Automated Detection of Proliferative Diabetic Retinopathy from Retinal Images

A dissertation submitted by

Roshan Alex Welikala

In fulfilment of the requirements for the Doctor of Philosophy degree in

Computer Science

Faculty of Science, Engineering and Computing,

Kingston University London, United Kingdom

September 2014

Director of Studies: Dr. Sarah A. Barman

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my supervisor Dr Sarah Barman, you have been a wonderful mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career have been priceless. I would also like to thank the rest of my supervisory team, Prof Jamshid Dehmeshki and Dr Andreas Hoppe for their support and advice.

I am grateful to consultant ophthalmologists Tom Williamson and Samantha Mann from St Thomas' hospital London for providing the retinal image dataset and for providing clinical guidance. I would also like to thank Dr Vikas Tah from Oxford eye hospital for the creation of the ground truths.

I would like to thank the Faculty of Science, Engineering and Computing, Kingston University London for funding this research project.

A special thanks to my mother and father, words cannot express how grateful I am for all of the sacrifices that you've made on my behalf. I would also like to thank my brothers, my friends and my fiancée who incentivised me to strive towards my goal.

ABSTRACT

Diabetic retinopathy (DR) is a retinal vascular disease associated with diabetes and it is one of the most common causes of blindness worldwide. Diabetic patients regularly attend retinal screening in which digital retinal images are captured. These images undergo thorough analysis by trained individuals, which can be a very time consuming and costly task due to the large diabetic population. Therefore, this is a field that would greatly benefit from the introduction of automated detection systems.

This project aims to automatically detect proliferative diabetic retinopathy (PDR), which is the most advanced stage of the disease and poses a high risk of severe visual impairment. The hallmark of PDR is neovascularisation, the growth of abnormal new vessels. Their tortuous, convoluted and obscure appearance can make them difficult to detect. In this thesis, we present a methodology based on the novel approach of creating two different segmented vessel maps. Segmentation methods include a standard line operator approach and a novel modified line operator approach. The former targets the accurate segmentation of new vessels and the latter targets the reduction of false responses to non-vessel edges. Both generated binary vessel maps hold vital information which is processed separately using a dual classification framework. Features are measured from each binary vessel map to produce two separate feature sets. Independent classification is performed for each feature set using a support vector machine (SVM) classifier. The system then combines these individual classification outcomes to produce a final decision. The proposed methodology, using a dataset of 60 images, achieves a sensitivity of 100.00% and a specificity of 92.50% on a per image basis and a sensitivity of 87.93% and a specificity of 94.40% on a per patch basis.

The thesis also presents an investigation into the search for the most suitable features for the classification of PDR. This entails the expansion of the feature vector, followed by feature selection using a genetic algorithm based approach. This provides an improvement in results, which now stand at a sensitivity and specificity

of 100.00% and 97.50% respectively on a per image basis and 91.38% and 96.00% respectively on a per patch basis. A final extension to the project sees the framework of dual classification further explored, by comparing the results of dual SVM classification with dual ensemble classification. The results of the dual ensemble approach are deemed inferior, achieving a sensitivity and specificity of 100.00% and 95.00% respectively on a per image basis and 81.03% and 95.20% respectively on a per patch basis.

PUBLICATIONS EMERGED FROM THE THESIS

Journal publications

- R.A. Welikala, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T.H. Williamson, S.A. Barman, Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification, Computer Methods and Programs in Biomedicine 114 (2014) 247-261.

- R.A. Welikala, M.M. Fraz, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T.H. Williamson, S.A. Barman, Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy, submitted to Computerized Medical Imaging and Graphics, final corrections submitted August 2014.

- R.A. Welikala, M.M. Fraz, T.H. Williamson, S.A. Barman, The automated detection of proliferative diabetic retinopathy using dual ensemble classification, submitted to Computer Methods and Programs in Biomedicine, submitted July 2014.

Conference papers

- R.A. Welikala, V. Tah, T.H. Williamson, A. Hoppe, J. Dehmeshki, S.A. Barman, Differing matched filter responsivity for the detection of proliferative diabetic retinopathy, Proceedings of the IASTED International Conference Signal Processing, Pattern Recognition and Applications (2013) 356-364.

Presentations

- Automated detection of proliferative diabetic retinopathy using dual classification, 1st Student conference of the interdisciplinary hub for the study of health and age-related conditions (2013).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Diabetic Patients are required to attend regular eye screening appointments in which Diabetic retinopathy (DR), a retinal disorder, can be assessed. From these appointments digital retinal images are captured, and these then undergo various stages of analysis (referred to as grading) by trained individuals. This can be a very time consuming and costly task due to the large number of diabetic patients. Therefore this is a field that would greatly benefit from the introduction of automated detection systems [1]. Not only would its implementation be more cost effective, but the screening programme and the National Health Service (NHS) would benefit in numerous other ways. Results would be produced more quickly, thus allowing patients to receive results as soon as possible hence minimising anxiety and also ensuring referrals to the hospital eye service are made according to time targets. Another benefit arises from the fact that human graders are subjective and can also become fatigued, whereas an automated system would provide consistent objective results.

1.1 Aims, Objectives and Contribution to Knowledge

This project involves the assessment and development of image processing techniques to analyse retinal images for the automated detection of diabetic retinopathy. Two years experience the author gained from working for the NHS as a diabetic retinopathy screener/grader has provided knowledge which has been advantageous to this project.

The main goal of DR screening programmes is the early detection of the disease to allow for timely intervention. However, DR is a disease of largely varying severity and patients attending screening may also present the later stages of the disease which need to be differentiated. Therefore it is the responsibility of the trained individuals to correctly label images according to the severity of the disease to ensure the appropriate outcome is achieved.

The literature so far shows that the vast majority of work has been performed on the automated detection of signs of early stage DR. Importantly, this can be applied to perform classification of DR disease/no DR disease (explained in section 1.2.3). The intention is that this can be used to remove all images without DR disease and therefore reduces the work load of images to be analysed by the trained individuals. Settling for automated systems that provide such a simple classification would provide only a small contribution to the analysis of retinal images. The aim is to develop automated systems that can identify the specific stage of the disease and therefore take up even more of the work load. The most important task of these systems will be the identification of the most advanced stage of the disease to ensure fast tracking for immediate attention as this stage poses the highest risk to loss of vision. Unfortunately only a small amount of work has been conducted in this area within the retinal image analysis field.

The main aim of this project will be the automated detection of proliferative diabetic retinopathy (PDR), which is the most advanced stage of the disease and is characterised by the presence of new vessels. These new vessels can be extremely difficult to detect as they are often finer in calibre and are often more obscure, tortuous and convoluted than normal vessels. The terms PDR and new vessels are used synonymously throughout this this report.

The objectives of this project can be summarized as follows:

- To develop and evaluate a methodology for the automated detection of PDR.
- The main goal of the algorithm is the decision of whether new vessels are present or not in the image. Results should meet the requirements for clinical integration, high specificity at maximum (100%) sensitivity.

- To assemble an image dataset for the evaluation of PDR detection. This should contain a varied selection of images.
- To characterise the appearance of new vessels and bring awareness of the importance of their detection. These are aspects/areas that are not very well understood by the retinal image processing community.
- To consider the quantification of new vessel growth as opposed to just the detection of their presence. This would help ophthalmologist to track the disease, further helping in understanding the progression of the disease and the effects of treatment.

The contribution to knowledge achieved from these objectives is:

- A novel algorithm for the automated detection of PDR. This will highlight the use of suitable image processing techniques and classification procedures for the detection of PDR.
- Identification of a specific set of PDR characteristics useful for their detection, in the form of a tailored feature set used by a classifier.

In this project, the main bulk of the methodology is based on the application of two different vessel segmentation approaches and dual classification. A standard line operator and a novel modified line operator were used for vessel segmentation. The latter, based on double sided thresholding, was designed to reduce false responses to non-vessel edges. Both generated binary vessel maps held vital information which were processed separately. This was achieved with a dual classification system, which is novel to retinal image analysis. Local morphology features were measured from each binary vessel map to produce two separate feature sets. Independent classification was performed for each feature set using a support vector machine (SVM) classifier. These individual classification outcomes are then combined by the system to produce a final decision.

This work is further extended by expanding the size of the feature sets, to include further morphology based features as well as intensity and gradient based features. This feature vector is fed into a genetic algorithm based feature selection approach with the objective of finding feature subsets and classifier parameters that improve

the performance of the classification. A final extension to the project sees the classifier changed from the SVM to an ensemble system of bagged decision trees. This brings with it its own technique for feature selection.

The remainder of this chapter will provide an overview of the mechanism of diabetic retinopathy and the screening process. This is followed by a literature review and then three chapters detailing the methodology and evaluation. The thesis ends with a final chapter of an overall discussion and conclusion.

1.2 Pathology and Screening Overview

Diabetes mellitus is a disorder of sugar metabolism caused by an impairment of insulin secretion and/or an insufficient action of insulin and hence is characterized by raised levels of glucose in the blood. High blood glucose levels (hyperglycemia) can damage the vessels that supply blood to vital organs. Diabetic retinopathy is the resultant disorder affecting the retinal vasculature, leading to progressive retinal damage that can end in loss to vision and blindness [2]. DR is recognized as the leading cause of blindness in the working-age population [3-5]. The problem is increasing in its scale, with diabetes having been identified as a significant growing global public health problem [6]; in fact in the United Kingdom three million people are estimated to have diabetes and this figure is expected to double in the next 15-30 years. 171 million people were estimated to have diabetes worldwide in the year 2000 and this figure is expected to rise to 366 million by the year 2030 [7].

If DR is detected early enough, laser treatment can diminish visual loss [8-10]. Also at the incipient stage, intervention such as improvement of the blood glucose control can help reverse or prevent further progression of the DR [11]. Often there are no symptoms in the early stages of DR, therefore screening is imperative for identification. Once DR is identified patients can be managed accordingly with the goal of the prevention of sight loss. In 2000, the introduction of a systematic national screening programme based on digital photography was proposed [12] which meet the Wilson and Jungner criteria for a screening programme [13]. The

programme is now in operation in the United Kingdom with diabetic patients aged 12 and above being invited at least annually for retinal screening using digital photography [14].

The paragraphs above clearly state the impact of DR on the population and the reasons why screening is necessary. Next, further details of the disease and the screening process shall be provided. A brief pathology section that follows will first help to achieve a basic understanding of the structure of the retina, and then a description of the clinical features of DR will be provided. This will be followed by a brief section detailing the tasks of a DR screening programme from the screening appointment to the manual analysis of images to the possible treatments.

1.2.1 Pathology

The retina is a complex multi-layered light sensitive tissue lining the inner surface of the eye. The retina contains millions of photoreceptors (rods and cones) that respond to focused light and converts it into electrical impulses. These impulses travel along the optic nerve to the brain where they are turned into images. The cross section of the human eye is shown in figure 1.1 which illustrates the location of the retina within the eye. The particular details from this figure that appear on retinal images are the retinal blood vessels, optic disc, macula and fovea. The macula occupies a substantial portion of the brain's visual capacity due to its high proportion of cone cells which are responsible for colour, fine visual acuity and central vision. The fovea is the central part of the macula and has the highest proportion of cone cells. A structure important to PDR that is shown in this figure, although is not visible in retinal images, is the vitreous body. This is an area found between the lens and retina which consists of vitreous humour which is a clear solution with a gel-like consistency.

Figure 1.2 shows examples of retinal images from healthy retinas; the main anatomical features are also labelled. This figure also illustrates some variations between the appearances of healthy retinas due to varying pigmentation for

different ethnicities and also the appearance of the tigroid pattern due to a thin retina making vessels from the choroid layer more visible. A final image is included to illustrate areas of reflection associated with young patients.

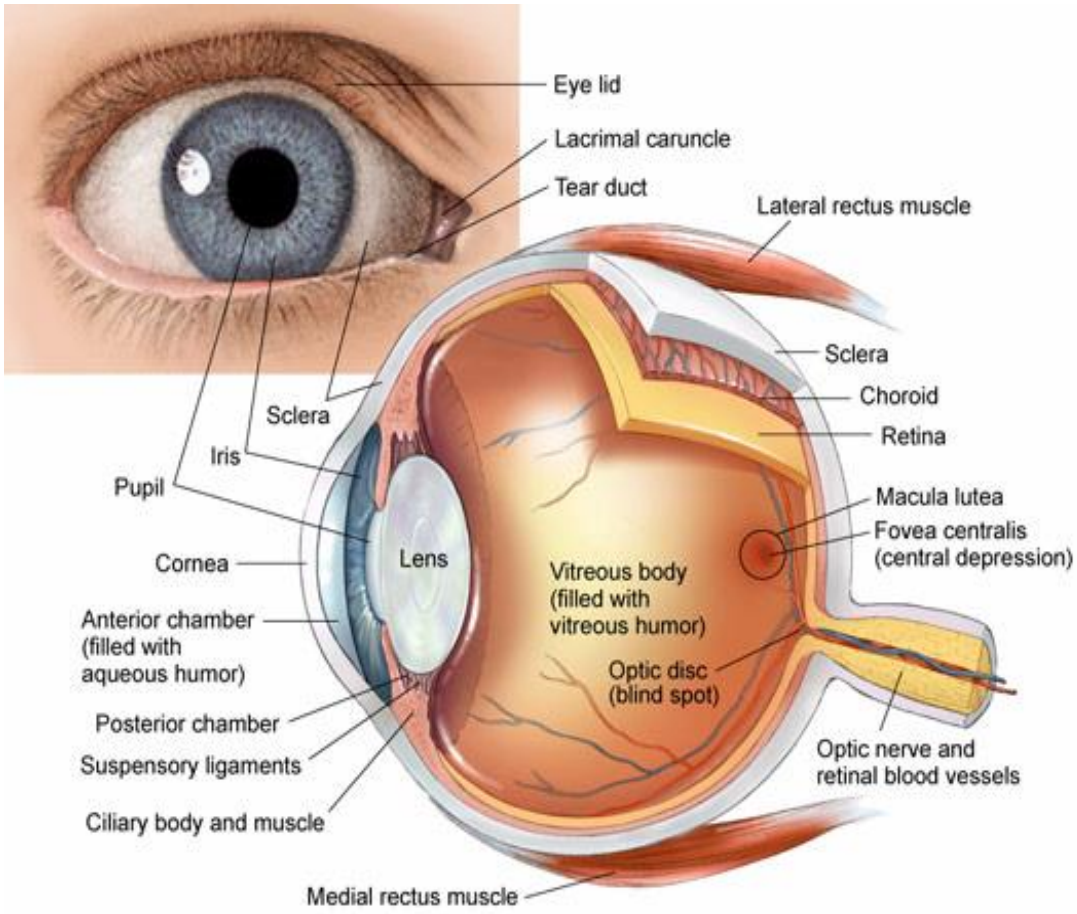


Figure 1.1: Anatomy of the eye. Image from www.biographixmedia.com.

Diabetic retinopathy is a microangiopathy affecting the retinal vasculature caused by hyperglycemia. The damage to the retinal blood vessels will cause blood and fluid to leak on the retina and forms features such as microaneurysms, haemorrhages, exudates, cotton wool spots and venous loops [15]. DR is a progressive disease; with progression the blockages and damage to blood vessels will cause areas of retinal ischaemia to develop. In an attempt of revascularization the growth of new blood

vessels is triggered. New vessels represent the advanced stages of DR, which poses a high risk of severe vision loss due to the new vessels' fragile nature and possible extensive bleeding. Below the significance and appearance of a few of the main features of DR will be described in more detail and corresponding images will be provided.

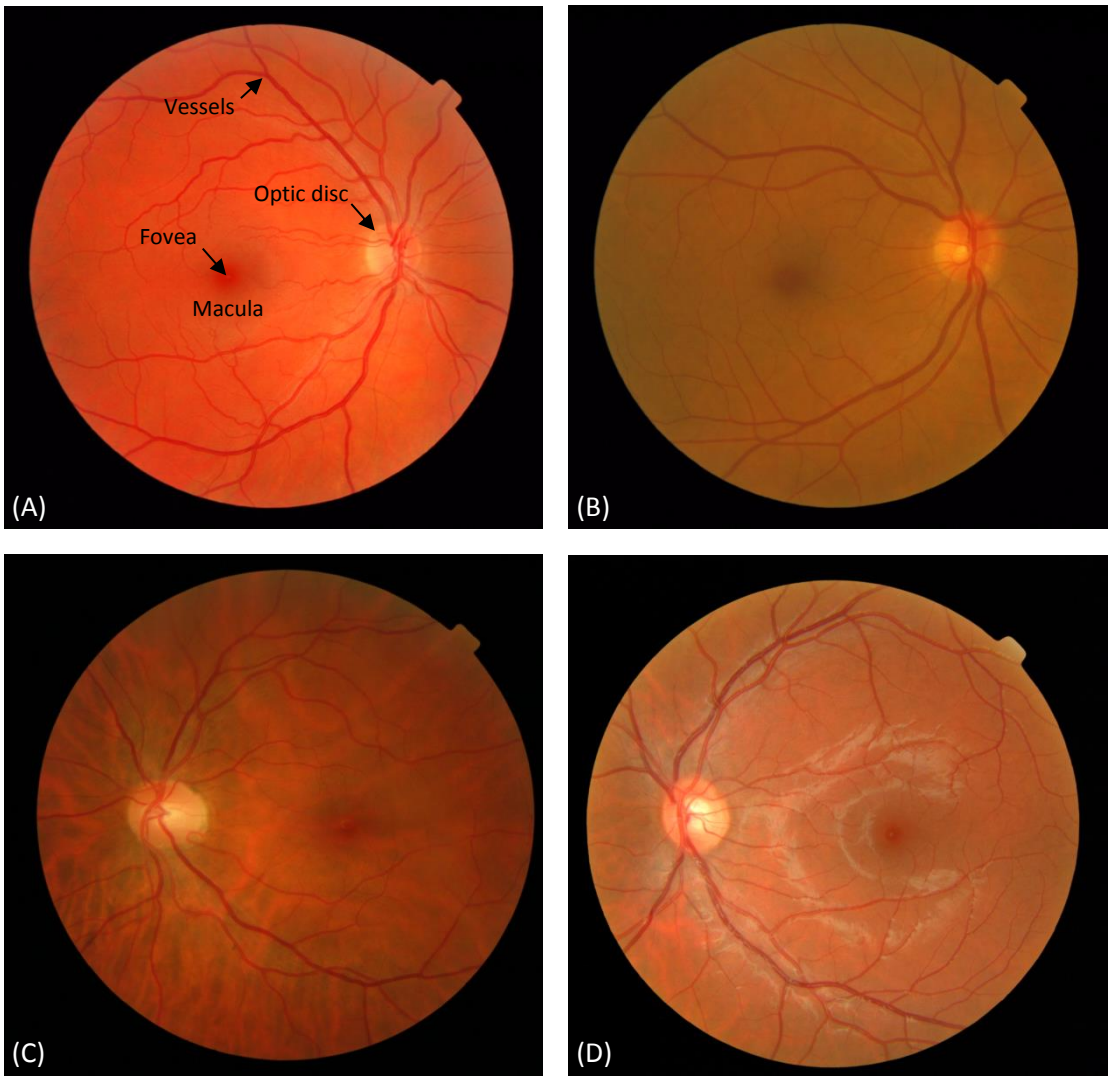


Figure 1.2: Healthy retinas **(A)** Caucasian retina, **(B)** South-Asian retina, **(C)** tigroid retina, **(D)** reflections associated with young patients. Images from [16].

Microaneurysms: These are balloon-like structures on the sides of capillaries which arise due to the weakening of capillary walls. As capillaries are not visible from conventional fundus images, microaneurysms appear like isolated red dots

unattached to any blood vessel. They are often the first signs of DR that can be detected. See figure 1.3(A).

Haemorrhages: Breakdown of the capillary walls results in the leakage of blood, which can take on various forms of size and shape depending on the retinal layer in which the vessels are located. These different forms are referred to as dot, blot or flame haemorrhages. See figure 1.3(B).

Exudates: Capillary breakdown can often result in the leakage of oedema. The build-up of oedema causes retinal thickening. If this build up occurs at the macular it is referred to as macular oedema and it is the most common cause of visual impairment amongst diabetic patients and if clinically significant will require laser treatment to reduce the risk of visual loss [17]. Oedema is a clear fluid, therefore it is not visible on a standard 2D retinal image. Exudates are the lipid residue from the oedema. They appear as waxy yellow lesions and take on various patterns including individual patches, tracking lines, rings (circinates) and macular stars. See figure 1.3(C)-(D). The detection of exudates at the macula acts as a marker for predicting the presence of macular oedema.

New Vessels: With the progression of capillary breakdown, areas of the retina can become ischemic and the growth of new vessels is triggered as an attempt of revascularization of the oxygen deprived tissue. There are two types, new vessels on the optic disc (NVD) and new vessels elsewhere (NVE). See figure 1.4. They are very fragile and can easily rupture resulting in extensive haemorrhaging in the form of pre-retinal and vitreous haemorrhages (see figure 1.5). These types of haemorrhages occur due to the fact that new vessels tend to grow away from the retinal surface towards the vitreous. Fibrous tissue can also develop around new vessels which can cause traction and lead to retinal detachment [2]. New vessels represent a high risk of severe loss of vision and laser treatment may be required to reduce this risk. The appearance of new vessels is notoriously difficult to characterise due to their variation in form. New vessels appear as unregulated growth appearing off or near a vein in the form of lacy networks which usually pass across the underlying veins and arteries. Normally new vessels are restricted to small regions of the retina where

ischemia has developed. If ischemia has become widespread across the image then it is also possible that the new vessels can span this larger region. These tortuous and convoluted vessels are usually fine in calibre, although on rare occasions they can possess a large calibre in which their diameter is bigger than the vessel they originated from. They can also often be extremely difficult to spot as they don't always appear as dense networks, they may appear as small networks or loops when they initially develop. New vessels can sometimes appear very faint and can be difficult to distinguish from the background. As mentioned above, new vessels tend to grow away from the retinal surface. This means they can appear out of the focal plane of the photograph, resulting in a very blurry and obscure appearance.

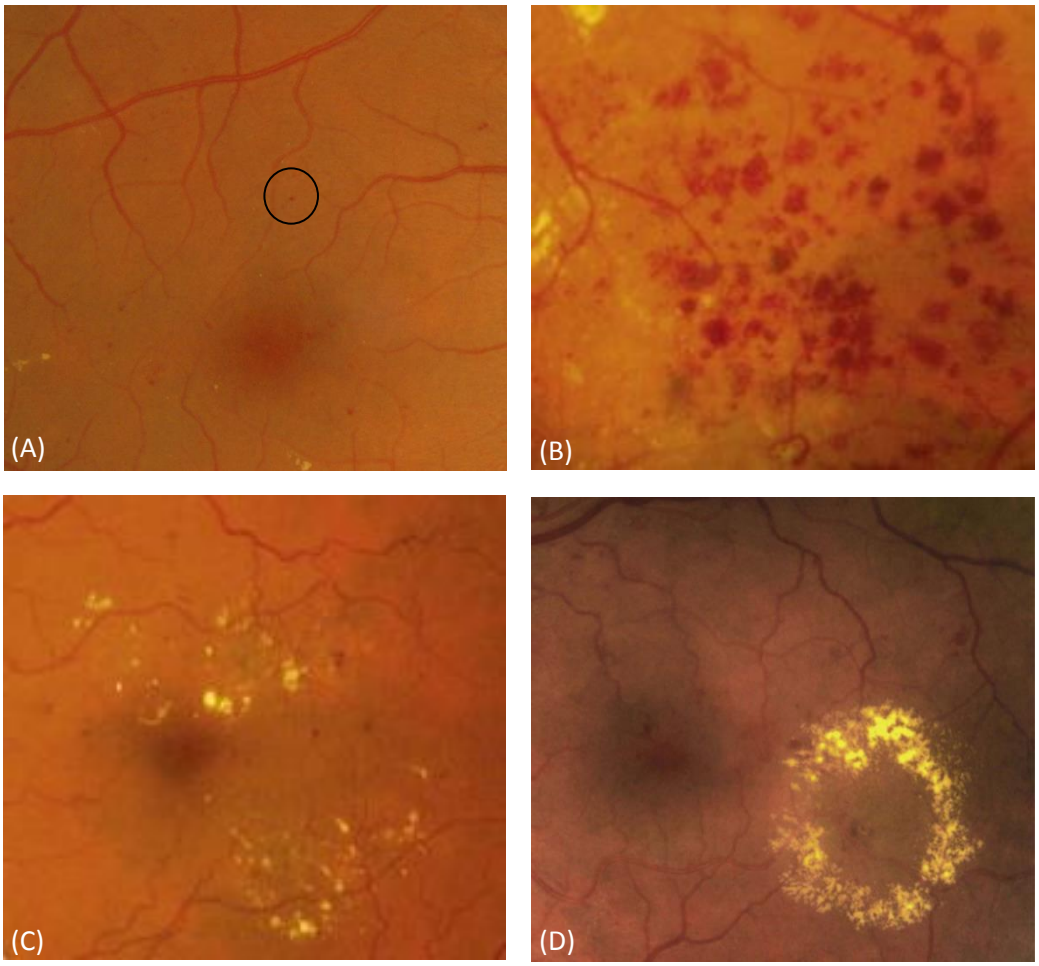


Figure 1.3: (A) Microaneurysms, **(B)** dot and blot haemorrhages, **(C)** exudates, **(D)** a ring of exudates (circinate). Images from [17].

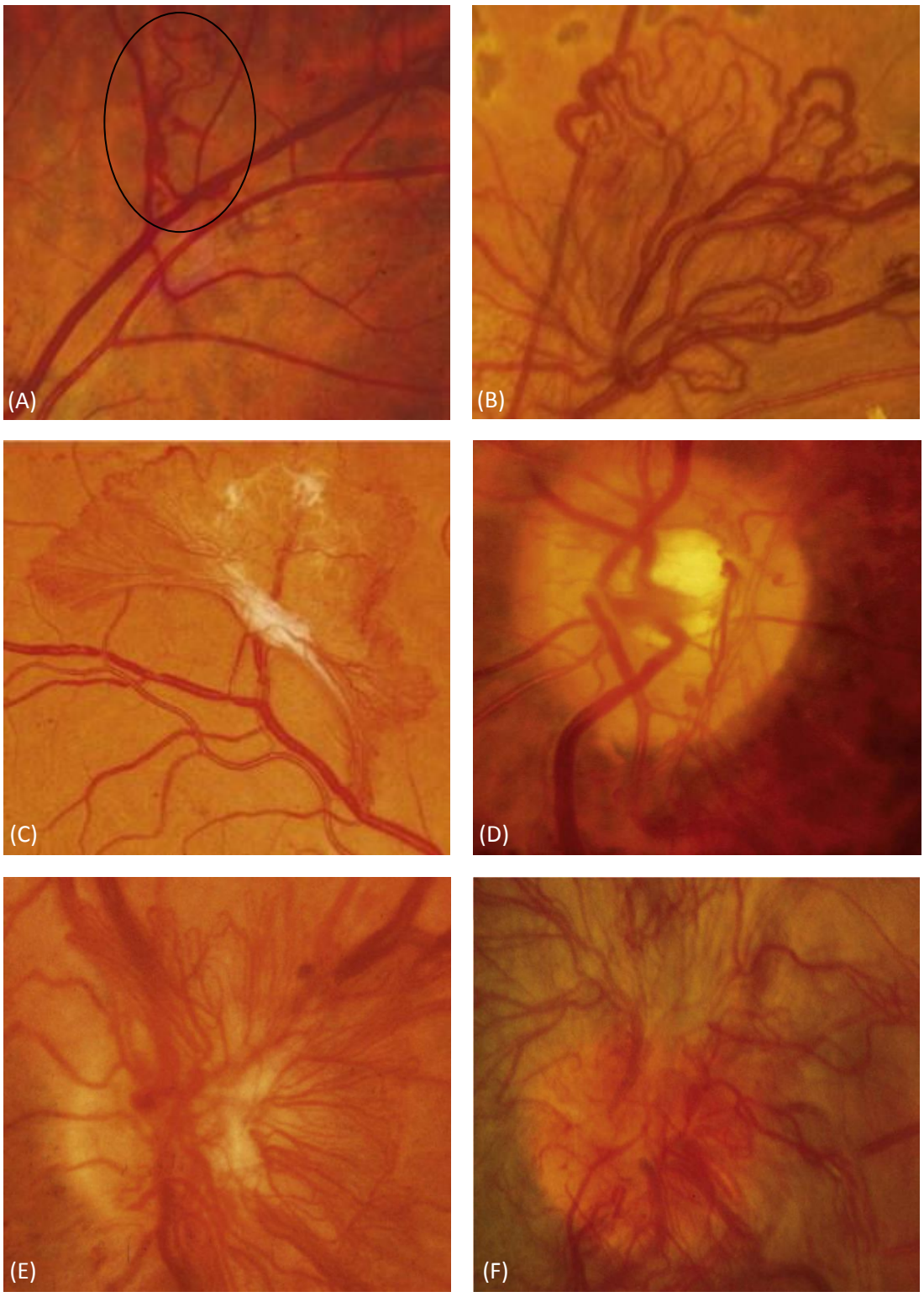


Figure 1.4: (A) Mild NVE, **(B)** severe NVE, **(C)** severe NVE with associated fibrosis, **(D)** mild NVD, **(E)** severe NVD, **(F)** very severe NVD. Images from [17].

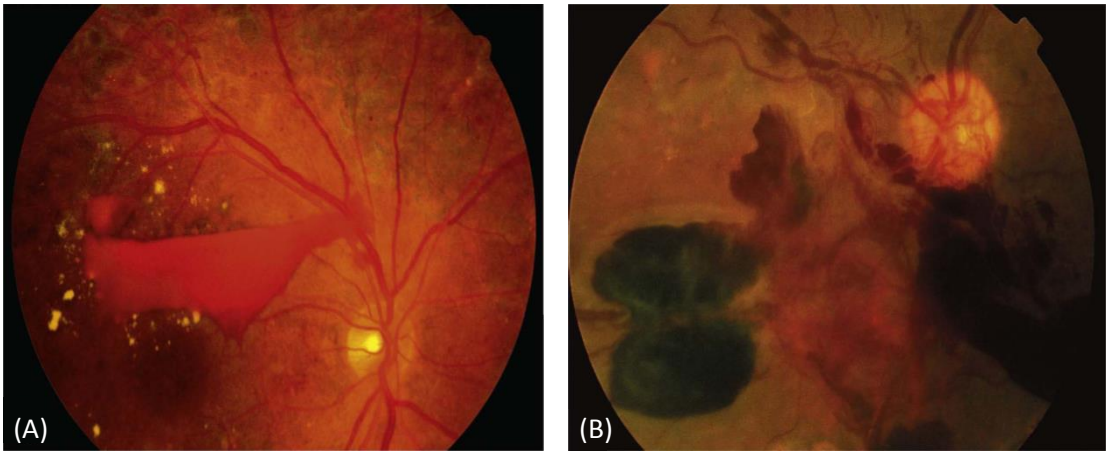


Figure 1.5: (A) Pre-retinal haemorrhage, **(B)** vitreous haemorrhage. Images from [17].

There are many more features of DR and table 1.1 provides an exhaustive list as well as the stage of the disease they present. Putting maculopathy to one side, progression of the disease increases with descent down the table. As this disease is progressive, clearly a stage of DR can contain all the features of the previous stage. Background DR is the earliest stage of DR and is not a threat to vision. Pre-proliferative DR represents progressive retinal ischaemia, with the increased risk of the progression of neovascularisation (new vessels). Proliferative DR is characterised by neovascularisation, it is the most advanced stage of the disease and can pose a high risk of severe loss of vision. It can be seen from table 1.1 that there are features other than new vessels that fall under proliferative DR; however, it should be remembered that these other features occur due to the presence of new vessels. Maculopathy can occur at any stage of DR, although it is more likely to occur as the disease progresses. The strict definition of maculopathy is the existence of any DR features at the macula, but is commonly reserved for vision-threatening macular oedema.

Table 1.1: Classification of diabetic retinopathy.

Stage of retinopathy	Features
Background	Microaneurysms. Dot and blot haemorrhages. Exudates.
Pre-proliferative	Multiple dot and blot haemorrhages. Cotton wool spots (CWS). Venous beading and loops. Intra-retinal microvascular abnormalities (IRMA).
Proliferative	New vessels elsewhere (NVE). New vessels at the disc (NVD). Pre-retinal/vitreous haemorrhage. Retinal detachment.
Maculopathy	Microaneurysms, haemorrhages, exudates at the macula. Macular oedema.

1.2.2 Screening

First a brief recap on the reason for DR screening shall be provided. DR has adverse effects and has been shown to be a major cause of vision loss. The disease responds best to treatment if it is detected at an early stage. However, there are generally no symptoms at an early stage of DR and the disease is often well advanced before symptoms develop. DR screening provides a detailed examination of the retina in order to detect early stage DR. Patients that attend screening can also present later stages of the disease which need to be differentiated and labelled. Once DR is identified patients can be managed accordingly with the goal of the prevention of sight loss. Screening is easily implemented using digital photography, which is an acceptable and cost effective test. Cost is actually a large factor and it has been shown that screening saves vision for a cost less than the disability payments

provided to people who would go blind in the absence of screening [1]. Screening is generally performed on an annual basis. This section will cover some of the main stages that a DR screening programme performs. This knowledge will be helpful when considering how automated detection systems are to be integrated into screening programmes.

Before proceeding, it should be understood that screening programmes are run and managed separately to the ophthalmology department. It is the purpose of the screening programme to monitor the condition of the patient's retinas from the large diabetic population and only refer those that have vision threatening DR to the ophthalmology department for specialist care.

The first stage of the appointment uses the Snellen chart (see figure 1.6(A)) to perform a visual acuity (VA) test which measures the sharpness of the patient's vision. VA measures are very important: for example they can be used to define maculopathy (table 1.2). Next eye drops are instilled in order to achieve pupil dilation, which provides a clearer view of the retina. Finally photographs are taken using a retinal camera, also known as a fundus camera. The fundus refers to the region of the retina opposite to the lens and includes the optic disc, macula and posterior pole. The camera produces a magnified view of the fundus using an optical design based on the principle of ophthalmoscopy. Topcon is an established manufacturer of retinal cameras; figure 1.6(B) shows a typical model. Photographs at two standard views are required per eye, the macular centred view and the optic disc centred view (see figure 1.7). It is the task of the screener (photographer) to ensure that an adequate standard of photographs are captured. This can be achieved by ensuring correct illumination of the retina (avoids dark images), correct lens focusing, and correct positioning within the retina. Also camera artefacts need to be avoided by keeping the lens clean and regularly servicing the camera. Another task of the screener is to briefly assess the images (not grading) and triage them according to severity of DR. This is done by simply putting the images into either a red, amber or green queue awaiting grading (discussed next).

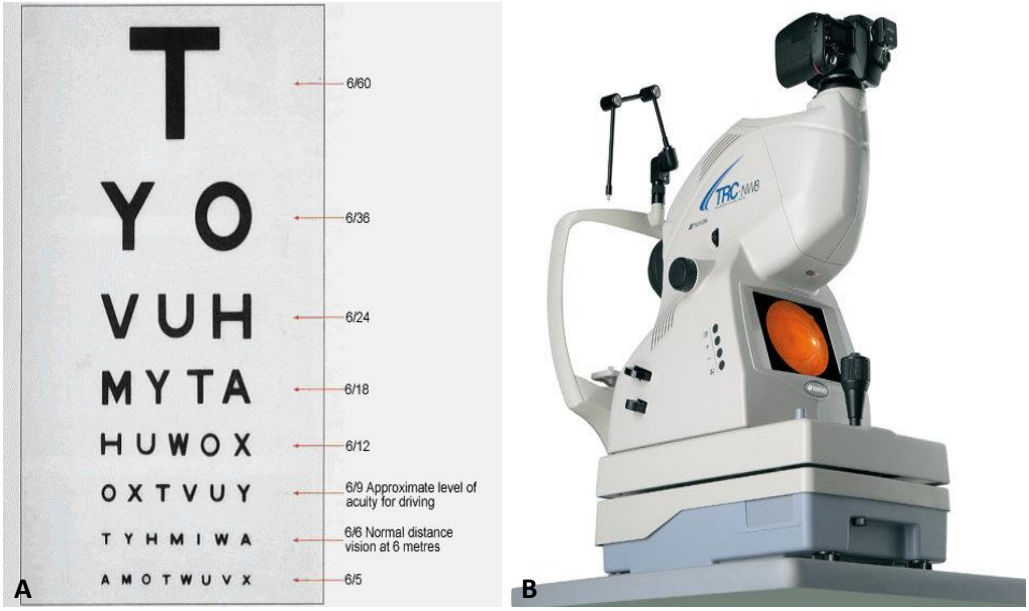


Figure 1.6: (A) Snellen test chart, **(B)** Topcon TRC NW8 retinal camera. Images from dwp.gov.uk and topcon-medical.co.uk.

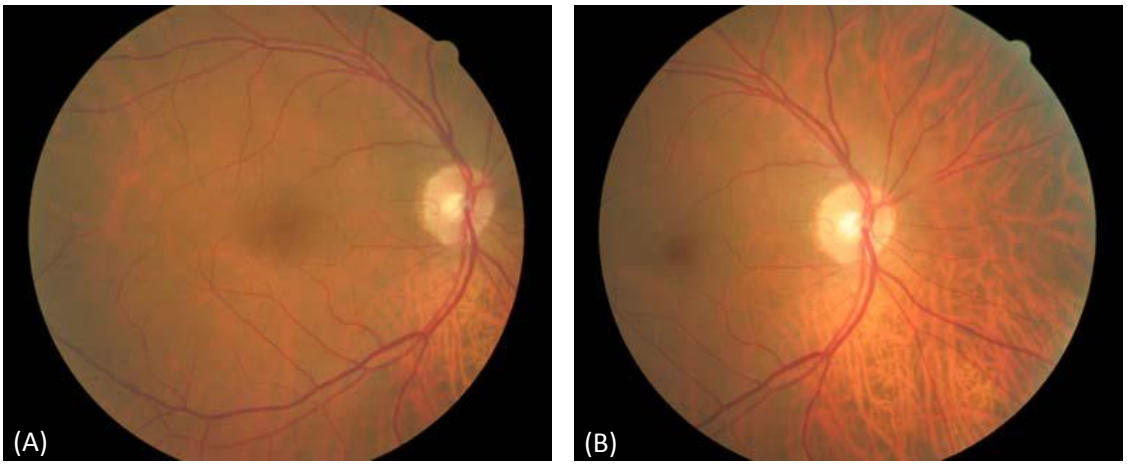


Figure 1.7: (A) Macular centred view, **(B)** optic disc centred view. Images from [2].

Once these retinal images are captured they then undergo thorough analysis in order to classify each image with a grade according to the stage of the disease. This process is referred to as grading and is performed by trained individuals referred to as graders. The UK National Screening Committee (NSC) have created strict grading guidelines (see table 1.2) which state the different features which constitute each stage of DR and each of these stages has been labelled with a grade (R0, M0, R1, M1, R2 and R3). Figure 1.8 shows some examples of retinal images representing these

grades. These grading guidelines are a more detailed version of table 1.1 in order to assist and standardise classification. For each eye both R and M grades are awarded; for example, a left eye's retina which displays pre-proliferative DR and maculopathy would be awarded a grade of R2M1. There are other grades: P is selected when there is evidence of previous laser treatment from the appearance of laser scars and U is selected when no grade is obtainable. There are various reasons why the grade U would be awarded, from poor quality photographs taken by the screener (discussed earlier), media opacities (e.g. cataracts and asteroids hyalosis) preventing a clear view of the retina and patients' inability to comply (physical and mental disabilities). When this occurs, the conventional screening process using a retinal camera is considered inadequate and the patient is booked in for a slit lamp bio-microscopy examination.

Table 1.2: National Screening Committee (NSC) grading standards [14].

Retinopathy (R)	
R0 None	
R1 Background	Microaneurysm(s) retinal haemorrhage(s) ± any exudate not within the definition of maculopathy.
R2 Pre-proliferative	Venous beading. Venous loop or reduplication. Intraretinal microvascular abnormality (IRMA). Multiple deep, round or blot haemorrhages. Cotton wool spots (CWS).
R3 Proliferative	New vessels on disc (NVD). New vessels elsewhere (NVE). Pre-retinal or vitreous haemorrhage. Pre-retinal fibrosis ± tractional retinal detachment.
Maculopathy (M)	
M0 None	
M1 Maculopathy	Exudate within 1 disc diameter (DD) of the centre of the fovea. Circinate or group of exudates within the macula. Retinal thickening within 1DD of the centre of the fovea (if stereo available). Any microaneurysm or haemorrhage within 1DD of the centre of the fovea only if associated with a best VA of ≤ 6/12 (if no stereo).
Photocoagulation (P)	Evidence of focal/grid laser to macula. Evidence of peripheral scatter laser.
Unclassifiable (U)	Unobtainable / ungradeable.

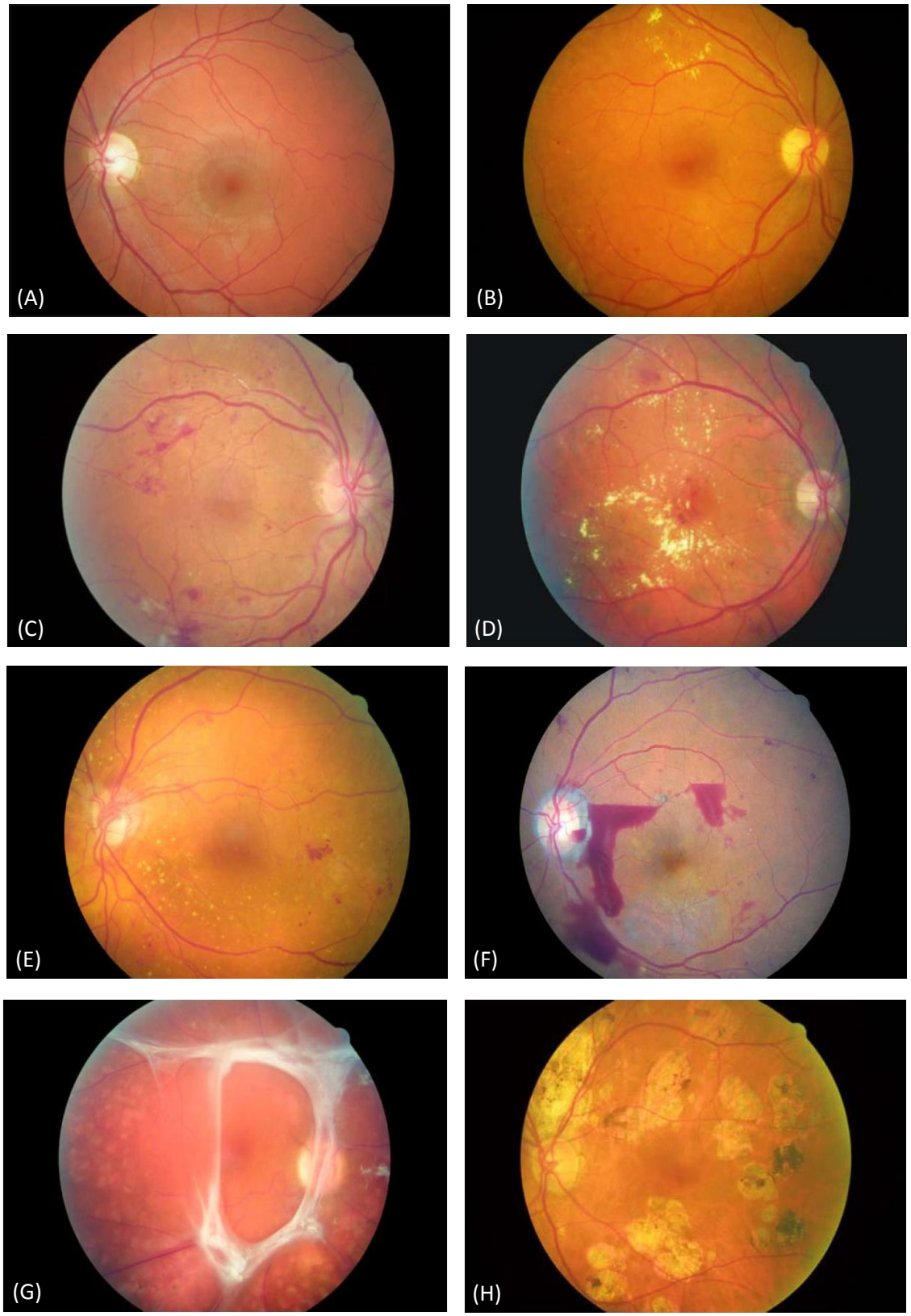


Figure 1.8: Image examples of grades from the National Screening Committee (NSC), [14]. **(A)** Normal retina with no DR - NSC grade R0. **(B)** Background DR with microaneurysms, haemorrhages and exudates - NSC grade R1. **(C)** Pre-proliferative DR with CWS, IRMA and multiple blot haemorrhages - NSC grade R2. **(D)** Diabetic maculopathy with haemorrhages

and circinate/exudates - NSC grade M1. **(E)** Proliferative DR with new vessels at the optic disc (NVD) - NSC grade R3. **(F)** Proliferative DR with pre-retinal and vitreous haemorrhages - NSC grade R3. **(G)** Proliferative DR with fibrous proliferation - NSC grade R3. **(H)** Evidence of previous laser therapy - NSC grade P.

The grades awarded to patients will determine the next course of action. If there is no threat to their vision then the patient will receive their next screening appointment in one year's time (annual recall). If the patient has vision-threatening DR, their grades would indicate that a referral to ophthalmology is required. The grades that require referral also have an associated time target in which they must be seen by ophthalmology. Time targets indicate the urgency of referral which correlates to the severity of the disease and therefore the risk to vision. This information is provided in table 1.3. An example is a patient whose retinal images contain new vessels, gets awarded a grade of R3 (proliferative DR) that signifies a high risk of severe loss of vision and therefore requires an urgent referral to the ophthalmology department with the time target of ideally being seen within 1 week from the date of screening. Time targets are a very important factor, hence grading needs to be incorporated into automated detection systems if they are to be introduced into the clinical environment.

Table 1.3: NSC DR grades and referral time targets [14].

Grade	Outcome
R0 (No DR)	Annual recall
R1 (Background DR)	Annual recall
M1 (Maculopathy)	Routine referral, 70% within 13 weeks and 100% within 18 weeks
R2 (Pre-proliferative DR)	Routine referral, 70% within 13 weeks and 100% within 18 weeks
R3 (Proliferative DR)	Urgent referral, 95% within 1 week, 100% within 2 weeks

Screening programmes are based on the concept of the detection of early stage DR to prevent the loss of vision. From the description provided so far it should be evident

that this doesn't so much apply to the earliest DR stage (R1) but more to detecting the vision threatening stages (M1,R2,R3) early enough to prevent loss of vision. However, the classification of the grade R1 still has huge importance; for example the presence of DR can be used to encourage patients into better control of their blood glucose level which can retard DR and therefore prevent vision-threatening DR from developing.

It is also evident that DR grading is by no means a simple task. The graders are skilled individuals who are able to essentially perform multi-level classification, identifying all different stages of DR. Also when assessing retinal images other retinal conditions may be identified. Therefore, graders are also trained to detect and refer various other non-DR lesions. These include branch retinal vein occlusion (BRVO), central retinal vein occlusion (CRVO), retinal arterial occlusion, emboli, hypertensive retinopathy, macroaneurysms, cataracts, glaucoma, age-related macular degeneration (ARMD) and naevi. Therefore, developing an automated system to mimic the performance of a human grader is a very complex task.

The grading process is carried out under strict regulations. The environment that graders work in is very important to ensure that grading is consistently accurate. This should be a darkened room to prevent glare on the computer screen which may interfere with images. Grading should not be carried out in-between patients in a busy clinic as full attention to the task must be given; therefore grading laboratories are usually assigned and these should be kept as quiet as possible. Graders regularly perform full 8 hour shifts of grading. This type of environment can be difficult to work under (despite regular breaks) and graders can become easily fatigued which could result in a decrease in accuracy. This is an aspect where automated systems show serious potential improvements as they can produce consistent and objective results. Graders possess various image manipulation tools to aid their image analysis. These include magnification, intensity control, contrast control and red-free manipulation. There is more than one stage of grading (discussed later), therefore the agreement between the stages of grading is regularly under review using the inter-grader report to ensure accuracy and consistency. A minimum level of 82.5% agreement should be met, individual graders who don't achieve this level will be

placed under review. Also serious cases of under-grading will be reviewed by the lead clinician. The most serious case is a grader under-grading a R3 image. This could result in a patient either not being referred or missing their referral time target to the ophthalmology department and therefore, as a consequence, the patient could potentially completely lose their vision.

As mentioned above, there is more than one stage of grading to ensure a high standard of accuracy and consistency (essentially quality control). There are two different grading pathways [14] and these are shown in figure 1.9 and figure 1.10. Images are triaged by the screener immediately after capture and therefore enter the grading pathway according to severity. We shall concentrate on pathway 2 as it is the most common employed of the two in the UK [14]. There are two main grading stages of this pathway, first a full disease grade and second a full disease grade performed by primary and secondary graders respectively. The term full disease grade means that the grade should specifically state the stage of DR (using R and M grades) as opposed to just a disease/no disease grade. Upon grading, the primary grader will send those images graded as R0M0 (no disease) straight through to the outcome of annual recall without the need to be sent to the secondary grader; this will remove the main bulk of images. All images with a grade indicating disease are triaged (according to grade) and sent to the secondary grader. The secondary grader performs grading blind to the results from the primary grader. All grade agreements that require referral are triaged and sent to the ophthalmology department. All agreements that do not require referral are marked for annual recall. All disagreements are sent to arbitration (also known as tertiary grading) which is performed by the clinical lead that makes the final decision to decide the outcome; again those requiring referral get triaged. At any stage of grading an image with an R3 grade can be urgently referred directly to ophthalmology without having to go through the whole grading pathway in order to prevent any possible delays. It should be noted that triaging is performed at every stage of grading in order to ensure that time targets are met.

Pathway 1 is exactly the same apart from the additional disease/no disease grading stage at the very start of the pathway where images without disease can be directly

marked as annual recall and therefore require no further grading. Approximately 50-70% of diabetic patients will present no DR; therefore this stage of grading will take a significant amount of the workload away from the primary grading stage. This stage of grading is a much simpler task than the other stages and therefore is performed by individuals without the same knowledge/training and thus will be on a lower pay band. However, these individuals should still have the ability to identify R3 cases and send them directly through to the ophthalmology department.

