

---

# **Incorporating Spatial and Temporal Information for Microaneurysm Detection in Retinal Images**

---

A dissertation submitted by

**Mohamed Mustafa Sayed Ahmed Habib**

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  
In collaboration with St. George's university in London

Director of studies: Professor Sarah Barman

August 2017

## **ACKNOWLEDGEMENTS**

I would like to extend my gratitude to, first of all, my parents and siblings for their continuous support and encouragement and to my supervisor, Professor Sarah A. Barman, for her patient guidance and valuable support throughout my study. She has always been available for guidance and advice throughout my PhD period. I am indeed indebted to her for inspiring and encouraging me during this challenging task.

I appreciate the help of Dr Andreas Hoppe and other faculty members who have provided continued guidance throughout the program. I am grateful to Professor Christopher Owen and Dr Alicja Rudnicka from St. George's University of London for providing their support and in helping us find and supplying us with suitable datasets without which this work could not have been completed. Furthermore, their clinical feedback on the work has been invaluable throughout the project. My colleague Dr Roshan Welikala has shared a lot of his experience with me and provided support and ideas during the project, and for that I am very grateful.

This research was funded by a research fellowship of the Faculty of Science, Engineering and Computing at Kingston University, London, UK. I am also grateful to the research administration office for providing me various opportunities for training and workshops during the program.

I am also thankful to the people whom I did not mention, but have a valuable contribution in this research.

Mohamed Habib  
August 2017

## **ABSTRACT**

The retina of the human eye has the potential to reveal crucial information about several diseases such as diabetes. Several signs such as microaneurysms (MA) manifest themselves as early indicators of Diabetic Retinopathy (DR). Detection of these early signs is important from a clinical perspective in order to suggest appropriate treatment for DR patients. This work aims to improve the detection accuracy of MAs in colour fundus images. While it is expected that multiple images per eye are available in a clinical setup, proposed segmentation algorithms in the literature do not make use of these multiple images.

This work introduces a novel MA detection algorithm and a framework for combining spatial and temporal images. A new MA detection method has been proposed which uses a Gaussian matched filter and an ensemble classifier with 70 features for the detection of candidates. The proposed method was evaluated on three public datasets (171 images in total) and has shown improvement in performance for two of the sets when compared to a state-of-the-art method. For lesion-based performance, the proposed method has achieved Retinopathy Online Challenge (ROC) scores of 0.3923, 2109 and 0.1523 in the MESSIDOR, DIARETDB1 and ROC datasets respectively.

Based on the ensemble algorithm, a framework for the information combination is developed and consists of image alignment, detecting candidates with likelihood scores, matching candidates from aligned images, and finally fusing the scores from the aligned image pairs. This framework is used to combine information both spatially and temporally. A dataset of 320 images that consists of both spatial and temporal pairs was used for the evaluation. An improvement of performance by 2% is shown after combining spatial

information. The framework is applied to temporal image pairs and the results of combining temporal information are analyzed and discussed.

## **PUBLICATIONS FROM THIS WORK**

### **Journals**

1. Habib, M.M., Welikala, R.A., Hoppe, A., Owen, C.G., Rudnicka, A.R. and Barman, S.A., 2017. Detection of microaneurysms in retinal images using an ensemble classifier. *Informatics in Medicine Unlocked* Volume 9, 2017, Pages 44-57.
2. Habib, M., Welikala, R., Hoppe, A., Owen, C.G., Rudnicka, A.R., Tufail, A., Egan, C. and Barman, S.A., 2017. Incorporating spatial information for microaneurysm detection in retinal images. *Advances in Science, Technology and Engineering Systems Journal*, 2(3)Volume 2, Issue 3, Page No 642-649, 2017.

### **Conferences**

1. Habib, M.M., Welikala, R.A., Hoppe, A., Owen, C.G., Rudnicka, A.R. and Barman, S.A., 2016, December. Microaneurysm detection in retinal images using an ensemble classifier. In *Image Processing Theory Tools and Applications (IPTA), 2016 6th International Conference on*. IEEE 12-15 Dec. 2016, Oulu, Finland.
2. SEC Student Conference poster presentation, "Microaneurysm Detection in Retinal Images Using an Ensemble Classifier and Spatial Information", 2017, Kingston University, London, United Kingdom.

### **Presentations**

1. M.M. Habib, A. Hoppe, A. R. Rudnicka, C. G. Owen, S. A. Barman, Automated Evaluation of Retinal Image Registration Algorithms, Interdisciplinary Hub for the Study of Health and Age-related conditions, 2014, Kingston University, London, United Kingdom.
2. M.M. Habib, R.A. Welikala, A. Hoppe, C.G. Owen, A.R. Rudnicka and S.A. Barman, Microaneurysm Detection in retinal Images using an Ensemble Classifier and Spatial Registration, Science, Engineering and Computing student conference, 2017, Kingston University, London, United Kingdom.
3. M.M. Habib, Retinal Image Analysis: Preventing Blindness, 3 minute thesis (3MT) presentation, 2017, Kingston University, London, United Kingdom (second place winner).



# CONTENTS

<b>CHAPTER 1 INTRODUCTION .....</b>	<b>1</b>
1.1 THE ANATOMY OF THE EYE.....	2
1.2 RETINAL IMAGING .....	4
1.3 DIABETIC RETINOPATHY.....	9
1.4 AIMS, OBJECTIVES AND CONTRIBUTION TO KNOWLEDGE .....	16
<b>CHAPTER 2 LITERATURE REVIEW.....</b>	<b>20</b>
2.1 RETINAL IMAGE DATASETS.....	20
2.1.1 DIARETDB .....	21
2.1.2 MESSIDOR .....	21
2.1.3 ROC Dataset .....	22
2.1.4 CHASE.....	23
2.1.5 STARE .....	23
2.1.6 DRIVE.....	23
2.1.7 ARIA.....	24
2.1.8 Hussein .....	24
2.2 MICROANEURYSM DETECTION.....	26
2.3 IMAGE REGISTRATION .....	36
2.4 MULTI-IMAGE ANALYSIS ALGORITHMS .....	44
2.4.1 Longitudinal Change Analysis.....	44
2.4.2 Retinal Atlases.....	45
2.4.3 Other Multi-image Algorithms .....	46
2.5 CONCLUSION.....	47
<b>CHAPTER 3 MICROANEURYSM DETECTION USING AN ENSEMBLE CLASSIFIER .....</b>	<b>48</b>
3.1 METHODOLOGY .....	49
3.1.1 Preprocessing .....	50
3.1.2 MA Candidate Detection .....	51

3.1.3 MA Candidate Classification .....	61
3.2 EXPERIMENTAL EVALUATION .....	67
3.2.1 Datasets .....	67
3.2.2 Evaluation .....	70
3.2.3 Feature Analysis .....	89
3.2.4 Algorithm Performance .....	93
3.3 CONCLUSION .....	98
<b>CHAPTER 4 MICROANEURYSM DETECTION USING SPATIAL INFORMATION .....</b>	<b>102</b>
4.1 METHODOLOGY .....	104
4.1.1 Image registration .....	106
4.1.2 Candidate Matching .....	107
4.1.3 Scores fusion .....	111
4.2 EXPERIMENTAL EVALUATION .....	112
4.2.1 Dataset .....	112
4.2.2 Image Registration .....	115
4.2.3 Quantitative Results .....	126
4.3 CONCLUSION .....	131
<b>CHAPTER 5 MICROANEURYSM DETECTION USING TEMPORAL INFORMATION .....</b>	<b>134</b>
5.1 METHODOLOGY .....	135
5.1.1 Image Registration .....	136
5.1.2 Candidate Matching .....	137
5.2 EXPERIMENTAL EVALUATION .....	144
5.2.1 Dataset .....	144
5.2.2 Image Registration .....	145
5.2.3 Quantitative Results .....	149
5.2.4 Combining Feature Vectors .....	153
5.3 CONCLUSION .....	155

<b>CHAPTER 6 SUMMARY AND CONCLUSIONS .....</b>	<b>157</b>
6.1 SUMMARY OF RESULTS.....	157
6.1.1 <i>Microaneurysm Detection Using an Ensemble Classifier</i> .....	158
6.1.2 <i>A Framework for Combining Information from Retinal Images</i> .....	159
6.2 CONCLUSION.....	162
6.3 FUTURE WORK .....	163
6.3.1 <i>Spatial Dimension</i> .....	164
6.3.2 <i>Temporal Dimension</i> .....	164
6.3.3 <i>Other Avenues</i> .....	165

# List of Figures

FIGURE 1.1 A) ANATOMY OF THE EYE. IMAGE FROM WWW.BIOGRAPHIXMEDIA.COM. B) EXAMPLES OF HEALTHY RETINAL IMAGES. LABELS OF THE MAIN RETINAL FEATURES ARE HIGHLIGHTED. IMAGES FROM [3]..... 4

FIGURE 1.2(A) MICROANEURYSMS, (B) DOT AND BLOT HAEMORRHAGES, (C) EXUDATES, (D) A RING OF EXUDATES (CIRCINATE). IMAGES FROM[44]..... 13

FIGURE 1.3 A) MILD NEW VESSELS ELSEWHERE (NVE), B) SEVERE NEW VESSELS ELSEWHERE (NVE), C) SEVERE NVE WITH ASSOCIATED FIBROSIS, D) MILD NVD. IMAGES FROM [44]. ..... 14

FIGURE 1.4 A COLOUR RETINAL IMAGE WITH MICROANEURYSMS OF VARIOUS CONTRASTS HIGHLIGHTED IMAGE FROM MESSIDOR. 15

FIGURE 3.1 AN EXAMPLE OF THE PREPROCESSING STAGE. A) THE COLOUR IMAGE, B) THE GREEN CHANNEL IMAGE, C) THE PREPROCESSED IMAGE..... 51

FIGURE 3.2A MICROANEURYSM’S INTENSITY PROFILE RESEMBLES THAT OF A 2D GAUSSIAN FUNCTION. A) A MICROANEURYSM COLOUR IMAGE, B) THE GREEN CHANNEL OF (A). C) THE INTENSITY PROFILE OF (B), D) A GAUSSIAN INTENSITY PROFILE ( $\sigma = 1$ ) ..... 52

FIGURE 3.3AN EXAMPLE OF THE CANDIDATE DETECTION PHASE A) THE PREPROCESSED IMAGE, B) THE RESULT OF THE BOTTOMHAT OPERATION, C) THE GAUSSIAN FILTER RESPONSE D) THE THRESHOLDED IMAGE E) THE RESULT OF THE REGION GROWING OPERATION. THE HIGHLIGHTED REGION IS A TRUE MICROANEURYSM..... 54

FIGURE 3.4 (A) AVERAGE SENSITIVITY PLOTTED AGAINST AVERAGE FALSE POSITIVES PER PIXEL FOR BOTH GAUSSIAN FILTER AND EXTENDED MINIMA TRANSFORM.(B) THE SAME GRAPH IN(A) BUT THE HORIZONTAL AXIS IN LOGARITHM SCALE. .... 61

FIGURE 3.5 OUT OF BAG (OOB) CLASSIFICATION ERROR AS A FUNCTION OF THE NUMBER OF TREES IN THE TREE ENSEMBLE CLASSIFIER. .... 71

FIGURE 3.6 EXAMPLES OF MICROANEURYSM DETECTION ALGORITHMS APPLIED TO THE MESSIDOR DATASET. THE FIRST COLUMN SHOWS THE COLOUR IMAGE PATCH. COLUMNS 2-4 SHOW THE PREPROCESSED IMAGE WITH THE ALGORITHM RESULTS HIGHLIGHTED. A BLUE CIRCLE REPRESENTS THE GROUNDTRUTH LABELLED MICROANEURYSM. A GREEN CIRCLE REPRESENTS AN MA CANDIDATE DETECTED BY THE ALGORITHM. A RED CIRCLE REPRESENTS ANMA CANDIDATE THAT WAS DETECTED AS A

CANDIDATE MA BUT CLASSIFIED AS A FALSE POSITIVE BY THE CLASSIFIER. THE YELLOW BOUNDARIES SHOWS THE VESSEL REGIONS DETECTED BY THE QUARTZ SOFTWARE [144].	73
FIGURE 3.7 AN EXAMPLE OF A CANDIDATE (DOTTED CIRCLE) THAT IS CONSIDERED TO MATCH A GROUNDTRUTH (SOLID CIRCLE). THERE IS A MATCH SINCE THE CENTRE OF THE CANDIDATE LIES WITHIN THE GROUNDTRUTH REGION.	75
FIGURE 3.8 AN EXAMPLE OF VARIABILITY IN RESULTS EVERYTIME A TREE ENSEMBLE MODEL IS BUILT FOR CLASSIFICATION. IN THIS EXAMPLE THE TREE ENSEMBLE CLASSIFIER WAS RUN 11 TIMES AND THE MAXIMUM, MINIMUM AND MEDIAN RESULTS ARE DISPLAYED.	76
FIGURE 3.9 FREE-RECEIVER OPERATING CURVE (FROC) FOR ALL MICROANEURYSM CANDIDATES IN THE TEST SET. IN F) THE BLACK LINES REPRESENT THE PERFORMANCE OF EACH RESPECTIVE MODEL ON THE DIARETDB1 TEST SET BY TRAINING ON THE DIARETDB1 TRAINING SET, WHILE THE BLUE LINES REPRESENT THE PERFORMANCE OF EACH MODEL ON THE DIARETDB1 TEST SET AFTER TRAINING ON THE MESSIDOR TRAINING SET.	77
FIGURE 3.10. EXAMPLES OF CANDIDATES FALSELY DETECTED ON VESSELS. RED CIRCLES REPRESENT INTIAL CANDIDATE MAs DETECTED BY THE PROPOSED METHOD.	82
FIGURE 3.11 ROC CURVES FOR PER-PIXEL EVALUATION OF THE PROPOSED ALGORITHM COMPARED WITH 2 OTHER METHODS. THE EVALUATIONS WERE PERFORMED ON MESSIDOR, DIARETDB1 AND ROC DATASETS.	85
FIGURE 3.12 EXAMPLES OF REGION GROWING STAGE APPLIED TO VARIOUS CANDIDATES IN THE ROC DATASET	89
FIGURE 3.13 FEATURE IMPORTANCE MEASURED USING THE MESSIDOR DATASET. X-AXIS REPRESENTS THE FEATURE NUMBER. Y-AXIS REPRESENTS THE FEATURE IMPORTANCE MEASURE.	90
FIGURE 4.1. A) A CONCEPTUAL DIAGRAM OF THE TWO SPATIAL VIEWS OF THE RETINA (OPTIC DISC AND FOVEA CENTERED). THE PATCHES IN B) AND C) DEMONSTRATE HOW THE SAME MICROANEURYSM IN DIFFERENT VIEWS OF THE RETINA CAN HAVE A VARYING LEVEL OF CONTRAST.	103
FIGURE 4.2. AN OVERVIEW OF THE PROPOSED METHODOLOGY.	106
FIGURE 4.3. THE MANUAL CONTROL POINT SELECTION PROCESS FOR THE IMAGE REGISTRATION PHASE.	107
FIGURE 4.4. AN ILLUSTRATION OF THE TOLERANCE ADDED WHILE MATCHING 2 CANDIDATES. THE DASHED CIRCLES REPRESENT TWO CANDIDATES FROM TWO VIEWS OF THE RETINA. IT IS ASSUMED THAT THEY ARE MISALIGNED DUE TO REGISTRATION	

INACCURACY. THE SOLID CIRCLE REPRESENTS A TOLERANCE REGION AROUND THE FIRST CANDIDATE AND USING THIS METHOD A MATCH IS MADE. ....	109
FIGURE 4.5. AN EXAMPLE OF THE CANDIDATE DETECTION RESULT. IN EACH COL: A) A COLOUR IMAGE FROM AN OPTIC DISC CENTERED IMAGE. B) THE FOVEA CENTERED VIEW OF (A). (C) THE GREEN CHANNEL IMAGE OF (A). D) THE GREEN CHANNEL IMAGE OF (B). THE NUMERIC ANNOTATIONS ON (B) REPRESENT THE RESULT OF THE MATCHING OPERATION WITH A). '-1' REPRESENTS NO MATCH. 1) CANDIDATE #113 IS A TRUE CANDIDATE AND IT HAS BEEN CORRECTLY MATCHED IN BOTH IMAGES (B). THE CANDIDATE HAS VARIABLE CONTRASTS IN BOTH IMAGES AS CAN BE SEEN IN (C) AND (D). THE MATCHING WILL HENCE IMPROVE THE CONFIDENCE REGARDING THIS CANDIDATE. 2) A VESSEL BIFURCATION POINT BEING FALSELY DETECTED IN BOTH VIEWS. 3) A FADED VESSEL SEGMENT BEING FALSELY DETECTED IN BOTH VIEWS. 4) A TRUE CANDIDATE BEING CORRECTLY MATCHED IN BOTH VIEWS. THE CANDIDATES MORE SUBTLE IN THE LEFT COLUMN.....	110
FIGURE 4.6. THE METHOD EMPLOYED FOR THE SCORES FUSION PHASE. $\beta 1$ AND $\beta 2$ ARE PARAMETERS SET FOR THE MODEL. ....	112
FIGURE 4.7 THE IMAGES AND VIEWS AVAILABLE FOR EACH PATIENT IN THE LONGITUDINAL DATASET. KEY: OD CENTRED - OPTIC DISC CENTRED IMAGE, ARROWS: IMAGES THAT WERE REGISTERED TOGETHER, DOTTED LINE: IMAGES THAT WERE GROUNDTRUTHED FOR MICROANEURYSMS. ....	113
FIGURE 4.8. AN ILLUSTRATION OF THE TYPES OF IMAGES AVAILABLE IN THE SPATIAL DATASET. OD CENTERED – OPTIC DISC CENTERED. ONLY IMAGES FROM 2013 WERE CONSIDERED FOR THE SPATIAL EXPERIMENT EVALUATION. ....	115
FIGURE 4.9. SAMPLE CHECKERBOARD PATCHES SHOWING REGISTRATION OF MULTIPLE TRANSFORMATION MODELS. ....	118
FIGURE 4.10 AN EXAMPLE OF OVERLAPPING VESSEL CENTERLINES FOR COMPUTING CENTERLINE ERROR AFTER REGISTRATION. THE DISTANCE BETWEEN EACH RED PIXEL AND THE NEAREST BLUE PIXEL IS COMPUTED AND THE AVERAGE OF ALL DISTANCES IS A MEASURE OF ALIGNMENT ACCURACY. ....	120
FIGURE 4.11. A BOX AND WHISKER PLOT OF THE CENTERLINE ERROR MEASURE (CEM) VALUES.....	121
FIGURE 4.12 (A) RETINAL IMAGE (B) EXTRACTION OF RETINA MASK (C) VESSEL SEGMENTS AND (D) VESSEL CENTRELINES (C). ....	123
FIGURE 4.13 CEM VALUES FOR IMAGE DM0001, WHERE 3(A), 3(B) AND 3(C) REFER TO THE METHODOLOGY STEPS MENTIONED ABOVE. ....	125
FIGURE 4.14 A COMPARISON OF PERFORMANCE BETWEEN THE TECHNIQUE APPLIED TO SINGLE IMAGES (SOLID) AND AFTER INCORPORATING SPATIAL INFORMATION (DASHED) .....	128

FIGURE 4.15 THE SCORES FROM EACH CORRESPONDING SPATIAL PAIR OF IMAGES (THESE ARE THE SAME SCORES AS ILLUSTRATED IN FIGURE 4.2). TRUE CANDIDATES ARE LABELLED IN BLUE CIRCLES WHILE FALSE DETECTIONS ARE LABELLED IN ORANGE CROSSES. THE ORANGE LINE REPRESENTS THE LINEAR LINE FOR THE EQUATION  $X=Y$ . ..... 129

FIGURE 5.1 EXAMPLES OF CANDIDATES MATCHING FROM 2013 TO 2012. THE YELLOW CIRCLES REPRESENT THE SEARCH REGION AROUND THE CENTRE OF A DETECTED CANDIDATE. THE LABELS ON THE TOP-RIGHT IMAGE REPRESENT WHICH LABEL IN THE TOP-LEFT IMAGE WAS MATCHED. A LABEL OF “-1” IN THE TOP-RIGHT IMAGE MEANS THAT THERE WAS NO MATCH FROM THE IMAGE IN 2012. A BLUE CIRCLE REPRESENTS A GROUNDTRUTHED IMAGE. 1) A CANDIDATE THAT IS VISIBLE AND DETECTED IN 2013 APPEARS WITH LESS CONTRAST IN 2012 2) A CANDIDATE WITH SIMILAR APPEARANCE IS DETECTED IN BOTH IMAGES 3) A SUBTLE CANDIDATE THAT IS CLOSE TO A VESSEL IS NOT DETECTED IN BOTH YEARS AND THEREFORE WILL NOT BENEFIT FROM THE PROPOSED APPROACH. .... 140

FIGURE 5.2 PROPOSED FRAMEWORK FOR COMBINING FEATURES PRIOR CLASSIFICATION. .... 143

FIGURE 5.3 AN ILLUSTRATION OF COMBINING FEATURES VECTOR (“FEATURE EXTENSION” IN FIGURE 5.2) BY EXTENDING MATCHED CANDIDATES  $(A_1, B_1)$  AND DUPLICATING THE FEATURE VECTORS OF CANDIDATES THAT HAVE NO MATCHES  $(A_2)$ . .... 143

FIGURE 5.4 AN ILLUSTRATION OF THE TEMPORAL DATASET AND HOW IT WAS USED TO EVALUATE THE TEMPORAL FRAMEWORK..... 145

FIGURE 5.5 EXAMPLES OF CHECKERBOARD PATCHES WHICH SHOW OTHER REGISTRATION RESULTS FOR DIFFERENT TRANSFORMATION MODELS. .... 146

FIGURE 5.6 BOX AND WHISKER PLOT OF THE CEM VALUES FOR 160 REGISTRATIONS OF TEMPORAL IMAGES. OUTLIERS ARE NOT SHOWN IN THE GRAPH SINCE IT WAS VISUALLY VERIFIED THAT THEY ARE DUE TO AN INACCURATE VESSEL SEGMENTATION WHILE COMPUTING THE CEM VALUES (FIGURE 5.7). .... 148

FIGURE 5.7 AN EXAMPLE OF AN OUTLIER RESULTS FOR THE CEM VALUES. IT CAN BE OBSERVED THAT THE HIGH VALUE OF CEM IS CAUSED DUE TO INACCURATE AUTOMATIC SEGMENTATION. THE PURPLE PIXELS REPRESENT VESSELS THAT ARE NOT ALIGNED WITH THE CORRESPONDING VESSELS IN THE OTHER PAIR. WHILE THE ALIGNMENT IS VERY GOOD IN THIS CASE, THE CEM VALUES WILL BE HIGH AS A RESULT OF INACCURATE SEGMENTATION. .... 148

FIGURE 5.8 SCATTER PLOT OF THE PROBABILITIES FOR THE SPATIAL IMAGE PAIRS. SCORE 2 REPRESENTS THE SCORE OF THE CANDIDATES BEING CLASSIFIED (2013 IMAGE) WHILE SCORE 1 AXIS REPRESENTS THE SCORES THAT ARE MATCHED TO SCORE 2 (2012 IMAGE). BASED ON THE PLOT ABOVE, THE THRESHOLD HAS BEEN CHOSEN TO BE BETWEEN 0.18 AND 0.3. .... 151

FIGURE 5.9 FROC PLOTS FOR THE ORIGINAL ENSEMBLE APPLIED TO A SINGLE IMAGE. THE OTHER FROC CURVE IS APPLIED TO THE IMAGE OF INTEREST AND THEN MATCHED WITH THE NEAREST CORRESPONDING CANDIDATE. THE MAXIMUM OF BOTH SCORES IS TAKEN IF THE ORIGINAL CANDIDATE SCORE (SCORE 2) LIES BETWEEN THE RANGE OF 0.18-0.3..... 151

