modern digital imaging systems offer very high resolutions, yielding in vivo retinal surface snapshots that are adequate for clinical analysis. During manual clinical analysis, each of these images is compared against healthy and pathological retinal specimen by several experts for quality assurance purposes. Diagnosis is essentially achieved through a process of matching instances or portions of retinal images to previously encountered cases. Besides being subjective, this approach is prone to slow processing, inconsistency and human error.

There has been a recent concerted effort towards automating this process through the use of image processing and pattern recognition techniques in order to overcome these challenges. A digital approach towards retinal diagnosis involves multiple steps such as image enhancement, registration, segmentation and diagnostic classification. This study focuses on retinal image enhancement and segmentation sub-tasks. We particularly focus

on retinal blood vessels as they are an essential indicator of the presence or onset of ocular disease. This work touches on a wide range of topics that may ultimately affect quality of blood vessel segmentation. After establishing the trends in previous literature, we propose new features for vessel enhancement and scrutinise existing ones. We not only focus on the effectiveness of methods but also propose ways of improving their efficiency.

1.2 Motivation

1.2.1 General Motivation for the Study

From its recent studies, the International Diabetics Federation (IDF) reports the following statistics about diabetes: out of every two people with diabetes, one does not know they have it; there are currently 387 million people living with diabetes; and an increase of over 205 million is expected by the year 2035[1]. These and other similar findings have motivated several countries to embark on national treatment and research initiatives to reduce the prevalence of diabetes and its related illnesses. Diabetic Retinopathy (DR) is caused by diabetes and is the primary cause of blindness in America and India [6].

The Aravind Eye Hospital[2] in India offers a good example of the challenges experienced during large scale DR screening programmes. It is one of the largest eye care facilities in the world that has embarked on a public initiative to detect diabetic retinopathy at an early stage. This task has proven to be extremely laborious in a setting where less than 15 doctors are responsible for manually diagnosing diabetic retinopathy in over 2 million retinal images per year [6]. Given that the majority of the screened patients show no signs of diabetic retinopathy, the initiative can be considered as counter productive since resources are mainly being used to confirm the absence of the illness as opposed to treating the confirmed cases. The automation of this diagnosis could greatly ease the pressure on the limited ophthalmologic experts and increase the efficiency of screening programmes.

In its 2011 bulletin [7], the World Health Organization (WHO) reports that 36 million lives are lost each year globally due to Non Communicable Diseases (NCDs) and encourages the identification of cost-effective interventions for the reduction of morbidity and mortality [5]. The need for interventions is pronounced in developing countries as they have to strike a balance between huge social grant responsibilities and infrastructure investment. Although South Africa is a developing country, it has a well established social grant system that currently has over 16 million beneficiaries [8] and a budget of over

[1] http://www.idf.org/worlddiabetesday/toolkit/gp/facts-figures
[2] http://www.aravind.org/

R259 billion [9]. These grants include a disability grant which is offered to individuals that are unable to derive a sustainable income due to circumstances such as ill-health and disability.

In a recent study, Khan et al. [5] provide evidence for the cost effectiveness of a national screening programme for diabetic retinopathy in South Africa. They calculate the Incremental Cost-Effectiveness Ratio (ICER) for diabetic retinopathy in 2013 as R10,500 per year for every case of blindness that could be averted. This is less than the 2013 annual disability grant of R12,120[3]. They also argue that if all cases of diabetic retinopathy related blindness were successfully screened and treated, 18,590 Years Lost due to Disability (YLD) [10] would be saved in South Africa. All these forecasts are based on a programme in which diabetic retinopathy is manually detected through non-mydriatic fundus photography; we postulate that greater savings in costs and time would be realised through automated screening.

Since blood vessels are one of the prominent landmarks of human retinal images, they play an important role in the diagnosis and clinical management of ocular diseases in general. In order to detect any deterioration of the retina or assess the progression of previously identified disorders, a longitudinal analysis of retinal images from the same patient is required. Image registration [11] attempts to align images captured from the same retina, at different times under different camera settings. The accuracy of such alignment initiatives relies on effective landmark identification. For example, a lesion which exists at a particular location relative to a prominent blood vessel, can easily be monitored if it is correctly located at subsequent clinical examinations. Retinal image registration can also assist with detecting anomalies such as the onset of vessel proliferation, which is an indication of advanced DR. In addition, retinal blood vessel segmentation can assist with the calculation of vascular properties such as diameter, texture, branch factor and tortosity which are important indicators of DR [12, 13]. In summary, segmentation of retinal blood vessels is an important precursor to effective diagnosis of DR and other ocular diseases. It is therefore essential to improve its accuracy in order to ensure reliability of tasks such as vessel detection and characterization, which are used in automated diagnosis.

1.2.2 Motivation for Adopted Approach

It is evident from the numerous previous studies that endeavour to automatically segment retinal images and/or diagnose diabetic retinopathy [4, 14, 15], that computer algorithms are yet to achieve the levels of accuracy exhibited by medical experts. A

[3]The current annual disability grant is R20,340 [9].

recent review [16] of modern developments in retinal image analysis, noted that future work needs to focus on image processing techniques as they have a direct bearing on relevant clinical information within the images. The effectiveness of Gaussian based filters is highlighted, with techniques such as Gabor wavelets [17] achieving accuracies of up to 94.7% on the DRIVE dataset.

While many other studies have explored automatic segmentation of retinal images with reasonable success and enlightening findings, this does not preclude the need for further investigation. Given the critical nature of medical diagnostics, it is imperative that more research be conducted to find new and more effective ways of improving the accuracy of present initiatives. Medical imaging is a vast field of science that draws expertise from domains such as image processing, Artificial Intelligence (AI), data mining, engineering and medicine. This means that its improvement can be achieved from various angles.

This study proposes a unique multi-step approach towards retinal image segmentation, that incorporates techniques such as Gabor filters, Random Forests (RFs) and automatic thresholding. An effort is made to either improve or maximize the potential of each of these techniques, before aggregating them. Hence, this study is not only relevant to the field of retinal image segmentation but it also contributes to image processing and machine learning in general. For example, our improvements on RF algorithm are not specific to retinal image segmentation but may well be applied to other fields seeking improved ensemble classification. Although some of these techniques may have been used separately in previous retinal image segmentation studies, it is envisaged that improving and combining them will ultimately yield an overall improved segmentation result.

1.3 Problem Statement

The typical sequence of steps towards fundus image segmentation includes image pre-processing, feature extraction and ultimately pixel categorization [15]. These processes operate at various levels of abstraction and subsequent processes are dependent on the output of the preceding ones. Therefore, the accuracy at each level is essential for the formulation of a reliable final classification. The accurate segmentation of vascular tissue in retinal images remains a challenging and unsolved problem due to the complex distribution of pixel intensities and potential existence of pathological landmarks in these images [12].

In this thesis, we focus on the problem of how to maximize Gabor filters for effective and efficient automatic retinal blood vessel segmentation. Gabor filters are capable of producing multiple responses and it is possible that combining these responses in new

ways could lead to more effective features. Improving the efficiency of Gabor filtering, feature selection and automatic thresholding methods may lead to a compromise in effectiveness due to the loss of vital information. On the other hand, a reduction in redundancy could produce more focused methods that yield improved effectiveness.

1.4 Thesis Objectives

The main aim of this study is to improve the segmentation of retinal blood vessels through the fusion of Gabor filters, RFs feature selection and automatic thresholding techniques. The specific objectives that will be pursued are as follows:

- To formulate efficient and reliable RF feature selection.
- To maximize the use of Gabor filter banks in retinal image segmentation.
- To improve the efficiency and effectiveness of Valley Emphasis automatic thresholding.

The fulfillment of these objectives is guided by experimental tests based on the following research questions:

- Does the use of purity gap gain and relative variable importance in RF induction yield an efficient and reliable feature selection mechanism?
- Do RF feature selection statistics confirm the superior discriminative ability of Gabor maximum response when compared with the proposed secondary features in the context of retinal images?
- How does the efficiency and effectiveness of the legacy Valley Emphasis method compare against its proposed variants in the context of retinal blood vessel segmentation?

1.5 Contributions of the Thesis

This section outlines the originality in our research and how this thesis makes a distinct contribution to the knowledge and understanding of the subject of retinal image segmentation. The major contributions of our approach towards retinal image blood vessel segmentation include formulation of new techniques and applying existing ones in new ways, for improved performance and/or understanding. We proceed to elaborate on each of these contributions.

1.5.1 Random Forest Feature Selection

Although legacy RFs [18] have been recommended as a viable feature selection strategy, their drawbacks in this context include penalizing correlated variables and a bias of their gini importance score towards multi-valued variables [19]. In this study we propose a new metric called the purity gap gain score, as an efficient alternative to the gini variable importance measure. In addition, we employ the concept of relative variable importance as a means of reducing bias towards multi-valued features.

1.5.2 Maximizing Gabor Features

The traditional application of Gabor filters [20] for image enhancement selects the maximum response from a filter bank whose members differ in terms of properties like orientation and wavelength. Although this enhanced image is generally easier to segment, it is worth pondering whether there may be benefit in aggregating several responses from a filter bank. In the traditional approach, only the maximum response is used for further image processing, while other responses are discarded. We firstly introduce new features such as response convexity and the orthogonal to maximum response, for use in the context of retinal image enhancement. We subsequently apply our RF feature selection scheme which confirms maximum response as the most effective feature out of the proposed small set of Gabor features, for later use in automatic thresholding.

1.5.3 Valley Emphasis Thresholding Adaptation

Although the Valley Emphasis Thresholding (VET) method [21] has proven to be an effective thresholding technique in industrial fault detection, its application to retinal image segmentation remains largely unexplored. We review some of the contemporary adaptations to VET and propose hybrid alternatives which are essentially adaptations from some of these methods. We identify some redundancy in the VET method, which is successfully reduced in our selective approach.

1.6 Thesis Outline

The remainder of this thesis is organised as follows. Chapter 2 presents the background on the human retina and Chapter 3 reviews the state-of-the-art in retinal image segmentation. Chapter 4 investigates the problem of applying RFs to feature selection. In Chapter 5, Gabor filter enhancement features and algorithms are designed, implemented

and the most effective feature is selected based on discriminative ability in retinal images. Chapter 6 reviews existing adaptations of the VET method, explores a hybrid alternative and presents the ultimate task of retinal image segmentation. Chapter 7 draws conclusions from this study and proposes possible areas for extending this work.

Chapter 2

The Human Retina

2.1 Introduction

Although Computer-Aided Diagnosis (CAD) based on medical images has been explored in various contexts such as dermatology [22], hepatology [23] and psychiatry [24], its application to ophthalmology [25, 26] has gained prominence due to its non-invasive and reliable image acquisition. Analysis of ocular images is not only useful for the diagnosis of eye disorders like cataracts and glaucoma but also conditions such as diabetics, cardiovascular disease, HIV and cancer [27] which may occur elsewhere in the body. Some of these conditions present with manifestations in the eye that signal the onset or presence of a secondary disorder. At times, these conditions are asymptomatic and can only be reliably diagnosed after examination by a specialist. This attests to the ancient proverb: "the eyes are the windows to the soul" and it is one of the reasons why regular visits to eye specialists for check-ups are encouraged.

This chapter presents a background on the human retina, which illustrates its detailed anatomy, complex physiology and associated disorders.

2.2 Retinal Anatomy

One of the earliest recorded studies of the human eye was by Aristotle, a Greek scientist from approximately 400BC [28]. He postulated that objects adjust the air around them and this alteration is propagated through the air to the eye, enabling vision [29]. Although studies of the eye in the past were limited due to the inability to inspect its anatomy at microscopic level, to date, the behaviour of ocular components can be researched owing to sophisticated imaging and surgical instruments. As new discoveries on

the anatomy of the eye continue to be made, prospects for more effective eye treatment increase [30].

Figure 2.1 shows a typical cross sectional view of the human eye and retina which shows numerous components which play a role in facilitating vision. The path travelled by light through the eye intersects with the cornea, aqueous humor, lens and vitreous humor before it reaches the retina [31]. The cornea and lens play the leading role in the refraction of light so that it converges onto the retina [32]. This effort is complimented by the medium in the aqueous and vitreous humor which also serves to focus light onto the retina, albeit to a lesser extent [33]. The iris adjusts the diameter of the pupil depending on the intensity of incident light such that it dilates in conditions of low light and constricts when exposed to abundant light [34]. The portion of the lens that is adjacent to the pupil is either thickened or stretched by the cilliary muscles depending on the proximity of the object of interest. Since the fovea is responsible for central vision, light from an object of interest needs to be focused on to it to enable detailed vision.

The retina derives its name from the latin word for 'net' which is befitting as its primary role is to absorb light and convert it into biochemical signals of visual information [34]. Vertebrate retinas are made up of 3 layers of neural cells which are joined by 2 synaptic layers: the outer and inner plexiform layers [36]. The outer nuclear layer contains photoreceptor cells which are responsible for grey scale night vision and full colour day light vision respectively [34]. The outermost boundary of the retina is an epithelium layer which contains melanin pigment in order to absorb excess light that is not captured by the photoreceptor cells [32]. The inner nuclear layer contains bipolar cells which propagate vertical signal transmission; and lateral neurons: horizontal and amacrine cells, which create links between vertical neurons resulting in increased perception of contrast [34]. The innermost part of the retina is made up of ganglion cells that relay visual stimulation to the brain using their axons via the optic nerve [32]. Blood vessels that enter the eye through the optic disk perforate the retina so as to nourish its inner layers [36]. These vessels are however absent in the fovea, which is a depression in the central part of the retina containing the highest density of cone photoreceptors so as to facilitate central vision [36].

Modern retinal examination is usually conducted through a digital fundus cameras and computer equipment that are typically set up as shown in Figure 2.2. The connection of the camera to a computer enables scalable storage, analysis and transmission of captured retinal images. A more advanced utilization of this equipment is for CAD, which can facilitate automated diagnosis of ocular diseases [4]. Although fully automated CAD systems for ocular diseases have not yet been realised [4], there are several existing

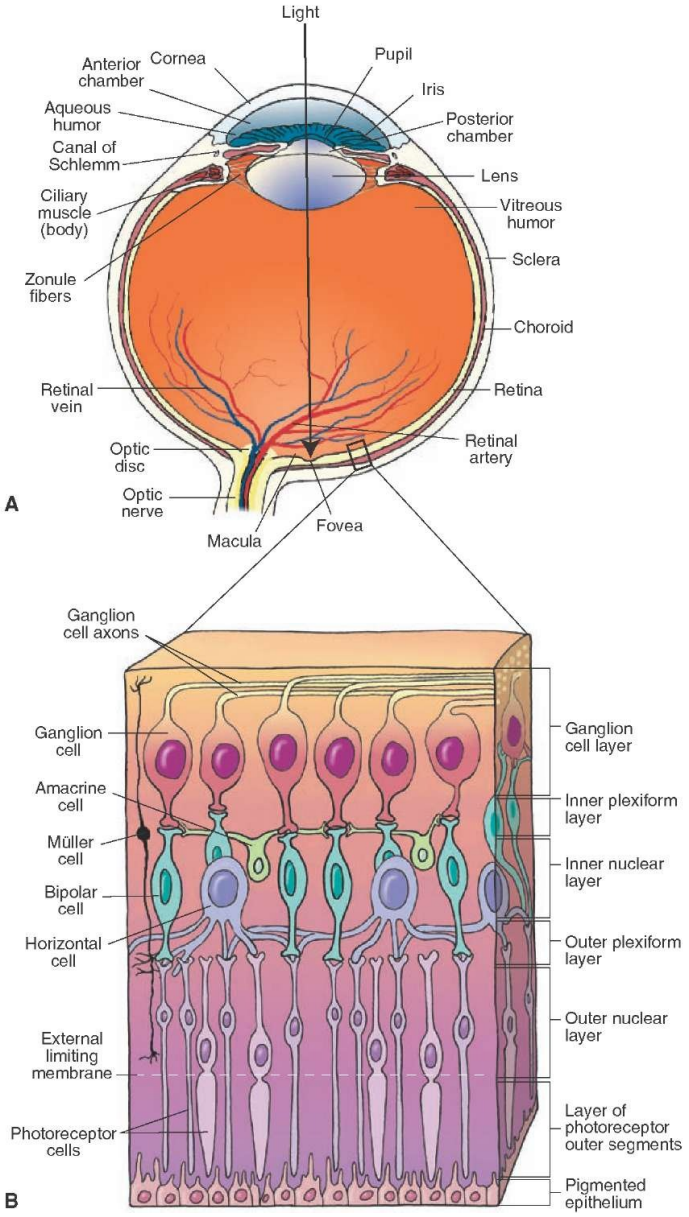


FIGURE 2.1: The anatomy of the human eye and retina [32]

semi-automated initiatives such as the ATLANTIA and AVATA laboratories run by IMED [37] in Singapore which focus on pathology detection and visual acuity assessment respectively.

In some cases digital fundus cameras have the advantage of capability to operate in a non-mydriatic setting in which eye examination is carried out without the need for medically induced pupil dilation [38]. Since pupil dilation takes time to induce and wear off, medical practitioners and patients may prefer non-mydriatic examinations which require less time to complete. Research has also shown non-mydriatic digital fundus photography to have higher inter-examination agreement among ophthalmologists and to require less training time for effective use when compared with other alternatives such as

FIGURE 2.2: Retina examination through using a fundus camera [35]

ophthalmoscopy [39]. Digital fundus cameras however have the disadvantage of limited mobility, given their bulkiness. This makes it difficult to use them on patients that are too ill to sit in front of the camera [39]. Current efforts aim to mitigate this challenge through mobile phone based fundus photography [40, 41]. Although this avenue has a lot of potential for clinical diagnosis and self assessment, its current limitation is the processing power and pixel resolution of mobile devices.

2.3 Diabetic Retinopathy

This section elaborates on the Diabetic Retinopathy (DR) disorder by outlining its systemic factors, pathophysiological processes, symptoms, diagnosis and treatment.

2.3.1 Diabetes Mellitus

DR is a retinal blood vessel disorder that is considered to be the most common complication of Diabetes mellitus [42]. Diabetic mellitus is a general term referring to metabolic

diseases that occur as a result of inadequate insulin secretion and/or diminished tissue response to insulin [43]. Since insulin is responsible for the metabolism of carbohydrates, fat and proteins, its deficient action results in excess glocuse levels in the blood, a condition known as hyperglycemia that leads to vascular degeneration [44].

There are two types of diabetes: Type 1 and Type 2 which typically manifest in childhood and adulthood years respectively. Type 1 diabetes (also known as juvenile-onset diabetes) is mainly caused by a hereditary autoimmune destruction of β cells of the pancreas which are responsible for insulin production [43]. Type 2 diabetes (also known as adult-onset diabetes) is caused by a combination of reduced tissue sensitivity to insulin action and insufficient compensatory insulin secretion [43]. Type 2 diabetes accounts for approximately 90% diabetes cases world wide and its prevalence is on the rise due to modern day factors such as increasing obesity, sedentary life styles and poor diet [42]. The long term effects of hyperglycemia of diabetes include damage of blood vessels leading to the dysfunction of various vital organs such as the eyes, kidneys and heart [43].

2.3.2 Pathology of Diabetic Retinopathy

Due to the intricate anatomy of the retina, DR is the most common microvascular complication of diabetes induced hyperglycemia [42]. Previous medical research shows that persistent high blood glucose levels lead to the interrelated processes of oxidative stress, vascular permeability and inflammation within the retina [1, 45, 46]. Oxidative stress refers to the body's inability to detoxify reactive molecules that contain oxygen leading to cell component damage [47]. The intra-retinal microvasculature forms part of the Blood Retinal Barrier (BRB) which consists of tightly joined cells that prevent certain substances from entering retinal tissue [42]. Vascular permeability occurs when there is a leakage of macromolecules into retinal tissue as a result of a compromise in the BRB [48]. This permeability also renders white blood cells unrestricted to enter surrounding tissue and activate hormones called prostaglandins to create blood clots in an attempt to heal any damaged tissue [49]. Inflammation apparently occurs as a result of neurovascular dysfunction [42] that yields aberrant cytokines which are normally responsible for cell signalling [50]. In this abnormal state cytokines interfere with the insulin signalling mechanism causing insulin resistance and spiked blood sugar [49]. These heightened sugar levels in turn trigger white blood cells to attack, causing inflammation trademarks such as increased blood flow, capillary permeability and swelling [51].

It can be inferred that DR is essentially a condition characterized by improper function of the retina either as a direct result of hyperglycemia symptoms or the retina's attempts

to mediate them [1, 52]. From an optical perspective, opaque products from retinal capillary leakage and subsequent neovascularization lead to impaired vision due to light obstruction [1]. From a cellular standpoint, fluid accummulation within the retina may alter action potentials required for the transmission of visual signals to the brain and alter neuronal metabolism leading to improper neurotransmission [1, 53, 54]. Moreover it is apparent how hyperglycemia leads to a chronic vicious cycle in which retinal tissue self destructs presenting symptoms such as blood clots and damaged vasculature that are characteristic of DR [42].

2.3.3 Symptoms of Diabetic Retinopathy

The diagnosis of DR is traditionally done with the aid of a process known as fluorescence angiography that requires a patient to be injected with a dye called fluorescein [55]. The desired effect of this dye is to ultimately cause blood vessels and associated lesions in the retina to have high contrast for easier visual analysis. The typical manifestations that characterize DR are as follows:

- red lesions - range in size from small microaneurysms, to moderate dot haemorrhages, and larger blot haemorrhages [15]. A lesion is defined as a circumscribed area of tissue that is altered due to pathology or injury [56]. Diabetic retinal microaneurysms are saccular dilations in capillaries caused by factors such as localized weakening of capillary walls and fat deposition [57]. The continued progression of microaneurysms and vascular degeneration eventually leads to blood vessel occlusion (blockage) and haemorrhage (bleeding) [58].
- whitish lesions - include lipid exudates (also known as hard exudates) and cotton wool spots (also known as soft exudates) [15]. An exudate is defined as a fluid released from the body due to pathological processes and containing high levels of protein, cells or tissues [56]. Hard exudates are yellowish, well defined granules that are formed by leakage of fluid from blood vessels and by products of inflammatory edema (swelling) [59]. Soft exudates are cloud-like formations from nerve fibre fluid leakage in the retina [42]; they occur mainly when diabetes mellitus is complicated by arterial hypertension [60]. The leakage of nerve fibre fluid in the retina is caused by ischaemia: a pathological condition occurring due to inadequate blood flow to tissue resulting in failure to meet cellular energy demands [61]. Ischemia deprives the affected tissue of oxygen, metabolic substrates, and removal of waste products; ultimately inducing injury to the tissue [61].
- proliferative vascularization - includes neovasculization which occurs when the body attempts to circumvent damaged vascular tissue by forming fragile and leaky

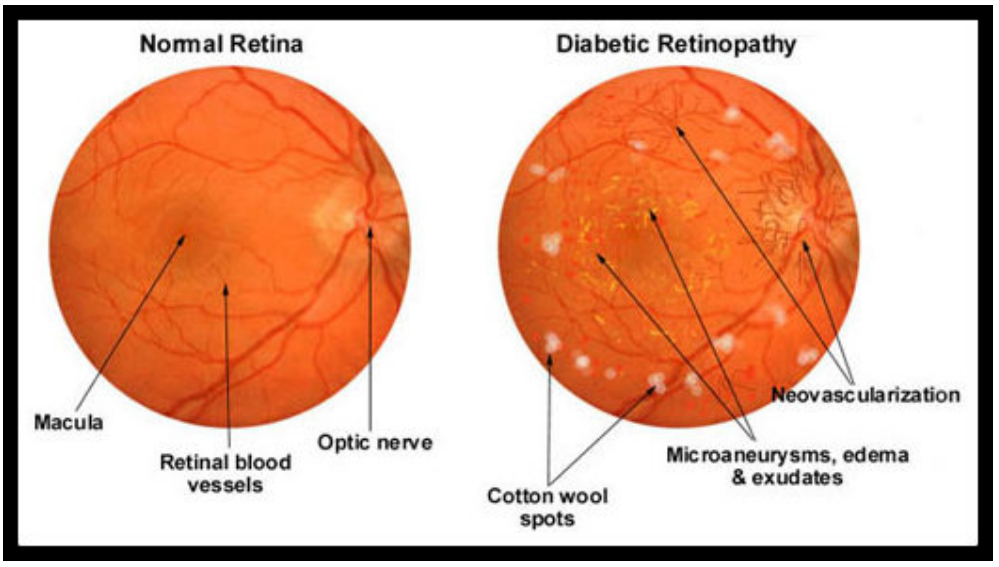


FIGURE 2.3: Normal Retina (left) and Typical DR Manifestations (right) [64]

vessels which if untreated can develop fibrous adhesive tissue leading to tractional retinal detachment [42]. Proliferative Diabetic Retinopathy (PDR) is at times preceded by venous abnormalities such as beading, looping and Intra-Retinal Microvascular Anomalies (IRMAs) [42, 62].

Figure 2.3 contrasts a visualization of a normal human retina against one showing typical DR symptoms such as microaneurysms, cotton wool spots and neovascularization which are discussed above. The typical manifestations of DR can be detected through visual clinical examinations without the need for surgical operations because of their distinctive optical characteristics; this makes their detection an interesting domain for machine learning and image processing techniques. The underlying factors of these manifestations, such as vessel permeability and inflammation are more difficult to observe using fundus photography. Although retinal edema/swelling can potentially be detected using cross-sectional imaging techniques such as Optical Coherence Tomography (OCT), equipment costs are normally exorbitant [63].