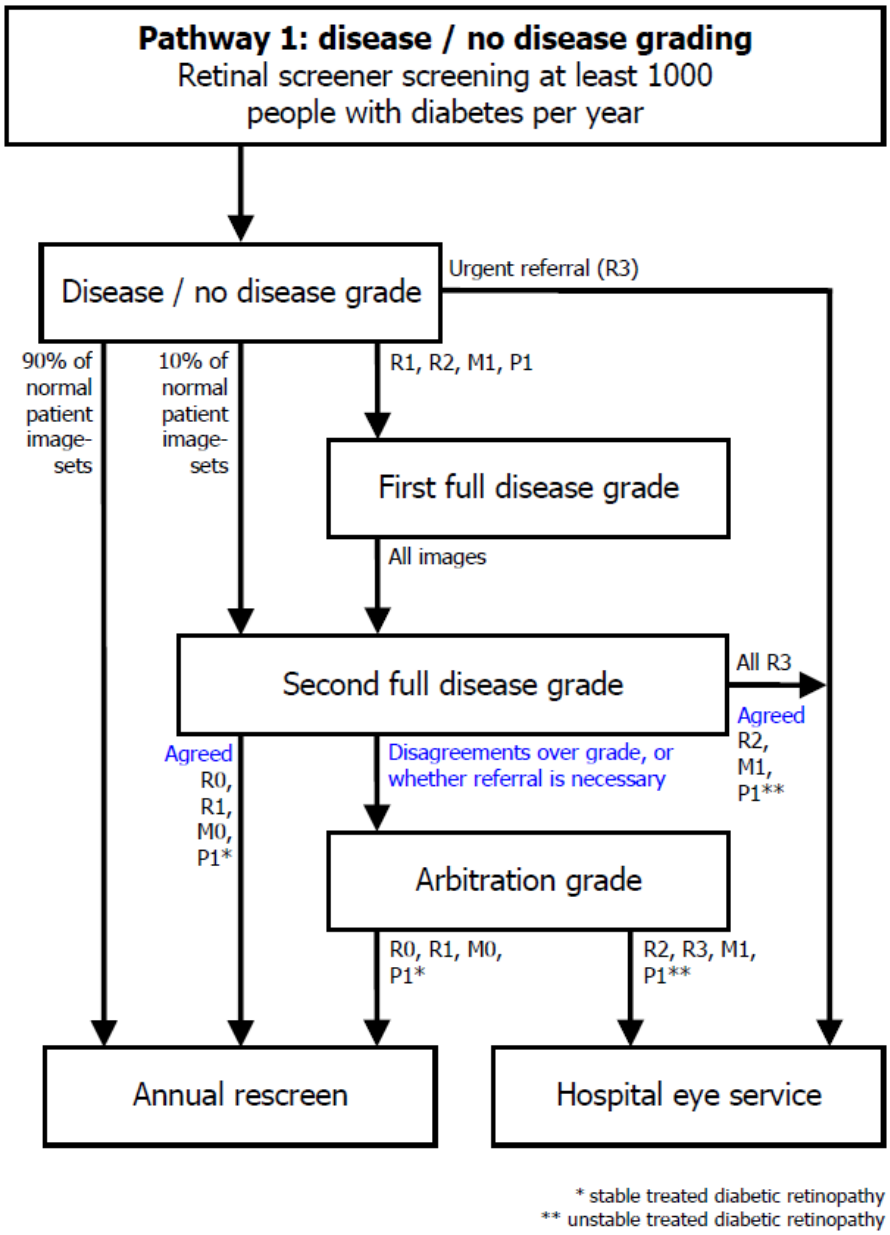


Figure 1.9: NSC grading pathway 1 [14].

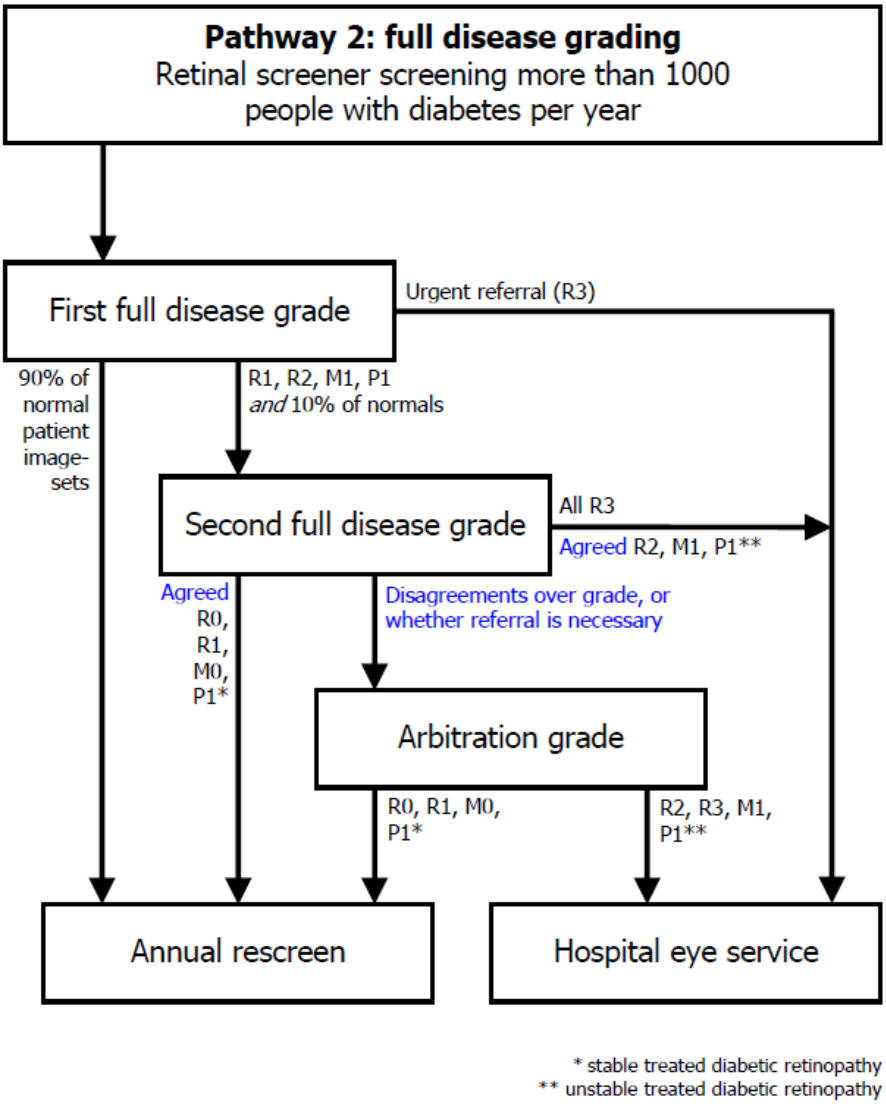


Figure 1.10: NSC grading pathway 2 [14].

Once patients are referred to ophthalmology they remain under their care until their condition becomes stable; at which point they can be referred back to the screening programme. Whilst under the care of ophthalmology it may be a case of simply providing closer monitoring of the patients' retina by specialists. Various follow-up tests may be done to achieve a more detailed assessment such as fluorescein angiography or optical coherence tomography (OCT). Treatment may be required and laser photocoagulation remains the main treatment for diabetic retinopathy. For cases of proliferative DR where vessels are considered at risk of bleeding then pan-retinal (scatter) photocoagulation may be required, which is the application of

hundreds (sometimes thousands) of small laser burns. The aim is to destroy ischaemic areas: ischaemic stimulus is lessened and the new vessels can then regress. If the maculopathy grade has been shown to be caused by clinically significant macular oedema then laser photocoagulation may be required. The aim is to target areas of leakage using focal laser burns or a grid of laser burns for areas of diffuse leakage. This reduces leakage and improves reabsorption of retinal oedema. Laser photocoagulation is unlikely to improve vision once it has decreased but it is effective in stabilizing vision. Laser treatment can have serious side effects, and therefore careful consideration of the situation needs to be taken first. Other treatments for DR include the intra-vitreal injection of drugs and vitrectomy surgery [17].

1.2.3 Automated Detection

So far we have established that human graders can become easily fatigued leading to a decrease in accuracy, although there are many other weaknesses. Despite having strict grading guidelines human graders are subjective, and therefore grades for a specific image can vary considerably amongst different graders. Human graders are relatively slow and aim to grade images for approximately 60-80 patients a day. Backlogs of images are formed at each stage of grading and often it can take several days for an image to be graded and move on to the next grading stage. Incidents such as a human grader taking sick leave can result in these backlogs becoming unacceptable in size. The introduction of automated systems has the benefits of producing consistent and objective results. They will reduce the workload for human graders and therefore provide a reduction in the cost of running the screening service. As automated systems can process images quickly, operate 24 hours a day and run on multiple computers it is likely that backlogs will be minimal for the stages of grading that they perform. This will increase the speed of the grading process, ensuring patient's test results letters are sent out more quickly and most importantly time targets for referrals are met.

It is clear from the pathology section that different DR features have their own distinct characteristics. It is also clear that graders use strict guidelines to categorise

retinal images into the different DR stages. How this knowledge is used to develop automated systems will be discussed in the literature review that follows.

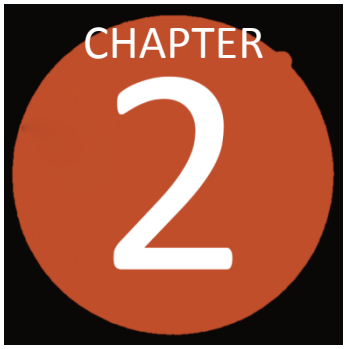
So how are automated systems to be introduced into the grading pathway? The majority of DR automated detection research has gone into the detection of microaneurysms. This provides the detection of early stage DR, but microaneurysms almost always appear to be present in any image containing DR regardless of the severity, and hence these algorithms also provide the classification of DR disease or no DR disease. As mentioned earlier, approximately 50-70% of the images show no signs of DR. Thus the classification system can be used to remove these and therefore significantly reduce the manual grading workload. Large-scale audits of disease/no disease automated grading systems have shown the benefits they provide [18].

Such disease/no disease automated systems would replace the initial stage from grading pathway 1, although it would be lacking in not having the ability to fast-track R3 images. Such an inability would be a substantial problem if this initial stage took considerable time to perform. This is not the case as this automated stage is fairly simple and is likely to process images very quickly; hence no images are likely to get held up in this stage as may be the case if performed by a human grader. Therefore, R3 fast-tracking can instead commence from the first full disease grade stage without any real delay. In simpler terms consider this stage as an almost instant filter removing no DR disease images prior to the first full disease grade and still retaining the initial triaging performed by the screener (photographer).

The aim is for automated systems to have a much more substantial role in the grading pathway than that mentioned above and thus take on more of the workload. The target is to produce a full DR disease automated detection system. Any system is unlikely to be sophisticated enough to take over the complete grading process and therefore human graders are always likely to remain in some context. Also keeping more than one stage of grading is beneficial for quality control. Therefore, a possible role will be to take over the first full disease grade (primary grader) from grading pathway 2. To achieve this, the system must be capable of removing all images

without DR and the greater challenge of detecting and differentiating the stages of DR to allow for triaging to ensure time targets are met. This will include the vital step of the detection of PDR (R3) for fast tracking (urgent referrals).

The grading pathways generally stress the importance of the detection of R3 for fast-tracking due to their high risk of severe vision loss. This supports the efforts of this project into the detection of new vessels (PDR), which would be ideally integrated into a full DR disease automated detection system.



2 LITERATURE REVIEW

The research field of retinal image analysis has attracted a lot of interest in the last couple of decades, with the automated detection of diabetic retinopathy having received a considerable share of this interest. Landmark detection is also an area that has received considerable interest. Landmarks consist of blood vessels, the optic disc and the fovea. This section will start with a brief review of the automated segmentation of blood vessels. Most DR detection methodologies use it as a prerequisite before identifying pathological entities, in particular new vessel detection methods. A brief review of the main methodologies used to detect the main DR features (microaneurysms, haemorrhages and exudates) will be provided. This will be followed with a section providing a detailed account on the detection of new vessels (proliferative DR), which is the main focus of this project. This chapter will end with a brief overview of the machine learning (classification, clustering, etc.) procedures that have been mentioned throughout this literature review.

Most methodologies start by pre-processing the images. The main pre-processing steps are applied to correct for poor illumination and poor contrast. Poor illumination is often tackled with a technique called shade correction [19,20], whereby an image approximating the background is subtracted from the original image. The background image is obtained with the application of a median filter whose size is significantly greater than the largest retinal feature. Poor contrast is frequently tackled with contrast limited adaptive histogram equalisation (CLAHE) [21,22]. This is a technique for local contrast enhancement which is preferred to global contrast enhancement. However, pre-processing can only correct up to a certain extent; therefore, it is the task of the screener (photographer) to ensure an

adequate standard of photographs are captured (detailed in section 1.2.2). To avoid an overly lengthy literature review, pre-processing steps shall be omitted from further discussions, although a comparative evaluation of pre-processing steps for retinal analysis is provided by Youssif [23].

It should be noted that the word “feature” will be used in two different contexts. The word “feature” refers to those components that make up a feature set used for classification. DR features on the other hand refer to pathological features such as microaneurysms and haemorrhages.

2.1 Vessel Segmentation

Some of the main attributes of vessels that are utilised in segmentation methods are their deep red colour, their contrast with the background, and their gradient at vessel edges. They can be approximated as piecewise linear and their cross section intensity profile approximates a Gaussian function. There are numerous methods reported in the literature, which can be generally categorised into the four main groups based on matched filtering, mathematical morphology, vessel tracking and machine learning. Vessel segmentation methods can struggle to detect fine, faint and very tortuous vessels. Vessel crossing and branching can complicate the profile model. Vessels can show strong reflections along their centreline, which can further complicate the profile model. Pathologies can also hamper accuracy, producing false positives. Vessel segmentation methods are assessed on a per pixel basis. Results will not be included in this section. The following sections describe vessel segmentation techniques according to algorithm type.

Matched filtering: Chaudhuri [24] first proposed the use of the matched filter for vessel segmentation and it has become one of the most popular techniques. It made use of the fact that the cross section of the vessels can be approximated by a Gaussian function, in addition to the fact that vessels may be considered as piecewise linear segments. Therefore the method used a two dimensional Gaussian filter; this differed from the standard isotropic Gaussian filter: instead a single Gaussian

function cross profile was repeated a number of times and stacked to make up the length of the filter (this gave the filter an aspect of linearity). The length of the filter was set to a distance in which vessel segments were assumed to approximate linearity. This filter, which resembled the shape of vessel segments, was convolved with the retinal image in order to "match" the blood vessel segments. The filter was also rotated in order to detect vessels of different orientations. This resulted in a vessel enhanced image, known as the matched filter response (MFR). This was followed by a global threshold to produce a binary vessel map. Al-Rawi [25] improved on the performance of the matched filter by using an optimization technique to automatically find the best filter parameters.

Unfortunately, the matched filtering technique responds not only to vessels but also non-vessel edges. The step edges created between exudates and the background cause the most nuisance. A single global threshold applied to the MFR is not sufficient and can result in many false positives. Therefore many modified matched filtering methods have been proposed. Hoover [26] stated a piecewise threshold probing technique using the vessel structural information. The algorithm probed the MFR and during each probe a set of criteria was used to determine the threshold for that region in order to segment vessels. L.Zhang [27] used the fact that the peak point of the MFR for a vessel is considerably greater than its neighbouring points on both sides, whereas for non-vessel edges the MFR peak point is not much greater than its neighbours on both sides. Therefore a double sided thresholding technique was proposed. B.Zhang [28] applied the first order derivative of the Gaussian (FDOG) kernel, which produced a response map to help differentiate vessel from non-vessel. Whilst vessels and non-vessel edges both produced strong responses for the MFR, they differed for the mean of the FDOG response, with vessels being close to zero (at vessel centre) and the non-vessel edges, in contrast, were high. The knowledge gained from the FDOG response was used to accordingly adjust the threshold of the standard MFR. [27] and [28] both applied a multi-scale approach, where multiple scales of the Gaussian function were used to ensure vessels of varying widths were detected.

Unless otherwise stated, the methods in this literature review that used filters, wavelets and operators for vessel segmentation all had an aspect of linearity and all were applied in multiple orientations.

Mathematical morphology: Mendonca [29] used difference of offset Gaussian filters (first order derivative filter) for the extraction of vessel centrelines. Then returning to the original pre-processed image, vessel enhancement was performed separately at different scales by using a modified top hat operator (morphological operator) with a disc structural element of increasing size used to enhance vessels with different widths. At each scale a binary vessel map was produced by applying the double threshold operator to perform morphological reconstruction. The final vessel segmentation was obtained using the vessel centrelines along with several binary vessel maps to perform iterative region-growing. Fraz [30] proposed a modified version of Mendonca [29]. Here the vessel centrelines were detected using the first order derivative of the Gaussian kernel. The significant alterations were after the top hat vessel enhanced image was obtained; bit plane slicing was used in which the sum of the higher order bit planes (containing the majority of visually significant data) were used to create a binary vessel map. Again this was followed by a similar region-growing procedure using the vessel centrelines and the binary vessel map.

Vessel tracking: Cree [31] tracked vessels by fitting a two dimensional Gaussian model. An initial vessel point had to be selected manually along with estimations of its width and orientation. A small local region was cut out about this point and a matching width and orientation Gaussian model was fitted using the optimisation procedure of non-linear least squares. Accurate measurements of vessel width and orientation were made from the fitted model. A small step was made in the direction of the vessel and the previous measurements were used as estimates to make a new fit and so on. Other approaches include a vessel tracking method based on using a probabilistic formulation [32] and a method based on multi-scale line tracking [33].

Machine learning: (Pixels classified as vessel or non-vessel).

Supervised: Sinthanayothin [34] first applied principal component analysis (PCA) [35] to reduce the image to just structural detail (texture and uncorrelated noise

minimised). The Canny edge detector was applied to the first principal component to measure edge strength. Values from the first principal components and the edge strength were used as input data for a neural network classifier. Staal [36] was based on the extraction of ridges which were used to produce image primitives in the form of line elements. These line elements were then used to partition the image into patches. A k-nearest neighbour was used to classify pixels using a vector of 27 features derived from properties of the patches and line elements. Soares [37] created a feature vector for each pixel from the pixel's intensity and from the response of applying a two dimensional Gabor wavelet at multiple scales and various orientations. Pixel classification was performed with a Bayesian classifier in which class likelihoods were described as a linear combination of Gaussian functions. Ricci [38] proposed the use of line operators which were based on calculating the pixel line strength from the evaluation of the average intensity value along lines passing through the target pixel at different orientations. The average value of a short orthogonal line was also calculated to help reduce false positives. These two values along with the pixel's intensity value were used to construct a feature vector used by a support vector machine to perform classification. Fraz [39] used an ensemble classifier of boosted and bagged decision trees with a feature vector based on gradient orientation analysis, morphological transformation, line strength measures and Gabor filter responses. Marín [40] applied Hu moments to detect linear structures (vessels). These are shape descriptors invariant to translation, rotation and scale. The Hu moments along with grey-level values formed a 7-D feature vector used by a neural network classifier.

Unsupervised: Tolias [41] proposed an unsupervised fuzzy algorithm for vessel tracking. The method was based on a fuzzy C-means clustering algorithm finding the membership functions of the two linguistic values (vessel and non-vessel). The proposed scheme used only intensity information, hence no assumptions for the shape of the vessels were made and no edge information was required (usually corrupted by noise). Kande [42] improved the contrast of blood vessels against the background by applying matched filtering. Labelling was implemented on the enhanced image using a spatially weighted fuzzy C-means clustering algorithm to

segment the vessels. The spatial weighting element took into account that the spatial distributions of grey levels as image pixel intensities are not independent of each other.

Studies have shown that vessel calibre relates to hypertension and cardiovascular disease [43]. The main driving force for accurate segmentation has been for the quantification of vessel calibre [44-46] allowing for cardiovascular studies. Numerous other morphological features can be measured from the binary vessel map, such as tortuosity, branching angle and vessel density. Thus, vessel segmentation also forms the backbone for many automated systems aimed at diagnosing ophthalmic disease. The simplest and most widely used tortuosity measure is simply the arc length over the chord length which was first proposed by Lotmar [47]. Numerous methodologies have been developed to assess these features [48], [49], [50] including Martinez-Perez [51] which proposed various methods to quantify geometrical and topological properties. Perez-Rovira [52] presented a software application called VAMPIRE, which brought together several recent methodologies. As vessel segmentation is effectively the identification of anatomical landmarks, it has been applied to make image registration a much simpler task, whether it is to combine images from different modalities [53] or to assist in assessing the change in DR after a time interval.

2.2 Microaneurysm and Haemorrhage Detection

Attributes that help define microaneurysms are their red colour (similar to vessels), contrast with the background, circular shape and small size. Hence microaneurysm detection methodology often ignores objects that deviate away from the classical description and as a result haemorrhages which are larger in size and can deviate to a great extent in shape (dot, blot and flame) are often overlooked. Other algorithms on the other hand concentrate on the combined detection of both microaneurysms and haemorrhages (referred to together as red lesions) and hence are not too specific in regard to strict shape definitions. This section does not include the detection of the much larger and more serious pre-retinal and vitreous

haemorrhages that constitute to PDR. Note that performance assessment for red lesion detection was performed either on a per pixel, per candidate, per sub-image or per image basis. Results will not be included in this section.

There are various areas where detection systems will generally struggle. Microaneurysms can often have a very subtle appearance, possessing a very low contrast with the background. Red lesions that appear very close to or attached to vessels can often be difficult to detect. Also camera artifacts produce specks on the image which are very similar in appearance to microaneurysms and can easily produce false positives. Pigment spots, which are common, can also easily be mistaken for microaneurysms or haemorrhages.

As microaneurysms are the first signs of DR these algorithms aid in the detection of early stage DR. Also microaneurysms almost always appear to be present in any image containing DR regardless of the severity, and hence these algorithms also provide the classification of DR disease or no DR disease. As mentioned previously, such a classification can be used as a filter to remove all the images showing no signs of DR and therefore substantially reduce the workload for the human graders. For this reason microaneurysm detection has received the vast majority of the research in the field of automated DR detection.

Spencer [19] developed a microaneurysm detection method that has become the basis for many other methods. The top hat operator using a linear structuring element was applied which first retained the vascular structure and then subtracted it from the original image to produce an image containing only circular objects. This was followed by matched filtering using a 2-d Gaussian kernel (isotropic) to match the shape of the microaneurysms and then thresholding to produce a binary image containing microaneurysm candidates. A region growing algorithm was performed which delineates each candidate allowing for shape and intensity analysis. This allowed various measurements of the candidates to be taken such as size, perimeter, complexity, aspect-ratio and intensities, which were used to compute a manually derived classification. Cree [54] improved the performance of this method by expanding the feature set to include 5 more features. This comprised of 4 more

intensity features and the fifth new feature was the matched filter response. Frame [55] investigated the use of different types of classification used for this method including machine learning approaches. The manually derived classifier produced results that deemed it superior. Despite this result, machine learning classification algorithms come to dominate this field due to ease of use, capability of handling very large data sets and their ability to recognize complex patterns. Hipwell [56] successfully applied the method to normal retinal images as it had previously only been applied on fluorescein angiograms.

The top hat operator is an important technique in the field of retinal image analysis, and therefore expansion on how it works in this context shall be provided. It is a morphological operator that applies opening (erosion followed by dilation) using a structuring element. For these methods a linear structuring element was used whose length was greater than the diameter of the largest red lesion, although not too long to ensure it was small enough to fit within all of the vessel structure. Therefore after the opening, only objects that were able to fully contain the structuring element were retained. To ensure all vasculature was retained the structuring element was applied in multiple orientations as vessels appear in many orientations. The resultant images were combined to produce a single vessel image which was subtracted from the original image to produce an image containing only small objects.

Niemeijer [20] developed a red lesion detection system also based on Spencer [19], but with extensive adaptations. A new candidate extraction system based on pixel classification was proposed. Pixels were classified as being foreground (red lesions and vessels) or background using a k-nearest neighbour (kNN) classifier with a feature vector containing Gaussian derivatives and pixel intensities. Then connected component analysis was applied and objects smaller than a specific size were retained and hence the vascular structure was removed. This new extraction system was run in conjunction with the existing extraction system based on the top hat operator and matched filtering to create a hybrid extraction system. A kNN classifier was used for candidate classification using the existing feature set developed by Spencer [19] and Cree [54], along with several new features; amongst these were Gaussian derivatives and colour features.

Gardner [57] divided images into small sub-images of 20 x 20 and 30 x 30 pixels and then investigated the use of artificial neural networks with intensity values as inputs in order to label the sub-images as red lesions as well as exudates, vessels and normal background. Sinthanayothin [58] developed an algorithm to detect red lesions (also exudates). Blood vessels were removed using a neural network method, for which inputs were derived from principal component analysis and the edge detection of the first principal components. A moat operator was applied to the image which had the effect of creating a trough (dip) around the red lesions, which resulted in the strengthening of their edges and hence aided the segmentation process. Recursive region growing was performed to identify similar pixels within a region, in which the merged pixels were represented with their median intensity value and thresholding was performed to identify red lesions. Zhang [59] applied multi-scale correlation filtering with Gaussian kernels to detect microaneurysm candidates. A 31-d feature vector largely based on [19,20] was created and used to perform manually derived classification with the use of a discrimination table for the different features.

Sopharak [60] applied the extended minima transform, which was a thresholding technique and was defined as the regional minima of the h-minima transform. The h-minima suppressed all intensity values that were less than or equal to a predefined threshold. The regional minima were connected pixels with the same intensity value whose external boundary pixels all have a higher value. The resultant was a binary image containing various white regions representing small dark objects and small vessel segments. Previously detected exudates and vessels were removed to leave microaneurysm candidates and objects with a size smaller than a specific small size were classified as microaneurysms. Sopharak [61] provided an extension to [60] with an 18-d feature vector, used by a naïve Bayes classifier to classify the pixels of the microaneurysm candidates. Features were based on intensity, colour, size and shape.

2.3 Exudate Detection

Attributes that characterise exudates are their colour (yellow), high intensity values, high contrast with background, and sharp edges. Unfortunately they vary in size and shape. They also vary greatly in their pattern of distribution, in which they can appear as individual patches, clusters, streaks (tracking lines) and large rings of multiple exudates (circinate). Exudates are often referred to as bright lesions. Exudate detection can struggle in terms of the retinal feature of drusen providing false positives, this will be further explained later. Note that performance assessment for exudate detection was performed either on a per pixel, per candidate, per sub-image or per image basis. Results will not be included in this section.

A considerable portion of exudate detection methods were based on the morphological reconstruction technique. The idea was to perform a rough segmentation of exudates, which was likely to be all exudates accompanied by surrounding areas of retinal background. The resultant was then subtracted from the original image to leave black empty regions in their place and this acted as the marker image. The original image was used as the mask image. Morphological reconstruction by dilation was performed and resulted in the black regions being filled by surrounding pixels, and hence these regions appeared as retinal background. The difference between the original image and the reconstructed image was thresholded and the result was exudate segmentation. Walter [62] and Sopharak [63] applied this morphological reconstruction technique in which the initial rough segmentation was achieved by thresholding the local variance image, which was based on exudates being characterised by high contrast. Sopharak [64] also applied the morphological reconstruction technique, although the initial rough segmentation was achieved from using fuzzy C-means clustering of pixels using the features of intensity, standard deviation of intensity, hue and number of edges.

Osareh [65] used fuzzy C-means clustering of pixels to segment exudates according to their colour. Further classification was then carried out using a neural network classifier with inputs consisting of 18 features of exudate candidates including size,

edge strength and various colour measurements. Sopharak [66] demonstrated the superiority of a support vector machine classifier when compared to naive Bayes and nearest neighbour classifiers. The investigation was based on 15 features. There was no prior segmentation to detect candidates and instead features were measured on a per pixel basis. These features included intensity, standard deviation of intensity, hue, number of edge pixels and difference of Gaussian filter responses.

Drusen is a retinal feature that is very similar in appearance to exudates in terms of their yellow colour, size and shape, although they tend to possess a lower contrast with the background and have weaker edges. Their presence generally doesn't pose any risk and does not indicate DR, although there has been some interest in their detection as they may represent an early stage manifestation of age-related macular degeneration. Duanggate [67] proposed a drusen detection algorithm based on a scale-space approach combined with feature-based analysis (colour, texture). However the importance of drusen in the context of DR lay in differentiating it from exudates and Niemeijer [68] proposed such a method. Initially pixels were assigned a probability of being part of a bright lesion (exudates and drusen etc.) using the response to a set of Gaussian derivative filters and a kNN classifier. Pixels with a high probability were grouped to form lesion pixel clusters. These clusters were assigned a probability of being true bright lesions using a kNN classifier and various features including size and contrast. Lesion pixel clusters with a high enough probability underwent a final classification which differentiated the type of bright lesion using a linear discriminate analysis classifier and features including number of red lesions and the probability for the cluster.

2.4 New Vessel Detection

So far the detection of microaneurysms/haemorrhages has been covered which enables the classification of early stage DR, as well as the classification of DR disease/no DR disease. As mentioned previously, an automated detection system should strive to take on a more substantial role in the grading pathway than this simple classification. Therefore the emphasis is now on the classification of the

severity of the disease, which will give the automated system the ability to triage and therefore take on a more substantial grading role (see section 1.2.3). Exudate detection has been covered, which takes the first step to indicating a grade of DR showing a higher level of severity. This section will cover the detection of proliferative DR (PDR), which is the highest severity of DR, possessing a high risk of severe loss of vision and hence patients need urgent referral (R3, 95% within 1 week, 100% within 2 weeks [14]). The grading pathways from figure 1.9 and 1.10 emphasise the importance in detection and fast tracking R3 images. Despite this there really only exists a few small pockets of research into its automated detection as its variability in appearance makes it an extremely difficult task to perform. Another reason for the lack of research is the fact that PDR makes up a low percentage of DR cases. This fact to a person without the required medical knowledge may be wrongly perceived as a factor in its identification holding less significance. Ultimately it is a problem that is not very well understood by the image processing community.

As mentioned previously, PDR is characterised by the presence of new vessels. Therefore the terms PDR and new vessels are used synonymously throughout the literature and most methods (including this project) stating the detection of PDR are actually claiming to specifically detect new vessels. In reality it is not as simple as this. Yes, new vessels will always be present in a PDR case, although they may be obscured or completely hidden from view by features associated with new vessels such as pre-retinal haemorrhages, vitreous haemorrhages, fibrosis tissue and tractional retinal detachment. These make up only a small percentage of cases, and therefore the main focus should remain on new vessel detection. However, in the future, development of the detection of these other new vessel associated DR features is required to achieve a full PDR detection system.

The majority of methodologies were only based on detecting networks of new vessels (new vessel regions). These regions can be generally characterised by an increase in vessel segments, which appear in multiple orientations and also possess a tortuous nature. Another form of characterisation was that new vessel regions could be deemed as possessing a more erratic pixel intensity distribution. However, referring

back to their description in the pathology section (1.2.1), their variation in appearance makes them a very difficult task to detect.

Results from the performance evaluations will be stated in this section. The irregularity and obscurity of new vessels can make accurate individual pixel labelling by a human observer or automated systems an overly complex and unnecessary task. All methods work by classifying either vessel segments, local regions or entire images as containing the presence of new vessels. From a clinical point of view performance measures for new vessel detection are best presented on a per image basis. For the purpose of screening the main target is simply a yes or no on the presence of new vessels. However, precision in terms of region delineation or segment classification can hold benefits in terms of quantifying the new vessel growth.

The majority of new vessel detection methods can be split into two main categories, based on whether vessel segmentation was performed or not. Those methods based on vessel segmentation were developed with the purpose of analysing the characteristics of the binary vessel map in search of abnormality. The other category is methods based on extracting textural information from the images and therefore avoiding the difficulties that arise from vessel segmentation.

The section will be categorised into the following sub-sections: PDR vessel segmentation, followed by the two main categories of segmented vessel map analysis and texture analysis, and ending with stage identification, other methods and performance comparisons. Figure 2.1 and 2.2 offer a breakdown of the reviewed articles in this section in terms number of articles in each methodology category and the yearly decomposition of articles. This section will end with a summary of the reviewed articles in the form of tables 2.2, 2.3, 2.4, and 2.5.

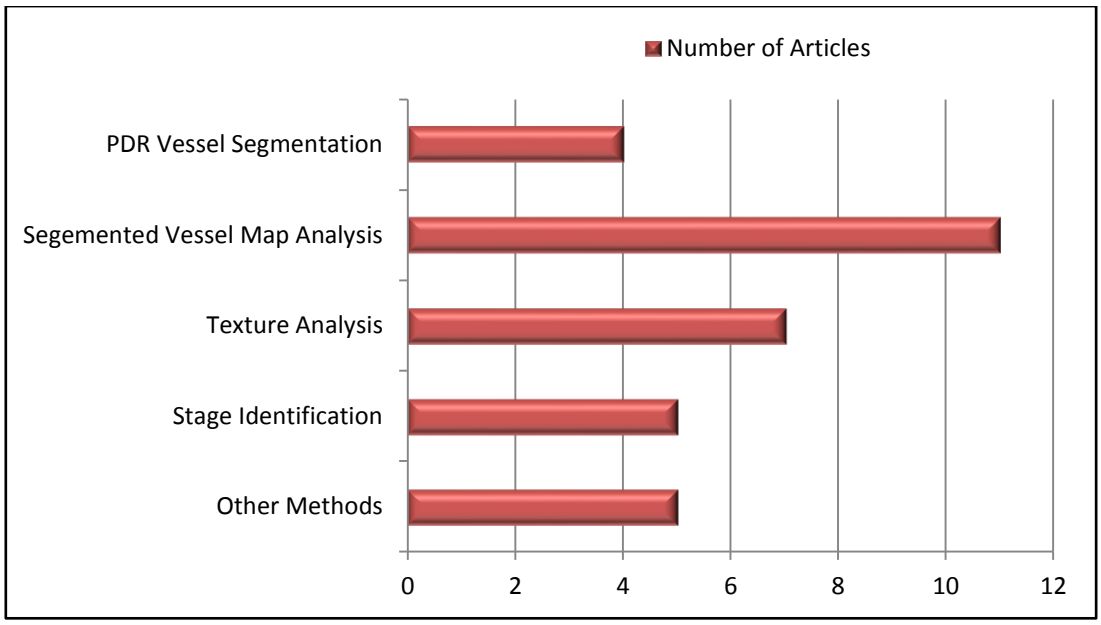


Figure 2.1: Categorization of reviewed articles in section 2.4 (New Vessel Detection).

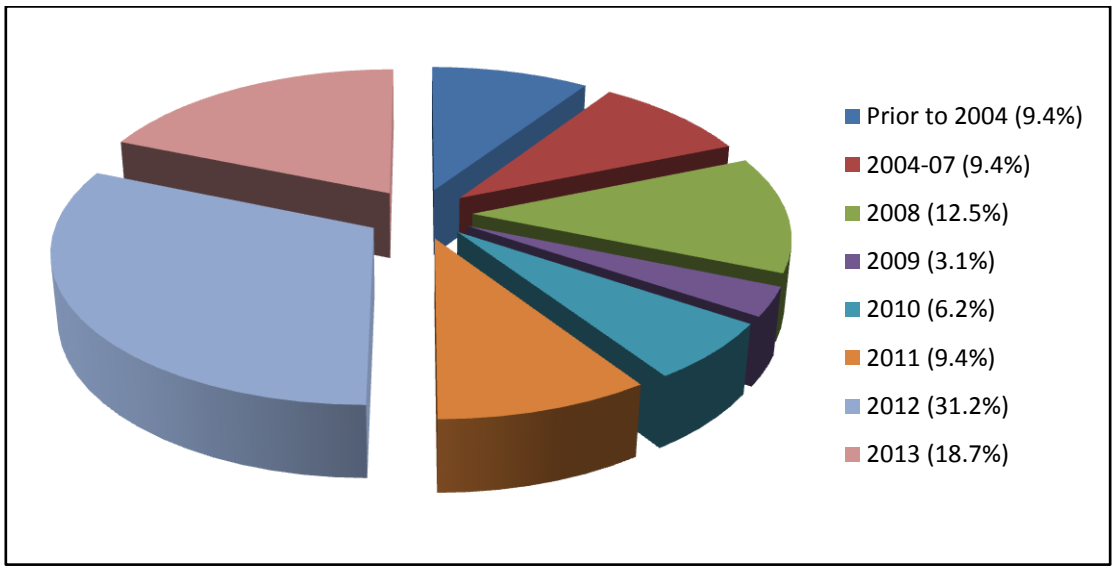


Figure 2.2: Yearly decomposition of review articles in section 2.4 (New Vessel Detection).

2.4.1 PDR Vessel Segmentation

One of the first steps of some PDR detection methods was vessel segmentation, with the intention to further analyse this vessel map in the search for the presence of new vessels. Vessel segmentation methods operate in terms of detecting structure and linearity, hence the tortuous and irregular nature of new vessels and their often obscure appearance (out of focus) can make this a very problematic task. The latter

point was a problem in respect to vessels being blurred into each other and therefore changing their appearance. Also their low contrast with the background further adds to the problem. Therefore if new vessel segments/regions were missing from the segmented vessel map then any further analysis was seriously jeopardised. Another problematic task was the false segmentation of non-vessels. There are only a few vessel segmentation methods that have been particularly designed for PDR detection and these shall be discussed next.

Vessel segmentation methods based on detecting linear vessel segments also respond to non-vessel edges. False positives caused by exudates can seriously hinder new vessel detection methods and therefore their reduction can assist in the performance of new vessel detection systems. If falsely segmented, exudates can be very similar to new vessels in terms of possessing large local densities and large curved edges. L.Zhang [27] proposed a method that put their main focus on tackling this particular problem, to provide an improved vessel segmentation method to be used as the initial step for PDR detection methods to provide an improved performance. B.Zhang [28] developed a method that also put clear emphasis on the reduction of false positives caused by exudates, although the benefits to PDR detection are not stressed. Both methods were based on Gaussian match filtering and have already been discussed in section 2.1.

Akram [69] was another method offering improved vessel segmentation to be used for PDR detection. The various angles and scales of a 2-dimensional Gabor wavelet were applied to achieve vessel enhancement. This was followed by a multi-layered thresholding technique. This worked by first applying an initial threshold and then performing thinning to produce a vessel skeleton. The threshold was then iteratively lowered and with each iteration all those vessels which were connected to the vessels segmented from the previous iterations were kept. Adaptive thresholding was further applied to improve segmentation accuracy. This vessel segmentation method was reported to work well for the detection of faint and thin vessels and therefore had improved suitability for new vessel detection.

Ramlugun [22] described a small vessel extraction technique; the main contribution was the varying of the clip limit for contrast limited adaptive histogram equalization (CLAHE). Large vessels were easily differentiated from the background due to their inherent high contrast, and therefore a low clip limit was applied. For small vessels a high clip limit was applied because of their low contrast difference to the background. This was followed by the use of Gabor filters for vessel enhancement, double-sided thresholding [27] and hysteresis thresholding.

So far steps have been discussed to improve the segmentation of faint and fine vessels, as well as reducing false positives. However, as already mentioned above, the main problem with vessel segmentation methods based on detecting linear vessel segments was that they will struggle with new vessels that are very tortuous and irregular in structure. This issue doesn't appear to be well addressed. A possible route of investigation would be to loosen/adjust the parameters of the detection system to increase sensitivity to new vessels. An unfortunate consequence of this would be the increase in the segmentation of other DR lesions and general spurious objects. This would worsen, the more the parameters are adjusted to accommodate for new vessels. Therefore this area requires considerable further investigation.

2.4.2 Segmented Vessel Map Analysis

The following new vessel detection methods are categorised as those performing vessel segmentation. The purpose of segmentation was the analysis of the segmented vessel map in search of abnormality (new vessels).

The majority of papers in this section used conventional vessel segmentation techniques. Hence they did not put emphasis on the fact that the segmentation should be designed for PDR detection, as documented in the previous section.

2.4.2.1 Morphological features

The main set of characteristics explored from the segmented vessel map were based on morphology. This is the study of form or structure and includes aspects such as area, perimeter, scale and orientations. Characteristics based on intensity and gradient information were also useful to extract from the segmented vessel map to assist in analysis.

Hassan [70] developed a method that essentially quantified the local vessel coverage as an indication of new vessels. A standard Gaussian matched filtering approach was applied for vessel segmentation despite documented modified approaches showing improvements. This was followed by various morphological operations and thresholding to produce a binary vessel map. Next all red lesions and bright lesions (e.g. exudates) were removed using the compactness measure and this was followed by thinning to produce a vessel skeleton. Local analysis was performed to detect the new vessels using a small square window of size 100 x 100 pixels which was used to scan across the image and at each stage the number of blood vessels and area of blood vessels within the box region was calculated. A higher number for both of these would indicate new vessels. The classification appears to be manually derived. Results showed a sensitivity of 63.9% and specificity of 89.4% on the detection of new vessels which was on a per pixel basis. To clarify, pixels were not classified as belonging to new vessels but instead as belonging to new vessel regions. Such pixels may just be retinal background or normal vessels; however they belonged to a region containing new vessels. These results illustrated how well the method was capable of delineating the new vessel regions, which does have its advantages, for example, to identify the location and quantify the spread of new vessel growth. The sensitivity result was low, although the obscurity of new vessel regions can make them extremely difficult to correctly delineate and therefore a high sensitivity was not viable. Results on a per image basis would have been interesting from a clinical point of view, although these were not provided as the image data set used was created with only images containing new vessels. Finally, as different parts of a normal retina can be quite varied in terms of how much vascular structure they contain, it would

suggest that the number and area of vessels within a region alone are not really sufficient for the classification of new vessels. However used appropriately and in conjunction with other features, vessel coverage can still pose as a very strong aid in the detection of new vessels.

Jelinek [71] developed an algorithm for new vessel detection which used the wavelet transform approach both for vessel segmentation and for the morphological characterization of new blood vessels. Various angles and scales of a 2 dimensional Gabor wavelet were applied to achieve vessel segmentation. Various post processing steps were applied to the vessel map including skeletonization. Seven morphological features were then determined from the overall vessel skeleton. Three of these were traditional features, which were area, perimeter and circularity. The other four features were all obtained from the application of the derivatives of Gaussian wavelets to the vessel skeleton. These were curvature, orientation entropy, continuous wavelet transform second moment (measure of dispersion) and correlation dimension (used to measure a type of fractal dimension). A linear discriminant analysis was applied on the feature set for classification, reporting a sensitivity of 94% and a specificity of 82% for the detection of PDR (new vessels) on a per image basis. The features used may all be indicative of new vessels, although they are all measured only on a global scale, and therefore the introduction of local analysis may aid in making this method more sensitive to new vessels. Also it should be remembered that any method that was based on only the use of global features would mean that the identification of where the new vessel regions were located within the image was not possible. A final point is that this method has the advantage of being applied to fluorescein angiograms which make the presence of new vessels much clearer due to higher contrast. Producing this type of image requires the injection of a contrast agent to the vascular system and is therefore invasive and time consuming and is not suitable for the large scale screening process.

Goatman [72] developed a method to detect new vessels on the optic disc using the combination of watershed lines and ridge strength to detect vessel segments and a comprehensive set of 15 features for the classification. Vessel detection at the optic disc was aided by its bright background, which provided a high vessel to background

contrast. Images were inverted and the watershed transform was applied which divided the image into regions based on a topographic map of image grey levels. The dividing lines between regions are known as watershed lines (or ridges) and represented vessel centre lines. All bifurcation pixels were removed to form vessel segments. The nature of the watershed transform is to create closed regions connected by watershed lines, and as a result of this some of the segments were non-vessels. These were removed using the ridge strength calculation along with thresholding to remove the weaker ridges (segments). The Gaussian filtered image (Gaussian function at a certain scale in accordance with vessel size) and various gradients (first and second order partial derivatives) were used to calculate ridge strength. Various measurements taken from these vessel segments were used to derive a feature vector for each vessel segment containing 15 components which was used by a support vector machine for classification of segments as normal or abnormal. Note that the feature set was not restricted to just aspects of the segment concerned, but also used aspects of the surrounding local vasculature. The top 5 features assisting the classification process were number of vessels, mean vessel wall gradient, normalized grey level, mean vessel width and mean ridge strength. Other features included vessel density, gradient variation, segment length, direction and 3 tortuosity measures amongst others. It was deemed that classification on a per vessel segment basis was relatively poor, and therefore the outcome was used to assess per image detection performance. An image labelled as containing new vessels required one or more segments to be classified as new vessels. A sensitivity of 84.2% and a specificity of 85.9% were achieved on a per image basis for the detection of new vessels.

Arputham [73] and Pavai [74] have both proposed alternations to Goatman [72], although this only entailed changing the method used for the initial detection of the optic disc. Goatman [72] applied a method described in [75]. First the optic disc's approximate location was obtained based on the confluence of major retinal vessels. This was then refined based on the circular edge of the optic disc, done using the Hough transform to identify circular forms within the image following gradient determination using the Sobel operator. Arputham [73] applied a region-based

active contour model to segment the optic disc. Pavai [74] performed optic disc segmentation using the texture descriptors of entropy, inertia and energy.

Akram [76] built on the work undergone in [69] to produce a new vessel detection system. The vessel segmentation method was identical with the use of Gabor wavelet, multi-layered thresholding and adaptive thresholding. A 5-d feature set was formed for each vessel segment; this consisted of energy, mean gradient, standard deviation gradient, mean intensity and intensity variation. These were mainly intensity and gradient based features used to characterise the vessel segment. A Bayesian classifier was used for classification to define each vessel segment as healthy or unhealthy (new vessel). A sensitivity of 98.93% and a specificity of 96.35% were reported on a per vessel segment basis. Goatman [72] had already shown that per vessel segment classification generated a relatively poor performance (the outcome was used accordingly to generate adequate per image results). Goatman, for each vessel segment, produced a rich feature set that not only described morphological, intensity and gradient aspects of the segment, but also described the morphology of the surrounding vasculature, which was still insufficient. Akram produced a feature set containing only a few intensity and gradient based measures, with no real emphasis on morphology. The feature set does not appear to contain sufficient information to distinguish new vessel segments, and therefore the good results that were reported are not what would be expected from such a feature set.

Akram [77] further builds on prior work [69,76]. The feature set had been extended, which included the original intensity and gradient based features as well as an additional 5 features based on the morphology of the vessel segment and the surrounding vasculature. Classification of vessel segments was performed using a multimodal m-Mediods based classifier. Mediods are representative objects in a cluster whose average dissimilarity to all the objects in the cluster is minimal. This classifier was a soft classification approach which caters for the presence of multimodal distributions, and therefore could deal with the complex, multimodal and overlapping nature of distributions of samples belonging to the two classes (new vessel and non-new vessel). Results from several datasets were listed separately; this

included a sensitivity of 98% and a specificity of 97% on a per segment basis for their largest dataset. A per image outcome with a sensitivity and specificity of 97% and 92% respectively for images with NVD and 96% and 94% respectively for images with NVE was achieved.

Saranya [78] created a feature vector that involved the use of Hu moments along with gradient, intensity and morphology based features for the detection of new vessels. The top-hat and bottom-hat operators were applied for vessel enhancement, followed by fuzzy C-means clustering for vessel segmentation. The feature vector was used to perform classification on a per segment basis using a K-nearest neighbour classifier. This outcome was used to give a per image performance of 96.25% and 89.65% for sensitivity and specificity respectively.

Welikala [79] developed a methodology based on multi-scale matched filtering with emphasis on using two different sets of parameters to allow for the detection of new vessels. Parameters were selected to first increase and then decrease the filters' response to new vessels, followed by thresholding to produce two separate binary vessel maps. The difference image resulted in most normal vasculature being removed and therefore leaving behind possible new vessels. A 5-d feature set based on the local morphology of the vasculature was calculated using a small scanning window of size 151 x 151 pixels. Classification was performed using a k-nearest neighbour classifier which achieved a sensitivity and specificity of 100% and 70% respectively on a per image basis.

2.4.2.2 Fractal Analysis

Fractals provide a method for describing natural objects that would be too complex for Euclidean geometry to describe. Fractals are typically self-similar patterns, meaning upon magnification there are subsets that look like the whole figure. This relationship can be represented by a single value called the fractal dimension, which is an index to magnification to describe at what factor the detail will be increased by.

Unlike ordinary Euclidean dimensions that are integers, fractal dimensions are non-integers. The fractal dimension is used to quantify the complexity of the fractal pattern, which can be considered as expressing how thoroughly the pattern fills the space. Hence an index of 1.7 would fill a 2-dimensional space more thoroughly than an index of 1.2. Fractal analysis is a study of morphology; however it can also be considered as a measure of texture.

Mainster [80] demonstrated that the retinal vasculature was a fractal and follows the theory of fractal geometry. This was presumed as a rough approximation, as natural objects exhibit scaling symmetry only over a limited range of magnifications and also tend to be only roughly self-similar. This paper stated that fractal dimensions provide an insight into the relationship between vascular patterns and retinal disease.

Daxer [81] applied fractal dimensions to characterise new vessels. The idea was that the presence of new vessels resulted in the whole pattern's convolutedness rising, hence filling the space more thoroughly and therefore possessing a higher fractal dimension. It could also be considered that the retina's fractal pattern would become upset at the local region containing new vessels. Instead of automated vessel segmentation the vessels were manually traced out. The density-density correlation function was used to calculate the fractal dimension. This value was calculated for just the superior temporal quadrant as opposed to the whole retinal image. The images assessed were just for a single patient at different stages in time, before new vessels, with new vessels and with the regression of new vessels. The results showed a significant change in the fractal dimension at each stage, with measurements of 1.665 before new vessels, 1.794 with new vessels and 1.708 with the regression of new vessels. Therefore this method was proposing the quantification of neovascularization, which could be used to evaluate the effect of treatment to PDR. What was also interesting was that this paper opened the door to the possibility for fractal dimensions to be applied to the detection of new vessels.

Karperien [82] applied fractal analysis for the detection of new vessels. A technique using a 2-dimensional Gabor wavelet was applied to achieve vessel segmentation. Also the paper expanded out their investigation by looking not only at global fractal

dimension measures, but also a local measure. The fractal dimension was calculated using the box counting method and the correlation dimension method to give two global measures. A local measure known as the local connected fractal dimension was also measured, which was calculated just like the box counting method but on a smaller scale, with the box reaching a maximum of 10% of the image size. A third global measure was used which was derived from the local measure by taking a mean across all pixels in the image. The global fractal dimension measurements produced showed a significant difference between the new vessels and control cases, although this was only shown for one of their global measures. A method of classification was not applied. Next the local fractal dimension measure was applied to every vessel pixel and showed that an image with new vessels contained a higher frequency of pixels at higher fractal dimensions. Again a method of classification was not applied. A further useful extension to this work would be to see a visual representation of how the local fractal dimension value varied across a single image containing new vessels, although unfortunately this was not included. This method was applied to fluorescein angiograms, which are not suitable for the large scale screening process.

Both Daxer [81] and Karperien [82] have demonstrated a difference between the absence and presence of new vessels using fractal analysis. The next step would be to apply this form of analysis to develop a new vessel detection system and to assess the level of results achieved. Using just a global measure would be problematic as the normal retinal vasculature can vary enormously from person to person; Karperien [82] stated literature that has shown the majority of values range from 1.60 to 1.88. Therefore the presence of new vessels in a lot of cases possibly may not possess a fractal dimension high enough to be outside this range. The local measure may possess slightly more prospects for future development. A possibly better approach to that already mentioned would be to calculate local region measures and assess them relative to the global measure. Therefore, any local region measures significantly higher than the global measure would indicate new vessels. It should not be presumed that for normal vasculature all local regions share the same fractal pattern as the whole; in fact many local regions may be empty and contain no detail. However, strategically placed local regions (normal) should theoretically possess the

same fractal pattern as the whole vasculature (normal). Unfortunately the retinal vasculature is only a rough approximation of a fractal and this is even more evident when observing the local regions.

2.4.3 Texture Analysis

There are several methodologies to assess image texture and they can vary considerably, hence in this report the texture methods have been categorised separately. These methods benefit from not requiring vessel segmentation and hence did not have to deal with difficulties associated with it for new vessel detection.

2.4.3.1 Statistical Texture Analysis

Image texture can be difficult to define, in fact there exist several different definitions of image texture attempted by researchers. There may be no universally agreed upon definition; however, a definition that is widely accepted defines image texture as the spatial variation of pixel intensities.

Frame [83] proposed the use of texture analysis for the detection of new vessels. The retinal surface and its features were deemed to be more stochastic than regular and hence statistical texture analysis was used. Information on the texture of a region is contained in the tone (pixel intensity) and structure (spatial relationship) of the pixels. Local analysis was performed on each pixel using a small local box region around the pixel and 14 texture measures were calculated (e.g. variance, entropy and contrast). Eight of these were first order statistical texture measures and the other six were second order statistical texture measures. First order measures can sometimes be misleading as they are based solely on the grey level intensity histogram and hence possess no spatial relationship information. Second order measures combat this problem as they are derived from the grey level co-occurrence matrix (GLCM), which is a tabulation of how often different combinations of pixel intensity values occur in a region according to a given direction and separation

distance. Haralick [84],[85] first suggested the use of the GLCM, which has become one of the most widely used texture analysis techniques. The idea was that texture measures can identify local regions in the image that exhibit the very irregular distribution of pixel intensities of a new vessel region. Linear discriminant analysis used the 14-element feature vector to classify the pixel's region as containing new vessels or not. No results were provided by this paper, instead some images displaying the results of their method were presented. However, the presentation of these images in their report is very poor in quality and hence makes it extremely difficult to make any visual assessment by the reader. This paper has concluded that the detection of new vessels has yielded unconvincing results, although a statistical difference does exist. A possible improvement would be to increase the size of the local box region, as a maximum of 8x8 pixels was used for an image of 1024x1024 pixels, which seems insufficient to really capture the intensity variation that new vessel regions provide.