FIGURE 5.10 AN ILLUSTRATION OF SEVERAL PATCHES AFTER CANDIDATE MATCHING AND ASSIGNING SCORES. IN EACH PATCH OF 2012 (A) THE ANNOTATIONS CORRESPOND TO THE CANDIDATE NUMBER FOLLOWED BY THE CLASSIFIER SCORE. IN EACH ANNOTATION OF (B) THE FIRST NUMBER CORRESPONDS TO THE MATCHED CANDIDATE FROM 2012 ('-1' IF NO MATCH EXISTS) FOLLOWED BY THE CLASSIFIER SCORE OF THE CANDIDATE). IN (1) AND (2) A TRUE CANDIDATE HAS BEEN MATCHED TO A PREVIOUS CANDIDATE AND THEREFORE THE SCORES WILL BE CORRECTLY FUSED. IN (3), DUE TO A MISMATCH OF REGISTRATION A CANDIDATE HAS BEEN MISALIGNED AND A NEIGHBORING CANDIDATE WAS MATCHED WITH THE 2012 CANDIDATE. THIS CAN CAUSE A FALSE POSITIVE TO BE MATCHED TO A TRUE POSITIVE WHICH WILL AFFECT THE RESULTS NEGATIVELY AFTER THE SCORES ARE FUSED. IN (4) A FALSE POSITIVE HAS BEEN CORRECTLY MATCHED WITH ANOTHER FALSE POSITIVE WITH A HIGHER SCORE. THIS IS THE MAIN REASON FOR THE LACK OF IMPROVEMENT OF THIS METHOD. .... 152

FIGURE 5.11 FROC CURVE FOR THE FEATURE EXTENSION ON THE TEMPORAL DIMENSION COMPARED WITH THE PROPOSED ENSEMBLE ALGORITHM ..... 154

# List of Tables

TABLE 1.1 CLASSIFICATION OF DIABETIC RETINOPATHY. TABLE FROM [8].	15
TABLE 2.1 SUMMARY OF PUBLICLY AVAILABLE RETINAL IMAGE DATASETS.	26
TABLE 2.2. SUMMARY OF MA DETECTION ALGORITHMS IN THE LITERATURE. THE PERFORMANCE SUPERSCRIPTS ARE DEFINED AS FOLLOWS: <sup>A)</sup> LESION-BASED MEASURE <sup>B)</sup> IMAGE-BASED MEASURE <sup>C)</sup> PIXEL-BASED MEASURE. KEY: AUC – AREA UNDER THE CURVE, FP/IMAGE - FALSE POSITIVES PER IMAGE, PPV – POSITIVE PREDICTIVE VALUE. TABLE FROM [31].	31
TABLE 2.3 IMAGE REGISTRATION ALGORITHMS IN THE LITERATURE.	38
TABLE 3.1 MICROANEURYSMS INITIAL CANDIDATES DETECTION METHODS IN THE LITERATURE.	55
TABLE 3.2 FEATURES LIST. THE SYMBOLS BELOW (G,T,SEED,C, $\Sigma$ ) ARE DEFINED BELLOW IN THIS SECTION. KEY: STD - STANDARD DEVIATION, MAX - MAXIMUM, MIN – MINIMUM.	62
TABLE 3.3 DISTRIBUTION OF DR GRADES (A) AND RESOLUTIONS (B) OF IMAGES IN THE DATASET.	69
TABLE 3.4 ROC SCORES FOR MULTIPLE CATEGORIES IN THE SET. BOLD REPRESENTS THE MAXIMUM SCORE.	79
TABLE 3.5 AREA UNDER THE CURVE FOR PER-PIXEL EVALUATION OF MICROANEURYSMS ACROSS VARIOUS DATASETS	86
TABLE 3.6 THE 5 MOST (A) AND LEAST (B) DISCRIMINATIVE FEATURES ACCORDING TO THE BAGGING FEATURE IMPORTANCE MEASURE.	92
TABLE 3.7 AVERAGE PERFORMANCE OF THE PROPOSED ALGORITHM ON BOTH MESSIDOR AND DIARETDB2 DATASETS. THE GIVEN TIMES CORRESPOND TO THE AVERAGE TIME PER IMAGE FOR EACH DATASET (WITH THE EXCEPTION OF THE CLASSIFICATION TIME).	93
TABLE 4.1 THE DISTRIBUTION OF IMAGE RESOLUTIONS IN THE LONGITUDINAL DATASET.	115
TABLE 4.2 THE NUMBER OF CONTROL POINTS NEEDED FOR EACH TRANSFORMATION MODEL	118
TABLE 4.3 CENTERLINE ERROR METRIC (CEM) MEAN AND STANDARD DEVIATION VALUES FOR VARIOUS TRANSFORMATION MODELS.	120
TABLE 4.4 CONFIDENCE INTERVAL FOR EACH OF THE 20 IMAGE SETS (MEAN $\pm$ STANDARD DEVIATION). THE THIRD COLUMN REPRESENTS THE % IMPROVEMENT OF 3C RELATIVE TO 3B (RADIC COMPARED TO AFFINE).	124
TABLE 4.5 SUMMARY OF THE GROUNDTRUTH CANDIDATES THAT CONTAIN MATCHES IN THE CORRESPONDING SPATIAL PAIR.	129

TABLE 5.1 CENTERLINE ERROR METRIC (CEM) MEAN AND STANDARD DEVIATION VALUES FOR VARIOUS TRANSFORMATION MODELS

..... 148

## **Glossary of Acronyms and Abbreviations**

1D	one-dimension/one-dimensional
AMD	Age-related macular degeneration
CEM	Centreline error measure
FN	False negatives
FOV	Field of view
FP	False positives
FROC	Free-receiver operating curve
DR	Diabetic retinopathy
KNN	K-nearest neighbours
MA	Microaneurysm
NHS	National health service
OCT	Optical coherence tomography
OD	Optic disc
PPV	Positive predictive value
RADIC	Radial distortion correction
RPE	Retinal pigment epithelium
ROC	Retinopathy online challenge (or receiver operating curve)
std	Standard deviation
SIFT	Scale-invariant feature
SVM	Support vector machines
TN	True negatives
TP	True positives

# 1

## INTRODUCTION

Our eye is a window that allows us to see the outside world. In addition, it is also a window that allows us to capture the retina, a very important tissue at the back of the eye that helps doctors diagnose several eye diseases related to diabetes. Diabetes patients are at risk of having several eye conditions such as Diabetic Retinopathy (DR), Diabetic Macular Edema (DME), cataract and glaucoma. [1]. All these forms of eye disease have severe potential to cause vision loss and blindness.

Diabetic Patients are required to attend regular eye screening appointments. The purpose of these screening appointments is to assess whether the patients suffer from Diabetic Retinopathy – a retinal disorder. From these appointments digital retinal images are captured, which then undergo various stages of analysis by trained *graders*. Grading is part of the screening process and it involves assessing the severity of Diabetic Retinopathy for a patient by observing images of their retina. Grading is a very time-consuming process which adds a high cost to the screening process due to the large number of patients. Therefore, automating the process of grading from retinal images can help reduce this time cost that is incurred. Other benefits that this automation could add to the National Health Service (NHS)

include: 1) Faster grading results means that patients can receive feedback instantly 2) Automated grading would be more objective since computers are not subject to human errors or fatigue.

This thesis addresses the problem of automated detection of one of the symptoms of Diabetic Retinopathy in patients. The next section will explain the anatomy of the human eye in depth. A list of the imaging techniques that are available to capture the retina is discussed next. Following that, a detailed account of Diabetic Retinopathy and its diagnosis will be presented. All of this background information will lead to the aims, objectives and contribution to knowledge of the work, concluding this chapter.

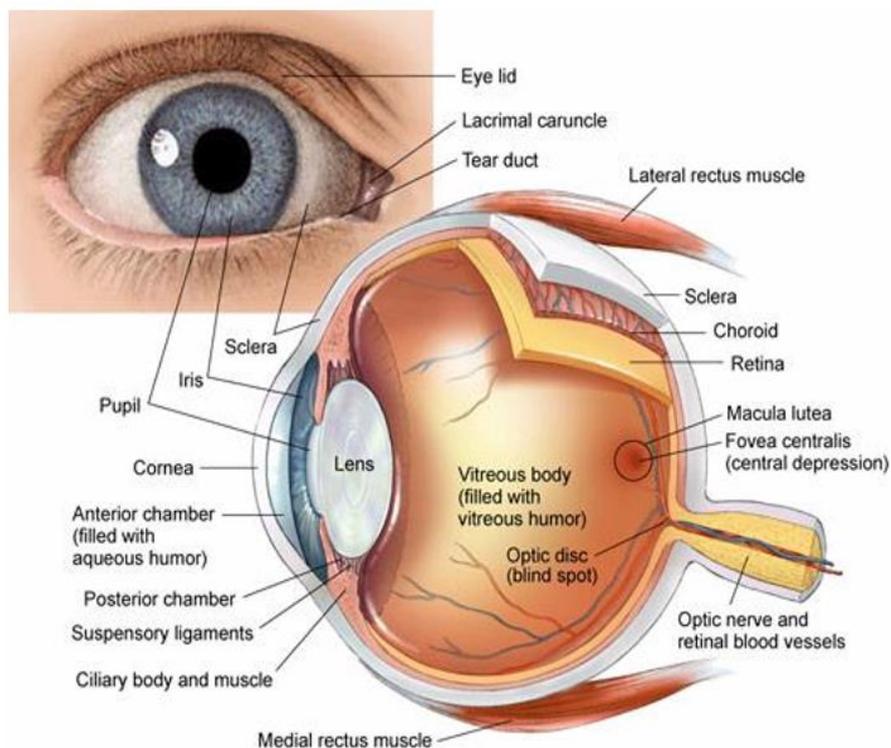
## **1.1 The anatomy of the Eye**

Our eyes allow us to see and interpret the world around us by processing the light that gets reflected by objects in the world. An illustration of the anatomy of the eye is presented in Figure 1.1. The outermost tough layer of the eye is called the sclera and it maintains the shape of the eye. Light waves from the outside world enter the eye first through the cornea, which is a clear dome at the front of the eye. The cornea is the eye's window since it allows light to enter the eye. Light enters the eye through the cornea and then progresses through the pupil – which is a circular opening in the centre of the iris. Light waves are first converged by the cornea and then they are further bent by the lens cortex, which is located immediately behind the iris and the pupil. The light converges to a focal point located immediately behind the back surface of the lens. The region where the light converges at the back of the eye is known as the fovea. This region contains millions of photoreceptors that capture light rays and convert them into electrical impulses. The impulses are transmitted through the optic nerve to the brain. The brain is responsible for interpreting

the light signals and “forming” the image. This means that we do not “see” with our eyes, but rather our brains are responsible for this interpretation of light waves[2].

Blood vessels in the human eye originate from the ophthalmic artery, which splits into the central retinal artery, the posterior ciliary artery and the anterior ciliary artery. Arteries transport oxygenated blood from the lungs to the rest of the body tissues. The central retinal artery enters via the optic nerve slightly behind the eyeball and appears at the optic disc opening. It further divides into four branches, each supplying blood to a quadrant of the retina.

(a)



(b)



Figure 1.1 a) Anatomy of the eye. Image from [www.biographixmedia.com](http://www.biographixmedia.com). b) Examples of healthy retinal images. Labels of the main retinal features are highlighted. Images from [3].

## 1.2 Retinal Imaging

The retina is one of the few regions in the human body where it is possible to perform non-invasive inspection of blood vessels. This opens up potential for a wide range of medical applications [4]. By utilising imaging techniques such as retinal fundus photography and Optical coherence tomography (OCT), it is possible to acquire high resolution imagery of the retina for offline analysis. This has had a lot of medical applications, including:

1. **Screening programs:** where patients with a suspected retinal disorder are assessed. Screening tools can aid clinicians and experts to identify regions of abnormality in an image by performing automated segmentation of the captured images. For example, screening tools may be able to perform vessel segmentation [5], vessel width measurement [2], exudate detection [6], microaneurysm detection [7] and new vessel detection [8].

2. **Medical diagnosis:** There is a correlation between certain abnormalities in the retina and the prevalence of some diseases. This fact can be used to perform medical diagnosis. For example, the presence of exudates and micro aneurysms is linked to diabetic retinopathy - a leading cause of blindness worldwide [7].
3. **Pathology research:** There are speculations that some abnormalities in the retina are correlated with certain diseases or are an indication of a risk factor of this disease. For example, the changes in morphology, tortuosity and other retinal blood vessel features are thought to be linked to various clinical disorders such as cardiovascular diseases [5]. However, in order to show that there is a correlation statistical analysis needs to be performed on a large number of images acquired from a large number of patients. Automated segmentation and measurement tools can help perform bulk segmentation and analysis of a large number of images to aid this task.

One of the early successful attempts to visualise the retina was the ophthalmoscope, which is a hand-held tool that allows direct (online) inspection of the retina. Fundus imaging is the process of capturing an image of the 3D semi-transparent retinal tissue using the image intensities to represent the amount of light reflected [9]. The interior surface of the eye – including the retina, optic disc, macula and posterior pole can be visualised in a fundus image. This is the only non-invasive method through which blood vessels can be visualized and inspected [4]. A fundus camera is used to capture a fundus image. It consists of a specialized low power microscope with an attached camera.

As its name suggests, colour fundus photography involves an examination of the retina in full colour under the illumination of white light. The image intensities are represented by the amount of reflected red, blue and green wavebands. The retinal pigment poorly absorbs

red light and hence the choroid and choroidal pattern can be visualized in this band. The choroid is the vascular layer behind the retina which contains connective tissues. It lies between the retina and the sclera. The green band provides good contrast for visualizing the retinal vascular network, haemorrhages, drusen and exudates. Drusen are yellow residue which form under the retina and they are made up of proteins. They are considered to be early signs of age-related macular degeneration. Haemorrhages and exudates are signs of diabetic retinopathy and will be explained in 1.3. In the blue channel, the anterior retinal layers of the retina can be viewed by providing a dark background against which the top layers of the retina are imaged [10]. Early treatment diabetic retinopathy study (ETRDS) group introduced a standard of 7-field photography gold standard for the diagnosis of diabetic retinopathy [11]. However, capturing 7 fields of the retina is a time-consuming process and requires skilled photographers as well as pupil dilation. As a result, 2 fields (45 degrees) was suggested an alternative of (macular centred and optic disc centred)[12].

Fluorescence angiography [13] is a technique for examining the blood circulation in the retina. Sodium fluorescent is injected into the circulation system and then a fluorescence angiogram is obtained by capturing a video sequence of the fluorescence as it progresses through the blood circulation system. The images are captured by emitting blue light at a wavelength of 490 nanometres. Indocyanine green (ICG) is another variant to Sodium fluorescent. It is captured at a wavelength of 800 nanometres. Light at this frequency penetrates the retinal layers and allows ICG angiography [14] to capture deeper patterns of the circulation compared to fluorescence angiography. This makes it a better option for capturing the choroidal circulation. Fluorescence angiography is especially useful in the diagnosis of Diabetic Retinopathy and Macular Degeneration since the blood flow can be

visualised and any barriers that obstruct the blood flow can be observed. This work is focused on the detection of one of the symptoms of Diabetic Retinopathy (DR) and hence this is described in section 1.3

In Stereo fundus photography, [15] multiple images from two or more different view angles are captured of the same retina in order to infer depth information and maximise photographic visual information. A modified fundus camera can capture sequential hyperspectral images of the retina. The intensities in these images represent the light reflected by several wavelength bands from the retina [16]. This is useful for non-invasively measuring the retinal biochemistry, and particularly to measure blood oximetry — the fraction of oxygen-saturated haemoglobin relative to the total haemoglobin in the blood, from the retinal vasculature [17].

Widefield-view imaging makes it possible to capture a much larger region of the retina in a single image. While single-field colour fundus photographs typically capture 45 degrees field-of-view of the retina, wide-field imaging is capable of capturing up to 200 degrees. The region around the posterior pole (optic disc and macular) of the eye is where the most important ocular diseases occur including glaucoma, macular degeneration, diabetic retinopathy (DR) and optic neuropathy [18]. The advantages of single-field fundus photography include convenience to the patient (requiring less time, in many cases alleviating the need for pupil dilation and less light exposure). A report published by the American Academy of Ophthalmology in 2004 suggests that there is sufficient evidence to show that single-field fundus images are sufficient for the screening of diabetic retinopathy [19]. Wide-field view retinal images can be obtained by one of three methods: 1) Creating a montage of multiple single-field images, 2) Using a special lens in front of a regular fundus

camera and 3) Using a special wide-angle camera[18]. The first wide-angle camera system was developed in 1977 and used a contact lens and fibre optic scleral trans illumination. It had a field of view of up to 148 degrees [20]. Recently, Optos (Optos, Dunfermline, United Kingdom) has introduced retinal imaging technology with non-contact scanning laser which enabled the capture of a larger field of view (200 degrees). Several studies that diagnosing Diabetic Retinopathy (DR) from wide-field view images resulted in a higher grade compared to the set of single-field images recommended by ETRDS group[21]–[26].

Recently portable mobile phone based photography has been introduced by the portable eye examination kit (PEEK) [27]. The use case for this is for developing countries in order to save the costs of eye examination. In developing areas, eye patients with poor access to roads may not get screened frequently enough. This is where portable eye examinations using handheld mobile phones may serve to fill this shortcoming. Moreover using PEEK retina device can economise when it comes to the equipment costs. At the time of writing this report, a PEEK retina device costs \$233, which greatly reduces the costs compared to a fundus camera which can cost thousands of dollars.

Optical coherence tomography (OCT) is a non-invasive, non-contact, fast and sensitive ophthalmic imaging technique to capture portions of the retina. OCT allows for the measurement of various ocular structures of the retina including the anterior segment, the optic nerve and the retinal tissue. OCT uses light to capture micrometre resolution and 3D images from biological tissues such as the retina. OCT is based on low-coherence interferometry. Light emitted from the light source is split into two arms – sample arm and a reference arm (usually a mirror). The reflected light is combined only if both arms have travelled the same optical distance. Using this fact, a reflectivity profile of the light sample

can be obtained (time domain of the OCT). This reflectivity profile is known as an A-scan and it contains information about the spatial dimensions and locations of the structures within the item of interest. Several A-scans can be combined laterally along the axial length to form a cross-sectional tomography (B-scan).

Adaptive Optics Retinal Imaging allows high resolution photography of the retina [28]. Adaptive Optics is a technology that can improve the performance of retinal imaging by reducing the effect of wavefront distortions by correcting for the deformations of an incoming wavefront. It works by deforming a mirror in order to correct for the eye's aberrations and providing nearly diffraction-limited resolution. As a result, it is possible to view the retina with a nearly cellular resolution. The availability of commercial systems and the mounting examples of clinical utility from research grade systems has greatly increased interest in Adaptive Optics imaging amongst practitioners. Currently Adaptive Optics can integrate with fundus cameras, OCT and Scanning Laser Tomography. Retinal structures that can be visualised include photoreceptors, retinal pigment epithelium (RPE) and retinal vasculature. It is possible to visualise even the smallest capillaries in the retinal vasculature and non-invasively measure blood velocity. Adaptive Optics can be used in the diagnosis of Glaucoma, Diabetic Retinopathy, Age-related macular degeneration and inherent retinal degradations.

### **1.3 Diabetic Retinopathy**

Diabetic retinopathy is a serious and common eye complication that is caused as a result of Diabetes [29]. It is a complication of diabetes that affects the retinal vasculature and leads to progressive retinal damage. This could end in loss to vision and ultimately blindness [29]. Diabetic Retinopathy (DR) DR has been identified as a significant global public health

problem [30]. Diabetic retinopathy is the most common cause of vision loss among people with diabetes and a leading causes of blindness among working age adults [1], [31], [32]–[34]. Diabetic Retinopathy is often not noticed and due to this reason diabetes patients need to get their retina regularly screened in order to increase the chances of early detection. If it is detected early, it can increase the chances of effective treatment by 95 percent [1]. Laser treatment can prevent sight loss if DR is detected early [35]–[37]. Prevention or further progression of DR can also be achieved at the early stages by controlling blood glucose level [38]. Computer automation and automated detection systems would greatly benefit the field of Diabetic Retinopathy [39]. The benefits include a more cost-effective diagnosis and treatment referral.

There are often no symptoms in the early stages of diabetic retinopathy and therefore screening is crucial for identification. Early identification will lead to patients being managed accordingly and hence contribute to sight loss prevention. In 2000, a systematic national screening programme was introduced based on digital photography[40]. This programme meets the Wilson and Jungner criteria for a screening programme [41]. A screening programme is currently in operation in the United Kingdom with diabetic patients aged 12 and above. The programme requires these patients to attend a retinal screening annually using digital photography [42].

Diabetic retinopathy is a microvascular disease which affects the retina and is caused by hyperglycemia. A damage to retinal blood vessels causes blood and other fluids to leak onto the retina. This leads to the formation of features such as microaneurysms, haemorrhages, exudates, cotton wool spots and venous loops [43]. As time goes by blockages and damage to blood vessels will cause areas of retinal ischemia to develop. As a result, the growth of

new blood vessels is triggered. The growth of new vessels happens in a later stage of the disease and poses a high risk of severe vision loss since new vessels are fragile in nature and are susceptible to bleeding. The main features that occur as a result of Diabetic Retinopathy are described next:

**Microaneurysms:** These are bulges that form on the sides of capillaries. They arise due to the weakening of capillary walls. Capillaries are not visible in fundus images. As a result, microaneurysms appear like isolated red dots, not attached to any blood vessels. These are one of the earliest signs of Diabetic Retinopathy that can be detected (Figure 1.2(a)). Detection of microaneurysms is the focus of this work. Due to the subtle nature of microaneurysm appearance and the variability of their contrast in images, the automated detection of these features can be challenging (Figure 1.4).

**Haemorrhages:** When capillary walls break down they result in the leakage of blood. This can take the form of various sizes and shapes, depending on the retinal layer in which these leaking vessels are located. They are often called dot, blot or flame haemorrhages, depending on appearance (Figure 1.2(b)).

**Exudates:** A leakage of fluid causes edema (swelling in the blood vessels) and can occur as a result of capillary breakdown. After the edema builds up retinal thickening can occur. The presence of edema in the macular region of the retina is known as macular edema – one of the most common causes of visual impairment amongst diabetic patients. Laser treatment can reduce the risk of visual sight loss [44]. Since the fluid is clear it is not visible on a standard 2D retinal image. However, the insoluble residue from edema is visible in fundus images and they are known as Exudates. They appear as waxy yellow lesions but their shape appearance varies. They can take various patterns including individual patches, tracking

lines, rings (circinates) and macular stars (Figure 1.2(c-d)). Hence the presence of exudates are considered signs of macular edema.

**New Vessels:** After the progression of capillary breakdown, areas of the retina can become ischemic and trigger the growth of new vessels in the affected region as an attempt to re-vascularize the oxygen deprived tissue. There are two kinds of new vessels: New vessels on the Optic Disc (NVD) and New Vessels Elsewhere (NVE) (Figure 1.3). New vessels are extremely fragile and can easily rupture, resulting in extensive haemorrhaging in the form of pre-retinal and vitreous haemorrhages. Haemorrhages occur due to the fact that new vessels tend to grow away from the retinal surface and towards the vitreous. Further complications such as retinal detachment may occur as a result. New vessels are considered late stage diabetic retinopathy and represent a high risk of severe loss of vision. Laser treatment is required to reduce the risk of blindness. Their appearance in fundus image exhibits small networks or loops of tortuous vessels when they initially develop. They can also be very difficult to distinguish from the background as they tend to grow away from the retinal surface – and are hence away from the focal plane of the fundus photograph. Due to this reason, they exhibit a blurry and obscure appearance.