One of the main setbacks of DR is that it is largely asymptomatic until the latter stages of the disease when there is a high risk of irreversible retinal damage and loss of vision [5]. It is therefore essential to undergo regular screening in order to detect the disease at its onset and enable more effective treatment.

2.3.4 Classification of Diabetic Retinopathy

In order to prescribe adequate therapy, the screening of DR needs to go beyond merely detecting the presence of pathology, instead it should categorize, classify and rank the

severity of the disorder [63]. Categorizing fundus images from different patients with similar pathology requires a screening framework that is detailed enough to avoid ambiguity and yet flexible enough to allow for generalization. Giving common labels and demarcations to pathological clusters of related cases allows for universal interpretation of symptoms and findings. While most lesions do not cause a direct obstruction to the visual information that is aggregated by the retina, those lying within or proximal to the macula pose a significant threat to vision [42]. Such cases should be prioritized for treatment, hence it is essential to have some aspect of severity staging in any DR screening programme. Although several initiatives have previously been made by medical experts to standardize the classification of DR, a consensus of opinions is difficult to attain [65].

If realized, generic classifications could be beneficial for facilitating unambiguous communication between various diabetic practitioners [63], verifying/validating clinical findings and providing a reference guide for the diagnosis of complex cases. One of the earliest such initiatives was the Airlie House symposium in 1968 which brought a group of experts together in order to consolidate existing knowledge of DR and develop a standard classification of the disease [63, 66]. The Early Treatment of Diabetic Retinopathy Study (ETDRS) [67] subsequently extended this classification by:

1. providing clinically labelled reference ophthalmologic stereo photographs;
2. classifying DR into 13 complexity levels ranging from absence of retinopathy (level 10) to severe hemorrhaging and retinal detachment (level 85); and
3. introducing and empirically defining the concept of Clinically Significant Macular Edema (CSME) as an advanced stage of DR [63, 68].

Efforts are ongoing to adapt existing DR grading/classification systems so that they are simpler and more practical for clinical use. A fairly recent example is the International Clinical Disease Severity Scale for Diabetic Retinopathy (ICDSSDR) [69] which is based on the ETDRS and the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) [63]. The ICDSSDR used standard photographs as benchmarks to recognize five stages of DR: no apparent retinopathy, mild Non-Proliferative Diabetic Retinopathy (NPDR), moderate NPDR, severe NPDR and Proliferative Diabetic Retinopathy (PDR) [69]. The simple and yet objective approach of the ICDSSDR makes it easy to recall and less prone to ambiguity [63].

Regardless of the grading scheme used, DR seems to be traditionally classified into two major clinical categories: NPDR and PDR, where the latter is considered more severe than the former. PDR is characterized by proliferation of delicate and leaky new blood vessels (neovascularization) that can form fibrous and adhesive connective tissue

over time, leading to complications such as vitreous haemorrhage and tractional retinal detachment [42]. The severity of NPDR is typically based on the prevalence and location of retinal lesions such as microaneurysms, cotton wool spots and IRMAs [42]. Lesions close to the macular are an indication of excessive vascular permeability and swelling, a condition known as Diabetic Macular Oedema (DMO) which is a big threat to central visual acuity and is the most common cause of blindness in diabetes [42, 70].

2.3.5 Treatment of Diabetic Retinopathy

The treatment options for DR are implemented by either controlling the systemic factors such as glycaemic and Blood Pressure (BP) levels that exacerbate it, or providing therapy to counteract/minimize the ocular damage caused by its pathological processes [42]. Therapy options include laser photo-coagulation [67], vitreoretinal surgery [71] and pharmaceutical therapy [42]. Although laser photo-coagulation is used to seal leaking blood vessels and destroy lesions, it inevitably damages retinal tissue in the process resulting in some level of visual loss [72].

Vitreoretinal surgery is normally reserved for sight threatening complications of PDR such as severe vitreous hemorrhage and retinal detachments as it is an expensive and complicated procedure that requires a high level of surgical precision [71, 72]. Pharmaceutical options include anti-VEGF therapy to suppress the over-expression of the VEGF signal protein in order to reduce BRB compromise [42]. Additionally, steroid treatment and Endothelial Progenitor Cell (EPC) therapy have been explored for stimulating vessel and ischaemic retina repair with some success [42, 72].

Current treatment options for DR do not guarantee restoration of any lost vision and effective treatment without side effects. It is therefore imperative to prioritize the development of mechanisms to assist in the early detection of DR so that its adverse effects can be minimized. Given the significant efforts that have been made to standardize the classification of DR based on its common visual symptoms and the universal acceptance of these standards, the stage is set for automated DR detection. Furthermore, it remains to be proved whether current medical image processing methods and techniques are adequate for solving this complex task effectively.

2.4 Conclusion

This chapter has highlighted the importance of the human eye as a convenient indicator of general wellness, since it can be inspected non invasively. The retina takes center stage

in CAD since it is directly responsible for light absorption and its blood vessels can be visually analysed. Although DR retinopathy is a leading cause of blindness, it is largely asymptomatic in its earlier stages. Hence, its screening programmes are a public health imperative in most developing countries. The fact that there are well documented DR classification schemes such as ETDRS and ICDSSDR, makes its automated diagnosis a possibility. Since current treatment options do not guarantee the return of lost sight, effective early diagnosis becomes very crucial. The next chapter looks into previous efforts towards automatic retinal image analysis.

Chapter 3

Literature Review

3.1 Introduction

The use of retinal digital imaging is now standard practise within DR screening programmes, for example, the UK National Screening Committee recommends that all diabetic patients aged 12 years and above, be screened annually using digital fundus photography [15]. Using digital retinal imaging allows the operations of screening programmes to be decentralized as images may be captured at locations convenient to patients then electronically transferred to remote sites for analysis by experts in a manner that promotes quality assurance [15]. Since grading digital retinal images for the presence of DR is largely a pattern recognition task, it is not surprising that a lot of research has recently gone into automatic DR detection. [14, 15].

When a retinal image is captured using a digital fundus camera, it is stored in an electronic format that can be interpreted as a matrix of pixel intensity values [14]. These values can be in a grey scale or a vectorized colour scheme such as the Red Green Blue (RGB) format. Under normal/ideal circumstances, each pixel may be surrounded by a neighbourhood of other pixels with either highly contrasting or similar intensity values depending on whether it lies within or at the boundary of a homogeneous object. Medical images are prone to have poor contrast and varying levels of illumination, making it difficult to precisely delineate their components; hence, a preprocessing stage is generally required to normalize image brightness and reduce noise [14]. Subsequent processing may involve the segmentation of retinal landmarks so that features can be extracted from them for characterizing the pathological properties of the image [14, 15]. CAD of DR can therefore be thought of as a process that uses a decision making protocol based on clinical domain knowledge or specimen images, to ascertain the level of pathology in a retinal image [4].

This section presents a narrative review of some previous and recent endeavours towards CAD in DR. The review period which spans from 1970 to date, is divided into 3 chronological groups: the early studies (1970 - 1989), the intermediate studies (1990 - 2009) and the recent studies (2010 to date) for referral convenience and to enable chronological comparison. Common retinal image datasets that serve as a benchmark for objective comparison and verification of reported results are explored, and typical evaluation metrics are introduced.

3.2 Early Studies

The history of ocular CAD dates back to the earliest paper in this area published in 1973 by Matsui et al. from the Central Research Laboratory, Hitachi Ltd., which focused on segmenting normal and abnormal retinal vessels using contour lines [73, 74]. Prompted by the wide usage of fundus photograph evaluation in mass screening programmes, researchers: Yamamoto et al., from the same institution, realised that the examination of these photographs is a pattern recognition task which requires meticulous attention and the increasing demand for this practise could exceed the capacity of available practitioners [75]. They noted the apparent lack of research on automatic analysis of fundus photographs and discussions about the technical possibility of this idea. An algorithm was therefore developed to detect retinal changes caused by hypertension and sclerosis over time using a longitudinal series of fundus photographs. The changes in question included streak patterns such as artery narrowing and vein crossings; and spot patterns such as haemorrhages and exudates. The algorithm used included smoothing, chromatic and contour line filters; and it relied on human input to indicate the location of vein crossings. The algorithm was tested against the findings of a physician using nine fundus photographs and it was shown to give positive results. The momentum from this initiative was however lost as the adoption and maturation of the general idea of ocular CAD was hampered by the lack of scientific evidence for disease characteristics and classification at the time [73].

In an effort to expedite mass diagnosis of adult disorders like hypertension and diabetes, Akita and Kuga studied the following problems in 1982: the extraction of retinal blood vessels; recognition of arteries and veins; and the detection of abnormalities such as haemorrhages, exudates and arterio-venous crossings [76]. Their work was presented as an extension of the work done by Yamamoto et al. [75], as it did not depend on human intervention. The approach used was based on chromatic discrimination and mathematical modelling of components. Experiments on fundus images were presented

visually without the use of evaluation metrics that give a score of overall effectiveness of the approach.

Some efforts during this period focused on specific aspects of automatic ocular analysis as opposed to carrying out a complete diagnosis. Lay et al. specifically focused on the automatic detection and counting of microaneurysms in retinal angiograms using a binary transformation filter and morphological shape operators [77]. The approach was tested on 25 angiograms and a good correlation with manual analysis by experts is reported. Peli and Lahav [78] formulated an objective method to measure the chronological progression of drusen in the retinal pigment epithelium in order to augment research in to automated Age-Related Maculopathy (ARM) detection. A reproducibility of 6.1% was reported on the evaluation of repeated photographs.

The progress of computer vision and artificial intelligence in the late 1980s in areas such as remote sensing inspired efforts to establish complete medical diagnosis systems [79]. Katz et al. designed software to automatically segment normal and abnormal retinal structures using features such as colour, shape and texture [79]. The output was a list of possible diagnoses and their associated confidence levels. The extension of this work led to the Structured Analysis of the REtina (STARE) project [80], which developed an image understanding system to support ophthalmologists by providing automated analysis and diagnosis of retinal images. The system initially identified structural components of the retina such as blood vessels, the optic disk and lesions; it then used a knowledge description language to provide symbolic representation of DR manifestations in an image, then integrate quantitative and contextual data extracted from the images to train neural networks on how to pathologically classify images. The dataset that was used contained 9 diagnoses, each of which had 11 patients. The neural network classifier that was deduced from this dataset, achieved an accuracy rate of 83% for diagnosing new cases. Other effort at the time included that of Ward et al., who developed an image analysis system to measure exudate prevalence in fundus photographs [81]. Preprocessing operators were used to enhance contrast and reduce variation in illumination while the actual segmentation of exudates was based on grey level thresholding. The system was found to be effective when tested on standard photographs from the ETDRS, which led to its use for monitoring the effectiveness of laser treatment in diabetic retinopathy patients at the time.

3.3 Intermediate Studies

Although efforts towards CAD of ocular diseases continued into the 1990s, Sharp et al. reported that they were unable to complete a 1998 systematic literature review on the

use of digital imaging technology in DR detection due to the lack of advancement in this field [15, 82]. Although the focus of most ocular CAD in the 1990s was on DR [15], the automatic detection of other disorders such as glaucoma [83, 84], hypertension/sclerosis [85] also received some research attention. The DR studies during the period 1990 - 2009 saw a steady increase in diversified research with various points of emphasis such as the segmentation of specific retinal components, investigation of AI techniques and comprehensive diagnosis [15]. In addition, the methods used were technically more advanced than those used in the early studies which relied mainly on grey level thresholding.

3.3.1 Optic Disk Detection

Computer aided tasks on the optic disk are normally centered around its localization and subsequent segmentation. While localization generally aims to identify the center of the optic disk, segmentation determines its boundary in order to compute the disk diameter which may be used for locating the fovea [15]. A simple approach to optic disk localization is to identify the largest cluster of bright pixels in a fundus image [14, 86]. This approach is however inadequate in the presence of other fundus components such as exudates, with similar intensities [87]. As such, a combination of features like colour and morphology may be effective in the presence of destructors [15]. Other approaches that were used for optic disk localization in the intermediate studies included algorithms for vessel confluence detection at the optic disk [14, 88]; and finding the location with the highest intensity variance of adjacent pixels within a neighbourhood window [89].

Segmentation of the optic disk is useful for spatially cognisant retinal image registration which supports blood vessel tracking [90] and localization of the fovea [91]. Examples of optic disk segmentation studies during the early 1990s include the work of Mendels et al. [92, 93] who investigated the effectiveness of morphological operators combined with active contours; they reported an accurate segmentation of the optic disk on all the images in a test set of 9 fundus photographs. The approach adopted by Osareh et al. [94] implemented parametric contour models derived from clinical reference standards with an accuracy rate of 90%. Other variants of the contour model approach include the use of a gradient vector flow technique by Xu and Prince [95], and elliptic deformable model with edge density sensitivity by Lowell et al. [91]; to detect the rim of the optic disk.

3.3.2 Blood Vessel Detection

Given the significant role that blood vessels play in retinal pathology, it is not surprising to learn that the structured literature review conducted by Winder et al. [15] found

retinal blood vessel segmentation to be the most popular focus area of DR automation studies in the period: 1998 - 2008. Although various approaches have been proposed for vascular segmentation, the independent reviews by Patton et al. in 2006 [14] and Winder et al. in 2009 [15] both came to the conclusion that efforts at the time could be categorized into 4 general methods: matched filtering, morphological processing, vessel tracking and pixel-based classification.

3.3.2.1 Blood Vessel Matched Filters

Hoover et al. proposed an automated method to segment the retinal blood vessel network in fundus images using local and global vessel properties [96]. A matched filter was used to enhance pixels with a neighbourhood profile that matched predefined blood vessel cross sectional Gaussian models. The enhanced images were thresholded at various grey levels yielding a Receiver Operating Characteristic (ROC) curve with a true positive rate of up to 75% when compared against hand-labeled ground truth segmentations from the STARE [80] dataset. When using matched filters, a two-dimensional linear filter is typically rotated in 30-45° increments in order to achieve rotation invariance [97] and the maximum filter correlation value is taken as the Matched Filter Response (MFR) [15].

The drawbacks of the matched filter include the computational overhead from applying a large convolution filter over several orientations; a lack of flexibility in the Gaussian models in order to dynamically match different vessel sizes; and false positives due to exudates and reflection artefacts in low contrast images [15]. Efforts to address some of these shortcomings include the use of non-linear tram-line filters which plot a line along the center of a vessel segment and contrast its longitudinal profile against that of adjacent tram-lines on either side [98]. The challenge with tram-line filters is determining the appropriate length of structuring element to use as long tram-lines may not fit well into tortuous vessels [14]. In 2007, Ricci and Perfetti proposed the use of a rotated fixed length line detector that is sensitive to vessel center line pixels and two orthogonal short line operators to construct a feature vector for classification using supervised Support Vector Machines (SVMs) [99]. The behaviour of the line detector in the presence of a central reflex was shown to be satisfactory and the overall algorithm was evaluated through ROC analysis, resulting in an Area Under the Curve[1] (AUC) of 0.9558 and 0.9602 for the DRIVE and STARE datasets, respectively [12, 99].

Staal et al. extended the concept of matched filters to form a ridge-based vessel segmentation method that detects image ridges using a kernel whose weights are derivatives

[1] The ROC curve.

of a Gaussian [100]. The ridges approximate line elements corresponding to vessel center lines and pixels neighbouring a particular line element form a patch, whose feature vectors are classified by a kNN classifier into vascular or non-vascular tissue categories. An AUC of 0.952 and an accuracy of 0.944 [2] were achieved on the STARE dataset. In 2009, Osareh and Shadgar proposed scale and orientation selective Gabor filters as a three-dimensional extension to matched filters [101]. The extracted features were classified using a Gaussian Mixture Model (GMM) and SVMs, yielding an AUC of 0.974, a sensitivity of 96.5% and a specificity of 97.1% on a dataset of 90 normal and abnormal images. Further experimentation on the DRIVE dataset yielded an AUC of 0.965 with an average processing time of 30 seconds per image.

3.3.2.2 Blood Vessel Morphological Processing

In line with the biological concept of morphology, which refers to the shape and structure of organisms; morphological concepts in image processing refer to shape representation and manipulation [12]. Morphological processing can therefore be looked at from two different perspectives:

1. Characterizing existing object shapes using metrics such as circularity ratio and eccentricity which quantify the similarity of a given boundary to a formal shape [102].

2. Enhancing existing object shapes through morphological operators that have various effects such as opening, closing, erosion and dilation [103]. These operators can also be combined to form algorithms such as top-hat and watershed transformations that are commonly used for medical image segmentation [12].

Examples of vessel shape characterization include the work of Zan and Klein in 2001 who managed to detect vessels using shape properties, connectivity and curvature [104]. Although the authors motivate the recognition of linear segments and connectivity as morphological processing, the concepts used are very similar to line detection except that in this case lines are detected using morphological operators with linear structuring elements. Similarly, cross curvature evaluation is synonymous with matched filtering save for the fact that a Laplacian is used to model the curvature of a vessel cross sectional profile. Cross curvature evaluation was used to prune false positives that may have been enhanced by linear segment detection based on the fact that vessel components tend to have linearly coherent curvature. Evaluation was done on the DRIVE dataset, yielding

[2] This is comparable to the effectiveness of the human observed who achieved an accuracy of 0.947 for the same task.

an accuracy and AUC of 0.9377 and 0.8984 respectively. Heneghan et al. extended this work by deriving structural properties such as vessel width and tortuosity which are useful indicators of conditions like Retinopathy Of Prematurity (ROP)[3] [105]. Given the variety of shape characterization features available to the image processing community, there seems to be room for the application of more advanced shape detectors, especially for purposes such as differentiating between lesions and blood vessels.

Mendonca and Campillo used Difference of Offset Gaussian (DoOG) filters in conjuction with scalable morphological reconstruction to segment retinal vasculature [106]. A modified top-hat filter with adjustable size was used to enhance vessels of various widths and four different orientations of the DoOG filter were used to extract pixel features for building a vessel center line classification model. Segmentation was carried out using an iterative region growing algorithm initiated from the extracted vessel center lines. The effectiveness of the approach was demonstrated through average accuracy values of 0.9463 and 0.944 on the DRIVE and STARE datasets respectively. Yang et al. used a multi-oriented top-hat filter in combination with opening and closing morphological operators to enhance retinal blood vessels [107]. The Fuzzy C-Means (FCM) clustering algorithm was used to segment the vessels followed by a post-processing stage to reduce weak edges and noise. The approach was tested on the STARE dataset and shown to yield results comparable to a specialist's segmentation using graphical plots of accuracy ratios.

3.3.2.3 Blood Vessel Tracking

Tracking algorithms can be used to segment image objects based on the connectedness and homogeneity of constituent pixels. Since detection is typically based on representative regions like object center lines as opposed entire objects, tracking is essentially reduced to local characterization and segmentation of such a region from a given starting point. Although tracking algorithms work at the level of single vessels, they provide accurate vessel width measurements and can be used to analyse attributes such as bifurcation and tortuosity [12]. Tracking algorithms can therefore perform vessel analysis efficiently as they focus only on the regions of interest. They are however dependent on seed points and can be misled by inconsistent center line profile and branching.

Gao et al. used twin Gaussian functions to model vessel cross-sectional profiles and track them in the direction of the prevailing longitudinal axis [108]. The tracking method was used to measure average vessel segment widths and recommended for possible retinal pathology diagnosis which is typically done by analysing bifurcation geometry properties

[3] Retinopathy in infants.

such as vessel diameters, branch angles and branch lengths [108, 109]. Lalonde et al. presented a novel tracking algorithm that is based on vessel edges as opposed to center lines [90]. After enhancement using the Canny edge operator, vessels were tracked by monitoring the connectivity of twin borders. Discontinuities in vessel edges were used to trigger new seed points for subsequent tracking which assisted with detecting bifurcations and disjointed vessels. Although several other studies such as [110–112] explored vessel tracking during the intermediate period (1990 - 2009), the lack of common objective evaluation criteria to assess tracking effectiveness is conspicuous.

3.3.2.4 Blood Vessel Pixel Classification

The lowest level at which machine learning techniques can be applied to image component segmentation is at pixel level. In the supervised approach, pre-labelled pixels are required for training a classifier to recognize typical class instances through a process of generalization. In the unsupervised approach, instances are normally clustered based on mutual proximity in the feature space and optimization algorithms that determine an effective partitioning of datasets into a predetermined[4] number of classes. Examples of supervised machine learning approaches include Naive Bayes, SVMs, kNN and Neural Networks while unsupervised machine learning is exemplified by C-Means Clustering, SOMs and GMMs [113, 114]. Pixel classification methods offer simplified feature extraction as classifiers can be fed with raw pixel values or filter outputs from a local neighbourhood. They are however prone to computational overhead as they operate at pixel level and may require long training times to capture the complex distinguishing inter-pixel relationships in a given class.

In 2004, Niemeijer et al. presented their work which used the DRIVE dataset to compare four previously published non-supervised[5] retinal vessel segmentation methods against a supervised pixel classification method [115]. The features used in the supervised approach included the image green channel as well as filter output from the Gaussian and its derivatives (first and second) at multiple scales. This yielded 31 features which were normalized to zero mean and unit variance then fed to a kNN classifier where $k = 30$. Preliminary investigations showed that filter output was superior to raw pixel values within a neighbourhood and the kNN classifier was more effective than linear and quadratic classifiers. The kNN classifier was used to deduce a probability map for each pixel, enabling soft classification which could be further quantized for binary segmentation through thresholding. Experiments on the DRIVE test set resulted in a Maximum Average Accuracy (MAA) rate and an AUC of 0.9416 and 0.9294 respectively.

[4] Although more complex algorithms can partition datasets without knowledge of a target number of classes, in most cases this number is known before hand.

[5] The authors describe this as “not trained with examples”.

Although this segmentation approach was superior to the other methods considered, the performance of the human observer was significantly better.

In 2006, Soares et al. proposed the use of the inverted image green channel and maximum responses from rotated two-dimensional Morlet wavelets (or Gabor wavelets) at different scales as feature vectors for supervised retinal vessel segmentation [116]. Training set examples were used to estimate a linear combination of class conditional probability density functions for a Bayesian classifier in order to model complex decision surfaces that effectively distinguish blood vessel pixels. Optimal results are achieved when 20 Gaussians are used to model each class. In this setting, an AUC and accuracy of 0.9598 and 0.9467 are obtained on the DRIVE dataset respectively while the STARE dataset yields an AUC and accuracy of 0.9651 and 0.9474 respectively. A significant drawback of the study is that in all experiments, one million pixels were randomly sampled to train the classifiers; it is therefore difficult to re-enact this study.

3.3.3 Retinal Pathology Detection

A review of automatic DR detection algorithms by Faust et al. [117] showed that previous studies in the intermediate era explored a wide spectrum of pathological features drawn from retinal components such as blood vessels, exudates, microaneurysms and haemorrhages. In order to extract useful pathological features from these retinal components, their accurate and reliable segmentation is usually required before pathological classification.

A project by Hayashi et al. [118] in 2001 developed a CAD system to assist physicians with the diagnostic analysis of blood vessels and abnormalities in fundus images. The blood vessel properties investigated were vascular intersections and widths which may indicate the presence of disorders such as arteriosclerosis and high blood pressure. Vascular intersections were detected using a filter that highlights local regions with vessels crossing the filter boundary at least four times. Misdetections were minimized by subtracting regions that match typical false positive vascular intersections. Variations in vascular width were determined through a process that involved morphological thinning, center-line fitting and perpendicular-line estimation. Blood vessel segments with a width that deviated from the average vessel width by more that a predetermined threshold were marked as abnormal sections. Red lesions were detected using a combination of morphological processing, HSV colour space contrast enhancement and binarization. Further validation was performed on candidate red lesion regions by ensuring that their contrast against neighbouring regions was greater that a given threshold. The system was tested on an unspecified collection of approximately 450 fundus images and its effectiveness

was justified using true positive rates and error ratios of component segmentation; actual diagnosis was left to the physician. True positive rates of 76.5%, 70% and 58.8% were obtained for abnormal segment widths, white and red lesion detection respectively while intersection detection showed an error rate of 25.1%.

Foracchia et al. [119] presented a computerized system for the quantitative characterization of vascular diagnostic properties of hypertensive retinopathy in fundus images. These properties include vessel tortuosity, narrowing and change in direction particularly at branch and intersection points in the vascular network. A tracking algorithm was first used to segment the blood vessels then detect bifurcations and crossings for the recognition of Gunn and Salus signs which are indicative of hypertensive retinopathy [120]. Although the actual diagnosis of hypertensive retinopathy was not implemented in this work, it boasts of an agreement of its algorithm output with the qualitative ordering of vessel tortuosity in a set of 22 hypertensive retinopathy images by three clinical experts.

Vallabha et al. [121] investigated the effectiveness of Gabor filter banks for detecting and classifying DR vascular abnormalities. An non-public collection of over 100 images exhibiting either normal retina or various levels of NPDR were obtained from an eye center and used to test the approach. The green channel of the images was filtered using a collection multiple scale and orientation Gabor filters in the frequency domain. The hypothesis proposed that blood vessels in an abnormal retina have multiple peaks as opposed to one dominant orientation which is likely to be shown by normal retina in an energy-orientation plot. Although visual illustrations of the correlation between NPDR symptoms and energy vs orientation plots are shown, only 4 test images are depicted in the work and no objective evaluation is presented for the over 100 images collected.