Nithyaa [86] proposed the use of statistical texture features for the detection of new vessels at the optic disc. No local box region/scanning window was used, as the texture analysis was performed only on the region local to the optic disc. Features included standard deviation, entropy and skewness, which were measured from the grey level intensity histogram. A neuro-fuzzy classifier (combination of neural networks and fuzzy logic) used a six element feature vector to perform classification. Sensitivity and specificity results were not stated. The area under the ROC curve of 0.947 was reported for the detection of new vessels on the optic disc on a per image basis. Many details were not presented in this work and future publication will be required to clarify the technique.

2.4.3.2 Spectral Texture Analysis

The texture of retinal images can also be described by their frequency content. Measures can indicate coarseness/fineness as well as directionality. These measures

can be used to characterise new vessel regions which contain vessels that are fine in calibre and possess multiple orientations.

Agurto [87] utilized a multi-scale amplitude modulation frequency modulation (AM-FM) method for the characterization of retinal structures. The algorithm was applied on small 40 x 40 pixel image patches which contained the retinal structures of either microaneurysms, haemorrhages, exudates, new vessel, retinal background or vessels. A filter bank (array of band pass filters) was used to create the image at various frequency scales. At each scale AM-FM demodulation was applied and from the components the instantaneous amplitude (IA) and instantaneous frequency (IF) were computed. The modulus IF is a measurement of the local frequency content, local image intensity variations are reflected in the IA and also relative angles are estimated locally as deviations from the dominant neighbourhood angle. For each image patch a cumulative distribution function (CDF) was produced for each of the three measures at each of the image scales, and these were used as the texture feature vector. Significant differences were shown statistically using the Mahalanobis distance to compute the differences between extracted feature vectors of the different types of retinal structures. Prior to this the dimensionality of the feature vectors was reduced using principal component analysis.

Agurto [87] further demonstrated the use of these features to perform whole retinal image classification. This involved the division of the image into small patches followed by the same extraction of features and the reduction of dimensionality (done for each patch). Next k-means clustering was used to produce a single feature vector for the whole image and finally the use of partial least squares for classification. Results for the classification of DR versus no DR showed a sensitivity of 92% with a poor specificity of 54% (per image). A second whole retinal classification was shown for the detection of just advanced forms of DR (risk 3) that required urgent referral and produced a sensitivity and specificity of 100% and 88% respectively. However, their description of risk 3 did not represent only PDR, it represented new vessels, significant macular oedema or a large amount microaneurysms/haemorrhages (greater than 15).

Agurto [88] extended their methodology based on AM-FM techniques to specifically detect NVD. However, this now included the use of vessel segmentation as well as granulometry. A sensitivity of 96% and a specificity of 83% were achieved on a per image basis.

Vallabha [89] proposed the use of a Gabor filter bank for identification of vascular abnormalities in DR. This method stated that its intentions were the detection of features such as venous beading and intra-retinal microvascular abnormality (IRMA) for classification as severe pre-proliferative DR. These may not be new vessels, although they represent a very severe stage of pre-proliferative DR where the growth of new vessels is imminent. The appearance of IRMA is very similar to that of new vessels, and therefore it is unlikely that any algorithm could differentiate them; therefore, the principles of this method can also apply to new vessel detection. Image patches of the retina of size 256 x 256 pixels were used. These patches were filtered through a Gabor filter bank consisting of several filters tuned to specific scales (frequencies) and orientations. Plots of energy versus orientation were produced for each scale. Image patches containing vascular abnormalities contained fine vessels with multiple dominant orientations, and hence the plots (at finer scales) contained multiple peaks. A classification method has yet to be applied.

2.4.3.3 Combined Texture Analysis

Vatanparast [90] compared the performance of six different texture based methods for the detection of neovascularisation. This was a region-based method, with a window of size 128 x 128 pixels used to create local regions (patches). Texture information extracted from these patches includes GLCM, AM-FM, Gabor filters, Contourlet transform and local binary patterns. Each of the six texture methods had its own feature set created and was evaluated separately. A support vector machine with a linear kernel was used for classification of the patches. The AM-FM technique produced the best results with a sensitivity of 99.62% and a specificity of 96.61% on a per patch basis.

Lee [91] proposed a new vessel detection method which combined statistical texture analysis, high order spectrum analysis (derived from the average Fourier spectrum signal) and fractal analysis. However, vessel segmentation was required for multi-fractal analysis (not for mono-fractal analysis). Classification was performed using a logistic regression classifier to achieve a sensitivity of 96.3% and a specificity of 99.1% on a per image basis.

2.4.4 Stage Identification

Methods discussed so far were all capable of identifying new vessel regions through analysis of the specific characteristics that new vessel regions exhibit. The methods in this section differ largely in terms of the main features that were used to assess the amount of all general DR pathology in the whole image as opposed to the type. The general idea was that the greater the amount of DR pathology present, the more severe the stage of DR, with PDR being the most severe stage. It should be noted that a large amount of DR pathology is only indicative and is by no means definitive of the presence of new vessels. For instance, new vessels are triggered by retinal ischemia, although the haemorrhages and other features associated with the capillary breakdown are not always apparent in such abundance and may in fact have started to fade away. Also new vessel growth may have been triggered only by a small restricted area having capillary breakdown and retinal ischemia.

Acharya [92] applied measures from the GLCM as well as the run length matrix [93] which, like the GLCM, also provides higher order statistical texture measures. The run length matrix contains information on the run of a particular grey-level, or grey-level range, in a particular direction. Examples of features are short run emphasis and long run emphasis. A feature set consisting of 5 texture measures was used by a support vector machine classifier in order to classify images into one of 4 different classes of DR which were normal, pre-proliferative, proliferative and macular oedema (maculopathy). A sensitivity of 98.9% and a specificity of 89.5% were reported on the correct class being achieved on a per image basis. The method was similar to Frame

[83], although it differs in terms of the use of global statistical texture measures as opposed to local. In this method no interest was taken in assessing local regions in search of these distinctive statistical measures that new vessels may exhibit. For global measures it is more likely to be the case of assessing the overall contribution from the summation of all DR features present in the image and therefore aids to indicate the amount of DR pathology present as opposed to the type (as mentioned above).

Anitha [94] used a statistical texture based feature set for multi-level pathology identification. The identification of PDR was not covered by the method, although other pathologies characterising vessel abnormalities were covered such as central neo-vascularisation membrane. This method differed in terms that its not detecting stages of DR, but instead it was detecting several different retinal pathologies. However, it was similar in that statistical texture measures were calculated globally. These extracted measures included energy, entropy and standard deviation. The feature set provided the inputs for a Kohonen neural network to perform classification achieving a sensitivity of 96% and a specificity of 98% on the correct pathology being detected on a per image basis.

The next three methods described all involved vessel segmentation, but the intention was to produce a binary map not containing only vasculature structure (including new vessels), but also other DR features such as microaneurysms and haemorrhages. Therefore, a simple feature like the area of the segmented map can be used to indicate the amount of pathology present. All methods use global features.

Nayak [95] had a feature set including the measures of vessel area and vessel perimeter of the binary map. This binary map was achieved using several morphological openings using a diamond and disc structuring element in order to make the image less noisy and this was followed by thresholding. The feature set was used by a neural network classifier in order to detect one of three different stages of DR including proliferative (new vessels), as well as normal and non-proliferative. A sensitivity of 90.32% and a specificity of 100% were reported for attaining the correct class on a per image basis. A sensitivity of 90.91% was reported

for the detection of PDR on a per image basis, specificity was not stated. The feature set included two other features of exudate area and contrast. These features required their own steps of image processing of the original image.

Yun [96] proposed a similar method to Nayak [95], with the binary map being achieved by the same method. Six features were used which consisted of vessel area and perimeter for each: the red, green and blue layers. Four stages of DR were identified using a neural network classifier, which included proliferative as well as normal, moderate non-proliferative and severe non-proliferative. Results showed a sensitivity of 91.7% and a specificity of 100% on the correct class being achieved on a per image basis. A sensitivity of 83% was reported for the detection of PDR on a per image basis, but no specificity was stated.

Priya [97] included a feature set containing the area of the binary map along with the two statistical texture measures of mean and standard deviation (obtained prior to segmentation). A Gaussian matched filtering approach was applied for vessel segmentation, although fuzzy C-means clustering was used to classify pixels as vessels or no vessels instead of the standard threshold approach. This approach was stricter in its vessel segmentation than the two previous methods and therefore did not have the same level of inclusion of other DR features in its binary vessel map. However, there still remained a sufficient inclusion to aid the classification process. The feature set was fed into a support vector machine for classification into one of three different stages of DR including proliferative as well as normal and non-proliferative DR. The results showed a sensitivity of 99.45% and a specificity of 100% for awarding the correct class on a per image basis. A sensitivity of 98.93% was reported for the detection of PDR on a per image basis, but no specificity was stated.

It is clear that the main features (with a couple of exceptions) used by these methods, along with the fact that they were calculated globally, means that were intended to define the general overall state of the image, in other words the amount of pathology present. Some of the methods from the previous sections were based on using feature sets containing only global measures and were capable of specifically identifying images with new vessel regions. An example is the curvature

of the entire retinal vasculature, which is a global feature used in a feature set by Jelinek [71] along with other global measures to specifically indicate the presence of new vessels. Therefore it is clear that it is the type of features along with the fact that they are measured globally that defines the fact that only the indication of the amount of pathology is expected.

It should also be noted that in this section there has been a strong emphasis that features measure the amount of pathology and the classification was performed in respect to that amount of pathology. In reality the patterns and boundaries that the classifiers and their feature set found are likely to be more complex.

2.4.5 Other Methods

Patients attend DR screening appointments on a regular basis, and therefore screening programmes develop image libraries for each patient containing images from previous appointments. Human graders often use these additional images to assess the change in the condition of the retina, which helps in their detection of DR. A few automated detection methods have also taken advantage of this fact. McRitchie [98] used image registration based on affine transformations and mutual information, followed by image subtraction for visualizing the changes which occur in retinal pathology over a one year interval. However, this work only explored changes occurring due to microaneurysms, haemorrhages and exudates. Amrutkar [99] proposed an image subtraction technique to detect new vessels, although a future publication will be required to clarify the technique and the results.

There exist techniques developed from other research topics that are relevant to PDR detection. Automated methods for the quantification of micro-vessel density within the inner surface of egg shells in order to study the angiogenesis in developing chick embryos is an active research field [100,101]. Measures included branching points, vessel length and GLCM textural information. A system using edge contour analysis was presented by Zutis [102] for detecting abnormal retinal capillary regions, with the focus on telangiectasia.

2.4.6 Performance Comparisons

Several publicly available retinal image databases exist (see table 2.1); these include the STARE database [26], the DRIVE database [36], the ARIA online database [103], the ImageRet database [104] and the MESSIDOR database [16]. In these databases each retinal image is also accompanied by an annotated image. The term annotated refers to the labelling of a retinal image to provide a ground truth, whether it is the manual segmentation of vessels, the manual segmentation of lesions, marking the location of lesions or an overall image DR grade. Therefore, this allows for the training and testing of algorithms, or just the testing for algorithms not using supervised machine learning. Hence a database will be often divided into a training set and a testing set. But most importantly these databases allow for performance comparisons. Several algorithms being evaluated on the same database allows for direct comparisons of their performances to be made. For example the DRIVE and STARE databases have become the standard databases for evaluating any vessel segmentation algorithm.

Table 2.1: Publicly available retinal image databases.

Database	Year	Retinal Images	Annotations
STARE [26]	2000	10 normal, 10 pathological	Manual segmentation of vessels
DRIVE [36]	2004	33 normal, 7 pathological	Manual segmentation of vessels
ARIA [103]	2006	61 normal, 151 pathological	Vessel tracing, optic disc and fovea manually marked
ImageRet [104]	2008	25 normal, 194 pathological	Microaneurysms, haemorrhages and exudates manually marked
MESSIDOR [16]	2008	1200 normal/pathological	Diabetic retinopathy image grade

Unfortunately there isn't a standard database used for evaluating PDR detection systems. This means that comparisons of different algorithms' performances can be difficult to make. New vessel detection systems perform evaluation using one of the following options. Option 1 is to create a new vessel image dataset from one of the

public databases or a combination of public databases; however, new vessel images are scarce in these databases. Option 2 is to create a new vessel dataset using images from a local hospital or a local screening programme. Option 3 is the use of a combination of both public and local databases.

Unfortunately there is a lack of new vessels images in the publicly available databases mentioned above; therefore the majority of reported methodologies adopted either option 2 or 3. These local new vessel databases have not gone on to be released publicly due to the strict regulations and the lengthy procedure involved in gaining the required ethical approval. Therefore, with no external access to such databases, it is difficult for us to know how challenging a task a local database presented to the methodology that it was used by. This is in respect to whether or not the database contained a sufficient variety of new vessel formations, as well as a sufficient range of other clinical conditions. From the details provided by the reported methodologies, it is evident that not enough emphasis was put on ensuring that the non-PDR images contained an adequate amount of images with other pathologies to accompany the images of healthy retinas.

Generally, the reported methodologies performed evaluation using databases that contained a relatively low number of PDR images. Hassan [70] used only 11 PDR images for evaluation, whilst Goatman [72] used 38 PDR images and 71 non-PDR images. The majority of the reported methodologies performed evaluation using between 10 and 40 PDR images (see table 2.5).

Akram [77] used four of the retinal public databases: DRIVE, STARE, ImageRet and MESSIDOR. In total, they extracted a large number of non-PDR images as well as 52 images with PDR from these databases to use for evaluation. Such numbers of publicly available PDR images gives the potential to create a standard database for evaluating PDR detection. However, the image file names/numbers that they used were not listed and we were unable to find the number of new vessel images they refer to.

As the field of PDR detection advances, a publicly available PDR database will be invaluable to help in the evaluation and direct comparison of different algorithms. This database would need to be sufficient in size. Images with PDR are a lot more varied in appearance than most of the retinal image processing community are aware. Therefore this database would need to provide a varied selection of PDR images to educate the researchers and to allow for more appropriate training and testing. This varied selection would also include single new vessels, obscure new vessels and also cases of when new vessels are completely hidden from view by new vessel associated features (e.g. pre-retinal haemorrhage). However, the database should be sub-divided to create a dataset in which all the PDR images include visible new vessel regions, as this will allow researchers to first accomplish tackling this task. This database would also need to provide normal images and a varied selection of other stages of DR.

Table 2.2 lists all methods discussed in section 2.4. Whilst these methods all relate to the detection of PDR, only a handful have proposed a completed system capable of such detection, and this is indicated by column 7 of the table 2.2. The performance metrics listed are only those that apply to PDR detection. Table 2.3 and table 2.4 list the performance measures of completed PDR systems. Table 2.3 includes only those methods designed to specifically detect new vessels. Table 2.4 includes the methods from section 2.4.4 which work along a very different principle to identify the stage of DR, which includes a stage for PDR. Calculation of performance measures for the stage of DR identification is not a binary classification and is no longer a standard and simple task; therefore, there could be discrepancies in its calculation amongst the algorithms in table 2.4. Finally, table 2.5 lists the database type and size used by the algorithms listed in table 2.3 and table 2.4.

Table 2.2: Categorization of methods from section 2.4. PDR = proliferative DR, SN = sensitivity, SP = specificity, Acc = accuracy, AUC = area under receiver operating characteristic (ROC) curve, PPV = positive predictive value.

Algorithm	Year	Title	Type of Image	Intended PDR Detect-ion	Type of PDR Detect-ion	Complete PDR Detection	Perform-ance Metrics	Section
L.Zhang [27]	2009	A modified matched filter with double-sided thresholding for screening proliferative diabetic retinopathy	Normal fundus images	Yes	-	No	-	PDR vessel segment-ation 2.4.1
B.Zhang [28]	2010	Retinal vessel extraction by matched filter with first-order derivative of Gaussian	Normal fundus images	No	-	-	-	
Akram [69]	2012	Automated segmentation of blood vessels for detection of proliferative diabetic retinopathy	Normal fundus images	Yes	-	No	-	
Ramlugun [22]	2012	Small retinal vessels extraction towards proliferative diabetic retinopathy screening	Normal fundus images	Yes	-	No	-	
Hassan [70]	2011	Detection of Neovascularization in Diabetic Retinopathy	Normal fundus images	Yes	NVE, NVD	Yes	SN, SP, AUC	Segmented vessel map analysis 2.4.2
Jelinek [71]	2007	Automated segmentation of retinal blood vessels and identification of proliferative diabetic retinopathy	Fluorescein angiograms	Yes	NVE, NVD	Yes	SN, SP, AUC	
Goatman [72]	2011	Detection of new vessels on the optic disc using retinal photographs	Normal fundus images	Yes	NVD	Yes	SN, SP, AUC	
Arputham [73]	2012	An Approach for the Detection of Proliferative Diabetic Retinopathy	Normal fundus images	Yes	NVD	Yes	SN, SP	
Pavai [74]	2013	Identification of proliferative diabetic retinopathy using texture segmentation	Normal fundus images	Yes	NVD	Yes	SN, SP	
Akram [76]	2012	Detection of neovascularization for screening of proliferative diabetic retinopathy	Normal fundus images	Yes	NVE, NVD	Yes	SN, SP, PPV	
Akram [77]	2013	Detection of neovascularization in retinal images using multivariate m-Mediods based classifier	Normal fundus images	Yes	NVE, NVD	Yes	SN, SP, PPV, Acc, AUC	
Saranya [78]	2011	A novel approach for the detection of new vessels in the retinal images for	Normal fundus images	Yes	NVE, NVD	Yes	SN, SP, Acc	

	2	screening diabetic retinopathy						2.4.2
Welikala [79]	2013	Differing matched filter responsivity for the detection of proliferative diabetic retinopathy	Normal fundus images	Yes	NVE, NVD	Yes	SN, SP	
Mainster [80]	1990	The Fractal Properties of Retinal Vessels: Embryological and Clinical Implications	Fluorescein angiograms	Yes	NVE, NVD	No	-	
Daxer [81]	1993	Characterisation of the neovascularisation process in diabetic retinopathy by means of fractal geometry: diagnostic implications	Normal fundus images	Yes	NVE, NVD	No	-	
Karperien [82]	2008	Automated detection of proliferative retinopathy in clinical practice	Fluorescein angiograms	Yes	NVE, NVD	No	-	
Frame [83]	1997	Texture analysis of retinal neovascularisation	Normal fundus images	Yes	NVE, NVD	Yes	Visual	Texture analysis 2.4.3
Nithyaa [86]	2012	Identification of the diabetic retinopathy by detecting new retinal vessel using fundus image	Normal fundus images	Yes	NVD	Yes	AUC	
Agurto [87]	2010	Multiscale AM-FM Methods for Diabetic Retinopathy Lesion Detection	Normal fundus images	Yes	NVE, NVD	Yes	SN, SP, AUC	
Agurto [88]	2012	Detection of neovascularization in the optic disc using an AM-FM representation, granulometry, and vessel segmentation	Normal fundus images	Yes	NVD	Yes	SN, SP, AUC	
Vallabha [89]	2004	Automated detection and classification of vascular abnormalities in diabetic retinopathy	Normal fundus images	Yes	NVE, NVD	No	-	
Vatanparast [90]	2012	A feasibility study on detection of neovascularization in retinal color images using texture	Normal fundus images	Yes	NVE, NVD	Yes	SN, SP	
Lee [91]	2013	Detection of neovascularization based on fractal and texture analysis with interaction effects in diabetic retinopathy	Normal fundus images	Yes	NVE, NVD	Yes	SN, SP, Acc, AUC	
Acharya [92]	2012	An Integrated Index for the Identification of Diabetic Retinopathy Stages Using Texture Parameters	Normal fundus images	Yes	Stage	Yes	Not stated	Stage identification 2.4.4
Anitha [94]	20	Automated multi-level pathology identification	Normal fundus	No	-	-	-	

	12	techniques for abnormal retinal images using artificial neural networks	images					2.4.4
Nayak [95]	2008	Automated identification of diabetic retinopathy stages using digital fundus images	Normal fundus images	Yes	Stage	Yes	SN	
Yun [96]	2008	Identification of different stages of diabetic retinopathy using retinal optical images	Normal fundus images	Yes	Stage	Yes	SN	
Priya [97]	2011	Review of automated diagnosis of diabetic retinopathy using the support vector machine	Normal fundus images	Yes	Stage	Yes	SN	
McRitchie [98]	2006	Image registration and subtraction for the visualization of change in diabetic retinopathy screening	Normal fundus images	No	-	-	-	Other Methods 2.4.5
Amrutkar [99]	2013	An efficient approach for the detection of new vessels in PDR using image subtraction and exudates in SDR using blobs detection	Normal fundus images	Yes	NVE, NVD	Yes	-	
Doukas [100]	2008	Computer-supported angiogenesis quantification using image analysis and statistical averaging	-	No	-	-	-	
De La Cruz [101]	2013	Preliminary studies on vascular network quantification in chick chorioallantoic membrane images	-	No	-	-	-	
Zutis [102]	2013	Towards automatic detection of abnormal retinal capillaries in ultra-widefield-of-view retinal angiographic exams	Fluorescein angiograms	No	-	-	-	

Table 2.3: Performance measures for the identification of new vessels (proliferative DR).

Algorithm	Year	SN	SP	Acc	AUC	Level
Hassan [70]	2011	63.90	89.40	-	0.7045	Pixel
Jelinek [71]	2007	94.00	82.00	-	0.900	Image
Goatman [72]	2011	84.2	85.9	-	0.911	Image
Arputham [73]	2012	84.70	86.10	-	-	Image
Pavai [74]	2013	88.89	91.30	-	-	Image
Akram [76]	2012	98.93	96.35	-	-	Segment
Akram [77]	2013	98.00	97.00	98.00	0.980	Segment
Akram [77]	2013	96.00	94.00	0.95	-	Image
Saranya [78]	2012	96.25	89.65	96.53	-	Image
Welikala [79]	2013	100.00	70.00	-	-	Image
Nithyaa [86]	2012	-	-	-	0.947	Image
Agurto [87]	2010	100.00	88.00	-	0.980	Image
Agurto [88]	2012	96.00	83.00	-	0.940	Image
Vatanparast [90]	2012	99.62	96.61	-	-	Patch
Lee [91]	2013	96.30	99.10	98.50	0.993	Image

Table 2.4: Performance measures for the identification of the stage of DR, which includes a stage for PDR. SN, SP and AUC are for the correct stage of DR being identified. PDR SN and PDR SP are specifically for the detection of the PDR stage. All algorithms produce results on a per image basis.

Algorithm	Year	Stages	SN	SP	AUC	PDR SN	PDR SP
Acharya [92]	2012	4	98.9	89.5	0.972	-	-
Nayak [95]	2008	3	90.32	100	-	90.91	-
Yun [96]	2008	4	91.7	100	-	83	-
Priya [97]	2011	3	99.45	100	-	98.93	-

A plot of the sensitivity and specificity of the completed PDR detection systems from the reviewed articles is illustrated in figure 2.3. However, as stated above, comparison between these different methodologies is difficult to assess due to there being no standard database for evaluating the detection of PDR. Figures 2.4-2.6 illustrate how comparison difficulties are further extended by variations in respect to the type of image used, the type of PDR detection performed and the level selected for performance evaluation.

Table 2.5: Database type and size used by the algorithms listed in table 2.3 and table 2.4.

Algorithm	Year	Database	Images	PDR images	Non-PDR images
Hassan [70]	2011	Combination of public + local	11	11	0
Jelinek [71]	2007	Local	27	16	11
Goatman [72]	2011	Local	109	38	71
Arputham [73]	2012	-	-	-	-
Pavai [74]	2013	Local	-	-	-
Akram [76]	2012	Public (ImageRet)	20	10	10
Akram [77]	2013	Combination of public	503	52	451
Saranya [78]	2012	Combination of public + local	50	20	30
Welikala [79]	2013	Public (MESSIDOR)	20	5	15
Nithyaa [86]	2012	Local	150	100	50
Agurto [87]	2010	Public (ETDRS [105])	30	-	-
Agurto [88]	2012	Local	57	27	30
Vatanparast [90]	2012	Local	43	31	12
Lee [91]	2013	Combination of public	137	27	110
Acharya [92]	2012	Public (MESSIDOR)	180	-	-
Nayak [95]	2008	Local	140	-	-
Yun [96]	2008	Local	124	22	102
Priya [97]	2011	Local	-	-	-

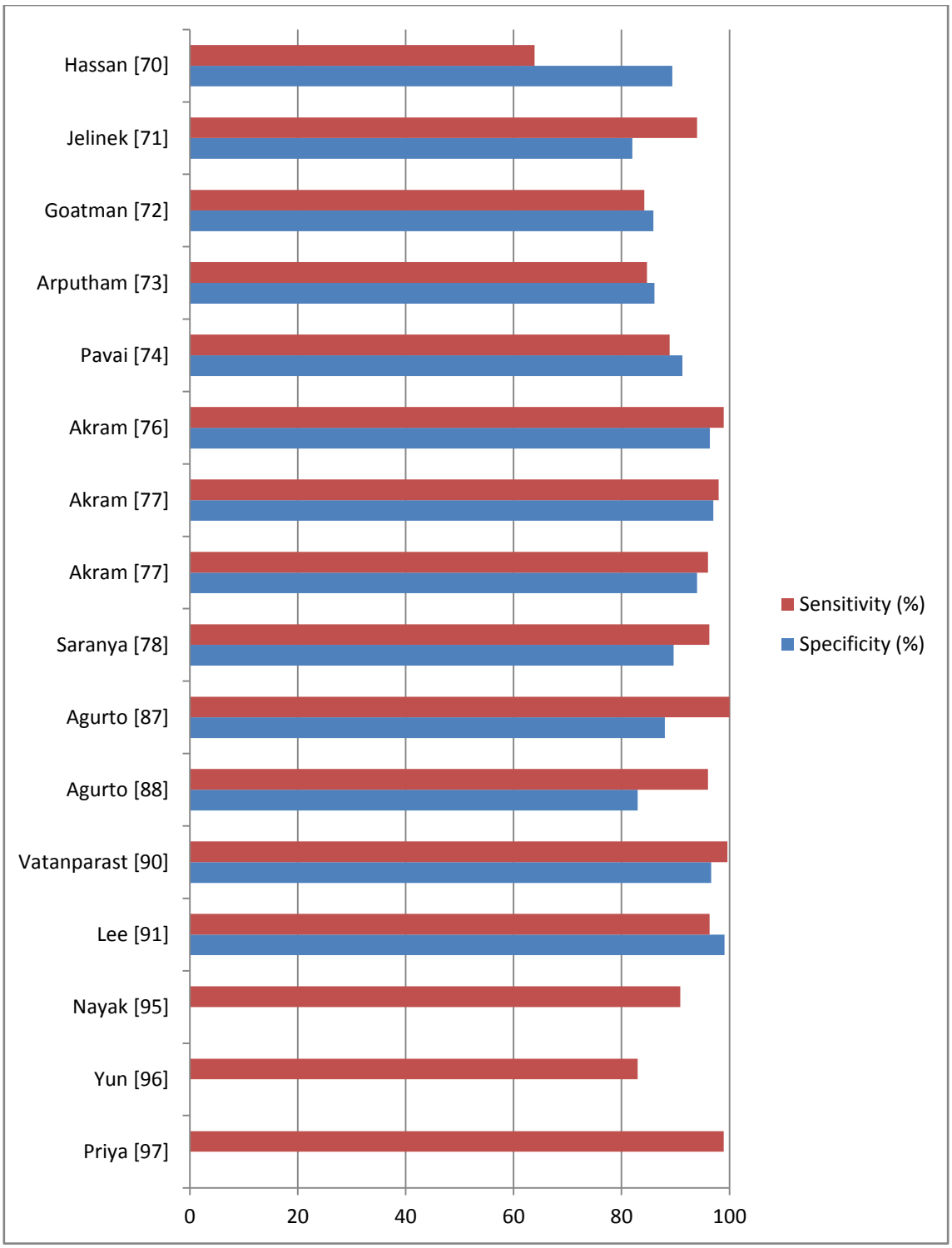


Figure 2.3: Sensitivity and specificity results of completed PDR detection methods reviewed in section 2.4.

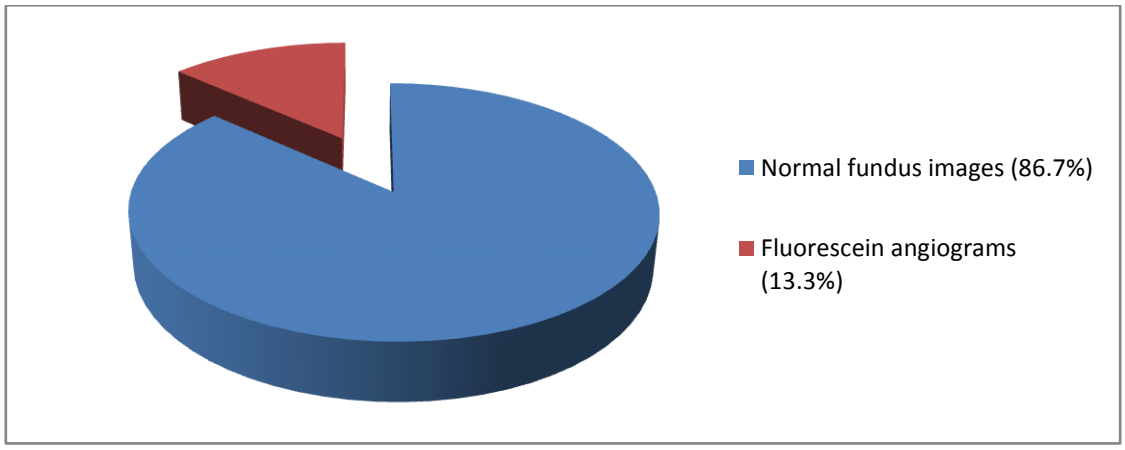


Figure 2.4: Type of images used by articles reviewed in section 2.4.

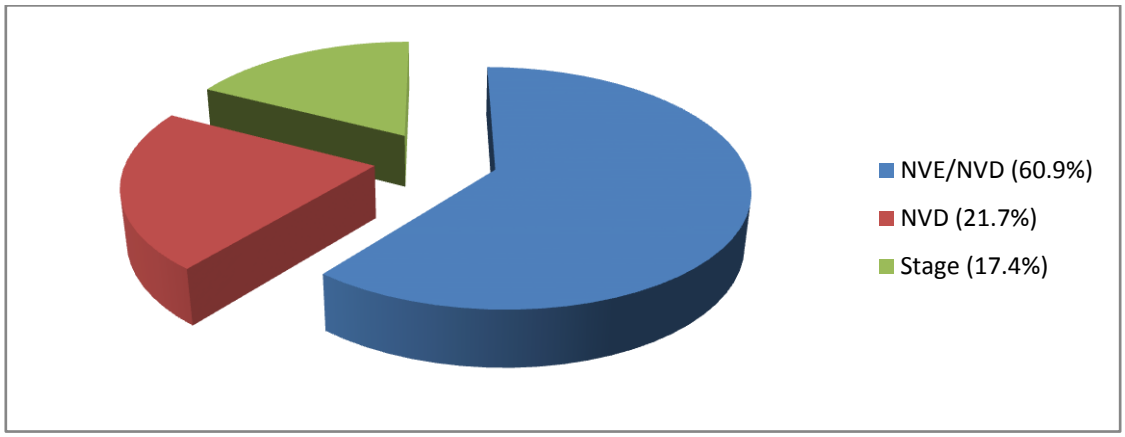


Figure 2.5: Type of PDR detection performed by articles reviewed in section 2.4.

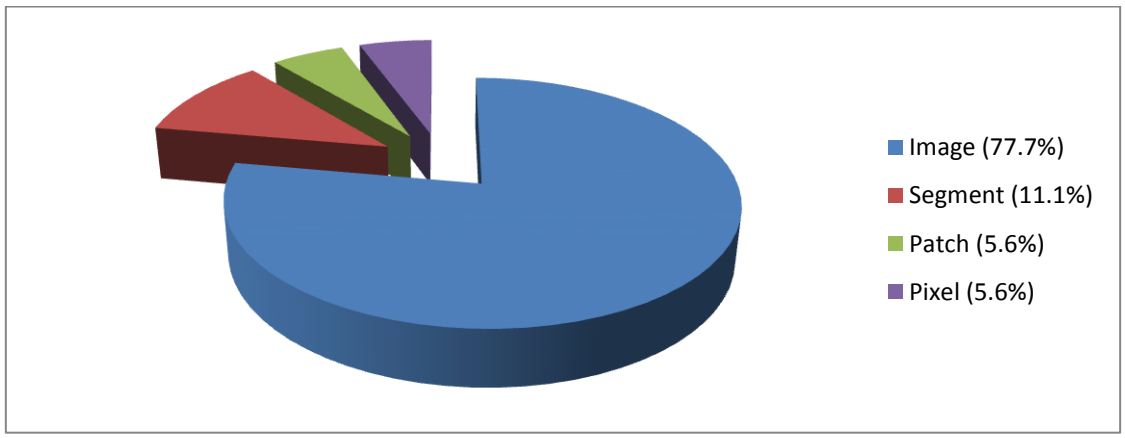


Figure 2.6: Levels selected for performance evaluation used by articles reviewed in section 2.4.

2.5 Machine Learning

The main focus of this literature review has been to detail the various image processing techniques involved in the field of retinal image analysis. The extraction of information/features has also been documented in detail. However, machine learning techniques that the features were used by have only been mentioned in passing. Therefore, this section will provide a brief overview of the main machine learning techniques used within this field. Just for clarity, we shall start with a basic definition of machine learning.

Machine learning is the construction and study of computer systems that can learn from data, and therefore act without explicitly programmed instructions. There are several different categories of machine learning, including (but not limited to) supervised learning and unsupervised learning. Classification is considered an instance of supervised learning, in which prior labelling information (achieved through training) is exploited to classify objects into known class labels. The corresponding unsupervised procedure is known as clustering, in which no class labels and no training data is provided. The goal of clustering is to unravel underlying similarities and group together similar objects.

A brief description of several machine learning techniques used in the field of retinal image analysis is given below. This includes the supervised learning techniques of the nearest neighbour classifier, the Bayes classifier, artificial neural networks and support vector machines, as well as the unsupervised learning technique of clustering.

Nearest Neighbour Classifier [106]**:** This is one of the simplest machine learning procedures that can be used for classification. It classifies an object based on the class of its nearest neighbour within the vector space defined by the features. A variant of the nearest neighbour algorithm is the k-nearest neighbour, where instead of finding just one nearest neighbour, k neighbours are found. The majority class of these k nearest neighbours is the class label assigned to the object.

The Bayes classifier: This uses the concept of probability to class objects. The classifier employs the posterior probabilities to assign the class label to an object. These probabilities are calculated by employing the Bayes theorem [107] which involves the use of prior probability and conditional probability. The naive Bayes classifier [108] simplifies matters by assuming the effect of a feature value on a given class is independent of the value of all other features in the feature vector.

Artificial neural networks [109]**:** This performs classification using networks of neurons based on the neural structure of the brain. Neurons are also known as nodes. The input to a node is weighted and summed and if this aggregate exceeds a threshold, the node outputs a signal which is sent as inputs to the next layer of nodes in the network. The progression of signals throughout the network leads to an output which assigns the object a class label. Training data is used to set the weights of the network.

Support Vector Machines [110]**:** Classification is performed by seeking a linear decision plane (hyperplane) that separates and creates the largest margin between two classes of objects. If the classes are not linearly separable the data is mapped into a higher dimensional space, where the separating linear decision surface exists and it is determined. The mapping is performed using a mathematical construction known as the kernel trick. There exist numerous different kernel functions and the selection of the most appropriate is important in order to maximise performance.

Clustering [111]**:** This is the task of grouping together similar objects, such that objects in the same group are more similar to each other than to those in other groups. Similarity is defined by the distance to the cluster centres in the vector space defined by the features. The procedure is an iterative process, where it often starts with an initial guess for cluster centres, and the centres of clusters get updated as objects are removed or added to them. Fuzzy C-means clustering is one of the more popular clustering algorithms. It is an overlapping clustering algorithm, where objects are grouped into a predefined number of clusters and can belong to every cluster to a certain degree (soft partitioning).

Table 2.6: The usage of machine learning techniques by articles reviewed in this chapter.

	Nearest Neighbour Classifier	Bayes Classifier	Artificial Neural Networks	Support Vector Machines	Clustering
Vessels	[36]	[37]	[34] [40]	[38]	[41] [42]
Microaneurysms and haemorrhages	[20]	[61]	[57]		
Exudates	[68]			[66]	[64] [65]
New vessels	[79] [78]	[76]	[94] [95] [96]	[72] [73] [74] [90] [92] [97]	[87] [97]

The choice of which machine learning technique to use is a critical step in the development of detection algorithms. Each technique has positives and negatives, and no one technique can be deemed to be superior in all cases. In terms of accuracy the support vector machines generally outperform the rest, followed by artificial neural networks. These two techniques also tend to be better at dealing with larger multi-dimensional feature vectors. The Bayes classifier and nearest neighbour classifier are better at dealing with the danger of overfitting. The Bayes classifier has an advantage in interpretability, as the acquired knowledge is transparent, whereas the nearest neighbour classifier, support vector machines and artificial neural networks are considered to be black boxes, as the acquired knowledge cannot be read in a comprehensible way. A comparison of machine learning algorithms is provided by Kotsiantis [112].

Table 2.6 provides a summary of the five main machine learning technique used by articles reviewed in this chapter. There are some methodologies that use more than one technique: Priya [97] uses fuzzy C-means clustering for vessel segmentation followed by support vector machines to classify the stage of DR. There are also methodologies (not listed in table 2.6) that use manually derived classification: Spencer [19] uses such an approach for the detection of microaneurysms and Hassan [70] does so for the detection of new vessels. Another approach which doesn't fall

into the main categories of table 2.6 is the use of an ensemble classifier of boosted and bagged decision trees for the segmentation of vessels [39].

Support vector machines are the newest supervised machine learning technique [110] and have been gaining popularity due to their promising performance. Aside from the qualities mentioned above, they have the ability to generate a good performance even with a small training set and they possess flexibility in modelling diverse sources of data. Sopharak [66] demonstrated that the support vector machine classifier outperformed the naive Bayes classifier and the nearest neighbour classifier for the detection of exudates. Table 2.6 shows the support vector machine to be the most popular technique for the detection of new vessels.

2.6 Discussion and Conclusion

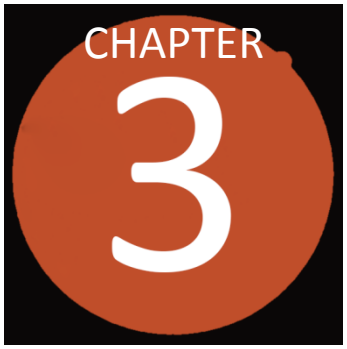
This chapter provides a survey of the current retinal image analysis methods developed for the automated detection of diabetic retinopathy. Vessel segmentation, as well as microaneurysm, haemorrhage and exudate detection are mature research fields. In contrast, little work has done to detect PDR despite its clinical importance.

The two main categories of PDR (new vessel) detection are segmented vessel map analysis and texture analysis, the former being the more popular, consisting of more published articles. The former required PDR vessel segmentation as an early step. However, this can be a very problematic task due to the tortuous and irregular nature of new vessels and requires more attention from the research community. Another significant problem with vessel segmentation are false positives caused by pathological lesions and camera artefacts. Following segmentation, the vascular map was analysed using mainly morphological based features along with intensity and gradient based features. Goatman [72] provided a comprehensive set of 15 features which were often reused by other articles. Aside from these features, the fractal

dimension is a measure that has received some attention in describing the morphology of the vascular map in order to detect new vessels.

Texture analysis benefits from not requiring vessel segmentation and hence does not have to deal with difficulties associated with it for new vessel detection. These methods were mainly based on the use of the grey level co-occurrence matrix (GLCM) for statistical texture analysis or amplitude modulation frequency modulation (AM-FM) for spectral texture analysis. Textural information can often be difficult to extract due to the lack of distinction that new vessels can exhibit.

From tables 2.3-2.4 and figure 2.3 the methodologies by Akram [76,77], Vatanparast [90] and Lee [91] appear to be superior in performance in terms of achieving operating points with reasonably high sensitivities and specificities. However, with no standard database for PDR detection evaluation, comparisons in performance can be difficult to make. Therefore, progress in this field would benefit from a publicly available PDR database. A range of machine learning techniques have been applied in the field of retinal image analysis, and table 2.6 shows that support vector machines is the most used technique for the detection of new vessels.



3 PROLIFERATIVE DIABETIC RETINOPATHY DETECTION USING A MODIFIED LINE OPERATOR AND DUAL SVM CLASSIFICATION

This chapter describes an automated method for the detection of new vessels in retinal images. Two vessel segmentation approaches were applied, using the standard line operator and a novel modified line operator. The latter, based on double sided thresholding, was designed to reduce false responses to non-vessel edges. Both generated binary vessel maps held vital information and were processed separately. This was achieved with a dual classification system, which is novel to retinal image analysis. Local morphology features were measured from each binary vessel map to produce two separate feature sets. Independent classification was performed for each feature set using a support vector machine (SVM) classifier. The system then combines these individual classification outcomes to produce a final decision.

The main purpose of the novel application of using two segmentation approaches and dual classification in retinal image analysis is to detect new vessels whilst reducing false responses caused by other common retinal features.

Work from this chapter was used to create the publication [113]. Results from this chapter do not correspond exactly to those reported in [113] due to a few alterations and additions to the methodology.

3.1 Methodology

The proposed system was adapted from the general approach of the use of vessel segmentation followed by analysis of the segmented vessel map to detect new vessels as documented in the literature review (section 2.4.2).

Extensive experimentation was undertaken in deriving the parameters and thresholds listed throughout the methodology. These were empirically derived by means of visual inspection. These parameters are relative to a retinal image size of 1479 x 1479 pixels.

3.1.1 Framework

The architecture of this system is shown is figure 3.1. Initial steps included spatial normalization to ensure the system's robustness with respect to image resolution and pre-processing to enhance the vasculature. Thereafter the system splits into two pathways for the two different vessel segmentation methods. Straight vessel removal was applied to remove large sections of normal vasculature. Structural analysis was applied to the resultant where local features associated with the morphology of the vasculature were measured. Each pathway had its own feature set produced, using the same set of local features. Independent classification was performed for each feature set using a linear support vector machine. The system produced a final decision by combining the two individual classification outcomes in which regions of the retina were labelled as new vessels or non-new vessels.

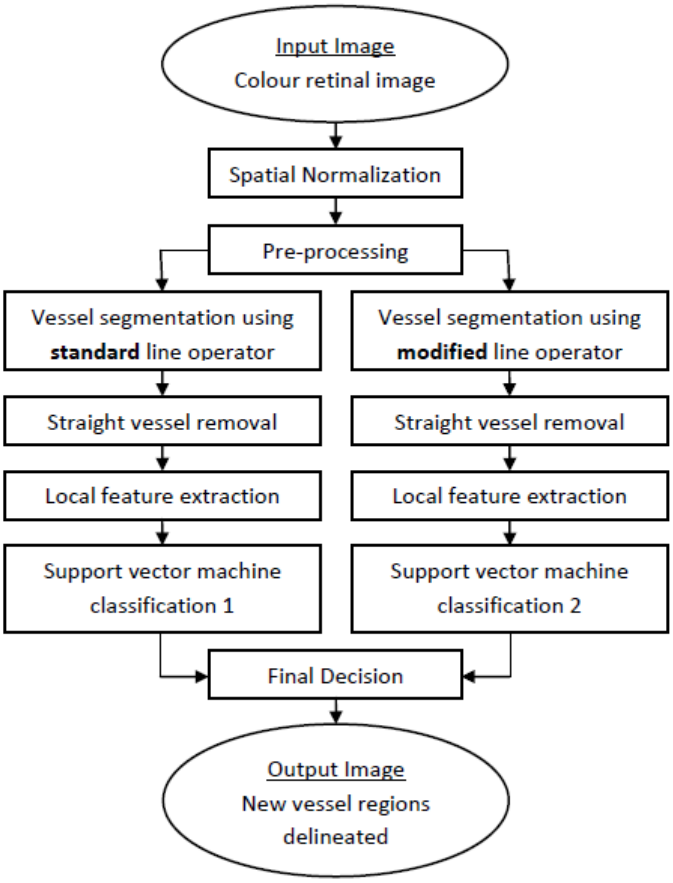


Figure 3.1: System architecture.

3.1.2 Pre-processing

Retinal images often show light variations, poor contrast and noise. This, along with the fact that new vessels tend to lack distinction meant that pre-processing was required. The green channel exhibits the best vessel/background contrast while the red and blue tend to be very noisy. Therefore the inverted green colour channel was used, where vessels appear brighter than the background. A median filter with a kernel size of 3 x 3 pixels was applied to reduce salt and pepper noise. Local contrast enhancement was achieved by applying contrast limited adaptive histogram equalisation (CLAHE) [21]. Shade correction was performed by subtracting an image approximating the background. This approximation was obtained by applying a median filter with a 105 x 105 pixel size kernel. This large size was chosen to ensure new vessel regions, as well as general structures, were preserved. A morphological top-hat transformation was used to produce an image containing small circular

objects (microaneurysms) [19] which were then subtracted from the image. The result of pre-processing is shown in figure 3.2. Figure 3.3 draws attention to the new vessel regions within the pre-processed image from the previous figure.

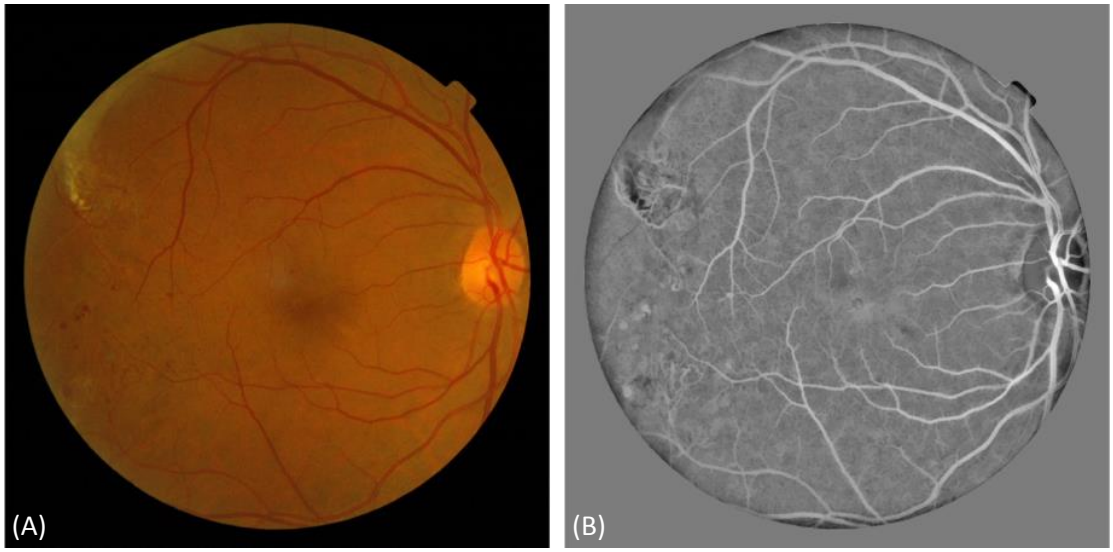


Figure 3.2: (A) Original image. **(B)** Pre-processed image.

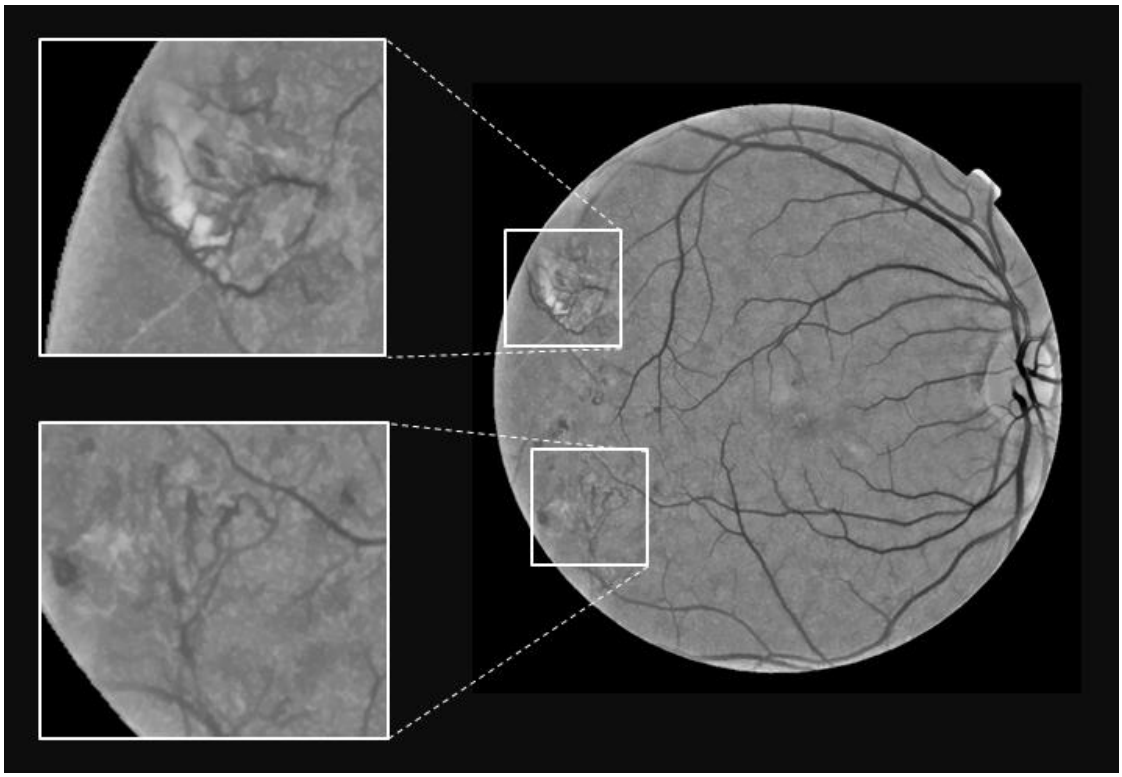


Figure 3.3: Magnified new vessel regions from a pre-processed image (inverted back for better visualization).

3.1.3 Line Operator /Modified Line Operator

The detection of linear structures has become a topic of significant importance in medical image analysis mainly due to the fact that vessels can be approximated as being piecewise linear. Ricci [38] applied line operators to detect linear structures in retinal images having been inspired by a method [114] which applied a line operator to detect linear structures in mammographic images.

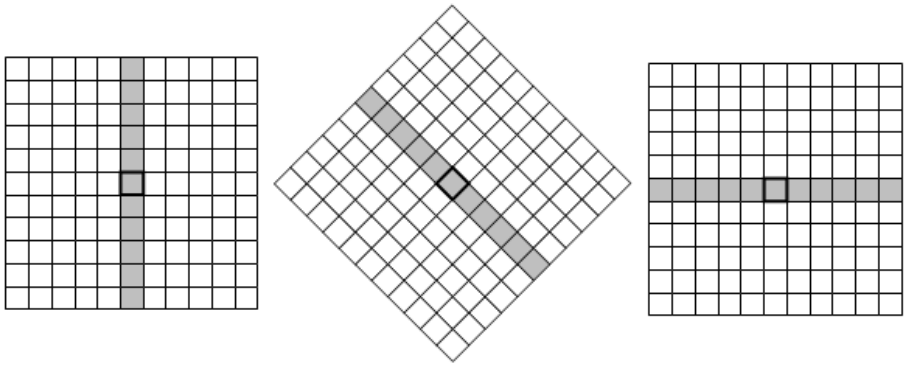


Figure 3.4: Standard line operator at 3 different orientations.