A comprehensive list of Diabetic Retinopathy features is presented in Table 1.1. The progression of the diseases in the table is listed in ascending order (With the exception of maculopathy). Since the disease is progressive these features accumulate as we move down the table. Background Diabetic Retinopathy is the earliest stage in Diabetic Retinopathy and does not pose a threat to vision. Pre-proliferative Diabetic Retinopathy represents progressive retinal ischaemia and increases the risk of progression of new vessels. Proliferative Diabetic Retinopathy is characterized by presence of new vessels and is the

most advanced stage of the disease. It poses severe risk of vision loss. Maculopathy can occur at any stage of Diabetic Retinopathy and is more likely to occur at a later progress of the disease. The definition of maculopathy is the existence of any Diabetic Retinopathy features at the macular regions. However, it is commonly used to refer to vision threatening macular edema.

The impact of diabetic retinopathy and the necessity of retinal screening has been discussed in this section. Further details about the screening process for diabetic retinopathy shall be provided next, followed by an explanation of the anatomy of the eye. The aims, objectives and contributions to knowledge will be discussed next.

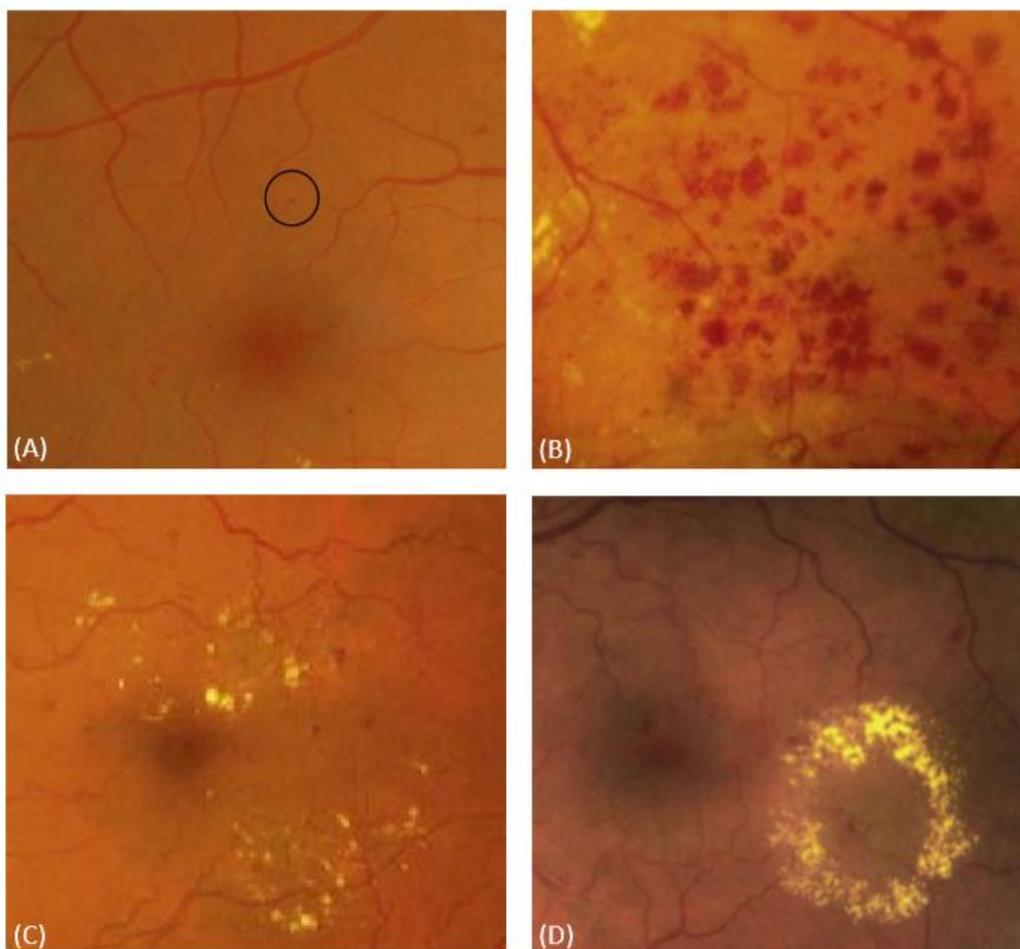


Figure 1.2(A) Microaneurysms, (B) dot and blot haemorrhages, (C) exudates, (D) a ring of exudates (circinate). Images from[44].

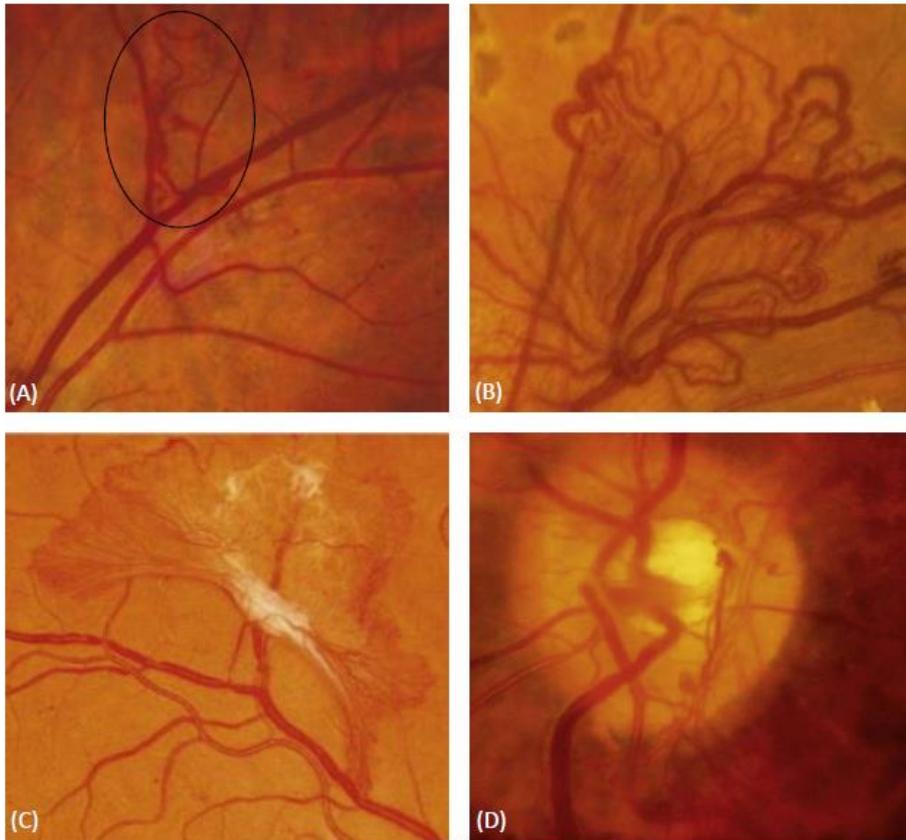


Figure 1.3 a) Mild New vessels Elsewhere (NVE), b) severe New vessels Elsewhere (NVE), c) severe NVE with associated fibrosis, d) mild NVD. Images from [44].

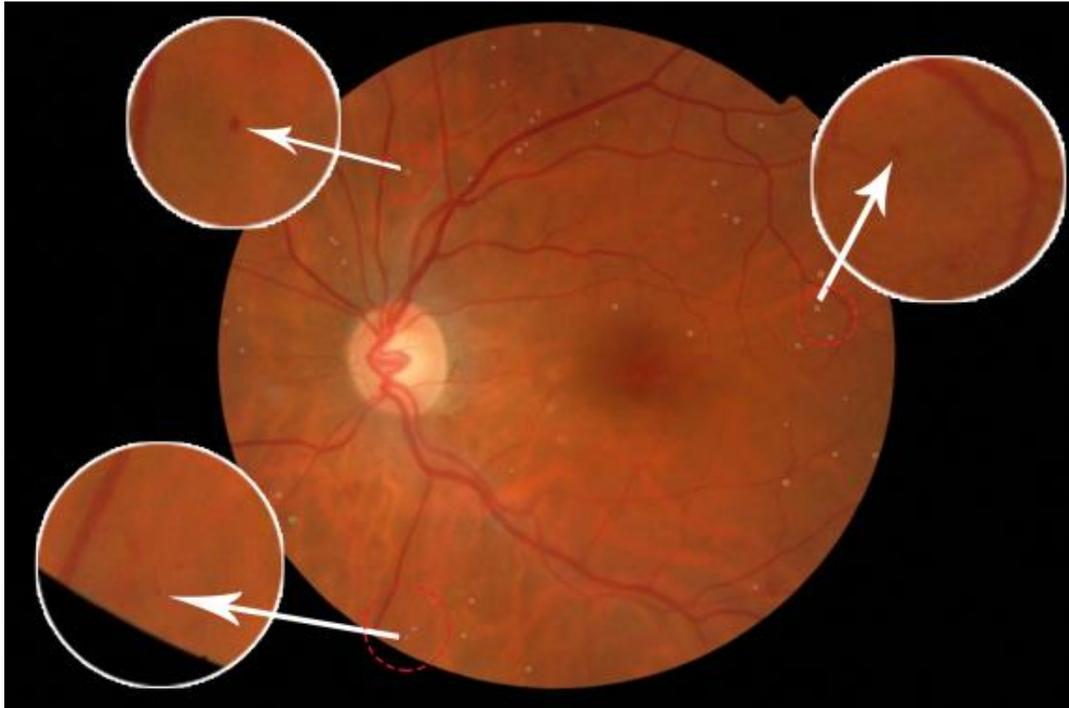


Figure 1.4 A colour retinal image with microaneurysms of various contrasts highlighted image from MESSIDOR.

Table 1.1 Classification of diabetic retinopathy. Table from [8].

Stage of retinopathy	Features
Background	Microaneurysms
	Dot and blot haemorrhages
	Exudates
Pre-proliferative	Multiple dot and blot haemorrhages
	Cotton Wool Spots
	Venous bleeding and loops
	Intra-retinal microvascular abnormalities (IRMA)
Proliferative	New vessels everywhere (NVE)
	New vessels at the disc (NVD)
	Pre-retinal/vitreous haemorrhages
	Retinal detachment
Maculopathy	Microaneurysms, haemorrhages and exudates at the macula
	Macular odema

## 1.4 Aims, Objectives and Contribution to Knowledge

The aim of this project is to develop techniques to detect microaneurysms from fundus retinal images. As mentioned earlier, the goal of Diabetic Retinopathy screening programmes is the early detection of disease to allow treatment in the early stages before the disease progresses.

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 with 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 for the patient is achieved. Microaneurysms are one of the earliest visible signs of Diabetic Retinopathy and therefore it is crucial to detect them reliably. As can be seen in Figure 1.4, microaneurysms appear as small red dots of varying sizes, shapes and contrast. Their sizes are very small and they can be often confused with noise or other artefacts in the image.

Several algorithms for microaneurysm detection have been proposed in the literature. The proposed microaneurysm detection algorithms have attempted to detect microaneurysm regions from given colour fundus images. The following points need to be considered regarding the current state of retinal imaging:

1. Screening programs for Diabetes patients capture at least 2 colour fundus images per patient (two 45-degree images, one optic disc-centred and another macula-centred). These images contain a common overlapping region.
2. Screening programs are frequent; this means that it is likely that a patient arriving for a screening session would have had his retinal images captured in the past.

3. There are different modalities used depending on application. These are captured based on clinical application as described in Section 1.2. Patients may have more than one modality captured depending on the clinical problem. An accurate diagnosis in this case may mean that doctors refer to information across modalities. For example, after a fundus screening, a doctor may refer a patient to an OCT image capture if there is possibility of age-related macular degeneration (AMD).

Inspired by the points mentioned above, the objectives of the work are:

**Objective 1:** Investigate and implement a microaneurysm detection algorithm that can be applied to a single image, but has the potential to be applied to datasets that contain spatial and temporal information. Evaluate the algorithm performance on a single image and compare to state-of-the-art techniques.

**Objective 2:** Create a framework based on the algorithm in objective 1 that is capable of combining information from an image pair so that a segmentation decision is performed collectively.

**Objective 3:** Apply the framework to a set of spatial image pairs (optic disc centred and fovea centred). Assess whether an improvement in accuracy is achieved when compared to application to a single image.

**Objective 4:** Apply the framework to a set of temporal image pairs. These are images of the same retina captured at different time intervals. Assess whether an improvement in accuracy is achieved when compared to application to a single image.

The novelty of this approach is that the problem of detection has been redefined in light of the availability of this extra information and algorithms have been developed according to

this situation. A further application of point number 3 is that it is possible that with the arrival of hand-held fundus photography devices such as PEEK vision [27], the combination of spatial information becomes increasingly important since these devices have a small field of view (FOV) and therefore multiple spatial views need to be captured, aligned and have their information combined. The contributions of this work are as follows:

- 1- Development of a new MA detection algorithm that uses an ensemble classifier to detect candidates from a *single* colour fundus retinal image. The ensemble classifier gave us a flexible framework to extend the analysis from a single image to more than one.
- 2- Development of a novel framework (based on the ensemble algorithm in point 1 for combining information from image pairs based on a pair of retinal images that contain an overlap.

The assessment of point number 2 has been made for both spatial and temporal pairs of images. The algorithm in point 1 was used as a baseline for measuring the performance and the improvement from using the framework was assessed compared to the baseline accuracy. In the next Chapter, an extended literature review is presented to summarise the work that has been done in the past related to microaneurysm detection. Since the project is related to image registration, the relevant literature includes image registration algorithms, and multi-image analysis work. Chapter 3 will present a proposed combination of algorithms for detection of microaneurysms from a single retinal image. In Chapter 4 the algorithm will be extended to make use of spatial information from retinal images. Chapter 5 will present experimental results for the combination of information from temporal images. Discussions and concluding remarks will be presented in Chapter 6.



# 2

## **LITERATURE REVIEW**

---

This chapter starts with a general survey of the field of Retinal Image Analysis and the various research areas involved in this field. An overview of the available retinal imaging datasets in the public domain will be presented. An extensive literature review of Microaneurysm detection – which is the focus of this thesis -- is then presented. The topic of this thesis is a combination of two fields: Microaneurysm Detection and Retinal Image Registration. Therefore, a survey of microaneurysm detection techniques is covered, and then a survey of image registration techniques will be discussed. A general survey of the field of image registration is discussed, followed by a deeper discussion of retinal image registration in the context of retinal images. Finally, for the sake of completeness, the field of Multi-image Analysis is discussed. This field includes image analysis algorithms that make use of multiple images to perform the analysis.

### **2.1 Retinal Image Datasets**

The existence of public datasets is crucial for the research in the field of retinal images. It is very important to be able compare reported performances of multiple proposed algorithms. Evaluation on public datasets enables this direct comparison. The challenges to making

medical image datasets public include the need for ethical approval. A summary of available retinal imaging datasets is mentioned and discussed in the following sections. A summary of the datasets, their resolutions, the groundtruth types available and the number of images available for each dataset is presented in Table 2.1.

### **2.1.1 DIARETDB**

The DIARETDB database was published in 2008 and 3 versions of the dataset were released: DIARETDB0, DIARETDB1 and DIARETDB1 v2.1 [45], [46]. DIARETDB0 consists of 130 retinal images, of which there are 20 normal images and the rest contain various symptoms of diabetic retinopathy. According to the dataset website, DIARETDB1 has been deprecated in favour of DIARETDB1 v2.1. This set consists of 89 images out of which 5 images are healthy and the remaining show at least signs of diabetic retinopathy. The images were marked by four experts for the presence of microaneurysms, haemorrhages, and both hard and soft exudates. The resolution of the images was 1500 x 1152 pixels and they were stored in PNG format. The images were captured at a 50 degrees FOV.

### **2.1.2 MESSIDOR**

The MESSIDOR database [3] is one of the largest databases of public datasets currently available online. It consists of 1200 retinal images and is provided courtesy of the MESSIDOR program partners. The images were captured at three different ophthalmology departments using a non-mydratic 3CCD camera (Topcon TRC NW6) at 45 FOV resolution. Non-mydratic cameras eliminate the need for pupil dilation before capturing retinal images of patients. Furthermore, they eliminate bright-light flashes in the patient's eye. The image resolutions were 1440 x 960, 2240 x 1488 and 2403 x 1536 pixels. The images were stored in TIFF format. 800 out of the 1200 images were captured with pupil dilation. The reference

standard that is bundled with the database contains the grading information for Diabetic Retinopathy (DR grade) and risk of macular edema for each image. Macular edema occurs when there is deposits of fluids or proteins close to the macula in the eye. These fluids may appear as exudates close to the macula region in retinal images. Moreover, there are no labels or markings for the location of individual marking labels that come with the database. As a part of this thesis, 32 images from the MESSIDOR dataset have been groundtruthed and the labels have been published online for the sake of comparative evaluation<sup>1</sup>[31].

### **2.1.3 ROC Dataset**

The Retinopathy Online Challenge (ROC) dataset is a set released as a part of an online competition for the detection of microaneurysms [47]. The database contains 100 digital colour fundus photographs with microaneurysms existing in some of the images. The dataset is divided into testing and training sets (50% each). Microaneurysm groundtruth labels are available only for the training set. The groundtruths are provided as circular MA objects: each object label consists of the center coordinates of the object and the radius of the object. The assumption is that all MA candidates are approximately circular. The test set groundtruths are not public since the contest organizers used those to evaluate submissions. Moreover, the groundtruth of this dataset has generated discussion in the literature [48], [49] due to the fact that many of the MA candidates marked in the groundtruth are invisible to the viewers or could not be seen by other expert observers<sup>2</sup>. The images are captured using either a TopCon NW100 or a Canon CR5-45NM camera at 45

---

<sup>1</sup> The groundtruth dataset can be downloaded using the following link:

<http://blogs.kingston.ac.uk/retinal/?p=311>

<sup>2</sup> The dataset authors have been contacted by me several times to provide the test set ground truths but they have not been responsive.

degrees FOV and was JPEG compressed in the camera. There are three different image sizes present in the database; 768×576, 1058×1061 and 1389×1383 pixels.

#### **2.1.4 CHASE**

The CHASE dataset [5], [50] contains a set of 14 image pairs (right and left eye) captured from 9-10 year old children of different ethnic origin – forming a total of 28 images. The groundtruths for blood vessels are distributed along with the dataset. 2 human observer vessel segments are available for each image in the dataset. Due to the young ages and diverse ethnic origins of the eyes captured, the images contain subtle differences in background levels of retinal pigmentation (being more pigmented in South Asians compared to White Europeans)

#### **2.1.5 STARE**

The STructured Analysis of the Retina(STARE) dataset is suitable for evaluating vessel segmentation algorithms in colour fundus images [51]. The dataset contains 20 images and 10 of those consist of pathology. The images were captured using a TopCon TRV-50 fundus camera at a 30 degrees field of view (FOV). The images have a resolution of 605 x 700 pixels. The approximate diameter of the FOV is 650 x 500 pixels. The images were segmented by two human observers. The first observer segmented ~10% of the pixels as vessel pixels while the second observer segmented 14.9% of the pixels. The reason for this discrepancy is that one of the observers marked more of the thinner vessels.

#### **2.1.6 DRIVE**

The Digital Retinal Images for Vessel Extraction (DRIVE) database[52] is a public dataset that consists of 40 colour fundus images. The photographs were obtained from a Diabetic Retinopathy screening program in the Netherlands. Each image has been JPEG compressed,

which is common practice in screening programs. 7 of the 40 images contain a pathology (exudates, haemorrhages and pigment epithelium changes). The images were acquired using a Canon CR5 non-mydratic 3-CCD camera with a 45 degree FOV. The resolution of the images was 768 x 584 pixels. The FOV of each image is circular and consists of a diameter of ~540 pixels. The set has been divided into training and test sets (50% each). Three human observers were involved in the groundtruth process. The observers were trained by an expert ophthalmologist. Each observer marked a mutually exclusive set of images. Vessel map groundtruths are made available for both the tests and training tests.

### **2.1.7 ARIA**

This dataset consists of three groups: age-related macular degeneration group (92 images), diabetes group (59 images) and a control group (61 images). Blood vessels, optic disc and fovea regions are marked in all of the images. The images have a resolution of 768 x 576 pixels in RGB colour. The images were captured using a Zeiss FF450+ fundus camera at a 50 degree FOV. The images are stored in an uncompressed TIFF format. This dataset was published in 2006 as a part of a research collaboration between Royal Liverpool University Hospital Trust, Liverpool, UK and the University of Liverpool [53].

### **2.1.8 Hossein**

Dr Hossein Rabbani[54] has published several retinal imaging datasets, including multimodal datasets. The modalities include fluorescein images, OCT images and colour fundus images. Examples of the published datasets are: A multimodal dataset of colour fundus and OCT images for 50 patients, A dataset of OCT images for 15 patients, a dataset of fluorescence angiogram images for 70 patients with diabetic retinopathy, a dataset of 33 colour fundus

images with signs of exudates and a dataset of 25 colour fundus images with both healthy eyes and images with diabetic retinopathy signs.

An important observation regarding the aforementioned datasets is that none of these datasets are suited for multi-image analysis evaluation. For example, none of these public datasets contain special pairs of colour fundus images. This factor has formed a bottleneck for the research community to produce and evaluate algorithms that make use of spatial information, for example. Another observation is that the evaluation goals of multiple datasets differ. For example, in the MESSIDOR dataset the goal is to detect the Diabetic Retinopathy grade of the patient. In contrast, the DRIVE dataset is more suited for vessel segmentation evaluation algorithms. In the context of microaneurysm detection, algorithms have used several metrics and methods for evaluating their algorithms. Each of these metrics has both advantages and disadvantages, and they will be discussed in the next section.

Generally in terms of the types of groundtruths they can be divided into three types: pixel groundtruths, object groundtruths and grade-based groundtruths. Pixel-based groundtruths are very difficult and time-consuming to generate, but they provide the most accurate representation of the objects being labelled. Object based labels approximate the shape of the underlying objects based on assumptions. These are useful for microaneurysms that are approximately circular in most cases. Finally, grade-based labels are used during the development of expert-based systems, where the objective is to analyse an image and predict the grade and severity of disease, just as how a doctor would perform. Since microaneurysms are one of the earliest signs of diabetic retinopathy, they can be used to evaluate whether or not DR exists for the patients. In this way, a microaneurysm detection

algorithm can be evaluated using the grades based on the accuracy of its detection of diseased versus healthy patients for each image.

Table 2.1 Summary of publicly available retinal image datasets.

Dataset name	Number of images	Image Resolutions	Groundtruth type
DIARETDB1.1v2	89	1500 x 1152	MA Pixel labels
MESSIDOR	1200	1440 x 960 2240 x 1488 2403 x 1536	DR grades
ROC	100	768×576 1058×1061 1389×1383	MA Object labels
CHASE	14	1280 x 960	Vessel pixel labels
DRIVE	40	650 x500	Vessel pixel labels
ARIA	110	768 x 576	Vessel pixel labels

## 2.2 Microaneurysm Detection

Microaneurysm (MA) detection algorithms that have been proposed in the literature consist of three main stages: 1) preprocessing 2) MA candidates detection and 3) MA candidates classification. *Preprocessing* involves correction for non-uniform illumination in retinal images and enhances the contrast of MA candidates in the image. The goal of the *MA candidate detection* is to detect an initial set of candidate regions where MAs are likely to exist. *MA candidate classification* uses machine learning techniques to improve the sensitivity of the algorithm. This is done by removing false detections that have been identified by the algorithm in the previous stage. Moreover, there are some proposed methods that are unsupervised – which means that they do not require a classification stage [49], [55]–[60]. A summary of MA candidate detection algorithms presented in the literature

is listed in Table 2.2. For each algorithm, the table describes image type, the initial candidates method, the classifier used, and the reported performance for each classifier. There is discrepancy in the ways of evaluation used by the proposed methods. Furthermore, a diverse range of datasets have been used to perform evaluation. These reasons make it difficult to accurately compare between different techniques unless they were implemented and directly compared by applying them on the same dataset.