One of the earliest works on retinal lesion detection was by Goldbaum et al. [87], who in 1990 investigated the discrimination of similarly coloured retinal objects as part of the STARE project. Their preliminary investigations showed that similarly coloured lesions such as exudates, cotton-wool spots and drusen can be computationally distinguished using colour features only. A retinal specialist (Dr Goldbaum) selected 30 retinal images from different patients such that the three lesion categories each had 10 representative sample images that observed random distribution over variables such as age and retinal pigmentation. Groups of lesion pixels were manually extracted from all the sample images and used to create a three-dimensional average colour (RGB) vector for each patient. To reduce luminance noise the vectors were further reduced to two-dimensional features through a spherical-coordinate transformation. A distance based Bayesian classifier was trained on this dataset of 30 two-dimensional colour vectors (10 instances for each class: exudates, cotton-wool spots and drusen). The prior probabilities of each class

were equivalent, owing to the training set distribution. The posterior probabilites were estimated using a two-dimensional Gaussian Probability Density Function (PDF) estimated from the training set. Using the Mahalanobis distance to threshold the Bayesian class probabilities, a leave-one-out cross validation approach yielded 70%, 70% and 50% accuracy for exudates, cotton-wool spots and drusen respectively.

In 2000, Wang et al. [122] extended the work of Goldbaum et al. [87] with the intention of specifically improving exudate detection in varying illumination conditions. The RGB colour space was transformed into an isoilluminant surface model with three spherical features that allow for more effective differentiation in luminance and colour differences. The Bayesian posterior probability was used to formulate a discriminating function known as the Minimum Distance Discriminant (MDD). The MDD is trained on a set of sample exudate and normal background retinal image patches. Before applying the MDD on the spherical coordinates of individual pixels, a brightness adjustment procedure was applied in order to enhance dim lesions existing in low contrast regions. While this preprocessing function improves the contrast of obscure lesions, it also yields background formations that may be wrongly classified as lesions. A further verification process that uses a pixel's local neighbourhood intensity distribution is employed to re-classify these potential classification errors. Experiments conducted on a set of 54 retinal images that contain abnormalities and exhibit varying levels of image quality demonstrate a 100% accuracy in terms of correctly labelling each image as either normal or abnormal. Further experiments on a set of 100 normal retinal images demonstrate an accuracy of 70% in terms of identifying truly normal retinal images. Although this study showed a combined classification accuracy of 85%, its major drawbacks include the unavailability of the dataset (both training and test sets) used for more objective analysis of performance and its classification at image level as opposed to pixel level which is more challenging given that judgements need to be made using limited local information.

Inspired by the work of Gardner et al. [123], Hunter et al. [124] applied neural networks to the task of exudate detection. In both studies red-free[6] retinal images were tiled using either 20×20 or 16×16 blocks; some of these blocks exhibited exudates while others had drusen amongst other retinal components. Aided by a hierarchical feature selection process called sensitivity analysis, Hunter et al. went further to use other features such as Fourier transform and Prewitt edge detection filters values to build a neural network with 11 inputs and 7 hidden units which was significantly smaller than the network of 400 input and 80 hidden units used by Gardner et al. From a data set of 16 fundus images, a training set (56 exudate and 44 drusen tiles) was used to build a

[6]Subjected to light that is free of the red channel instead of white light for contrast enhancement purposes.

classifier which when evaluated on a testing set (60 exudate and 51 drusen tiles), yielded sensitivity and specificity values of over 91%.

Osareh et al. [125] used features derived from various colour space statistics and shape descriptors on Fuzzy C-Means (FCM) segmented exudate candidate regions. The FCM segmentation is based on zero-crossings of the second derivative of the histogram of each colour band at multiple scales. The zero-crossings were used to define possible colour clusters in the images and each pixel assigned a membership value to reflect its degree of belonging to a particular class. The selected features were implemented on neural network and SVM classifiers using an unspecified data set of 142 retinal images; neural networks proved superior with an exudate classification accuracy of 93.4%.

Microaneurysms manifest as small blobs in the retinal capillaries [57] and are known to be the earliest visually recognizable symptom of DR [117]. Because microaneurysms occur within and have virtually the same colour as blood vessels, their automated detection and subsequent analysis are potentially more challenging than that of exudates. Bernardes et al. [126] analysed microaneurysms through a method called MA-tracker which used image registration and chronological change detection techniques. Images were preprocessed using principal component analysis and contrast enhancement in the RGB and HSV colour spaces to yield enhanced fundus images which could then be manually earmarked for microaneurysms by trained graders. Using a collection of 235 colour fundus photographs and 145 fluorescein angiograms collected at 6 monthly intervals over a period of 5 years from NPDR patients, it was shown that although microaneurysms remained stable over time, only 29.4% invariably stayed in the same location. Jelinek et al. [127] compared the effectiveness of the Waikato automated microaneurysm detector against that of optometrists at diagnosing DR based on ground truth findings as determined by a team of ophthalmologists. The automated detector enhanced interest points using shade correction, top hat vessel removal, matched filter thresholding and region growing techniques. A combination of colour (RGB and HSV) and morphological (area, complexity and aspect ratio) features are extracted from interest points and used to train standard classifiers such as SVMs, ADtrees and logistic regression [128]. While the optometrists achieved 97% sensitivity at 88% specificity, the automated detector recorded 85% sensitivity at 90% specificity at detecting DR on a set of 758 fundal images from diabetic patients. Although the automated detector had lower effectiveness than optometrists its performance met the prescribed National Health and Medical Research Council[7] guidelines.

[7] Australia

Larsen et al. [129] compared the performance of a patented DR detection algorithm from a commercial fundus analysis system, RetinaLyze (https://retinalyze.com/) against visual lesion annotation by six specialist ophthalmologists. A collection of 400 fundus photographs from diabetic patients in the Welsh Community Diabetic Retinopathy Study were used for evaluation; the fundus status of the images ranged from normal to early proliferative retinopathy. The manual lesion annotation by specialists was done using software that allowed regular shape and freehand drawing to mark red lesions after full colour stretching for each RGB channel to enhance visual contrast. RetinaLyze applies modeling techniques on the green channel of fundus images to identify candidate red lesions; blood vessels and the optic disk are extracted prior to enlarging the candidate lesions through region growing methods. The automated system yielded an accuracy, sensitivity and specificity of 79%, 96.7% and 71.4% respectively when using predetermined system threshold settings. When these settings were varied, an AUC of 90.3% was achieved. Additionally, Cohen's kappa was used to measure classification agreements with the gold standard overall visual DR grading. The automated system and ophthalmologists had agreements of 0.659 and 0.648 respectively; this showed that the performance of the automated system was comparable to human experts.

Acharya et al. [130] used morphological image processing techniques and a SVM classifier to automatically diagnose DR as either non-existent, mild NPDR, moderate NPDR, severe NPDR and PDR. A collection of 331 retinal photographs showing either normal retina or varying levels of DR were obtained from the National University Hospital in Singapore. Adaptive histogram equalization was performed on the inverted intensity levels to improve the contrast of the images before a series of morphological operator, thresholding, edge detection, image subtraction and median filtering techniques were applied in variation to enhance the different components of the retina. The features extracted were the pixel areas of blood vessels, microaneurysms, exudates and haemorrhages. An average accuracy, sensitivity and specificity of 85%, 82% and 86% respectively were obtained on an independent test set derived from the image collection. Although different data sets were used, these results are comparable with other similar studies in this era such as Sinthanayothin et al. [131] and Singalavanija et al. [132].

3.4 Recent Studies

While previous research has improved the effectiveness of retinal vascular segmentation, the same degree of success is yet to be realised in the detection of retinal manifestations such as red and white lesions. It is typically these manifestations that form the basis of DR diagnosis and subsequent grading of its severity. This section focuses on some of

the recent methods that have been investigated for use in actual automatic diagnosis of DR as well as the current state of the art in retinal image segmentation. To increase the likelihood of reproducibility and/or objective comparative analysis, deliberate effort has been made to focus on studies that use publicly available datasets. Some current commercial systems and prospective technology for automated DR detection are also explored.

3.4.1 Computer Aided Diagnosis

In their review of recent CAD methods for detecting DR using Digital Fundus Photographs (DFPs), Zhang et al. [4] established that the segmentation of candidate vessel abnormalities and subsequent extraction of features from them, are foundational steps towards DR diagnosis. After the localization of lesions and neovascularization, information from individual detections is fused to predict the presence or severity of DR. Supervised classification seems to be the predominant approach for facilitating automated diagnosis; individual studies normally make scientific contributions by either proposing new pathological image processing indicators or formulating hybrid or unique[8] classifiers. An example of the latter is the work of Akram et al. [133] in 2013 which used morphological operators and Gabor filters to highlight potential microaneurysms. A hybrid classifier which comprised of a GMM, SVM and SOM was supplied with shape, colour and statistical features from candidate microaneurysms. The proposed method is said to have achieved accuracy rates of up to 99.49% on the DIARETDB0 and DIARETDB1 datasets; the specific details of training and test set partitions are however not given.

Other researchers have focused on experimenting with new and existing feature operators for effective pathological discrimination. In 2014 Ashraf et al. [134] proposed the use of LBP texture descriptors for Regions Of Interest (ROIs) in a DFP. LBPs were selected due to their invariance to illumination fluctuations in an image and thus used as justification for not applying a preprocessing filter to erode potential pathological candidate regions. Different variations of LBPs were used in combination with SVMs to detect normal and abnormal (containing either haemorrhages and microaneurysms) ROIs. Accuracy rates of up to 86.15% were achieved on the DIARETDB1 dataset using 10-fold cross-validation. Research published by Seoud et al. [135] in April 2016 reported an effective new set of shape features called Dynamic Shape Features that are useful for DR detection and screening. A series of operations are used for contrast enhancement and illumination normalization in the DFP. Using the derived green channel, candidate microaneurysms and haemorrhages are extracted by thresholding connected components with local minimal intensity then discarding candidates with low contrast and/or high

[8] Previously unused or novel.

intensities. Morphological flooding is then iteratively applied to the processed image and at each flooding level, conventional shape features such as eccentricity and circularity are computed on resulting catchment basins. The evolution of these features over successive flooding is tracked for each candidate lesion resulting in a curve for each candidate lesion feature. Classification is done using a Random Forest (RF) classifier based on the shape feature evolution curves for each candidate lesion. On the Retinopathy Online Challenge dataset, a Free-response Receiver Operating Characteristic (FROC) score of 0.42 is achieved while on the Messidor dataset, experiments show an AUC of 0.899. Examples of other recent studies that have applied machine learning methods and computer vision features for DR detection can be found in the review by Amin et al. [136].

Pires et al. [137] used a Content Based Image Retrieval (CBIR) and machine learning approach for lesion classification. A Bag of Visual Words (BoVW) was constructed using the Speeded-Up Robust Feature (SURF) local descriptor as a low level feature set and a SVM classifier. When tested on the DR1, DR2 and Messidor datasets, the algorithm achieves AUCs of up to 97.8% and 93.5% for detecting images containing exudates and red lesions respectively. Although the algorithm is shown to be a flexible framework for characterizing DFPs using a predetermined feature set of codewords and no landmark specific preprocessing, it is unable to identify the precise location, size and quantity of lesions in a DFP.

3.4.2 Recent Retinal Image Segmentation Studies

Although efforts towards retinal image based diagnosis have strengthened, low level research into landmark segmentation continues to date. This is because progress in the latter generally improves the former. A recent study by [138] reviewed previous work with publicly available implementations of retinal image segmentation. After extensive experiments on datasets such as ARIADB, CHASEDB1, DRIVE, HRF and STARE, it was found that although potential still exists for both supervised and unsupervised approaches to retinal image segmentation, unsupervised methods are greatly dependent on preprocessing.

Some recent unsupervised initiatives include the work of Neto et al. [139] who used a methodology combining Gaussian smoothing, morphological operators, contrast enhancement and adaptive local thresholding. Experimental evaluations were done on the DRIVE and STARE datasets, achieving an average sensitivity and balanced accuracy of 0.7819 and 0.8702 respectively. Aguirre-Ramos et al. [140] used a Gabor filter bank

and Gaussian fractional derivative to enhance blood vessels. Thereafter automatic segmentation was conducted using a bi-modal Gaussian threshold, which was shown to be superior to the Otsu method. An average specificity, accuracy and balanced accuracy of 0.7854, 0.9503 and 0.8758 respectively were obtained on the DRIVE dataset.

Recent advances in machine learning such as deep learning, seem to have also found relevance in retinal image segmentation. Liskowski and Krawiec [141] used a deep neural network trained on 27×27 contrast normalized image patches, selected at random from retinal images. A total of 400000 and 380000 patches are drawn from the DRIVE and STARE datasets respectively and fed into the neural network. Testing is also done at patch level; in both cases, only patches that completely fit into the FOV are considered. A ROC and accuracy of up to 0.99 and 0.97 respectively are reported. Bala and Vijayachitra [142] used an unsupervised segmentation approach followed by a supervised diagnosis method. Segmentation was done through histogram equalization, matched filters and modified local entropy thresholding. An Extreme Machine Learning (EML) approach which comprised of a feedforward neural network trained using statistical texture features extracted from the automatically thresholded images, was used to diagnose Proliferative Diabetic Retinopathy (PDR). The EML classifier achieved diagnosis accuracies of 97.5% and 95% on DIARECTDB0 and DRIVE datasets respectively. The success of this PDR screening shows the impact that effective retinal image segmentation can have on automatic diagnosis.

A comprehensive review of recent techniques used in retinal image analysis can be found in [16]. Although a wide range of supervised and unsupervised approaches are reported, we note the absence of ensemble techniques such as Random Forests (RFs) [18] in this review.

3.4.2.1 Retinal Image Datasets

Some of the most commonly encountered datasets in previous literature are summarized in Table 3.1. The use of these datasets entails either blood vessel segmentation, landmark localization/characterization and pathological diagnosis. Each dataset is supplied with a ground truth which either shows the precise locations of landmarks or specifies the relevant diagnostic details. The images in the DRIVE dataset were obtained from a Netherlands based DR screening programme. The screened population consisted of 400 diabetic subjects between 25-90 years of age. The photographs in the dataset were randomly selected from a pool of subjects between 25 and 90 years old, 33 do not indicate diabetic retinopathy while 7 have signs of mild early diabetic retinopathy. This dataset comes with a mask for each image, to identify the field of view.

Dataset	Size	Dimensions	Purpose
DRIVE [100]	40	565×584	Vessel segmentation
STARE [96]	400 80 20	700×605	Vessel segmentation, optic disk location and diagnosis
CHASEDB1 [143, 144]	28	999×960	Vessel segmentation and tortuosity measurement
DIARETDB [145, 146]	130 89	1500×1152	Diagnosis
MESSIDOR [147]	1200	1440×960 2240×1488 2304×1536	Diagnosis

TABLE 3.1: Details of common retinal image datasets

The STARE dataset images and clinical data originate from the Shiley Eye Center at the University of California. The dataset comes in 3 forms for use in either vessel segmentation, optic disk localization and diagnosis. Each of these are only provided as a test set and there are no FOV masks. The vessel segmentation subset is the most commonly used and it has 10 out of 20 images with pathological signs.

The CHASEDB1 dataset was collected during a Child Heart Health Study in England (CHASE), during which a cardiovascular health survey was conducted in 200 primary schools in London, Birmingham, and Leicester. The images in the dataset are the left and right fundus images of 14 child subjects recruited from a school in North-East London.

The DIARETDB0 dataset contains 110 out of 130 retinal images which show some symptoms of DR. The DIARETDB1 dataset contains 84 out of 89 images which exhibit signs of mild proliferative DR. Gold standards informed by expert opinion, indicate locations of landmarks such as microaneurysms, hemorrhages and exudates. Both datasets were obtained from the Kuopio university hospital in Finland.

The MESSIDOR dataset which is sponsored by the Messidor programme partners is currently the largest publicly available retinal image repository. The images are taken from 3 ophthalmologic departments under varying conditions such as image resolution and the use of pupil dilation. Although this dataset does not contain a gold standard showing locations of pathological landmarks, it comes with a medical diagnosis for each image.

3.4.2.2 Evaluation Metrics in Retinal Image Segmentation

Generally the main objective of retinal image segmentation is to differentiate the retinal background from important landmarks such as blood vessels and pathological objects. This task has proved to challenging in cases where there is low contrast between the background and important retinal tissue. For blood vessel segmentation there is the added difficulty of differentiating vascular tissue from other non-vascular tissue which may have similar profiles.

In order to determine how successful blood vessel segmentation is, a gold standard (typically constructed by several opthalmological experts) is required and verification is normally done at pixel level. If the goal is to identify vessel tissue, four classification possibilities arise [12]. A True positive (TP) classification occurs when a pixel is identified as a vessel in both the gold standard (also known as ground truth) and the algorithmic image segmentation. A True negative (TN) classification occurs when the gold standard and the segmented image both classify a pixel as non-vascular. Misclassifications arise when the ground truth and segmented images have differing labels for a given pixel. A False negative (FN) occurs when a pixel is erroneously classified as non-vessel in the segmented image but as experts identify it as a vessel pixel in the ground truth image. A False Positive (FP) occurs when a pixel is labelled as non-vessel in the ground truth image but marked as a vessel in the segmented image.

These four basic classification categories form the foundation of composite evaluation metrics which occur at image level. The True Positive Rate (TPR) represents the proportion of correctly classified vessel pixels while the False Positive Rate (FPR) represents the rate at which pixels are mislabelled as being vascular. The segmentation accuracy (Acc) is simply the proportion of correct classifications (true positives and true negatives) over all the pixels of a retinal image. The ability to correctly detect vascular and non-vascular pixels is represented by Sensitivity (SN) and Specificity (SP) respectively. Given a segmented retinal image Img containing n pixels, its sensitivity and specificity can be defined as follows:

$$Sensitivity(Img) = \frac{\sum_{p=0}^{n} I_{TP}(Img(p))}{\sum_{p=0}^{n} I_{TP}(Img(p)) + \sum_{p=0}^{n} I_{FP}(Img(p))}, \tag{3.1}$$

$$Specificity(Img) = \frac{\sum_{p=0}^{n} I_{TN}(Img(p))}{\sum_{p=0}^{n} I_{TN}(Img(p)) + \sum_{p=0}^{n} I_{FN}(Img(p))}, \tag{3.2}$$

where I_{TP}, I_{FP}, I_{TN} and I_{FN} represent indicator functions for TP, FP, TN and FN respectively. The grey level at a pixel p is represented by $Img(p)$.

Neto et al. [139] used the concept of a balanced accuracy as shown in equation 3.3, to reduce the effect of vessel and non-vessel pixel count disproportions in typical fundus images.

$$BAcc(Img) = \alpha * Sensitivity(Img) + \beta * Specificity(Img). \tag{3.3}$$

The authors argue that the traditional accuracy metric is a subset of balanced accuracy. In the former, α and β are set to the probability of vascular and non-vascular tissue respectively, as shown in equations 3.4 and 3.5. In the latter, both weights are set to 0.5.

$$\alpha = \frac{\sum_{p=0}^{n} I_{TP}(Img(p)) + \sum_{p=0}^{n} I_{FP}(Img(p))}{n}. \tag{3.4}$$

$$\beta = \frac{\sum_{p=0}^{n} I_{TN}(Img(p)) + \sum_{p=0}^{n} I_{FN}(Img(p))}{n}. \tag{3.5}$$

A receiver operating characteristic (ROC) curve assists with comprehensively evaluating an algorithm's segmentation ability under varying settings. It plots a curve of TPR against FPR and measures the area under this curve (AUC) as a normalized value between 0 and 1. A curve that approaches the top left corner is indicative of good segmentation performance which closely approaches an AUC of 1. For the retinal images occuring in most datasets, these metrics are generally computed over only the pixels within the FOV.

3.5 Conclusion

Although there is an increasing interest in automated retinal image diagnosis, research into improved retinal landmark segmentation continues to be relevant, due to its capacity for objective computation of landmark location, size and shape. Highly accurate blood vessel segmentation can yield reliable measurements of features such as tortousity, width and branching factor which can assist in DR diagnosis [143].

The objective assessment of blood vessel segmentation algorithm performance is assisted by the availability of several public benchmark datasets. These repositories enable comparative research as well as re-enactment and verification of published experimental results. Although existing retinal image datasets showcase a variety of fundus images with varying levels of pathology, image quality and structure; there seems to be a lack of images from the African continent and third world countries in general. This is of concern given that the effect of ocular diseases is generally more pronounced in these settings due to lack of skilled medical experts and access to health facilities. Previous epidemiological and digital retinal imaging studies indicate that environmental and genetic factors may significantly contribute towards retinal vascular quality [148]. This

means that retinal image analysis systems trained on images from a specific geographical location, may not necessarily be relevant in a different location that has a different genetic pool and environmental context.

Although concepts suh as Gabor filters and automatic thresholding seem to have been widely explored in combination with various other techniques, their use is generally in the form of direct application without any enhancements. There may be room for innovative adaptations in some of these concepts, which may ultimately benefit the general image processing community. Experimentation with Gabor filters [101, 140] is normally in the form of parameter tuning with little interogation done on the effectiveness of its default features. Thresholding is typically enforced through entropy [142] maximization or the legacy Otsu method [140] without any form of adaptation.

Although there is evidence of the application of recent advances in machine learning such as deep learning [141], to retinal image analysis, these concepts are yet to be comprehensively explored. We have only encountered a few studies which utilize ensemble classification [144, 149, 150], a supervised machine learning technique that has been proved to be highly robust to noisy training data. It may be, that the challenge in retinal image analysis lies more at the feature extraction level, hence the wide diversity in algorithms that are applied at that level. In the next chapter we explore the use of RFs for selecting the most discriminative feature in a dataset.

Chapter 4

Random Forest Feature Selection

4.1 Introduction

Although feature selection [151] is commonly applied to avoid the curse of dimensionality in contexts such as text and DNA microarray analysis, it is generally useful as a machine learning preprocessing step. It ensures that all variables that are included in a predictive model are not redundant and therefore, meaningfully contribute towards its discriminative ability. The benefits of feature/variable selection include improved prediction accuracy and efficiency, reduced overfitting, as well as easier understanding, interpretation and control over model training [152].

Algorithms used for establishing a set of useful features can be categorized as either filter, wrapper or embedded methods [153]. Filter methods ensure feature usefullness using statistical metrics such as chi square, fisher score and variance; which evaluate intrinsic properties of features without considering their relationship to the target variable. Wrapper methods use cross-validation techniques to select features that optimize classifier performance; examples include genetic algorithms, sequential selection and recursive elimination. Embedded algorithms use the model learning process to gather useful statistics on the utility of each feature; example classifiers that can be used to gather such statistics include decision trees and SVMs [154].

The Random Forest (RF) classifier has the capacity to facilitate both wrapper and embedded feature selection through the Mean Decrease Accuracy (MDA) and Mean Decrease Impurity (MDI) methods respectively [155]. MDI is known to be biased towards predictor variables with multiple values whilst the MDA is stable in this regard [19]. MDA is accordingly the generally preferred option for RF-based feature selection despite its higher computational overhead when compared to MDI. This research ultimately

intends to apply RF feature selection to retinal images at pixel level, hence computational overhead is of significant importance. We propose two improvements to the MDI method to overcome its shortcomings. The first is using our proposed Purity Gap Gain (PGG) measure which has an emphasis on computational efficiency, as an alternative to the Gini Importance (GI) metric. The second is incorporating a Relative Mean Decrease Impurity (RMDI) score, which aims to offset the bias towards multi-valued predictor variables through random feature value permutations. Experiments are conducted on UCI datasets to establish the effectiveness of PGG and RMDI.

In the remainder of this chapter, we begin with a detailed overview of RFs and then proceed to introduce our proposed improvements to MDI feature selection. Experimental details are then presented and results discussed.

4.2 Random Forests

Ensemble classification methods are premised on a training phase which generates several individual base classifiers and a testing phase which aggregates classifier predictions into a consolidated result [156]. The most common examples include boosting [157], bagging [158] and Random Forests (RFs) [18], all of which are based on decision trees. In boosting, trees are trained in succession and emphasis is placed on instances that are misclassified by previous classifiers; the ensemble is therefore a committee of evolving trees that increasingly adapt towards learning how to correctly predict the most difficult training instances. In bagging, trees can be trained concurrently and emphasis is placed on creating bootstrapped samples; the ensemble is therefore a committee of trees which have random similarities and differences. RFs complement bagging by introducing an extra level of stochastic behaviour in the selection of features involved in node split point optimization. As a result, RFs are known to 1) have reliable classification performance that is comparable to Support Vector Machines (SVMs), 2) be tolerant of noise and robust to over fitting, and 3) be fast and simple to use since they only rely on 2 main parameters [156].

We proceed to outline the adopted notation for representing the various components of a RF and introduce some of its internal measures. As a supervised machine learning algorithm, RFs receive a training set X, of N instances, each of which can be represented by $X(i, \_) = (\mathbf{x}, y)$, where $1 \leq i \leq N$ and $\mathbf{x} = (x_1, x_2, \ldots, x_A)$, is a vector of A components. Each instance $\mathbf{x}$ has a corresponding label y such that $1 \leq y \leq C$, from the set of C possible class values. X can be presented as a table in which the rows and columns correspond with the dataset's instances and features respectively. Accordingly,

the notations $X(i,\_)$ and $X(\_,j)$ represent views of the i^{th} instance and j^{th} feature respectively. A set of samples $\{X_k, 1 \leq k \leq T\}$, each of size N, is drawn with replacement from X[1] and used to create a RF classifier $h(x)$ made up of T corresponding decision tree classifiers $h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_T(\mathbf{x})$. The RF algorithm can therefore be said to accept a training set X as input and output an ensemble classifier $h(x)$. During the classification stage, $h_k(x) = 0$ denotes a correct prediction whereas $h_k(x) = y$ specifies that x has been incorrectly labelled as a member of class y.