The standard line operator is illustrated in figure 3.4. The average grey-level of the pixels lying on a line passing through the target pixel (i, j) was calculated for multiple orientations. The orientation giving the largest value was found and its value was denoted with $\boldsymbol{L}(i, j)$. The line strength of the pixel, $\boldsymbol{S_1}(i, j)$, is given by

$$\boldsymbol{S_1}(i, j) = \boldsymbol{L}(i, j) - \boldsymbol{N}(i, j) \tag{3.1}$$

where $\boldsymbol{N}(i, j)$ is the average grey-level of the similarly orientated neighbourhood. Conversely to [38], the whole of the line operator window was orientated as opposed to keeping a fixed window and orientating only the line. At certain orientations, the line's path could not be exactly matched by the pixel grid, thus line and region averages were found by interpolation. Nearest neighbour interpolation was preferred due to its simplicity.

The line strength was large if the winning line was aligned with a vessel. In figure 3.5(E)-(H), the line strength images corresponding to the images in figure 3.5(A)-(D)

are shown. The line operator parameters of length 15 pixels and width 25 pixels were empirically derived to ensure an adequate response to new vessels was achieved (see appendix I). The line operator was applied over 12 different orientations (angular resolution of 15°). The square dimensions of the line operator from [38] did not have to be adhered to, the operator width was selected in accordance to vessel width and the operator length was selected to be relatively short as new vessels tend to be tortuous. An empirically derived threshold, T1, was applied to the line strength image to produce segmentation of the vessels. T1 was chosen to be relatively low to ensure that faint and obscure new vessel segments were retained. Figure 3.5(I)-(L) shows the binary vessel maps corresponding to the images in figure 3.5(A)-(D). The new vessels have been segmented with reasonable accuracy, thus meaning their properties could be adequately analysed. Considering some simple properties of new vessels such as high local density and large curvatures, it is evident from figure 3.5 that new vessels were distinguishable from normal vasculature.

A well-documented problem of vessel segmentation is that they respond not only to vessels but also to non-vessel edges. Bright lesions cause the most misclassifications. Areas of glare or reflection artefact, which are common on retinal images of younger individuals, also cause false responses. Figure 3.5(G)-(H) illustrates the strong line strength response to vessels as well as the edges of the bright lesions and figure 3.5(K)-(L) shows that after thresholding both the vessels and the edges of the bright lesions were detected. These false detections caused large local densities and large curvatures which were indistinguishable from new vessels.

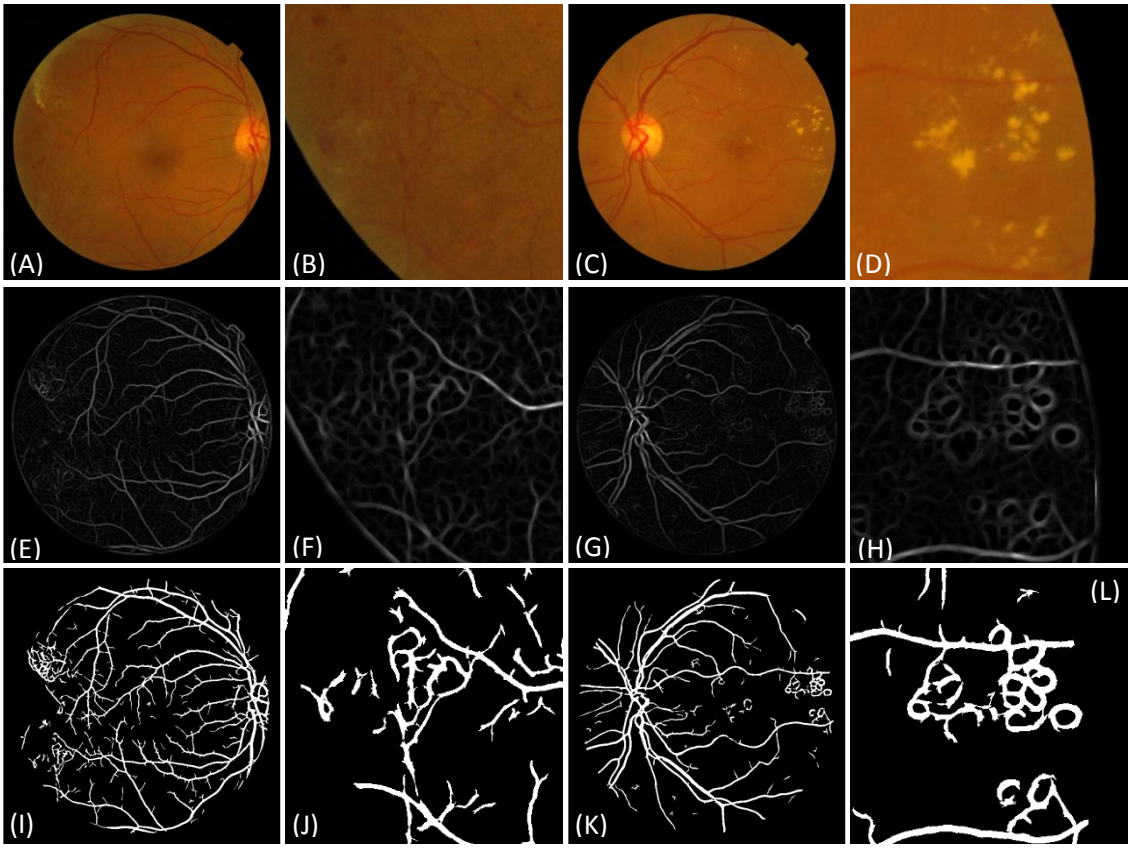


Figure 3.5: (A)-(D) Original images. **(A)** Retinal image with new vessels. **(B)** New vessels, zoom-in image of **(A)**. **(C)** Retinal image with bright lesions. **(D)** Bright lesions, zoom-in image of **(C)**. **(E)-(H)** Line strength maps corresponding to **(A)-(D)**. **(I)-(L)** Binary vessel maps corresponding to **(A)-(D)**, standard line operator.

A novel modified line operator was developed to reduce the false responses to non-vessel edges, inspired by L.Zhang [27] where a matched filter with double sided thresholding was proposed. Three modified line strength measures were derived,

$$\boldsymbol{S_2}(i,j) = \boldsymbol{L}(i,j) - \boldsymbol{N_R}(i,j) \tag{3.2}$$

$$\boldsymbol{S_3}(i,j) = \boldsymbol{L}(i,j) - \boldsymbol{N_L}(i,j) \tag{3.3}$$

$$\boldsymbol{S_4}(i,j) = \boldsymbol{L}(i,j) - \boldsymbol{N_B}(i,j) \tag{3.4}$$

where $\boldsymbol{N_R}(i,j)$ is the average grey-level of just the right side of the similarly orientated neighbourhood, $\boldsymbol{N_L}(i,j)$ is the average grey-level of just the left side of the similarly orientated neighbourhood and $\boldsymbol{N_B}(i,j)$ is the median value of a large neighbourhood (not orientated). Figure 3.6 illustrates these three measures. The

operator parameters of length and width remained unchanged, the size of the large neighbourhood was set to 151 x 151 pixels. An empirically derived threshold, T2, was applied to all 3 modified line strength measures and followed by the logical AND operator to define the pixel as a vessel.

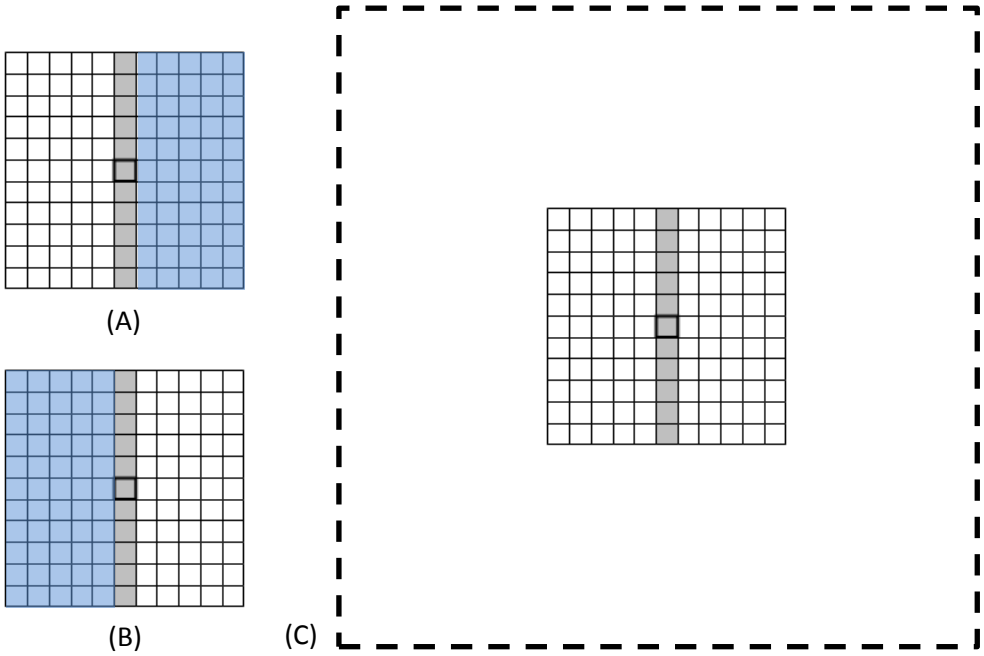


Figure 3.6: Modified line operator. **(A)** Average grey-level of the right side, $\boldsymbol{N_R}(i,j)$. (B) Average grey-level of the left side, $\boldsymbol{N_L}(i,j)$. **(C)** Median value of a large neighbourhood, $\boldsymbol{N_B}(i,j)$.

To simplify the explanation of the modifications, figure 3.7 shows a 1D cross section of a vessel (left of the trace) and a bright lesion (right of the trace) from a pre-processed image. The thresholding of the line strength done in the standard line operator approach was simply stating that $\boldsymbol{L}(i,j)$ must be a value T1 greater than the grey-level average of the similarly orientated neighbourhood. Looking at the trace it is clear that the vessel points are significantly greater than its local neighbourhood. Unfortunately, this is also this case for the edges of the bright lesion due to the large dip in value caused by the bright lesion. This was the cause of the false responses from the standard approach. The trace shows that the difference between the two cases is that vessels are significantly greater than their local neighbourhood on both sides, unlike non-vessel edges. The first modification was that now $\boldsymbol{L}(i,j)$ was

required to be a value of T2 greater than the grey-level average of both the right and the left side of the similarly orientated neighbourhood, which rectified this problem. However, consider two bright lesions in close proximity. The space in-between them is significantly greater than both sides. We could distinguish this case by the fact that the pixel value of these points are likely to be similar to that of the retinal background. The median value of a large neighbourhood was used to calculate the retinal background value. Therefore the next modification was that $\boldsymbol{L}(i,j)$ was required to be a value of T2 greater than the local retinal background value.

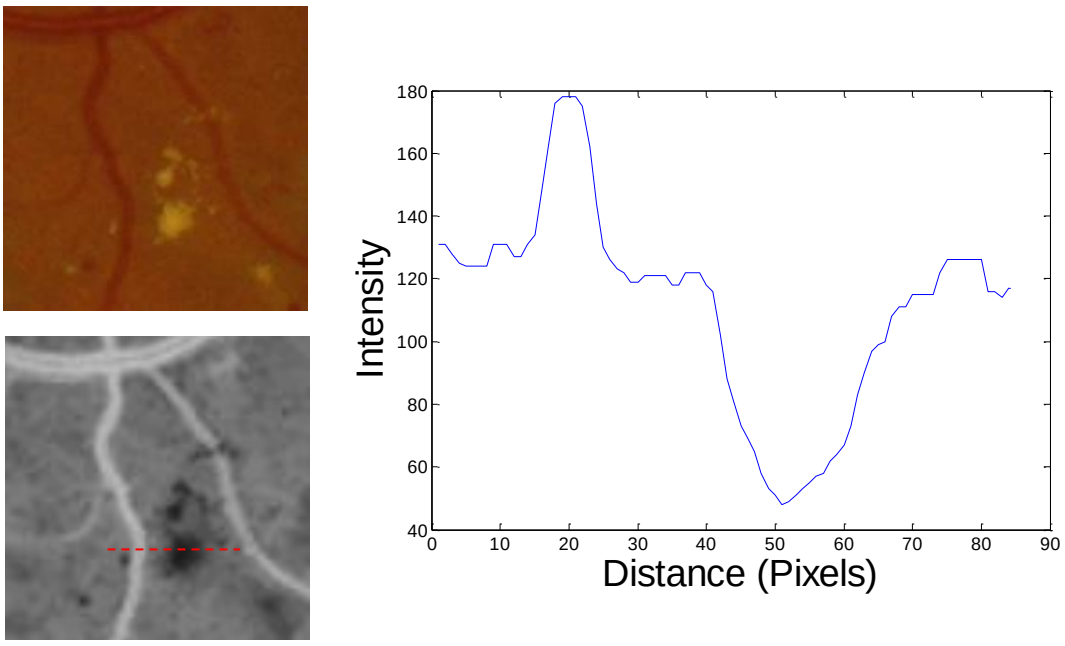


Figure 3.7: Cross section of a vessel and a bright lesion.

The binary vessel maps generated by the modified line operator approach are shown in figure 3.8. The false responses caused by non-vessel edges were now significantly reduced to the extent that non vessel edges were distinguishable from new vessels. Unfortunately the segmentation of new vessel had become slightly degraded. This is a trait of all vessel segmentation techniques: the more emphasis put on the reduction of non-vessel responses, the greater the risk of damage to the segmentation of the vessels. This meant that new vessels were no longer so distinguishable from the normal vasculature.

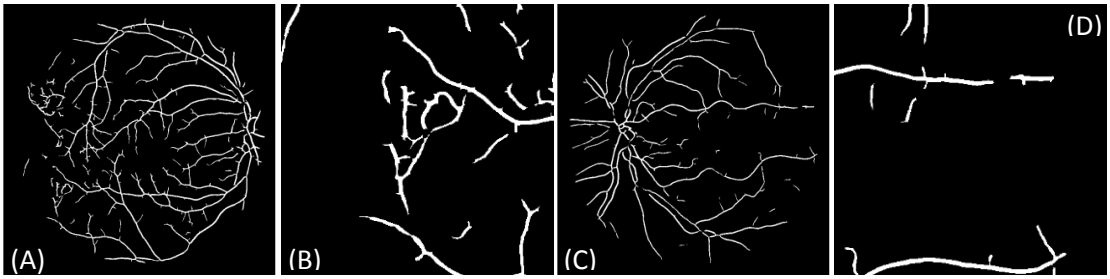


Figure 3.8: Binary vessel maps corresponding to the images in figure 3.5(A)-(D), modified line operator.

Both of the segmentation methods showed disadvantages and therefore neither method alone was suitable for the detection of new vessels. However, each of the produced binary maps held vital information. The standard approach provided the information to distinguish new vessel from normal vasculature and the modified approach provided information to distinguish new vessels from non-vessel edges. Extraction of information from both maps could be used effectively to detect new vessels, and therefore both segmentation methods were applied. Further examples of the results from both segmentation methods are provided in figure 3.9.

The segmented results shown in the figures so far also included an additional step to remove any falsely detected microaneurysms and haemorrhages, known as dark lesions. The line operator parameters and the low threshold that were selected to ensure increased sensitivity to new vessels also caused an increased sensitivity to red lesions. A simple measure of circularity (see equation 3.5) and area from the objects in the binary vessel map was used to distinguish dark lesions and other spurious objects in order to remove them.

$$Circularity = 4\pi.area/perimeter^2 \tag{3.5}$$

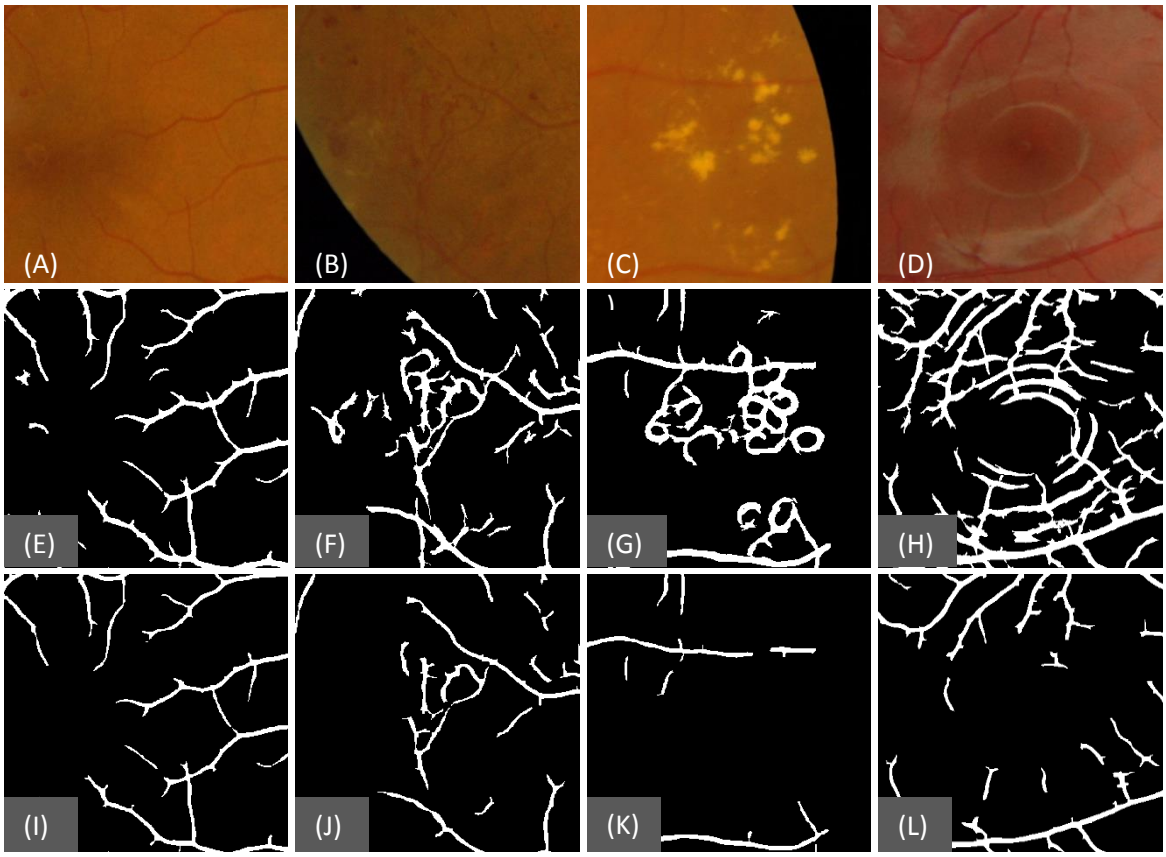


Figure 3.9: (A) Normal vessels. **(B)** New vessels. **(C)** Exudates. **(D)** Reflection artefacts. **(E)-(H)** Binary vessel map of **(A)-(D)** using the standard line operator. **(I)-(L)** Binary vessel map of **(A)-(D)** using the modified line operator.

3.1.4 Straight Vessel Removal

Many sections of the normal vasculature possess high local densities, which occur at or near the optic disc, at bifurcation points and at crossover points (see figure 3.10). Therefore the claim that, from the binary vessel map, new vessels could be distinguished from the normal vasculature was only possible if large sections of the normal vasculature were first removed. This was done using a novel technique proposed by Welikala [79], which involves the segmentation of the straight vessels. The standard line operator was applied as before, but the operator length was increased to an empirically derived length of 81 pixels (see appendix I). The operator was no longer sensitive to the tortuous vessels and was instead only sensitive to

relatively straight vessels. An empirically derived threshold, T3, was applied and the resultant straight vessel maps are shown figure 3.11(C)-(D).

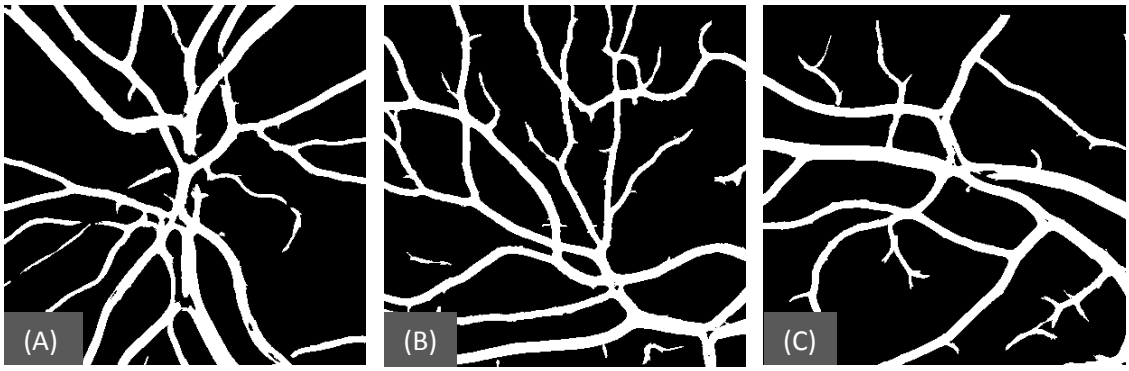


Figure 3.10: Segmented normal vasculature with high local densities. **(A)** Optic disc. **(B)-(C)** Bifurcation points and crossover points.

The binary vessel maps from the standard and modified approach (from section 3.1.3) were both skeletonised by means of morphological thinning, to ensure they were thinner than the straight vessel map. Following this the straight vessel map was subtracted from each. The following sections will describe how the binary vessel maps, which now contained only partial vasculature, were assessed for new vessels.

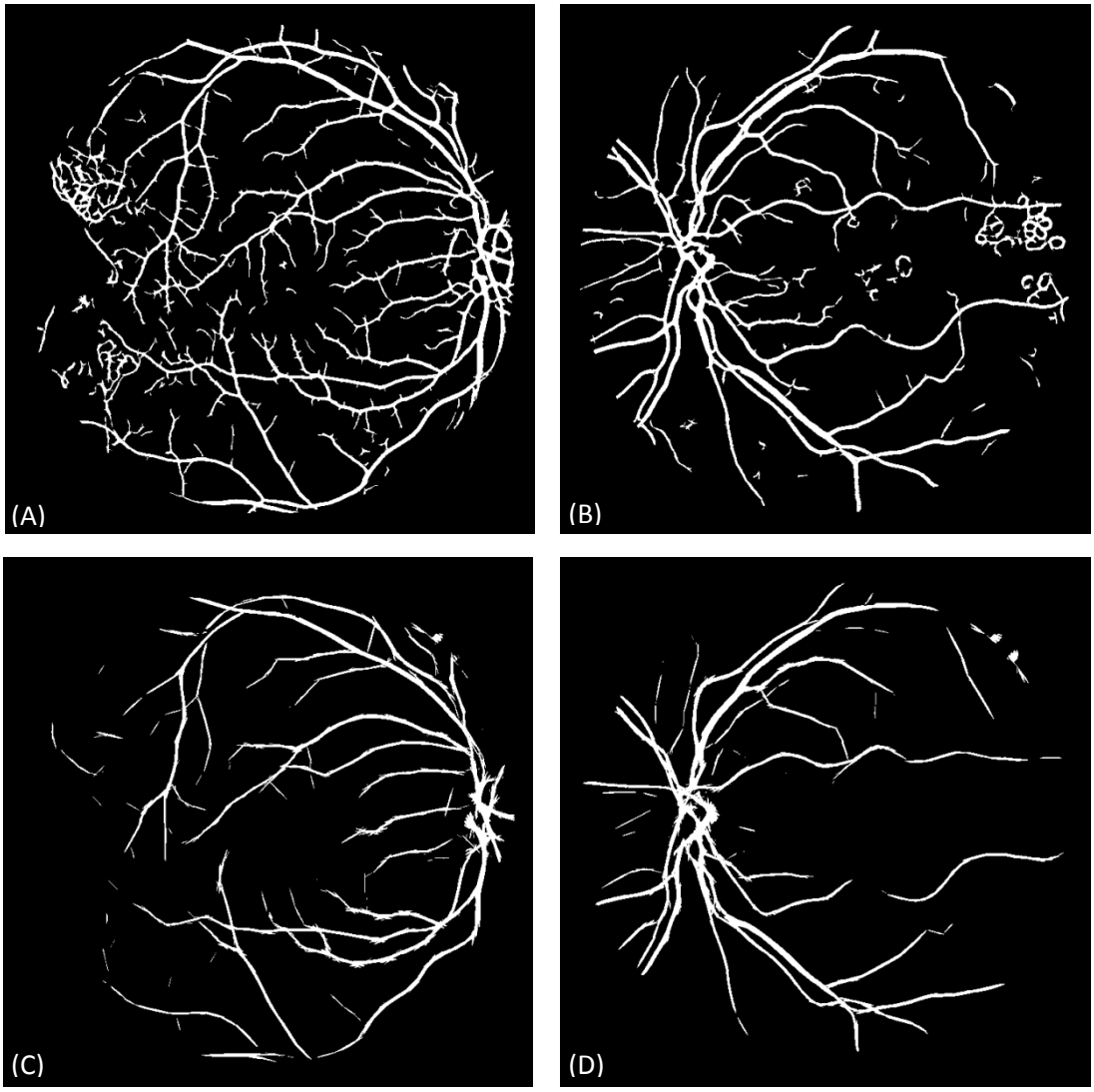


Figure 3.11: (A)-(B) Vessel maps created using the standard line operator with a length of 15 pixels. **(C)-(D)** Straight vessel maps corresponding to (A)-(B), created using standard line operator with a length of 81 pixels.

3.1.5 Feature Extraction

The design of this method is aimed at the classification of image regions that contain new vessels. These image regions can be described as containing many vessel segments, which are closely spaced, appear in multiple orientations and possess a tortuous nature. There was no intention in this work to identify individual new vessel pixels or segments.

The binary vessel maps were converted into vessel segments prior to measurements of features. The vasculature was a single pixel in thickness as skeletonization had been performed in the previous stage. Vessel segments were created by removing bifurcation points which were pixels with more than two eight-way neighbours. Finally, small segments consisting of fewer than ten pixels were discarded.

A sub-window of size 151 x 151 pixels was created in order to calculate local features associated with the morphology of the vasculature. This sub-window was scanned through the image and at increments of 20 pixels (reduces overlapping), candidate pixels were selected. At each candidate pixel position four features were calculated. The same set of features was measured from the binary vessel map from each of the standard and modified approaches to produce two separate feature sets for each candidate pixel.

The description of the features is listed below. The equations for these features are detailed in chapter 4, section 4.1.1. The term segment refers to the skeletonised vessel segments. All segments were labelled.

1) **Number of vessel pixels**

   The sum of all segment pixels within the sub-window.

2) **Number of vessel segments**

   The number of whole and partially included segments within the sub-window.

3) **Number of vessel orientations**

   The end points of a segment were connected by a straight line. The angle the line made with the x-axis that fell within the range -90° to 90° of the unit circle was calculated. The calculated angle was accordingly dropped into one of eight bins, each representing a range of angles. This was done for each segment within the sub-window and the number of non-empty bins represented the number of orientations.

4) **Vessel density mean**

A segment was dilated with a disk structuring element with a radius of 20 pixels. The number of pixels from all segments within the dilated area was divided by the number of pixels within the segment to give its vessel density. This was done for each segment within the sub-window and the mean vessel density was calculated.

3.1.6 Dual SVM Classification

The conventional approach for classification is an individual classifier that uses a single feature set. Ensemble based systems combine multiple classifiers in order to enhance the performance of the individual classifier. Fraz [39] described a multiple classifier approach using bagging and boosting techniques, which used a single feature set. Of more relevance to this work are the multiple classifier approaches that use multiple feature sets. Polikar [115] describes an overview of multiple classifier systems. This includes data fusion where the nature of features are different (heterogeneous features). Therefore the features are separately grouped and separately classified, followed by the combination of the outcomes. Chim [116] proposed a dual classification system that used two different feature sets. The features could have been combined to produce a single feature set, but to achieve a better performance they were kept separate and independent classifications were performed which were then combined to produce a final decision.

The proposed methodology adopted a dual classification approach. All features were normalised so that each feature had zero mean and unit standard deviation. Independent classification was performed for each of the two feature sets using a support vector machine (SVM) classifier [110,117]. The optimal SVM kernel and parameters were determined by a cross-validation grid search, with the kernel types of linear, Gaussian radial basis function and polynomial being tested. Each classifier independently labelled the candidate pixel as new vessels or non-new vessels. The system produced a final decision by combining the outcomes. The candidate pixel

achieved a new vessel label only when both classifications agreed on its identity being new vessels; otherwise it achieved a non-new vessel label. When complete, all candidate pixels labelled as new vessels were morphologically dilated with a structuring element the size of the sub-window (used in feature extraction) to illustrate the new vessel regions.

Whilst only two classes were used, new vessels and non-new vessels, both feature sets and their independent classification were not intended to distinguish the same two cases. Classifier 1, associated with the feature set measured from the standard line operator approach, was intended to distinguish new vessels from normal vessels. Classifier 2, associated with the feature set measured from the modified line operator approach, was intended to distinguish new vessels from exudates/reflections. Combining the outcomes then removed the false new vessel responses that each classifier made.

Consider that there are three classes, new vessels, normal vessels and exudates/reflections. Hence, a one-versus-one multiclass SVM has similarities to our approach as it is based on independent classifications between each of the classes. However, it would be disadvantaged as each classification would be made within a combined single feature space, whereas the dual approach used only the features relevant to each classification.

To validate our dual classification system its performance was compared to the performance from a single SVM classification and a one-versus-one multiclass SVM classification. Note that for both of these approaches the two feature sets have been combined to create a single feature set.

SVM classification was adopted in this project as it is a state-of-the art method which is reported to possess a good generalization performance even with a small training set, has the ability to deal with high dimensional data and the flexibility in modelling diverse sources of data [117]. Section 2.5 states how SVM classification has gained in popularity due to its high performance accuracy and it has become the most popular technique for new vessel detection.

We end this subsection by briefly touching on the main principle of SVM classification. SVMs seek a linear decision surface (hyperplane) that can separate classes of objects (solid red line in figure 3.12). In a d-dimensional space, the hyperplane can be represented by

$$\vec{w} \cdot \vec{x} + b = 0 \tag{3.6}$$

were $\vec{x}$ and $\vec{w}$ are d-dimensional vectors, $\vec{x}$ is the feature vector, $\vec{w}$ is a weight vector and b is a scalar. Two classes are linearly separable if we can find $\vec{w}$ and b such that

$$\vec{w} \cdot \vec{x} + b > 0 \tag{3.7}$$

for all objects belonging to one class

$$\vec{w} \cdot \vec{x} + b < 0 \tag{3.8}$$

for all objects belonging to the other class.

An infinite number of such hyperplanes may exist. SVMs find the hyperplane that maximizes the distance (gap or margin) between the border-line objects (that are also called support vectors). The two hyperplanes that pass through the support vectors (dashed black lines in figure 3.12) are parallel to $\vec{w} \cdot \vec{x} + b = 0$, and therefore they will only differ in the scalar b value. After rescaling $\vec{w}$ and b by the same factor, the two hyperplanes are given by equations of the form

$$\vec{w} \cdot \vec{x} + (b - 1) = 0 \tag{3.9}$$

$$\vec{w} \cdot \vec{x} + (b + 1) = 0 \tag{3.10}$$

Or equivalently

$$\vec{w} \cdot \vec{x} + b = +1 \tag{3.11}$$

$$\vec{w} \cdot \vec{x} + b = -1 \tag{3.12}$$

and the distance (D) between these two planes is given by formula in equation 3.13. Full details on how this is derived are provided by [118].

$$D = |b_1 - b_2| / \|\vec{w}\| \tag{3.13}$$

Therefore

$$D = 2/\|\vec{w}\| \tag{3.14}$$

Since we want to maximize the gap between the two planes, we want to minimize$\|w\|$. This is an optimization problem.

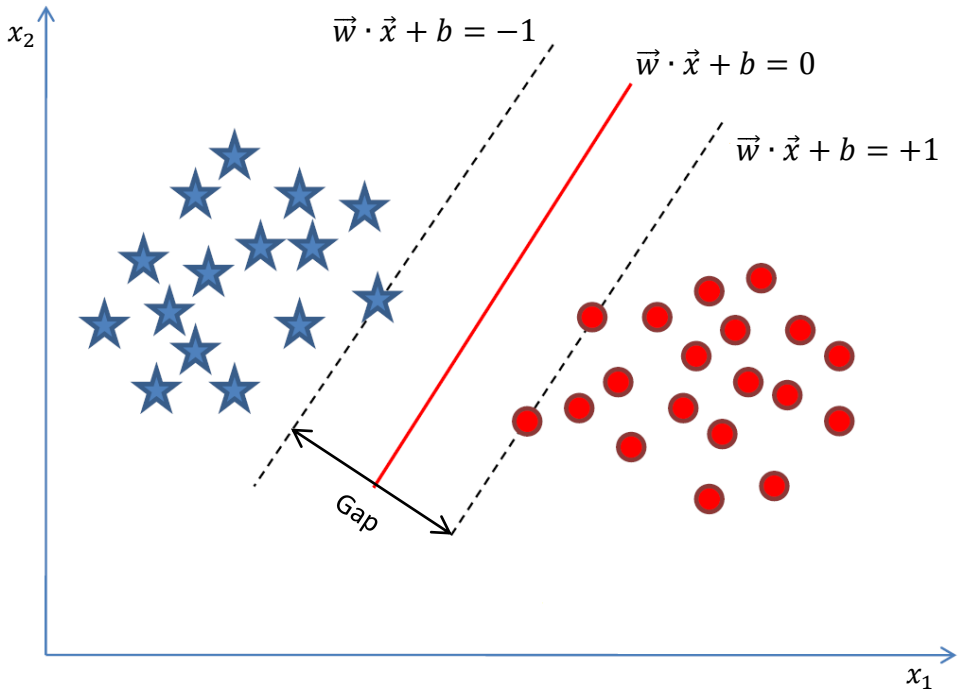


Figure 3.12: Illustration of SVMs. Linear decision surface that can separate the two classes and has the largest gap between border-line objects. Image adapted from [118].

If the classes are not linearly separable the SVM maps the data into a higher dimensional space known as the feature space, where the separating linear decision surface exists and it is determined (see figure 3.13). The feature space results from a mathematical construction known as the kernel trick. There are numerous different kernel functions. Besides the standard linear kernel, the most popular kernel functions are Gaussian radial basis function kernel (rbf) and the polynomial kernel.

The majority of kernels possess parameters which need to be selected. Another parameter associated with SVMs is the soft margin parameter C to deal with noisy measurements and outliers. Therefore, the effectiveness of SVMs depends on the

selection of the kernel function, the kernel's parameters and the soft margin parameter. A grid search is the conventional approach for selecting the optimal SVM parameters, and as stated above this was used in our proposed method.

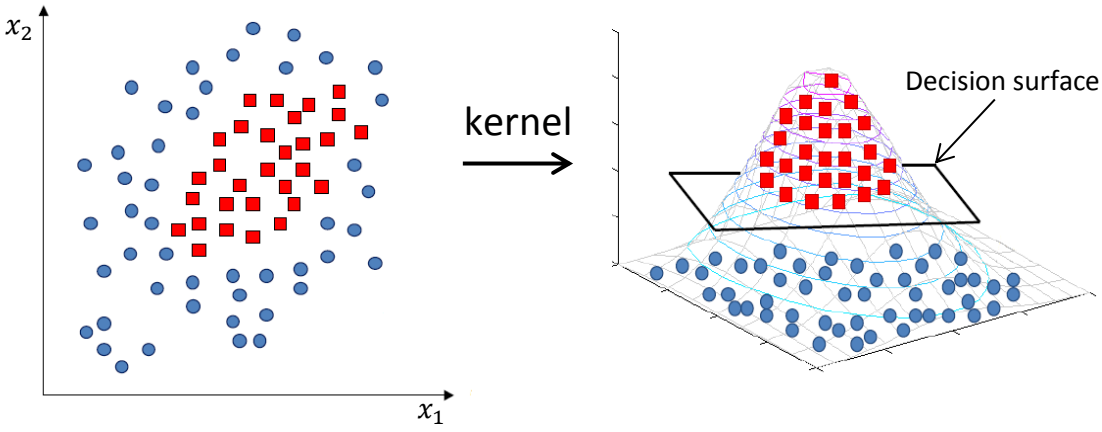


Figure 3.13: Illustration of SVMs. The kernel trick maps data into a higher dimensional space where a separating decision surface is found. Images adapted from [118].

3.2 Experimental Evaluation

3.2.1 Materials

Due to the low prevalence of new vessels in the screening population, the proposed method was evaluated using images collected from two sources (public and local). These were the publicly available MESSIDOR retinal image database, provided by the Messidor program partners [16], and the St Thomas' Hospital Ophthalmology Department. A total of 60 images (see appendix II) were included in the dataset: 20 with confirmed new vessels and a further 40 images without new vessels. The image data from each source were as follows:

1) MESSIDOR: 5 new vessel images, 20 normal images and 20 images with the large majority showing other DR pathology (mainly bright lesions) and the remainder showing strong reflection artefacts. These images were acquired

from a colour video 3CCD camera on a Topcon TRC NW6 fundus camera with a 45 degree field of view (FOV) and an image resolution of 2240 x 1488 pixels.

2) St Thomas' Hospital: 15 new vessel images acquired with a Nikon D80 digital SLR camera on a Topcon TRC NW6 fundus camera with a 45 degree FOV and an image resolution of 2896 x 1944 pixels. Ethical approval was obtained for the use of these images.

Images were scaled to the same size using a spatial normalization technique proposed by [119] along with bicubic interpolation and anti-aliasing. This was based on normalizing the FOV width, with the requirement that all images were captured with the same FOV angle. All images were normalized to have a FOV width of 1379 pixels. This was followed by cropping to remove some of the surrounding black border to produce images of size 1479 x 1479 pixels.

For training data, a specific selection of pixels was chosen from the dataset. It can be convenient to refer to a pixel as an image patch, considering features were extracted using information from the local neighbourhood contained within the sub-window centred over the target pixel. These image patches were labelled as either new vessels or non-new vessels by an ophthalmologist. Separate training data was used for each classifier. Classifier 1 was trained with 50 new vessel patches and 50 normal vessel patches. Classifier 2 was trained with 50 new vessel patches and 50 patches made up of a variety of bright lesions, dark lesions and reflection artefacts. Note that each new vessel patch correlated to a pixel selected in the centre of a new vessel region.

Testing was performed across the whole of each retinal image, in terms of the classification process being performed at every candidate pixel location (increments of 20 pixels). Because of the limited size of the dataset, splitting the data set to create separate training and testing sets was not suitable. Instead, both the training and testing sets were created using the same selection of images (entire dataset). Evaluation in this manner would clearly cause over fitting of the model, and therefore the leave-one-out cross validation method [120] was applied. This meant the classifiers were trained using all the selected patches from all the images except

those from the single test image, and this process was repeated for each image. The feature value normalization was also recalculated each time, leaving out the test image.

As mentioned above, the system made decisions on a pixel basis using information extracted from the local neighbourhood. Therefore, in order to visualise these local neighbourhoods, all positive candidate pixels were then morphologically dilated with a structuring element the size of the sub-window used in feature extraction. This resulted in the delineation of the new vessel regions. This procedure is illustrated in figure 3.14. However, the performance from a per image basis is more useful from a clinical point of view. An image simply achieved a new vessel label if it contained any delineated regions. Prior to this, all images had been labelled by an ophthalmologist using the same labels as before (new vessels and non-new vessels) but on a per image basis.

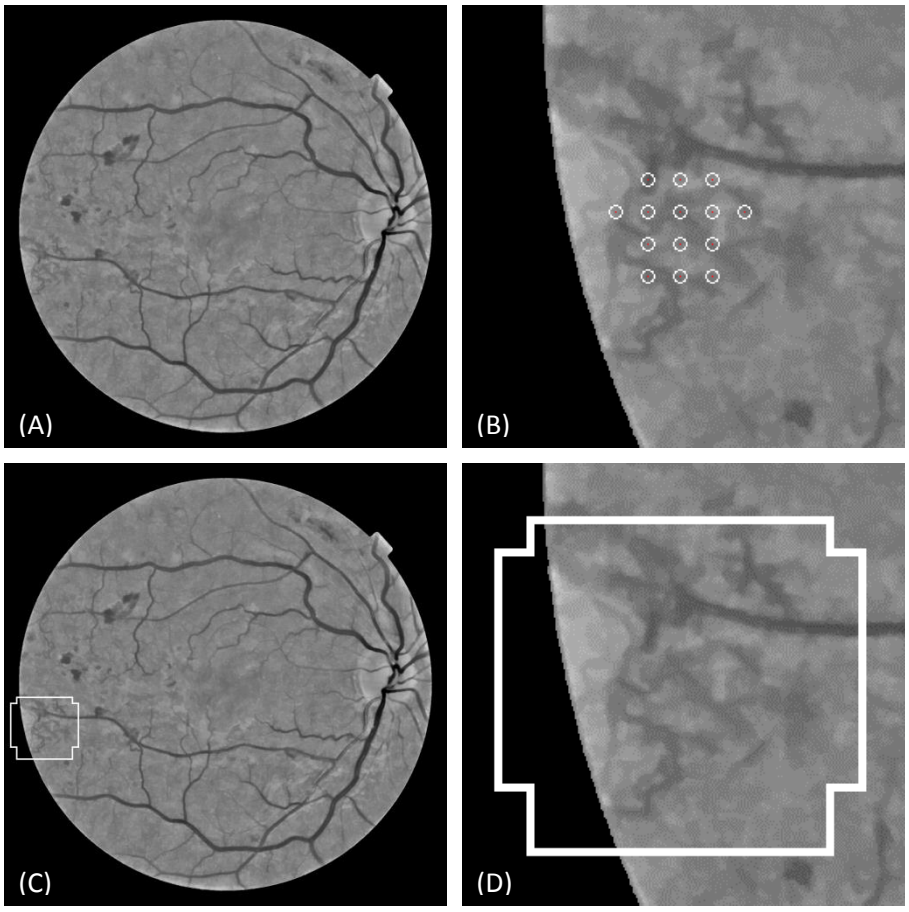


Figure 3.14: (A) Positive candidate pixels marked in red, not visible without zoom. **(B)** Zoom-in region of **(A)**, each positive candidate pixels is indicated with a circle (for the purpose of visualization in this figure only). **(C)** Positive candidate pixels are dilated with a square

structuring element of size 151 x 151. Only the outline of the delineated region is retained. **(D)** Zoom-in region of **(C).**

To get a more detailed insight into the system's performance the obvious choice would be to also evaluate the performance of the delineation of new vessel regions. This involves assessing the outcome for every candidate pixel across all images. However, the performance of delineation was not evaluated and the reason why shall be discussed in section 3.3. Instead, an alternative evaluation was achieved by performing testing on the selection of pixels used for training using the leave-one-out cross validation method. A few additional pixels from outside the training data were also tested. As stated above a pixel can be referred to by the term patch. Therefore, we refer to this assessment as evaluation on a per patch basis.

3.2.2 Performance Measures

As mentioned above, the performance was separately assessed on a per image and per patch basis. Any image or patch was classified as either new vessels or non-new vessels. Consequently there are four outcomes, two classifications and two misclassifications which are defined in table 3.1(A). The algorithm was evaluated in terms of sensitivity (SN), specificity (SP) and accuracy (Acc). These are often used in machine learning and are measures of the quality of binary classification. These metrics are defined in table 3.1(B).

Table 3.1: (A) New vessel classification. **(B)** Performance measures for new vessel detection.

	New vessels present	**New vessels absent**
New vessels detected	True positive (TP)	False positive (FP)
New vessels not detected	False negative (FN)	True negative (TN)

(A)

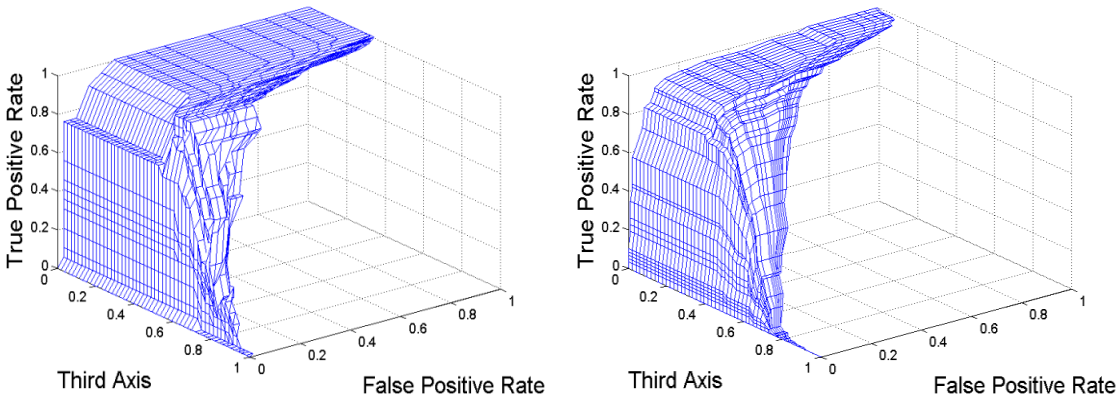
Measure	**Description**
SN	TP/(TP+FN)
SP	TN/(TN+FP)
Acc	(TP+TN)/(TP+FP+TN+FN)

(B)

The use of the receiver operating characteristic (ROC) curve allows for the visualization of the performance of a binary classifier system, expressing the trade-off between increased detection and false alarms. This was created by plotting the true positive rate (SN) versus the false positive rate (1-SP) at various threshold levels of the probability score of the classifier. The SVM calculated a new vessel probability score using the distance to the decision boundary.

With a dual classifier approach and therefore two probability scores, the creation of ROC curves was not a straight-forward task. This problem was tackled by creating a 3D ROC surface. The majority of the literature relating to higher dimensional ROC analysis relates to multi-class analysis [121,122] and are not closely related to our work. Of more relevance are studies that add a third axis which represents varying the threshold of an additional parameter [123]. To the best of our knowledge, there appears to be no available literature concerning ROC analysis for dual classifiers.

The addition of a third axis to the conventional 2D ROC plot accommodates for varying the threshold of the probability score of the additional classifier that arises in the dual classification approach. The resultant was a 3D ROC surface that explored all combinations of thresholds for the dual classification. Figure 3.15 shows the 3D ROC surface representing performance of the proposed method on both a per image and per patch basis.



(A) (B)

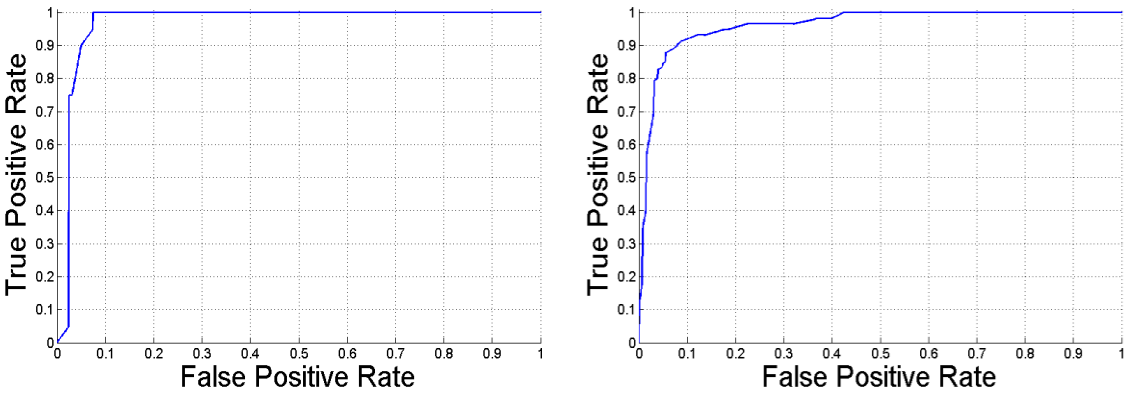
Figure 3.15: 3D ROC surfaces. **(A)** Performance on a per image basis. **(B)** Performance on a per patch basis. The third axis is required to vary the threshold of the probability score of the additional classifier that arises in a dual classification approach.

Information from this 3D ROC surface was extracted to create a 2D ROC curve. For each false positive rate value its maximum true positive rate value was found by searching along the third axis of the 3D ROC surface. From the 2D plot generated, the area under the curve (AUC) was extracted and used as a performance measure.

The operating point that achieved the highest accuracy (known as maximum Acc) from ROC curve was selected as the optimal operating point. However, this was only the case for per patch assessment. For per image assessment it was evident that the algorithm could reach a sensitivity of 100% at a high specificity. This was important as from a clinical point of view a sensitivity of 100% was considered an essential requirement from a per image basis. Maximum accuracy may not always equate to an operating point with a sensitivity of 100%. Therefore an application specific performance measure was created for the per image basis, in which the operating point with the highest specificity at a sensitivity of 100% was selected as the optimal operating point.

3.2.3 Results

The ROC curves for performance of the proposed method on a per image and per patch basis are depicted in figures 3.16(A)-(B). The AUC value for the per image basis is 0.9693. The optimal operating point according to the application specific performance measure is a sensitivity of 100.00% and a specificity of 92.50%. For a per patch basis the AUC value is 0.9616. The operating point with maximum accuracy of 92.35% gives a sensitivity of 87.93% and a specificity of 94.40%. These results, along with the grid search determined optimal kernel and parameters, are presented in table 3.2.



(A) (B)

Figure 3.16: ROC curves. **(A)** Performance on a per image basis. **(B)** Performance on a per patch basis.

Table 3.2: Performance analysis of the proposed method.

Method	Level	SN	SP	Acc	AUC	Kernel	Parameters
Proposed Method	Image	100.00	92.50	95.00	0.9693	Linear	C=0.01
Proposed Method	Patch	87.93	94.40	92.35	0.9616	Linear	C=1

The performance of the proposed dual classification was assessed against single classification and multiclass classification, and the results are presented in table 3.3.

The table also lists the SVM kernel type and SVM parameters selected to reach optimal performance for each classification method, determined using a grid search. The comparison of classification methods is assessed on a per image basis.

Table 3.3: Comparison between single, multiclass and dual SVM classification on a per image basis.

Classification Method	SN	SP	Kernel	Parameters
Single	100.00	87.50	Polynomial	C=0.01, Order=2
Multiclass	100.00	90.00	Linear	C=0.01
Dual	100.00	92.50	Linear	C=0.01

Examples of classified images are given in figure 3.18 and figure 3.19. The number in the bottom right hand corner of the figures specifies the number of candidate pixels classified as new vessels. For the purpose of visualisation, all positive pixels were then morphologically dilated with a structuring element the size of the sub-window used in feature extraction, and the resultant was the delineation of new vessel regions (indicated with a white boundary). Images containing any candidate pixels classified as new vessels, and therefore any delineated regions are classified as new vessel images. As stated earlier, the performance evaluation of delineation was not assessed. However, figure 3.17 provides the manual delineation of new vessel regions marked by an ophthalmologist in order to allow for a visual comparison to the delineation shown in figure 3.18. In addition to this, table 3.4 provides the number of candidate pixels classified as new vessels and the number of new vessel regions that have been successfully detected for each image in the dataset. The image numbers in table 3.4 correspond to those listed in appendix II, in which images 1-20 are PDR images and 21-40 are non-PDR images.

Examples of classified patches are shown in figure 3.20. This illustrates the classification performance on individual patches, which have later been assembled together to form a single image only for visualization purposes. To elaborate,

patches on the left side of the white straight line represent new vessel patches and those on the right side represent non-new vessel patches. White boxes indicate those patches that have been by classified with a new vessel label by the proposed method. As stated previously, the term patch can be used to refer to a pixel. Therefore figure 3.20 essentially represents the classification of the central pixel of each image patch.