Spencer proposed one of the earliest techniques for MA detection from fluorescein angiograms [61]. The initial set of candidates was detected using a Gaussian matched filter. A template matched filter that matches the shape of an object detects it with optimum signal-to-noise ratio provided that only white noise exists[62]. A Gaussian template is a good approximation for the shape of MA candidates, that are approximately circular. Furthermore, since the preprocessing stages aim to eliminate non-uniform illumination by estimating the background and remove vessels whose shapes match that of a Gaussian, it is reasonable to use a Gaussian filter for noise removal since a lot of the noise has been removed during spencer's preprocessing stage. Each initial candidate that was detected is classified as either a true candidate or a spurious one using a rule-based classifier, and this produces the final set of candidates. Cree [63] applied Spencer's technique to a longitudinal set of florescence images in order to detect the 'MA turnover' – the progression of MA objects over time.

Recent techniques have addressed the problem of MA detection from colour fundus images. This is mainly due to the fact that colour images are more common in screening programs [19]. Furthermore, capturing fundus images is non-invasive and only requires pupil dilation.

A comprehensive survey of literature that has dealt with MA detection in colour fundus images is presented in the following paragraphs.

Spencer has been the basis for many of the proposed methods (applying morphological vessel removal followed by a Gaussian matched filter). A modification of Cree [63] was performed by Hipwell [64] in order to apply the algorithm to colour fundus images. Streeter's method [65] adapts the method proposed by Cree [63], however, during the classification phase, 80 features are extracted and Linear Discriminant Analysis (LDA) was used to perform the classification. Feature Selection was performed to filter the features down to 16 features. *Feature Selection* is a process to reduce redundant features in order to reduce computational time and decrease chances of overfitting. Fleming [66] is another Spencer-based approach. This technique introduced a novel region-growing step based on gradient values, rather than a simple threshold. In addition a paraboloid was fitted to each candidate using a parameter optimization process. The paraboloid parameters are used to compute many of the features used in the candidate classification phase. Zhang [67] replaced Fleming's Gaussian matched filter with multiple Gaussian filters at multiple scales and computed the maximum response to produce a probability map of the likelihood of presence of MA candidates. This probability map was then thresholded to produce the initial set of MA candidates. Finally a rule-based classifier using 30 features was used to perform the final classification. Li [58] introduced an unsupervised method based on a Multi-orientation Sum of Matched Filter (MSMF). This filter is a modification of the classical Gaussian Matched filter. This modified filter is anisotropic in nature and is applied in multiple directions. Due to this reason, this filter is better at suppressing responses to blood

vessels than the Gaussian Matched filter. Finally, Wu [68] modified the MSMF filter to take into account the varying size of MAs.

Sánchez used a mixture model-based clustering technique to detect the initial MA candidate regions [60]. The technique fits three normal distribution histograms to the retinal image histogram. The histograms should correspond to foreground pixels, background pixels and outliers. The foreground histogram pixels are considered to be the initial set of MA candidate regions. In the final classification step, logistic regression was used to classify each MA region as belonging to either a foreground or background region. Quéllec [69] based his technique on wavelet transforms applied in different sub-bands of the colour image.

A double-ring filter was used in Mizutani [48] to detect the initial candidates. MAs are dark circular regions within a brighter region to detect the MA candidates. Convolution filter was based on this property of MAs. It consists of an inner ring and an outer ring. A given pixel is considered to be a MA pixel if the average intensity of the inner ring is smaller than the average intensity of the outer ring. After the initial candidates are detected, classification is performed using 12 extracted features and an Artificial Neural Network (ANN).

A simple thresholding technique was used to detect the Initial set of candidates in Giancardo [49], [59]. A novel Radon-based transform was used to extract the features of the initial candidates and a Support Vector Machine (SVM) classifier was used to perform the final classification. An initial set of 31 features were computed for classification. The dimensionality of the features was reduced to 10 dimensions using Principal Component Analysis (PCA), and this reduced representation was used to perform the classification. A reduced dimension for the features reduces the risk of overfitting and also makes the classification more computationally efficient.

A 'moat operator' was used by Sinthanayothin [70] in order to enhance red lesions in the image and then these regions were segmented. Vessel regions were then removed to produce the final set of candidates. This method detected both MAs and haemorrhages. Unfortunately the moat operator was not defined in the paper and the exact definition was not found in the literature.

AbdelAzeem [71] used a Hessian matrix in order to detect the initial MA candidate set. A rule based classifier was used to classify the MA detections. The classification is based on a rule referred to as candidate 'energy'. The exact definition of the computed 'energy' was not mentioned in the paper, however, it is likely to be the same definition as in Fleming [66]. A Hessian matrix was used by Inoue [72] in order to detect the initial candidates and an Artificial Neural Network (ANN) was used to classify the features. A group of 126 features were fed into the ANN for classification. However this group of features was reduced using Principal Component Analysis (PCA) in order to reduce computational complexity and avoid overfitting. Moreover, Srivastava [73] used the eigenvalues of the hessian matrix in order to detect the initial candidates. Recently, Adal [74] has used a hessian matrix in order to detect the initial set of MA candidates. A combination of SURF, Radon and scale-space features were extracted from the initial candidates. Multiple classifiers (Support Vector Machines, K-Nearest-Neighbours, Naive Bayes and Random Forest) were also experimented with in this technique.

An adaptation of Spencer [61] is presented in Niemeijer [74]. The two main contributions are: A pixel based classification system for the initial candidate detection phase and an extended set of features used for pixel classification.

Table 2.2. Summary of MA detection algorithms in the literature. The performance superscripts are defined as follows: <sup>a)</sup> Lesion-based measure <sup>b)</sup> Image-based measure <sup>c)</sup> Pixel-based measure. Key: AUC – Area Under the Curve, FP/image - False positives per image, PPV – Positive Predictive Value. Table from [31].

Paper	Image Type	Initial candidates method	Classifier used	Reported Performance	
				Dataset	Performance
Spencer, 1995 [61]	Florescence	Gaussian Filter	Rule-based	Private dataset (4 images)	Sensitivity <sup>a</sup> : 0.25 FP/image <sup>a</sup> : 1.0
Cree, 1997 [63]	Florescence	Gaussian Filter	Rule-based	Private dataset (20 images)	Sensitivity <sup>a</sup> : 0.6 FP/image <sup>a</sup> : 1.0
Hipwell, 2000 [64]	Colour	Basic Thresholding	Rule-based	Private dataset (3783 images)	Sensitivity <sup>a</sup> : 0.6 FP/image <sup>a</sup> : 1.0
Sinthanayothin, 2002 [70]	Colour	Moat operator	N/A	Private dataset (14 images)	Sensitivity <sup>b</sup> : 0.885 Specificity <sup>b</sup> : 0.997
AbdelAzeem, 2002 [71]	Florescence	Hough transform	Rule-based	Private dataset (3 images)	Sensitivity <sup>a</sup> : 0.6 FP/image <sup>a</sup> : 17.67
Streeter, 2003 [65]	Colour	Gaussian filter	Linear Discriminant Analysis	Private dataset	Sensitivity <sup>a</sup> : 0.3 FP/image <sup>a</sup> : 1.0
Niemeijer, 2005 [21]	Colour	Gaussian Filter pixel classification	K-Nearest-Neighbours	Private dataset (100 images)	Sensitivity <sup>a</sup> : 0.83 FP/image <sup>a</sup> : 1.0 Sensitivity <sup>b</sup> : 1.0 Specificity <sup>b</sup> : 0.5
Fleming, 2006 [66]	Colour	Gaussian Filter	K-Nearest-Neighbours	Private dataset (1441 images)	Sensitivity <sup>a</sup> : 0.51 FP/image <sup>a</sup> : 1.0 Sensitivity <sup>b</sup> : 0.91 Specificity <sup>b</sup> : 0.5
Quellec, 2008 [69]	Colour	N/A	N/A	ROC dataset	Sensitivity <sup>c</sup> : 0.90 Specificity <sup>c</sup> : 0.898
Mizutani, 2009 [48]	Colour	double-ring filter	Neural network	ROC dataset	Sensitivity <sup>c</sup> : 0.15 PPV <sup>c</sup> : 1.0
Sánchez, 2009 [60]	Colour	Mixture model-based clustering	N/A	ROC dataset	ROC score: 0.332 Sensitivity <sup>a</sup> : 0.30 FP/image <sup>a</sup> : 1.0
Zhang, 2010 [67]	Colour	Multiscale Gaussian	Rule-based	ROC dataset	Sensitivity <sup>a</sup> : 0.11 FP/image <sup>a</sup> : 1.0 ROC: 0.201
Giancardo, 2010 [59]	Colour	Basic Thresholding	N/A	ROC dataset	Sensitivity <sup>a</sup> : 0.22 FP/image <sup>a</sup> : 1.0
Lazar, 2011 [55]	Colour	Local Maxima scanlines	N/A	ROC dataset	Sensitivity <sup>a</sup> : 0.38 FP/image <sup>a</sup> : 1.0 ROC score: 0.355

Sopharak, 2011	Colour	extended-minima	Naïve Bayes	Private dataset (45 images)	Sensitivity <sup>f</sup> : 0.816 Specificity <sup>f</sup> : 0.99
Giancardo, 2011 [49]	Colour	Basic Thresholding	N/A	ROC dataset	Sensitivity <sup>a</sup> : 0.43 FP/image <sup>a</sup> : 1.0 ROC: 0.375
Lazar, 2013 [56]	Colour	Local Maxima scanlines	N/A	ROC dataset	Sensitivity <sup>a</sup> : 0.41 FP/image <sup>a</sup> : 1.0 ROC: 0.423
Rocha, 2012 [76]	Colour	N/A	Support Vector Machine	DIARETDB1 v1	Sensitivity <sup>f</sup> : 0.91 Specificity <sup>f</sup> : 0.5
				MESSIDOR	Sensitivity <sup>f</sup> : 0.93 Specificity <sup>f</sup> : 0.5
Sopharak, 2013 [77]	Colour	extended-minima	Bayesian	Private dataset (80 images)	Sensitivity <sup>f</sup> : 0.86 Specificity <sup>f</sup> : 0.99
Akram 2013 [7]	Colour	Gabor filter	Hybrid classifier	DIARETDB0, DIARETDB1 v1	Sensitivity <sup>a</sup> : 0.99 Specificity <sup>a</sup> : 0.997 Accuracy <sup>a</sup> : 0.994
Li, 2013 [58]	Colour	Multi-orientation Gaussian (MSMF)	N/A	ROC dataset	Sensitivity <sup>a</sup> : 0.05 FP/image <sup>a</sup> : 1.0
Junior, 2013 [57]	Colour	Extended Minima	N/A	DIARETDB1v1	Sensitivity <sup>f</sup> : 0.87 Specificity <sup>f</sup> : 0.92
Inoue, 2013 [72]	Colour	Hessian Matrix Eigenvalues	Neural network	ROC dataset	Sensitivity <sup>a</sup> : 0.18 FP/image <sup>a</sup> : 1.0
Adal, 2014 [78]	Colour	Hessian Matrix Eigenvalues	Support Vector Machines, K-Nearest-Neighbours, Naïve Bayes, Random Forest	ROC dataset	ROC score: 0.363 Sensitivity <sup>a</sup> : 0.364 FP/image <sup>a</sup> : 1.0
Ram, 2015 [79]	Colour	Morphological reconstruction	K-Nearest-Neighbours	ROC dataset	Sensitivity <sup>a</sup> : 0.31 FP/image <sup>a</sup> : 1.0
				DIARETDB1 v1	Sensitivity <sup>a</sup> : 0.73 FP/image <sup>a</sup> : 1.0
				Private dataset	Sensitivity <sup>a</sup> : 0.18 FP/image <sup>a</sup> : 8.0
Wu, 2015 [68]	Colour	Multiscale Multi-orientation Gaussian (MMMMF)	Support Vector Machines, K-Nearest-Neighbours, Linear Discriminant Analysis	ROC dataset	Sensitivity <sup>a</sup> : 0.23 FP/image <sup>a</sup> : 1.0 Sensitivity <sup>f</sup> : 0.92 Specificity <sup>f</sup> : 0.50
Srivastava, 2015 [73]	Colour	Frangi-based filters	Support Vector Machines	MESSIDOR+ DIARETDB1 v1	Sensitivity <sup>f</sup> : 1.00 Specificity <sup>f</sup> : 0.50

Romero, 2015 [80]	Colour	Hit-or-miss transform	Neural networks	DIARETDB1v2.1	Sensitivity <sup>c</sup> : 0.93 Specificity <sup>c</sup> : 0.94
				ROC dataset	Sensitivity <sup>c</sup> : 0.88 Specificity <sup>c</sup> : 0.97
Haloi, 2015 [81]	Colour	N/A	Nearest-mean classifier	DIARETDB1v2.1	Sensitivity <sup>c</sup> : 0.88 Specificity <sup>c</sup> : 0.97
				ROC dataset	AUC <sup>c</sup> : 0.98

An unsupervised method was introduced in Lazar [55], [56] (does not require any training or classification steps). The reported results of this technique are comparable to other supervised methods, which make it a promising method. The essence of this technique is to discriminate between vessels and MAs by using a 1D scanline at different directions for each pixel. While a MA will have local minima in all directions of the rotated scanlines, a vessel will have only one minimum corresponding to when the scanline is perpendicular to the vessel. Hence, using this property, a probability map is produced at each pixel and then simple thresholding is applied to produce the final set of candidates.

Garcia [82] compared the accuracy of four neural network variants: Multilayer Perceptron (MP), Radial Basis Function (RBF), Support Vector Machine (SVM) and Majority Voting (MV). The initial candidates were detected using a local thresholding technique based on the mean pixels of the entire image compared to mean intensity in a small window around a pixel. According to their experiments, the RBF was suggested as the preferred classifier among all 4. An interesting approach that relies on visual dictionaries was presented in Rocha [76]. The use of visual dictionaries (bag of words) makes this approach more generisable since it does not rely on specific features during the classification. Therefore, the same approach can be used to perform detection of lesions other than MAs as well. The disadvantage of this is that it requires a larger training set. Haloi [81] recently applied deep

neural networks to detect MAs in colour images. Deep neural networks have gained popularity in the field of computer vision in the recent years since they do not require manual feature engineering (selection of features). Moreover, algorithms based on deep learning have produced results that out-perform other state-of-the-art algorithms in other computer vision applications. However, deep learning requires massive datasets for training [83] and such large labelled datasets are not yet available for retinal images.

Ram [79] used a dual classifier in order to classify the initial candidates. The initial candidates were detected using a simple thresholding operation after preprocessing. Two classification stages are then applied. The first classification stage was applied in order to separate MAs from vessels. The features used for this purpose are a second derivative Gaussian at multiple orientations, difference of Gaussians and inverted Gaussians. The second classification stage was applied in order to further separate MAs from other types of noise. for the MA classification on a pixel level rather than at a candidate level. This means that each pixel gets classified as either an MA or not, rather than each initial candidate as a whole. After preprocessing, the extended-minima transform is used to detect the initial candidates, and a Bayesian classifier was used to perform the pixel-based MA classification. Similarly, Junior [57] presents the same technique as Sopharak [77], but does not apply a classification stage. Akram [7] used a Gabor filter for the detection of the initial candidates. The Gabor filter is applied at multiple scales and rotated at various angles, and the maximum response is computed. This causes a large response for vessels, microaneurysms and haemorrhages. Vessel segments are removed using a vessel segmentation technique. A hybrid classifier is used to reduce the false positives in the initial candidates. The technique

has reported a lesion-based specificity and accuracy measure, even though it is not possible to measure the number of true negatives at the lesion level [61].

In conclusion, the majority of the proposed algorithms consist broadly of three stages: preprocessing, candidates detection and classification. The authors suggest that the bottleneck stage is the candidates detection – since if a candidate was not detected in this stage, it will not be detected during the classification phase (the purpose of classification is to decrease the amount of false positives while maintaining the true candidates present in the set).

A large portion of the proposed methods, including Spencer, use a gaussian matched filter or a variation of this filter. Many of the modifications to the Gaussian matched filter aim to reduce the amounts of false positives. The author argues that while attempting to reduce the number of false positives there is a risk of also losing some of the true positives during the process. Furthermore, as discussed earlier, the use of a matched filter that matches the shape (intensity profile) of the underlying candidate maximises the signal-to-noise ratio in the presence of white noise. For these reasons, the authors recommend the use of a Gaussian filter for enhancement, right after vessel removal. Alternatives to the Gaussian filter include scanlines [55] and the radon transforms [49].

In terms of the classifier use, a variety of classifiers have been used in literature. The choice of classifier is affected by contextual information such as the number of features used and the amount of training data present. Desired performance of the algorithm is also a factor to take into consideration; some classifiers are known to be slow to train and classify and hence would not be an appropriate choice for a real-time algorithm.

## 2.3 Image Registration

This section will review some of the work that has been done on retinal image registration. Work done on multi-image analysis will also be reviewed. During the literature review, the search was made for registration methods that are applied to retinal images. However, some of the relevant general registration techniques have also been included in this review.

Image Registration techniques can be broadly categorised into two main types: area-based techniques and feature-based techniques. Area-based approaches try to match each pixel from one image to its corresponding location in the base image location. They are based on pixel intensities and certain optimized functions, such as least mean square error, cross correlation, phase correlation or mutual information[84], [84]–[89]. Area-based methods are often applied to multi-modal retinal images. This is because their similarity functions are usually based on the image intensities which makes them applicable across multiple modalities. Feature-based approaches assume an underlying model of transformation and try to estimate the parameters in order to perform the model transformation. Factors such as non-uniform illumination and large homogenous areas in retinal image may degrade the performance of area-based methods. Feature-based methods are recommended if the images to be registered have distinguishable features while area based methods are recommended in images that are not so rich in particular detail[90]. Since retinal images contain detail in the form of the vessel tree structure, the optic disk, and the fovea, our focus in the literature search was on feature-based methods.

There are several clinical applications of retinal image registration. These include real-time tracking for treatment of blindness-causing conditions [91], change detection to measure the progress of a disease [92] or the impact of treatments [93], and multimodal integration

to aid in diagnosis and surgical planning[94].” [95]. Registration applications can be broadly categorised as follows [85]:

1. **Multimodal registration:** registration of an overlapping pair of images of the retina that were captured using different sensors (example: Colour/OCT or Colour/Fundus image pairs).
2. **Viewpoint registration (spatial registration):** Registration of two pairs of images captured using the same sensor but with different viewpoints (example: Colour Fundus – Optic Disc-centred and macula-centred).
3. **Temporal registration:** Registration of two pairs of images that have been captured using the same sensor and of the same patient but at different time intervals.
4. **Template registration:** Registration of an image to a reference template which is usually used as an “atlas” reference. An example of this is Retinal Atlas images [15], [96], [97].

The vascular tree structure is an appropriate representation of the retina for the purpose of image registration. This is due to three reasons: 1) It extends throughout the whole retina; 2) It does not move from its position, except in very few diseases; and 3) Landmark points can be localised within the vessel tree structure [98].

Table 2.3 summarises the retinal image registration techniques that have been proposed in the literature. For each proposed technique, the table also lists the types of images that the method is intended for (colour or fluorescence), the type of registration technique (feature based or similarity based), the optimization strategy used (usually for similarity-based techniques), the kind of landmarks used (for feature-based methods) and the transformation model.

Early work on retinal image registration [99], [84], [85], [100],[86],[101]has used a lower-order transformation model (similarity or affine model) to perform the registration. Since the eye is approximately spherical in shape, a higher order quadratic model is more appropriate to estimate the distortion caused during the process of retinal imaging (intuitively) [102]. However, since the retina is almost planar on small fields of view (such as fundus photographs, where the field of view captured is around 45 degrees of the retina), these registration methods produce modest errors in such scenarios [102]. Majority of the proposed feature-based methods have relied on vessel bifurcation and crossover points (Table 2.3).

Table 2.3 Image registration algorithms in the literature.

Method	Applied to	Registration Type	Optimization method	Similarity Measure	Landmarks	Transformation Model
Hart, 1994 [99]	Colour	Feature	N/A	N/A	Bifurcation points	Affine
Ritter, 1999 [85]	Stereo, temporal	Area	Simulated annealing	Mutual information		
Zana, 1999 [100]	Fluorescence, colour (multimodal)	Feature	N/A	Percentage vessel overlap	Bifurcation points	Affine
Lloret, 2000 [86]	Stereo, Fluorescence	Area	Simplex Algorithm	vessel structure	N/A	Similarity
Can, 2002 [102]	Colour	Area	N/A	Similarity-weighted histogram	Vessel centreline	Quadratic
Can, 2002 [103] (mosaic)	Colour	Feature	N/A	Similarity-weighted histogram	Bifurcation points	Quadratic
Heneghan, 2002 [101]	Colour	Feature	N/A	N/A	Bifurcation & crossover points	Affine
Tsai, 2002 [95]	Colour	Feature	N/A	N/A	Bifurcation points	Quadratic
Zhang, 2002 [104]	Fluorescence, colour	Feature	N/A	N/A	Bifurcation & crossover points	Affine

Shen, 2003 [105] (online, hierarchal)	Colour	Feature	N/A	N/A	Bifurcation points	Quadratic
Stewart, 2003 [106]	Colour	Feature	N/A	N/A	Bifurcation points	Quadratic
Chanwimaluang, 2006 [98]	Colour	Hybrid	Multi-resolution search	Vascular tree entropy correlation coefficient	Bifurcation points	Affine/ Quadratic
Ghassabi, 2013 [107]	Multi-modal	Feature	N/A	N/A	UR-SIFT	Quadratic
Reel, 2014 [108]	Colour	Area	N/A	Expectation Maximisation for Principle Component Analysis with Mutual Information	N/A	Similarity
Wong, 2016 [109]	Fluorescence, Colour (multi-modal)	Feature	N/A	N/A	low-dimensional step pattern analysis (corner points)	Affine
Xian, 2016 [110]	Fundus	Hybrid	Brute force	Mutual Information	Optic disc centre	Similarity
Guo, 2017 [111] (inaccessible)	-	-	-	-	SIFT	-
Matas, 2017 [112]	Colour	Feature	Particle Swarm Optimisation	N/A	SIFT	Similarity

Ritter [85] uses an area-based method in order to register a pair of images. A simulated annealing optimisation function is used to minimise a mutual information similarity function. The method has been applied to both stereo and temporal image pairs.

Zana[100] used a global affine transformation to align a pair of temporal images. The method is also applied to align a multimodal pair of images. Vessel bifurcation points are used as features and Hough transform is used to find feature correspondences. Multiple transformation likelihoods are computed based on the percentage of the overlaid vessels

that are superimposed after alignment. The transformation that achieves the highest percentage overlay is selected to produce the final alignment.

Lloret,2000 [86] have developed an algorithm that focuses on the registration of multimodal retinal images (fluorescence angiograms and colour images). The method uses an area-based method to perform the alignment of images. The authors argue against the use of vessel bifurcation points as landmarks. Instead they have used a measure of cross-correlation between vessel maps in order to perform the alignment of images.