The sample X_k is used to create a root node Θ_{k0}[2] of the k^{th} tree (h_k), which is built by recursively splitting nodes until given conditions are satisfied[3]. We denote any node at a valid depth $d \geq 0$ in h_k and its size by Θ_{kd} and $|\Theta_{kd}|$ respectively. It can therefore be said that $\Theta_{kd} \in h_k$. Let $\Phi(\Theta_{kd}, j)$ represent the set of all possible ways of partitioning node Θ_{kd} into B branches based on attribute j and $\phi \in \Phi(\Theta_{kd}, j)$ be one such partition. ϕ^* is the adopted partition as it is the most optimal in terms of decreasing the impurity of child nodes. When a partition is enforced, $\phi(z), 1 \leq z \leq B$ is the resulting child node of Θ_{kd} at depth $d+1$. For the sake of simplicity we represent $\phi(z)$ as Θ_{kd+1}. The probability of landing on branch z from Θ_{kd} when splitting based on ϕ is represented by $p(z)$. The probability of class y in node Θ_{kd} and the probability of class y in branch z of Θ_{kd} are represented by $p(y)$ and $p(y|z)$ respectively.

4.2.1 Gini Impurity

Individual decision trees within a RF, use conditional rules to recursively partition a sample Θ_{kd} in an attempt to produce pure child nodes with instances belonging to the same class. Although the Gini Index (GI) [159] was the original measure chosen by Brieman [18] for measuring the level of decrease in impurity along a RF decision tree path, it is still widely used due to its general effectiveness [160, 161]. The Gini index of a node Θ_{kd} is defined as follows:

$$GI\left(\Theta_{kd}\right) = 1 - \sum_{y=1}^{C} p\left(y\right)^2. \tag{4.1}$$

Given that a partition $\phi \in \Phi(\Theta_{kd}, j)$ splits Θ_{kd} into B child nodes ($\phi(z)$), the resulting decrease in impurity is calculated as follows:

$$\delta GI\left(\phi\right) = -\sum_{y=1}^{C} p\left(y\right)^2 + \sum_{z=1}^{B} p\left(z\right) \sum_{y=1}^{C} p\left(y|z\right)^2. \tag{4.2}$$

[1] Note that although X_k is drawn from X, the latter is a set while the former is a multiset.
[2] The terms sample and node are used interchangeably in this text since they are closely related.
[3] See section 4.4 for the list of conditions that lead to the end of node splitting.

Here, the GI is used to measure the impurity of Θ_{kd} and the weighted impurity of its child nodes ($\phi(z)$). $\delta GI(\phi)$ is therefore a measure of the change in impurity caused by the partition ϕ; it is commonly referred to as the Gini impurity [153] metric for simplicity[4]. The optimal partition ϕ^* for splitting a node Θ_{kd} on attribute j^* is deduced as follows:

$$\phi^* = \underset{\phi \in \Phi(\Theta_{kd}, j)}{\operatorname{argmax}} \ \delta GI(\phi). \tag{4.3}$$

Note that attribute j^* is the feature that yields the highest value of $\delta GI(\phi)$ after considering all possible split values from all features j. The process of training a RF classifier is essentially an iterative algorithm (commonly referred to as Forest-RI) in which a node Θ_{kd} is split into its child nodes. Each child node Θ_{kd+1} is likewise split and this process continues until specified conditions such a maximum node purity level or depth, are met. During each iteration, a predefined number of features is chosen randomly, without replacement. This extra level of stochastic behaviour distinguishes the RF algorithm from bagging and facilitates greater variation among the trees of a RF.

4.2.2 Random Forest Strength

In the typical RF algorithm, N instances are drawn from the training set with repetition. On average, a third of the training set is not used by each tree, due to this bootstrapped sampling [159]. This excluded collection of instances is known as the Out Of Bag (OOB) set, which is useful for determining feature importance as well as RF internal estimates for error, strength and correlation; in the context of the training set [18].

Let $P_o(\mathbf{x}, y)$ be the proportion of votes allocated to class y when an OOB instance $(\mathbf{x}, y), \mathbf{x} \notin \Theta_{k0}$ is classified by the RF trees. This proportion is defined by:

$$P_o(\mathbf{x}, y) = \frac{\sum_{k=1}^{T} I(h_k(\mathbf{x}) = y; (\mathbf{x}, y) \in O_k)}{\sum_{k=1}^{T} I((\mathbf{x}, y) \in O_k)}, \tag{4.4}$$

where O_k is the OOB set for classifier $h_k(\mathbf{x})$. It follows that $X \setminus X_k = O_k$. $I(c)$ is an indicator function which returns 1 if the condition c is true and 0 otherwise. The margin on an OOB instance reflects the extent to which votes for the correct class exceed votes for any other class and is represented by:

$$mr(\mathbf{x}, y) = P_o(\mathbf{x}, 0) - \max_{y=1}^{C} P_o(\mathbf{x}, y). \tag{4.5}$$

[4] Although it should ideally be referred to a the change in Gini impurity, we also follow this convention for the sake of consistency. Going forward, GI and δGI are used to denote Gini Index and Gini impurity, respectively.

The strength of a RF is an internal measure of the average expected margin over the whole training set [159]:

$$s(X) = \frac{1}{N}\sum_{i=1}^{N} mr\left(X\left(i,\_\right)\right). \tag{4.6}$$

4.2.3 Random Forest Variable Selection Methods

Although RFs are predominantly used as classifiers, one of their important byproducts is variable importance measures. There are two such common measures: MDI and MDA.

4.2.3.1 Mean Decrease Impurity

MDI is also referred to as Gini importance and it measures the average Gini impurity δGI for a given training set X and feature j, as shown below:

$$MDI(X,j) = \frac{1}{T}\sum_{k=1}^{T} \sum_{\Theta_{kd}\in h_k:\Theta_{k0}=X_k} \sum_{\phi\in\Phi(\Theta_{kd},j)} \delta GI\left(\phi\right) I(\phi = \phi^* \wedge j = j^*), \tag{4.7}$$

where $I(\phi = \phi^* \wedge j = j^*)$ is an indicator function call which returns 1 if j yields the optimal split ϕ^* and 0 otherwise. Equation 4.7 is essentially computing the average δGI arising from all the cases during the induction of $h(x)$ when feature j yields an optimal split. An attribute j will have a high MDI value if it frequently yields high δGI values during RF induction. MDI is a typical example of an embedded feature selection method since it utilizes information generated in the RF learning process; this makes it a computationally attractive option. Its overhead in terms of space is of the order A since it needs to keep track of total Gini impurity values for each attribute. This method is however not often used because of its bias towards variables with multiple values [19], this means that any real valued variable is likely to generate lower impurity scores than a binary valued one. This drawback makes the feature rankings that are generated by this method less objective [153, 162], since typical dataset features do not necessarily have the same value distribution.

4.2.3.2 Mean Decrease Accuracy

MDA evaluates variable importance based on the relative OOB error rate when an entire feature is replaced with its identically distributed perturbed version. In order to formally describe this process, we introduce some extra notation. Let $\bar{X}$ represent a modified version of a training set X such that $\bar{X}(\_, j)$ is a randomly generated permutation of

$X(\_, j)$, where $1 \leq j \leq A$. Let $\tilde{X}^{(v)}$ represent another modified version of X such that $\tilde{X}^{(v)}(\_, v) = \bar{X}(\_, v)$ and $\tilde{X}^{(v)}(\_, j) = X(\_, j)$ for $j \neq v$, where $1 \leq j, v \leq A$. We define a mapping function $\wp(x, v) : X \mapsto \tilde{X}^{(v)}$ that generates elements $\wp(x, v) \in \tilde{X}^{(v)}$ which are adaptations of $x \in X$ such that the v^{th} attribute is perturbed. We use the term perturb to refer to the process of replacing every instance's value for attribute v with the corresponding permuted value in $\bar{X}(\_, v)$.

Given a classifier h_k, we define its OOB error as the rate of incorrect classifications over O_k as follows:

$$OOB\_Error\,(k) = \frac{\sum_{x \in X} I\left(h_k\left(\mathbf{x}\right) \neq y; (\mathbf{x}, y) \in O_k\right)}{|O_k|}. \tag{4.8}$$

A similar error rate can be formulated for $\tilde{X}^{(v)}$ as follows:

$$OOB\_Error^{(v)}\,(k) = \frac{\sum_{x \in X} I\left(h_k\left(\wp(\mathbf{x}, v)\right) \neq y; (\mathbf{x}, y) \in O_k\right)}{|O_k|}. \tag{4.9}$$

The MDA of an attribute j with respect to a classifier h_k aggregates the abovementioned error rates as follows:

$$MDA(j) = \sum_{k=1}^{T} \frac{OOB\_Error^{(j)}\,(k) - OOB\_Error\,(k)}{T}. \tag{4.10}$$

Any attribute j will demonstrate the highest MDA if it has the highest dissimilarity between $X(\_, j)$ and $\tilde{X}^{(j)}(\_, j)$. Conversely, if a random permutation of j yields little difference in MDA when compared against its original version, j is assumed to be of little importance since its specific values do not seem to make any difference with respect to classification accuracy. MDA is generally preferred over MDI because of its better tolerance of the different levels of feature value granularity [19, 153, 162]. It is however computationally expensive since it requires the training of T trees followed by the testing of each of these trees, A times; since each attribute has to be perturbed.

4.3 Proposed MDI Improvements

We propose two avenues for improving MDI. The first is the replacement of the Gini impurity with a similar but computationally more effecient alternative, Purity Gap Gain (PGG). The second is the incorporation of the concept of random feature value permutation observed from MDA. We ultimately seek to reduce the susceptibility of MDI to feature value granularity, so as to jointly optimize the effectiveness and efficiency RF feature selection.

4.3.1 Purity Gap Gain

The first difference between PGG and Gini impurity is the use of a product of node proportions instead of the Gini Index (GI). One of these proportions is known as node purity [163] and is defined as follows:

$$\hat{p}\left(\Theta_{kd}\right) = \max_{y=1}^{C} p(y). \tag{4.11}$$

The other proportion is simply the minimum class proportion within a node. The concept of purity gap effectively rewards a smaller difference between the maximum and minimum proportions within a node as shown below:

$$PG\left(\Theta_{kd}\right) = 1 - \max_{y=1}^{C} p(y) \min_{y=1}^{C} p(y). \tag{4.12}$$

PG is maximized when the minimum and maximum proportion in a node are the same. In the case of two classes, this occurs when each class has a proportion of $\frac{1}{2}$. Thus PG models GI in that it is maximized when class proportions are equal in a node and it is minimized when there is a clear bias towards one of the class proportions. A notable difference between Equations 4.1 and 4.12 is the computational simplicity of the latter over the former. PG simply deduces the maximum and minimum class proportions, while the Gini index requires the calculation of the sum of all class proportions squared within a node. PGG is the weighted average change in PG over a branch from Θ_{kd} to Θ_{kd+1} as defined below:

$$PGG\left(\phi\right) = \sum_{z=1}^{B} p\left(z\right) PG\left(\phi(z)\right) - PG\left(\Theta_{kd}\right). \tag{4.13}$$

Although we have previously explored the concept of PGG for measuring RF tree confidence [164], its implementation in this study is simpler[5] and geared more towards mimmicking GI in an efficient manner.

4.3.2 Relative Mean Decrease Impurity

Inspired by MDA, we adopt the concept of Relative Mean Decrease Impurity (RMDI) proposed by Nguyen et al. [165], which captures the difference between the MDI values derived from $X(\_, j)$ and $\tilde{X}^{(j)}(\_, j)$ as shown below:

$$RMDI(j) = MDI(X, j) - MDI(\tilde{X}^{(j)}, j). \tag{4.14}$$

[5]Information about node depth is not considered in the current implementation.

In contrast to MDA, feature permutations are introduced at the start of the learning phase of a RF. Once a permutation of each feature is generated, these are injected into the dataset such that the number of features is now $2A$. If the number of features randomly selected during each iteration of Forest-RI is predetermined as A, the modified learning algorithm would generally take twice as long to train. It is however important to note that if common convention [18, 159] is followed, the computational overhead of the modified algorithm is reduced to $\sqrt{2A} - \sqrt{A}$, since $\sqrt{2A}$ features are now selected instead of $\sqrt{A}$.

Although the split values of permuted attributes are actually computed, nodes are never split based on such attributes, in order to retain the integrity of RF training. ϕ^* is therefore restricted to original features only; permuted features are mainly used for offsetting the total split scores of original features. Instead of simply observing the MDI of an attribute as a measure of attribute importance, we now consider how different this MDI is from the MDI of a permuted version of the attribute. We envisage that the incorporation of this relative comparison will offset the MDI's bias towards multi-valued attributes. Although RMDI may result in a slight increase in computational requirements when compared to MDI, the resulting time complexity will still be superior to that of MDA.

While the emphasis of the study by Nguyen et al. [165] was on reducing RF bias in the context of high dimensional data, the concept of RMDI is incorporated into this study mainly for its reported effectiveness in reducing bias towards multivalued features. Our ultimate goal is a suitable means of selecting the most discriminative feature from a set of proposed Gabor features for use in automatic thresholding. It is important that such a selection should be robust to the varying granularity of these features, to avoid a skewed outcome.

4.4 Experimental Protocol

The experiment used in this study is designed to achieve 2 objectives: 1) to assess the reliability of RMDI as a feature selection strategy and 2) to provide a preliminary comparison between Gini impurity and PGG and split metrics during RF induction. All experiments were conducted using our own C++ implementation of dichotomous ($B = 2$) RFs that use the following standard settings [18, 159]:

- Each RF is made out of $T = 100$ trees.
- The number of randomly selected attributes at each node of a tree is set to $\sqrt{A}$.

- The size of each bootstrapped sample Φ_{k0} is N, the size of X.
- Node splitting is stopped when either $|\Theta_{kd}| \leq 5$, $\hat{p}(\Phi_{kd}) = 1$, or a depth $d = 30$ is reached.

Experiments were conducted on a Windows PC[6] and took a combined running time of approximately 1 hour 30 minutes over all data sets.

4.4.1 Experiment Description

To achieve the first objective, we perform repeated random sub-sampling validation 30 times. In each case, the dataset is split using an 80%:20% split ratio into training and validation data respectively. Feature ranking is based on RMDI scores achieved on the training data. We partition the feature set into strong and weak features using the top and bottom half of features in the RMDI rank respectively. The corresponding validation data is then used as a training sample for strong and weak features separately. We then calculate the strengths of the strong and weak feature based RFs, as defined in equation (4.6). We use the paired T-test to determine whether the mean difference between the two RF strengths is zero. We formally define our hypotheses as follows:

- The null hypothesis (H_0) assumes that the true mean difference (μ_d) between the strengths of the strong and weak feature based RFs is zero.
- The alternative hypothesis (H_1) assumes that μ_d is non zero.

If RMDI is a reliable means of feature selection, the experimental results should demonstrate sufficient evidence for us to reject (H_0) in favor of (H_1). Since the two forests utilize features of different combined usefullness, we do not expect their average strengths to be equal.

To achieve the second objective, we compare the strengths and running times of the RFs based on strong features for both Gini impurity and PGG. The same dataset splits are used in each case; the main difference is the split criteria used for RF induction. We are primarily interested in comparing the effectivess and efficiency of these two metrics.

4.4.2 Data Sets

The effectiveness of the methods proposed in this study is tested using 4 data sets from the UCI repository [166]. These data sets are drawn from Robnik [159] and Breiman's

[6] 2 cores running at 2.70GHz and 2GB of RAM.

[18] studies on RFs. The characteristics of the chosen data sets are summarized in Table 4.1, which reveals the diversity of the problems represented, in terms of data set size (N^*)[7], number of features (A) and number of classes (C).

TABLE 4.1: UCI Datasets

Dataset	N^*	A	C	**Dataset**	N^*	A	C
ecoli	336	7	8	segmentation	2310	19	7
ionosphere	351	34	2	sonar	208	60	2

4.5 Results

After features are sorted using RMDI scores derived from RF induction, features in the top half of this sorted list are presumed to be weaker[8] than those in the bottom half. Similarly, subsequent RFs built from the top half of this sorted list should ideally be weaker than those built from the bottom half. Tables 4.2 - 4.5 present the strengths of 30 RF pairs, using top and bottom half features. RFs are induced twice, using δGI and PGG as alternative split metrics. The headings δGI_T and PGG_T identify RFs built using the top half features when sorting is based on δGI and PGG based RMDI respectively. Likewise δGI_B and PGG_B are associated with the bottom half features. RFs that were built using all features[9], consistantly showed lower strengths than those built from either top or bottom half features. This emphasizes the importance of feature selection for improved classification.

We observe that the relative strengths of the majority of RF pairs were in line with our expectations. The ecoli and segmentation datasets each had only one anomaly in a PGG based RF. Out of 60 RF inductions, the isonosphere and sonar datasets had 10 and 6 odd cases respectively. It can be noted that the datasets with one anomaly have an odd number of features (A). A likely reason for this outcome is the fact that the top half of sorted features was smaller than the bottom half. Since the bottom half had stronger and more features than the top half, it is expected that the strengths of RFs corresponding to the former would be higher than those of the latter.

We use the mean (μ) strength over 30 RF inductions to rate the relative effectiveness of the 4 RF categories. This comparison is objective since for each run, the same Θ_{k0} is used to induce 4 different RFs. These mean values also confirm that the top half features were weaker than the bottom half for all datasets. A comparison of δGI_B and PGG_B mean RF strengths shows that the δGI based RFs were stronger than their PGG

[7] This value is larger than N, the size of a training set, X which is drawn from a data set.
[8] Sorting is done in ascending order, hence the list is arranged from weakest to strongest.
[9] These results have not been reported in Tables 4.2 - 4.5 to avoid redundancy.

counterparts in all cases except the sonar dataset. These findings are also confirmed by the paired T-test which yields p-values less that 1% for each of the comparisons: δGI_T vs δGI_B, PGG_T vs PGG_B and δGI_B vs PGG_B. This indicates that there is less than 1% likelihood that the observed trends are a mere coincidence, hence we reject the null hypothesis (H_0) in favor of the alternative hypothesis (H_1).

TABLE 4.2: Ecoli Dataset Random Forest Strengths

Run	δGI_**T**	δGI_**B**	**PGG_T**	**PGG_B**
1	-0.027146	0.389795	0.047715	0.070317
2	0.152976	0.436591	0.043644	0.076737
3	0.152096	0.387711	0.050019	0.104831
4	0.145038	0.365383	0.087261	0.149182
5	0.061024	0.416723	0.094282	0.161514
6	0.181508	0.444353	0.160363	0.07783
7	-0.023615	0.408102	0.071638	0.117691
8	-0.00364	0.391702	0.083588	0.142523
9	0.02483	0.415895	0.092955	0.137078
10	0.170839	0.441432	0.057166	0.107488
11	0.019756	0.415132	0.079084	0.106102
12	-0.007194	0.431746	0.08335	0.124272
13	0.150296	0.442049	0.080319	0.114403
14	0.076479	0.445552	0.049069	0.132312
15	0.058208	0.441629	0.089712	0.102199
16	0.157126	0.459481	0.084636	0.13876
17	0.003931	0.421854	0.081421	0.092023
18	-0.01861	0.420028	0.04713	0.083462
19	0.042306	0.430415	0.091708	0.162491
20	0.182107	0.45673	0.050223	0.104296
21	0.049584	0.420147	0.078497	0.142797
22	0.029947	0.423455	0.032072	0.117826
23	0.021469	0.414482	0.065759	0.105622
24	0.004867	0.387004	0.091882	0.130458
25	0.013595	0.406981	0.04701	0.134774
26	-0.004538	0.371943	0.047229	0.093141
27	0.085616	0.428243	0.082603	0.136401
28	0.098499	0.449171	0.024366	0.094602
29	0.066641	0.415959	0.062593	0.100947
30	-0.035408	0.400474	0.068714	0.125563
μ	0.0609529	0.419338733	0.070866933	0.116254733

In proposing PGG as a split metric, we sought to provide an effective and efficient alternative to δGI. Figure 4.1 demonstrates that over all datasets, experiments using PGG had lower computational overhead than those using δGI. The difference in running times seems to get more pronounced as the dataset size increases. We observe that the mean strengths for δGI_B and PGG_B were close in value (0.5795 vs 0.5453), with fewer anomalies for the latter than the former over the isonosphere dataset. This could be an indication that the PGG split metric is promising in such contexts.

TABLE 4.3: Isonosphere Dataset Random Forest Strengths

Run	δGI**_T**	δGI**_B**	**PGG_T**	**PGG_B**
1	0.556447	0.582385	0.514891	0.511525
2	0.550569	0.590034	0.481161	0.555759
3	0.560526	0.570626	0.484594	0.543271
4	0.516025	0.556717	0.525872	0.524723
5	0.520416	0.599395	0.490189	0.549813
6	0.568939	0.560921	0.53242	0.533997
7	0.561062	0.586017	0.510265	0.560203
8	0.519336	0.582497	0.482497	0.556673
9	0.530046	0.61077	0.50587	0.551568
10	0.508288	0.56497	0.541541	0.540246
11	0.589797	0.568449	0.524403	0.500343
12	0.585577	0.574657	0.540212	0.524845
13	0.545405	0.570824	0.524202	0.558661
14	0.541053	0.582383	0.485662	0.544718
15	0.545074	0.612689	0.50907	0.543123
16	0.513623	0.566544	0.488532	0.565566
17	0.541315	0.576777	0.493366	0.539206
18	0.556248	0.596467	0.476581	0.568932
19	0.546211	0.590314	0.484315	0.564066
20	0.551437	0.542613	0.473955	0.523157
21	0.541913	0.569445	0.515962	0.527477
22	0.582216	0.557931	0.500887	0.528861
23	0.52593	0.560144	0.55522	0.557496
24	0.567995	0.587164	0.47623	0.549819
25	0.539891	0.573105	0.507317	0.533809
26	0.528445	0.589206	0.505377	0.535757
27	0.553018	0.586033	0.504947	0.572588
28	0.538827	0.584156	0.497916	0.552165
29	0.597805	0.589895	0.501625	0.575789
30	0.545177	0.602018	0.525964	0.566263
μ	0.547620367	0.579504867	0.5053681	0.5453473

TABLE 4.4: Segmentation Dataset Random Forest Strengths

Run	δGI**_T**	δGI**_B**	**PGG_T**	**PGG_B**
1	0.280749	0.451078	0.026682	0.216214
2	0.382937	0.443338	0.065954	0.210795
3	0.150113	0.471911	0.02725	0.190696
4	0.307423	0.430391	0.084253	0.147266
5	0.275586	0.436293	0.070093	0.148077
6	0.184008	0.46554	0.06357	0.174442
7	0.292195	0.442552	0.06355	0.155224
8	0.173639	0.446667	0.057788	0.16547
9	0.353221	0.455366	0.074285	0.154391
10	0.269949	0.451071	-0.000421	0.217403
11	0.385196	0.449063	0.035755	0.202914
12	0.387249	0.423394	0.085668	0.176955
13	0.373037	0.457085	0.129758	0.125576
14	0.261068	0.445298	0.044706	0.182459
15	0.286878	0.473829	0.063416	0.148759
16	0.365537	0.443747	0.06488	0.170588
17	0.033245	0.471337	0.076697	0.16636
18	0.244264	0.461377	0.063161	0.149092
19	0.315729	0.426779	0.072084	0.154189
20	0.205651	0.472926	0.079125	0.167418
21	0.140948	0.434108	0.125577	0.130802
22	0.215783	0.41543	0.08794	0.169451
23	0.149275	0.461224	0.047319	0.149366
24	0.276067	0.434075	0.099273	0.10654
25	0.301123	0.49464	0.067388	0.167487
26	0.242784	0.496258	0.042917	0.181709
27	0.021848	0.458347	0.041644	0.169971
28	0.196497	0.448907	0.082877	0.128287
29	0.289368	0.424298	0.083893	0.142953
30	0.317692	0.443497	0.053572	0.17874
μ	0.255968633	0.4509942	0.0660218	0.164986467

TABLE 4.5: Sonar Dataset Random Forest Strengths

Run	δGI**_T**	δGI**_B**	**PGG_T**	**PGG_B**
1	0.277886	0.267121	0.196934	0.345602
2	0.271027	0.279378	0.189782	0.30134
3	0.218538	0.287594	0.221478	0.268406
4	0.293588	0.264731	0.215746	0.320396
5	0.250705	0.307684	0.248996	0.302134
6	0.252636	0.259464	0.197973	0.296295
7	0.302454	0.30526	0.204969	0.320364
8	0.2361	0.28515	0.268867	0.317392
9	0.207769	0.337482	0.317218	0.310252
10	0.1418	0.289696	0.226896	0.316504
11	0.187644	0.263049	0.23491	0.283686
12	0.216668	0.325378	0.224855	0.364993
13	0.217332	0.285277	0.209871	0.317247
14	0.262298	0.329419	0.244875	0.313514
15	0.247052	0.264555	0.264563	0.334389
16	0.271854	0.287196	0.290693	0.290879
17	0.283797	0.262356	0.227599	0.289655
18	0.212502	0.272545	0.170465	0.326425
19	0.184329	0.274708	0.226859	0.325275
20	0.264698	0.271629	0.305203	0.26897
21	0.249948	0.274471	0.282349	0.299947
22	0.279594	0.322566	0.169293	0.295569
23	0.241776	0.257334	0.205737	0.294691
24	0.199526	0.287055	0.284262	0.313362
25	0.273647	0.299759	0.266106	0.333127
26	0.258242	0.288529	0.263361	0.271273
27	0.259743	0.281895	0.309879	0.301405
28	0.255122	0.325952	0.200246	0.306815
29	0.256761	0.285733	0.192828	0.293787
30	0.191721	0.313998	0.315653	0.28764
μ	0.242225233	0.288565467	0.2392822	0.307044467

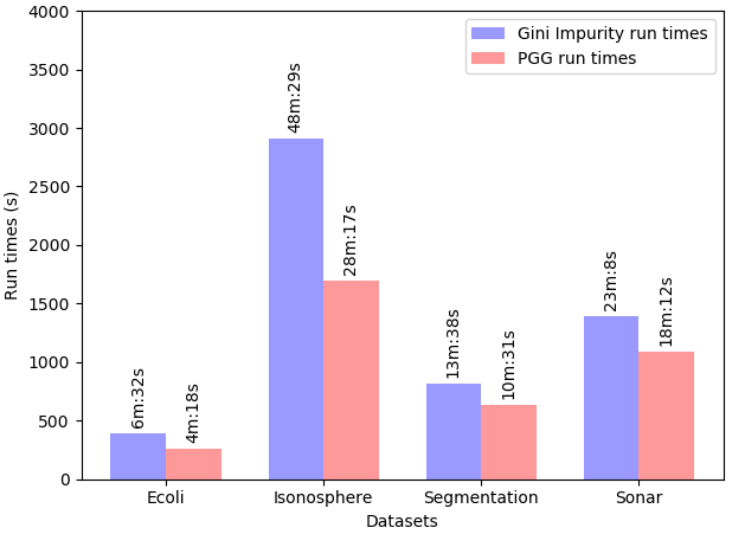


FIGURE 4.1: Running times by dataset and split metric

In order to evaluate the potential of using RMDI scores for feature ranking, we observe a sample of actual feature ranks from the ecoli dataset[10] in Table 4.6. Although the δGI metric seems to produce varying ranks, some intuitive finding can be made. It can be noted that features 2 and 3 are generally weak, since they consistently occur within the top two rank positions. In contrast to this, feature 5 potentially has high discriminative power since it prominantly occurs within the bottom half of the ranking order. The PGG metric showed greater stability than δGI, with only one variation in ranking order throughout all 30 runs. This variation happens to be the only anomaly for the PGG metric over the ecoli dataset. This observation once again demonstrates the potential of PGG as a split criteria in RFs.