Finally, table 3.5 restates the results along with the reported results from other new vessel detection methods. Note that the performance on a per image basis is more useful from a clinical point of view.

This methodology was implemented using Matlab R2013a on an Intel(R) core(TM)2 Quad CPU Q9300 at 2.5 GHz. The Matlab Code took 450 seconds to process each image.

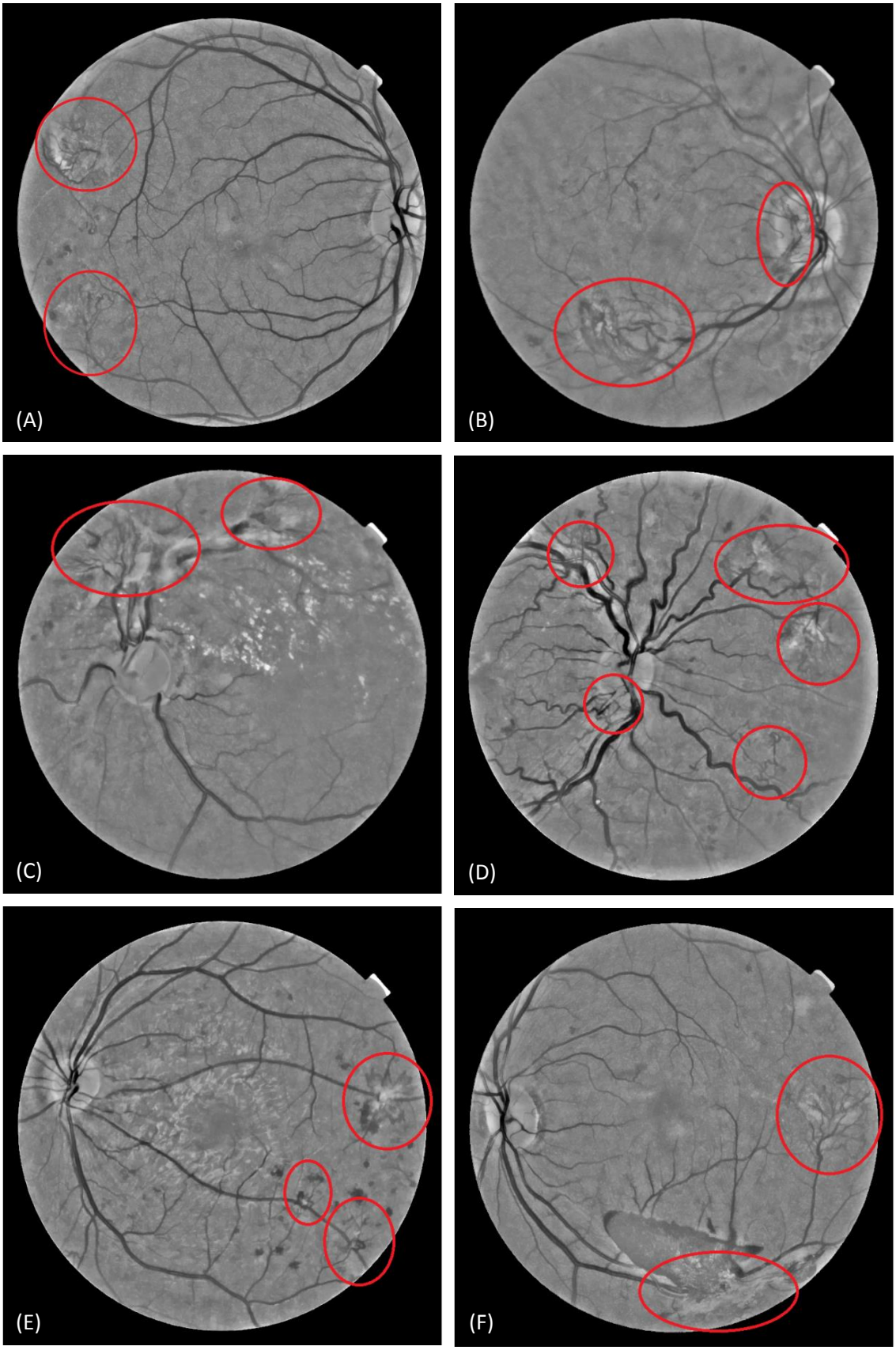


Figure 3.17: Location and delineation of new vessel regions marked by an ophthalmologist.

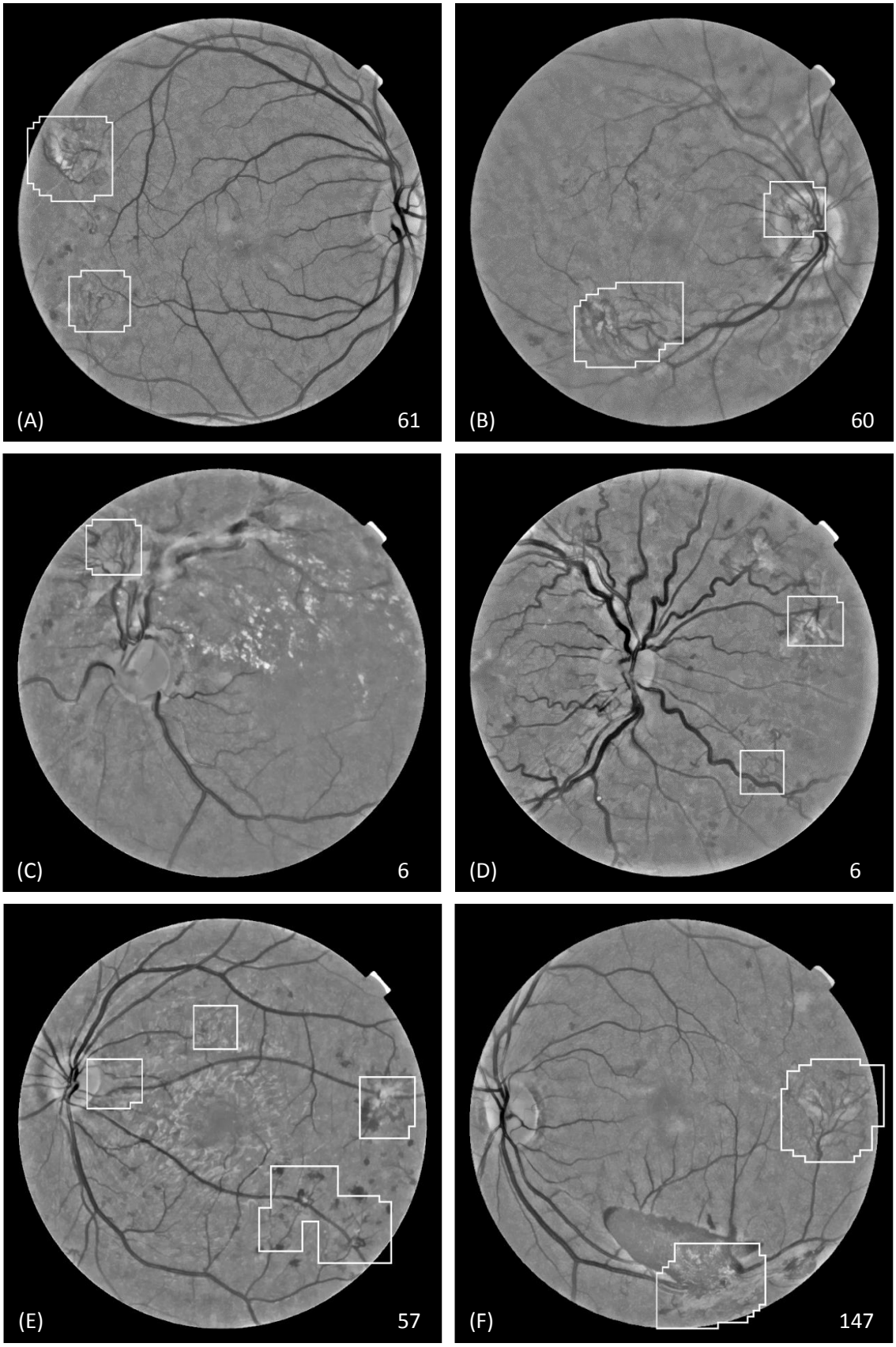


Figure 3.18: Results of the proposed system on a per image basis. True positive images. **(A)-(F)** Corresponds to figures 3.17**(A)-(F)** respectively.

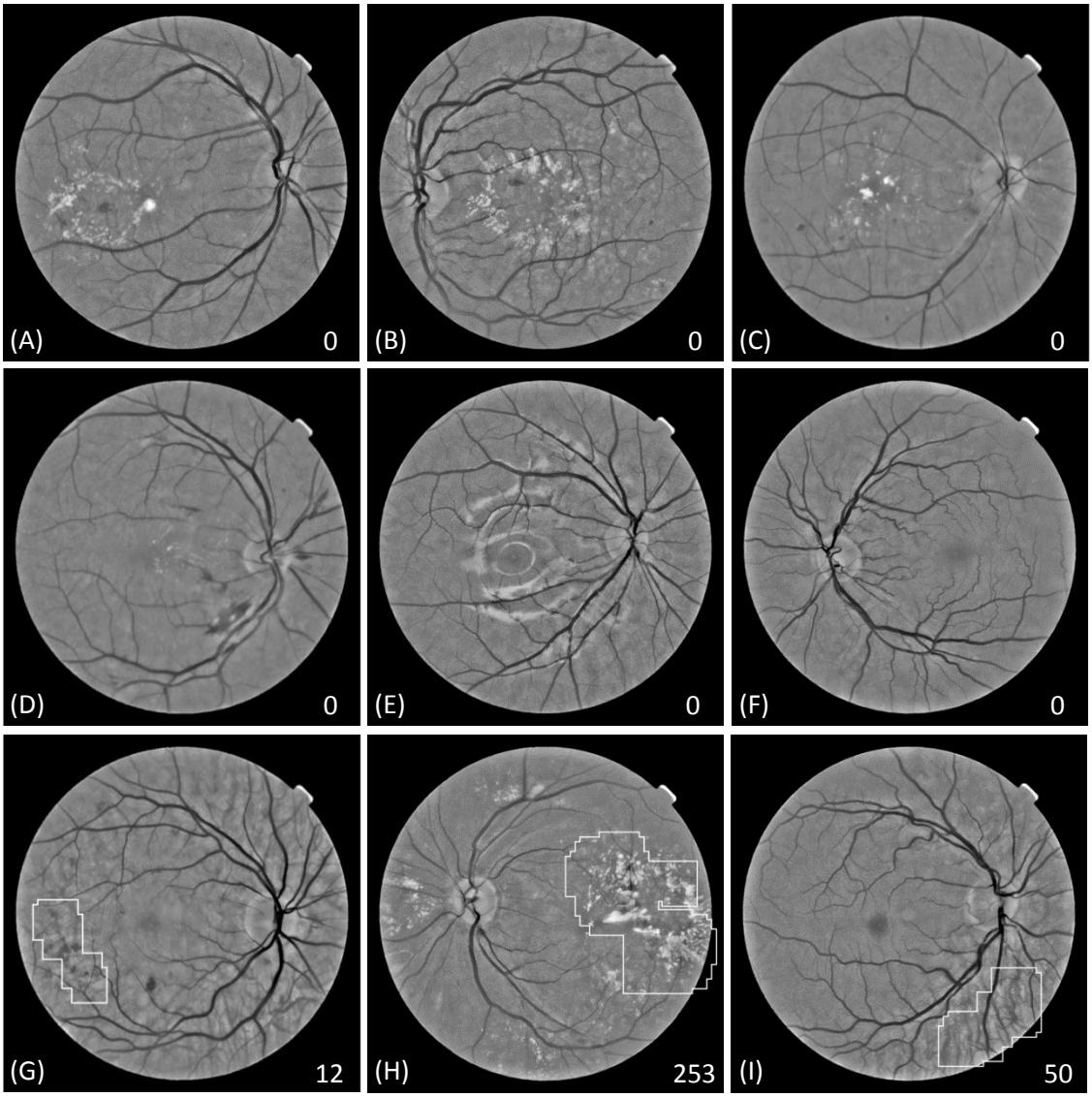
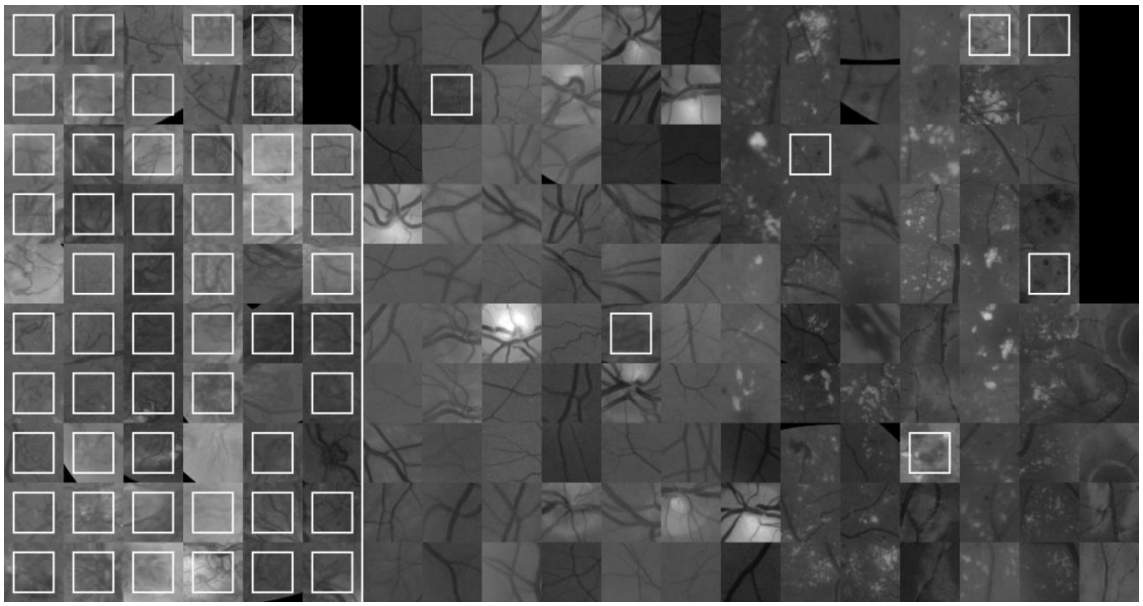


Figure 3.19: Results of the proposed system on a per image basis. **(A)-(F)** True negative images. **(G)-(I)** False positive images.

Table 3.4: The number of candidate pixels classified as new vessels and the number of correctly detected new vessel regions during assessment on a per image basis.

Image N^0	Manually marked new vessel regions	Candidate pixels classified as new vessels	Correctly detected new vessel regions
1	3	57	3
2	4	155	4
3	5	6	2
4	7	314	4
5	2	73	2
6	3	140	1
7	1	128	1
8	2	61	2
9	2	60	2
10	3	34	1
11	2	6	1
12	5	234	3
13	2	26	1
14	2	77	2
15	1	14	1
16	3	167	2
17	1	264	1
18	3	212	2
19	2	147	2
20	3	172	2
21	0	0	-
22	0	0	-
23	0	50	-
24	0	0	-
25	0	0	-
26	0	0	-
27	0	0	-
28	0	0	-
29	0	0	-
30	0	0	-
31	0	0	-
32	0	0	-
33	0	0	-
34	0	0	-
35	0	0	-
36	0	0	-
37	0	0	-
38	0	0	-
39	0	0	-
40	0	0	-
41	0	0	-
42	0	0	-
43	0	0	-
44	0	0	-
45	0	0	-
46	0	0	-
47	0	0	-
48	0	0	-
49	0	0	-
50	0	0	-
51	0	0	-
52	0	0	-
53	0	0	-
54	0	0	-
55	0	0	-
56	0	253	-
57	0	12	-
58	0	0	-
59	0	0	-
60	0	0	-



(A) (B)

Figure 3.20: Results of the proposed system on a per patch basis. **(A)** New vessel patches, 51 TPs and 7 FNs out of 58 patches. **(B)** Non-new vessel patches, 118 TNs and 7 FPs out of 125 patches. White boxes indicate patches classified with a new vessel label.

Table 3.5: Reported results for new vessel detection methods.

Algorithm	SN	SP	Acc	AUC	Level
Hassan [70]	63.90	89.40	-	0.7045	Pixel
Jelinek [71]	94.00	82.00	-	0.900	Image
Goatman [72]	84.2	85.9	-	0.911	Image
Arputham [73]	84.70	86.10	-	-	Image
Pavai [74]	88.89	91.30	-	-	Image
Akram [76]	98.93	96.35	-	-	Segment
Akram [77]	98.00	97.00	98.00	0.980	Segment
Akram [77]	96.00	94.00	95.00	-	Image
Saranya [78]	96.25	89.65	96.53	-	Image
Welikala [79]	100.00	70.00	-	-	Image
Nithyaa [86]	-	-	-	0.947	Image
Agurto [87]	100.00	88.00	-	0.980	Image
Agurto [88]	96.00	83.00	-	0.940	Image
Vatanparast [90]	99.62	96.61	-	-	Patch
Lee [91]	96.30	99.10	98.50	0.993	Image
Proposed Method	100.00	92.50	95.00	0.9693	Image
Proposed Method	87.93	94.40	92.35	0.9616	Patch

3.3 Discussion and Conclusion

In this chapter we have presented an effective new vessel detection method based on a dual classification approach and a four-dimensional feature vector used to analyse the morphology of the local retinal vasculature.

The main contribution of this method is the novel application of creating two different segmented vessel maps, along with a dual classification framework used to independently process information extracted from each of the maps. The objective of this method is to detect new vessels whilst reducing false responses caused by bright lesions and other retinal features. Segmentation methods include the standard line operator and a novel modified line operator. The latter targets the reduction of false responses to non-vessel edges. In addition, emphasis is put on ensuring new vessels are adequately segmented with the correct selection of parameters for vessel segmentation. Another important point emphasised in this chapter, is that high vessel area alone is not enough to identify new vessels. A more detailed analysis of the morphology of the vasculature is required as well as the removal of straight vessels.

Table 3.3 compares the dual classification system's performance against the performance from a single SVM classification and a one-versus-one multiclass SVM classification. Note that for both the single and multiclass SVMs the two feature sets were combined to create a single feature set. These results validate the dual classification framework, showing this process to be superior to the others.

From the examples of classified images shown in figure 3.18 it is evident that the algorithm responds well to a variety of new vessel formations. This includes both new vessels elsewhere (NVE) and new vessels at the optic disc (NVD), new vessels with associated fibrosis and obscure new vessels. Also evident, from figures 3.19(A)-(F), is the algorithm's ability to avoid false responses despite the presence of bright lesions, dark lesions and reflection artefacts. Figure 3.19(I) shows a false positive image caused by vessels from the layer beneath the retina (choroid) being visible. This is an unusual case because when vessels from this layer are visible they normally possess little contrast to the background and often have a wider calibre, and

therefore they do not get segmented. Another cause of false responses are dilated capillaries known as intra-retinal microvascular abnormalities (IRMA), which are very difficult to differentiate from new vessels. IRMA represents a stage of DR that indicates a high risk of the progression of new vessels. A final area of difficultly concerns the spacing between bright lesions of close proximity. The modified line operator was designed to tackle this problem and does so successfully for the majority of cases. However, on rare occasions (see figure 3.19(H)) these areas appear significantly darker than the retinal background and therefore the modified line operator can struggle to avoid their segmentation.

In 2002, studies from the UK [124,125] reported that the prevalence of PDR is 3.7% for patients with type 1 diabetes and 0.5% for patients with type 2 diabetes. When taking into account the proportions of patients with each type of diabetes in the UK, an overall value of 0.82% can be derived. Although the prevalence of PDR is low, the associated risk of the rapid onset of vision loss means it must be detected reliably. In UK screening programmes it is considered a serious breach in protocol if an image with PDR is either missed or delayed in referral. Therefore, the maximum sensitivity of 100.00% that our algorithm reaches on a per image basis whilst achieving a respectable specificity of 92.50% gives it potential for clinical application. However, when considering the low prevalence of PDR, this specificity value is lacking. The results indicate that if the algorithm is applied in screening programmes, for every correctly identified patient with PDR there would be approximately 9 incorrectly identified patients. This calculation does not consider that the 40 non-PDR images from the dataset used to calculate specificity were chosen to make this dataset more challenging and is therefore not a true representative of a screening population.

It should be noted that the proportion of PDR images to non-PDR images used in our dataset does not correspond to the prevalence of PDR in the population. This is also the case for all methodologies in the current literature. To match the PDR prevalence of 0.82%, the 20 PDR images in our dataset would have to be accompanied by 2439 non-PDR images. Theoretically this should not affect the sensitivity and specificity results, apart from the specificity now being measured against a larger and therefore more representative/varied selection of images. However, the correct proportions of

images would mean the accuracy (Acc) result would be a more correct representation in respect to the population.

The proposed method does achieve better performance metrics than most of the other published methods, as shown in table 3.5. True comparisons are difficult to make as there exists variability in terms of their application. Goatman [72] and Agurto [88] seek to detect only NVD. Jelinek [71] applied their methods on fluorescein images as opposed to conventional retinal images. Also the level (segment, image etc.) selected for performance evaluation varies amongst the reported methods. More importantly there exist no standard datasets that have been used for testing. Our particular dataset was not solely created from publicly available datasets due to their limited inclusion of images containing new vessels. Akram [77] created a dataset containing 52 images with new vessels from the four main publicly available retinal image databases. However, the image file names used are not available and we were unable to find the number of new vessel images reported.

Our algorithm's reported optimal performance on each level (image and patch) does not correspond to the same operating point. The reported per image performance of 100.00% and 92.50% for sensitivity and specificity respectively corresponds to a per patch performance of 51.72% and 98.40% for sensitivity and specificity, respectively. This shows, for this per image performance, that the system puts no emphasis on detecting and correctly delineating all new vessels. Instead, identifying any part of any new vessels in the image is sufficient to achieve a positive image label. Figure 3.18(D) illustrates how there is no requirement to identify all new vessels in the image, with only two out of the five new vessel networks being identified. Figure 3.17(D) shows the same image with the location of all five new vessel networks marked by an ophthalmologist. Such an approach assists in obtaining a higher specificity. Niemeijer [20] follows this same approach but in respect to dark lesion detection. Such an approach may hold certain risks, although it is still a viable option for clinical application. The optimal performance for evaluation on a per patch basis is very different, with the emphasis to detect as many new vessel regions as possible (see figure 3.20).

The quantification of new vessel growth as opposed to just the detection of their presence was stated as a possible objective in section 1.1. Quantification can be achieved through delineation. Whilst our algorithm delineates new vessel regions to a certain degree, it was decided not to assess evaluation of its performance in these terms. Striving to achieve a more complete delineation of new vessel regions would make the algorithm more susceptible to false positives. Therefore, our main concern is the identification of new vessel regions as opposed to their accurate delineation, and thus our aim is to detect the centre of new vessel regions. Hence all new vessel patches used for training correlated only to sub-windows directly centred over the middle of new vessel regions. This is a more targeted approach which suits the clinical requirements for the identification of new vessels. Hence, this is why evaluation on a per patch basis (testing on training patches using leave-one-out cross validation) was preferred to the evaluation of delineation to get a more detailed insight into the system's performance than that provided by performance on a per image basis.

Previously mentioned were other vessel segmentation methods [27,28] that specialize in the reduction of false responses caused by bright lesions. However, the reduction has to achieve a more comprehensive level if they are to be successfully applied in PDR detection. Our modified line operator achieves such required levels, particularly due to the additional step taken to resolve the false response caused by the space in-between bright lesions of close proximity. Such a comprehensive level inevitably risks damage to the segmented vessels and therefore brought around the requirement of a dual system.

Future developments should involve the extraction of further information regarding the morphology of the vasculature in order to create a higher dimensional feature set. For example, our feature set lacked any tortuosity measures. Such an important new vessel characteristic should be included. Non-morphology based features should also be explored. The next chapter (section 4.1.1) addresses these issues with the expansion of the feature vector. Another intended development concerns vessel segmentation. Currently a global threshold is applied, although a more sophisticated

approach such as adaptive thresholding or a supervised machine learning approach may yield better results.

All existing PDR detection methods define new vessels as dense lacy networks of unregulated vessel growth. This description does not match their appearance at their initial formation, when they can appear as loops or small networks. These changes in the vasculature can be so subtle and it is likely that all existing methods, including our work, would fail in detecting such cases. These cases must still receive urgent referral to a specialist, although they possess far less risk of a rapid onset of vision reduction in comparison to large regions of abnormal vessel growth. Another difficultly of PDR detection, as mentioned in section 2.4, is caused by retinal features associated with new vessels which may on rare occasions obscure or completely hide them from view. This includes pre-retinal and vitreous haemorrhages caused by the rupture of new vessels and tractional retinal detachment caused by fibrosis. The dataset used for evaluation of our methodology is limited in such cases. This is an issue also not fully addressed by other published methodologies. This is a problem that will have to be tackled in the future; however, for now the main focus should remain on the detection of visible new vessels.

To assist in the development of PDR detection algorithms it is vital that a large new vessel dataset becomes publicly available. New vessel formations can vary greatly in appearance, and thus this should be represented in the dataset. With a standard dataset, comparisons of published methods will be possible.

In conclusion, this chapter has demonstrated an automated system that is capable of detecting the presence of new vessels whilst reducing false responses to bright lesions and other retinal features. The proposal of introducing automated disease/no disease systems into DR screening programmes to reduce the manual grading workload has been considered. The addition of PDR detection will greatly strengthen the proposal by ensuring images requiring urgent referrals are automatically prioritized.

4 FEATURE VECTOR EXPANSION AND GENETIC ALGORITHM BASED FEATURE SELECTION

This chapter describes the exploration of features suitable for the classification of PDR and works as an extension to methodology detailed in the previous chapter. The general framework remains very similar to that depicted in figure 3.1 and is shown in figure 4.1. The feature extraction stage has been developed, with the expansion of the 4-D feature vector into a 21-D feature vector that includes morphology, gradient and intensity based features. This is followed by feature selection using a genetic algorithm (GA) based approach with the objective of finding feature subsets that improve the performances of the classification. In addition, the genetic algorithm is also used to search for the optimal SVM parameters. Work from this chapter was used to contribute to the publication [126].

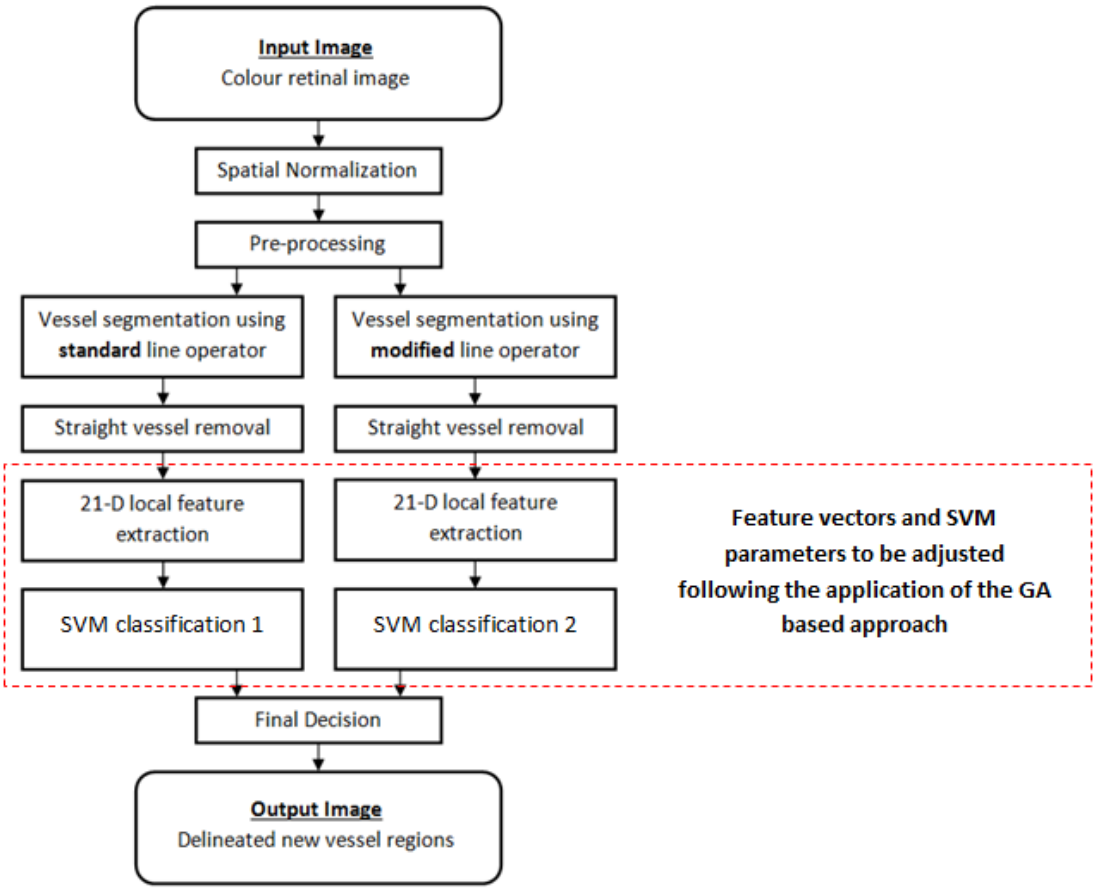


Figure 4.1: System architecture.

4.1 Methodology

4.1.1 Feature Vector Expansion

In the previous chapter a simple 4-D feature vector was built. This was based on the description of new vessel regions as containing many vessel segments, which are closely packed and appear in multiple orientations. However, there are other characteristics that can help to distinguish new vessels. New vessels tend to appear finer in calibre, shorter in length and possess a tortuous nature. New vessels also appear less homogeneous then normal vessels. Therefore, additional features were created based on these characteristics. Features were also pulled in from the current literature. This formed a 21-D feature vector which contained morphology, intensity and gradient based features.

All features were measured from information extracted from the sub-window. This sub-window was scanned through the image and at increments of 20 pixels, candidate pixels were selected. At each candidate pixel position a 21-D feature vector was calculated. This same set of features was measured for each of the pathways to produce two separate feature vectors for each candidate pixel. Some features were designed specifically for a particular classification, but for simplicity both feature vectors remained identical.

The description of the features is listed below. The term segment refers to the skeletonised vessel segments. All segments were labelled.

1) Number of vessel pixels $(\boldsymbol{n_p})$

The sum of all segment pixels within the sub-window, n_p, is defined as

$$n_p = \sum_{i \in S_p} 1 \tag{4.1}$$

where S_p is the set of segment pixels in the sub-window.

2) Number of vessel segments $(\boldsymbol{n_s})$

The number of whole and partially included segments within the sub-window, n_s, is defined as

$$n_s = \sum_{i \in S_s} 1 \tag{4.2}$$

where S_s is the set of segments in the sub-window.

3) Number of vessel orientations $(\boldsymbol{n_{or}})$

The end points of a segment were connected by a straight line. The angle the line made with the x-axis that fell within the range -90° to 90° of the unit

circle was calculated. The calculated angle was accordingly dropped into one of eight bins, each representing a range of angles. This was done for each segment within the sub-window. A non-empty bin was awarded the value of 1, which is defined as

$$b_i = \begin{cases} 0 & , \ if \ \emptyset \\ 1 & , \ otherwsie \end{cases} \tag{4.3}$$

where b_i is the ith bin. The number of non-empty bins represented the number of orientations, n_{or}, which is defined as

$$n_{or} = \sum_{i}^{8} b_i \tag{4.4}$$

4) Vessel density mean $(\boldsymbol{vd_m})$

A segment was dilated with a disk structuring element with a radius of 20 pixels. The number of pixels from all segments within the dilated area was divided by the number of pixels within the segment to give its vessel density. This was done for each segment within the sub-window and the mean vessel density was calculated. This is expressed through equations 4.5 to 4.8.

The dilated segment, D, is defined as

$$D = v \oplus B \tag{4.5}$$

where v is a single segment and B is the structuring element. The intersection, I, between D and all segments in the sub-window, V, is defined as

$$I = D \cap V \tag{4.6}$$

The number of pixels in I divided by the number of pixels in the single segment, v, gives the vessel density, vd, for that segment and is defined as

$$vd = {\sum_{i \in S_{p1}} 1} \Big/ {\sum_{i \in S_{p2}} 1} \tag{4.7}$$

where S_{p1} is the set of pixels in I and S_{p2} is the set of pixels in v. This was calculated for each segment within the sub-window. The vessel density mean, vd_m, is defined as

$$vd_m = \left(\sum_{i \in S_s} vd_i \right) / n_s \tag{4.8}$$

where vd_i is the vessel density of the ith segment.

5) Tortuosity mean (T_m)

The tortuosity of each segment was calculated using the true length (measured with the chain code) divided by the Euclidean length. The mean tortuosity, T_m, was calculated from all the segments within the sub-window, defined as

$$T_m = \left(\sum_{i \in S_s} tL_i \, / eL_i \right) / n_s \tag{4.9}$$

where tL_i is the true length and eL_i is the Euclidean length of the ith segment. Lengths were measured in pixels.

6) Tortuosity max (T_{mx})

The maximum tortuosity, T_{mx}, amongst all segments in the sub-window is defined as

$$T_{mx} = \max_{i \in S_s} (tL_i / eL_i) \tag{4.10}$$

7) Vessel length mean $(\boldsymbol{tL_m})$

The mean true length of all segments within the sub-window, tL_m, is defined as

$$tL_m = \left(\sum_{i \in S_s} tL_i \right) / n_s \tag{4.11}$$

8) Number of bifurcation points $(\boldsymbol{n_{bi}})$

The number of bifurcation points removed within the sub-window when creating segments, n_{bi}, is defined as

$$n_{bi} = \sum_{i \in S_{bi}} 1 \tag{4.12}$$

where S_{bi} is the set of bifurcation points removed from the sub-window.

9) Grey level mean $(\boldsymbol{GL_m})$

The mean grey level of all segment pixels within the sub-window, GL_m, is defined as

$$GL_m = \left(\sum_{i \in S_p} gl_i \right) / n_p \tag{4.13}$$

where gl_i is the grey level (pre-processed image) of the ith segment pixel.

10) Grey level coefficient of variation $(\boldsymbol{GL_{cv}})$

The standard deviation of the grey level of all segments pixels within the sub-window, σ_{GL}, is defined as

$$\sigma_{GL} = \sqrt{\frac{1}{n_p}\sum_{i \in S_p}(gl_i - GL_m)^2} \quad (4.14)$$

The grey level coefficient of variation, GL_{cv}, is the ratio of the standard deviation, σ_{GL}, to the mean of the grey level of all segment pixels within the sub-window,GL_m, which is defined as

$$GL_{cv} = \sigma_{GL}/GL_m \quad (4.15)$$

11) Gradient mean (G_m)

The gradient magnitude, G, calculated using the Sobel gradient operator, is defined as

$$G = \sqrt{{G_x}^2 + {G_y}^2} \quad (4.16)$$

where G_x and G_y are the first derivatives achieved from convolution of the pre-processed image with the Sobel horizontal mask and vertical mask respectively. The mean gradient magnitude along all segment pixels within the sub-window, G_m, is defined by

$$G_m = \left(\sum_{i \in S_p} G_i\right)/n_p \quad (4.17)$$

where G_i is the gradient of the ith segment pixel.

12) Gradient coefficient of variation (G_{cv})

The standard deviation of the gradient of all segments pixels within the sub-window, σ_G, is defined as

$$\sigma_G = \sqrt{\frac{1}{n_p}\sum_{i \in S_p}(G_i - G_m)^2} \quad (4.18)$$

The gradient coefficient of variation, G_{cv}, is the ratio of the standard deviation, σ_G, to the mean of the gradient of all segment pixels within the sub-window, G_m, which is defined as

$$G_{cv} = \sigma_G / G_m \tag{4.19}$$

13) Line strength mean $(\boldsymbol{LS_m})$

The mean line strength of all segment pixels within the sub-window, LS_m, is defined as

$$LS_m = \left(\sum_{i \in S_p} LS_i \right) / n_p \tag{4.20}$$

where LS_i is the line strength of the ith segment pixel.

14) Vessel width mean $(\boldsymbol{W_m})$

Skeletonization correlates to vessel centre lines. The distance from a segment pixel to the closest boundary point of the vessel using the vessel map prior to skeletonization was measured. This gives the half-width at that point which was then multiplied by 2 to achieve the full vessel width. The mean vessel width, W_m, was calculated for all segment pixels within the sub-window, which is defined as

$$W_m = \left(\sum_{i \in S_p} 2hW_i \right) / n_p \tag{4.21}$$

where hW_i is the vessel half-width of the ith segment pixel.

15) Vessel wall gradient mean (wG_m)

As for the vessel width above, the closest boundary point was assumed to be the vessel wall. The sum of all vessel wall pixels within the sub-window, n_{wp}, is defined as

$$n_{wp} = \sum_{i \in S_{wp}} 1 \tag{4.22}$$

where S_{wp} is the set of wall pixels in the sub-window. The mean gradient magnitude along all vessel wall pixels within the sub-window, wG_m, is defined as

$$wG_m = \left(\sum_{i \in S_{wp}} G_i \right) / n_{wp} \tag{4.23}$$

where G_i is the gradient of the ith wall pixel.

16) Vessel wall gradient coefficient of variation (wG_{cv})

The standard deviation of the vessel wall gradient along all vessel wall pixels within the sub-window, σ_{wG}, is defined as

$$\sigma_{wG} = \sqrt{\frac{1}{n_{wp}} \sum_{i \in S_{wp}} (G_i - wG_m)^2} \tag{4.24}$$

The vessel wall gradient coefficient of variation, wG_{cv}, is the ratio of the standard deviation, σ_{wG}, to the mean of the vessel wall gradient, wG_m, along all vessel wall pixels within the sub-window, which is defined as

$$wG_{cv} = \sigma_{wG} / wG_m \tag{4.25}$$

17) Compactness (C)

The full binary vessel map prior to skeletonization and straight vessel removal was used. Area (a) and perimeter (p) within the sub-window were measured and are defined by equations 4.26 and 4.27.

$$a = \sum_{i \in S_{fp}} 1 \tag{4.26}$$

where S_{fp} is the set of full binary vessel map pixels in the sub-window.

$$p = \sum_{i \in S_{pp}} 1 \tag{4.27}$$

where S_{pp} is the set of full binary vessel map perimeter pixels in the sub-window. Compactness, C, is defined as

$$C = 4\pi a / p^2 \tag{4.28}$$

Circularity is also a measure of compactness, and hence this is the same measure as that from equation 3.5.

18) Linkage (L)

The full binary vessel map prior to skeletonization and straight vessel removal was used. The number of unconnected binary objects within the sub-window, n_b, is defined as

$$n_b = \sum_{i \in S_b} 1 \tag{4.29}$$

where S_b is the set of full binary vessel map objects in the sub-window. The linkage, L, is defined by equation 4.30 in which vessel area (a) is divided by the number of objects (n_b) within the sub-window.

$$L = a / n_b \tag{4.30}$$

19) Local grey level mean (LGL_m)

The sum of all pixels within the sub-window, n_{ap}, is defined as

$$n_{ap} = \sum_{i \in S_{ap}} 1 \tag{4.31}$$

where S_{ap} is the set of all pixels in the sub-window. The mean grey level of all pixels within the sub-window, LGL_m, is defined as

$$LGL_m = \left(\sum_{i \in S_{ap}} gl_i \right) / n_{ap} \tag{4.32}$$

where gl_i is the grey level (pre-processed image) of the ith pixel.

20) Local grey level max (LGL_{mx})

The maximum grey level amongst all pixels within the sub-window using the pre-processed image, LGL_{mx}, is defined as

$$LGL_{mx} = \max_{i \in S_{ap}} (gl_i) \tag{4.33}$$

21) Local grey level variation (σ_{LGL})

The standard deviation of the grey level of all pixels within the sub-window using the pre-processed image, σ_{LGL}, is defined as

$$\sigma_{LGL} = \sqrt{\frac{1}{n_{ap}} \sum_{i \in S_{ap}} (gl_i - LGL_m)^2} \tag{4.34}$$

4.1.2 Genetic Algorithm Based Feature Selection and Parameter Optimization

Feature selection is the process of selecting the smallest subset of features that is necessary for accurate prediction (classification or regression). It achieves this by the

removal of redundant and irrelevant features. Redundant features are those which provide no further information than the currently selected features, and irrelevant features provide no useful information in any context. The benefits of feature selection are improving the classifier's performance and providing a better understanding of the underlying process that generated the data [127]. Using a small number of features will also save significant computation time and builds models that generalise better for unseen data.

An exhaustive search of all possible subsets of features to find the one which maximises the classifier's performance is the simplest approach. However, this approach is far too computationally expensive and is only suitable for the smallest of feature sets. Filter methods are also simple, as well as computationally fast. A statistical test is performed for individually ranking the features according to their relevance and low ranked features are then removed (filtered). However they suffer from ignoring feature dependencies and also ignore interaction with the classifier. Wrapper methods do interact with the classifier, utilizing them to score feature subsets and also model feature dependencies. Two of the most popular wrapper methods are greedy search strategies known as sequential forward selection and sequential backward selection. Genetic algorithms can also be applied as a wrapper method and shall be discussed further later. [128] provides a comprehensive review of feature selection techniques.

Feature selection does not feature heavily in the retinal image analysis articles reviewed in chapter 2. Goatman [72] applied a filter method to remove features according to their individual rank, which was performed using a statistical test known as the Wilcoxon rank sum test. Staal [36] performed feature selection using the wrapper method of sequential forward selection. Vatanparast [90] applied principal component analysis (PCA) [35] to reduce the dimensionality of the feature vector. However, this is not a method of feature selection; it was performed to avoid the disadvantage of high dimensionality.

Before proceeding, it is worth recapping on the main principles of SVM classification. SVMs seek a linear decision surface that can separate classes of objects and has the

largest distance between border-line objects. If the classes are not linearly separable the SVM maps the data into a higher dimensional space using the kernel trick, where the separating linear decision surface exists and can be determined. There are numerous different kernel functions and the majority possess parameters which need to be selected. The soft margin parameter C is another parameter associated with SVMs and is used to deal with noisy measurements and outliers. Therefore, the effectiveness of SVMs depends on the selection of the kernel function, the kernel parameters and the soft margin parameter C. A grid search is the conventional approach for selecting the optimal SVM parameters.

The feature subset selected influences the appropriate SVM parameters and vice versa [129]. To elaborate, each time a new feature subset is selected the input space is altered. Therefore the SVM parameters have to be re-explored to find an optimal selection. However, performing a grid search of SVM parameters each time a new feature subset is selected is a very time consuming task. To solve this problem the methodology by Huang [130] was followed which used genetic algorithms to allow for the exploration of the optimal feature subset and SVM parameters to occur simultaneously. The only requirement was the pre-selection of the kernel function type. Another important factor which applies specifically to dual classification is that this approach allows for both feature vectors to be explored simultaneously.

Genetic algorithms (GA) is a general adaptive optimization search methodology first presented by Bledsoe [131] and mathematically formalized by Holland [132], which takes inspiration from Darwin's theory of evolution. Each gene represents a variable and a sequence of genes is referred to as a chromosome. Each chromosome may be a potential solution to the optimization problem. A population of chromosomes is randomly initialized. Chromosomes are evaluated for their quality according to a predefined fitness function. A new population of chromosomes is produced by selecting high performing chromosomes (process known as selection) to produce offspring which retain many aspects of their parents. These offspring are formed by using genetic operators: crossover and mutation. Crossover is a mechanism for exchanging genes between two selected chromosomes to create new offspring. Mutation operates by modifying one or more components of a selected

chromosome. Selection is then performed on the new population of chromosomes followed by the production of offspring. Therefore, chromosomes are competing with each other and only the fittest survive, with the GA obtaining the chromosome providing the optimal or near-optimal solution after a series of iterative computations. The GA was chosen in preference to a large variety of other optimization procedures due to its ability to deal with large search spaces efficiently. Therefore, convergence is more likely to avoid local sub-optimal solutions and find the global optimal solution.

4.1.2.1 Genetic Operators

The role of the genetic operators of selection, crossover and mutation are explained above. There exists a variety of techniques for each of these operators which can be found documented by Sivanandam [133]. The techniques performed, which are detailed below, were roulette selection, single point crossover and uniform mutation. These were chosen as they are commonly used techniques, which are simple to implement and hence are very transparent. Therefore, there remains further scope to perform comparative assessment in the performance of different operators for this task.

Roulette selection chooses parents (high performing chromosomes) by simulating a roulette wheel, the size of each individual interval corresponds to the fitness function value of the associated individual chromosome. For example, in figure 4.2 the sum of all the individual's fitness function values make up the circumference of the roulette wheel. This sum equates to the range [0, Sum]. The algorithm selects a random number in the range [0, Sum] and the individual whose interval spans this random number is selected. Therefore, the probability of selection is equal to the interval's size.

Figure 4.2: Roulette selection, shown for an example with a population of 10 chromosomes.

The simplest form of crossover is that of single-point crossover. Consider two parent chromosomes. A random integer position is chosen between 1 and the number of genes (variables). Genetic information is exchanged between the two parents about this point, resulting in two new offspring being produced. Figure 4.3 illustrates the procedure. The probability of crossover is set at 0.75.

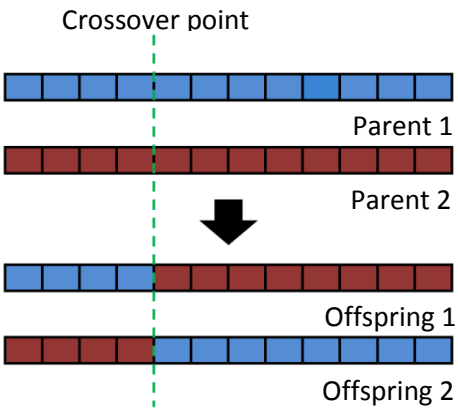


Figure 4.3: Single point crossover.

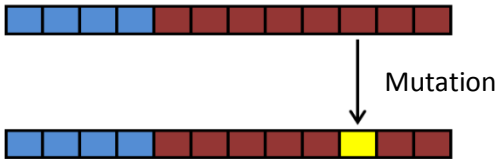


Figure 4.4: Mutation.

Uniform mutation has two steps. First, the algorithm selects a fraction of the genes of an individual chromosome for mutation, where each gene has a low probability of being mutated. This probability rate is set at 0.01. In the second step, each selected gene is replaced by a random number selected uniformly from the range for that

entry. Figure 4.4 illustrates the process of mutation where a single gene has been selected.

4.1.2.2 Chromosome Design

In the context of our optimization problem, each chromosome was comprised of three parts. This was the feature subsets, the soft margin parameter C and the kernel parameters. These are each described in further detail below and figure 4.5 displays the chromosome design.

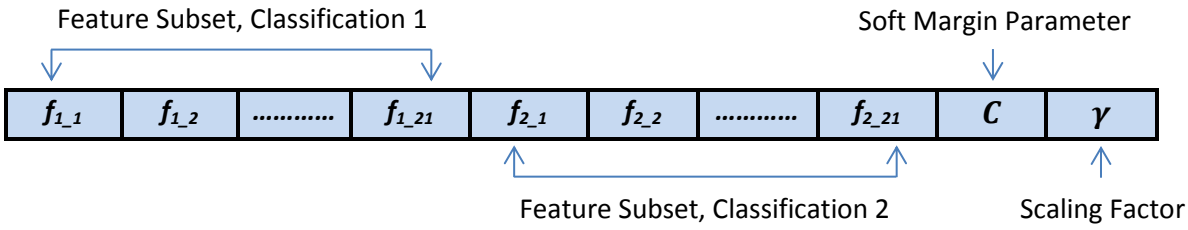


Figure 4.5: Chromosome design for the RBF kernel.

Feature subsets:

Two feature subsets were required, one for each classification in the dual process. The chromosome used 42 genes to represent the two sets of 21 features. These genes were denoted with value '1' to indicate the feature was selected or '0' to indicate the feature was not selected.

Soft margin parameter C:

This was represented with a gene with an integer value ranging from 1-5 which corresponded to the C values of 0.01, 0.1, 1, 10, 100.

Kernel parameters:

The type of kernel parameters depend on the kernel function used, with kernel types of linear, Gaussian radial basis function (RBF) and polynomial being assessed. The type of kernel function used was pre-selected. No kernel parameters were required for the linear kernel. For the RBF kernel the scaling factor (γ) was required and was represented by a gene with an integer value ranging from 1-5 which corresponded to γ of 0.0001, 0.001, 0.01, 0.1, 1. For the polynomial kernel the order of the polynomial was required and this was represented by a gene with an integer value ranging from 1-4 which corresponded to the order.

It should be noted that C and γ are continuous variables. However, they have been represented by discrete values in order to simplify the optimization problem by the reduction of the search space.

The SVM kernels that were chosen to be assessed, as well as the value range for the SVM parameters were based on a strategy for SVM model selection provided by Statniov [134].

4.1.2.3 Fitness Function

The optimal solution that the GA produced was a chromosome with the optimal feature subsets and SVM parameters that maximised the fitness function. The criterion used to design the fitness function is shown in equation 4.35, with p representing the classification performance (units of percent) and n representing the combined number of selected features from both feature sets. Thus, a chromosome with a high classification performance and a small number of features produced a high fitness value. An empirically derived scaling factor of 1/100 was applied to n in order to ensure the main priority of the optimization procedure was on attaining a high classification performance, and hence a small number of features was a secondary matter. A logical starting point in deriving the scaling factor value was to ensure the maximum variation of n equated to a value just under the value of a single unit of p, and therefore initial investigations started with a scaling factor of

1/50. Thereafter, the scaling factor was varied in increments and it was found that the scaling factor of 1/100 achieved the best outcome. The classification performance was assessed from the final decision achieved from combining the two classification outcomes. The measure used to quantify the classification performance shall be discussed in section 4.2.1.

$$Fitness\ function = p - \frac{n}{100} \tag{4.35}$$

4.1.2.4 System Architecture for the Genetic Algorithm Based Approach

A flow chart presenting the system's architecture is provided in figure 4.6. All features were normalised so that each feature had zero mean and unit standard deviation. The kernel function type was pre-selected, and therefore this process was separately performed for the linear, RBF and polynomial kernel functions. An explanation of the main steps follows:

(1) Consider the box (from figure 4.6) labelled as population as the starting point. A population of chromosomes was randomly initialized. Each chromosome comprised of a feature subset and SVM parameters.

(2) All chromosomes in the population were each evaluated individually with the following procedure. The selected feature subset was used to accordingly alter the training and testing sets. The training set along with the selected SVM parameters were used to train the SVM classifier, while the testing set was used to calculate the classification performance. The chromosome was evaluated by the fitness function which was derived using the classification performance along with number of selected features.

(3) The fitness functions of all chromosomes in the population were checked to see if the stopping criteria was reached. The stopping criteria was set at the maximum possible fitness function (equates to p = 100 and n = 2). If reached the process was ended, otherwise it proceeded to the genetic operations.

(4) The genetic operations produced the next generation (new population). This was done by first selecting the elite chromosomes as parents, and then crossover and mutation were used to produce offspring (new population).

Steps (2)-(4) were performed iteratively until the stopping criteria was met or the pre-set maximum number of generations was reached.

The GA settings were chosen at a population size of 40 (initial population of 80), elite parent count of 2 and the number of generations of 10.

The efficiency of GA is greatly dependent on the parameters selected. There are several publications that state recommended standard parameter settings. However, no recommended settings are truly universal; instead, parameters have to be generally tuned to specific problems. Therefore, the GA parameter values stated above, as well as the crossover and mutation probabilities stated in section 4.1.2.1 were achieved by starting with the standard values provided De Jong [135] and Grefenstette [136]. This was then followed by parameter tuning. A limitation was the number of generations had to remain low due to time constraints, and this will be explained in the section 4.3.