Can [102] estimated a second-order polynomial transformation in order to model the distortion caused by the retinal imaging process as a quadratic distortion. The bifurcation points of the blood vessels are selected as control points in both images. Since a quadratic transformation needs to be estimated, there are 12 unknown parameters that need to be estimated. The estimation is done by a three-stage hierarchical approach. The three stages estimate using a translation, similarity and quadratic transformation respectively. An issue that arises with this approach is that it may cause distortion in non-overlapping regions (since one image is warped on to the anchor image) [96]. Due to this, the fused image tends to be more accurate in the overlapping regions than the non-overlapping regions [113].The method is extended to be applied to a mosaic of images in [103]. A mosaic is formed by registering 2 or more images collectively.

Tsai, 2002 [95] estimates a higher order model based on a small number of landmarks. The proposed algorithm is conceptually similar to the dual-bootstrap algorithm [106], [114]. It starts with a lower-order model for an initial alignment. The initial alignment is accurate in a small region of the retina, and using an iterative procedure this region is expanded to cover a higher order model of the entire retina. The initial model is based on a translation model

alone. Landmarks are based on vessel bifurcation points and are matched together using an 8-dimensional feature vector. Refinement to a higher order quadratic model happens using an iterative procedure from an initial small region.

Zhang, 2002 [104] introduced a feature-based automated registration method to align a colour (red-free) image with a fluorescence image. The method relies on the vessel bifurcation and crossover points as features. The main challenge in this case is the feature matching process since the images belong to multiple modalities. A Singular Value Decomposition (SVD) based algorithm is used for the alignment of both images using the reduced feature points in both the reference and moving image.

Shen, 2003 [105] has applied retinal image registration in the context of identifying the spatial location of a surgical tool such as a laser pointed at the human retina. The problem is treated as a registration problem – building a spatial map of the retina using diagnostic images and then registering each image against this map. Initial landmarks are found using vessel bifurcation points. The orientation of each bifurcation point (known as quasi-invariant feature vectors) are used in order to find feature correspondences. A quadratic transformation is used to align the initial set of diagnosis images to which the remaining images are aligned during the surgery. A hierarchical approach is used for starting from an initial similarity model and then an affine and finally a quadratic model for the alignment of the images in the mosaic.

The dual bootstrap registration algorithm is applied to retinal images in Stewart [106]. The algorithm uses vessel bifurcation points as landmarks. In dual-bootstrap, an initial lower-order global transformation is found as an initial alignment for the image pair. One or more initial regions (known as bootstrap regions) based on where the lower order transformation

is most accurate. An iterative process follows where a higher order model is selected (if it refines the transformation estimate) followed by an expansion of the bootstrap region to a larger area.

Chanwimaluang [98] proposed a hybrid registration technique that relies on 4 stages: 1) extracting the vascular tree, 2) Initial alignment using a lower order model (translation) 3) Assessment of quality assessment based on ETRDS field definition[115] and 4) Landmark extraction and higher order (affine/quadratic) alignment. During the initial alignment phase, a function of the vessel masks is used as a similarity measure for the initial alignment. This initial alignment is an area-based method where the optimization function is to minimize two values from the vessel masks: Entropy Correlation Coefficient and Binary Local Entropy. An additional quality assessment step exists to check whether the registration quality was sufficient enough for the zero-order model to determine whether a higher-order model should be applied. The higher order model is either an affine or quadratic model, and uses the Iterative Closest Point (ICP) algorithm. The feature points used in this case are vessel bifurcation and crossover points extracted from the vessel tree. If there are no sufficient features, a gridline is used to extract more sample feature points from the vessel centrelines. The advantage of this method is the flexibility of its application, since the initial alignment does not require a specific minimum number of feature points to be detected.

A comparative survey of multiple global transformation models is presented in [116]. Vessel bifurcation points are investigated as the landmarks of the information.

A quadratic model is used by [117] to perform registration. Therefore the transformation model is identical to [102]. The main difference between both these approaches lies in the control point selection method. Unlike [102], the employed method does not rely on

bifurcation points. This is useful in the cases when the vascular tree structure is distorted, such as when there is tortuosity or bleeding. However, according to [118], the technique does not work well with multi-modal images.

Lee [119] attempts to resolve the issue of image distortion caused by the algorithm in [102]. This is done by applying an initial radial distortion correction that corrects the distortion due to sphere-to-plane projection. After this correction is applied, it is possible to use an affine model after the radial distortion has been corrected. Experimentally, this method exhibits superior performance to the quadratic model, especially in non-overlapping regions[113].

Recent methods have made use of other feature descriptors such as SIFT [111], [112] or uniform robust SIFT (UR-SIFT) [107]. The advantages of these feature descriptors is that multiple points can be found even when there is a small overlap between the images. Furthermore, they can also be utilised to discover features from multi-modal images[107].

In this section, a comprehensive literature review of retinal images has been presented. Image registration is a crucial step for the development of a segmentation algorithm that relies on multiple images. The importance of image registration will be evident during the incorporation of spatial (Chapter 4) and temporal (Chapter 5) information.

In conclusion, a variety of image registration algorithms have been proposed in literature. For most tasks, a lower order affine or similarity model as long as the fundus image is captured from a small field of view. As long as the overlap between the images is not very large, it is possible to use a feature-based approach and extract landmarks from the vessel tree structure, while establishing correspondences using some sort of similarity measure based on the vessel centreline[120]. Finally, for to achieve accurate registration results in localised regions, the algorithm proposed by Stewart [106] is known to produce good

results. It is based on the Iterative closest point (ICP) method and employs an iterative process which starts with a lower-order model and iteratively moves to higher order models in local regions if possible.

## **2.4 Multi-image Analysis Algorithms**

Multi-image analysis is defined as the use of information from more than one image in order to perform image analysis. In this section, two approaches for multiple image analysis are discussed: Longitudinal change analysis and retinal atlases.

### **2.4.1 Longitudinal Change Analysis**

Longitudinal change analysis detects changes that have occurred to a single patient over time. This is useful to detect the progression of a disease or to help increase the accuracy of a segmentation algorithm. Interestingly, there has not been a large amount of work performed on longitudinal change analysis on retinal images.

One reason may be the lack of publicly available longitudinal datasets. None of the datasets mentioned in the review paper by Fraz [121] are longitudinal datasets. Without the existence of any datasets to work on it becomes impossible to validate research on the subject. The second reason may be due to the multi-disciplinary nature of the task. The task involves areas such as Image Registration, Change Analysis (which requires input of clinical expertise) and Machine Learning, each of which is a research area in its own right. This makes it difficult to perform research in longitudinal change analysis since expertise in multiple fields is required.

McRitchie [122] introduces a technique to visualize change in retinal images to make it easier to spot when something has changed over a period of time. Each image pair is

registered using a global affine transformation model technique and using mutual information as a similarity measure. After the images are aligned, the older image is then subtracted from the more recent image pair in order to highlight the regions that have changed over time. If a region gets brighter with time, it will be highlighted as a bright region. If a region gets darker with time it will be highlighted as a dark region. Regions that have not changed will remain grey. The resulting difference image will help clinicians identify regions of change. The main issue with this technique is that it is sensitive to the accuracy of the registration that was performed.

Narasimha [123] used longitudinal change analysis to detect non-vascular anomalies such as exudates and microaneurysms. After preprocessing, a Bayesian classifier is used to detect changes in image colour. A “redness” increase indicates the appearance of microaneurysms. Similarly, an increase in white or yellow indicates the appearance of exudates.

#### **2.4.2 Retinal Atlases**

Retinal atlases describe a trend of a collection of retinal datasets. Atlases are usually formed from a dataset of images that are related to each other in some way (for example, a dataset of images from a group of people with healthy retinas). Retinal atlases have only been studied quite recently and have a wide range of potential applications such as segmentation, registration and to categorise anatomies in a dataset.

Lee [96] introduces the concept of retinal atlases and how they can be generated. The methodology followed is to register a large number of retinal images together. The RADIC model [119] is used to perform the registration. The features used for the registration are the centre of the optic disc, centre of the fovea and the main blood vessel arch. Following this registration, any new image may then be aligned to this and its deviation from

normality" can be measured. There can be several methods for this deviation measure, depending on the target application.

Ali [97] uses retinal atlases to perform exudate segmentation. Since exudates appear as bright white or yellow spots on a retina surface, they are considered as outlier regions. Therefore, if a retinal atlas represents the average intensities of a dataset, exudates are detected by producing an intensity difference map between the atlas and a target image. Ideally, exudates will appear as the brightest spots in the difference map. Hence, a threshold can be used to easily detect these exudates.

### **2.4.3 Other Multi-image Algorithms**

Other related Multi-image analysis techniques have been explored in the literature and these are mentioned here for the sake of completeness. A few methods have addressed the problem of detecting or measuring change from multiple images for the purpose of disease identification. In Connor [25], vessel segmentation was performed on a series of fundus images. Vessel tortuosity and width were both measured. This was done in order to find a correlation between these measures and some signs including Diabetic Retinopathy. Arpenik [124] used fractal for the purpose of distinguishing between normal and abnormal vascular structures in a human retina. Patterson [125] developed a statistical approach in order to quantify change in the optic nerve head topography using a Heidelberg Retinal Tomograph (HRT). This purpose of this was to measure the disease progression in glaucoma patients. Artes [126] investigated the temporal relationship between visual field and optic disc changes in glaucoma patients. Bursell [127] reported on the difference in blood flow changes between insulin-dependent diabetes mellitus (IDDM) patients compared to healthy patients in video fluorescein angiography.

While the problem of analysing “change” and “progression” of disease has been studied in the literature, to the best of our knowledge, the combination of a spatial pair of retinal images for the improvement of detection of MAs has not yet been explored.

The literature discussed in this section indicates that there is limited exploration on analysing multiple images. However, analysis of multiple images has the potential to improve current techniques since, intuitively, having more information has potential to improve the accuracy of existing algorithms. Addressing this shortcoming in this area was the main motivation behind this work.

## **2.5 Conclusion**

In this section a detailed account of the relevant literature for microaneurysm detection, image registration and multi-image analysis has been presented in depth. Equipped with this information, the next chapters will go through the experiments that have been performed and the results of these experiments. Chapter 3 will discuss the experiments performed to assess existing microaneurysm detection algorithms, and a proposed improvement of an existing algorithm. Chapters 4 and 5 will discuss experiments that were performed to combine information from multiple images in spatial and longitudinal dimensions, respectively.

# 3

## **MICROANEURYSM DETECTION USING AN ENSEMBLE CLASSIFIER**

---

The understanding of existing microaneurysm detection work that operates on single images is a prerequisite to build more sophisticated techniques that rely on multiple images. This is because understanding the limitations of the existing methods will suggest ways to improve them after information is added from multiple images. For instance, if there is a common cause for false positives in a certain algorithm, how can this be eliminated after combining information from multiple images? Similarly, if there is a true negative that exists in one of the images, how can more information be utilised to increase our confidence in this particular true negative? Another way to think about this is to find out how to combine information in a way that will increase the number of true positives and decrease false detections.

In this chapter, an algorithm that detects microaneurysms from colour fundus images will be proposed. The motivation for the development of this algorithm has been to use a tree ensemble classifier during the classification stage, which has been very effective in other

segmentation applications [5], [128]. A large number of features are extracted for each microaneurysm candidate for the purpose of classification. These features will be analysed to understand which features are more distinctive for classifying microaneurysms. Since the ensemble classifier will output the likelihood score for each candidate in the form of a probability, this will be a useful feature while incorporating both spatial and temporal information (as will be seen in Chapter 4 and Chapter 5). The algorithm will be evaluated on multiple datasets and compared with other state-of-the-art techniques. The work of this chapter has links to publications resulting from the work described that have explained the proposed detection method and presented its evaluation results [129], [130].

### **3.1 Methodology**

The proposed method is based on the method suggested by Fleming [66]. The main modifications that were made to Fleming's algorithm will be stated throughout this chapter. This work is an extension of the algorithm published in [130] and includes a more extensive evaluation as well as detailed feature analysis. The proposed methodology consists of three phases: 1) preprocessing 2) MA Candidate Detection and 3) MA Candidate Classification. During the preprocessing stage, non-uniform illumination is removed from the image using background subtraction. Noise removal is also performed during this stage. In the MA Candidate Detection phase an initial set of MA candidates are detected. Ideally all the candidates in the image should be detected with as few false positives as possible. Most of these false positives should then be removed during the Candidate Classification phase. The three stages of the proposed algorithm will be explained in the following sections.

Despite being published in 2006, Fleming's reported per-lesion performance on a large private dataset is comparable to recently published methods. This makes it reasonable to

use Fleming as a baseline for comparison with the proposed technique. These methods include Wu (2015) [81], Adal (2014) [80], Inoue (2013) [73] and Li (2013) [68]. This is also illustrated in the literature review chapter (Table 2.2).

### 3.1.1 Preprocessing

The preprocessing steps are as follows: Given a colour retinal image (Figure 3.1a) the green channel is extracted (Figure 3.1b) since MA candidates appear with high contrast in this channel. Salt-and-pepper noise is removed by applying a 3x3 median filter. Contrast-Limited Adaptive Histogram Equalisation (CLAHE) [131] is applied in order to improve the contrast in the image. Further noise removal is performed by applying a 3x3 Gaussian filter to the image. Let the result of the previous operations be  $I_{adapt}$ . Shade correction ( $I_{shade}$ ) is performed by dividing the image by an estimate of the background:

$$I_{shade} = I_{adapt}/I_{bg} \quad (3.1)$$

Where  $I_{bg}$  is the background estimate calculated by applying a 68x68 median filter to  $I_{adapt}$ . The filter size is chosen to be large enough in order to eliminate vessels and other features in the image. This parameter is set in accordance with the field of view and the size of the large structures in the image such as the main blood vessels. Finally, global contrast normalization is performed on the resulting image by dividing it by its standard deviation:

$$I_{con} = \frac{I_{shade}}{std(I_{shade})} \quad (3.2)$$

Where  $std(I_{shade})$  represents the standard deviation of the shade corrected image. The result of these operations is illustrated in Figure 3.1(c). Following these operations it is necessary to detect an initial set of MA candidates from the preprocessed image. This is described in the following section.

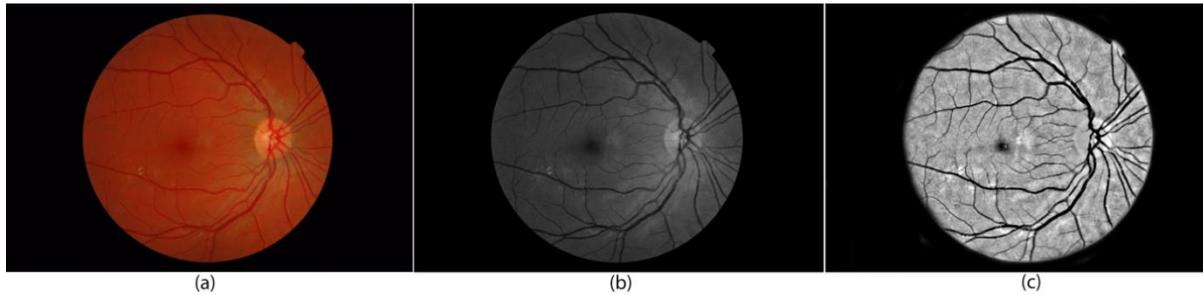


Figure 3.1 An example of the preprocessing stage. a) The colour image, b) the green channel image, c) the preprocessed image.

### 3.1.2 MA Candidate Detection

After performing noise removal and shade correction, an initial set of MA candidates can be detected. The method used is based on that proposed by Fleming [66]. A Gaussian matched filter ( $\sigma = 1$ ) is used in order to enhance circular dark regions in the image. Empirically it was found a sigma value of 1 produces high responses for even the smaller candidates – and also for noise. However, the filtering of false detections is left for the classification phase.

As discussed earlier, the shape of microaneurysms can be assumed to be circular. Furthermore, in colour images, the intensity profile of an MA candidate fades from a darkest point in the centre and gets brighter as it fades to the background. This intensity profile resembles the shape of a Gaussian bell curve (Figure 3.2). It has been previously stated in Section 2.1.2 that a template matched filter that matches the shape of an object detects it with optimum signal-to-noise ratio provided that only white noise exists [62]. Since a lot of preprocessing steps have been performed in the previous section to remove non-uniform illumination and enhance candidate contrast, we assume that this is the only noise present in the preprocessed image.

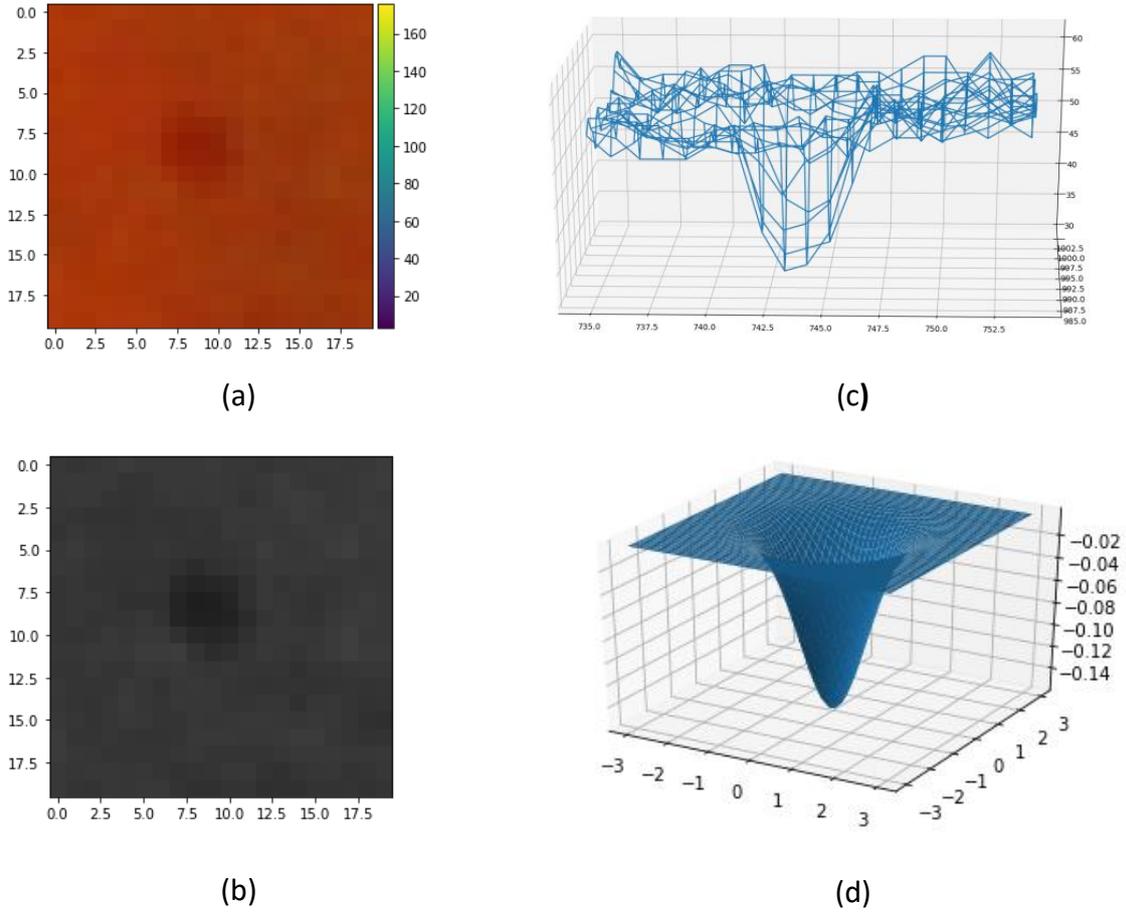


Figure 3.2A microaneurysm's intensity profile resembles that of a 2D Gaussian function. a) A microaneurysm colour image, b) The green channel of (a). c) The intensity profile of (b), d) A Gaussian intensity profile ( $\sigma = 1$ )

Since blood vessel cross-sections have intensity profiles similar to MAs, they need to be removed before applying the Gaussian matched filter. The following morphological operations are applied for vessel removal.

A closing operation is applied using a linear structuring element in multiple directions. The minimum of the application of closing operation at multiple operations was then subtracted from the shade corrected image [66].

$$I_{bothat} = \min_{i=0..7} \left( I_{shade} \circ \text{strel} \left( \frac{\pi i}{8}, n \right) \right) - I_{shade} \quad (3.3)$$

Where  $strel(x,n)$  represents a linear structuring element at an angle of  $x$  degrees and of length  $n$ . The size of the structuring element should be chosen to be larger than the largest vessel in the images (in our case a size of 11 pixels was selected through direct measurement in the images). This parameter was adjusted in accordance with to the average widths of the vessels – too large may miss some of the tortuous vessels and too small may not cause the larger vessels to be eliminated. This operation causes vessels to be removed from the image while retaining circular objects which resemble the shape of MAs (Figure 3.1 (a) and Figure 3.1(b)).

A Gaussian matched filter ( $\sigma = 1.0$ ) is then convolved with  $I_{bothat}$  in order to enhance circular dark regions  $I_{gauss} = I_{bothat} * gauss(1.0)$  (Figure 3.1(c)). The resulting response probability is then thresholded as follows in Equation (3.4):

$$I_{thresh} = thresh(I_{gauss}, 5\tau) \quad (3.4)$$

The value of  $\tau$  is chosen to be the threshold value at which the top 5 percent of pixels are selected[66]. A region growing operation based on Fleming [66] is performed in order to enhance the shapes of the detected MA candidates. The set of initial candidates are used as input. The procedure involves iteratively growing along the 8-connected pixels from the minimum intensity pixel of the candidate until a stopping condition. In our case, the stopping condition is when a maximum point of the “energy function” is reached. The energy function is defined as the average value of the gradients around the boundary of the grown region. All the parameters at this stage have been kept the same (Gaussian filter size and median filter size and  $\sigma$ ) except the maximum grown size.

Through our experiments it was found that the maximum grown size of 3000 pixels suggested by Fleming resulted in large blood vessel regions being falsely identified. It was

empirically found that a value of 100 pixels was a more suitable value for the maximum area and this parameter modification decreased the number of false positives appreciably, while achieving almost the same sensitivity. The value was chosen to be over twice the size of the average MA size in the groundtruth images. More examples of the region growing process are shown in Figure 3.12. This value needs to be adjusted based on the sizes of the largest microaneurysm candidates in the image.

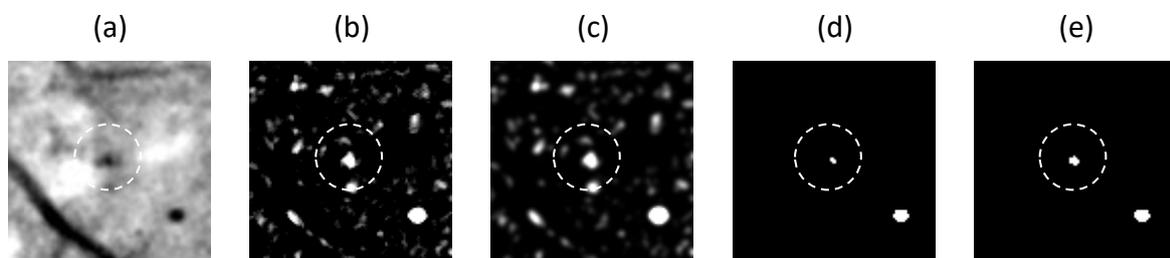


Figure 3.3 An example of the candidate detection phase a) The preprocessed image, b) The result of the bottom-hat operation, c) the Gaussian filter response d) The thresholded image e) the result of the region growing operation. The highlighted region is a true microaneurysm.

### ***3.1.2.1 Parameter selection for Initial Candidates Detection Stage***

The method proposed in this section for initial candidates detection uses a Gaussian filter to enhance the MA candidate responses, followed by a thresholding operation. Next, a region growing step follows to enhance the shapes of the detected candidates. In this section, other alternative methods for detecting initial candidates are discussed and compared with the Gaussian filter method used.