TABLE 4.6: A Sample of Ecoli Dataset Feature Ranks

Run	**Feature Ranks**	
	δGI	**PGG**
1	2, 3, 4, 0, 1, 6, 5	0, 1, 2, 3, 4, 5, 6
2	2, 3, 6, 0, 5, 1, 4	0, 1, 2, 3, 4, 5, 6
3	3, 2, 0, 4, 6, 1, 5	0, 1, 2, 3, 4, 5, 6
4	3, 2, 0, 1, 4, 5, 6	0, 1, 2, 3, 4, 5, 6
5	2, 3, 1, 6, 0, 4, 5	0, 1, 2, 3, 4, 5, 6
6	3, 2, 6, 4, 1, 0, 5	1, 5, 3, 6, 4, 2, 0

[10]This dataset is easier to analyse since it has fewer features. Only the first 6 runs have been shown; displaying all 30 runs is redundant as PGG yielded the same outcome except for run 6

4.6 Conclusion

This study sought to improve the MDI RF feature selection approach by proposing a new split metric (PGG) and applying the RMDI feature score. PGG has shown potential for use in large datasets as its effectiveness and efficiency seemed more pronounced in this context. Given the fact that RFs comprising features with higher RMDI scores had higher strength and the datasets used included multi-valued predictor variables, we conclude that our approach was reliable in partitioning strong and weak features. The observed rankings were however generally very fluid, making it difficult to establish a clear order of features in terms of discriminative ability.

In the next chapter, we apply the proposed methods to retinal image data and continue our quest for a suitable feature ranking mechanism based on RFs in that context.

Chapter 5

Retinal Image Enhancement through Gabor Features

5.1 Introduction

Gabor filters are based on Gaussian modulated sinusoidal waves [167, 168], making them effective in contexts such as vessel enhancement, which rely on both texture and line detection. Although Gabor filters are a suitable vessel enhancement operator, potential drawbacks include dependence on response normalization and parameter tuning. Sub-optimal normalization can lead to image blurring, poor contrast and false vessel detection. Poor parameter selection can lead to the suppression of vessels that should otherwise be enhanced. Kyrki et al. [169] highlight the significance of illumination invariance for object recognition and recommend normalized Gabor filters for implementing it. The robustness of Gabor filters has traditionally been ensured through rotation and scale invariance. Li et al. [170] modify Gabor filters to include a scale parameter in a bid to increase the detection of vessels of varying width.

Previous applications of Gabor filters to retinal image segmentation [17, 101, 171–174] have typically made use of a bank of filters, each with varying orientation and scale. Vessel enhancement is generally achieved at pixel level by superimposing the maximum real or wavelet response over an entire bank of filters. The maximum response is generally used without empirical justification and consideration of other possible features. In this chapter we introduce two new Gabor features: average magnitude response and response convexity, in a bid to further enhance retinal blood vessels. We subsequently apply our RF feature selection scheme on a retinal image dataset in order to confirm the most effective feature for later use in automatic thresholding.

5.2 Gabor Filters

Gabor filters convolve images with a sinusoidal Gaussian modulated function that is sensitive to orientation, frequency and bandwidth [168]. The filter response is a complex number with real and imaginary components that are orthogonal [175]. This results in four options for representing the filter response, namely the real, imaginary, phase and magnitude components. The real component can be represented as follows:

$$g_{real}(x, y, \lambda, \theta, \psi, \sigma, \gamma) = G(x', y', \sigma, \gamma) W_1(x', \lambda, \psi), \tag{5.1}$$

where G, W_1, x' and y' are defined by the respective equations (5.2-5.6) below.

$$G(x', y', \sigma, \gamma) = \exp\left[-\pi\left(\frac{x'^2}{\sigma} + \frac{\gamma y'^2}{\sigma}\right)\right]. \tag{5.2}$$

$$W_1(x', \lambda, \psi) = \cos\left(2\pi\frac{x'}{\lambda} + \psi\right). \tag{5.3}$$

$$W_2(x', \lambda, \psi) = \sin\left(2\pi\frac{x'}{\lambda} + \psi\right). \tag{5.4}$$

$$x' = x\cos\theta + y\sin\theta. \tag{5.5}$$

$$y' = -x\sin\theta + y\cos\theta. \tag{5.6}$$

The imaginary component, $g_{imaginary}$ is defined in a similar manner to equation 5.1, except W_1 is replaced with W_2. The phase (g_{phase}) and magnitude ($g_{magnitude}$) are calculated by arctan $\frac{g_{imaginary}}{g_{real}}$ and $\sqrt{g_{real}^2 + g_{imaginary}^2}$ respectively. The Gaussian envelope is represented[1] by G [169] while the sinusoidal wave is represented by W. The wavelength and phase-offset of the sinusoidal wave are symbolized by λ and ψ respectively. The angle of orientation is represented by θ while the sigma and aspect ratio of the Gaussian envelope are represented by σ and γ respectively. The co-ordinates of the point to be filtered are represented by x and y. The half response bandwidth, b of the Gabor filter is measured in octaves and is used to regulate the spread and shape of the filter. It is related to σ and λ according to equation (5.7) [176]. Given λ and b, it is therefore possible to calculate σ based on equation (5.7).

$$\frac{\sigma}{\lambda} = \frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}\left[\frac{2^b + 1}{2^b - 1}\right]. \tag{5.7}$$

[1] This representation was chosen because it emphasizes the individual contributions of the x and y coordinates on the filter orientation.

Gabor filters are suited for texture segmentation and are known to be a good model of how the human visual system processes light signals [177]. Their effectiveness is however hindered by the tedious task of parameter selection. Research that implements Gabor filters therefore usually either uses predetermined parameters or has an elaborate optimal parameter selection process. Cross sections of blood vessels which are perpendicular to the direction of blood flow usually have a Gaussian grey level distribution. As such, Gabor filters are seen as effective features for vessel detection.

5.2.1 Gabor Filter Parameters

The Gabor filter requires several parameters to be tuned based on the context of application. Having explored the real component of the Gabor filter in a previous study [178], we consider the magnitude response in this work as it aggregates both real and imaginary components. We compare its discriminative ability against that of a few other Gabor features in order to validate its superiority. A phase offset of 0 is used due to the fact that every retinal pixel is overlapped by the center of a kernel window when it is convolved. This means that the peak of the Gabor wave will correspond to the pixel being convolved. We essentially measure the correlation between a two dimensional neighbourhood profile centered on a given pixel and the Gabor kernel.

In standard data sets such as the DRIVE and STARE collections, the width of a retina vessel may lie within the range of [2, 10] pixels [179]. In order to detect a complete vessel profile at a given pixel, a wave needs to be centered on the vessel's cross section and be oriented in the direction of the vessel. Additionally, the kernel window needs to be large enough to capture the vessel's neighbourhood so that the Gabor filter response is effectively contextualized. A kernel size of 21x21 pixels was used in this study. Kernel sizes larger than this were explored but they did not lead to any improvement in results. The spread of the Gaussian function in the kernel is ensured by using a bandwidth of 0.25 octaves in all experiments. The ellipticity of the Gabor filter is implemented using an aspect ratio of 0.75. The wavelengths that are used in this study are 4, 8 and 12. Wavelengths of less than 4 pixels were found to be susceptible to noise interference.

The orientations are rotated by 15 degrees in the range [0, 180) degrees as is the case in previous literature [12]. All possible combinations of wavelength and orientation are considered, resulting in a filter bank of size 36. Each of these wavelength-orientation combinations is used to create a Gabor kernel of size: 21x21 pixels. Every retinal pixel is convolved with each of these kernels, resulting in 36 Gabor filter responses. Non-retinal pixels within the kernel's corresponding image window are not convolved. The response

image comprises of the highest Gabor filter response for each pixel. This image will subsequently be thresholded in order to yield a binary image.

Given the fact that Gabor filter banks generally yield a wealth of information describing pixel neighbourhoods, we ponder on the possibility of other ways of aggregating this information into a single feature value. We propose two new features: average magnitude and response convexity and compare their discriminative ability against the traditional maximum response approach. The RF techniques from the previous chapter are enhanced to enable the identification of superior features based on the correlation of reciprocal ranks with RF strengths. The primary objective is to establish/confirm the relative discriminative ability of the maximum response against the proposed features.

5.2.2 Gabor Filter Normalization

The Gabor Filter Response (GFR) may be affected by the general illumination of a pixel's neighbourhood. Hence, normalization is essential to ensure illumination invariant responses [169]. In general, normalization has the effect of scaling responses down to a specific range whilst attempting to either preserve desired characteristics of an image or suppress noise.

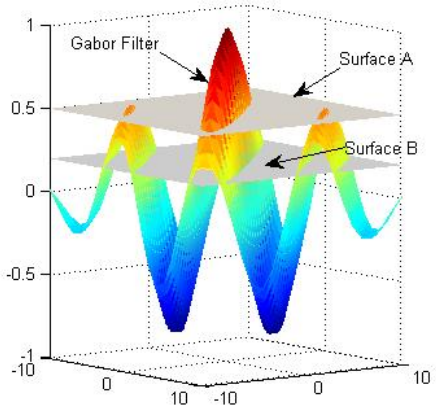


FIGURE 5.1: Gabor kernel superimposed on 2 surfaces of constant intensity.

Illumination variation is suppressed when regions of constant illumination generate a zero response. In this case, non-zero responses are strictly a result of line detection as opposed to high pixel intensities. Figure 5.1 shows a surface plot of a Gabor filter kernel with a size of 21x21 pixels. This filter has been generated using equation (5.1) with a bandwidth of 0.25 octaves, a wavelength of 8 pixels, an aspect ratio of 0.75 and an orientation of 45 degrees. The filter is superimposed on two surfaces of constant intensity. Surface A represents an image of size 21x21 pixels, each of which has an

intensity of 0.5. Surface B represents a slightly darker image of the same size with an intensity of 0.2. If this Gabor filter were illumination invariant it would produce the same response when filtering each surface. It however produces a response of 1.0173 and 0.4069 for surface A and B respectively. An effective normalization method is needed to complement this filter such that it is illumination invariant.

This study incorporates base-offset encoding to Gabor filter normalization in a bid to avert illumination inconsistency and improve vessel enhancement in retinal images [178]. The base-offset encoding scheme is a common lossless compression method that is adequate for medical images [180].

A given neighbourhood in an image exhibits smoothness when its pixels have a high similarity with each other. In such a case, the difference between the maximum and minimum intensity in the neighbourhood is small. This property may be used to efficiently encode the intensities by representing them based on their relative distance from a chosen base intensity. In this study, the median of an image window is used as the base. Effectively, this technique reduces the importance of the actual intensity whilst emphasizing on its relative intensity within a neighbourhood. Algorithm 1 describes the proposed new method in the case of a median offset. The algorithm accepts an input image that has been convolved with the scaled Gabor equation defined as:

$$g'(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \sqrt{\frac{\gamma}{2\pi\sigma^2}} g(x, y, \lambda, \theta, \psi, \sigma, \gamma). \tag{5.8}$$

The scale factor represents the inverse of the volume under the Gaussian surface [181] and this volume is equivalent to the sum of intensities on the surface. The scale factor therefore ensures that the integral of the Gaussian surface in the Gabor kernel window sums up to one. The median is calculated from a window that is cropped around a center pixel and has the same size as the Gabor kernel. This median is used as the base for calculating offsets within the cropped sub-image because it generally preserves raw intensity values as its computed value is one of the actual intensities in an image window. The resulting filtered image is therefore likely to have a high resemblance of the original image. This is not the case with bases such as the mean as they do not necessarily correspond to an existing intensity in an image window. The median offset yields both negative and positive values when used as a base. This is desirable in order to capture relative intensity information within an image window.

Algorithm 1 Calculation of Median-offset GFR, I_{MGR}, of an image, I_{src}

Require: I_{src}: Source Image

Require: $K_{w\times w}$: Normalized Gabor kernel of size $w \times w$

Ensure: I_{MGR}: Median-offset normalized Gabor filter Response image

for all pixel(x, y) in I_{src} **do**
  $r_{xy} := 0$
  $C_{w\times w}$ = Square sub-image of I_{src} centered at (x, y) and of width w
  m_{xy} = median value of $C_{w\times w}$
  for $i := -\frac{w}{2}$ to $\frac{w}{2}$ **do**
    for $j := -\frac{w}{2}$ to $\frac{w}{2}$ **do**
      $offset := I_{src}(x + i, y + j) - m_{xy}$
      $r_{xy} := r_{xy} + (offset \times K_{w\times w}(i, j))$
    end for
  end for
  $I_{MGR}(x, y) := r_{xy}$
end for

5.3 Proposed Secondary Gabor Features

A Gabor filter bank convolves an image surface and produces a 2-dimensional output corresponding to a predefined orientation and scale space, for each pixel. Besides the maximum response, we propose two new descriptors of this output space, these are described below. We also consider the response generated in the orientation orthogonal to that of the maximum response, at the same wavelength. This is a simple feature that may be useful at differentiating normal blood vessels from pathological landmarks which may not have the same normal to orthogonal response ratios.

5.3.1 Average Magnitude Response

This feature captures the general magnitude response across the entire parameter space. Algorithm 2 outlines how an input image I_{src} is enhanced using a filter bank B, comprising $|\Theta| * |\Lambda|$ Gabor kernels of the form $B(\theta, \lambda)_{w\times w}$.

Algorithm 2 Calculation of Average Magnitude Response, I_{AMR}, of an image, I_{src}

Require: I_{src}: Source Image
Require: $B(\theta, \lambda)_{w \times w}$: Gabor kernel of size $w \times w$, orientation θ and wavelength λ
Require: $\Theta = \{\theta_0, \ldots \theta_t\}$: Orientation space
Require: $\Lambda = \{\lambda_0, \ldots \lambda_l\}$: Wavelength space
Ensure: I_{MGR}: Average Magnitude Response image

for all pixel(x, y) in I_{src} **do**
  $r_{xy} := 0$
  $C_{w \times w}$ = Square sub-image of I_{src} centered at (x, y) and of width w
  m_{xy} = median value of $C_{w \times w}$
  for $i := -\frac{w}{2}$ to $\frac{w}{2}$ **do**
    for $j := -\frac{w}{2}$ to $\frac{w}{2}$ **do**
      $offset := I_{src}(x + i, y + j) - m_{xy}$
      for all θ in Θ **do**
        for all λ in Λ **do**
          $r_{xy} := r_{xy} + (offset \times B(\theta, \lambda)_{w \times w}(i, j))$
        end for
      end for
    end for
  end for
  $I_{MGR}(x, y) := \frac{r_{xy}}{|\Theta| * |\Lambda|}$
end for

It thus measures the uniformity of responses over the entire orientation and wavelength spectrums. Small values will generally indicate the presence of pixels in non-linear and non-scalable objects, and vice versa for large values.

5.3.2 Response Convexity

This feature can be interpreted as a measure of how regular a pixel's neighbourhood is. We expect vessels to have more regular neighbourhoods than pathological structures which may have irregular boundaries. We measure the cummulative magnitude response for each predefined orientation then we plot a radar chart based on these totals as shown in Figure 5.2. Convexity is common image processing shape descriptor that is defined as the ratio of the surrounding convex hull perimeter over an object's perimeter [102]. In this case, the object is a radar chart generated from Gabor filter responses. We expect that a normal blood vessel will have a high convexity since magnitude responses should

be highest along the vessel's orientation and they gradually decrease as the corresponding orthogonal orientation is approached.

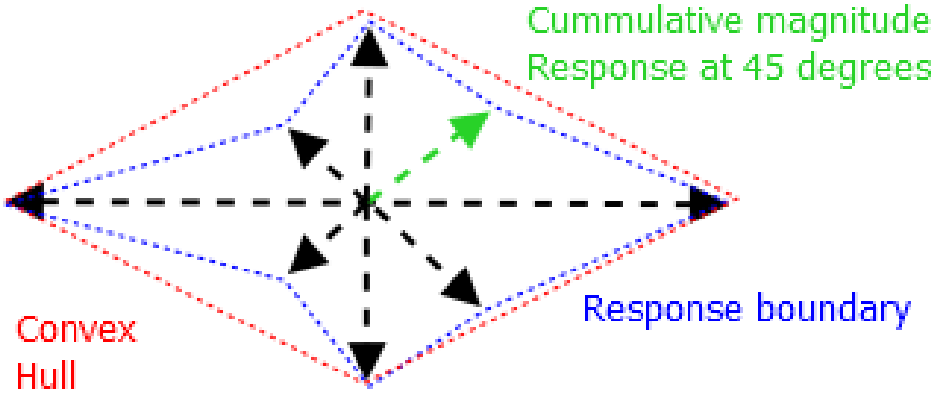


FIGURE 5.2: Convex hull based on magnitude responses

5.4 Computing RMDI on Retinal Gabor Features

We intend using RMDI as an importance measure that selects the feature with greatest potential for effectiveness during automatic thresholding. We first describe our proposed method for selecting the most effective feature then outline how the identified Gabor features: maximum response, average magnitude response, response convexity and orthogonal magnitude; are extracted from the training set images of the DRIVE dataset.

5.4.1 Reciprocal Ranks and Random Forest Strength

After observing the significant association between RMDI and RF strength in the previous chapter, we seek a method for quantifying feature effectiveness in the light of occasional feature rank and RF strength correlation anomalies. We have established that RMDI feature ranks can be generally unstable[2] and are therefore unsuitable for direct use as selectors of the most discriminative features. We propose using the correlation of reciprocal feature ranks with RF strengths as indicators of how much features influence RF strength. A reciprocal rank is simply the inverse of a rank; it is a concept used in Information Retrieval (IR) to score IR systems based on the accessibility of relevant responses within search query output lists [182].

[2] As shown in Table 4.6.

We postulate that an effective feature will influence RF strength mostly when it has a low rank[3]. and vice versa for features with high ranks. We measure how well a feature's reciprocal rank correlates with RF strength over 30 runs of RF induction. The highest correlation should ideally indicate the most discriminative feature.

5.4.2 Experimental Protocol

Since RMDI is a RF based feature selection method, we start by noting an important observation from the few previous studies that were found to employ RFs in the context of retinal images. When applying RFs to blood vessel segmentation, Wang et al. [150] used convolutional neural network features enforced at pixel level. The authors noted that due to the large volume of pixels, RF classifier training is susceptible to large computational time and huge storage space requirements. An automatic superpixel based sample selection method was introduced to reduce this overhead. Chang et al. [149] utilized a pool of hybrid features consisting of modern and classical features for characterizing retinal image pixels. To train each RF tree, 10 000 separate vessel and non-vessel samples were randomly selected from each training image. As a result only about 6% of the available pixels are used when training each RF tree, due to the sampling. Despite the use of sampling methods during classifier training, the studies reported segmentation accuracy rates of 97.67% and 94.74% respectively.

Inspired by these previous studies, we utilize a dlib [183] function used to pick initial points in clustering algorithms like the K-Means method. In a nutshell, it selects a set of reasonable starting points by picking points that are all well-separated from each other within the feature space. We use this function to select 10 000 separate vessel and non-vessel samples from each training image as was done in [149]. To further simplify this process, we only utilize images 37 - 39 of the DRIVE training set as these seem to be representative of normal, poorly contrasted and pathology containing retinal images. We then apply our proposed reciprocal rank correlation metric on our selected Gabor features.

5.4.3 Results

Tables 5.1 and 5.2 present the reciprocal rank and RF strength correlations from our retinal image Gabor feature data. We illustrate the intuition behind this presentation with an example. Run 1 using the GI metric ranked feature 3 as the strongest, followed by features 0 and 1; feature 2 was the weakest. From the results of the first run, the

[3]In this case, features are sorted in descending order of RMDI such that a ranks of 1 and 2 yield reciprocal ranks of 1 and $\frac{1}{2}$

corresponding reciprocal ranks of features 0 to 3 are 0.5, 0.33, 0.25 and 1 respectively. The correlation of RF strength and reciprocal rank for each feature is calculated at the bottom of each table. Both Tables 5.1 and 5.2 confirm feature 0 (maximum magnitude response) as having the highest correlation with RF strength.

It is interesting to note that both the GI and PGG metrics, reveal the same outcome for highest and second highest correlation, maximum magnitude response and average magnitude response respectively. The PGG metric is unable to compute a correlation for response convexity as a DIV/0 error is generated since, the reciprocal rank is constant for all 30 runs. It is therefore not possible to determine the third and fourth most effective feature in this context.

Figure 5.3 showcases the image enhancements produced by the maximum magnitude response feature on images 37 - 39 of the DRIVE dataset. Although Figure 5.3(a) presents some pathology in the form of hard exudates next to the fovea, our image enhancement is able to effectively isolate this landmark. Although Figures 5.3(b) and 5.3(c) have different levels of background contrast both have managed to highlight blood vessels and ignore the fovea. There is however some false enhancements around the optic disk of each image.

Figure 5.4 shows the enhancements of image 37 of the DRIVE dataset using our proposed Gabor features. Feature 2, exhibits a lot of noise while Features 1 and 3 are comparable with Feature 0. Feature 2, Response Convexity is hence not yet suitable for use in image segmentation.