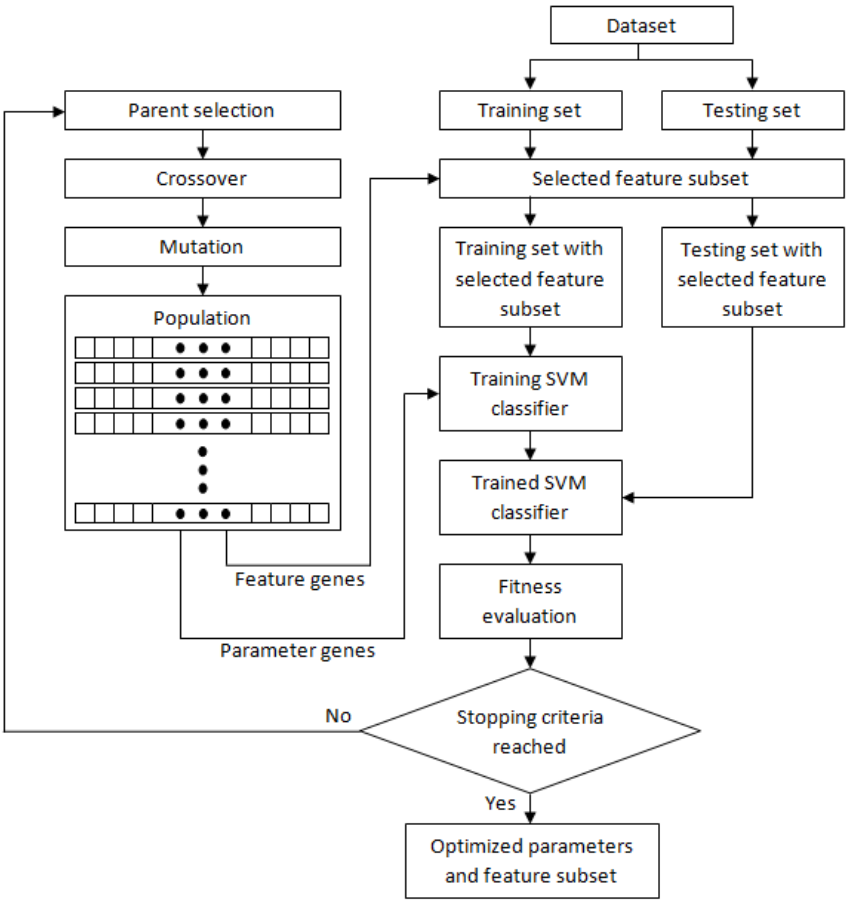


Figure 4.6: Architecture of the GA based feature selection and SVM parameter selection system, based on [130].

4.2 Experimental Evaluation

4.2.1 Materials and Performance Measures

The materials and performance measures used for this chapter are identical to those detailed in the previous chapter (sections 3.2.1 and 3.2.2). To summarize:

- 60 image dataset (new vessel images = 20, non-new vessel images = 40).
- Leave-one-out cross validation.
- Selection of pixels used for training data (selection remains identical).
- Evaluation on a per image and per patch basis.
- Performance metrics of sensitivity (SN), specificity (SP), accuracy (Acc), area under the ROC curve (AUC).

- 3D ROC surface used to assess dual classification.
- Per patch basis, optimal operating point = maximum accuracy.
- Per image basis, optimal operating point = highest specificity at a sensitivity of 100%.

To clarify the GA based approach was applied separately to find the optimal feature subset and SVM parameters for each per image and per patch evaluation.

The measures used to quantify p (classification performance) for involvement in the fitness function (equation 4.35) used by the GA based approach are those stated above for the optimal operating point (maximum accuracy, highest specificity at a sensitivity of 100%).

The evaluation procedures of leave-one-out cross validation, ROC analysis and the selection of the optimal operating point had to be performed for every single chromosome in the GA based approach.

4.2.2 Results

The Wilcoxon rank sum test, a statistical test, was used to infer the discrimination ability of the features. This is done by assessing whether the median feature values for the two classes differ significantly. This method of feature ranking does not interact with the classifiers. All calculations for the Wilcoxon rank sum test are performed using only the training data, and therefore this is the ranking of features prior to classification. Results for both classification pathways are shown in table 4.2, with a smaller p-value indicating better discrimination ability. These calculations were not used for feature selection in this project, however they were performed to provide a useful insight.

Table 4.1: Summary of feature vector.

Feature Number	Feature Name	Feature Number	Feature Name
1	Number of vessel pixels	12	Gradient coefficient of variation
2	Number of vessel segments	13	Line strength mean
3	Number of vessel orientations	14	Vessel width mean
4	Vessel density mean	15	Vessel wall gradient mean
5	Tortuosity mean	16	Vessel wall gradient coefficient of variation
6	Tortuosity max	17	Compactness
7	Vessel length mean	18	Linkage
8	Number of bifurcation points	19	Local grey level mean
9	Grey level mean	20	Local grey level max
10	Grey level coefficient of variation	21	Local grey level variation
11	Gradient mean		

Table 4.2: Performance of the 21 features for each classification pathway, assessed using the Wilcoxon rank sum test. Listed in order of p-value, with the lowest value at the top of the table.

Wilcoxon rank sum test			
Classification 1		Classification 2	
Feature	p-value	Feature	p-value
8	2.0889×10^{-17}	2	4.3315×10^{-16}
2	2.1501×10^{-17}	1	1.8680×10^{-14}
4	6.7195×10^{-17}	3	7.1091×10^{-14}
1	1.4700×10^{-16}	4	6.4934×10^{-11}
3	7.1501×10^{-16}	17	1.3213×10^{-11}
17	7.2511×10^{-14}	8	1.1658×10^{-10}
7	9.4306×10^{-13}	19	6.8748×10^{-8}
14	1.8945×10^{-12}	6	1.0735×10^{-5}
6	5.3935×10^{-11}	16	3.3293×10^{-4}
16	2.6465×10^{-10}	20	5.0906×10^{-4}
10	1.0324×10^{-8}	7	0.0011
11	8.1172×10^{-7}	18	0.0018
21	5.7432×10^{-4}	10	0.0019
13	0.0017	14	0.0060
18	0.0030	13	0.0084
12	0.0193	21	0.0122
5	0.1030	9	0.0189
9	0.3293	11	0.0204
19	0.4340	5	0.1411
15	0.7174	15	0.1938
20	0.8066	12	0.3683

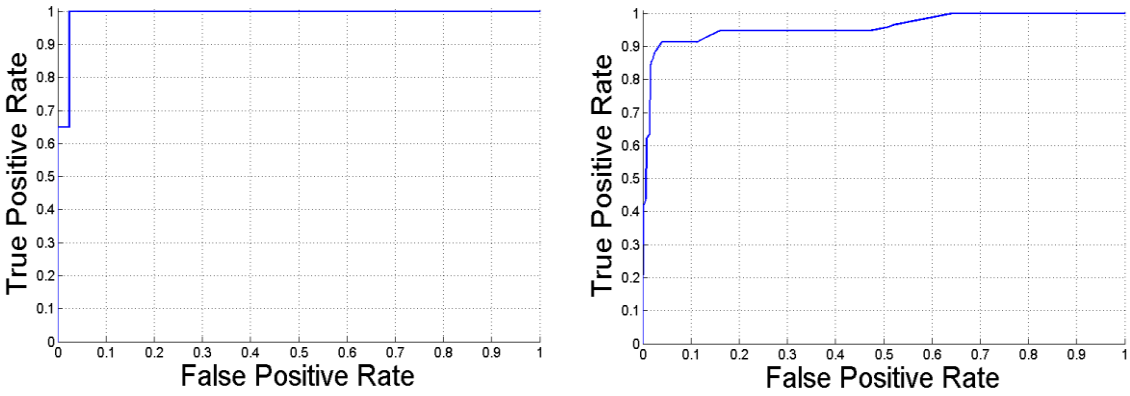
The features and SVM parameters selected for each classification that maximise the fitness function of the GA based system are shown in table 4.3. The resulting ROC curves of the proposed system under these settings on a per image and per patch basis are depicted in figures 4.7(A)-(B). The AUC value for the per image basis is 0.9914. The optimal operating point according to the application specific performance measure is a sensitivity of 100.00% and a specificity of 97.50%. For a per patch basis the AUC value is 0.9600. The operating point with maximum accuracy of 94.54% gives a sensitivity of 91.38% and a specificity of 96.00%. Table 4.4 shows these results along with the reported results from other new vessel detection methods.

Examples of classified images are given in figure 4.9 and figure 4.10. Recall, the number in the bottom right hand corner of the figures specifies the number of candidate pixels classified as new vessels. All positive pixels were then morphologically dilated with a structuring element the size of the sub-window used in feature extraction, and the resultant was the delineation of new vessel regions (indicated with a white boundary). Any image containing any candidate pixels classified as new vessels, and therefore any delineated region is classified as a new vessel image. As stated earlier, the performance evaluation of delineation was not assessed. However, figure 4.8 provides the manual delineation of new vessel regions marked by an ophthalmologist in order to allow for a visual comparison to the delineation shown in figure 4.9. In addition to this, table 4.5 provides the number of candidate pixels classified as new vessels and the number of new vessel regions that have been successfully detected for each image in the dataset. The image numbers in table 4.5 correspond to those listed in appendix II, in which images 1-20 are PDR images and 21-40 are non-PDR images.

Examples of classified patches are shown in figure 4.11, with patches on the left side of the white straight line representing new vessel patches and those on the right side representing non-new vessel patches. White boxes indicate those patches that have been by classified with a new vessel label by the proposed method. Note that the performance from a per image basis is more useful from a clinical point of view.

Table 4.3: Features and SVM parameters selected for each classification by the GA based system.

Level	Classification	Kernel	SVM parameters	Features
Image	1	RBF	C=1, γ=1	4,8,10,13,14,16,19
	2	RBF	C=1, γ=1	1,2,4,6,8,15,17
Patch	1	RBF	C=10, γ=1	2,3,4,7,9,10,14,16,20
	2	RBF	C=10, γ=1	1,6,7,14,16,20



(A) (B)

Figure 4.7: ROC curves. **(A)** Performance on a per image basis. **(B)** Performance on a per patch basis.

Table 4.4: Reported results for new vessel detection methods.

Algorithm	SN	SP	Acc	AUC	Level
Hassan [70]	63.90	89.40	-	0.7045	Pixel
Jelinek [71]	94.00	82.00	-	0.900	Image
Goatman [72]	84.2	85.9	-	0.911	Image
Arputham [73]	84.70	86.10	-	-	Image
Pavai [74]	88.89	91.30	-	-	Image
Akram [76]	98.93	96.35	-	-	Segment
Akram [77]	98.00	97.00	98.00	0.980	Segment
Akram [77]	96.00	94.00	0.95	-	Image
Saranya [78]	96.25	89.65	96.53	-	Image
Welikala [79]	100.00	70.00	-	-	Image
Nithyaa [86]	-	-	-	0.947	Image
Agurto [87]	100.00	88.00	-	0.980	Image
Agurto [88]	96.00	83.00	-	0.940	Image
Vatanparast [90]	99.62	96.61	-	-	Patch
Lee [91]	96.30	99.10	98.50	0.993	Image
Chapter 3	100.00	92.50	95.00	0.9693	Image
Chapter 3	87.93	94.40	92.35	0.9616	Patch
Proposed	100.00	97.50	98.33	0.9914	Image
Proposed	91.38	96.00	94.54	0.9600	Patch

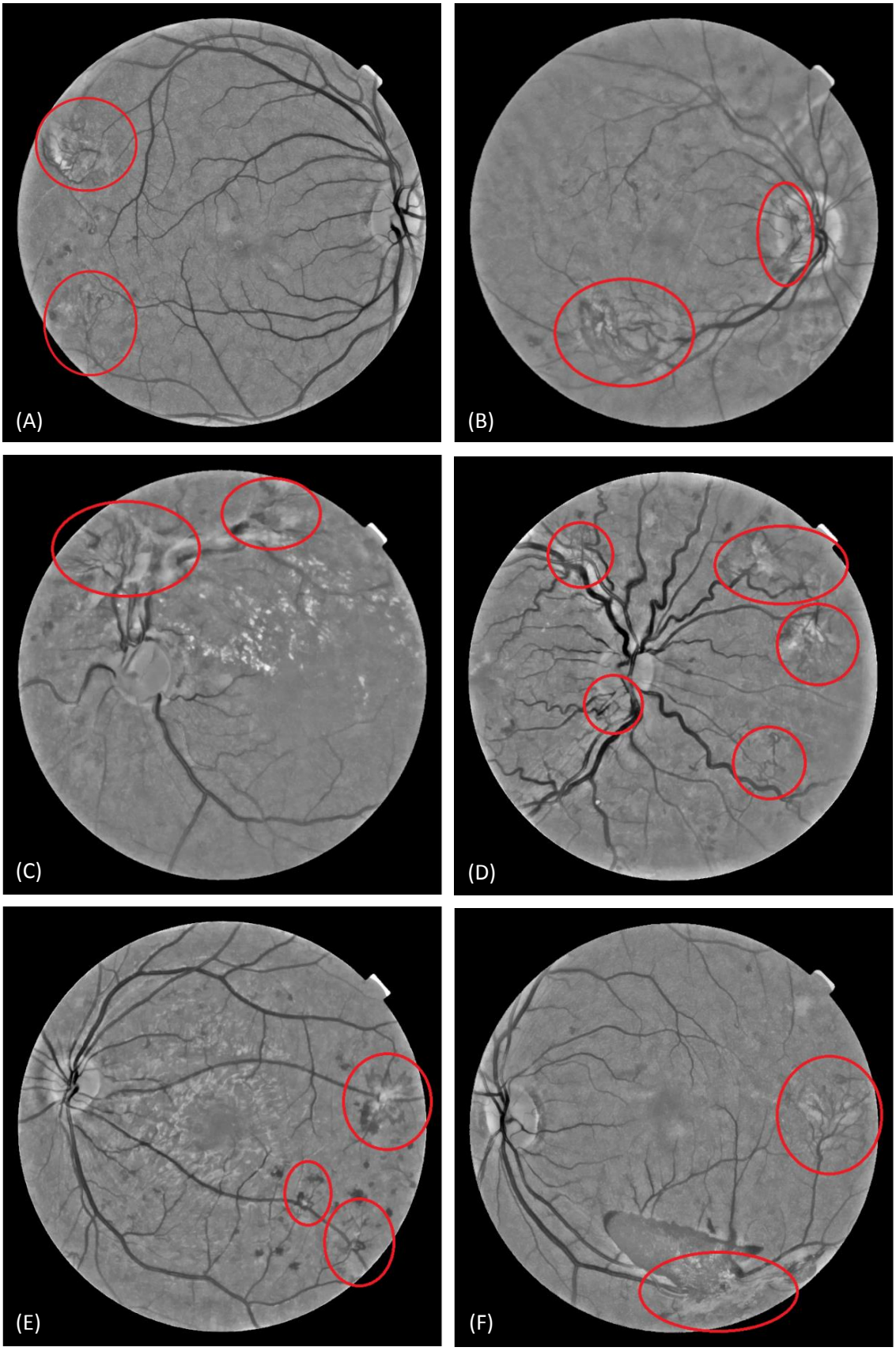


Figure 4.8: Location and delineation of new vessel regions marked by an ophthalmologist.

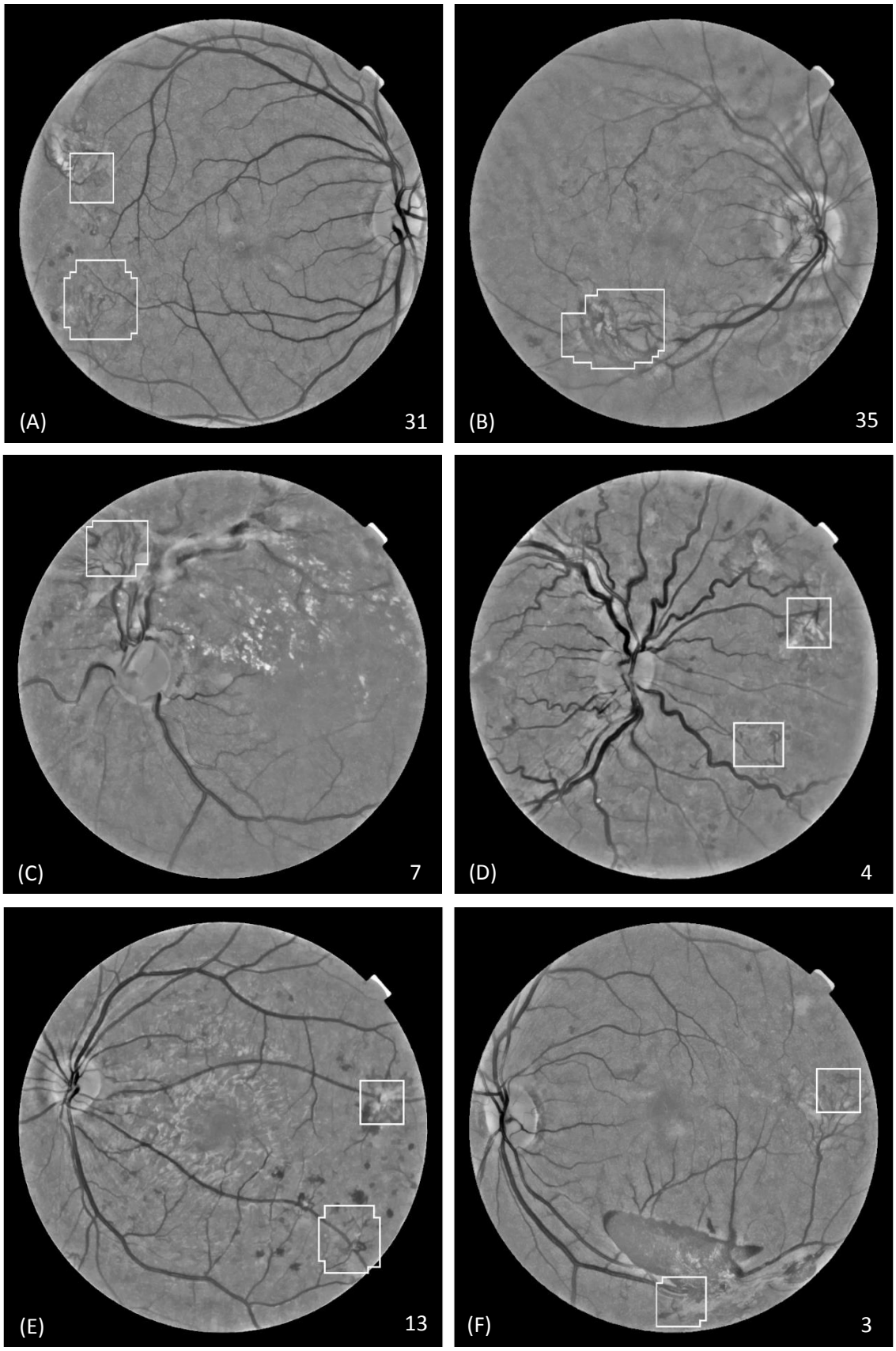


Figure 4.9: Results of the proposed system on a per image basis. True positive images. **(A)-(F)** Correspond to figures 4.8**(A)-(F)** respectively.

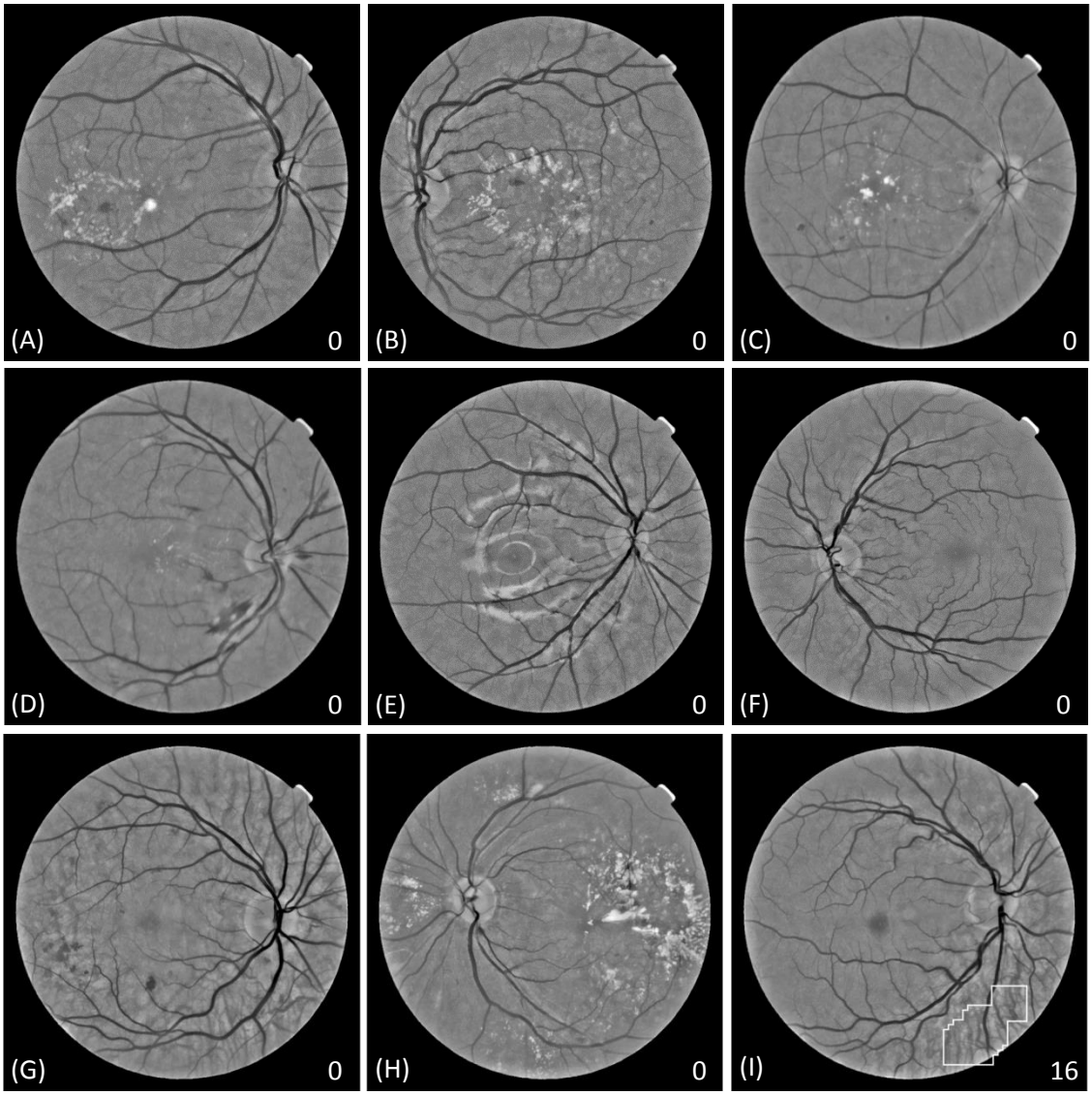
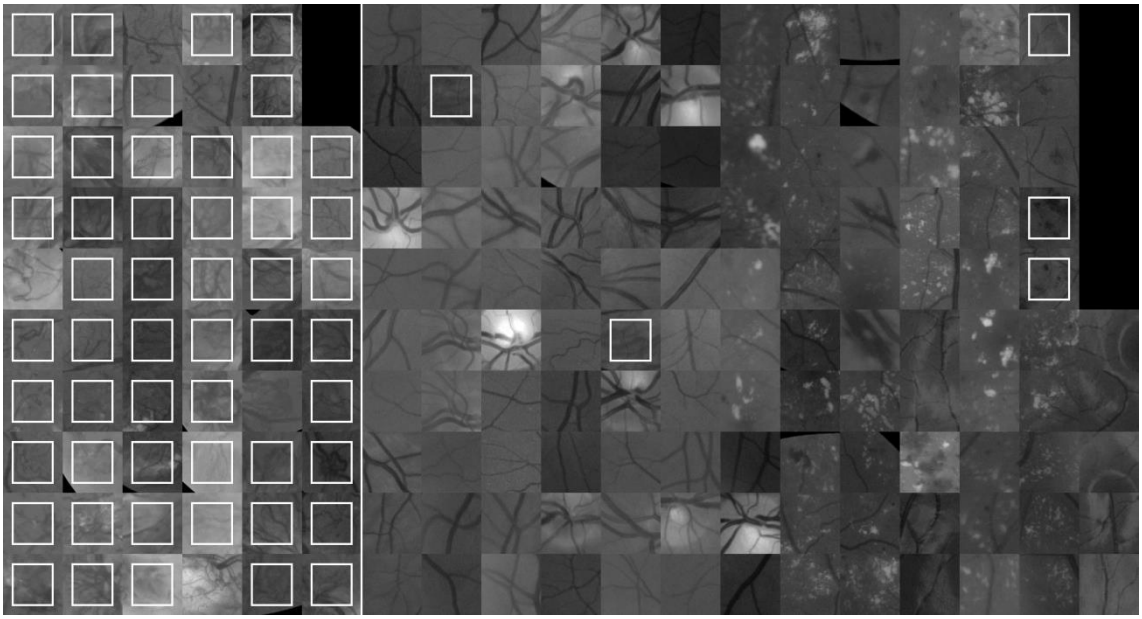


Figure 4.10: Results of the proposed system on a per image basis. **(A)-(H)** True negative images. **(I)** False positive image.

Table 4.5: The number of candidate pixels classified as new vessels and the number of correctly detected new vessel regions during assessment on a per image basis.

Image N^0	Manually marked new vessel regions	Candidate pixels classified as new vessels	Correctly detected new vessel regions
1	3	13	2
2	4	59	3
3	5	4	2
4	7	83	5
5	2	57	1
6	3	45	2
7	1	24	1
8	2	31	2
9	2	35	1
10	3	12	1
11	2	7	1
12	5	124	3
13	2	3	1
14	2	8	1
15	1	2	1
16	3	106	1
17	1	56	1
18	3	98	2
19	2	3	2
20	3	60	1
21	0	0	-
22	0	0	-
23	0	16	-
24	0	0	-
25	0	0	-
26	0	0	-
27	0	0	-
28	0	0	-
29	0	0	-
30	0	0	-
31	0	0	-
32	0	0	-
33	0	0	-
34	0	0	-
35	0	0	-
36	0	0	-
37	0	0	-
38	0	0	-
39	0	0	-
40	0	0	-
41	0	0	-
42	0	0	-
43	0	0	-
44	0	0	-
45	0	0	-
46	0	0	-
47	0	0	-
48	0	0	-
49	0	0	-
50	0	0	-
51	0	0	-
52	0	0	-
53	0	0	-
54	0	0	-
55	0	0	-
56	0	0	-
57	0	0	-
58	0	0	-
59	0	0	-
60	0	0	-



(A) (B)

Figure 4.11: Results of the proposed system on a per patch basis. **(A)** New vessel patches, 53 TPs and 5 FNs out of 58 patches. **(B)** Non new vessel patches, 120 TNs and 5 FPs out of 125 patches. White boxes indicate patches classified with a new vessel label.

The MATLAB Code took 721 seconds on an Intel(R) core(TM)2 Quad CPU Q9300 at 2.5 GHz to process each image using the full 21-D feature vectors. This was reduced to 527 seconds once the feature vectors had been reduced in dimensionality.

4.3 Discussion and Conclusion

In this chapter we have presented an extension to the dual classification new vessel detection method described in chapter 3, by including feature vector expansion and feature selection. This involves the exploration of a 21-D feature vector which contained morphology, intensity and gradient based features.

The first main contribution of this chapter is the creation of a large feature set in combination with the dual classification method. The 21-D feature set looks at many different aspects in order to find suitable features for discrimination, and these include morphology, intensity and gradient based features. The Wilcoxon test results in table 4.2 shows that morphology based features generally appear to have better

discrimination ability than the gradient and intensity based features. The original 4 morphology features from chapter 3, as well as the morphology features of compactness, number of bifurcation points and tortuosity max all possess good discrimination ability for both classifications. Other features show particularly good discrimination ability for just a single classification; for example, features 19 and 20, which are based on the local grey levels, have a significantly better discriminating ability for classification 2 than for classification 1.

The second main contribution of this chapter is the selection of features in combination with the dual classification method. A small feature subset not only has the potential to improve the classifier's performance, but is also better for generalization. However, with feature selection, overfitting can still occur and some techniques are more prone than others. The results from the Wilcoxon test were not used to apply a filter approach for feature selection as this approach ranks features individually and therefore ignores feature dependencies. For example, after filtering only high rank features would remain and some of these features may be highly correlated. Thus resulting in the possible inclusion of many redundant features and therefore failing to detect the most useful features. Wrapper methods are more thorough at feature selection as they interact with the classifier and can model feature dependencies. GA is a wrapper method, and therefore it has a better chance of finding the most useful features than the filter approach; however it is more at risk of overfitting. GA is preferred to greedy search strategies (popular wrapper methods) as GA is less prone to local sub-optima. The GA approach is also highly beneficial as it allows for SVM parameter selection to be incorporated.

Another advantage of the GA approach is its ease of dealing with feature selection for two feature vectors. It explores, using the final decision achieved from combining the two classification outcomes, both feature vectors simultaneously (each feature vector has its own combination of features). An alternative would be to use a feature selection approach to explore each feature vector and its corresponding classifier separately. However, this would include the limitation of only being able to use test data that the specific individual classifier was designed to discriminate.

Table 4.3 provides the selection achieved from this GA based approach. The selected feature subsets are relatively small in size. On a per image basis the feature subsets contained 7 features for each classification. Feature subsets of 6 features and 9 features are selected on a per patch basis. The RBF kernel with a γ of 1 was selected on both levels (image and patch).

From table 4.3, it is apparent that feature subsets contained a mix of features, but are more heavily orientated towards morphology based features. Many of the features that were deemed to have a high discriminating ability by the Wilcoxon test have been excluded from the feature subset due to their redundancy. Also it's apparent that almost all features possessing a low discriminating ability by the Wilcoxon test have been excluded due to them being irrelevant. However, each feature subset does still retain a couple of lower ranked features. This was expected because features which are completely irrelevant individually can provide significant performance improvements when considered with other features.

The proposed method does achieve better performance metrics than most of the other published methods, as shown in table 4.4. However, as stated in chapter 3, comparisons are difficult to make as there exists variability in terms of their application and no standard datasets exist for testing. The results in table 4.4 show the performance of the proposed system is superior to that of chapter 3, achieving a sensitivity of 100.00% and a specificity of 97.50% compared to a sensitivity of 100.00% and a specificity of 92.50% on a per image basis. The proposed method also achieves a maximum accuracy of 94.54% with a sensitivity of 91.38% and a specificity of 96.00% compared to a maximum accuracy of 92.35% with a sensitivity of 87.93% and a specificity of 94.40% for the method in chapter 3 on a per patch basis. From the examples of classified images shown in figure 4.9 and figure 4.10 and the classified patches shown in figure 4.11 it is clear that the algorithm responds well to a range of new vessel formations and has the ability to avoid false responses despite the presence of other pathology (bright lesions, dark lesions etc.) and reflection artefacts.

The results that achieve maximum performance on each level (image and patch) do not correspond to the same operating point. This was also documented in chapter 3. The reported per image performance of 100.00% and 97.50% for sensitivity and specificity, respectively, corresponds to a per patch performance of 31.03% and 99.20% for sensitivity and specificity, respectively. This illustrates that identifying any part of any new vessel region in the image is sufficient for the image to achieve a new vessel label.

From figure 4.11 it is clear that the emphasis on a per patch basis is to detect as many new vessel patches as possible. This further highlights how the process at the two levels (image and patch) differs. The actual classification process in terms of SVM boundaries is designed to label pixels, and this equates to a patch level (recall a pixel can be referred to with the term patch). Therefore, unlike on a per image basis, enhancing the results on a per patch basis is based on a direct enhancement of the SVM boundaries. The fact that the features selected for the per image and per patch basis (table 4.3) vary significantly further suggests that the requirements for classification on each level differ significantly. Evaluation on a per image basis may be appropriate for current clinical application; however, from a computer science point of view the development and evaluation from a per patch basis should guide future work.

When comparing the figures for the results on a per patch basis from chapter 3 and chapter 4, it is apparent that the FNs and FPs patches of figure 4.11 are not a direct subset of the FNs and FPs patches in figure 3.20. Four out of the five FNs from figure 4.11 match those from figure 3.20 and four out of five FPs from figure 4.11 match those from figure 3.20. Direct subsets of the FNs and FPs patches were not expected as the methodology in chapter 4 is not a straight forward extension to the methodology of chapter 3. This is evident from table 4.3, in which the final feature vectors used in chapter 4 are not created by simply adding to those from chapter 3.

Future developments of this method will include further exploration into the selection of SVM parameters. Currently a single set of SVM parameters were selected by the GA based system for both classifications 1 and 2. These are two

separate classifications with their own distinct feature sets, and therefore the optimal SVM parameters should be searched individually for each. This should lead to further enhancement of the classifier's performance. This holds the risk of causing overfitting, whilst our current approach is better in terms of generalization. Another aspect that needs attention is the number of generations. This value needs to be set significantly higher to explore whether the fitness function could be further increased. Current limitations restricting the number of generations are the computational expense of performing assessment by leave-one-out cross validation and the large amount of operating points to assess that arise from dual classification. Note that evaluation has to be performed for every single chromosome in the GA based approach. These limitations were the reason why the optimization problem was simplified with the use of discrete values to represent continuous variables.

This chapter provides a framework to demonstrate the use of GAs for feature and parameter selection. However, there is potential to enhance the system's performance through the further exploration of GA settings. This ranges from the techniques used for selection, crossover and mutation to the rates used for crossover and mutation.

In conclusion, this chapter has further enhanced the automated system that is capable of detecting the presence of new vessels whilst reducing false responses to bright lesions, dark lesions and reflection artefacts. This involved the use of feature vector expansion followed by a genetic algorithm based approach to perform feature selection and SVM parameter selection.

5 PROLIFERATIVE DIABETIC RETINOPATHY DETECTION USING DUAL ENSEMBLE CLASSIFICATION

This chapter further explores the framework of dual classification, which is now combined with an alternative classifier and feature selection technique in order to investigate if classification accuracy can be improved. The SVM classification used in the previous chapters is replaced with a new supervised method for the detection of new vessels by using an ensemble classifier of boot strapped (bagged) decision trees. The classifier based on the boot strapped decision trees is a classic ensemble classifier, which has been broadly applied in many application areas of image analysis [39], but has not been utilized for detecting neovascularization. An important feature of the bagged ensemble classifier is that the classification accuracy can be estimated during the training phase, without supplying the classifier with test data. The importance of each feature in the classification can also be predicted during the training phase, and this information allows for a feature selection procedure to be performed. Feature selection is performed on the same 21-D feature vector derived in chapter 4.

The general framework remains very similar to that depicted in figure 3.1 and is shown in figure 5.1. Work from this chapter was used to create the publication [137].

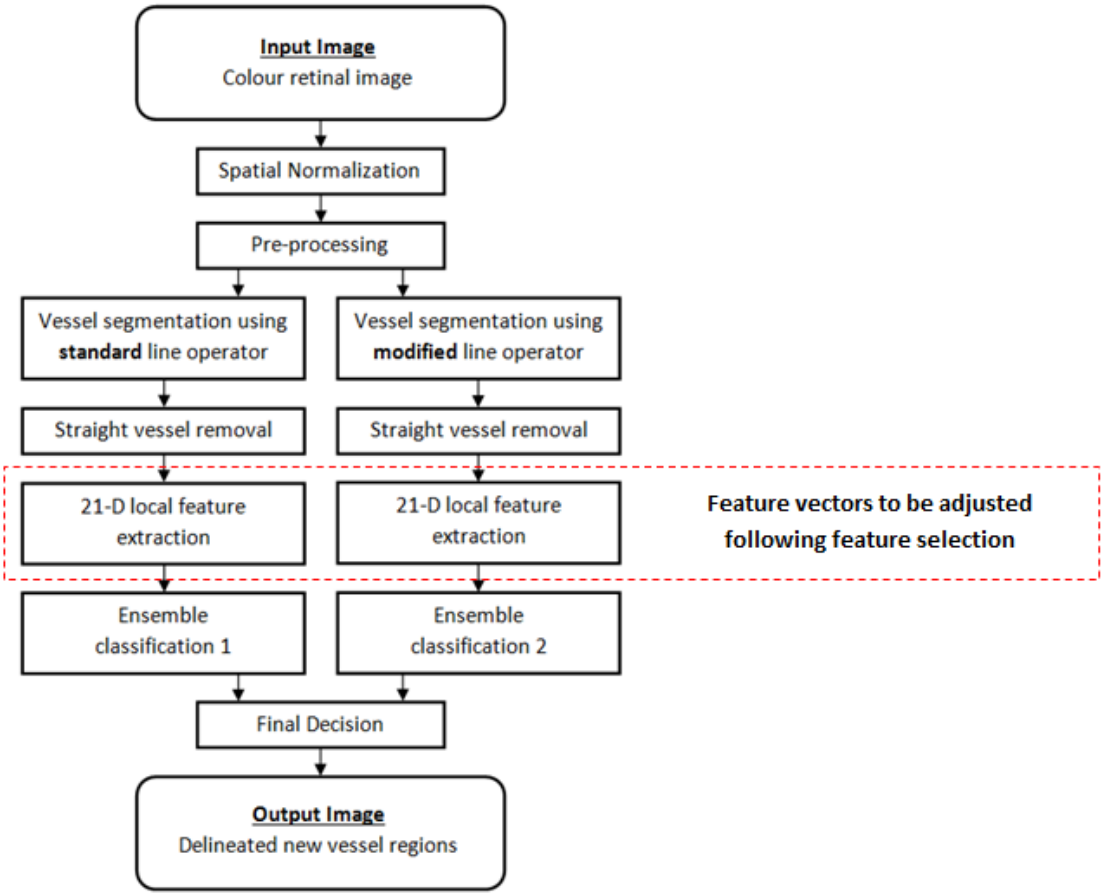


Figure 5.1: System architecture.

5.1 Methodology

5.1.1 Dual Ensemble Classification

This follows the same mechanism of dual classification as described in the previous chapters, but SVM classification is replaced with ensemble classification. Two separate 21-D feature vectors were produced. All features were normalised so that each feature had zero mean and unit standard deviation. Independent classification was performed for each feature vector using an ensemble system of bagged decision trees. These two independent outcomes were then combined to a produce a final decision. The classification worked on a pixel level. When complete, all pixels labelled as new vessels were morphologically dilated with a structuring element the size of the sub-window to illustrate the new vessel regions.

5.1.2 Ensemble Classifier

The process of consulting multiple experts or seeking multiple opinions ahead of final decision making is almost our second nature. The wide-spread gains of such a procedure in automated decision making applications give rise to the ensemble classification framework.

In ensemble classification [115], multiple classifiers and models are tactically generated and combined in order to give the solution to a machine learning problem, with a goal of obtaining better performance than could be obtained from any of the constituent classifiers/models. This process is used to abbreviate the likelihood of inadequate or unfortunate selection while improving the performance of the classifier. We use this strategy instinctively in our day to day activities, where we consider the opinion from several experts, evaluate and mingle their recommendations for establishing a well optimized and well-versed conclusion. In the same manner, the ensemble methods utilize multiple classifiers/models to accomplish gain in classification performance by mixing/aggregating the outcomes from several weak learners into one high-class classifier, with the goal of reducing the variance and amplifying the confidence in the decision. In this approach we used the decision trees as the classification model and the results of these weak learners were combined using Bootstrap aggregation, also known as bagging.

Breiman's bagging [138], also known as bootstrap aggregating, is among the most primitive ensemble methods, which is most perceptive and simplest to implement, with surprisingly fine results. In bagging, the component classifiers (in this case the decision trees) are developed on the bootstrap replicas of the training dataset. This is done by a random selection of "N" training instances out of "N" with replacement, where *"N"* is the size of the training set. Majority voting is used to combine the responses of the individual component classifiers (the decision trees). The decision of the ensemble is the class chosen by most of the component classifiers. Traditionally, the component classifiers are of the same general form; for example, all hidden Markov model, all neural networks or all decision trees, which was the case in this work.

Given the original training set T, multiple sets of training data T_b are created, where $b = 1,2, \ldots . . B$, by randomly sampling T with replacement. "B" is the number of component classifiers used in the ensemble system. On average, each training set T_b only contains two-thirds of the original samples. The bagging algorithm as explained in [139] is illustrated below.

Inputs for bagging algorithm

- Training data $T = \{x_1, x_2, \ldots \ldots \ldots, x_N\}$, where $x_i \in X$ is the feature vector of the ith instance in the feature space X, provided with correct class labels $w_i \in \Omega = \{w_1, \ldots \ldots, w_C\}$ for a C-class problem. N is the size of the training set.
- WeakLearn, (the weak learning algorithm).
- The Integer B, specifies the total number of iterations.

Do b = 1 . . . . B

1. Acquire the bootstrap sample T_b by randomly choosing N observations, with replacement, from the training set T.
2. Call the routine WeakLearn with T_b and obtain the hypothesis from the classifier
   $h_b: X \to \Omega$.
3. Add h_b to the ensemble, E.

End Do Loop

Test the algorithm with simple majority voting

Given the unlabelled data instance z

1: Estimate the ensemble $E = \{h_1, \ldots \ldots, h_B\}$ on z.

2: Suppose the classifier h_b gives its vote to the class w_j as,

$$v_{b,j} = \begin{cases} 1 & \text{if } h_b \text{ picks class } w_j \\ 0 & \text{otherwise} \end{cases},$$

3: The total votes obtained by each of the classes are:

$$V_j = \sum_{b=1}^{B} v_{b,j} \ , \quad j = 1, \dots, C.$$

4: The final classification/label is the class that gains the majority of the votes.

5.2 Experimental Evaluation

5.2.1 Materials and Performance Measures

The materials and performance measures used for this chapter are identical to those detailed in chapters 3. To summarize:

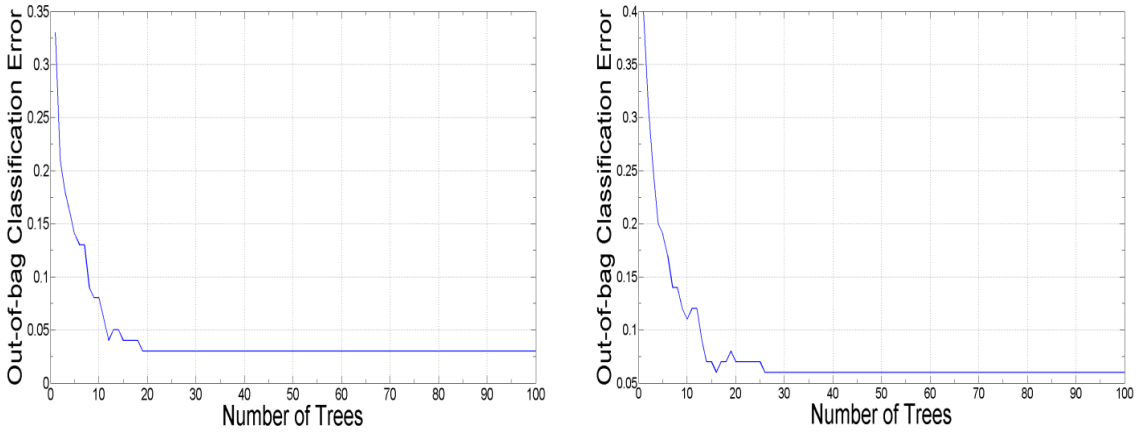
- 60 image dataset (new vessel images = 20, non-new vessel images = 40).
- Leave-one-out cross validation.
- Selection of pixels used for training data (selection remains identical).
- Evaluation on a per image and per patch basis.
- Performance metrics of sensitivity (SN), specificity (SP), accuracy (Acc), area under the ROC curve (AUC).
- 3D ROC surface used to assess dual classification.
- Per patch basis, optimal operating point = maximum accuracy.
- Per image basis, optimal operating point = highest specificity at a sensitivity of 100%.

5.2.2 Ensemble Classifier Evaluation

An important feature of the bagged ensemble is that the classification accuracy can be estimated during the training phase without supplying the test data. Moreover, the importance of each feature in classification can also be predicted during the training phase. The estimates of classification accuracy and the feature importance during the training of the classifier is a smart feature of bagging.

5.2.2.1 Out-of-Bag Classification Error

In Bagging, the component classifiers (in this case the decision trees) were developed on the bootstrap replicas of the training dataset. This was done by a random selection of N training instances out of N with replacement, where N is the size of the training set. The phenomenon of choosing the N out of N training instances with replacement leaves out 37% of instances on average for every component classifier in the ensemble. The left-overs were called the out-of-bag (OOB) observations and were used in the estimation of the predictive power of the ensemble. The OOB predicted responses were compared against the observed responses of all the training instances to estimate the average OOB error, which is an unbiased estimator of the true ensemble error. Figure 5.2(A) and figure 5.2(B) show the plot of the OOB classification error computed for the classifiers comprised of 100 decision trees for each of the classifications respectively. 25 and 30 trees were the optimal choices for classification 1 and classification 2, respectively, as OOB classification error remained constant with further increase in the number of decision trees in the ensemble.



(A) (B)

Figure 5.2: Out-of-Bag classification error for (A) classification 1 and (B) classification 2.

5.2.2.2 Feature Analysis

For establishing the importance of each feature, the feature importance index was computed during the classifier training phase. The OOB observations were used to determine the significance of each individual feature from the feature vector in the

classification. In order to predict the feature importance, the OOB data was randomly permutated across one variable for all instances and the increase in OOB error due to this permutation was estimated. This increase was directly proportional to the importance of the feature in the classification. The larger the increase, the more important the feature was in the classification.

Table 5.1: Summary of feature vector.

Feature Number	Feature Name	Feature Number	Feature Name
1	Number of vessel pixels	12	Gradient coefficient of variation
2	Number of vessel segments	13	Line strength mean
3	Number of vessel orientations	14	Vessel width mean
4	Vessel density mean	15	Vessel wall gradient mean
5	Tortuosity mean	16	Vessel wall gradient coefficient of variation
6	Tortuosity max	17	Compactness
7	Vessel length mean	18	Linkage
8	Number of bifurcation points	19	Local grey level mean
9	Grey level mean	20	Local grey level max
10	Grey level coefficient of variation	21	Local grey level variation
11	Gradient mean		

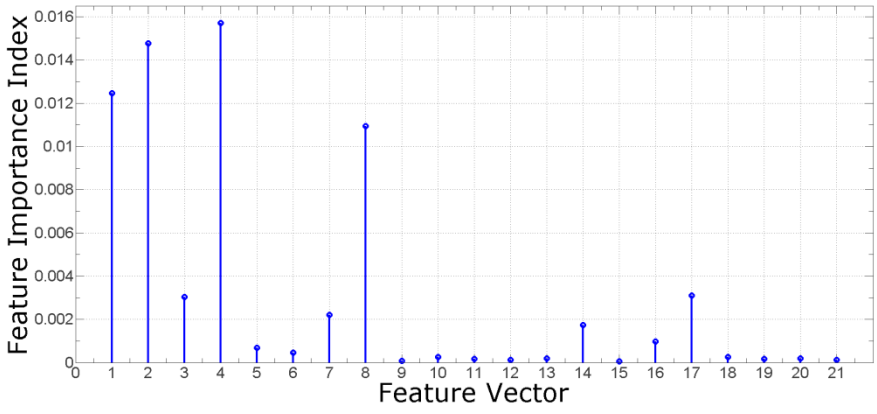


Figure 5.3: Feature Importance Index, classification 1.

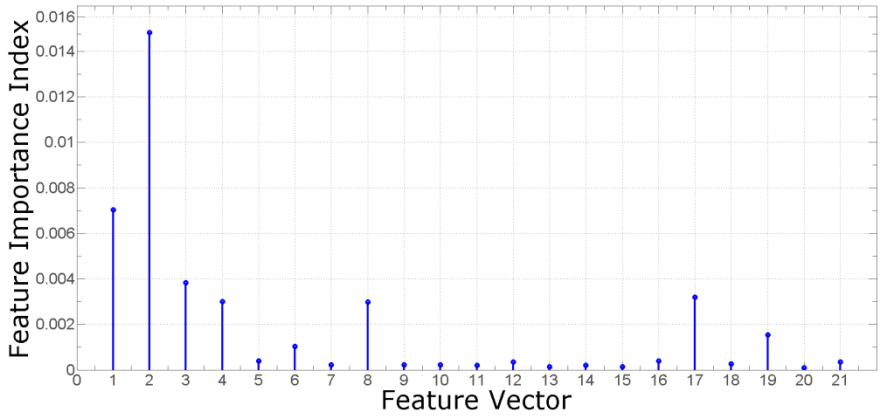


Figure 5.4: Feature Importance Index, classification 2.

Figure 5.3 and figure 5.4 shows the graphs for feature importance index (FII) calculated from classifiers created with 100 decision trees. The lowest ranked feature was removed (without return) from each feature vector simultaneously and the system's performance was assessed from the final decision achieved from combining the two classification outcomes. This procedure continued until there was any drop off in the optimal operating point, and the feature vectors prior to this drop off were chosen. This resulted in 33% of the most significant features being kept for both feature vectors for evaluation on a per image basis and 24% of the most significant features were kept for both feature vectors for evaluation on a per patch basis.

5.2.3 Results

ROC curves of the proposed system for evaluation on a per image basis and per patch basis are depicted in figures 5.5(A)-(B). These represent the performance using the feature vectors achieved from feature selection. The features selected are shown in table 5.2. The AUC value for the per image basis is 0.9505. The optimal operating point according to the application specific performance measure is a sensitivity of 100.00% and a specificity of 95.00%. For a per patch basis the AUC value is 0.9308. The operating point with maximum accuracy of 90.71% gives a sensitivity of 81.03%

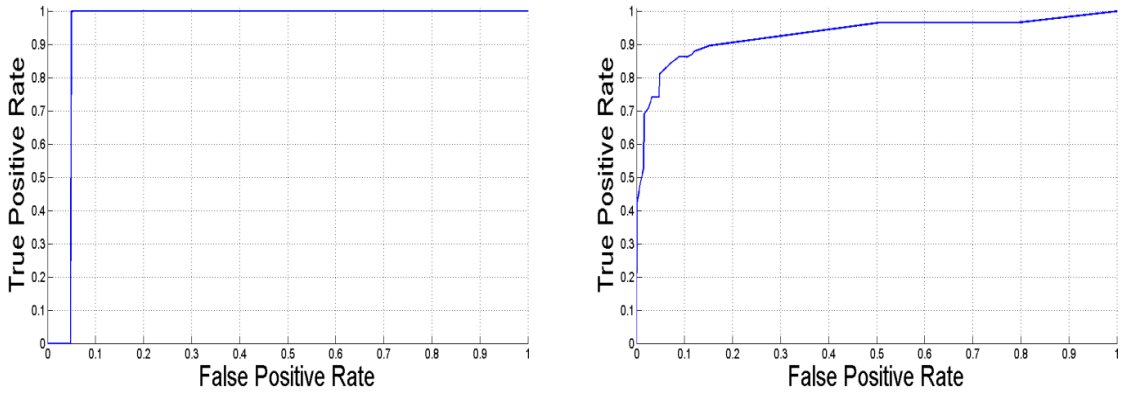
and a specificity of 95.20%. Table 5.3 shows these results along with the reported results from other new vessel detection methods.

Examples of classified images are given in figure 5.7 and figure 5.8. Recall, the number in the bottom right hand corner of the figures specifies the number of candidate pixels classified as new vessels. All positive pixels were then morphologically dilated with a structuring element the size of the sub-window used in feature extraction, and the resultant was the delineation of new vessel regions (indicated with a white boundary). Any image containing any candidate pixels classified as new vessels, and therefore any delineated region is classified as a new vessel image. As stated earlier, the performance evaluation of delineation was not assessed. However, figure 5.6 provides the manual delineation of new vessel regions marked by an ophthalmologist in order to allow for a visual comparison to the delineation shown in figure 5.7. In addition to this, table 5.4 provides the number of candidate pixels classified as new vessels and the number of new vessel regions that have been successfully detected for each image in the dataset. The image numbers in table 5.4 correspond to those listed in appendix II, in which images 1-20 are PDR images and 21-40 are non-PDR images.