Table 3.1 shows some of the initial candidate methods proposed in the literature. One of the early methods to detect initial candidates is the Gaussian matched filter. Several variants of this technique are discussed in the literature such as [67], [58] and [68]. Another

recent method uses the extended-minima morphological operator which detects local minima regions in an image.

This subsection presents a comparison of two methods for the initial candidate detection of microaneurysms (MAs), namely the extended-minima transform [77] and the Gaussian[66] methods. Experiments were performed to evaluate the performance of both methods using part of the MESSIDOR dataset. The objectives of performing this evaluation are:

1. To find the optimal parameters for each of these methods when they are applied on the MESSIDOR dataset.
2. To compare the performance (sensitivity and false positive rate) of both methods with respect to the MESSIDOR dataset.

Table 3.1 Microaneurysms initial candidates detection methods in the literature.

Paper	Image Type	Initial candidates method
Spencer [61]	Fluorescence	Gaussian Filter (sigma=1)
Hipwell [64]	Colour	Basic Thresholding
Cree[63]	Fluorescence	Gaussian Filter (Sigma=2)
Fleming [66]	Colour	Gaussian Filter (sigma=2)
Niemeijer [74]	Colour	Gaussian Filter (sigma=1) + pixel classification
Sinthanayothin [70]	Colour	Moat operator
Sopharak [77]	Colour	extended-minima (alpha={0.01,0.03,0.05,0.07,0.09})
Sopharak [132]	Colour	extended-minima (alpha=0.05)
Zhang [67]	Colour	Multiscale Gaussian
Giancardo [49]	Colour	Basic Thresholding
Giancardo [59]	Colour	Basic Thresholding
Rocha [76]	Colour	N/A
Hornero [60]	Colour	Mixture model-based clustering
Lazar [133]	Colour	1D Gaussian (multiple directions)
Lazar[55]	Colour	N/A
Quelle [69]	Colour	N/A

Wu [68]	Colour	Multiscale Gaussian (MSMF)
Srivastava [73]	Colour	Frangi-based filters
Mizutani [48]	Colour	double-ring filter
Ramero [80]	Colour	hit-or-miss transform
Streeter [65]	Colour	Gaussian filter (sigma=1.5 and sigma=2.38)
Li [58]	Colour	Multiscale Gaussian (MSMF)
Junior [57]	Colour	Extended Minima
Inoue [72]	Colour	Hessian Matrix Eigenvalues
M Adal [78]	Colour	Hessian Matrix Eigenvalues

The dataset used for the evaluation of the techniques was extracted from the MESSIDOR dataset. A description of the dataset is described in detail in 3.2. The dataset consists of 32 images extracted from the MESSIDOR dataset and microaneurysm candidates have been labeled in the images. During the evaluation only 16 images were considered since they are the ones that contain at least one microaneurysm and were hence suitable for the current evaluation.

### Extended-minima Method

The extended minima was applied after a preprocessing stage to enhance the contrast of the candidates as documented in [77]. The procedure to apply the extended minima is described below:

- Given an input colour image  $I_{RGB}$ , the green channel  $I_G$  is extracted.
- A  $3 \times 3$  median filter is applied to  $I_G$  for the purpose of noise removal.
- Contrast-adaptive histogram equalization is then applied to enhance the contrast in the image.
- In order to correct for non-uniform illumination, shade correction is applied to the contrast-enhanced image. During shade correction, the background is estimated by

applying a  $35 \times 35$  median filter. The resulting image of the aforementioned operations is the preprocessed image  $I_{pp}$ .

- The extended-transform is then applied to detect local minima regions in the image,

$I_{hmin}$ :

$$I_{hmin} = EM(I_{pp}, h) \quad (3.5)$$

Where  $EM(f, x)$  is the extended minima function applied to image  $f$  using a threshold of  $x$ .

The value of  $x$  controls the maximum height of the minima regions that will be detected by  $EM$ .

$I_{hmin}$  is a binary image where candidate microaneurysms are represented by 1 and all other regions by 0. Note that  $I_{hmin}$  will contain blood vessel regions and other noise falsely detected as candidate microaneurysms, since this operation detects all local minima up to a certain height. In order to avoid this blood vessel regions and any bright lesions (exudates) in the image are removed from  $I_{hmin}$ , since microaneurysms will never lie in these regions.

$$\overline{I_{hmin}} = I_{hmin} \wedge (\neg I_{ves}) \wedge (\neg I_{les}) \quad (3.6)$$

Where  $I_{ves}$  is a binary image representing the vessel regions and  $I_{les}$  is a binary image representing the bright lesion regions. Both of these are negated ( $\neg$ ) and a logical AND ( $\wedge$ ) is used to remove these regions from  $I_{hmin}$ . The image  $I_{les}$  and  $I_{ves}$  were calculated using the segmentation methods in [134] and [135] respectively.

Experiments were performed by varying the parameter  $h$  and computing the results of the extended-minima transform ( $I_{hmin}$ ) after vessel & exudates removal. The threshold value ( $h$ ) was varied between 0.05 to 0.3 with an interval of 0.05. This means that each image was processed six times ( $h \in \{0.05, 0.10, 0.15, 0.20, 0.25, 0.30\}$ ) with different values of  $h$  each

time. As the threshold values increases, the extended-minima transform becomes less sensitive to the noise in the image but also misses the more subtle microaneurysms. A threshold value ( $h$ ) which achieves the balance between robustness to noise and accuracy of detection should be selected. Therefore, for each image the following is calculated:

1. The sensitivity is based on individual objects (not pixels)[47]for each microaneurysms. This is calculated by counting the number of MA candidates that have at least 1 pixel detected by the extended-minima transform. The sensitivity is computed by dividing this by the total number of MA objects:

$$s = \frac{\text{Number of true MA objects detected}}{\text{Total number of true MA objects}} \quad (3.7)$$

2. The false positive pixels per image. These are the number of pixels that have been falsely detected as candidate MA pixels.

### **Gaussian Method**

The Gaussian matched filter has been used in several techniques ([67], [58] and [68]) as an initial detection method. The same procedure documented in Section 3.1 was followed. The method is summarized below:

- Given an input colour image  $I_{RGB}$  , the green channel  $I_G$  is extracted.
- A  $3 \times 3$  median filter is applied to  $I_G$  for the purpose of noise removal.
- A  $3 \times 3$  Gaussian filter ( $\sigma = 2$ ) is convolved with the result for further noise reduction.
- Contrast-adaptive histogram equalization is then applied to enhance the contrast in the image.
- In order to correct for non-uniform illumination, shade correction is applied to the contrast-enhanced image.

- During shade correction, the background is estimated by applying a 68x68 median filter. Global contrast normalization is performed on the resulting image by dividing it by its standard deviation.
- Finally, the blood vessels are removed from the image by applying a linear morphological tophat in 8 directions. The length of the tophat structuring element was chosen to be 11px long (measured as the width of the largest blood vessel in the images).
- The resulting image of the aforementioned operations is the preprocessed image  $I_{pp}$ .
- A 15x15 Gaussian filter is applied to detect local minima regions in the image. The values of sigma were varied to measure the effect of this parameter ( $\sigma_g \in \{0.5, 1.0, 1.5, 2.0\}$ ). The result of the Gaussian filter is a probability map between 0 and 1 where higher values indicate higher chance of that pixel being an MA.

$$I_{\text{gauss}} = I_{\text{pp}} * \text{gauss}(\sigma_g), \quad \sigma_g \in \{0.5, 1.0, 1.5, 2.0\}; \quad (3.8)$$

- In order to get the initial candidates a thresholding operation is performed. The top 5% of pixels with the highest values in the probability map are considered to be microaneurysm candidates:

$$I_{\text{thresh}} = \text{thresh}(I_{\text{gauss}}, \tau) \quad (3.9)$$

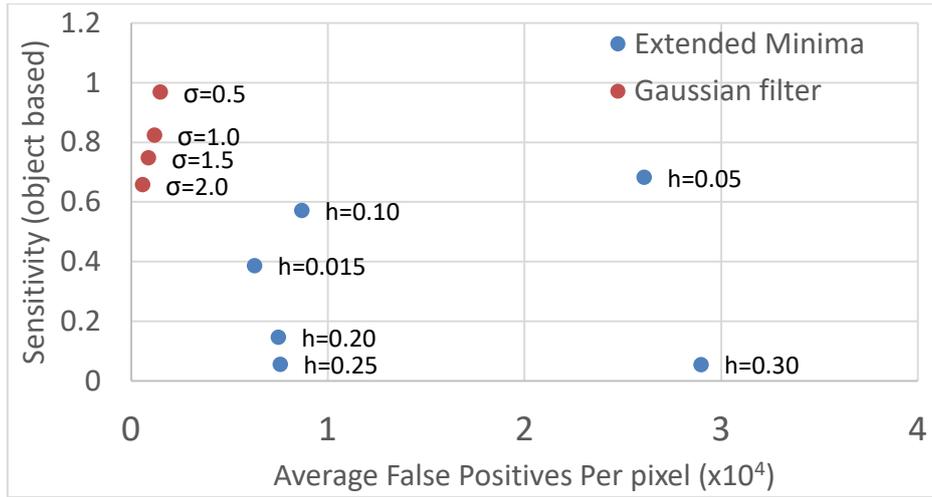
Where  $\text{thresh}(f, t)$  is the thresholding operation applied to image  $f$  using threshold  $t$ . The value of  $\tau$  is chosen such that only 5% of the pixels will belong to the MA candidate regions.

The procedures mentioned earlier for both the Gaussian matched candidate and the extended minima were applied to the dataset. The parameters of  $\sigma \in \{0.5, 1.0, 1.5, 2.0\}$  and  $h \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.3\}$  were varied for each of the Gaussian filter and the

extended minima respectively. The average sensitivities and false positives were computed for each experimental run. The results are visualized in Figure 3.4. Since the points are scattered unevenly, a log scale horizontal axis plot is presented in Figure 3.4(b). The log scale axis was plotted in order to better highlight the relative location of the points, since the scatter was uneven in Figure 3.4(a) and it was difficult to visualize these plots. It is observed that the Gaussian matched filter is superior for detection since all the points have achieved a lower false positive per pixel rate and a higher sensitivity.

In the scatter plots for the Gaussian filter of Figure 3.4(a), it is observed that increasing the value of  $\sigma$  decreases the value of the false positives as well as the sensitivities. This means that any of the  $\sigma$  values can be chosen depending on the current priority. For example, if maximization of the sensitivity is intended, a value of  $\sigma = 0.5$  or  $\sigma = 1.0$  may be selected. However if minimization of false positives is intended, a value of  $\sigma = 2.0$  may be selected. A trade-off between both the sensitivity and false positives rate can be attained by selecting a  $\sigma$  value between both extremes. In this chapter a value of  $\sigma = 1.0$  for the Gaussian filter is used to achieve a balance between the number of a high sensitivity and low number of false positive pixels.

(a)



(b)

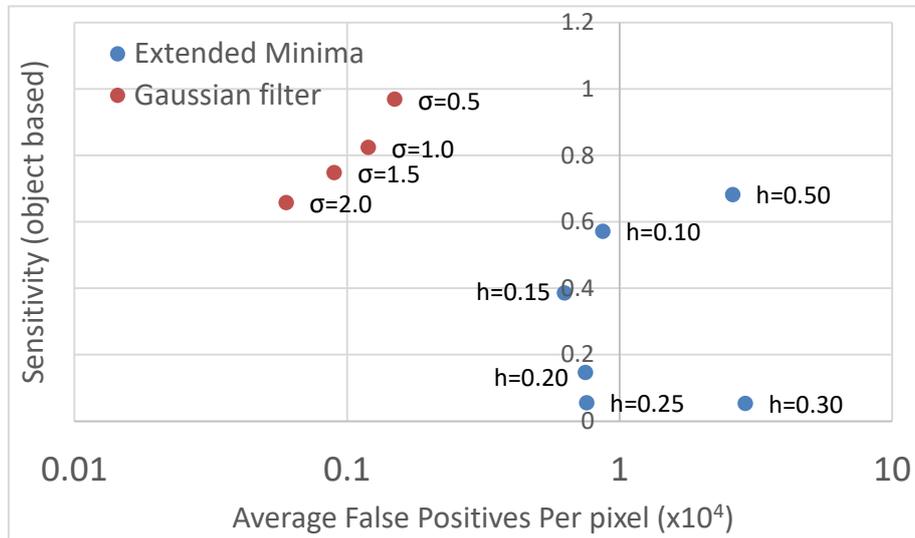


Figure 3.4 (a) Average Sensitivity plotted against average false positives per pixel for both Gaussian filter and Extended minima transform.(b) The same graph in(a) but the horizontal axis in logarithm scale.

### 3.1.3 MA Candidate Classification

The initial candidate detection phase will inevitably produce false positives. Common reasons for this are: 1) vessel cross sections or vessels that were not removed before the Gaussian filter and 2) noise in the image that looks similar to MAs. For these reasons a

classification phase was required in order to reduce the number of false positives that were detected during the candidate detection phase.

Table 3.2 Features list. The symbols below ( $G, t, seed, c, \sigma$ ) are defined below in this section. Key: std - standard deviation, max - maximum, min – minimum.

Category	Index	Description	Parameters	Feature count
Fleming	1	Number of peaks	N/A	1
	2	Major Axis length	N/A	1
	3	Mean of minor and major axis	N/A	1
	4	Eccentricity	N/A	1
	5	Depth of candidate in the original image	N/A	1
	6	Depth of candidate in the preprocessed image	N/A	1
	7	Energy	N/A	1
	8	candidate depth / mean diameter of MA candidate	N/A	1
	9	Energy with depth correction	N/A	1
Moment Invariants	10	7 moment invariant features	N/A	7
Shape features	11	Aspect ratio	N/A	1
	12	Major axis length	N/A	1
	13	Minor axis length	N/A	1
	14	Perimeter	N/A	1
	15	Area	N/A	1
	16	Eccentricity	N/A	1
	17	Compactness	N/A	1
Gaussian features	18	Gaussian seed pixel response: $G_{\sigma}(seed(c))$	$\sigma = 1$	1
	19	$\text{mean}(G_{\sigma}(x, y))_{(x,y) \in c}$		1
	20	$\text{std}(G_{\sigma}(x, y))_{(x,y) \in c}$		1

Gaussian Features 1D	21	Max 1D Gaussian response at various angles: $\max_{t \in \theta} (G_{1,t}^{1D}(x, y))$	$\theta \in \{0,10,20, \dots, 180\}$	1
	22	Min 1D Gaussian response at various angles: $\min_{t \in \theta} (G_{1,t}^{1D}(x, y))$		1
	23	Mean 1D Gaussian response at various angles: $\text{mean}_{t \in \theta} (G_{1,t}^{1D}(x, y))$		1
	24	Std of 1D Gaussian response at various angles: $\text{std}_{t \in \theta} (G_{1,t}^{1D}(x, y))$		1
	25	1D Gaussian response at angle perpendicular to the maximum response (30)		N/A
	26	max ( 29,33 )	N/A	1
Intensity features	27	Sum of candidate intensities	Applied to red, blue, green, hue, saturation, value and preprocessed channels.	7
	28	mean candidate intensity		7
	29	standard deviation of the candidate intensity		7
	30	Range (Max - min candidate value)		7
	31	candidate contrast		7
Morphological features	32	maximum candidate response of the morph close ratio	N/A	1
	33	minimum candidate response of the morph close ratio	N/A	1
	34	mean candidate response of the morph close ratio	N/A	1
TOTAL				70

The proposed method uses a Tree Ensemble classifier for classification. A Tree Ensemble classifier is an ensemble classifier based on decision tree learning. An ensemble classifier combines the decisions of multiple weak classifiers. Our main motivation for the use of this

classifier are: 1) Successful application in other fields in signal processing [136]–[139]; 2) In a tree ensemble classifier, each tree is trained on a bootstrapped sample of the observations. On average, 37% of the observations are omitted for each decision tree. These are known as “out-of-bag” observations and can be used to estimate the predictive performance of the classifier as well as the importance of each individual feature [5]. Therefore, it can rank features while performing classification, giving insights about the most important features; 3) robustness to outliers and the ability to cope with small training sets [140]; 4) Since decision tree classifiers have high variance, the use of an ensemble of decision trees helps reduce this variance [136]; 5) Decision trees are very efficient for both training and classification (as will be seen in Section 3.2.4).

Given a training set  $L$  consisting of data  $\{(y_n, x_n), n = 1 \dots, N\}$  where  $y$  represents the classification label (1 or 0 in our case), a given **CART (Classification And Regression Trees)** classifier  $T(x, L)$  will predict  $y$  given unlabeled data  $x$ .

However, in the case of an ensemble of trees a sequence of training sets  $\{L_k, k = 1 \dots, K\}$  is given and a sequence of classifiers  $\{T_k(x, L_k)\}$  is produced. The  $j^{th}$  classifier  $T_j(x, L_j)$  in the set will produce a label  $y_j$ . The set of labels  $\{y_k\}$  produced by the  $K$  classifiers need to be aggregated to produce a final label  $y$  for unlabelled data  $x$ . In our case a majority vote of the set of class labels  $\{y_k\}$  is used to produce the final classification  $y$ . It has been shown that combining the results of a set of weak classifiers  $\{T_k(x, L_k)\}$  often outperforms using a single classifier on the whole training set  $T(x, L)$ [141].

The final point that needs to be addressed is that given a training set  $L$ , how can a set of training sets  $\{L_k\}$  be produced in order to train each tree classifier  $T_j(x, L_j)$ . A sampling technique known as Bootstrap aggregation (or bagging) [37] was used in order to sample

the training data during the training process. In bagging, the  $j^{th}$  training set  $L_j$  is obtained by drawing  $M$  samples (with replacement) from the set of  $N$  training data,  $L$  (where  $M \leq N$ ). In practice, in order to produce  $L_j$ , a set of  $M$  random numbers  $\{r_m; r_m \leq N, m = 1 \dots M\}$ , and then  $L_j$  is drawn using  $L_j = \{(y_{r_m}, x_{r_m})\}$ . There is no restriction that the generated random numbers are unique and therefore each sample in the set  $\{(y_n, x_n)\}$  may be used more than once or not at all in  $L_j$ . After producing  $K$  training sets from  $L$ , there will be a set of samples in  $L$  that have not been drawn in any of the samples in the  $j^{th}$  classifier  $L_j$ . These unused features can be used to estimate the classification error (out-of-bag-error) for each tree and also estimate the “importance” of each feature (based on each tree and then averaged over all trees) [5]. The bagging approach is used to increase the diversity of training samples across the trees, which leads to increased prediction accuracy for unstable classifiers (including decision trees) [141], [142].

Fleming’s [66] feature set of 10 features was extended to include a set of 70 features. These were based on the features applied to MA classification that have been reported in the literature. Table 3.2 displays a list of the 70 features that were fed into the classifier. These features are explained below in the same order of appearance as Table 3.2:

- **Fleming’s features:** These are the features introduced by Fleming in his technique [66]. These features rely on fitting a paraboloid to each candidate’s intensity profile in order to estimate some parameters from the paraboloid. These features are based on both the shape and intensity of the object. A detailed explanation of these features can be found in the original paper [66].
- **Shape features & Moment Invariants:** These features describe various shape properties of the detected candidates. Moment Invariants (10) are 7 features that

represent various shape properties of an object [143]. Other shape features include Aspect Ratio (major axis length / minor axis length), major axis length, minor axis length, Perimeter, Area, Eccentricity, Compactness. Some of these are common to Fleming's features, however these are calculated at a pixel level rather than after fitting a paraboloid to the candidate. To elaborate, Fleming estimates a paraboloid for each candidate and then computes the values of eccentricity and major & minor-axis length from the paraboloid. In contrast, these features are calculated from the binary image.

- **Gaussian Features:** Features that are based on Gaussian filters have been extensively used in the literature [66]–[68], [77], [82]. A parameter value of  $\sigma = 1$  was used since that is parameter used during the initial candidates detection phase. Some definitions related to these features will follow. The symbols mentioned below also appear in Table 3.2. Let  $I_{shade}$  be the shade corrected image (Section 3.1.1). Define a Gaussian response  $G_\sigma$  as a convolution with the shade corrected image as in Equation (3.10):

$$G_\sigma = I_{shade} * gauss(\sigma) \quad (3.10)$$

where  $\sigma$  is the standard deviation of the matched filter.  $G_\sigma(x, y)$  represents the filter response at coordinates  $(x, y)$ . Let  $\bar{C}$  be a set of initial candidates detected (after region growing). Each candidate  $(c \in \bar{C})$  is a set of coordinates  $(x_i, y_i)$ . Let  $seed(c)$  be the coordinates  $(x_s, y_s)$  of the minimum intensity defined as follows:

$$seed(c) = (x_s, y_s) = \underset{(x,y) \in c}{\operatorname{argmin}}(I_{shade}(x, y)) \quad (3.11)$$

A 1-Dimensional Gaussian is a special case of  $G_\sigma$  applied linearly in one direction.  $G_{s,t}^{1D}(x,y)$  is the 1D Gaussian applied at angle  $t$  and a scale (standard deviation) of  $s$ . The 1D Gaussian filter was applied at a constant scale ( $s = 1$ ). Let the set  $\theta$  be the set of angles applied at each coordinate. In the proposed method:

$$\theta \in \{\theta_i : \theta_i = 10 * i; \quad i = [0..18]\} \quad (3.12)$$

- **Intensity Features:** These are calculated directly from the intensity in the image at multiple bands: the red (R), blue (B), green (G) band in the RGB colour space; the Hue (H), saturation (S) and value (V) bands of the HSV space [67].
- **Morphological Features:** These three features are based on applying a linear morphological close operator (15 px size has been empirically chosen to be larger than the largest vessel in the dataset) at different angles ( $\theta \in \{\theta_i : \theta_i = 22.5 * i; \quad i = [0..7]\}$ ) and are aimed at discriminating vessels from MAs. This is because the linear structures of vessels would respond differently to different angles of the linear operator, while the circular nature of MA objects would cause the response of the morphological operator to be more uniform.

A description of the algorithm has been given in this section. The performance of the algorithm has been evaluated and analysed and this will be discussed in the next section.

## 3.2 Experimental Evaluation

### 3.2.1 Datasets

An overview of current retinal imaging datasets available in the public domain was described in section 2.1. Of these datasets, the following are suitable for evaluating microaneurysm detection algorithms:

- i. Retinopathy online Challenge (ROC) dataset (100 images)
- ii. DIARETDB dataset (189 images)
- iii. MESSIDOR dataset (1200 images)

The ROC dataset contains 100 images split into 50 training images and 50 test images. Groundtruths are only available for the training set. The test set groundtruths are not public since the contest organizers used those to evaluate submissions. Moreover, the groundtruth of this dataset has generated discussion in the literature [48], [49] due to the fact that many of the MA candidates marked in the groundtruth are invisible to the viewers or could not be seen by other expert observers. This made it difficult to rely on this dataset as for the evaluation of the proposed method.

The DIARETDB has released three versions of the dataset. However according to their website, the first two versions of the dataset have been “deprecated”. Therefore the work described in this thesis relies on the latest version of their dataset for evaluation (DIARETDB1 v2.1).