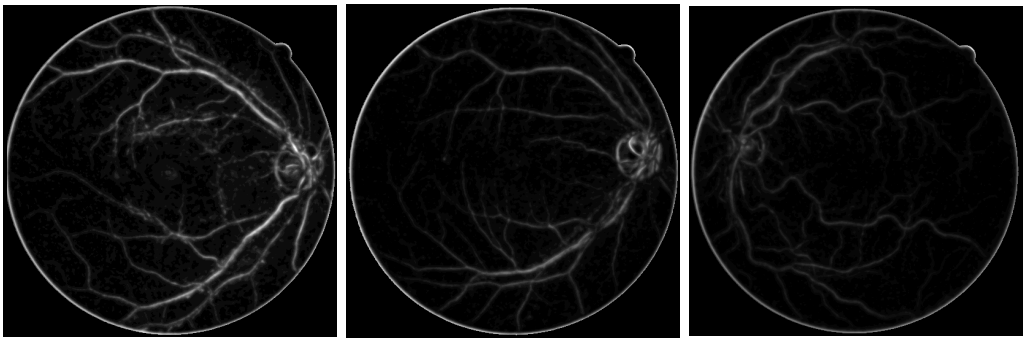
TABLE 5.1: Reciprocal Rank RF Strength Correlations based on GI Metric

Run	**Rank**				**RF Strength**	**Feature Reciprocal Rank**			
	4	**3**	**2**	**1**		**0**	**1**	**2**	**3**
1	2	1	0	3	0.169546	0.50	0.33	0.25	1.00
2	0	1	3	2	0.17335	0.25	0.33	1.00	0.50
3	0	2	3	1	0.176465	0.25	1.00	0.33	0.50
4	2	1	3	0	0.172212	1.00	0.33	0.25	0.50
5	1	0	2	3	0.170586	0.33	0.25	0.50	1.00
6	2	3	1	0	0.177599	1.00	0.50	0.25	0.33
7	2	3	0	1	0.166957	0.50	1.00	0.25	0.33
8	3	2	0	1	0.173942	0.50	1.00	0.33	0.25
9	0	3	2	1	0.17302	0.25	1.00	0.50	0.33
10	0	2	3	1	0.174144	0.25	1.00	0.33	0.50
11	0	2	1	3	0.174228	0.25	0.50	0.33	1.00
12	2	1	0	3	0.168828	0.50	0.33	0.25	1.00
13	2	1	0	3	0.171027	0.50	0.33	0.25	1.00
14	2	0	3	1	0.167512	0.33	1.00	0.25	0.50
15	2	3	0	1	0.17182	0.50	1.00	0.25	0.33
16	1	2	3	0	0.170612	1.00	0.25	0.33	0.50
17	2	3	1	0	0.176553	1.00	0.50	0.25	0.33
18	2	0	1	3	0.170706	0.33	0.50	0.25	1.00
19	2	0	3	1	0.171597	0.33	1.00	0.25	0.50
20	1	3	0	2	0.167254	0.50	0.25	1.00	0.33
21	2	3	0	1	0.174186	0.50	1.00	0.25	0.33
22	2	3	0	1	0.164737	0.50	1.00	0.25	0.33
23	2	1	3	0	0.172776	1.00	0.33	0.25	0.50
24	1	3	2	0	0.175263	1.00	0.25	0.50	0.33
25	2	0	1	3	0.171024	0.33	0.50	0.25	1.00
26	2	1	0	3	0.171941	0.50	0.33	0.25	1.00
27	1	2	3	0	0.165692	1.00	0.25	0.33	0.50
28	0	2	3	1	0.170351	0.25	1.00	0.33	0.50
29	2	0	1	3	0.163986	0.33	0.50	0.25	1.00
30	2	3	1	0	0.169495	1.00	0.50	0.25	0.33
					SUM	16.50	18.08	10.33	17.58
					STD	0.29	0.32	0.19	0.29
					COR	0.12	0.06	0.02	-0.21

TABLE 5.2: Reciprocal Rank RF Strength Correlations based on PGG Metric

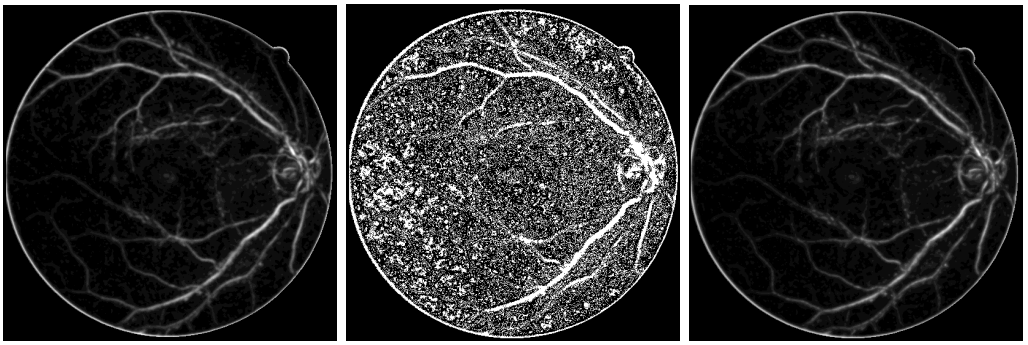
Run	**Rank**				**RF Strength**	**Feature Reciprocal Rank**			
	4	**3**	**2**	**1**		**0**	**1**	**2**	**3**
1	2	3	1	0	0.17427	1.00	0.50	0.25	0.33
2	2	3	0	1	0.17011	0.50	1.00	0.25	0.33
3	2	0	3	1	0.165754	0.33	1.00	0.25	0.50
4	2	0	1	3	0.162623	0.33	0.50	0.25	1.00
5	2	1	3	0	0.166112	1.00	0.33	0.25	0.50
6	2	0	3	1	0.171672	0.33	1.00	0.25	0.50
7	2	3	1	0	0.160423	1.00	0.50	0.25	0.33
8	2	0	3	1	0.163049	0.33	1.00	0.25	0.50
9	2	3	1	0	0.170793	1.00	0.50	0.25	0.33
10	2	3	0	1	0.162367	0.50	1.00	0.25	0.33
11	2	0	1	3	0.165896	0.33	0.50	0.25	1.00
12	2	0	1	3	0.159626	0.33	0.50	0.25	1.00
13	2	0	3	1	0.16835	0.33	1.00	0.25	0.50
14	2	1	0	3	0.169256	0.50	0.33	0.25	1.00
15	2	0	1	3	0.163665	0.33	0.50	0.25	1.00
16	2	0	3	1	0.167954	0.33	1.00	0.25	0.50
17	2	1	0	3	0.159729	0.50	0.33	0.25	1.00
18	2	1	3	0	0.162361	1.00	0.33	0.25	0.50
19	2	0	1	3	0.166069	0.33	0.50	0.25	1.00
20	2	3	0	1	0.163502	0.50	1.00	0.25	0.33
21	2	0	1	3	0.164907	0.33	0.50	0.25	1.00
22	2	3	1	0	0.165738	1.00	0.50	0.25	0.33
23	2	1	3	0	0.170556	1.00	0.33	0.25	0.50
24	2	3	1	0	0.164326	1.00	0.50	0.25	0.33
25	2	3	0	1	0.162806	0.50	1.00	0.25	0.33
26	2	3	1	0	0.171716	1.00	0.50	0.25	0.33
27	2	0	3	1	0.171437	0.33	1.00	0.25	0.50
28	2	3	0	1	0.167544	0.50	1.00	0.25	0.33
29	2	3	0	1	0.165099	0.50	1.00	0.25	0.33
30	2	0	3	1	0.169438	0.33	1.00	0.25	0.50
					SUM	17.33	20.67	7.5	17
					STD	0.29	0.28	0	0.28
					COR	0.16	0.14	NA	-0.31

(A) Gray Image 37 (B) Gray Image 38 (C) Gray Image 39



(D) Gabor Image 37 (E) Gabor Image 38 (F) Gabor Image 39

FIGURE 5.3: Filtering DRIVE Images 37 -39 Using Gabor Maximum Magnitude Response



(A) Feature 1: Average Magnitude (B) Feature 2: Response Convexity (C) Feature 3: Orthogonal Response

FIGURE 5.4: Filtering DRIVE Image 37 using proposed Gabor Features

5.5 Conclusion

This chapter has presented a RF based means of assessing feature effectiveness based on reciprocal ranks and RF strengths. The results presented confirm the general trend in previous literature which singles out the maximum response as the preferred choice for segmentation. In the next chapter we present work on our segmentation of retinal images using this feature and Valley emphasis thresholding.

Chapter 6

Automatic Thresholding

6.1 Introduction

To facilitate effective visual tissue analysis in medical image segmentation, the region representing biological tissue of interest is normally isolated from its background digitally. Although this task can be accomplished through manual interaction with an image histogram, the automated approach may be prefered due to its potential for autonomous and scalable operation.

The Otsu method [184] is a widely used automatic thresholding algorithm which, despite being relatively simple, is reliable in the context of images with well formed bimodal distributions. The effectiveness of this method is however significantly reduced when it is presented with image histograms that tend towards unimodal distributions.

The Valley Emphasis (VE) method [21] improves the Otsu method by exagerating valley points which are likely candidates for optimal thresholds in both unimodal and bimodal distributions. It has been found to be effective in industrial defect detection applications and competes favourably with other Otsu method adaptations. This study explores some recent improvements to the VE method and applies them to the challenging context of retinal blood vessel segmentation.

An attempt is made to build a simplified hybrid of VE adaptations so as to leverage their individual strengths. We note that the VE bears the computational overhead of considering all possible thresholds in search for the optimal candidate. We argue that since emphasis is ultimately placed on grey levels with lowest probability, it might be worthwhile excluding the highest peaks in the histogram from the search. We formulate the following research question to guide this study: can the proposed hybrid approach

jointly improve the effectiveness and efficiency of VE in the context of Gabor filtered retinal images.

The remainder of this chapter is structured as follows. Section 2 gives a brief description of the Otsu method and various VE variations including our hybrid approach. Section 3 outlines the experimental details of the study while Section 4 presents the results. We conclude by reflecting on our research question in the light of the reported findings.

6.2 Automatic Thresholding Methods

The following notation is assumed in our descriptions of various automatic thresholding methods. Given an $M \times N$ grey level image G, $G(x, y) = g$ represents the intensity of a pixel within the image, where $0 \leq x < M$, $0 \leq y < N$ and $0 \leq g < L$. The image has $n = M \times N$ pixels and L different grey level intensities. These intensities are represented by indexes as opposed to actual pixel values. The frequency/likelihood of a grey level g in a normalized histogram of G is represented by:

$$h(g) = \frac{n_g}{n}, \tag{6.1}$$

where n_g is the number of pixels with such an intensity in the image. Furthermore the average grey level within a portion of the image histogram is represented by $\mu_{(l_0,l_1)}$ where $0 \leq l_0 < l_1 < L$. Likewise the probability of a portion of the histogram is $p_{(l_0,l_1)}$, such that $p_{(0,L-1)} = 1$. In both cases, the portion of the histogram under consideration, is the range $[l_0, l_1]$ of grey levels.

6.2.1 Otsu Method

Although a variety of automatic thresholding techniques have been proposed [185], the Otsu method is still widely used to date due to its simplicity and moderate effectiveness. We outline its main concepts below.

If the image G is binarized using a threshold t, the probabilities and mean values of the two segments can be calculated as follows:

$$p_{(0,t)} = \sum_{g=0}^{t} h(g), \tag{6.2}$$

$$p_{(t+1,L-1)} = \sum_{g=t+1}^{L-1} h(g), \tag{6.3}$$

$$\mu_{(0,t)} = \sum_{g=0}^{t} \frac{gh(g)}{p_{(0,t)}}, \tag{6.4}$$

$$\mu_{(t+1,L-1)} = \sum_{g=t+1}^{L-1} \frac{gh(g)}{p_{(t+1,L-1)}}, \tag{6.5}$$

while the mean value of the whole image is:

$$\mu_{(0,L-1)} = \sum_{g=0}^{L-1} gh(g). \tag{6.6}$$

The between-class variance $\sigma_O^2(t)$ which ultimately measures the disparity between two classes is then calculated as follows:

$$\sigma_O^2(t) = p_{(0,t)}(\mu_{(0,t)} - \mu_{(0,L-1)})^2 + p_{(t+1,L-1)}(\mu_{(t+1,L-1)} - \mu_{(0,L-1)})^2. \tag{6.7}$$

This representation can be further simplified to:

$$\sigma_O^2(t) = p_{(0,t)}\mu_{(0,t)}^2 + p_{(t+1,L-1)}\mu_{(t+1,L-1)}^2. \tag{6.8}$$

The optimal threshold t^* which maximizes $\sigma_O^2(t)$, is defined as follows:

$$t^* = \underset{0<t<L-1}{\operatorname{argmax}}\, \sigma_O^2(t). \tag{6.9}$$

6.2.2 Valley Emphasis Method

Ng proposed the VE method [21] as an improvement to the Otsu method, that accommodates the automatic selection of optimal thresholds within unimodal distributions. This adaptation is facilitated by simply scaling the between-class variance with a weight that favours low thresholds as shown below:

$$\sigma_V^2(t) = (1 - h(t))(p_{(0,t)}\mu_{(0,t)}^2 + p_{(t+1,L-1)}\mu_{(t+1,L-1)}^2), \tag{6.10}$$

$$t^* = \underset{0<t<L-1}{\operatorname{argmax}}\, \sigma_V^2(t). \tag{6.11}$$

6.2.3 Valley Emphasis Method Adaptations

One of the recent adaptations of the VE method is by Fan and Lei [186] who note its weakness in scenarios where the variance of the object is significantly different from that of the background. This is said to be pronounced where the proportion of the

object/foreground is small in comparison to that of the background. In such situations where the background has a high mean grey level intensity, the scaling factor in the VE method is insufficient to favour thresholds with low probabilities. This finding is of particular interest to us, since the context of our thresholding is dominated by a low proportion of blood vessels in comparison to the retinal background.

The proposed adaptation, Neighbourhood Valley Emphasis (NVE) method, modifies the scaling factor $(1 - h(t))$ such that it considers the valley as a neighbourhood instead of a point. The probability of the valley is expressed as follows:

$$\hat{h}(t) = h(t-m) + \cdots + h(t-1) + h(t) + h(t+1) + \cdots + h(t+m), \tag{6.12}$$

where its neighbourhood is of size $z = 2m + 1$. The corresponding objective function and optimal threshold are defined by:

$$\sigma_N^2(t) = (1 - \hat{h}(t))(p_{(0,t)}\mu_{(0,t)}^2 + p_{(t+1,L-1)}\mu_{(t+1,L-1)}^2), \tag{6.13}$$

$$t^* = \underset{0<t<L-1}{\operatorname{argmax}}\, \sigma_N^2(t). \tag{6.14}$$

The new scaling factor $(1 - \hat{h}(t))$ ensures that t^* will always be at a value which is not only low but occurs in a neighbourhood of low probabilities. The authors however note the influence of z on the quality of segmentation and advise against using low or high neighbourhood lengths. From the conducted experiments, the recommended setting is $z = 11$ while threshold values such as 3 and 35 are considered to be too small or too big respectively. We argue that the choice of z should not be generalized but should instead, be based on the context under consideration. For example, low neighbourhood sizes could be suited for images with a small grey level range and low histogram variance. In such cases, it is likely that the valley separating classes is narrow. The total probability of such a narrow valley would still be favoured when faced with a high average background grey level, since the image intensity range is small.

Ng et al. subsequently proposed an alternative way [187] of incorporating the neighbourhood of a valley in the scaling factor, by introducing a Gaussian weighting mechanism as shown below.

$$\tilde{h}(t) = \sum_{i=-m}^{m} h(t-i) \exp^{-\frac{(t-i)^2}{2\sigma^2}} \tag{6.15}$$

$$\sigma_G^2(t) = (1 - \tilde{h}(t))(p_{(0,t)}\mu_{(0,t)}^2 + p_{(t+1,L-1)}\mu_{(t+1,L-1)}^2), \tag{6.16}$$

$$t^* = \underset{0<t<L-1}{\operatorname{argmax}}\, \sigma_G^2(t). \tag{6.17}$$

This approach which we refer to as the Gaussian Valley Emphasis (GVE) method also considers grey levels within a valley neighbourhood but differs in that proximity to the valley-center is taken into account. It also has the benefit of removing noise from the image histogram due to the smoothing effect of its Gaussian function. The details of the Gaussian window size and the prefered σ are however not specified and left for experimental determination.

A recent alternative approach which employs the concept of valley deepness, has been proposed by Ng et al. [188]; we refer to it as the Deepened Valley Emphasis (DVE) method. Given a prospective histogram valley-point b, the points a and c which represent the highest points to the left and right of b respectively, will be higher than b, if b is indeed a valley-point. In other words, $h(a) > h(b)$ and $h(c) > h(b)$ if b is a valley point. Equation 6.18 shows that, if b is a confirmed valley point, then its valley-depth $\check{h}(b)$ can be modelled as the average between the heights of the peaks at a and c, relative to $h(b)$, otherwise there is no valley-depth to measure.

$$\check{h}(b) = \begin{cases} \frac{\max\limits_{0 \le a < b}(h(a)-h(b)) + \max\limits_{b < c < L}(h(c)-h(b))}{2}, & \text{if } h(a) > h(b) \text{ and } h(c) > h(b) \\ 0, & \text{otherwise.} \end{cases} \tag{6.18}$$

A new objective function was proposed, based on a scaling factor which incorporates how low and how deep a valley point is, as shown in equation 6.19.

$$\sigma_D^2(t) = (1 - h(t) + \check{h}(b))(p_{(0,t)}\mu_{(0,t)}^2 + p_{(t+1,L-1)}\mu_{(t+1,L-1)}^2), \tag{6.19}$$

$$t^* = \underset{0<t<L-1}{\operatorname{argmax}}\, \sigma_D^2(t). \tag{6.20}$$

In a nutshell, the between class variance $\sigma_D^2(t)$, defaults to $\sigma_V^2(t)$ when t is not a true valley point but adds $\check{h}(t)$ to $\sigma_V^2(t)$ for true valley points, in order to favour points with low $h(t)$ values. Due to the sensitivity of this approach to noise in the image histogram, a 1-Dimensional Gaussian filter is proposed for smoothing the histogram before this objective function can be applied. The main drawback of this approach is its computational overhead due to the consideration of all possible permutations of the points a, b and c. Since we seek to improve the efficiency of the automatic thresholding, we have therefore not incorporated the features of this method in our proposed approach.

6.2.4 Proposed Methods

Although adaptations to the VE method are reported to generally improve the effectiveness of automatic segmentation in various contexts, they do so at the expense of additional computation. Based on the observation that all VE methods typically favour

thresholds that have low $h(t)$ values, we anticipate that it may not be necessary to consider all possible values of t when searching for t^*. We propose sorting candidate thresholds in ascending order of their $h(t)$ values and then only focusing on a portion at the top of the list. On the one hand, this line of reasoning may exclude potentially good thresholds that lie outside the focus area, on the other hand, it may do well by excluding potential false positives in terms of optimal thresholds.

In order to test the viability of our proposal, which we call the Selective Valley Emphasis approach, we apply it to two extensions of the VE method that have different levels of complexity as follows.

- **Selective VE** (SVE): creates a list of candidate thresholds sorted in ascending order of $h(t)$, then maximizes $\sigma_V^2(t)$ over the top half of the sorted list.
- **Selective Gaussian Neighbourhood VE** (SGNVE): creates a new between-class variance $\sigma_{GN}^2(t)$ which is a blend of the Gaussian and Neighbourhood VE methods. This variance is maximized over the top quarter of the sorted thresholds list with the lowest $h(t)$ values. We use a fixed Gaussian window with the following weights: $\left[\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right]$, such that $\hat{h}(t) = \frac{1}{4}h(t-1) + \frac{1}{2}h(t) + \frac{1}{4}h(t+1)$. Missing boundary values are compensated for by reflecting the existing neighbouring values of t.

6.3 Experiment Details

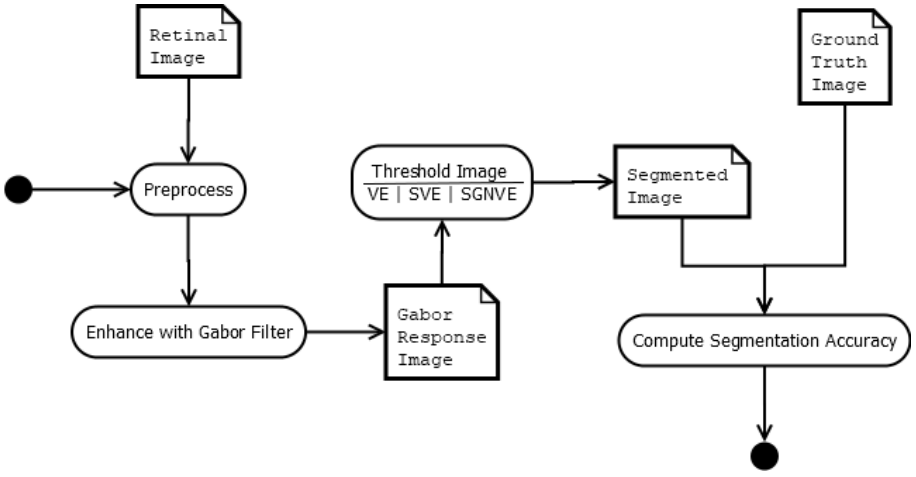


FIGURE 6.1: Retinal image segmentation steps

We evaluate the comparative effectiveness of our proposed VE methods by following the sequence of steps shown in Figure 6.1. Colour retinal input images are preprocessed by extracting the green channel, which is known to preserve the most detail for use in grey

level mode [99]. We are primarily interested in determining whether selective thresholding can improve the efficiency of VE method and if so, whether there is any compromise in effectiveness. The enhanced Gabor images are thresholded using the VE, SVE and SGNVE methods while noting running times and segmentation effectiveness. Based on our previous Gabor feature ranking, we focus on the maximum Gabor magnitude response of each retinal image pixel. We propose a minor alteration to the calculation of the Gabor filter magnitude, instead of the conventional euclidean distance, we take its squared value. Thus, the alteration is simply a matter of dropping the square root function in the calculation of $g_{magnitude}$. This alteration yields a simpler computation while retaining the integrity of the ($\leq$ and $\geq$) boolean relational operators. It does however affect the granularity of the response values and it would be interesting to note whether there is a significant influence on automatic thresholding. Our experiments are conducted on the DRIVE and STARE datasets to enable objective comparison with other studies.

6.4 Results

Table 6.1 shows the processor clock times for calls to functions implementing VE, SVE and SGNVE algorithms. The average execution time of SVE is less than that of VE on both the DRIVE and STARE datasets, with the paired t-test yielding p-values of approximately 0.007% and 0.11% respectively. The difference in execution times between VE and SVE is thus significant for the DRIVE and not for the STARE dataset. We also note that the STARE dataset execution times are consistently greater than those of the DRIVE dataset. This leads us to assume the impact of the SVE is more prominent in the context of lower VE computation overhead[1]. Although an attempt was made to improve the efficiency of SGNVE, its execution times were consistently greater than those of VE and SVE algorithms. Although the average execution times of these three algorithms differ, their thresholded output was identical in all cases, leading us to prefer the SVE algorithm since it is as effective as the other approaches but more efficient.

Table 6.2 compares our final segmentation results with those of previous work. It is interesting to note that the squared magnitude response yields a higher average accuracy than the magnitude and real response approaches, in both the DRIVE and STARE datasets. It is observed that the squared magnitude response value ranges are generally narrower than those of the magnitude response; this effect seems to favour the former approach. The results of the present study compare favourably with previous work; in some cases superior results are achieved. Sensitivity and specificity outcomes of up to

[1]From Table 3.1, we note that DRIVE images have smaller dimensions than STARE images.

Table 6.1: VE Algorithm Execution Times (ms)

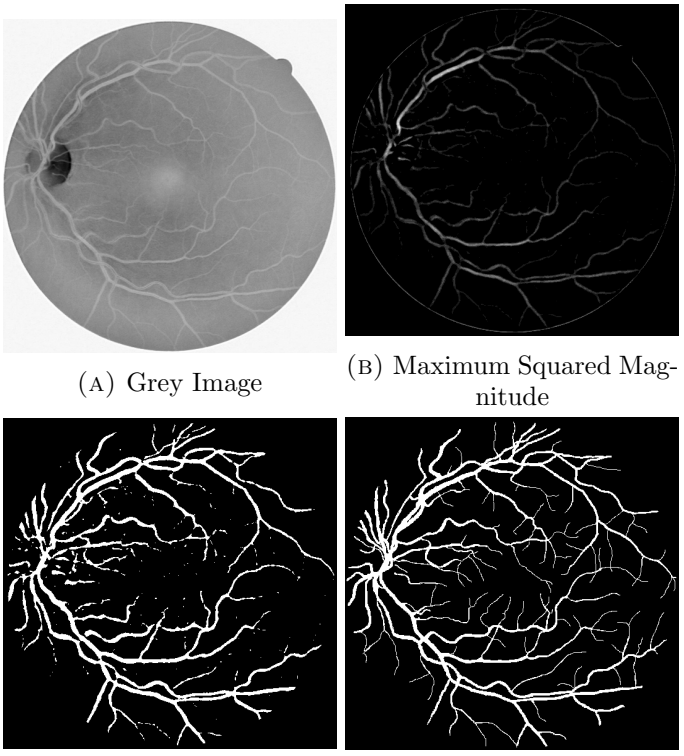
Image	**DRIVE**			**STARE**		
	VE	**SVE**	**SGNVE**	**VE**	**SVE**	**SGNVE**
1	25	25	151	35	34	212
2	29	27	146	32	31	192
3	28	22	127	31	32	204
4	36	32	165	28	29	163
5	24	25	156	41	36	190
6	26	28	148	39	41	231
7	29	26	149	34	35	195
8	24	24	149	32	33	212
9	26	26	166	46	34	188
10	27	22	129	35	34	201
11	37	30	154	38	36	206
12	25	23	140	34	37	204
13	24	23	136	35	34	189
14	24	22	153	34	34	200
15	26	23	155	36	33	191
16	25	25	137	28	32	193
17	25	23	147	38	34	211
18	31	30	158	32	32	197
19	23	22	132	32	32	174
20	26	30	158	33	31	174
μ	27	25.4	147.8	34.65	33.7	196.35

76.8% and 97.9% as well as 78.8% and 97.8% are achieved on normal images in the DRIVE and STARE datasets respectively; images exhibiting pathological traits remain a challenge. It is worth noting that although the most recent work of Neto et al. reports a very high sensitivity which surpasses that of the human observer, those results are obtained using a different gold standard called the average human observer.

We conclude our results section by exhibiting some visual segmentations achieved by our approach on normal and pathological images in Figures 6.2 - 6.5. A notable strength of our approach is the effective suppression of the optic disk and fovea regions. Although thick vessels are well located, thin vasculature remains a challenge. Image 8 of the DRIVE dataset and image 139 of the STARE dataset presented some microaneurysms. Although these were largely detected as blood vessels, further processing using shape features could possibly isolate them as pathological landmarks.