The MATLAB Code took 683 seconds on an Intel(R) core(TM)2 Quad CPU Q9300 at 2.5 GHz to process each image using the full 21-D feature vectors. This was reduced to 498 seconds once the feature vectors had been reduced in dimensionality.

Table 5.2: Selected features.

Level	Classification	Features
Image	1	4, 2, 1, 8, 17, 3, 7
	2	2, 1, 3, 17, 8, 4, 19
Patch	1	4, 2, 1, 8, 17
	2	2, 1, 3, 17, 8



(A) (B)

Figure 5.5: ROC curves. **(A)** Performance on a per image basis. **(B)** Performance on a per patch basis.

Table 5.3: Reported results for new vessel detection methods.

Algorithm	SN	SP	Acc	AUC	Level
Hassan [70]	63.90	89.40	-	0.7045	Pixel
Jelinek [71]	94.00	82.00	-	0.900	Image
Goatman [72]	84.2	85.9	-	0.911	Image
Arputham [73]	84.70	86.10	-	-	Image
Pavai [74]	88.89	91.30	-	-	Image
Akram [76]	98.93	96.35	-	-	Segment
Akram [77]	98.00	97.00	98.00	0.980	Segment
Akram [77]	96.00	94.00	0.95	-	Image
Saranya [78]	96.25	89.65	96.53	-	Image
Welikala [79]	100.00	70.00	-	-	Image
Nithyaa [86]	-	-	-	0.947	Image
Agurto [87]	100.00	88.00	-	0.980	Image
Agurto [88]	96.00	83.00	-	0.940	Image
Vatanparast [90]	99.62	96.61	-	-	Patch
Lee [91]	96.30	99.10	98.50	0.993	Image
Chapter 3	100.00	92.50	95.00	0.9693	Image
Chapter 3	87.93	94.40	92.35	0.9616	Patch
Chapter 4	100.00	97.50	98.33	0.9914	Image
Chapter 4	91.38	96.00	94.54	0.9600	Patch
Proposed (21 features)	100.00	95.00	96.67	0.9734	Image
Proposed (21 features)	79.31	95.20	90.16	0.9557	Patch
Proposed (7 features)	100.00	95.00	96.67	0.9505	Image
Proposed (5 features)	81.03	95.20	90.71	0.9308	Patch

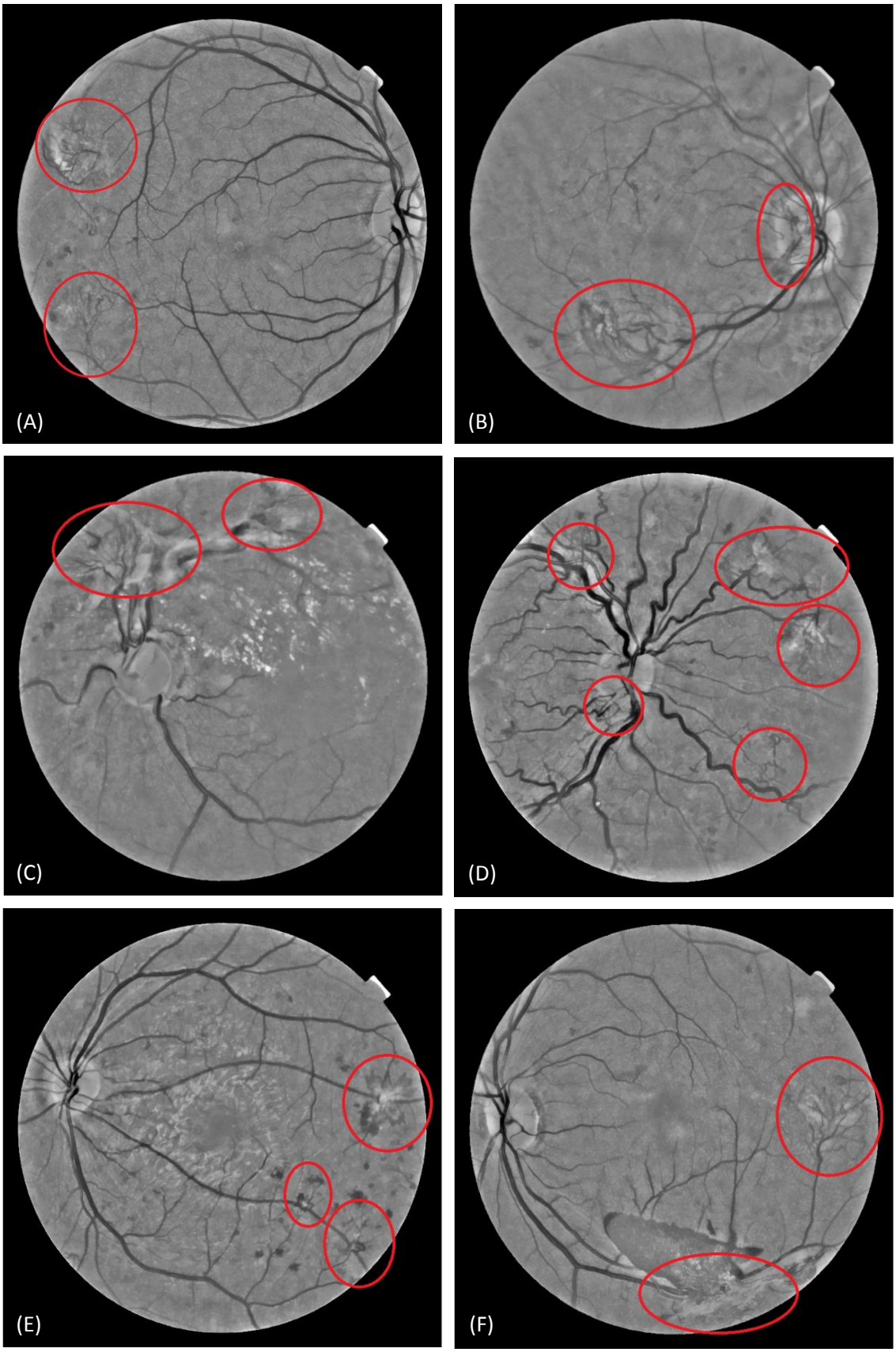


Figure 5.6: Location and delineation of new vessel regions marked by an ophthalmologist.

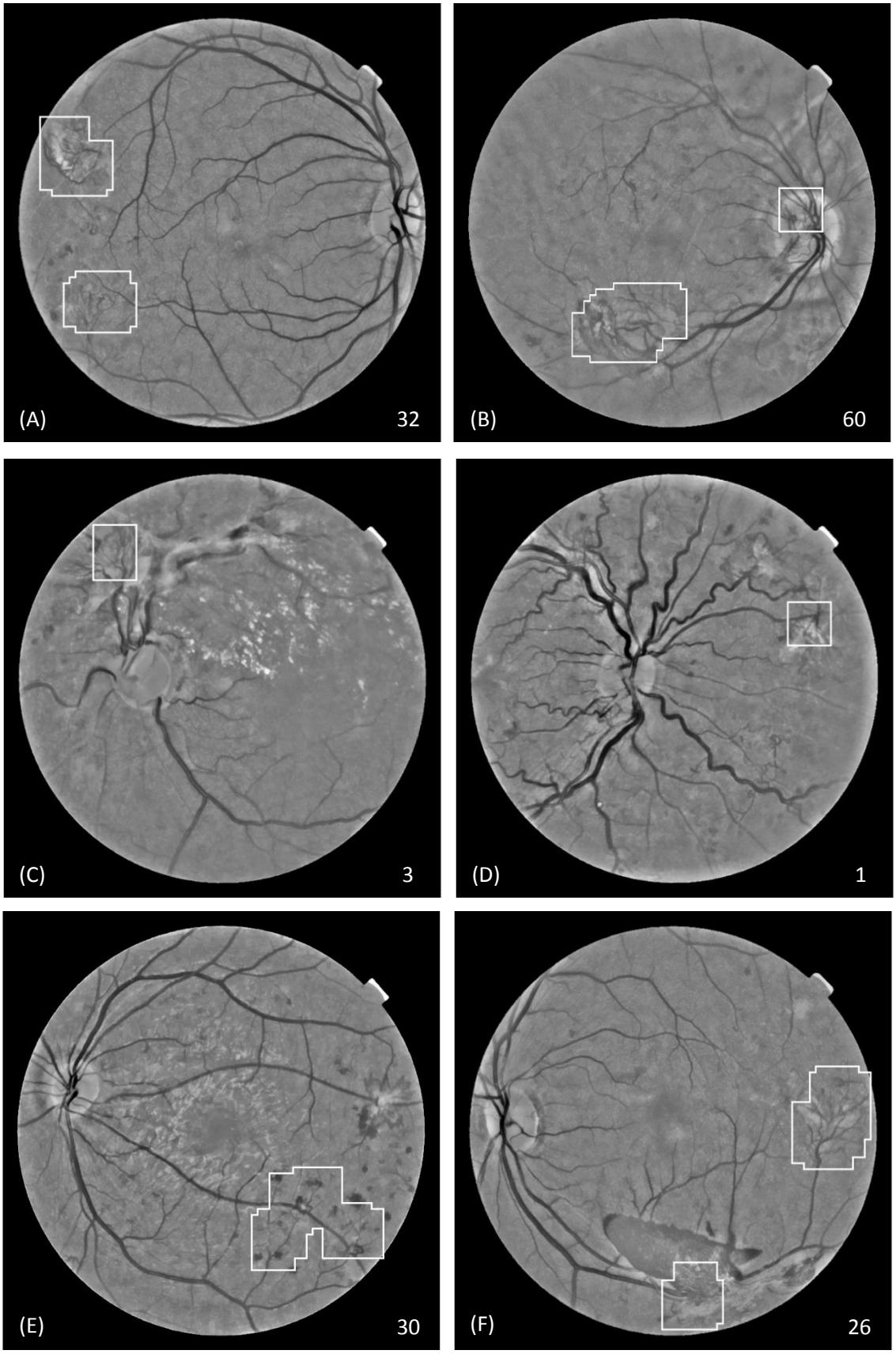


Figure 5.7: Results of the proposed system on a per image basis. True positive images. **(A)-(F)** Corresponds to figures 5.6**(A)-(F)** respectively.

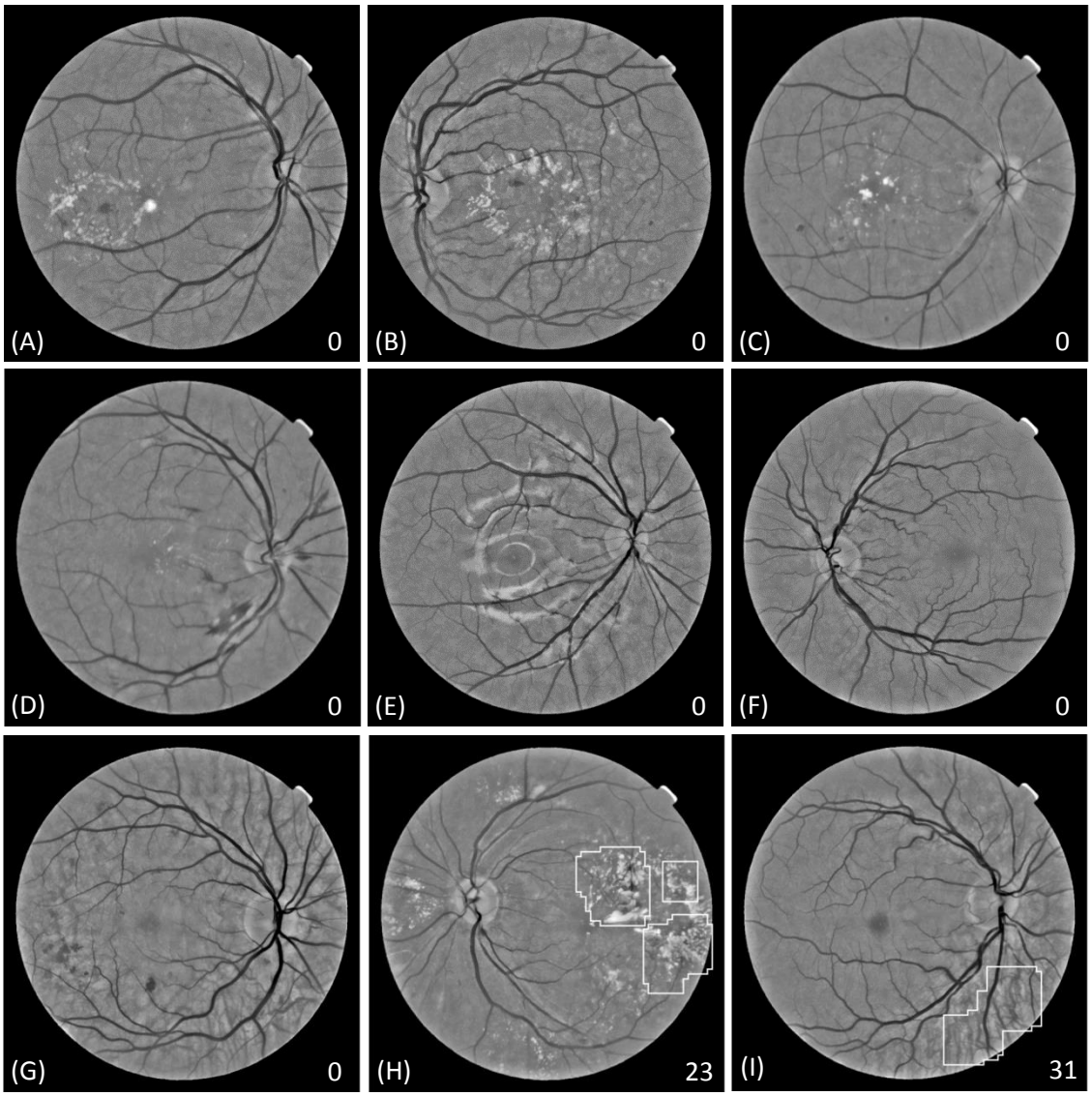


Figure 5.8: Results of the proposed system on a per image basis. **(A)-(G)** True negative images. **(H)-(I)** False positive images.

Table 5.4: The number of candidate pixels classified as new vessels and the number of correctly detected new vessel regions during assessment on a per image basis.

Image N^0	Manually marked new vessel regions	Candidate pixels classified as new vessels	Correctly detected new vessel regions
1	3	30	2
2	4	30	2
3	5	1	1
4	7	165	3
5	2	46	2
6	3	62	1
7	1	31	1
8	2	32	2
9	2	60	2
10	3	23	1
11	2	3	1
12	5	157	3
13	2	19	1
14	2	37	1
15	1	3	1
16	3	60	1
17	1	176	1
18	3	133	2
19	2	26	2
20	3	148	2
21	0	0	-
22	0	0	-
23	0	31	-
24	0	0	-
25	0	0	-
26	0	0	-
27	0	0	-
28	0	0	-
29	0	0	-
30	0	0	-
31	0	0	-
32	0	0	-
33	0	0	-
34	0	0	-
35	0	0	-
36	0	0	-
37	0	0	-
38	0	0	-
39	0	0	-
40	0	0	-
41	0	0	-
42	0	0	-
43	0	0	-
44	0	0	-
45	0	0	-
46	0	0	-
47	0	0	-
48	0	0	-
49	0	0	-
50	0	0	-
51	0	0	-
52	0	0	-
53	0	0	-
54	0	0	-
55	0	0	-
56	0	23	-
57	0	0	-
58	0	0	-
59	0	0	-
60	0	0	-

5.3 Discussion and Conclusion

In this chapter we have further investigated the dual classification new vessel detection method. An alternate classifier and feature selection technique were applied in order to provide a comparison with the results from the previous chapter. This classifier has previously not been utilized for detecting neovascularization and was tested on the basis of its promising performance from other work in the field of retinal image analysis [39].

The contribution of this chapter was the integration of ensemble classification which was compared to the support vector machine classification used in chapters 3 and 4. The ensemble method utilized multiple classifiers to accomplish gain in classification performance by mixing the outcomes from several weak learners into one high-class classifier. Decision trees were used as the classification model and the results of these weak learners were combined using bootstrap aggregation (bagging). A feature of the bagged ensemble was its ability to predict the importance of each feature during the training phase. This was useful to identify the most relevant features, allowing for feature selection to be performed. An advantage of this feature selection procedure was its ease of implementation in comparison to the GA based approach of chapter 4.

From table 5.3 it is evident that the proposed method does achieve better performance metrics than most of the other published methods. However, as stated in chapter 3 and 4, comparisons are difficult to make as there exists variability in terms of their application and no standard datasets exist for testing. The results from table 5.3 show that the performance of the proposed system is inferior to that of chapter 4, achieving a sensitivity of 100.00% and a specificity of 95.00% compared to a sensitivity of 100.00% and a specificity of 97.50% on a per image basis. Also, a maximum accuracy of 90.71% with a sensitivity of 81.03% and a specificity of 95.20% is achieved compared to a maximum accuracy of 94.54% with a sensitivity of 91.38% and a specificity of 96.00% on a per patch basis. However, the performance of the proposed system is superior to that of chapter 3 on a per image basis, achieving a sensitivity of 100.00% and a specificity of 95.00% compared to a sensitivity of

100.00% and a specificity of 92.50%. From the examples of classified images shown in figure 5.7 and figure 5.8 it is clear that the algorithm responds well to a range of new vessel formations and has the ability to avoid false responses despite the presence of bright lesions, dark lesions and reflection artefacts.

Following feature selection using the feature importance index, the system's performance on a per image basis remained the same with a sensitivity of 100.00% and a specificity of 95.00%. On a per patch basis, feature selection provided a small improvement in performance with the maximum accuracy moving up from 90.16% to 90.71%. Therefore, feature selection did not provide significant improvements in respect to its main objective (classification performance). However, the feature vectors were significantly reduced in dimensionality which in turns decreases the computational cost of feature generation, classifier training and classification. On a per image basis more than 66% of features were removed from the 21-D feature vector, with only 7 features being retained for each feature vector. This resulted in the removal of 27% of the computational time needed to process each image. However, the feature selection technique in chapter 4 resulted in similar reductions in computational time. On a per patch basis more than 76% of the features were removed, retaining only 5 features for each feature vector.

The feature importance index feature selection technique used in this chapter, like that of chapter 4, interacts with the classifier and models feature dependencies. However, unlike that of chapter 4, the selected features more closely correlate to those that would have been achieved by applying a filter approach to the Wilcoxon rank sum test results (see section 4.2.2). The selected features are almost all morphological based features, whereas those selected in chapter 4 contained slightly more of a mixture of features.

Whilst the performance of the proposed system is inferior to that of chapter 4, ensemble classification does show potential for use in new vessel detection. Prior to feature selection, with use of the full 21-D feature vector, dual ensemble classification (chapter 5) achieved a sensitivity of 100% and a specificity of 95.00%, outperforming the dual SVM classification (chapter 4) which could only achieve a

sensitivity of 100.00% and a specificity of 82.50%. This suggests that the ensemble classification is highly effective. However, the proposed system of chapter 4 employs a superior feature selection technique which then allows it to overtake and outperform the proposed system of this chapter.

Both feature selection techniques from chapters 4 (GA based approach) and this chapter (OOB feature importance index based approach) share similarities in their application. These are: allowing both feature vectors to be altered simultaneously, where each feature vector has its own combination of features and assessment is performed using the final decision achieved from combining the two classification outcomes. The difference is that the approach of chapter 4 is a lot more thorough in its exploration of the combinations of features. The approach from chapter 5 is a lot more restricted in its exploration, where the number of features from both feature sets are fixed to be equal and supposedly weaker features are removed early on without any reconsideration.

As extensively documented in chapter 3 and 4, the results that achieve maximum performance on each level (image and patch) do not correspond to the same operating point. The reported per image performance of 100.00% and 95.00% for sensitivity and specificity, respectively, corresponds to a per patch performance of 34.48% and 99.20% for sensitivity and specificity, respectively. Therefore, the classification requirements on each level differ significantly. This is further shown by the number of features being used at each level being different.

Further developments of this method will involve the inclusion and evaluation of the boosting technique for comparison to our current methodology. Boosting [140] also creates an ensemble of classifiers by re-sampling the data, which is then combined by majority voting, but it takes a different re-sampling approach than bagging. Another aspect to be explored would be to assess whether the results of the dual ensemble classification could be improved with the use of the genetic algorithm based feature selection approach.

In conclusion, this chapter has further justified the use of the algorithm's framework incorporating the use of the SVM classifier and a genetic algorithm based feature

selection approach. This was achieved by showing its performance to be superior to that of a framework with the incorporation of an alternative state-of-the-art classifier, ensemble classification of boot strapped decision trees.

6 DISCUSSION AND CONCLUSION

In this thesis we have presented the development of a methodology designed for the automated detection of proliferative diabetic retinopathy (PDR). This chapter includes the justification of the project, main contributions, results, final conclusions and future work.

Diabetic retinopathy (DR) is a sight threatening disease. The DR screening process allows for the early detection of the disease and therefore allows for timely intervention in order to prevent vision loss. The integration of automated detection systems has numerous benefits including significantly reducing the manual grading workload and helping towards ensuring time targets are met for referrals. The grading pathways (figures 1.9 and 1.10) stress the importance of the detection of R3 (PDR) for fast tracking due to their high risk of severe loss of vision. Therefore, if automated systems are to take on a substantial role in the grading pathway then they must be capable of detecting PDR. This supports the efforts of this project into the automated detection of new vessels (PDR), which is a research area that hasn't received the attention it deserves.

From the literature review and initial investigations, the best course of action was to build a framework based on vessel segmentation tailored towards PDR, followed by analysing the characteristics of the segmented binary vessel map in search of abnormality. This approach was followed, but with the novelty of creating two different segmented vessel maps, which works towards the objective of detecting new vessels whilst reducing false responses caused by bright lesions and other retinal features. Segmentation methods included a standard line operator approach and a

novel modified line operator approach. The former targeted accurate segmentation of new vessels and the latter targeted the reduction of false responses to non-vessel edges. Each segmented map held vital information, and therefore both maps were used in the detection of new vessels. The performance of the algorithm was shown to be superior when the information extracted from each of these two maps were processed separately using a dual SVM classification framework as opposed to single SVM classification and multiclass SVM classification. SVM classification has gained in popularity in the field due to its high performance accuracy and was therefore chosen as the classification technique of preference.

When extracting information from the segmented vessels maps in the form of local characteristics, it was evident that high vessel area alone was not enough to identify new vessels. A more detailed analysis of the local characteristics of the vasculature was required as well as the removal of straight vessels. This led to the initial creation of a 4-d feature vector to be used by the SVM classifiers, which was later expanded to a 21-d feature vector which included morphology, intensity and gradient based features. Feature vector reduction was performed in order to enhance the performance of the system as well as save on computation time. This was performed by using a genetic algorithm (GA) based feature selection approach which was very thorough in its search for the optimum feature subsets and also had the capability of optimising the parameters of the SVM classifiers. This GA based approach also had the advantage of being able to thoroughly explore both feature vectors simultaneously. Following feature selection, the derived feature subsets were heavily orientated towards morphology based features.

The proposed methodology achieves a good standard of results for the detection of new vessels, in which it responds well to a variety of new vessel formations and has the ability to avoid false responses despite the presence of bright lesions, dark lesions and reflection artefacts. The dual SVM classification framework that used a 4-d feature vector (chapter 3) achieves sensitivity of 100.00% and a specificity of 92.50% on a per image basis. For a per patch basis, it achieves a maximum accuracy of 92.35% that corresponds to a sensitivity of 87.93% and a specificity of 94.40%. The results were improved upon with the incorporation of a 21-d feature vector along

with the GA based feature selection approach (chapter 4). Results now stand at a sensitivity of 100.00% and a specificity of 97.50% on a per image basis. On a per patch basis, a maximum accuracy of 94.54% that corresponds to a sensitivity of 91.38% and a specificity of 96.00% was achieved. The performance on a per image basis is more useful from a clinical point of view.

A dual ensemble classification framework was created to investigate if the performance achieved by the dual SVM classification framework could be improved upon. The ensemble classifier was based on boot strapped decision trees, which has the ability to predict feature importance, and therefore facilitates an easy to implement feature selection procedure. The results were shown to be inferior to those achieved by dual SVM classification, achieving a sensitivity of 100.00% and a specificity of 95.00% on a per image basis. On a per patch basis, a maximum accuracy of 90.71% with a sensitivity of 81.03% and a specificity of 95.20% was achieved. Ensemble classification does show potential for use in the detection of new vessels; however, it lacked in overall performance as the associated feature selection procedure was not as thorough as the GA based feature selection approach.

The results of the proposed system (dual SVM classification and GA based feature selection, chapter 4) achieves better performance metrics than those reported by other published new vessel detection methods, besides Vatanparast [90] and Lee [91]. However, true comparisons are difficult to make as there exist no standard datasets for testing. Therefore, with each reported methodology creating its own dataset, it is difficult to know how challenging a task that dataset provides. The dataset used to evaluate our proposed system was created from a public and local source. It was created to be a challenging dataset, containing a large amount of other pathology in the non-new vessel images in order to demonstrate the capabilities of the proposed algorithm in avoiding false responses and it also contained a varied selection of new vessel images. Akram [77] used four publicly available retinal image databases to create a dataset containing 52 new vessel images. Such numbers of publicly available images gives the potential to create a standard database for evaluating PDR detection. Our intention was to evaluate our proposed system on this dataset to enable a comparison against the results of Akram

[77]. However, this was not possible as the image file names used are not available (not listed, and the authors were unresponsive over email) and we were unable to find the number of new vessel images that were reported. As the field of PDR detection advances, a publicly available PDR dataset will be invaluable as it could be used as a standard dataset which would help in the evaluation and direct comparison of different algorithms. The requirements of such a database are detailed in section 2.4.6.

Following on from the paragraph above, further reasons why true comparisons are difficult to make is that there exists variability in terms of the application of the methodologies. Our proposed system aims to detect both NVE and NVD, whereas Goatman [72], Arputham [73], Pavai [74], Nithyaa [86] and Agurto [88] aim only to detect NVD. Our proposed system is applied to conventional retinal images, whereas Jelinek [71] applied their method to fluorescein images, taking advantage of their higher contrast. The level at which performance evaluation is carried out also varies. Our proposed system has been evaluated on a patch and image level, whereas Hassan [70] has been evaluated on a pixel level and Akram [76] on a segment level. There remain a few reported methods [77,78,87,90,91] that are similar to our proposed system in terms of these aspects. The source codes relating to these algorithms have not been made public by the authors, and therefore we were unable to apply these methods to our dataset, which would have enabled true comparisons to be made. Alternatively, recreating these methodologies in order to be tested on our dataset was an option. However, the high level of complexity of these particular methodologies combined with the incomplete information regarding parameter settings would make this a substantial task with no guarantees of achieving performances that would accurately represent the reported methods. None of the reviewed articles on new vessel detection provide such comparisons. Therefore, like the rest of the field, we await the release of a publicly available new vessel dataset.

The dual classification framework requires the two independent classification outcomes to be combined in order to achieve a final decision. The approach taken is that both probability scores are separately thresholded and the sample is awarded a new vessel label only when both classifications agree on its identity being new

vessels. An alternative would be to combine the two probability scores into a single score and then threshold this in order to achieve the label of the sample. This approach wasn't taken as bright lesions and reflection artefacts may possess a low probability of being new vessels from classification 2, but they can receive a very high probability from classification 1; therefore, the latter would sway a combined score towards achieving an incorrect new vessel label.

The novel modified line operator plays a significant role in the proposed method. It is a vessel segmentation approach that is targeted at the reduction of false responses to non-vessel edges such as those caused by bright lesions. An alternative approach would be to remove bright lesions and reflection artefacts prior to the application of vessel segmentation. Niemeijer [20], in a method designed to detect red lesions, removes all bright lesions by removing all pixels above a specific value following shade correction. Aside from bright lesions and reflection artefacts, this will also remove fibrosis (fibrous tissue associated with new vessels) and the optic disc. Alternatively, the use of a specific exudate detection method [62-66,68] could be applied to target just the removal of exudates. However, the removal of image content in these alternatives could jeopardise the contrast and general appearance of vessels and therefore is avoided by vessel segmentation methods.

With a strong framework established any future developments would include further exploration into elements used by the framework as opposed to changing the actual framework. This includes further expansion of the feature vector and further comparative work into the integration of alternative classifiers. The GA based feature selection approach has shown its superiority in comparison to the feature selection approach associated with the ensemble classifier. However, it is possible that the results achieved by the GA based approach can be further enhanced. The GA based approach currently searches for a single set of SVM parameters for both classifications 1 and 2, when in fact the system performance could potentially be enhanced if the search is altered to find two sets of SVM parameters (one for each classification). Also the GA based approach, due to time constraints, is currently limited by the number of generations being set at 10. This should be ideally set at a value over 10,000, allowing more opportunity for the solution to evolve. This can be

achieved by increasing the computational efficiency of our proposed system, as well as performing evaluation on a larger dataset and thereby removing the need for leave-one-out cross validation.

A difficult dilemma is deciding whether evaluation on a per image or patch basis should lead future work. Requirements on each level differ significantly, to the extent that different feature subsets are preferred for each level. Evaluation on a per image basis may be currently more useful from a clinical point of view. However, this level of evaluation puts no emphasis on identifying all new vessel regions within the image and instead identifying any part of any new vessel region in the image is sufficient to achieve a positive image label. This strategy may be effective as it keeps specificity high, although it brings with it a certain level of risk. The ideal scenario would be the identification of all new vessel regions within the image. With this in mind, evaluation on a per patch basis should guide future work, as on a patch level the emphasis is to detect as many new vessel patches as possible. The optimum per patch performance currently stands at a sensitivity of 91.38% and a specificity of 96.00%; remember this does not correspond to the optimum per image performance. Striving to maintain this high sensitivity whilst getting the specificity to reach above 99.00% would be a suitable goal for future work. A per patch performance of this capability would mean the system would be capable of achieving a similar optimum per image performance to what it currently does (sensitivity = 100.00%, specificity = 97.50%), coupled with the possibility of almost all of the new vessel regions being identified within the images.

Our final goal is to develop an interactive retinal image analysis software tool focusing on the detection of PDR for use in screening programmes allowing for PDR images to be fast tracked to ophthalmology. This could be combined with a basic microaneurysm detection strategy allowing for the classification of DR disease/no DR disease. Our PDR detection system prevents bright lesions (exudates) from being mistakenly identified as new vessels. However, exudates can be significant in their own right, particularly as they can indicate possible clinically significant macular oedema. Therefore, exudates also require referral (less urgent), and for this reason

the potential software could also include exudate detection. Ideally the software would be a full DR disease detection system detecting all DR grades.

Prior to considering a software tool, the computational efficiency of the system would have to be improved. Currently the MATLAB code takes 527 seconds on average to process each image on an Intel(R) core(TM)2 Quad CPU Q9300 at 2.5 GHz. The MATLAB code was written with the intention to be used for research purposes only, and therefore there remains a large scope for improvements in computational efficiency.

The results of the proposed system indicate a good potential for clinical application, with a sensitivity of 100.00% and a specificity of 97.50% on a per image basis. Theoretically, these results should be maintained if the system is tested on a large population. However, such testing is essentially if the system is to be validated for clinical application. Even though the dataset used in this project was carefully selected to include a range of challenging cases, a large population may highlight various other overlooked challenging cases.

In conclusion, this thesis has demonstrated a PDR automated detection system that is capable of detecting the presence of new vessels whilst reducing false responses to bright lesions, dark lesions and reflection artefacts. PDR detection is a vital requirement if automated systems are to take on a substantial role in the grading pathway for diabetic retinopathy screening.

REFERENCES

[1] D.C. Klonoff and D.M. Schwartz, An economic analysis of interventions for diabetes, Diabetes Care 23 (2000) 390.

[2] P.M. Dodson, Oxford Diabetes Library: Diabetic Retinopathy, Oxford University Press (2008).

[3] T.A. Ciulla, A.G. Amador, B. Zinman, Diabetic retinopathy and diabetic macular edema: pathophysiology, screening, and novel therapies. Diabetes Care 26 (2003) 2653-2664.

[4] I. Kocur and S. Resnikoff, Visual impairment and blindness in europe and their prevention, British Journal of Ophthalmology 86 (2002) 716.

[5] D.S. Fong, L. Aiello, T.W. Gardner, G.L. King, G. Blankenship, J.D. Cavallerano, et al., Retinopathy in diabetes, Diabetes Care 27 Suppl 1 (2004) S84.

[6] H.R. Taylor and J.E. Keeffe, World blindness: a 21st century perspective, British Journal of Ophthalmology 85 (2001) 261-266.

[7] S. Wild, G. Roglic, A. Green, R. Sicree, H. King, Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 27 (2004) 1047-1053.

[8] T. Bek and M. Erlandsen, Visual prognosis after panretinal photocoagulation for proliferative diabetic retinopathy. Acta Ophthalmologica Scandinavica 84 (2006) 16.

[9] S.J. Lee, C. Sicari, C.A. Harper, P.M. Livingston, C.A. McCarty, H.R. Taylor, et al., Examination compliance and screening for diabetic retinopathy: a 2-year follow-up study. Clin Experiment Ophthalmol 28 (2000) 149-152.

[10] Photocoagulation for diabetic macular edema. Early Treatment Diabetic Retinopathy
Study report number 1. Early Treatment Diabetic Retinopathy Study research group. Arch Ophthalmol 103 (1985) 1796-806.

[11] J.P. De La Cruz, J.A. Gonzalez-Correa, A. Guerrero, Pharmacological approach to diabetic retinopathy, Diabetes/Metabolism Research and Reviews 20 (2004) 91-113.

[12] J.T. Gillow and J.A. Muir Gray, The National Screening Committee review of diabetic retinopathy screening, Eye 15 (2001) 1-2.

[13] J.M.G. Wilson and G. Junger, Principles and practice of screening for disease, Public Health Papers No. 34 (1968).

[14] U.K National Screening Committee, Essentials Elements in Developing a Diabetic Screening Programme. Workbook Version 4.3, (2009).

[15] R.N. Frank, Diabetic retinopathy, Progress in Retinal and Eye Research 14 (1995) 361-392.

[16] MESSIDOR [Online], http://messidor.crihan.fr.

[17] J.J. Kanski and B. Bowling, Clinical Ophthalmology: A Systematic Approach. Seventh edition. Elsevier Limited (2011).

[18] A.D. Fleming, K.A. Goatman, S. Philip, G.J. Prescott, P.F. Sharp, J.A. Olson, Automated grading for diabetic retinopathy: a large-scale audit using arbitration by clinical experts, British Journal of Ophthalmology 94 (2010) 1606-1610.

[19] T. Spencer, J.A. Olson, K.C. McHardy, P.F. Sharp, J.V. Forrester, An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus, Computers and Biomedical Research 29 (1996) 284-302.

[20] M. Niemeijer, B. van Ginneken, J. Staal, M.S.A. Suttorp-Schulten, M.D. Abramoff, Automatic detection of red lesions in digital color fundus photographs, IEEE Transactions on Medical Imaging 24 (2005) 584-592.

[21] S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, et al., Adaptive histogram equalization and its variations, Computer Vision, Graphics, and Image Processing 39 (1987) 355-368.

[22] G.S. Ramlugun, V.K. Nagarajan, C. Chakraborty, Small retinal vessels extraction towards proliferative diabetic retinopathy screening, Expert Systems with Applications 39 (2012) 1141-1146.

[23] A.A.A. Youssif, A.Z. Ghalwash, A.S. Ghoneim, A comparative evaluation of preprocessing methods for automatic detection of retinal anatomy, Proceedings of the Fifth International Conference on Informatics & Systems (2007) 24-30.

[24] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, M. Goldbaum, Detection of blood vessels in retinal images using two-dimensional matched filters, IEEE Transactions on Medical Imaging 8 (1989) 263-269.

[25] M. Al-Rawi, M. Qutaishat, M. Arrar, An improved matched filter for blood vessel detection of digital retinal images, Computers in Biology and Medicine 37 (2007) 262-267.

[26] A.D. Hoover, V. Kouznetsova, M. Goldbaum, Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response, IEEE Transactions on Medical Imaging 19 (2000) 203-210.

[27] L. Zhang, Q. Li, J. You, D. Zhang, A modified matched filter with double-sided thresholding for screening proliferative diabetic retinopathy, IEEE Transactions on Information Technology in Biomedicine 13 (2009) 528-534.

[28] B. Zhang, L. Zhang, L. Zhang, F. Karray, Retinal vessel extraction by matched filter with first-order derivative of Gaussian, Computers in Biology and Medicine 40 (2010) 438-445.

[29] A.M. Mendonca and A. Campilho, Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction, IEEE Transactions on Medical Imaging 25 (2006) 1200-1213.

[30] M.M. Fraz, S.A. Barman, P. Remagnino, A. Hoppe, A. Basit, B. Uyyanonvara, et al., An approach to localize the retinal blood vessels using bit planes and centerline detection, Computer Methods and Programs Biomedicine 108 (2011) 600-616.

[31] M.J. Cree, D. Cornforth, H.F. Jelinek, Vessel segmentation and tracking using a two-dimensional model, Proceedings of the Image and Vision Computing Conference (2005) 345-350.

[32] Y. Yin, M. Adel, S. Bourennane, Retinal vessel segmentation using a probabilistic tracking method, Pattern Recognition 45 (2012) 1235-1244.

[33] M. Vlachos and E. Dermatas, Multi-scale retinal vessel segmentation using line tracking, Computerized Medical Imaging and Graphics 34 (2010) 213-227.

[34] C. Sinthanayothin, J.F. Boyce, H.L. Cook, T.H. Williamson, Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images, British Journal of Ophthalmology 83 (1999) 902-910.

[35] S. Daultrey, Principal Components Analysis, Geo Abstracts Limited Norwich (1976).

[36] J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, B. van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Transactions on Medical Imaging 23 (2004) 501-509.

[37] J.V.B. Soares, J.J.G. Leandro, R.M. Cesar Jr, H.F. Jelinek, M.J. Cree, Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification, IEEE Transactions on Medical Imaging 25 (2006) 1214-1222.

[38] E. Ricci and R. Perfetti, Retinal blood vessel segmentation using line operators and support vector classification, IEEE Transactions on Medical Imaging 26 (2007) 1357-1365.

[39] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, et al., An ensemble classification based approach applied to retinal blood vessel segmentation, IEEE Transactions on Biomedical Engineering 59 (2012) 2538-2548.

[40] D. Marín, A. Aquino, M.E. Gegúndez-Arias, J.M. Bravo, A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features, IEEE Transactions on Medical Imaging 30 (2011) 146-158.

[41] Y.A. Tolias and S.M. Panas, A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering, IEEE Transactions on Medical Imaging 17 (1998) 263-273.

[42] G.B. Kande, P.V. Subbaiah, T.S. Savithri, Unsupervised fuzzy based vessel segmentation in pathological digital fundus images, Journal of Medical Systems 34 (2010) 849-858.

[43] M.K. Ikram, F.J. de Jong, J.R. Vingerling, J.C.M. Witteman, A. Hofman, M.M.B. Breteler, et al., Are retinal arteriolar or venular diameters associated with markers for cardiovascular disorders? The Rotterdam Study, Investigative Ophthalmology & Visual Science 45 (2004) 2129-2134.

[44] M. Saez, S. Gonzalez-Vazquez, M. Gonzalez-Penedo, M.A. Barcelo, M. Pena-Seijo, G.C. de Tuero, et al., Development of an automated system to classify retinal vessels into arteries and veins, Computer Methods and Programs in Biomedicine 108 (2012) 367-376.

[45] A. Bhuiyan, R. Kawasaki, E. Lamoureux, K. Ramamohanarao, T.Y. Wong, Retinal artery– vein caliber grading using color fundus imaging, Computer Methods and Programs in Biomedicine 111 (2013) 104-114.

[46] M.M. Fraz, P. Remagnino, A. Hoppe, A.R. Rudnicka, C.G. Owen, P.H. Whincup, et al., Quantification of blood vessel calibre in retinal images of multi-ethnic school children using a model based approach, Computerized Medical Imaging and Graphics 37 (2013) 48-60.

[47] W. Lotmar, A. Freiburghaus, D. Bracher, Measurement of vessel tortuosity on fundus photographs, Albrecht Von Graefes Arch Clin Exp Ophthalmol 211 (1979) 49-57.

[48] C.G. Owen, A.R. Rudnicka, R. Mullen, S.A. Barman, D. Monekosso, P.H. Whincup, et al., 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 (2009) 2004-2010.

[49] C. Sinthanayothin, P. Panitsuk, B. Uyyanonvara, Automatic retinal vessel tortuosity measurement, International Conference on Electrical Engineering/Electronics Computer Telecommunications and Information Technology (2010) 505-507.

[50] Y. Hatanaka, T. Hara, H. Fujita, M. Aoyama, H. Uchida, T. Yamamoto, Automated analysis of the distributions and geometries of blood vessels on retinal fundus images, Medical Imaging 2004 5370 (2004) 1621-1628.

[51] M.E. Martinez-Perez, A.D. Highes, A.V. Stanton, S.A. Thorn, N. Chapman, A.A. Bharath, et al., Retinal vascular tree morphology: a semi-automatic quantification, IEEE Transactions on Biomedical Engineering 49 (2002) 912-917.

[52] A. Perez-Rovira, T. MacGillivray, E. Trucco, K.S. Chin, K. Zutis, C. Lupascu, et al., VAMPIRE: Vessel assessment and measurement platform for images of the REtina, Conf Proc IEEE Eng Med Biol Soc 2011 (2011) 3391-3394.

[53] N. Mouravliansky, G.K. Matsopoulos, K. Delibasis, K.S. Nikita, Automatic retinal registration using global optimization techniques, Engineering in Medicine and Biology Society. Proceedings of the 20th Annual International Conference of the IEEE 20 (1998) 567-570.

[54] M.J. Cree, J.A. Olson, K.C. McHardy, P.F. Sharp, J.V. Forrester, A fully automated comparative microaneurysm digital detection system, Eye (London, England) 11 (Pt 5) (1997) 622-628.

[55] A.J. Frame, P.E. Undrill, M.J. Cree, J.A. Olson, K.C. McHardy, P.F. Sharp, et al., A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms, Computers in Biology and Medicine 28 (1998) 225-238.

[56] J.H. Hipwell, F. Strachan, J.A. Olson, K.C. McHardy, P.F. Sharp, J.V. Forrester, Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool, Diabetic Medicine 17 (2000) 588-594.

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

[58] C. Sinthanayothin, J.F. Boyce, T.H. Williamson, H.L. Cook, E. Mensah, S. Lal, et al., Automated detection of diabetic retinopathy on digital fundus images, Diabetic Medicine 19 (2002) 105-112.

[59] B. Zhang, X. Wu, J. You, Q. Li, F. Karray, Detection of microaneurysms using multi-scale correlation coefficients, Pattern Recognition 43 (2010) 2237-2248.

[60] A. Sopharak, B. Uyyanonvara, S.A. Barman, Automatic microaneurysm detection from non-dilated diabetic retinopathy retinal images using mathematical morphology methods, IAENG International Journal of Computer Science 38 (2011) 295.

[61] A. Sopharak, B. Uyyanonvara, S.A. Barman, Simple hybrid method for fine microaneurysm detection from non-dilated diabetic retinopathy retinal images, Computerized Medical Imaging and Graphics (2013).

[62] T. Walter, J. Klein, P. Massin, A. Erginay, A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina, IEEE Transactions on Medical Imaging 21 (2002) 1236-1243.

[63] A. Sopharak and B. Uyyanonvara, Automatic exudates detection on thai diabetic retinopathy patients' retinal images, Proceedings of the ECTI International Conference (2006) 709-712.

[64] A. Sopharak, B. Uyyanonvara, S.A. Barman, Automatic exudate detection from non-dilated diabetic retinopathy retinal images using fuzzy C-means clustering, Sensors 9 (2009) 2148-2161.

[65] A. Osareh, M. Mirmehdi, B. Thomas, R. Markham, Automated identification of diabetic retinal exudates in digital colour images, British Journal of Ophthalmology 87 (2003) 1220-1223.

[66] A. Sopharak, M.N. Dailey, B. Uyyanonvara, S.A. Barman, T. Williamson, K.T. Nwe, et al., Machine learning approach to automatic exudate detection in retinal images from diabetic patients, Journal of Modern Optics 57 (2010) 124-135.

[67] C. Duanggate, B. Uyyanonvara, S.S. Makhanov, S.A. Barman, T. Williamson, Drusen detection based on scale-space with feature stability, Medical Image Understanding and Analysis (2011) 317-321.

[68] M. Niemeijer, B. Van Ginneken, S.R. Russell, M.S.A. Suttorp-schulten, M.D. Abràmoff, Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis, Investigative Ophthalmology & Visual Science 48 (2007) 2260-2267.

[69] M.U. Akram, I. Jamal, A. Tariq, J. Imtiaz, Automated segmentation of blood vessels for detection of proliferative diabetic retinopathy, IEEE-EMBS International Conference on Biomedical and Health Informatics (2012) 232-235.

[70] S.S.A. Hassan, D.B.L. Bong, M. Premsenthil, Detection of neovascularization in diabetic retinopathy, Journal of Digital Imaging : the Official Journal of the Society for Computer Applications in Radiology (2011).

[71] H.F. Jelinek, M.J. Cree, J.J. Leandro, J.V. Soares, R.M. Cesar Jr, A. Luckie, Automated segmentation of retinal blood vessels and identification of proliferative diabetic retinopathy, Journal of the Optical Society of America. Optics, Image Science, and Vision 24 (2007) 1448-1456.

[72] K.A. Goatman, A.D. Fleming, S. Philip, G.J. Williams, J.A. Olson, P.F. Sharp, Detection of new vessels on the optic disc using retinal photographs, IEEE Transactions on Medical Imaging 30 (2011) 972.

[73] J.S. Arputham, G. Tamilpavai, S. Tamilselvi, An approach for the detection of proliferative diabetic retinopathy, IJCA Proceedings on International Conference in Recent trends in Computational Methods, Communication and Controls ICON3C (2012) 25-29.

[74] G.T. Pavai and S.T. Selvi, Identification of proliferative diabetic retinopathy using texture segmentation, Journal of Computer Science 9 (2013) 358-367.

[75] A.D. Fleming, K.A. Goatman, S. Philip, J.A. Olson, P.F. Sharp, Automatic detection of retinal anatomy to assist diabetic retinopathy screening, Physics in Medicine and Biology 52 (2007) 331-345.

[76] M.U. Akram, A. Tariq, S.A. Khan, Detection of neovascularization for screening of proliferative diabetic retinopathy, In Proceedings of the 9th international conference on Image Analysis and Recognition - Volume Part II (2012) 372-379.

[77] M.U. Akram, S. Khalid, A. Tariq, M.Y. Javed, Detection of neovascularization in retinal images using multivariate m-Mediods based classifier, Computerized Medical Imaging and Graphics 37 (2013) 346.

[78] K. Saranya, B. Ramasubramanian, S. Kaja Mohideen, A novel approach for the detection of new vessels in the retinal images for screening Diabetic Retinopathy, International Conference on Communications and Signal Processing (2012) 57-61.

[79] R.A. Welikala, V. Tah, T.H. Williamson, A. Hoppe, J. Dehmeshki, S.A. Barman, Differing matched filter responsivity for the detection of proliferative diabetic retinopathy, Proceedings of the IASTED International Conference Signal Processing, Pattern Recognition and Applications (2013) 356-364.

[80] M.A. Mainster, The fractal properties of retinal vessels: embryological and clinical implications, Eye 4 (1990) 235-241.

[81] A. Daxer, Characterisation of the neovascularisation process in diabetic retinopathy by means of fractal geometry: diagnostic implications, Graefe's Archive for Clinical and Experimental Ophthalmology 231 (1993) 681-686.

[82] A. Karperien, H.F. Jelinek, J.J.G. Leandro, J.V.B. Soares, R.M. Cesar Jr, A. Luckie, Automated detection of proliferative retinopathy in clinical practice, Clinical Ophthalmology 2008 (2008) 109.

[83] A.J. Frame, P.E. Undrill, J.A. Olson, K.C. McHardy, P.F. Sharp, J.V. Forrester, Texture analysis of retinal neovascularisation, IEE Colloquium on Pattern Recognition (Digest No. 1997/018) (1997) 5/1-5/6.

[84] R.M. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man, and Cybernetics 3 (1973) 610-621.

[85] R.M. Haralick, Statistical and structural approaches to texture, Proceedings of the IEEE 67 (1979) 786-804.

[86] S. Nithyaa and S. Karthikeyen, Identification of the diabetic retinopathy by detecting new retinal vessel using fundus image, International Journal of Advanced Research in Technology 2 (2012) 55-59.

[87] C. Agurto, V. Murray, E. Barriga, S. Murillo, M. Pattichis, H. Davis, et al., Multiscale AM-FM methods for diabetic retinopathy lesion detection, IEEE Transactions on Medical Imaging 29 (2010) 502-512.

[88] C. Agurto, Y. Honggang, V. Murray, M.S. Pattichis, S. Barriga, W. Bauman, et al., Detection of neovascularization in the optic disc using an AM-FM representation, granulometry, and vessel segmentation, Annual International Conference of the IEEE, Engineering in Medicine and Biology Society (EMBC) (2012) 4946-4949.

[89] D. Vallabha, R. Dorairaj, K. Namuduri, H. Thompson, Automated detection and classification of vascular abnormalities in diabetic retinopathy, Conference Record of the Thirty-Eighth Asilomar on Signals, Systems and Computers 2 (2004) 1625-1629.

[90] M. Vatanparast and A. Harati, A feasibility study on detection of Neovascularization in retinal color images using texture, 2nd International eConference on Computer and Knowledge Engineering (ICCKE) (2012) 221-226.

[91] J. Lee, B.C.Y. Zee, Q. Li, Detection of neovascularization based on fractal and texture analysis with interaction effects in diabetic retinopathy, PLoS One 8 (2013) e75699.

[92] U.R. Acharya, E.Y.K. Ng, J. Tan, S.S. Vinitha, K. Ng, An integrated index for the identification of diabetic retinopathy stages using texture parameters, Journal of Medical Systems 36 (2012) 2011-2020.

[93] M.M. Galloway, Texture analysis using gray level run lengths, Computer Graphics and Image Processing 4 (1975) 172-179.

[94] J. Anitha, C.K.S. Vijila, A.I. Selvakumar, A. Indumathy, D.J. Hemanth, Automated multi-level pathology identification techniques for abnormal retinal images using artificial neural networks, Br J Ophthalmol 96 (2012) 220-223.

[95] J. Nayak, P.S. Bhat, U.R. Acharya, C.M. Lim, M. Kagathi, Automated identification of diabetic retinopathy stages using digital fundus images, J Med Syst 32 (2008) 107-115.

[96] W.L. Yun, U.R. Acharya, Y.V. Venkatesh, C. Chee, L.C. Min, E.Y.K. Ng, Identification of different stages of diabetic retinopathy using retinal optical images, Information Sciences 178 (2008) 106-121.

[97] R. Priya and P. Aruna, Review of automated diagnosis of diabetic retinopathy using the support vector machine, International Journal of Applied Engineering Research 1 (2011) 844-863.

[98] I.N. McRitchie, P.M. Hart, R.J. Winder, Image registration and subtraction for the visualization of change in diabetic retinopathy screening, Computerized Medical Imaging and Graphics : the Official Journal of the Computerized Medical Imaging Society 30 (2006) 139-145.

[99] N. Amrutkar, Y. Bandgar, S. Chitalkar, S. Tade, An efficient approach for the detection of new vessels in PDR using image substraction and exudates in SDR using blobs detection, International Journal of Emerging Trends in Electrical and Electronics 3 (2013) 10-15.

[100] C.N. Doukas, I. Maglogiannis, A.A. Chatziioannou, Computer-supported angiogenesis quantification using image analysis and statistical averaging, IEEE Transactions on Information Technology in Biomedicine 12 (2008) 650-657.

[101] L.T. De La Cruz and N.S.T. Hirata, Preliminary studies on vascular network quantification in chick chorioallantoic membrane images, Workshop of Works in Progress (WIP) in SIBGRAPI 2013 (XXVI Conference on Graphics, Patterns and Images) (2013).