The MESSIDOR dataset is a very large dataset with 1200 images. However, the microaneurysm labels are not supplied with the dataset, which makes it not suitable for evaluating microaneurysm segmentation algorithms specifically (candidate based evaluation). For this reason, a small subset of the dataset has been groundtruthed for the purpose of evaluation<sup>3</sup>, and this will be used in the next section. Thirty-two images were selected from the MESSIDOR dataset to cover a wide range of retinopathy. A summary of the images in dataset in terms of retinopathy grade and number of MAs included in the set

---

<sup>3</sup> The groundtruth dataset can be downloaded using the following link:  
<http://blogs.kingston.ac.uk/retinal/?p=311>

is shown in Table 3.3. The grade is predominantly based on the number of MAs (the presence of haemorrhages and new vessels is also factored in) [3]. Sixteen healthy images (without MAs) were included in order to maintain a balanced dataset while performing per-image MA evaluation (evaluating whether or not an MA candidate exists for each image). The images all had the same resolution of 1440 x 960 pixels. The images were groundtruthed by an expert grader. The dataset has also been made publicly available<sup>3</sup>. All the visible MAs were marked during the process. A circular marker was used rather than pixel-based marker [47]. The main reason for the use of a circular marker is that majority of the literature has relied on lesion-based metrics to measure the accuracy of detection. Using lesion-based metrics makes the results more sensible since the measure is informative of the amount of MA candidates that were detected by a given algorithm. In contrast, reliance on pixel-based metrics can be misleading due to the imbalance in proportion between very few MA pixels and a large number of background pixels.

Table 3.3 Distribution of DR grades (a) and resolutions (b) of images in the dataset.

Retinopathy Grade	Number of Microaneurysms	No. of Images (training)	No. of Images (test)
DR0	0	8	8
DR1	1-5	3	4
DR2	6-14	3	3
DR3	>15	2	1
	TOTAL	16	16

Motivated by the Retinopathy Online Challenge [47], each MA candidate was labelled using one of the following categories: Obvious, Regular or Subtle and Close-to-Vessel. The labels

Obvious, Regular or Subtle are based on the relative visibility and/or local contrast of the corresponding MA in the image. Close-to-Vessel is a label given to MA candidates that lie close to a blood vessel. A detailed explanation of each category is mentioned in [47].

### 3.2.2 Evaluation

The public MESSIDOR dataset mentioned in the previous section was used to train and measure classifier model error. Hence, the models have been built using the training set and accuracy of the model was measured using the 16 images in the MESSIDOR test set. In order to ensure that the model is not overfitting the MESSIDOR dataset the performance of the model was also measured on the DIARETDB1 test set. In the case of an overfit model the error on the DIARETDB1 test set would be greater than the error on the MESSIDOR test set [137]. Hence, it was important to ensure that the error on both the MESSIDOR and the DIARETDB1 sets was similar for the model. The following procedure was followed in order to perform the evaluation on the dataset:

- The MESSIDOR dataset was split into 16 images for building the model (training set) and 16 images for measuring the model classification error (test set) as shown in Table 3.3.
- The procedure outlined in Section 3.1.2 was used to generate the 70 features. The training groundtruth was used to label the features. These features were used to train the Tree Ensemble classifier and generate the model. Note that the training set was undersampled in order to maintain a balance between the positive and negative samples. One parameter that needs to be set for the Tree Ensemble classifier is the number of generated trees ( $N$ ). Figure 3.5 shows the out-of-bag (OOB) classification error as a function of the number of trees in the Tree Ensemble classifier. A value of

$N = 40$  was selected based on Figure 3.5. This is also within the range recommended by Breiman [141]. The value of  $N$  was chosen from the plot at the point where there is no more significant decrease in the OOB error.

- For each image in the MESSIDOR test set, the procedure outlined in Section 3.1.2 was used to find the set of candidates and their corresponding features. Each candidate feature vector was then fed into the classifier in order to classify it as a true candidate or a false positive.
- The final classified set of candidates was then compared against the ground truths in order to assess the performance.

In addition to the proposed algorithm, Fleming’s algorithm [66] was also implemented in order to compare it against the proposed technique. Fleming uses a K-Nearest-Neighbours classifier with 9 features. This will be known as the “basic feature set”. In addition, the K-Nearest-Neighbours classifier has been trained with all 70 features, and this will be known as “extended feature set”.

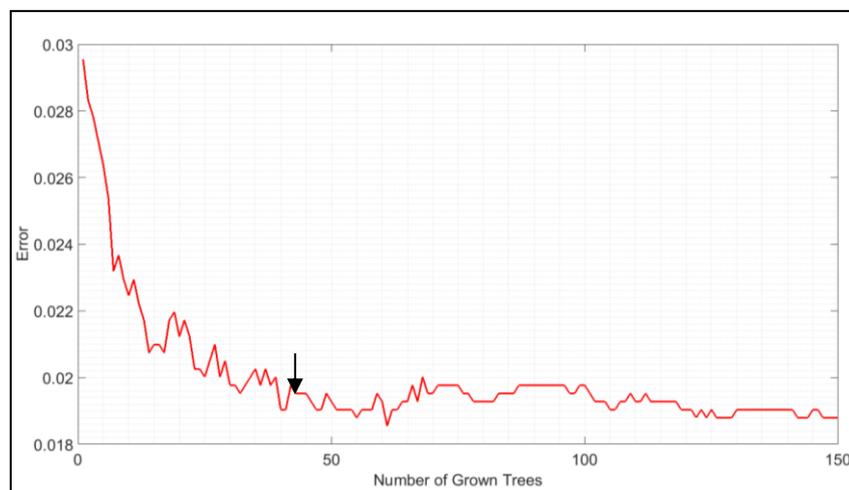
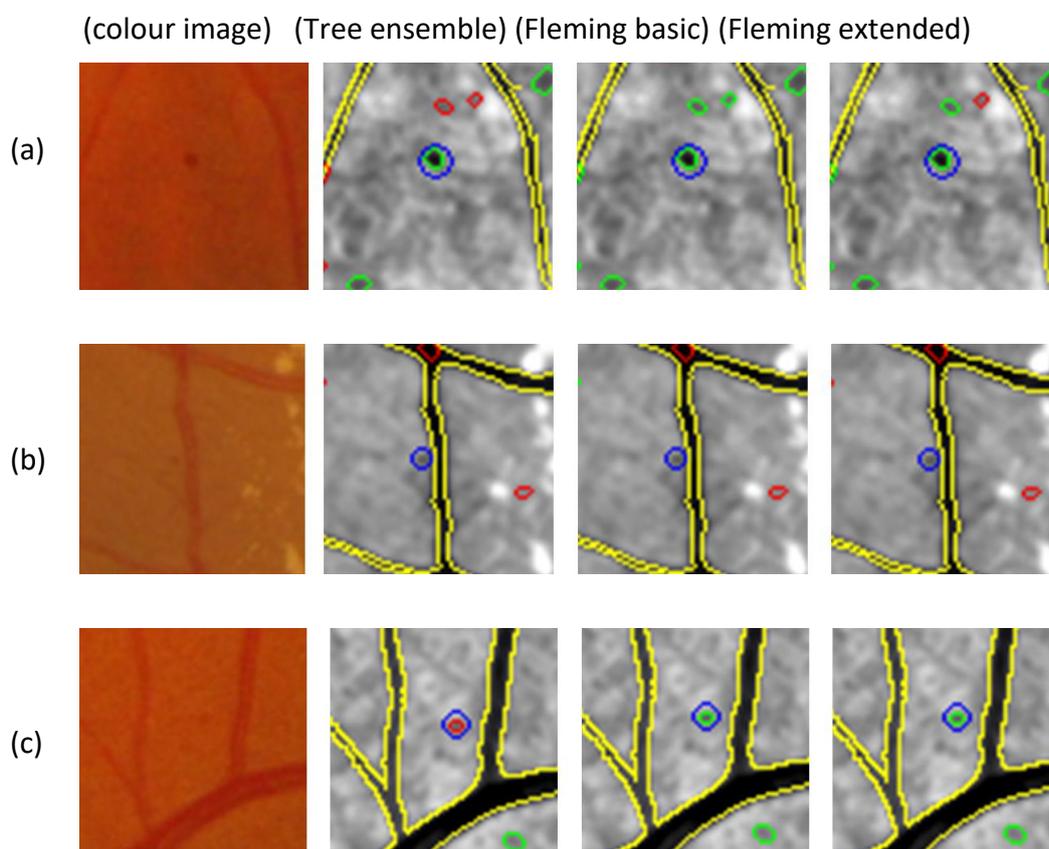


Figure 3.5 Out of bag (OOB) classification error as a function of the number of trees in the Tree Ensemble classifier.

Figure 3.6 shows example patches from the MESSIDOR images for the three methods mentioned above. The figure shows several patches from multiple colour images. The patches are scaled by 200% and for each patch the groundtruth, the MA candidates (after region growing) and the result of the classification have been highlighted. For the purpose of comparison the results are shown for the proposed method, Fleming’s algorithm with the basic feature set, and Fleming’s algorithm with the extended feature set.



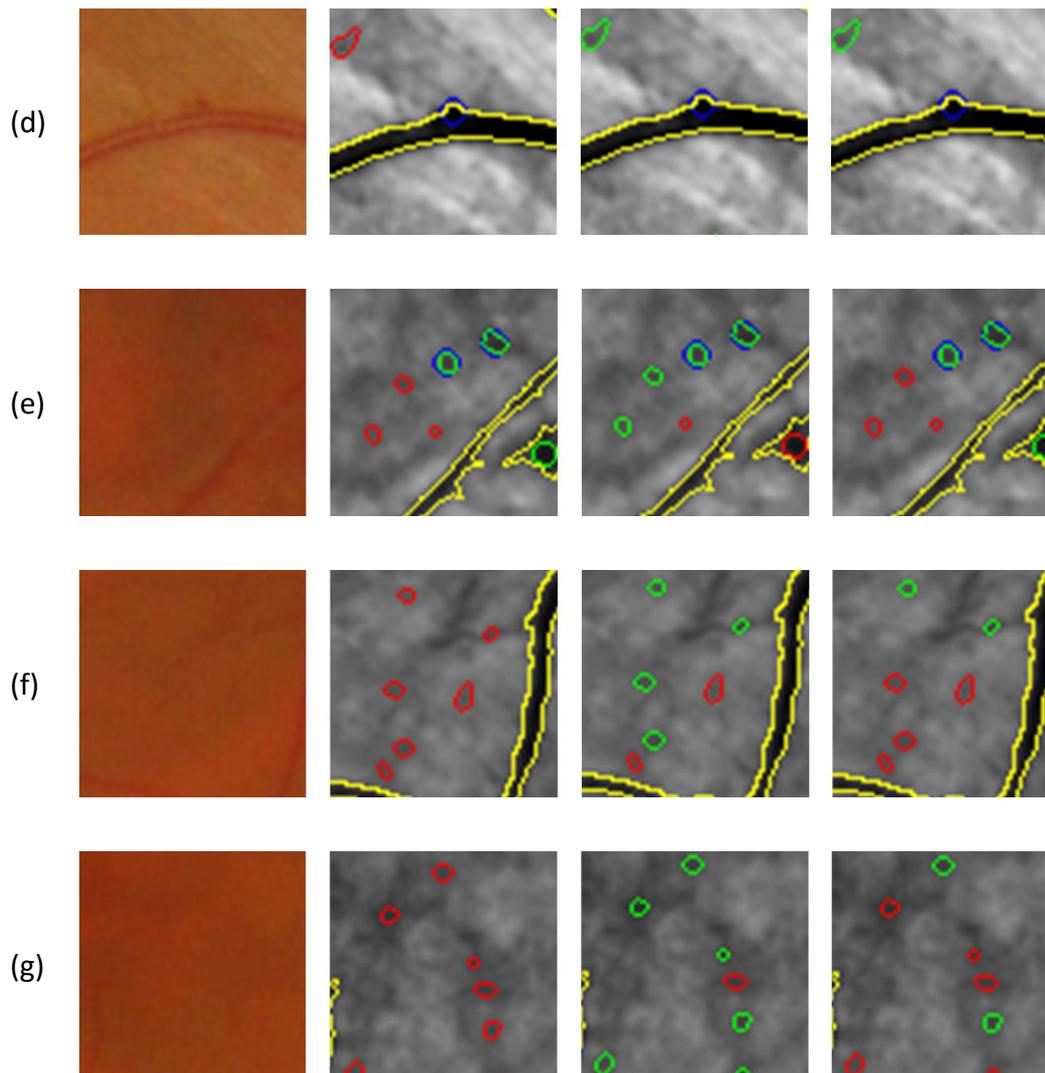


Figure 3.6 Examples of microaneurysm detection algorithms applied to the MESSIDOR dataset. The first column shows the colour image patch. Columns 2-4 show the preprocessed image with the algorithm results highlighted. A blue circle represents the groundtruth labelled microaneurysm. A green circle represents an MA candidate detected by the algorithm. A red circle represents an MA candidate that was detected as a candidate MA but classified as a false positive by the classifier. The yellow boundaries shows the vessel regions detected by the QUARTZ software [144].

The colour codes of the labels in Figure 3.6 are as follows: A blue circle represents the groundtruth labeled microaneurysm. A green circle represents an MA candidate detected by the algorithm. A red circle represents an MA candidate that was detected as a candidate MA but classified as a false positive by the classifier. Further analysis of these visual patches is presented below along with the quantitative results.

In order to measure the accuracy of the algorithm, the sensitivity of the proposed method was measured [61]. Given image  $I_i$  in a test set (for  $i = [1..t]$ , where  $t$  is the number of images in the test set), let  $G_i$  be the set of true MA objects (groundtruth) for image  $I_i$  and  $C_i$  be the set of detected candidates after classification (Section 3.1.3) for image  $I_i$ . The sensitivity is defined as:

$$Sensitivity = \frac{\sum_{i=1}^t |G_i \cap C_i|}{\sum_{i=1}^t |G_i|} \quad (3.13)$$

Where  $|x|$  represents set cardinality of  $x$ . Thus the sensitivity is the proportion of true candidates detected in proportion to the total number of true candidates. A candidate  $c \in C$  is considered to be equivalent to  $g \in G$  if the pixel coordinates of  $g$  and  $c$  overlap by at least 1 pixel (Figure 3.7). Note that the sensitivity was measured at a candidate level rather than at a pixel level (lesion-based sensitivity). Since the number of true negatives cannot be determined, a Free Receiver Operating Curve (FROC) was used rather than a traditional ROC curve [9]. In an FROC curve, the x-axis is replaced with the average number of false positive candidates per image instead of the specificity. Figure 3.9 shows FROC curves for both the proposed method (using the Tree Ensemble classifier) (solid) and Fleming's method (using K-Nearest Neighbours) (dotted) for multiple categories in the dataset. Each curve was generated by evaluating the trained model on the test set. The dotted curve represents the performance of Fleming's state-of-the-art algorithm on the MESSIDOR dataset, the dashed curve represents the performance of Fleming's algorithm (using the K-Nearest-Neighbours classifier) with the extended feature set.

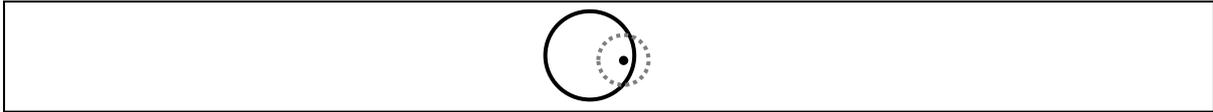


Figure 3.7 An example of a candidate (dotted circle) that is considered to match a groundtruth (solid circle). There is a match since the centre of the candidate lies within the groundtruth region.<sup>3</sup>

Lesion-based sensitivity was used for the purpose of evaluation. A value of  $K = 15$  was used for the K-Nearest-Neighbours classifier [66]. Each classifier produces a probability value ( $P$ ) between 0 and 1 representing the likelihood of it being a true one. A threshold value  $P_t$  was used to produce the final classification (i.e. class = 0 if  $P \leq P_t$ , otherwise class = 1). The value of  $P_t$  was varied in order to generate the FROC curve. Tree Ensembles generate trees at random and generate the attribute splits at random as well [136], [137]. Due to this feature of the classifier, every run produces results with slightly different accuracy (Figure 3.8). To overcome the varying results, the Random Forest classifier was applied multiple times and, for each run in the curve, the Area Under each Curve (AUC) was computed. Finally, the curve with the median AUC value was displayed. This helps reduce the variability in the FROC curve. For our experiments it was found that applying the classifier 11 times was sufficient to reduce the variability in the results. In an experiment run on the MESSIDOR dataset the Tree Ensemble classifier was run 100 times and the average mean squared error (MSE) for all the curves was found to be 0.0124, which shows that the variability in the classifier can be considered negligible.

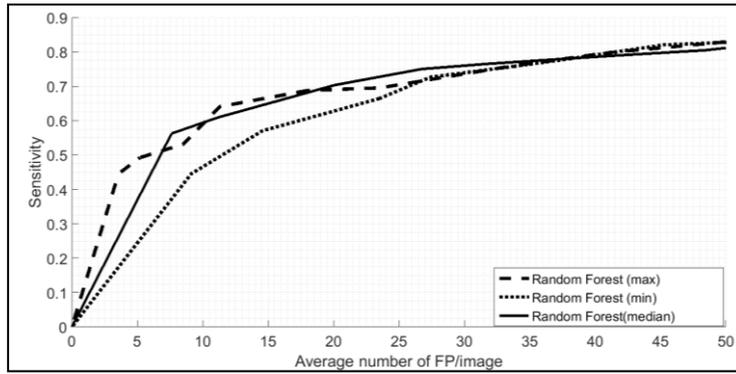
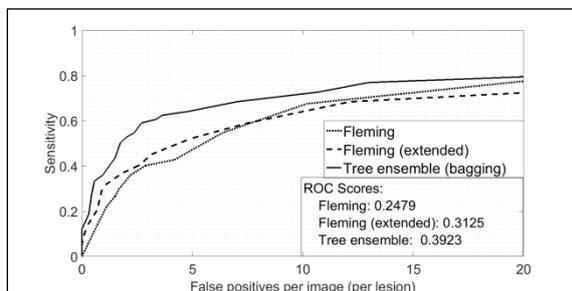


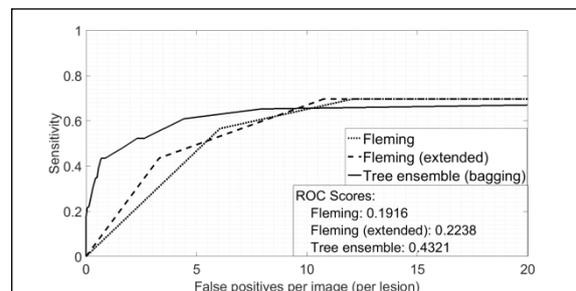
Figure 3.8 An example of variability in results everytime a Tree ensemble model is built for classification. In this example the Tree ensemble classifier was run 11 times and the maximum, minimum and median results are displayed.

Figure 3.9 shows the FROC curves for the three algorithms: Tree ensemble, Fleming (basic feature set) and Fleming (extended feature set). The model was built using the training set of the MESSIDOR dataset and the performance was measured using the MESSIDOR test set. The first FROC curve in Figure 3.9 was generated by evaluating the classification model performance on the test set (16 images, 128 microaneurysms) after training using the entire set of MA labels in the training set (16 images, 128 microaneurysms).

a) MESSIDOR - All Candidates



b) MESSIDOR - Obvious Candidates



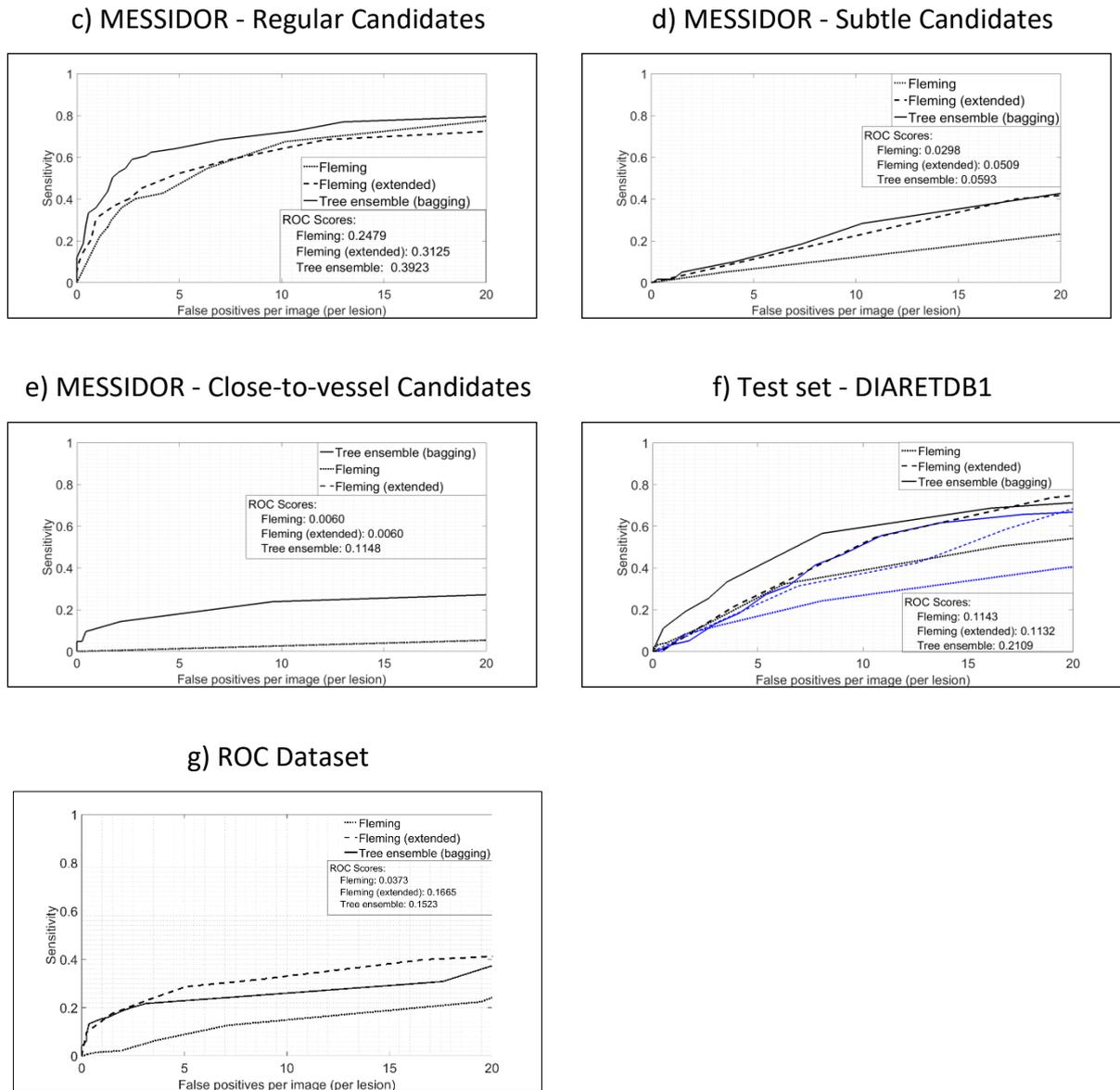


Figure 3.9 Free-Receiver operating curve (FROC) for all microaneurysm candidates in the test set. In f) the black lines represent the performance of each respective model on the DIARETDB1 test set by training on the DIARETDB1 training set, while the blue lines represent the performance of each model on the DIARETDB1 test set after training on the MESSIDOR training set.