TABLE 6.2: Results of Retinal Image Thresholding

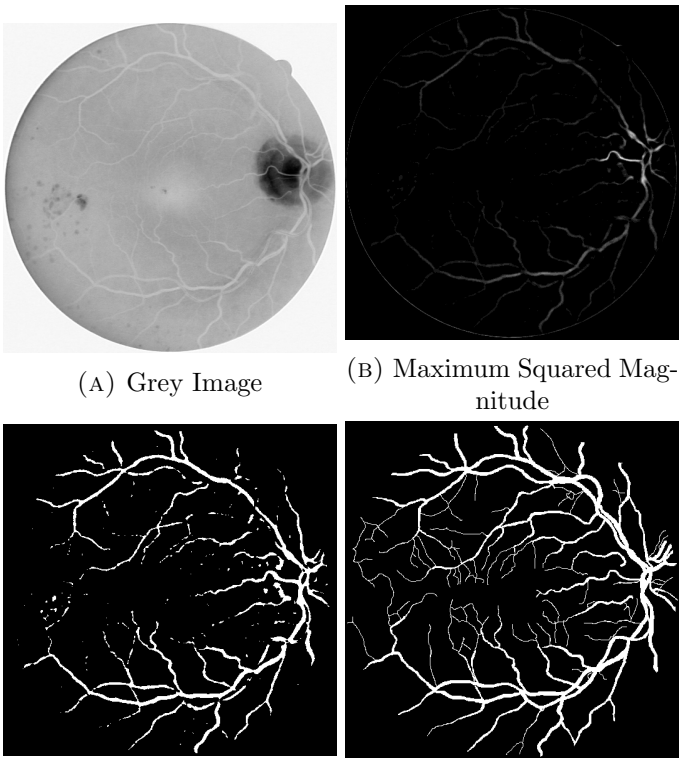
Approach			Sensitivity	Specificity	Accuracy
DRIVE	VE	Real [178]	68.3	97.22	93.54
		Magnitude	61.73	97.94	93.3
		Squared Magnitude	66.8	97.82	93.84
	Previous Work	Human observer [12]	77.63	97.23	94.70
		Yin et al. [189]	62.52	97.10	92.67
		Zhang et al. [190]	71.20	97.24	93.82
		Niemeijer et al. [115]	68.98	96.96	94.17
		Neto et al. [139]	79.42	96.31	-
STARE	VE	Real [178]	68.85	96.15	93.2
		Magnitude	69.16	96.07	93.2
		Squared Magnitude	68.6	96.39	93.58
	Previous Work	Human observer [12]	77.63	97.23	94.70
		Hoover et al. [96]	67.51	95.67	92.67
		Zhang et al. [190]	71.77	97.53	94.84
		Mendonca et al. [106]	69.96	97.3	94.4
		Neto et al. [139]	76.95	95.37	-



(A) Grey Image (B) Maximum Squared Magnitude

(C) SVE Segmentation (D) Ground Truth

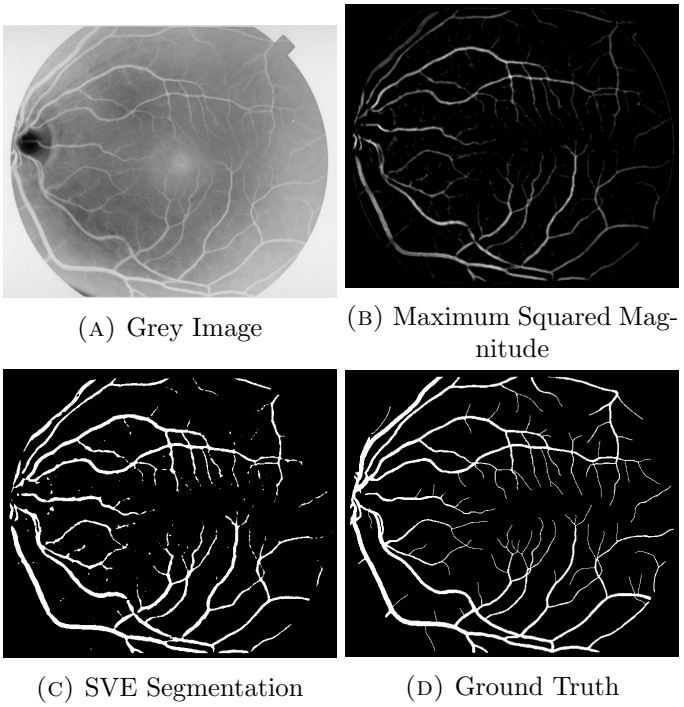
FIGURE 6.2: Segmentation of Image 1 in DRIVE Dataset



(A) Grey Image (B) Maximum Squared Magnitude

(C) SVE Segmentation (D) Ground Truth

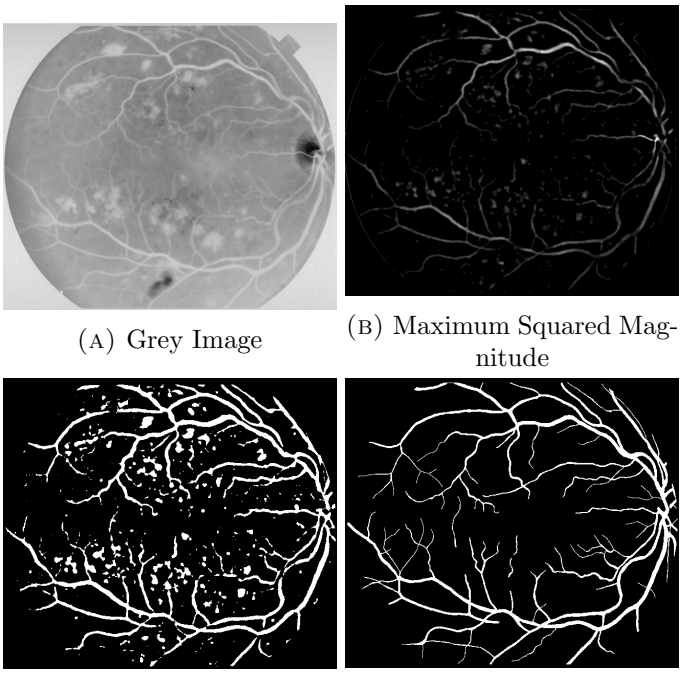
FIGURE 6.3: Segmentation of Image 8 in DRIVE Dataset



(A) Grey Image (B) Maximum Squared Magnitude

(C) SVE Segmentation (D) Ground Truth

FIGURE 6.4: Segmentation of Image 77 in STARE Dataset



(A) Grey Image (B) Maximum Squared Magnitude

(C) SVE Segmentation (D) Ground Truth

FIGURE 6.5: Segmentation of Image 139 in STARE Dataset

Chapter 7

General Conclusion

7.1 Research Overview

This thesis has presented a series of methods and techniques culminating in the automatic segmentation of retinal images. In Chapter 2, the background on the human retina was presented in order to introduce medical terms used throughout the thesis and set the context of the study. Chapter 3 reviewed the state-of-the-art in retinal image segmentation and highlighted the lack of robust engagement with computational algorithms. In Chapter 4, we propose a method that uses Random Forests (RFs) to identify strong and weak features based on their Relative Mean Decrease Impurity (RMDI). Although new Gabor features are presented in Chapter 5, it is ultimately the modification of an existing feature (squared magnitude) that proves most effective in vessel segmentation. We used the correlation of reciprocal feature ranks with RF strength to identify the most effective feature out of a set of four Gabor features. In Chapter 5, we apply a Selective Valley Emphasis (SVE) thresholding algorithm to segment retinal images that were enhanced using the maximum magnitude response. Although segmentation through SVE matches the effectiveness of the ordinary VE method, the computational effectiveness of the former is demonstrated in our experimental results.

7.2 Research Milestones

Based on the application of our RF feature selection framework on some University of California, Irvine (UCI) datasets, we are able to conclude that our Purity Gap Gain (PGG) metric in conjunction with RMDI provides an efficient and reliable feature selection mechanism. This approach is however not conducive for use in standalone feature ranking as the generated rankings can be unstable over different iterations. Using our

feature correlation method, we are able to confirm the general trend in literature, where the maximum response of the Gabor filter is perceived as being the most effective feature. In addition we manage to improve retinal image enhancement using a minor alteration to the computation of magnitude response which ultimately improves classification accuracy. When comparing the legacy VE method against its proposed variants in the context of retinal blood vessel classification, we establish that it is less efficient but equally effective.

The main aim of this study was to improve the segmentation of retinal blood vessels through the fusion of Gabor filters, RFs feature selection and automatic thresholding techniques. This has been achieved through the selection of maximum magnitude response as the most effective feature and the ultimate use of its squared magnitude in conjunction with SVE thresholding. We thus conclude that the objectives of the study have been met since we have managed to: 1) formulate efficient and reliable RF feature selection, 2) maximize the use of Gabor filter banks in retinal image segmentation, and 3) improve the efficiency and effectiveness of Valley Emphasis automatic thresholding.

7.3 Possible Extensions

One of the most significant challenges in medical image processing is overcoming computational overhead, although several methods for improving efficiency have been proposed in this thesis, future work could focus on the use of parallel computing methods like Graphics Processing Units (GPUs) which this study did not have access to. Future work could also incorporate techniques such as deep learning which were not the focus of this study. Although a feasibility study for a Diabetic Retinopathy (DR) screen programme was conducted in South Africa [5], more work in this area could assist in the creation of retinal image datasets from the African continent. The existence of a greater diversity of datasets could pave the way for studies into genetic factors associated with visual retinal properties. Other possible extension could include the comparison of our proposed feature selection strategy against other existing methods. Further experiments could also be conducted to establish the effect of our proposed Gabor filter enhancement techniques on various classification techniques.

Bibliography

[1] David A. Antonetti, Alistair J. Barber, Sarah K. Bronson, Willard M. Freeman, Thomas W. Gardner, Leonard S. Jefferson, Mark Kester, Scot R. Kimball, J. Kyle Krady, Kathryn F. LaNoue, et al. Diabetic retinopathy seeing beyond glucose-induced microvascular disease. *Diabetes*, 55(9):2401–2411, 2006.

[2] Donald Fong, Lloyd P. Aiello, Frederick L. Ferris, and Ronald Klein. Diabetic retinopathy. *Diabetes Care*, 27(10):2540–2553, 2004.

[3] Joanna M. Tarr, Kirti Kaul, Mohit Chopra, Eva M. Kohner, and Rakesh Chibber. Pathophysiology of diabetic retinopathy. *International Scholarly Research Notices Ophthalmology*, 2013:1–13, 2013. doi: https://doi.org/10.1155/2013/343560.

[4] Zhuo Zhang, Ruchir Srivastava, Huiying Liu, Xiangyu Chen, Lixin Duan, Damon Wing Kee Wong, Chee Keong Kwoh, Tien Yin Wong, and Jiang Liu. A survey on computer aided diagnosis for ocular diseases. *BMC Medical Informatics and Decision Making*, 14(1):80, 2014.

[5] Taskeen Khan, Melanie Y. Bertram, Ruxana Jina, Bob Mash, Naomi Levitt, and Karen Hofman. Preventing diabetes blindness: Cost effectiveness of a screening programme using digital non-mydriatic fundus photography for diabetic retinopathy in a primary health care setting in South Africa. *Diabetes Research and Clinical Practice*, 101(2):170 – 176, 2013. ISSN 0168-8227.

[6] Nathan Silberman, Kristy Ahrlich, Rob Fergus, and Lakshminarayanan Subramanian. Case for automated detection of diabetic retinopathy. In *Association for the Advancement of Artificial Intelligence Spring Symposium: Artificial Intelligence for Development*, 2010.

[7] Dermot Maher and Nathan Ford. Action on noncommunicable diseases: balancing priorities for prevention and care. *Bulletin of the World Health Organization*, 89 (8):547–547, 2011.

[8] Gabrielle Kelly and GroundUp staff. Everything you need to know about social grants, March 2015. URL http://groundup.org.za/article/everything-you-need-know-about-social-grants_820. Accessed: 2018-07-15.

[9] National Treasury and South African Revenue Service. Budget 2018 people's guide, February 2018. URL http://www.treasury.gov.za/documents/national%20budget/2018/guides/2018%20Peoples%20Guide%20English.pdf. Accessed: 2018-07-15.

[10] Wilhelm Kirch, editor. *Encyclopedia of Public Health*. Springer Netherlands, 2008. ISBN 978-1-4020-5613-0. doi: 10.1007/978-1-4020-5614-7_3818. URL http://dx.doi.org/10.1007/978-1-4020-5614-7_3818.

[11] T. Teng, M. Lefley, and D. Claremont. Progress towards automated diabetic ocular screening: A review of image analysis and intelligent systems for diabetic retinopathy. *Medical and Biological Engineering and Computing*, 40(1):2–13, Jan 2002. ISSN 1741-0444. doi: 10.1007/BF02347689. URL https://doi.org/10.1007/BF02347689.

[12] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, and S.A. Barman. Blood vessel segmentation methodologies in retinal images – a survey. *Computer Methods and Programs in Biomedicine*, 108(1):407 – 433, October 2012.

[13] Kuryati Kipli, Mohammed Enamul Hoque, Lik Thai Lim, Muhammad Hamdi Mahmood, Siti Kudnie Sahari, Rohana Sapawi, Nordiana Rajaee, and Annie Joseph. A review on the extraction of quantitative retinal microvascular image feature. *Computational and Mathematical Methods in Medicine*, 2018, 2018.

[14] Niall Patton, Tariq M. Aslam, Thomas MacGillivray, Ian J. Deary, Baljean Dhillon, Robert H. Eikelboom, Kanagasingam Yogesan, and Ian J. Constable. Retinal image analysis: concepts, applications and potential. *Progress in Retinal and Eye Research*, 25(1):99–127, 2006.

[15] R. John Winder, Philip J. Morrow, Ian N. McRitchie, J.R. Bailie, and Patricia M. Hart. Algorithms for digital image processing in diabetic retinopathy. *Computerized Medical Imaging and Graphics*, 33(8):608–622, 2009.

[16] Kirsty C. Jordan, Matteo Menolotto, Nigel M. Bolster, Iain A.T. Livingstone, and Mario E. Giardini. A review of feature-based retinal image analysis. *Expert Review of Ophthalmology*, 12(3):207–220, 2017.

[17] M. Akram and Shoab Khan. Multilayered thresholding-based blood vessel segmentation for screening of diabetic retinopathy. *Engineering with Computers*, 29: 1–9, 04 2013. doi: 10.1007/s00366-011-0253-7.

[18] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[19] Robin Genuer, Jean-Michel Poggi, and Christine Tuleau-Malot. Variable selection using random forests. *Pattern Recognition Letters*, 31(14):2225–2236, 2010.

[20] Itzhak Fogel and Dov Sagi. Gabor filters as texture discriminator. *Biological Cybernetics*, 61(2):103–113, 1989.

[21] Hui-Fuang Ng. Automatic thresholding for defect detection. *Pattern Recognition Letters*, 27(14):1644–1649, 2006.

[22] Giuseppe Di Leo, Gabriella Fabbrocini, Alfredo Paolillo, Orsola Rescigno, and Paolo Sommella. Towards an automatic diagnosis system for skin lesions: estimation of blue-whitish veil and regression structures. In *6th International Multi-Conference on Systems, Signals and Devices. SSD'09.*, pages 1–6. IEEE, 2009.

[23] Abou Zaid Sayed Abou Zaid, Mohamed Waleed Fakhr, and Ahmed Farag Ali Mohamed. Automatic diagnosis of liver diseases from ultrasound images. In *The 2006 International Conference on Computer Engineering and Systems*, pages 313–319. IEEE, 2006.

[24] Laura Igual, Joan Carles Soliva, Sergio Escalera, Roger Gimeno, Oscar Vilarroya, and Petia Radeva. Automatic brain caudate nuclei segmentation and classification in diagnostic of attention-deficit/hyperactivity disorder. *Computerized Medical Imaging and Graphics*, 36(8):591–600, 2012.

[25] Luis Alberto Carvalho. Preliminary results of neural networks and zernike polynomials for classification of videokeratography maps. *Optometry & Vision Science*, 82(2):151–158, 2005.

[26] Huiqi Li, Joo Hwee Lim, Jiang Liu, Damon Wing Kee Wong, Ngan Meng Tan, Shijian Lu, Zhuo Zhang, and Tien Yin Wong. An automatic diagnosis system of nuclear cataract using slit-lamp images. In *Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, 2009. EMBC 2009.*, pages 3693–3696. IEEE, 2009.

[27] David Freeman. What your eyes say about your health, August 2011. URL http://www.webmd.com/what-your-eyes-say-about-your-health. Accessed: 2018-07-15.

[28] Geoffrey Ernest Richard Lloyd. *Early greek science: Thales to Aristotle*. Random House, 2012.

[29] Rosanna Gorini. Al-haytham the man of experience. first steps in the science of vision. *Journal of the International Society for the History of Islamic Medicine*, 2 (4):53–55, 2003.

[30] Harminder S. Dua, Lana A. Faraj, Dalia G. Said, Trevor Gray, and James Lowe. Human corneal anatomy redefined: A novel pre-descemet's layer (Dua's layer). *Ophthalmology*, 120(9):1778 – 1785, 2013. ISSN 0161-6420. doi: http://dx.doi.org/10.1016/j.ophtha.2013.01.018. URL http://www.sciencedirect.com/science/article/pii/S0161642013000201.

[31] Clyde W. Oyster. The human eye. *Ophthalmic and Physiological Optics*, 1999.

[32] Visual system (sensory system) part 1. URL http://what-when-how.com/neuroscience/visual-system-sensory-system-part-1/. Accessed: 2018-07-15.

[33] David A. Atchison and George Smith. *Optics of the human eye*. Butterworth-Heinemann Oxford, 1st edition, February 2000. ISBN 9780702038099.

[34] Thomas C. Litzinger and Katia Del Rio-Tsonis. Eye anatomy. *Encyclopedia of Life Sciences*, 2002.

[35] Fundus photo. URL http://www.obecano.com/fundus-photo. Accessed: 2018-07-15.

[36] Helga Kolb. Simple anatomy of the retina. *Webvision: The Organization of the Retina and Visual System*, 1995.

[37] Intelligent medical imaging. URL http://dicom.i2r.a-star.edu.sg/iMed/. Accessed: 2018-07-15.

[38] P. Massin, A. Erginay, A. Ben Mehidi, E. Vicaut, G. Quentel, Z. Victor, M. Marre, P.J. Guillausseau, and A. Gaudric. Evaluation of a new non-mydriatic digital camera for detection of diabetic retinopathy. *Diabetic Medicine*, 20(8):635–641, 2003.

[39] Beau Bruce. Non-mydriatic fundus photography, March 2015. URL http://content.lib.utah.edu/cdm/ref/collection/ehsl-nam/id/8248. Accessed: 2018-07-15.

[40] Robi N. Maamari, Jeremy D. Keenan, Daniel A. Fletcher, and Todd P. Margolis. A mobile phone-based retinal camera for portable wide field imaging. *British Journal of Ophthalmology*, 98(4):438–441, 2014.

[41] Nishtha Panwar, Philemon Huang, Jiaying Lee, Pearse A. Keane, Tjin Swee Chuan, Ashutosh Richhariya, Stephen Teoh, Tock Han Lim, and Rupesh Agrawal. Fundus photography in the 21st century-A review of recent technological advances and their implications for worldwide healthcare. *Telemedicine and E-Health*, 2015.

[42] Alan W. Stitt, Noemi Lois, Reinhold J. Medina, Peter Adamson, and Timothy M. Curtis. Advances in our understanding of diabetic retinopathy. *Clinical Science*, 125(1):1–17, 2013.

[43] American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*, 28:S37, 2005.

[44] Ronald Klein. Hyperglycemia and microvascular and macrovascular disease in diabetes. *Diabetes Care*, 18(2):258–268, 1995.

[45] David A. Antonetti, Erich Lieth, Alistair J. Barber, and Thomas W. Gardner. Molecular mechanisms of vascular permeability in diabetic retinopathy, 1999.

[46] Alan W. Stitt. The role of advanced glycation in the pathogenesis of diabetic retinopathy. *Experimental and Molecular Pathology*, 75(1):95–108, 2003.

[47] Brooks M. Hybertson, Bifeng Gao, Swapan K. Bose, and Joe M. McCord. Oxidative stress in health and disease: the therapeutic potential of Nrf2 activation. *Molecular Aspects of Medicine*, 32(4):234–246, 2011.

[48] Lobo C.L., Bernardes R.C., and Cunha-Vaz J.G. Alterations of the blood-retinal barrier and retinal thickness in preclinical retinopathy in subjects with type 2 diabetes. *Archives of Ophthalmology*, 118(10):1364–1369, 2000. doi: 10.1001/archopht.118.10.1364. URL http://dx.doi.org/10.1001/archopht.118.10.1364.

[49] Jessie Szalay. Inflammation: Causes,symptoms & anti-inflammatory diet, September 2015. URL http://www.livescience.com/52344-inflammation.html. Accessed: 2018-07-15.

[50] Jun-Ming Zhang and Jianxiong An. Cytokines, inflammation and pain. *International Anesthesiology Clinics*, 45(2):27, 2007.

[51] Maximilian Zeyda and Thomas M. Stulnig. Obesity, inflammation, and insulin resistance – a mini-review. *Gerontology*, 55(4):379–386, 2009.

[52] Ramzi S. Cotran, Vinay Kumar, Tucker Collins, and Stanley Leonard Robbins. *Pathologic basis of disease*. WB Sunders Company, 1999. ISBN 978-0721673356.

[53] V.C. Greenstein, A. Shapiro, Q. Zaidi, and D.C. Hood. Psychophysical evidence for post-receptoral sensitivity loss in diabetics. *Investigative Ophthalmology & Visual Science*, 33(10):2781–2790, 1992.

[54] Tamás T Várkonyi, Tünde Petõ, Rózsa Dégi, Katalin Keresztes, Csaba Lengyel, Márta Janáky, Péter Kempler, and János Lonovics. Impairment of visual evoked potentials: An early central manifestation of diabetic neuropathy? *Diabetes Care*, 25(9):1661–1662, 2002.

[55] Thitiporn Chanwimaluang. *Advanced retinal imaging: feature extraction, 2-D registration, and 3-D reconstruction*. PhD thesis, Oklahoma State University, 2006.

[56] Donald Venes. *Taber's cyclopedic medical dictionary*. FA Davis, 2013.

[57] J. Reimer Wolter. Diabetic capillary microaneurysms of the retina. *Archives of Ophthalmology*, 65(6):847–854, 1961.

[58] Norman Ashton. Studies of the retinal capillaries in relation to diabetic and other retinopathies. *The British Journal of Ophthalmology*, 47(9):521, 1963.

[59] J.G. Arroyo. Cotton-wool spots may challenge diagnosis. *Review of Ophthalmology*, 11:4, 2004.

[60] V. Esmann, K. Lundbaek, and P.H. Madsen. Types of exudates in diabetic retinopathy. *Acta Medica Scandinavica*, 174(3):375–384, 1963.

[61] Neville N. Osborne, Robert J. Casson, John P.M. Wood, Glyn Chidlow, Mark Graham, and José Melena. Retinal ischemia: mechanisms of damage and potential therapeutic strategies. *Progress in Retinal and Eye Research*, 23(1):91–147, 2004.

[62] Toke Bek. Venous loops and reduplications in diabetic retinopathy prevalence, distribution, and pattern of development. *Acta Ophthalmologica Scandinavica*, 77 (2):130–134, 1999.

[63] Lihteh Wu, Priscilla Fernandez-Loaiza, Johanna Sauma, Erick Hernandez-Bogantes, and Marissé Masis. Classification of diabetic retinopathy and diabetic macular edema. *World Journal of Diabetes*, 4(6):290, 2013.

[64] Mousami Paver. Better eyes and better health. URL http://npoptometry.com/better-eyes-and-better-health/. Accessed: 2018-07-15.

[65] J. Haut, J.Y. Redor, E. Abboud, G. van Effenterre, and F. Moulin. Classification of diabetic retinopathy. *Ophthalmologica*, 195(3):145–155, 1987.

[66] Morton F. Goldberg and Lee M. Jampol. Knowledge of diabetic retinopathy before and 18 years after the airlie house symposium on treatment of diabetic retinopathy. *Ophthalmology*, 94(7):741–746, 1987.

[67] Early Treatment Diabetic Retinopathy Study Research Group et al. Treatment techniques and clinical guidelines for photocoagulation of diabetic macular edema: Early treatment diabetic retinopathy study report number 2. *Ophthalmology*, 94 (7):761–774, 1987.

[68] Early Treatment Diabetic Retinopathy Study Research Group et al. Grading diabetic retinopathy from stereoscopic color fundus photographs-an extension of the modified airlie house classification: Etdrs report number 10. *Ophthalmology*, 98(5):786–806, 1991.

[69] C.P. Wilkinson, Frederick L. Ferris, Ronald E. Klein, Paul P. Lee, Carl David Agardh, Matthew Davis, Diana Dills, Anselm Kampik, R. Pararajasegaram, Juan T. Verdaguer, et al. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*, 110(9):1677–1682, 2003.

[70] A. Joussen, N. Smyth, and C. Niessen. Pathophysiology of diabetic macular edema. In *Diabetic Retinopathy*, volume 39, pages 1–12. Karger Publishers, 2007. ISBN 978-3-8055-8243-8.

[71] Paul G. Arrigg and Jerry Cavallerano. The role of vitrectomy for diabetic retinopathy. *Journal of the American Optometric Association*, 69(11):733–740, 1998.

[72] Rafael Simó and Cristina Hernandez. Advances in the medical treatment of diabetic retinopathy. *Diabetes Care*, 32(8):1556–1562, 2009.

[73] Michael D. Abramoff and Meindert Niemeijer. Detecting retinal pathology automatically with special emphasis on diabetic retinopathy. In Herbert Jelinek and Michael J. Cree, editors, *Automated Image Detection of Retinal Pathology*, pages 67–77. CRC Press, 2009. ISBN 978-0-8493-7556-9.

[74] M. Matsui, T. Tashiro, K. Matsumoto, and S. Yamamoto. [A study on automatic and quantitative diagnosis of fundus photographs. i. detection of contour line of retinal blood vessel images on color fundus photographs (author's transl)]. *Nippon Ganka Gakkai Zasshi*, 77(8):907–918, 1973.