[102] K. Zutis, E. Trucco, J.P. Hubschman, D. Reed, S. Shah, J. van Hemert, Towards automatic detection of abnormal retinal capillaries in ultra-widefield-of-view retinal angiographic exams, International Conference on Engineering in Medicine and Biology (2013).

[103] ARIA Online; Retinal Image Archive [Online], http://www.eyecharity.com/aria_online.

[104] IMAGERET, optimal detection and decision-support diagnosis of diabetic retinopathy [Online], http://www2.it.lut.fi/project/imageret.

[105] Fundus Photograph Reading Center, University of Wisconsin-Madison [Online], http://eyephoto.ophth.wisc.edu/ResearchAreas/Diabetes/DiabStds.htm.

[106] T.M. Cover and P.E. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory IT-13 (1967) 21-27.

[107] A. Papoulis, Bayes' theorem in statistics and Bayes' theorem in statistics (reexamined), Probability, random variables, and stochastic processes. 2nd ed. New York, NY: McGraw-Hill (1984) 38-114.

[108] P. Domingos and M. Pazzani, On the optimality of the simple Bayesian classifier under zero-one loss, Machine Learning 29 (1997) 103-130.

[109] G.P. Zhang, Neural networks for classification: a survey, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 30 (2000) 451-462.

[110] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, Proceedings of the fifth annual workshop on Computational Learning Theory (1992) 144-152.

[111] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM computing surveys (CSUR) 31 (1999) 264-323.

[112] S.B. Kotsiantis, Supervised machine learning: A review of classification techniques, Informatica 31 (2007) 249-268.

[113] R.A. Welikala, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T.H. Williamson, et al., Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification, Computer Methods and Programs in Biomedicine 114 (2014) 247-261.

[114] R. Zwiggelaar, S.M. Astley, C.R.M. Boggis, C.J. Taylor, Linear structures in mammographic images: detection and classification, IEEE Transactions on Medical Imaging 23 (2004) 1077-1086.

[115] R. Polikar, Ensemble based systems in decision making, Circuits and Systems Magazine, IEEE 6 (2006) 21-45.

[116] Y.-. Chim, A.A. Kassim, Y. Ibrahim, Dual classifier system for handprinted alphanumeric character recognition, Pattern Analysis and Applications 1 (1998) 155-162.

[117] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and other Kernel-Based Learning Methods, Cambridge University Press (2000).

[118] A. Statnikov, C.F. Aliferis, D.P. Hardin, I. Guyon, A Gentle Introduction to Support Vector Machines in Biomedicine, World Scientific Publishing (2011).

[119] X. Zhang, G. Thibault, E. Decencière, G. Quellec, G. Cazuguel, A. Erginay, et al., Spatial normalization of eye fundus images, IEEE International Symposium on Biomedical Imaging (2012).

[120] M. Stone, Cross-validatory choice and assessment of statistical predictions, Journal of the Royal Statistical Society. Series B (Methodological) (1974) 111-147.

[121] T. Landgrebe and R. Duin, A simplified extension of the area under the ROC to the multiclass domain, Seventeenth Annual Symposium of the Pattern Recognition Association of South Africa (2006).

[122] M.S. Wandishin and S.J. Mullen, Multiclass ROC analysis, Weather and Forecasting 24 (2009) 530-547.

[123] D.L. Simon, A three-dimensional receiver operator characteristic surface diagnostic metric, Annual Conference of the Prognostics and Health Management Society (2010).

[124] N. Younis, D.M. Broadbent, S.P. Harding, J.P. Vora, Incidence of sight-threatening retinopathy in Type 1 diabetes in a systematic screening programme, Diabetic Medicine 20 (2003) 758-765.

[125] N. Younis, D.M. Broadbent, J.P. Vora, S.P. Harding, Incidence of sight-threatening retinopathy in patients with type 2 diabetes in the Liverpool Diabetic Eye Study: a cohort study, The Lancet 361 (2003) 195-200.

[126] R.A. Welikala, M.M. Fraz, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, et al., Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy, submitted to Computerized Medical Imaging and Graphics.

[127] I. Guyon and A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research 3 (2003) 1157-1182.

[128] Y. Saeys, I. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (2007) 2507-2517.

[129] H. Frohlich, O. Chapelle, B. Scholkopf, Feature selection for support vector machines by means of genetic algorithm, Proceedings. 15th IEEE International Conference on Tools with Artificial Intelligence (2003) 142-148.

[130] C.L. Huang and C.J. Wang, A GA-based feature selection and parameters optimization for support vector machines, Expert Systems with Applications 31 (2006) 231-240.

[131] W.W. Bledsoe, The use of biological concepts in the analytical study of systems, Paper presented at ORSA-TIMS National Meeting, San Francisco (1961).

[132] J.H. Holland, Adaptation in Natural and Artificial Systems, MIT Press (1992).

[133] S.N. Sivanandam and S.N. Deepa, Genetic Algorithm Optimization Problems, Springer (2008).

[134] A. Statnikov, D. Hardin, I. Guyon, C.F. Aliferis, A Gentle Introduction to Support Vector Machines in Biomedicine, World Scientific (2011).

[135] K.A. De Jong and W.M. Spears, An analysis of the interacting roles of population size and crossover in genetic algorithms, in parallel problem solving from nature, Anonymous, Springer (1991) 38-47.

[136] J.J. Grefenstette, Optimization of control parameters for genetic algorithms, Systems, Man and Cybernetics, IEEE Transactions on 16 (1986) 122-128.

[137] R.A. Welikala, M.M. Fraz, T.H. Williamson, S.A. Barman, The automated detection of proliferative diabetic retinopathy using dual ensemble classification, submitted to Computer Methods and Programs in Biomedicine.

[138] L. Breiman, Bagging predictors, Machine Learning 24 (1996) 123-140.

[139] R. Polikar, Bootstrap - Inspired Techniques in Computation Intelligence, Signal Processing Magazine, IEEE 24 (2007) 59-72.

[140] R.E. Schapire, The boosting approach to machine learning: An overview, Nonlinear Estimation and Classification (2003) 149-171.

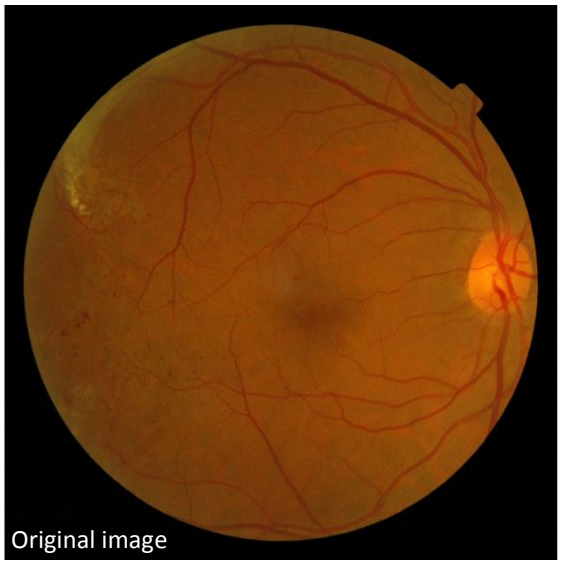
APPENDIX I: LINE OPERATOR PARAMETERS

This section provides a brief insight into how visual inspection was used to empirically derive the parameters of the line operator. The descriptions below correspond to the following pages which contain multiple line strength images (including a zoom-in new vessel region) for a single retinal image. Parameters are relative to a retinal image size of 1479 x 1479 pixels.

Line operator width: Chosen in accordance with vessel calibre. New vessels are predominantly small/medium in calibre, therefore an operator width of 25 pixels was chosen in accordance. An operator width below 25 pixels tends to enhance noise and incorrectly cause two separate strong responses for medium/large vessels. An operator width above 25 pixels tends to struggle when vessels are closely clustered.

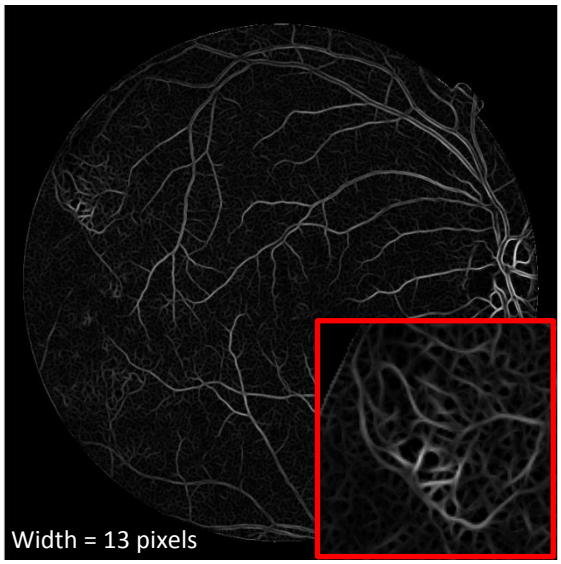
Line operator length: Chosen in accordance with vessel length. New vessels tend to be relatively short and tortuous in nature. An operator length of 15 pixels was chosen in accordance. An operator length below 15 pixels tends to be sensitive to noise. An operator length above 15 pixels tends to struggle with enhancing tortuous vessels.

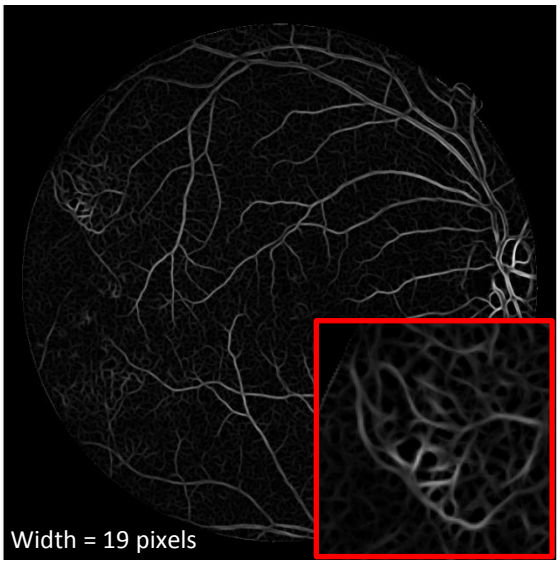
Line operator length (straight vessel removal): An operator length of 81 pixels was chosen in order to enhance only relatively straight vessels. An operator length below 81 pixels still retains some sensitivity to tortuous vessels. An operator length above 81 pixels has a reduction in sensitivity to straight vessels.

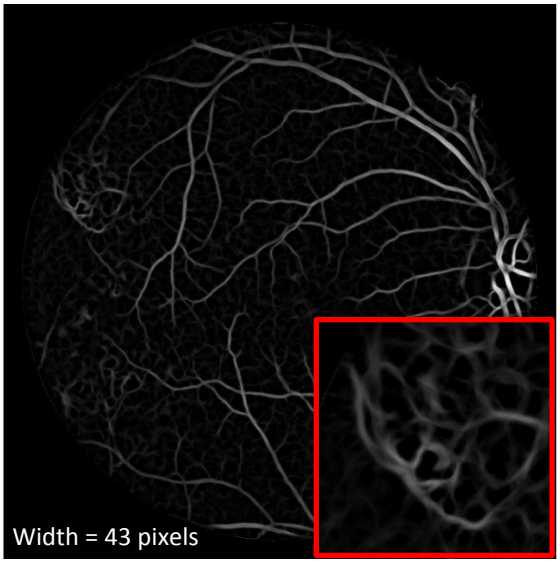


LINE OPERATOR WIDTH

Optimal width = 25 pixels. A pre-fixed length of 15 pixels is used for these examples.

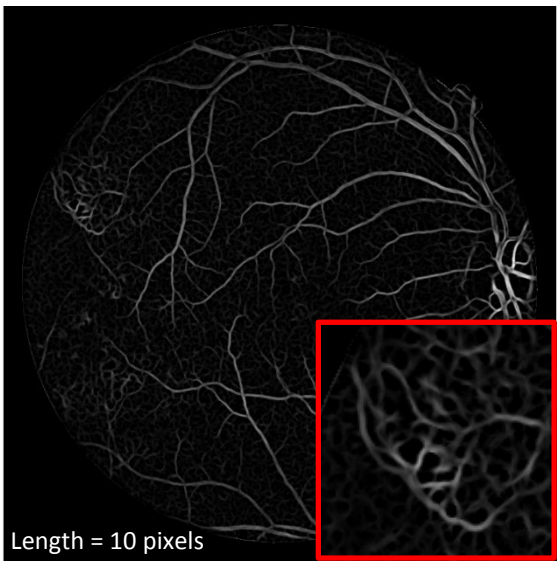


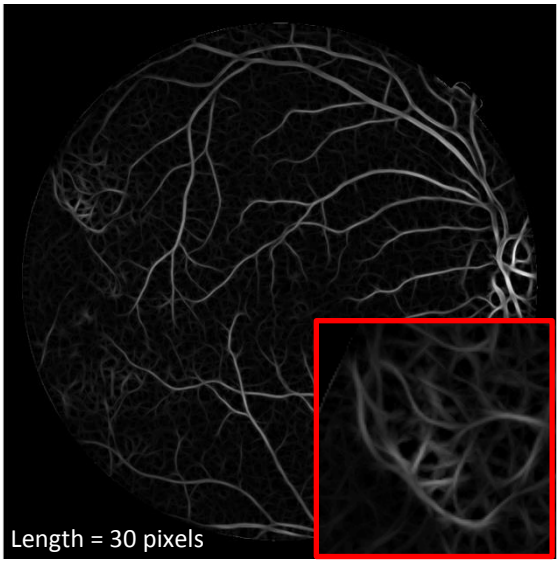




LINE OPERATOR LENGTH

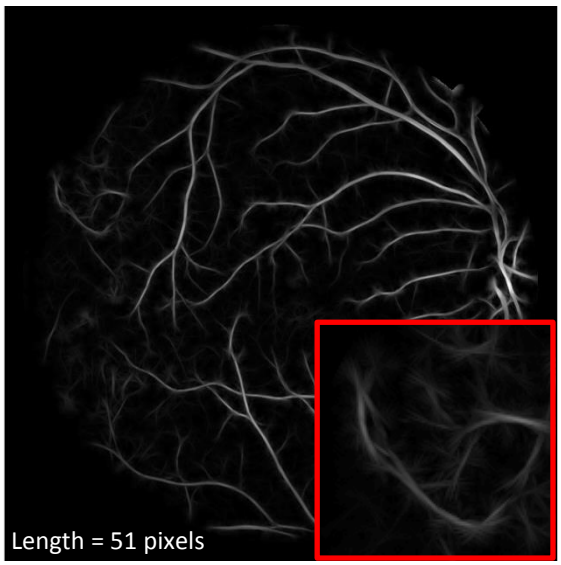
Optimal length = 15 pixels. A pre-fixed width of 25 pixels is used for these examples.

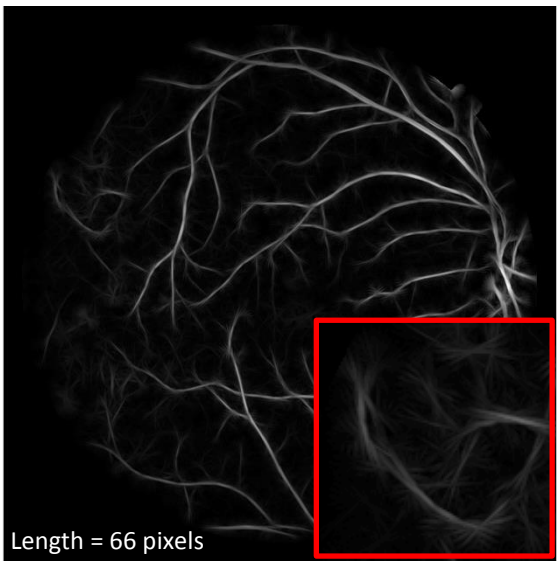


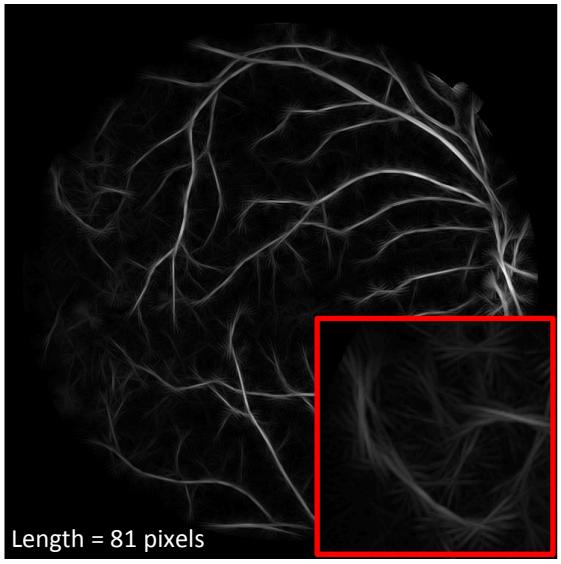


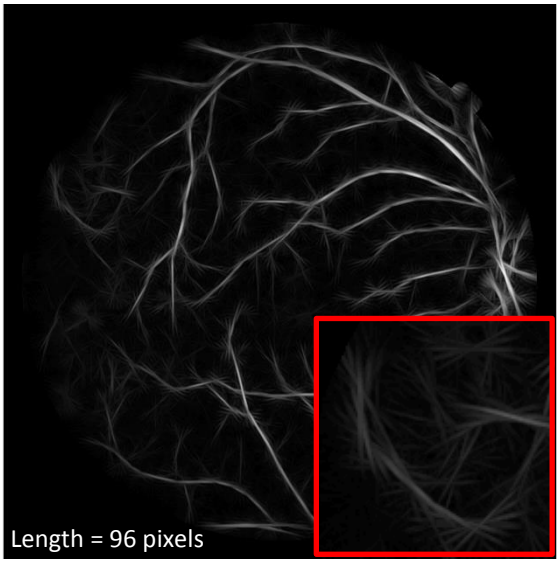
LINE OPERATOR LENGTH (STRAIGHT VESSEL REMOVAL)

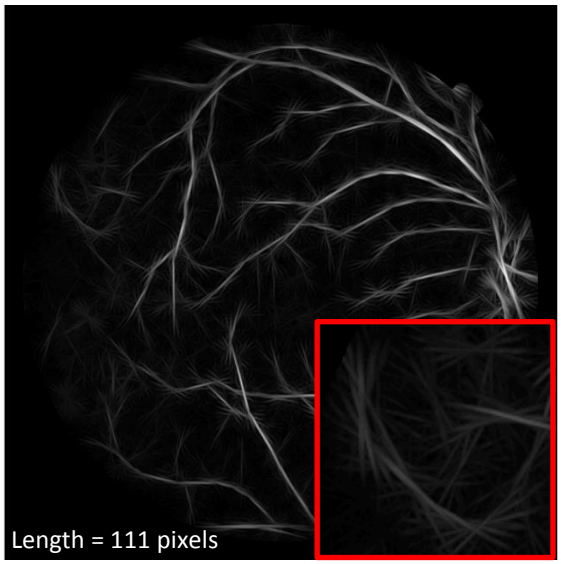
Optimal length = 81 pixels. A pre-fixed width of 25 pixels is used for these examples.

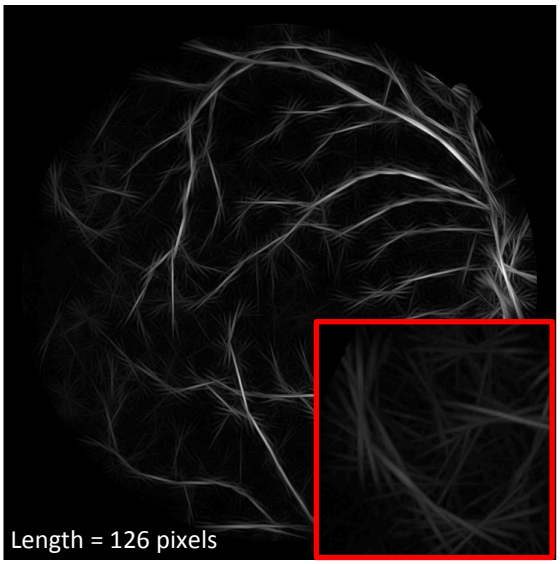












APPENDIX II: DATASET

Retinal image database

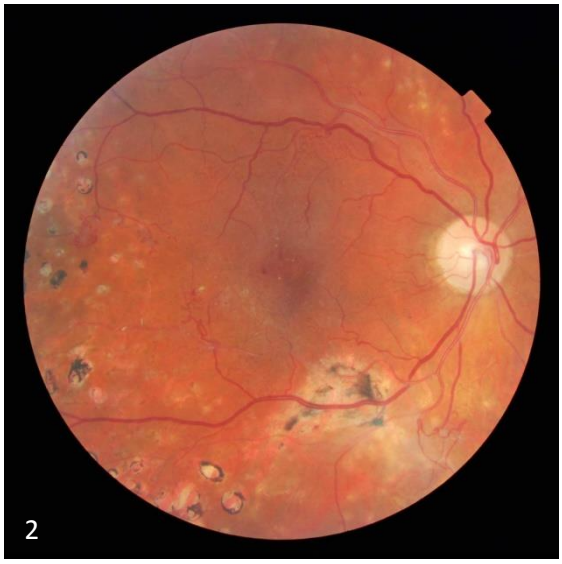
New vessel images, 1-20:

Images contain new vessels elsewhere (NVE) and new vessels at the optic disc (NVD).

Non new vessel images, 21-60:

21-40: Normal images.

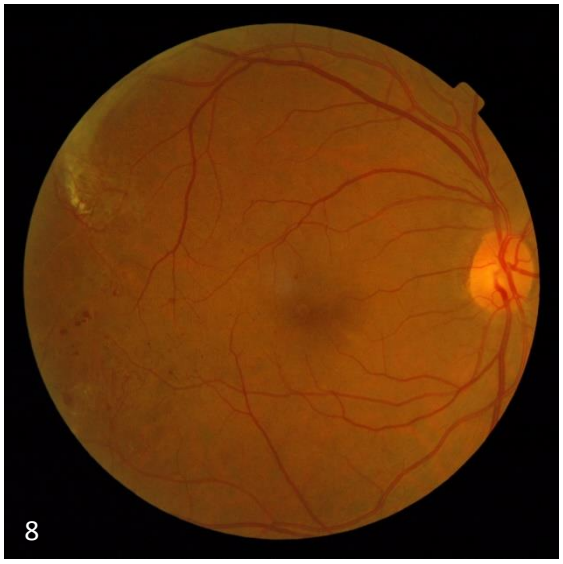
41-60: Images containing bright lesions, dark lesions and reflection artefacts.



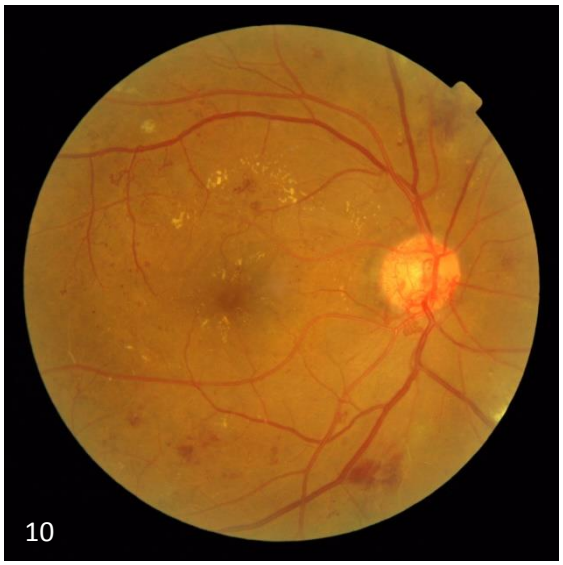

5

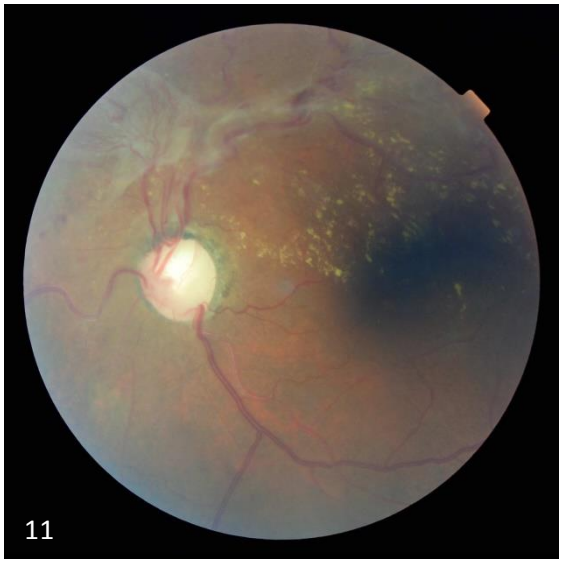

6


7


8


9


10


11


12


13


14

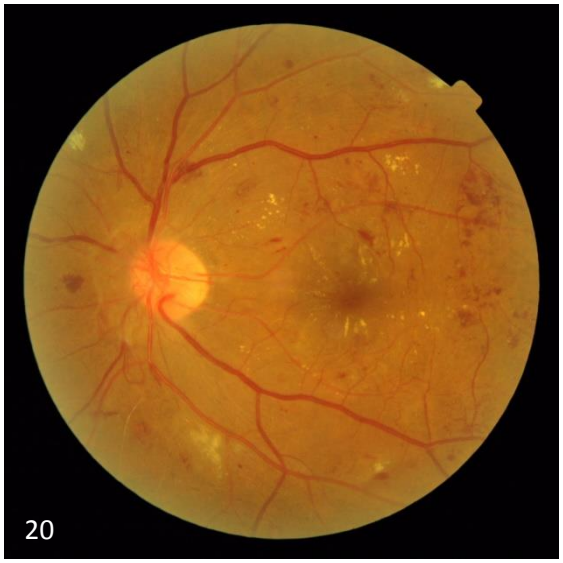

15


16


17


18


19


20

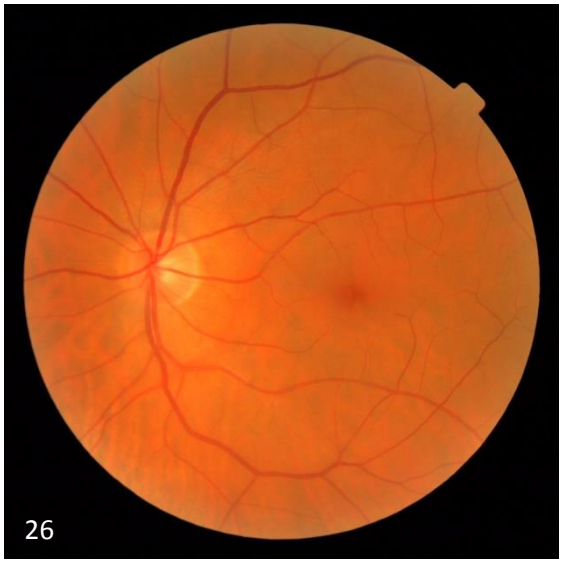

21


22

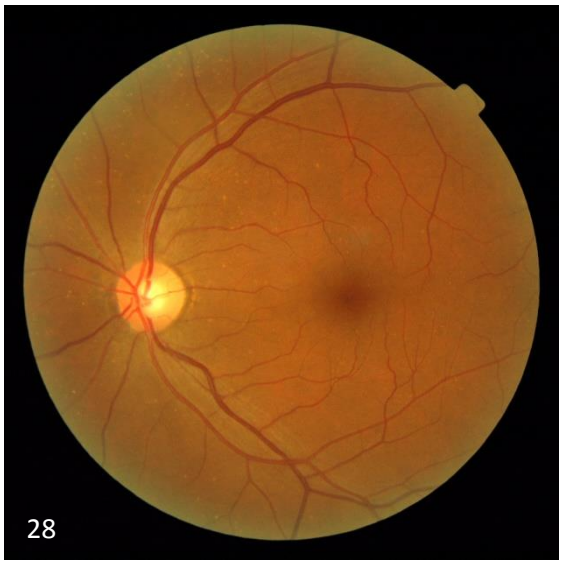

23


24


25


26

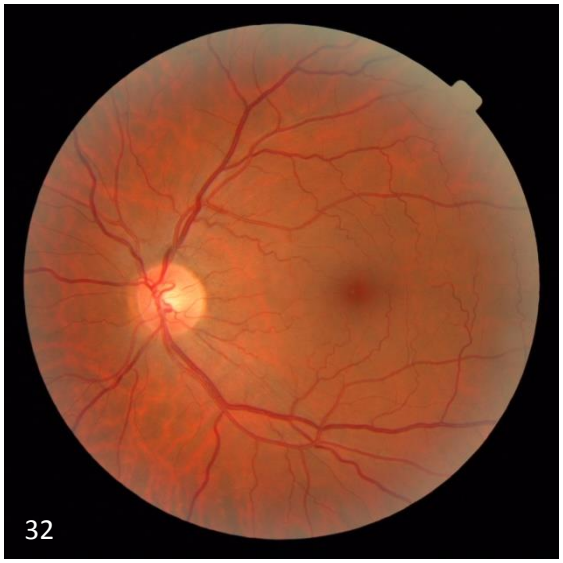

27


28


29


30


31


32

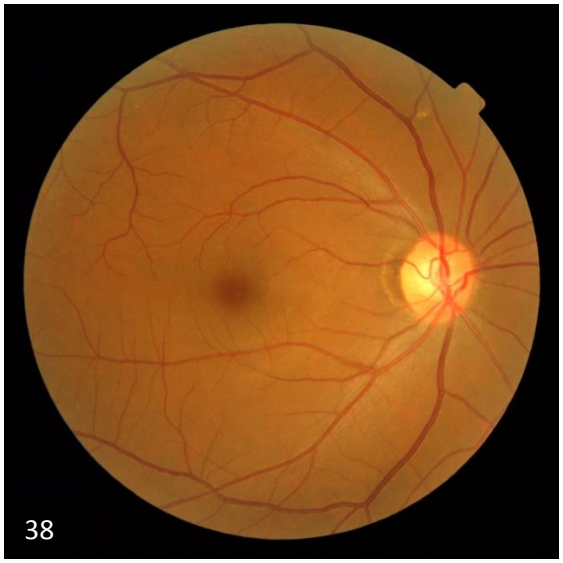

33


34

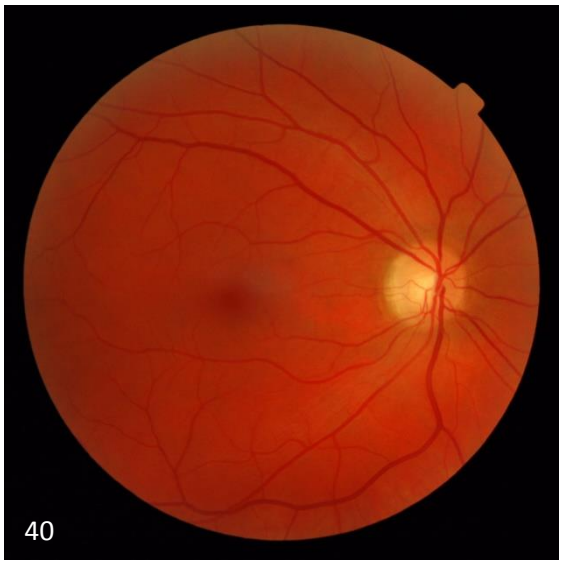

35


36


37


38


39


40


41


42


43


44

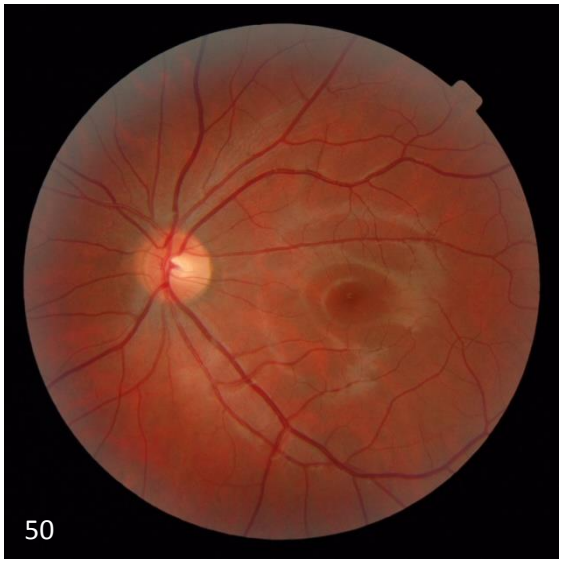

45


46

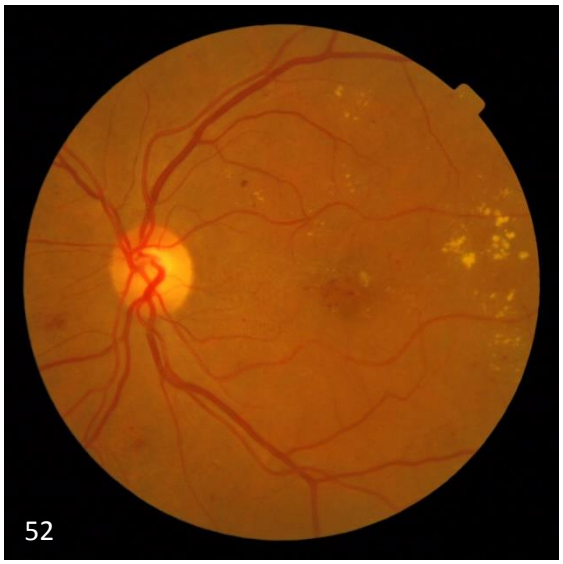

47


48


49


50


51


52


53


54


55


56


57

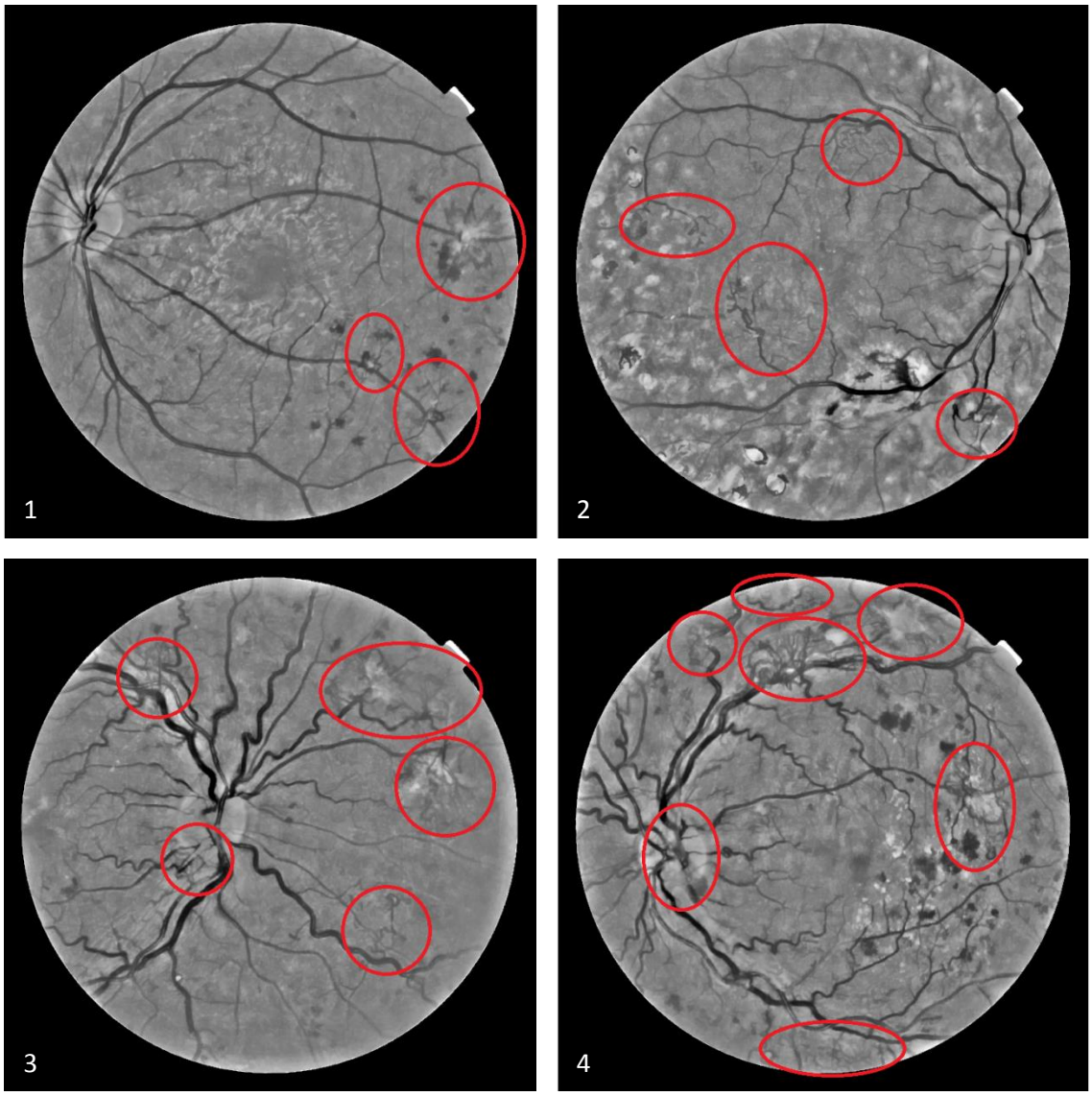

58


59

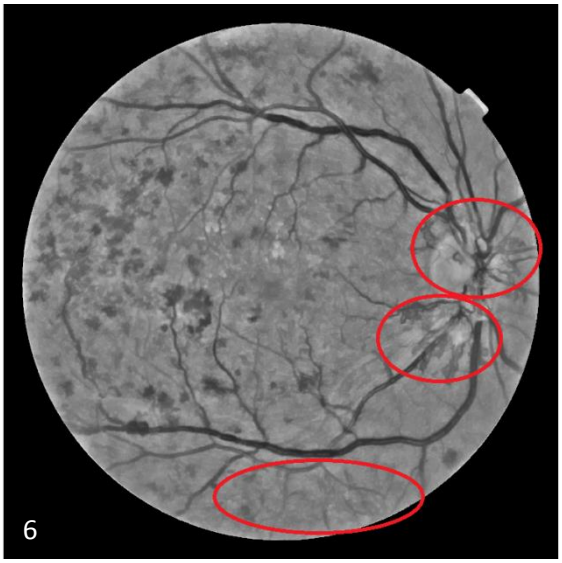

60

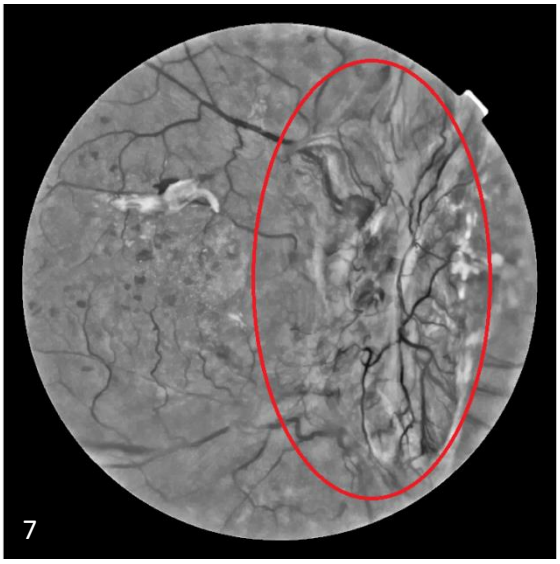
APPENDIX III: MANUAL DELINEATION

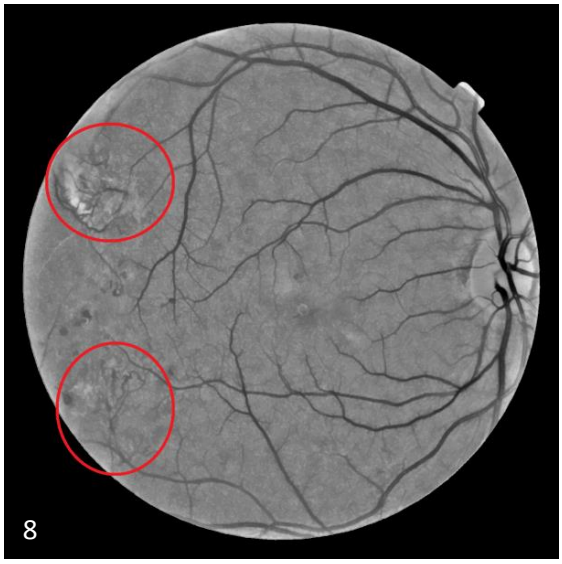
Location and delineation of new vessel regions within the 20 new vessel images; marked by an ophthalmologist. Shown on pre-processed images for better visualization.

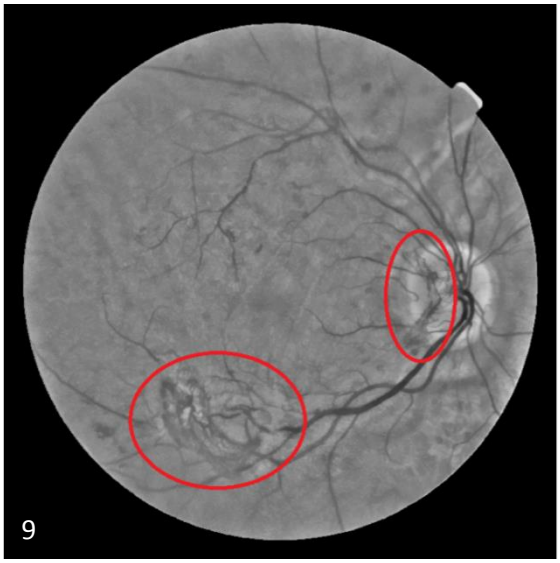


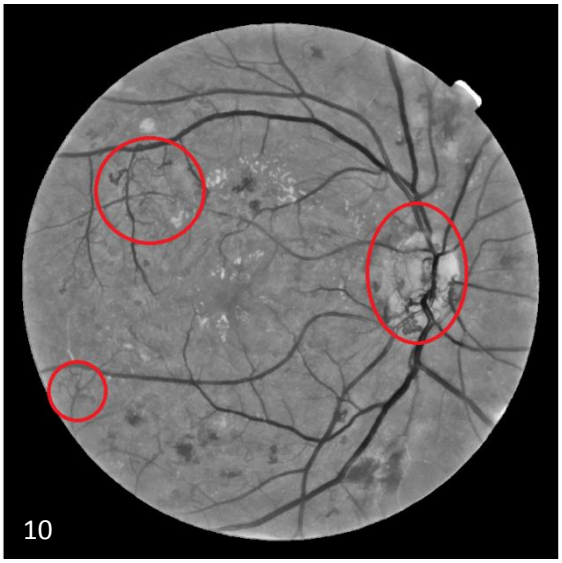

5

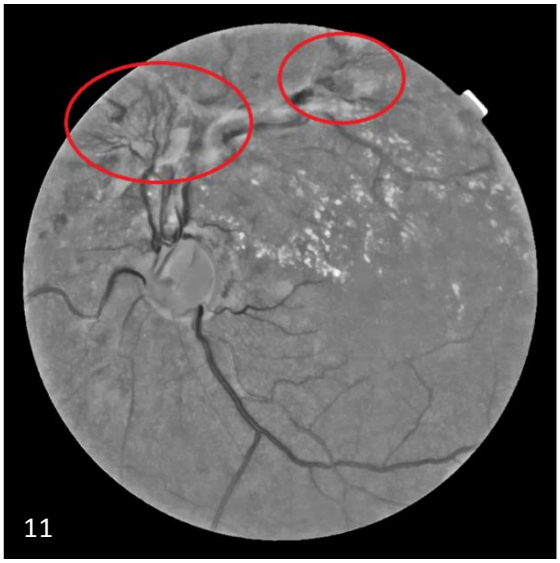

6

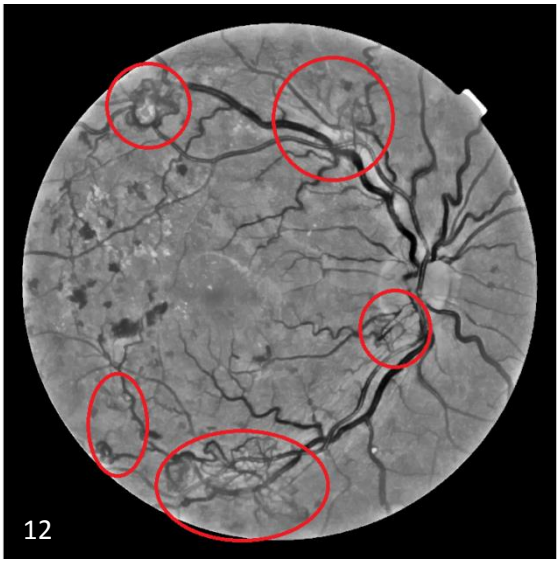

7


8

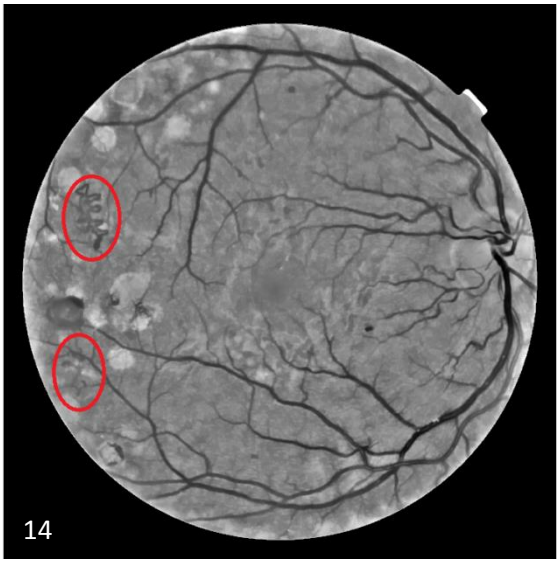

9

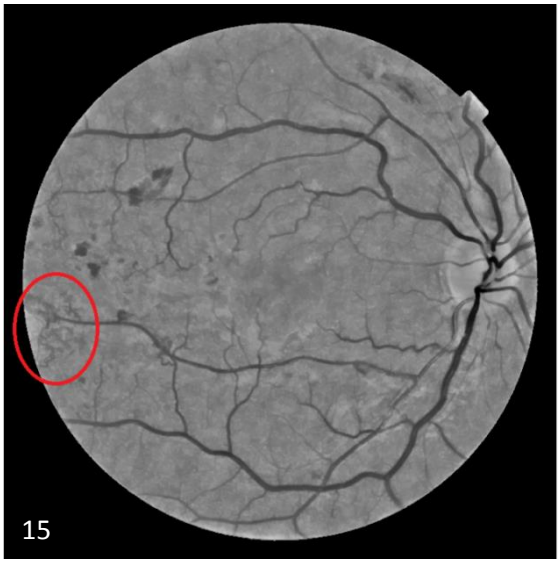

10

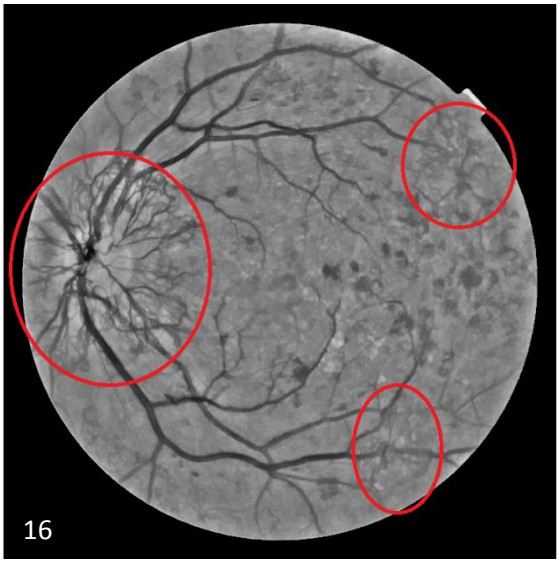

11

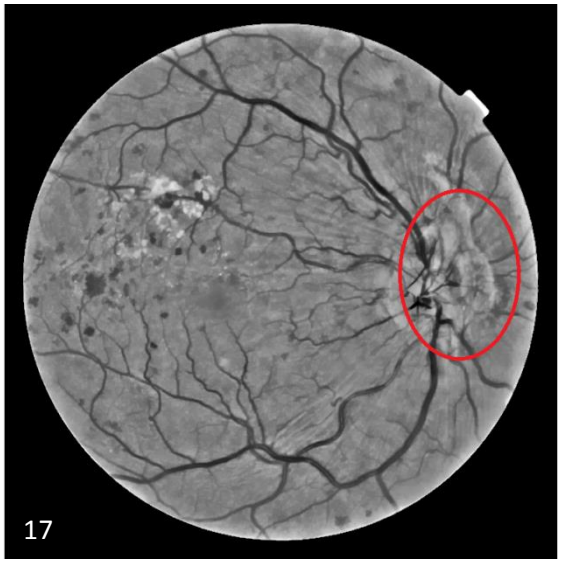

12

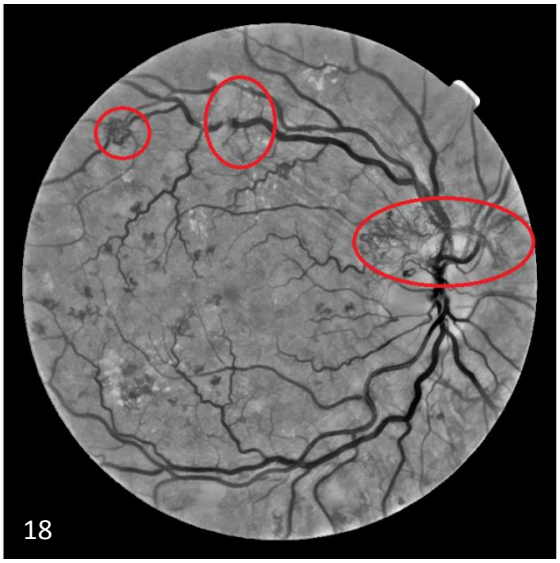

13

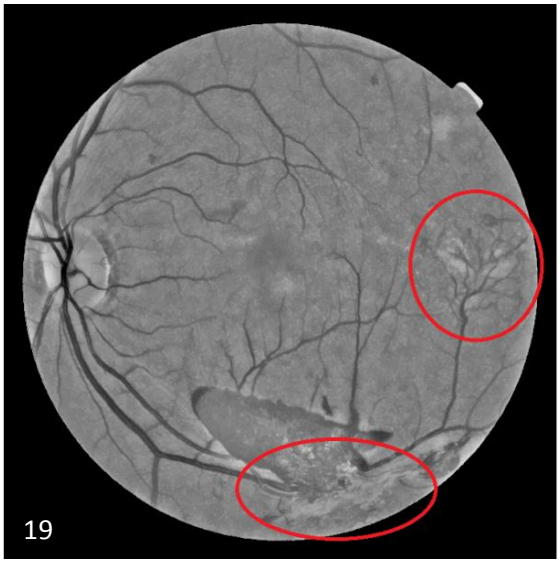

14

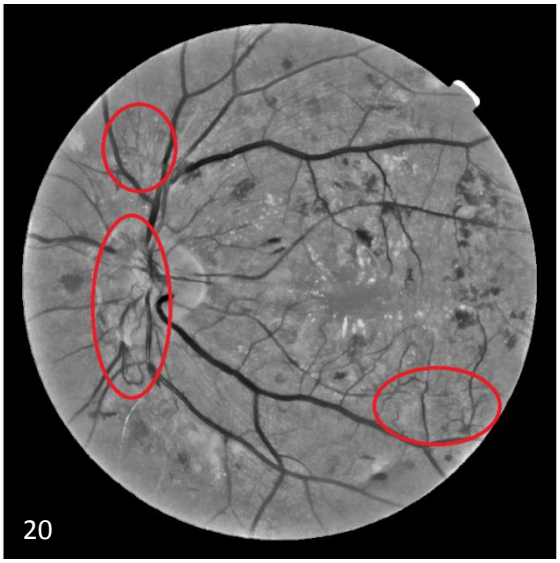

15


16


17


18


19


20