[75] S. Yamamoto and H. Yokouchi. Automatic recognition of color fundus photographs. In Jr. Preston, K. and M. Onoe, editors, *Digital processing of biomedical images*, pages 385–398. Springer US, 1976. ISBN 978-1-4684-0771-6. doi: 10.1007/

978-1-4684-0769-3_24. URL http://dx.doi.org/10.1007/978-1-4684-0769-3_24.

[76] Koichiro Akita and Hideki Kuga. A computer method of understanding ocular fundus images. *Pattern Recognition*, 15(6):431–443, 1982.

[77] Bruno Lay, Claude Baudoin, and Jean-Claude Klein. Automatic detection of microaneurysms in retinopathy fluoro-angiogram. In *27th Annual Technical Symposium*, pages 165–173. International Society for Optics and Photonics, 1984.

[78] Eli Peli and Moshe Lahav. Drusen measurement from fundus photographs using computer image analysis. *Ophthalmology*, 93(12):1575–1580, 1986.

[79] Norman Katz, Michael Goldbaum, Mark Nelson, and Subhasis Chaudhuri. An image processing system for automatic retina diagnosis. In *1988 Los Angeles Symposium–OE/LASE'88*, pages 131–137. International Society for Optics and Photonics, 1988.

[80] Michael H. Goldbaum, Norm P. Katz, S. Chaudhuri, and Mark Nelson. Image understanding for automated retinal diagnosis. In *Proceedings of the annual symposium on computer application in medical care*, pages 756–760. American Medical Informatics Association, 1989.

[81] Nicholas P. Ward, Stephen Tomliivson, and Christopher J. Taylor. Image analysis of fundus photographs: the detection and measurement of exudates associated with diabetic retinopathy. *Ophthalmology*, 96(1):80–86, 1989.

[82] Peter F. Sharp, J. Olson, F. Strachan, J. Hipwell, A. Ludbrook, M. O'donnell, S. Wallace, K. Goatman, A. Grant, N. Waugh, et al. *The value of digital imaging in diabetic retinopathy.* NCCHTA, 2003.

[83] Mo Yulong and Xiao Dingru. Recognizing the glaucoma from ocular fundus image by image analysts. In *Proceedings of the Twelfth Annual International Conference of the IEEE, Engineering in Medicine and Biology Society, 1990.*, pages 178–179. IEEE, 1990.

[84] Simon Lee and Michael Brady. Integrating stereo and photometric stereo to monitor the development of glaucoma. *Image and Vision Computing*, 9(1):39–44, 1991.

[85] Larry D. Hubbard, Rosemary J. Brothers, William N. King, Limin X. Clegg, Ronald Klein, Lawton S. Cooper, A. Richey Sharrett, Matthew D. Davis, Jianwen Cai, Atherosclerosis Risk in Communities Study Group, et al. Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the atherosclerosis risk in communities study. *Ophthalmology*, 106(12): 2269–2280, 1999.

[86] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum. Automatic detection of the optic nerve in retinal images. In *Proceedings of the IEEE International Conference on Image Processing*, volume 1, pages 1–5, 1989.

[87] Michael H. Goldbaum, N.P. Katz, M.R. Nelson, and L.R. Haff. The discrimination of similarly colored objects in computer images of the ocular fundus. *Investigative Ophthalmology & Visual Science*, 31(4):617–623, 1990.

[88] Adam Hoover and Michael Goldbaum. Fuzzy convergence. In *Proceedings of 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*, pages 716–721. IEEE, 1998.

[89] Chanjira Sinthanayothin, James F. Boyce, Helen L. Cook, and Thomas H. Williamson. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *British Journal of Ophthalmology*, 83 (8):902–910, 1999.

[90] Marc Lalonde, Langis Gagnon, and Marie-Carole Boucher. Non-recursive paired tracking for vessel extraction from retinal images. In *Vision Interface*, pages 61–68, 2000.

[91] James Lowell, Andrew Hunter, David Steel, Ansu Basu, Robert Ryder, Eric Fletcher, and Lee Kennedy. Optic nerve head segmentation. *IEEE Transactions on Medical Imaging*, 23(2):256–264, 2004.

[92] F. Mendels, C. Heneghan, P. Harper, R. Reilly, and J. Thiran. Extraction of the optic disk boundary in digital fundus images. *Proceedings Biomedical Engineering Society/Engineering in Medicine and Biology Society 1999*, 2(EPFL-CONF-86620):1139, 1999.

[93] François Mendels, Conor Heneghan, and J. Thiran. Identification of the optic disk boundary in retinal images using active contours. In *Proceedings of Irish Machine Vision and Image Processing Conference (IMVIP) 1999*, number EPFL-CONF-86621 in Cerebrovascular Diseases, pages 103–115. IEEE, 1999.

[94] Alireza Osareh, Majid Mirmehdi, Barry Thomas, and Richard Markham. Comparison of colour spaces for optic disc localisation in retinal images. In *Proceedings of 16th International Conference on Pattern Recognition, 2002.*, volume 1, pages 743–746. IEEE, 2002.

[95] Chenyang Xu and Jerry L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998.

[96] Adam Hoover, Valentina Kouznetsova, and Michael Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210, 2000.

[97] Subhasis Chaudhuri, Shankar Chatterjee, Norman Katz, Mark Nelson, and Michael Goldbaum. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transactions on Medical Imaging*, 8(3):263–269, 1989.

[98] Andrew Hunter, J. Lowell, D. Steel, A. Basu, and R. Ryder. Non-linear filtering for vascular segmentation and detection of venous beading. Technical report, University of Durham, 2002.

[99] Elisa Ricci and Renzo Perfetti. Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Transactions on Medical Imaging*, 26(10):1357–1365, 2007.

[100] Joes Staal, Michael D. Abràmoff, Meindert Niemeijer, Max A. Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004.

[101] A. Osareh and B. Shadgar. Automatic blood vessel segmentation in color images of retina. *Iranian Journal of Science and Technology*, 33(B2):191–206, 2009.

[102] Mingqiang Yang, Kidiyo Kpalma, and Joseph Ronsin. A survey of shape feature extraction techniques. *Pattern Recognition*, pages 43–90, 2008.

[103] Rafael C. Gonzales, Richard E. Woods, and Steven L. Eddins. *Digital image processing using MATLAB*. Pearson Prentice Hall, 2004.

[104] Frederic Zana and Jean-Claude Klein. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Transactions on Image Processing*, 10(7):1010–1019, 2001.

[105] Conor Heneghan, John Flynn, Michael O'Keefe, and Mark Cahill. Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. *Medical Image Analysis*, 6(4):407–429, 2002.

[106] Ana Maria Mendonca and Aurelio Campilho. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Transactions on Medical Imaging*, 25(9):1200–1213, 2006.

[107] Yong Yang, Shuying Huang, and Nini Rao. An automatic hybrid method for retinal blood vessel extraction. *International Journal of Applied Mathematics and Computer Science*, 18(3):399–407, 2008.

[108] Xiaohong Gao, Anil Bharath, Alice Stanton, Alun Hughes, Neil Chapman, and Simon Thom. A method of vessel tracking for vessel diameter measurement on retinal images. In *Proceedings of 2001 International conference on Image processing*, volume 2, pages 881–884. IEEE, 2001.

[109] A.V. Stanton, B. Wasan, A. Cerutti, S. Ford, R. Marsh, P.P. Sever, S.A. Thom, and A.D. Hughes. Vascular network changes in the retina with age and hypertension. *Journal of Hypertension*, 13(12 Pt 2):1724, 1995.

[110] Francis K.H. Quek and Cemil Kirbas. Vessel extraction in medical images by wave-propagation and traceback. *IEEE Transactions on Medical Imaging*, 20(2): 117–131, 2001.

[111] Hong Shen, Badrinath Roysam, Charles V. Stewart, James N. Turner, and Howard L. Tanenbaum. Optimal scheduling of tracing computations for real-time vascular landmark extraction from retinal fundus images. *IEEE Transactions on Information Technology in Biomedicine*, 5(1):77–91, 2001.

[112] Kelvin Poon, Ghassan Hamarneh, and Rafeef Abugharbieh. Live-vessel: Extending livewire for simultaneous extraction of optimal medial and boundary paths in vascular images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pages 444–451. Springer, 2007.

[113] Peter Dayan. Unsupervised learning. *The Massachusetts Institute of Technology Encyclopedia of the Cognitive Sciences*, 1999.

[114] Saurabh A. Shah and N.C. Chauhan. Techniques for detection and analysis of tumours from brain mri images: A review. *Journal of Biomedical Engineering and Medical Imaging*, 3(1):09, 2016.

[115] Meindert Niemeijer, Joes Staal, Bram van Ginneken, Marco Loog, and Michael D Abramoff. Comparative study of retinal vessel segmentation methods on a new publicly available database. In *Medical Imaging 2004*, pages 648–656. International Society for Optics and Photonics, 2004.

[116] Joao V.B. Soares, Jorge J.G. Leandro, Roberto M. Cesar Jr, Herbert F. Jelinek, and Michael J. Cree. Retinal vessel segmentation using the 2-d Gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging*, 25(9):1214–1222, 2006.

[117] Oliver Faust, Rajendra Acharya, Eddie Yin-Kwee Ng, Kwan-Hoong Ng, and Jasjit S. Suri. Algorithms for the automated detection of diabetic retinopathy using digital fundus images: a review. *Journal of Medical Systems*, 36(1):145–157, 2012.

[118] Jun-Ichiro Hayashi, Takamitsu Kunieda, Joshua Cole, Ryusuke Soga, Yuji Hatanaka, Miao Lu, Takeshi Hara, and Hiroshi Fujita. A development of computer-aided diagnosis system using fundus images. In *Proceedings of Seventh International Conference on Virtual Systems and Multimedia, 2001.*, pages 429–438. IEEE, 2001.

[119] M. Foracchia, E. Grisan, and A. Ruggeri. Extraction and quantitative description of vessel features in hypertensive retinopathy fundus images. In *Book Abstracts 2nd International Workshop on Computer Assisted Fundus Image Analysis*, volume 6, 2001.

[120] Jeffrey Wigdahl, Pedro Guimarães, Georgios Leontidis, Areti Triantafyllou, and Alfredo Ruggeri. Automatic gunn and salus sign quantification in retinal images. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 5251–5254. IEEE, 2015.

[121] Deepika Vallabha, Ramprasath Dorairaj, Kamesh Namuduri, and Hilary Thompson. Automated detection and classification of vascular abnormalities in diabetic retinopathy. In *Thirty-Eighth Asilomar Conference on Signals, Systems and Computers.*, volume 2, pages 1625–1629. IEEE, 2004.

[122] Huan Wang, Wynne Hsu, Kheng Guan Goh, and Mong Li Lee. An effective approach to detect lesions in color retinal images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.*, volume 2, pages 181–186. IEEE, 2000.

[123] G.G. Gardner, D. Keating, T.H. Williamson, and A.T. Elliott. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *British Journal of Ophthalmology*, 80(11):940–944, 1996.

[124] Andrew Hunter, James Lowell, Jonathan Owens, Lee Kennedy, and David Steele. Quantification of diabetic retinopathy using neural networks and sensitivity analysis. In *Artificial Neural Networks in Medicine and Biology*, pages 81–86. Springer, 2000.

[125] Alireza Osareh, Majid Mirmehdi, Barry Thomas, and Richard Markham. Comparative exudate classification using support vector machines and neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 413–420. Springer, 2002.

[126] Rui Bernardes, Sandrina Nunes, Ivânia Pereira, Teresa Torrent, Andreia Rosa, Dalila Coelho, and José Cunha-Vaz. Computer-assisted microaneurysm turnover in the early stages of diabetic retinopathy. *Ophthalmologica*, 223(5):284–291, 2009.

[127] Herbert J. Jelinek, Michael J. Cree, David Worsley, Alan Luckie, and Peter Nixon. An automated microaneurysm detector as a tool for identification of diabetic retinopathy in rural optometric practice. *Clinical and Experimental Optometry*, 89(5):299–305, 2006.

[128] Lee Streeter and Michael J. Cree. Microaneurysm detection in colour fundus images. *Image and Vision Computing New Zealand*, pages 280–284, 2003.

[129] Michael Larsen, Jannik Godt, Nicolai Larsen, Henrik Lund-Andersen, Anne Katrin Sjølie, Elisabet Agardh, Helle Kalm, Michael Grunkin, and David R. Owens. Automated detection of fundus photographic red lesions in diabetic retinopathy. *Investigative Ophthalmology & Visual Science*, 44(2):761–766, 2003.

[130] U.R. Acharya, C.M. Lim, E.Y.K. Ng, C. Chee, and T. Tamura. Computer-based detection of diabetes retinopathy stages using digital fundus images. *Journal of Engineering in Medicine*, 223(5):545–553, 2009.

[131] Chanjira Sinthanayothin, J.F. Boyce, T.H. Williamson, H.L. Cook, E. Mensah, S. Lal, and D. Usher. Automated detection of diabetic retinopathy on digital fundus images. *Diabetic Medicine*, 19(2):105–112, 2002.

[132] Apichart Singalavanija, Jirayuth Supokavej, Parapan Bamroongsuk, Chanjira Sinthanayothin, Suthee Phoojaruenchanachai, and Viravud Kongbunkiat. Feasibility study on computer-aided screening for diabetic retinopathy. *Japanese Journal of Ophthalmology*, 50(4):361–366, 2006.

[133] M. Usman Akram, Shehzad Khalid, and Shoab A. Khan. Identification and classification of microaneurysms for early detection of diabetic retinopathy. *Pattern Recognition*, 46(1):107–116, 2013.

[134] Muhammad Nadeem Ashraf, Zulfiqar Habib, and Muhammad Hussain. Texture feature analysis of digital fundus images for early detection of diabetic retinopathy. In *11th International Conference on Computer Graphics, Imaging and Visualization (CGIV).*, pages 57–62. IEEE, 2014.

[135] Lama Seoud, Thomas Hurtut, Jihed Chelbi, Farida Cheriet, and J.M. Pierre Langlois. Red lesion detection using dynamic shape features for diabetic retinopathy screening. *IEEE Transactions on Medical Imaging*, 35(4):1116–1126, 2016.

[136] Javeria Amin, Muhammad Sharif, and Mussarat Yasmin. A review on recent developments for detection of diabetic retinopathy. In press, 2016. URL http://downloads.hindawi.com/journals/scientifica/aip/6838976.pdf.

[137] Ramon Pires, Herbert F. Jelinek, Jacques Wainer, Eduardo Valle, and Anderson Rocha. Advancing bag-of-visual-words representations for lesion classification in retinal images. *PloS one*, 9(6):e96814, 2014.

[138] Pavel Vostatek, Ela Claridge, Hannu Uusitalo, Markku Hauta-Kasari, Pauli Fält, and Lasse Lensu. Performance comparison of publicly available retinal blood vessel segmentation methods. *Computerized Medical Imaging and Graphics*, 55:2–12, 2017.

[139] Luiz Câmara Neto, Geraldo L.B. Ramalho, Jeová F.S. Rocha Neto, Rodrigo M.S. Veras, and Fátima N.S. Medeiros. An unsupervised coarse-to-fine algorithm for blood vessel segmentation in fundus images. *Expert Systems with Applications*, 78: 182–192, 2017.

[140] Hugo Aguirre-Ramos, Juan Gabriel Avina-Cervantes, Ivan Cruz-Aceves, José Ruiz-Pinales, and Sergio Ledesma. Blood vessel segmentation in retinal fundus images using Gabor filters, fractional derivatives, and expectation maximization. *Applied Mathematics and Computation*, 339:568–587, 2018.

[141] Paweł Liskowski and Krzysztof Krawiec. Segmenting retinal blood vessels with deep neural networks. *IEEE Transactions on Medical Imaging*, 35(11):2369–2380, 2016.

[142] M. Ponni Bala and S. Vijayachitra. Extraction of retinal blood vessels and diagnosis of proliferative diabetic retinopathy using extreme learning machine. *Journal of Medical Imaging and Health Informatics*, 5(2):248–256, 2015.

[143] Christopher G. Owen, Alicja R. Rudnicka, Robert Mullen, Sarah A. Barman, Dorothy Monekosso, Peter H. Whincup, Jeffrey Ng, and Carl Paterson. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program. *Investigative Ophthalmology & Visual Science*, 50(5):2004–2010, 2009.

[144] Muhammad Moazam Fraz, Paolo Remagnino, Andreas Hoppe, Bunyarit Uyyanonvara, Alicja R. Rudnicka, Christopher G. Owen, and Sarah A. Barman. Ensemble classification system applied for retinal vessel segmentation on child images containing various vessel profiles. In *International Conference Image Analysis and Recognition*, pages 380–389. Springer, 2012.

[145] Tomi Kauppi, Valentina Kalesnykiene, Joni-Kristian Kamarainen, Lasse Lensu, Iiris Sorri, Hannu Uusitalo, Heikki Kälviäinen, and Juhani Pietilä. Diaretdb0:

Evaluation database and methodology for diabetic retinopathy algorithms. Technical report, Machine Vision and Pattern Recognition Research Group, Lappeenranta University of Technology, Finland, 2006.

[146] R.V.J.P.H. Kälviäinen and H. Uusitalo. Diaretdb1 diabetic retinopathy database and evaluation protocol. In *Medical Image Understanding and Analysis*, volume 2007, page 61. Citeseer, 2007.

[147] Etienne Decencière, Xiwei Zhang, Guy Cazuguel, Bruno Lay, Béatrice Cochener, Caroline Trone, Philippe Gain, Richard Ordonez, Pascale Massin, Ali Erginay, et al. Feedback on a publicly distributed image database: the Messidor database. *Image Analysis & Stereology*, 33(3):231–234, 2014.

[148] Cong Sun, Jie Jin Wang, David A. Mackey, and Tien Y. Wong. Retinal vascular caliber: systemic, environmental, and genetic associations. *Survey of Ophthalmology*, 54(1):74–95, 2009.

[149] Erkang Cheng, Liang Du, Yi Wu, Ying J. Zhu, Vasileios Megalooikonomou, and Haibin Ling. Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features. *Machine Vision and Applications*, 25(7):1779–1792, 2014.

[150] Shuangling Wang, Yilong Yin, Guibao Cao, Benzheng Wei, Yuanjie Zheng, and Gongping Yang. Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing*, 149:708–717, 2015.

[151] Manoranjan Dash and Huan Liu. Feature selection for classification. *Intelligent Data Analysis*, 1(3):131–156, 1997.

[152] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar):1157–1182, 2003.

[153] Hideko Kawakubo and Hiroaki Yoshida. Rapid feature selection based on random forests for high-dimensional data. *Expert Systems with Applications*, 40:6241–6252, 2012.

[154] Thomas Navin Lal, Olivier Chapelle, Jason Weston, and André Elisseeff. Embedded methods. In *Feature extraction*, pages 137–165. Springer, 2006.

[155] Gilles Louppe, Louis Wehenkel, Antonio Sutera, and Pierre Geurts. Understanding variable importances in forests of randomized trees. In *Advances in Neural Information Processing Systems*, pages 431–439, 2013.

[156] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R News*, 2(3):18–22, 2002.

[157] Robert E. Schapire, Yoav Freund, Peter Bartlett, and Wee Sun Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. *Annals of Statistics*, pages 1651–1686, 1998.

[158] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[159] Marko Robnik-Šikonja. Improving random forests. In *European Conference on Machine Learning*, pages 359–370. Springer, 2004.

[160] Chunyu Hu, Yiqiang Chen, Lisha Hu, and Xiaohui Peng. A novel random forests based class incremental learning method for activity recognition. *Pattern Recognition*, 78:277–290, 2018.

[161] Yisen Wang and Shu-Tao Xia. Unifying attribute splitting criteria of decision trees by tsallis entropy. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*., pages 2507–2511. IEEE, 2017.

[162] Carolin Strobl, Anne-Laure Boulesteix, Achim Zeileis, and Torsten Hothorn. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BioMed Central Bioinformatics*, 8(1):25, 2007.

[163] Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[164] Mandlenkosi Victor Gwetu, Serestina Viriri, and Jules-Raymond Tapamo. Purity and out of bag confidence metrics for random forest weighting. In *International Conference on Computational Collective Intelligence*, pages 491–502. Springer, 2018.

[165] Thanh-Tung Nguyen, Joshua Zhexue Huang, and Thuy Thi Nguyen. Unbiased feature selection in learning random forests for high-dimensional data. *The Scientific World Journal*, 2015, 2015.

[166] C.L. Newman, Blake D.J., and C.J. Merz. UCI repository of machine learning databases, 1998. URL http://www.ics.uci.edu/$\sim$mlearn/MLRepository.html. Accessed: 2018-07-15.

[167] Dennis Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

[168] John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, 2(7):1160–1169, 1985.

[169] Ville Kyrki, Joni-Kristian Kamarainen, and Heikki Kälviäinen. Simple Gabor feature space for invariant object recognition. *Pattern Recognition Letters*, 25(3): 311–318, 2004. ISSN 0167-8655. doi: http://dx.doi.org/10.1016/j.patrec.2003.10.008.

[170] Qin Li, Jane You, Lei Zhang, and Prabir Bhattacharya. A multiscale approach to retinal vessel segmentation using Gabor filters and scale multiplication. In *IEEE International Conference on Systems, Man and Cybernetics, 2006. SMC'06.*, volume 4, pages 3521–3527. IEEE, 2006.

[171] D. Wu, Ming Zhang, Jyh-Charn Liu, and W. Bauman. On the adaptive detection of blood vessels in retinal images. *IEEE Transactions on Biomedical Engineering*, 53(2):341–343, 2006. ISSN 0018-9294. doi: 10.1109/TBME.2005.862571.

[172] R.M. Rangayyan, F. Oloumi, F. Oloumi, P. Eshghzadeh-Zanjani, and F.J. Ayres. Detection of blood vessels in the retina using Gabor filters. In *Canadian Conference on Electrical and Computer Engineering, 2007. CCECE 2007.*, pages 717–720, 2007. doi: 10.1109/CCECE.2007.184.

[173] A. Bhuiyan, B. Nath, J. Chua, and R. Kotagiri. Blood vessel segmentation from color retinal images using unsupervised texture classification. In *IEEE International Conference on Image Processing, 2007. ICIP 2007.*, volume 5, pages V – 521–V – 524, 2007. doi: 10.1109/ICIP.2007.4379880.

[174] G. Azzopardi and N. Petkov. Trainable cosfire filters for keypoint detection and pattern recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):490–503, 2013. ISSN 0162-8828. doi: 10.1109/TPAMI.2012.106.

[175] Jesper Juul Henriksen. 3d surface tracking and approximation using Gabor filters. Technical report, South Denmark University (March 28, 2007), 2007.

[176] N. Petkov and E. Subramanian. Motion detection, noise reduction, texture suppression and contour enhancement by spatiotemporal Gabor filters with surround inhibition. *Biological Cybernetics*, 97(5–6):423–439, 2007. doi: 10.1007/s00422-007-0182-0.

[177] J.G. Daugman. Complete discrete 2-d Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36(7):1169–1179, 1988. ISSN 0096-3518. doi: 10.1109/29.1644.

[178] Mandlenkosi Victor Gwetu, Jules Raymond Tapamo, and Serestina Viriri. Segmentation of retinal blood vessels using normalized Gabor filters and automatic thresholding. *South African Computer Journal*, 55(1):12–24, 2014.

[179] Mohammed Al-Rawi, Munib Qutaishat, and Mohammed Arrar. An improved matched filter for blood vessel detection of digital retinal images. *Computers in Biology and Medicine*, 37(2):262 – 267, 2007. ISSN 0010-4825. doi: http://dx.doi.org/10.1016/j.compbiomed.2006.03.003.

[180] Zahia Brahimi, K.A. Saadi, and N. Baraka. A lossless image compression algorithm using variable block size segmentation. In *Proceedings of the 24th Annual Conference of the IEEE, Industrial Electronics Society, 1998. IECON'98.*, volume 3, pages 1471–1476. IEEE, 1998.

[181] B. Nikolic. Calculating the area under a two-dimensional Gaussian. URL http://www.mrao.cam.ac.uk/~bn204/_downloads/twodgaussian.pdf. Accessed: 2018-07-15.

[182] Manning Christopher D., Raghavan Prabhakar, Schütze Hinrich, and Mark Sanderson. Introduction to information retrieval. *Natural Language Engineering*, 16(1):100–103, 2010.

[183] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

[184] Hetal J. Vala and Astha Baxi. A review on Otsu image segmentation algorithm. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 2(2):pp–387, 2013.

[185] C.I. Chang, Yingzi Du, J. Wang, S.M. Guo, and P.D. Thouin. Survey and comparative analysis of entropy and relative entropy thresholding techniques, 2006.

[186] Jiu-Lun Fan and Bo Lei. A modified valley-emphasis method for automatic thresholding. *Pattern Recognition Letters*, 33(6):703–708, 2012.

[187] Hui-Fuang Ng, Davaajargal Jargalsaikhan, Hao-Chuan Tsai, and Chih-Yang Lin. An improved method for image thresholding based on the valley-emphasis method. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. APSIPA. 2013*, pages 1–4. IEEE, 2013.

[188] Hui-Fuang Ng, Cheng-Wai Kheng, and Jim-Min Lin. A weighting scheme for improving Otsu method for threshold selection. *Journal of Computers, Taiwan*, 27(2):12–21, Jul 2016. ISSN 1991-1599.

[189] Yi Yin, Mouloud Adel, and Salah Bourennane. Retinal vessel segmentation using a probabilistic tracking method. *Pattern Recognition*, 45(4):1235–1244, 2012.

[190] Bob Zhang, Lin Zhang, Lei Zhang, and Fakhri Karray. Retinal vessel extraction by matched filter with first-order derivative of Gaussian. *Computers in Biology and Medicine*, 40(4):438–445, 2010.