In addition, evaluations for the subsets of the MESSIDOR groundtruths are also presented: obvious candidates, regular candidates, subtle candidates and close-to-vessel candidates. Each reported performance for a subset of the MESSIDOR dataset was trained on the respective subset of microaneurysm groundtruths in the dataset. This was done to highlight the variation in performance for each category of microaneurysms. Finally, the classification

models were tested on the test set of the DIARETDB1 set. In order to demonstrate the overfitting process, two models were generated for each classifier: first by training on the MESSIDOR training set and second by training on the DIARETDB1 training set. The models were then evaluated on the DIARETDB1 test set (61 images, 169 microaneurysms). In Figure 3.9(f) the black curves represent the performance of each respective model on the DIARETDB1 test set using a model which was trained on the DIARETDB1 training set (28 images, 85 microaneurysms), while the blue curves represent the performance of each model on the DIARETDB1 test set using a model which was trained on the MESSIDOR training set. It is observed that the blue curve performance is comparable to the dashed black curve, which represents the performance of the Fleming (extended feature set) algorithm on the DIARETDB1 dataset. This shows that the Tree Ensemble classifier is generalizable across datasets since the performance is still comparable using a model trained on a different dataset. It should be noted that while the classifier has been shown to be generalisable between training and test sets, there are still some parameters in the prior stages that need to be set and are dependent on both the image resolution and the field of view. These parameters include:

- The median filter applied for background estimation (preprocessing). The size of this filter needs to be larger than all the features in the image in order to eliminate these structures.
- The  $\sigma$  parameter of the Gaussian filter. Empirically a value of  $\sigma = 1$  has produced good responses across the three datasets being evaluated.
- The structuring element width needs to be set to a value which is larger than the large blood vessel structures in the images

- The maximum region growing size. Empirically it was found that a value larger than the largest MA candidates produces acceptable results.

The final plot in Figure 3.9 represents the performance benchmarks of the dataset on the Retinopathy Online Challenge dataset. Since the training set groundtruths are not available for this dataset, the training set was split into 2 parts (25 images each) and they were used for training and evaluating the algorithm. It is observed that the performance is relatively low on the ROC dataset compared to both MESSIDOR and DIARETDB1. This can be explained by the fact that the groundtruth of this dataset has generated discussion in the literature [48], [49] due to the observation that many of the MA candidates marked in the groundtruth are invisible to the viewers or could not be seen by other expert observers.

In order to quantify the results further, the ROC Scores for each method are presented in Table 3.4. The ROC Score [47] calculates the average sensitivity of the curve at multiple False Positive Rate intervals (1/8, 1/4, 1/2, 1, 2, 4, 8). In other words, the ROC score measures the average sensitivity of a technique at low False Positive rates. The ROC score simply captures the first section of the FROC curve (until 8 FP/image) as a simple quantifiable result. The ROC score focuses on the algorithm performance at low false positive rates.

Table 3.4 ROC scores for multiple categories in the set. Bold represents the maximum score.

Category	Method		
	Fleming (basic feature set)	Fleming (extended feature set)	Proposed method (Tree ensemble)
MESSIDOR - All candidates	0.2479	0.3125	<b>0.3923</b>
MESSIDOR (Obvious Only)	0.1916	0.2238	<b>0.4321</b>
MESSIDOR (Regular Only)	0.2479	0.3125	<b>0.3923</b>

MESSIDOR (Subtle Only)	0.0298	0.0509	<b>0.0593</b>
MESSIDOR (Close-to-vessel Only)	0.0060	0.0060	<b>0.1148</b>
DIARETDB1 (test set)	0.1143	0.1132	<b>0.2109</b>
ROC dataset	0.0373	<b>0.1665</b>	0.1523

Fleming [66] is used as a baseline for comparison with the proposed technique . Fleming’s reported per-lesion performance on a large private dataset is comparable to recently published methods. This makes it reasonable to use it as a baseline for comparison. Recent methods which are comparable to Fleming include Wu (2015) [81], Adal (2014) [80], Inoue (2013) [73] and Li (2013) [68]. The sensitivities for these methods at 1 FP/image are: Fleming (2006) 0.51, Wu (2015) 0.23, Adal (2014) 0.36, Inoue (2013) 0.23. This value of 1 FP/image was chosen since it is the median value of the 8 samples used while computing the ROC Score [47]. Quantitative and visual analysis of the results of the proposed method are discussed below.

The FROC curves in Figure 3.9 can be understood by observing the patches in Figure 3.6. It is observed that the Tree ensemble classifier is superior to the K-Nearest-Neighbours classifier in terms of eliminating the false positives in the image. This is observed more evidently in rows (e), (g) and (f), since the second column shows more red circles that do not intersect with a true candidate (blue circle). To elaborate, a red circle which does not intersect with a blue circle is a true negative. A blue circle which intersects with a green circle is a true positive. A blue circle which intersects with a red circle, or does not intersect with anything is a false negative.

The patches show that most of the time, the three methods equally detect the true positive candidates in the image. In fact, Figure 3.6(c) shows that the proposed method has marked a true candidate MA as a false positive while the K-Nearest-Neighbours classifier correctly marked it as a true candidate. The conclusion drawn is that all methods are almost equivalent in terms of maintaining true positive candidates while the proposed method is superior in terms of eliminating false positives. This is important from a clinical perspective since a reduction in the number of false positives while maintaining the same sensitivity will avoid over-referral of the patients. One more interesting observation is that of Figure 3.6(d) which shows an example of a close-to-vessel candidate. The figure shows that during the preprocessing phase the candidate merges with the vessel and therefore remains undetected as a candidate. This indicates why the performance of close-to-vessel candidates is very low for all methods.

The analysis of the patches in the previous paragraph will help to explain the FROC curves in Figure 3.9. The FROC curves show that the proposed method performs better when all candidates are considered. In addition, it is also better at distinguishing close-to-vessel candidates. However, the curves intersect in the case of Obvious, Regular and Subtle candidates. This makes it difficult to judge which method performs better. For this purpose, a numerical measure is needed in order to compare the curves in a more objective manner.

The ROC Score [47] calculates the average sensitivity of the curve at multiple False Positive Rate intervals (1/8, 1/4, 1/2, 1, 2, 4, 8). Table 3.4 shows the ROC scores for the corresponding ROC curves in Figure 3.9. As illustrated by the FROC curves, the ROC score for the proposed method is better in terms of all candidates and close-to-vessel candidates. It also achieves a better score for regular, obvious and subtle candidates.

Figure 3.9(f) shows that the proposed method can build a model that is generalizable across datasets. In this figure the black lines represent the performance of each respective model on the DIARETDB1 test set by training on the DIARETDB1 training set, while the blue curves represent the performance of each model on the DIARETDB1 test set after training on the MESSIDOR training set. It can be seen that the proposed method is much more generalizable than the Fleming technique since the performance does not deteriorate significantly even when the classifier is trained on a different dataset (training on MESSIDOR and testing DIARETDB1). This fact increases the confidence that overfitting does not occur on the model that was trained on the MESSIDOR dataset.

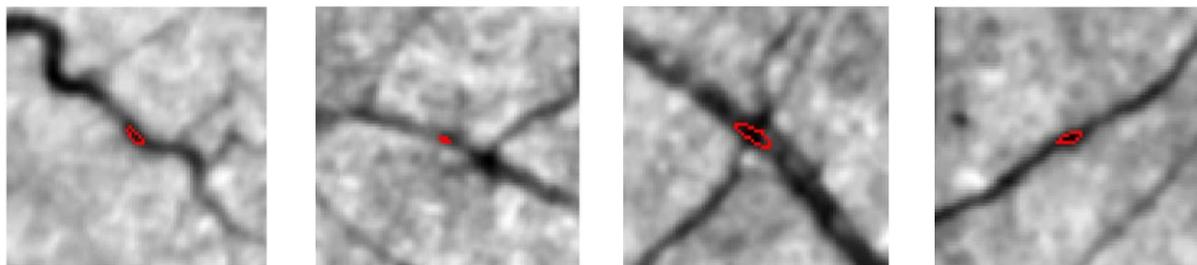


Figure 3.10. Examples of candidates falsely detected on vessels. Red circles represent initial candidate MAs detected by the proposed method

One of the common causes of false positives is the vessel cross sections and vessel regions that have not been removed (Figure 3.10) [66], [78]. This is one of the opportunities where combining spatial information can reduce the number of false positives in the image.

### *Per pixel evaluation*

In addition to a per-lesion based evaluation, a per-pixel evaluation can give insights about the performance of the algorithm at a pixel level. Figure 3.11 shows ROC curves for the performance of the proposed algorithm on various datasets at the pixel level. The curves are

plotted for sensitivity against the positive predictive value (PPV). For each pixel, the algorithm labels it as either a candidate or not. There are 4 possible outcomes for each label:

- True positive (TP): Algorithm labels a pixel as a candidate pixel and it is a candidate pixel.
- False positive (FP): Algorithm labels a pixel as a candidate pixel and it is not a candidate pixel.
- True negative (TN): Algorithm labels a pixel as background pixel and is a background pixel
- False negative (FN): Algorithm labels a pixel as a background pixel and it is not a background pixel.

Based on the number of True positives, false positives and false negatives, the sensitivity and PPV values can be calculated as follows:

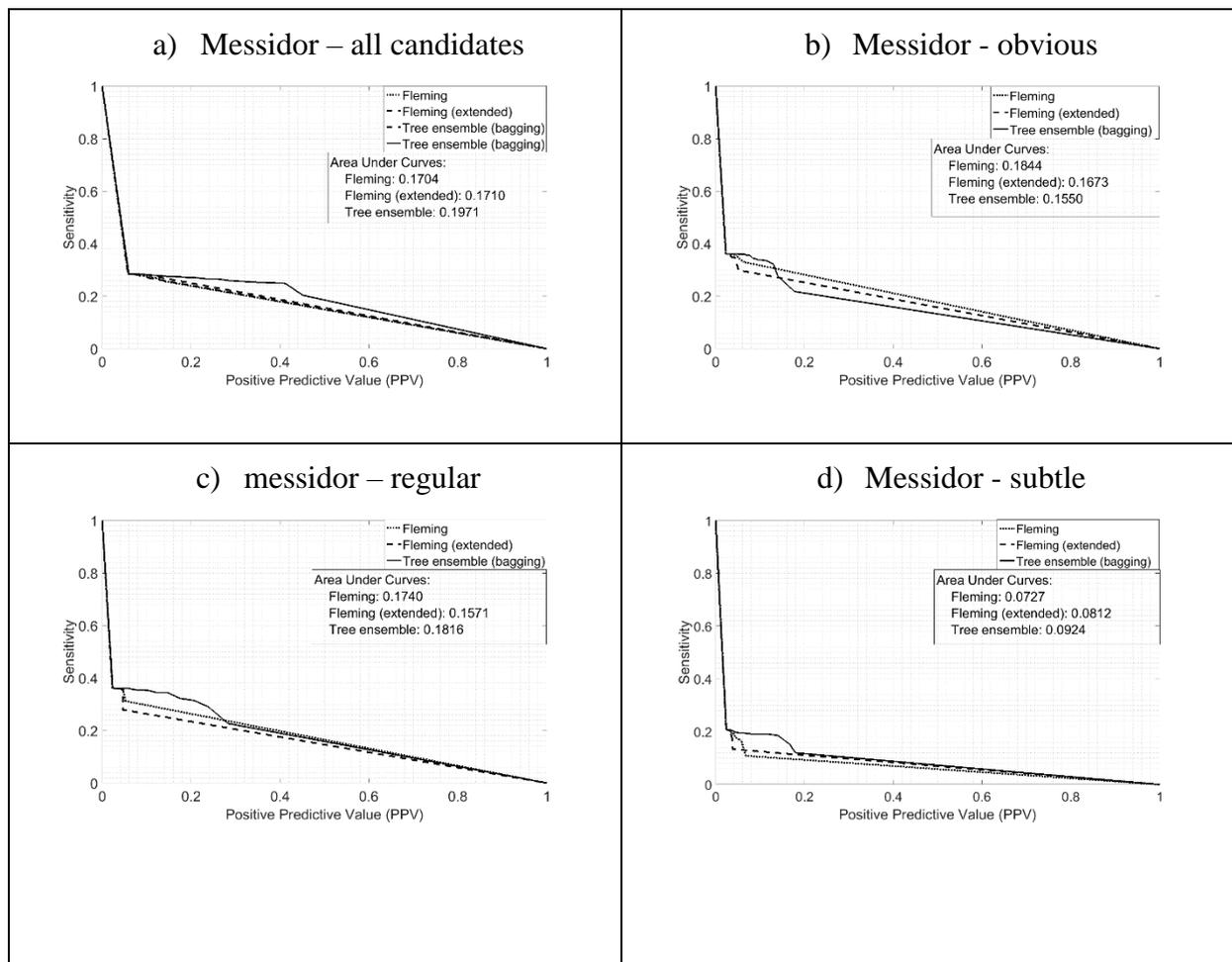
$$sensitivity = \frac{TP}{TP + FN} \quad (3.14)$$

$$PPV = \frac{TP}{TP + FP} \quad (3.15)$$

Where  $TP$ ,  $FP$ ,  $TN$ ,  $FN$  represent the number of pixel occurrences of each category respectively. Finally, a threshold value can be varied based on the classifier result for each candidate and this can be used to generate the ROC curves. Note that a lot of methods have used the specificity measure for the ROC score [47], [60], [60], [69]. However, the specificity value can be biased in this case since there is a large imbalance between the number of background pixels and the number of foreground pixels. There are very few microaneurysm

pixels compared to the number of background pixels. Since the specificity value measures the proportion of background pixels that are correctly measured by the algorithm, this measure will be quite high for most algorithms since the false positives will account for a very small proportion of the total set of background pixels.

Figure 3.11 shows the MESSIDOR, DIARETDB1 and the ROC datasets. The evaluations have been performed for Fleming, Fleming with extended feature sets and the proposed algorithm which uses a tree ensemble classifier. Every algorithm produces a set of candidates and the likelihood of each classifier being a true candidate. A threshold parameter is varied in order to generate each ROC curve.



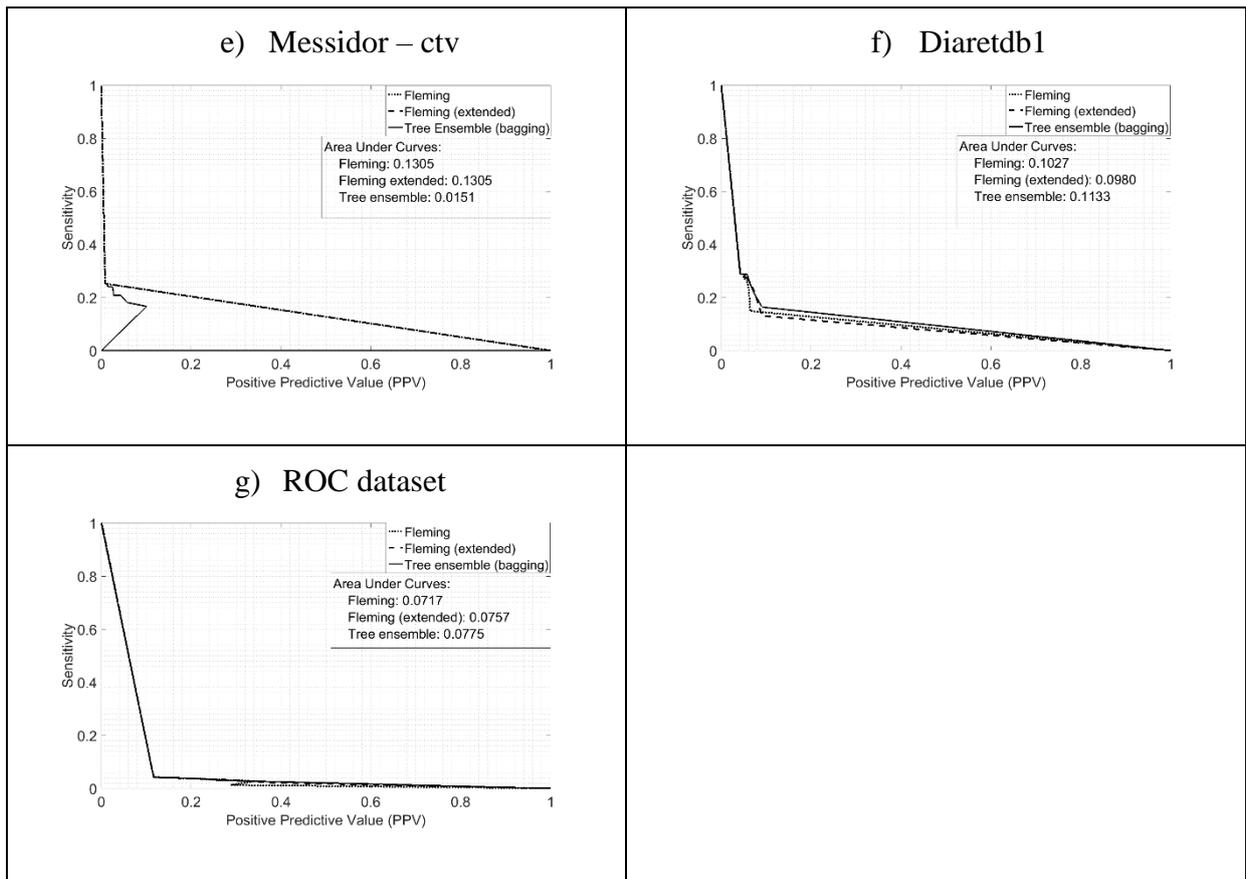


Figure 3.11 ROC curves for per-pixel evaluation of the proposed algorithm compared with 2 other methods. The evaluations were performed on MESSIDOR, DIARETDB1 and ROC datasets.

The value of the sensitivity represents the proportion of true negative ground truth pixels that have been detected by the algorithm. The PPV value represents the probability of the algorithm performing a positive detection – given that a pixel was selected. Similar to the per-lesion evaluation, the variants of the MESSIDOR dataset are also considered (obvious, regular, subtle and close-to-vessel candidates).

In general, it is observed that the pixel-based sensitivities in Figure 3.11 are much lower than the lesion based sensitivities (Figure 3.9). Intuitively, one would expect the sensitivities of the lesion-based measure to be close to the sensitivities of the pixel-based measure considering that the algorithms would match the pixels of each detected lesion. However, it is observed that this is not the case. For example, for the MESSIDOR dataset, the knee-jerk

of the curve reaches a sensitivity of just above 0.2 (Figure 3.11(a)), whereas in the per-lesion case it reaches a sensitivity of over 0.6 (Figure 3.9(a)). This suggests that only 30% of the candidates' pixels are detected on average ( $0.2 = 0.6 * 30\%$ ). Upon further investigation, it was noted that many of the groundtruth labels assume that the shapes of microaneurysm candidates are circular[47]. Furthermore, it was observed that a lot of the groundtruths are less conservative about the actual boundaries of each candidate (Figure 3.12).

Since a lot of the curves overlap, it is difficult to compare the performance across the algorithms. While not perfect, the area under the curve for each plot can be used to provide a direct comparison between multiple algorithms. Table 3.5 shows the area under the curve (AUC) for each plot in Figure 3.9. The proposed algorithm achieves a higher AUC in the case of the ROC dataset and the MESSIDOR dataset. It achieves a higher AUC for the regular and subtle categories of the MESSIDOR dataset. The method proposed by Fleming achieves a higher AUC in the case of MESSIDOR (obvious) and DIARETDB1. The reason for this is probably due to some candidates which have been detected by the KNN classifier in Fleming but detected as false candidates by the tree ensemble classifier. As noted in the previous section, the tree ensemble classifier performs better at removing false positives but may also remove more true positives from the image. In contrast, the KNN classifier may contain more false positives, but could also correctly mark some candidates as true while the tree ensemble classifier would eliminate these candidates.

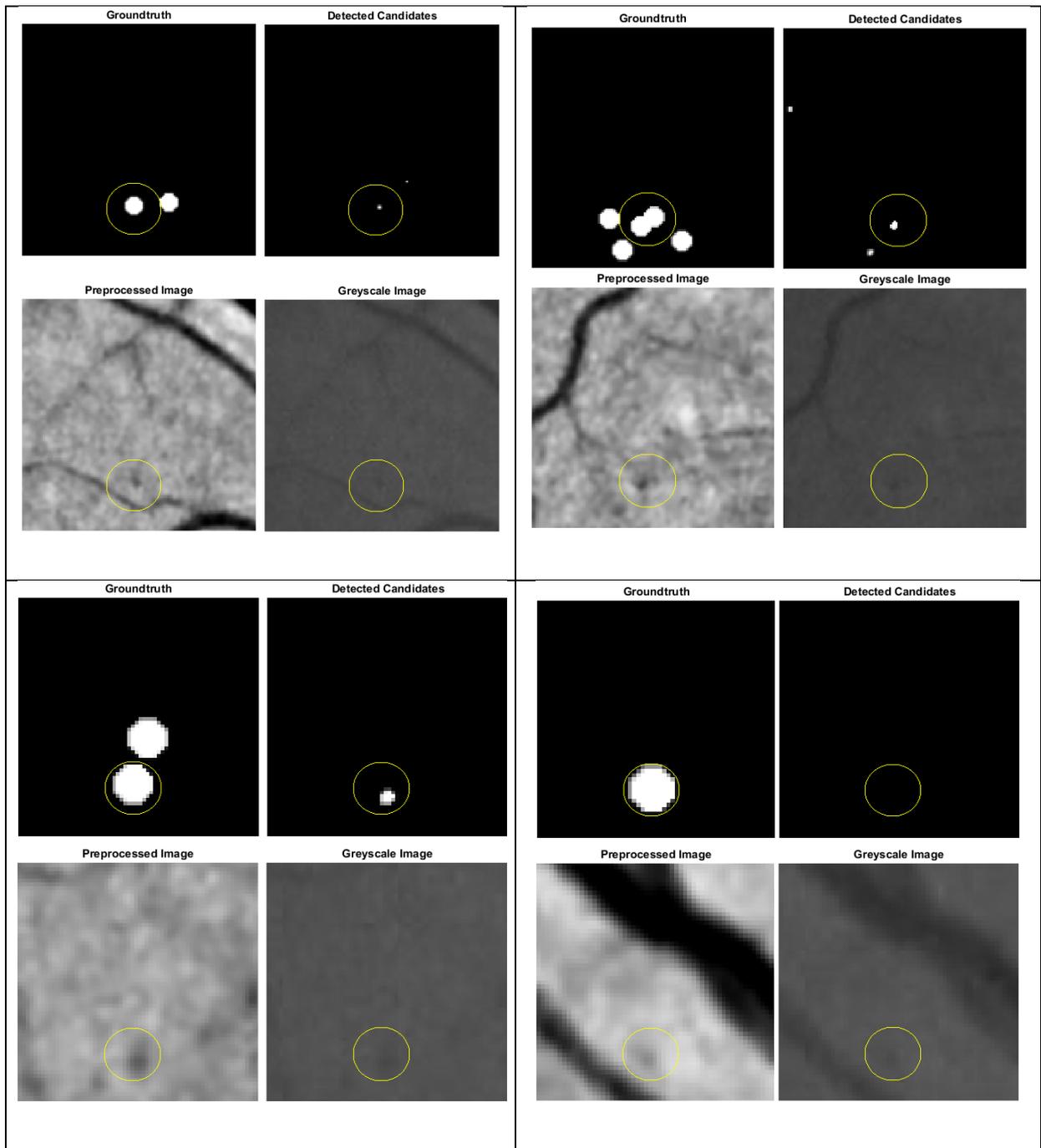
Table 3.5 Area Under the Curve for per-pixel evaluation of microaneurysms across various datasets

Category	Method		
	Fleming (basic feature)	Fleming (extended feature)	Proposed method (Tree ensemble)

	set)	set)	
MESSIDOR - All candidates	0.1704	0.1710	<b>0.1971</b>
MESSIDOR (Obvious Only)	<b>0.1844</b>	0.1673	0.1550
MESSIDOR (Regular Only)	0.1740	0.1571	<b>0.1816</b>
MESSIDOR (Subtle Only)	0.0727	0.0810.2	<b>0.0924</b>
MESSIDOR (Close-to-vessel Only)	<b>0.1305</b>	<b>0.1305</b>	0.0151
DIARETDB1	<b>0.1207</b>	0.0980	0.1133
ROC dataset	0.0717	0.0757	<b>0.0775</b>

Due to the ambiguity of pixel labelling for the groundtruths of microaneurysms, it is preferable to use a per-lesion measure for comparison since this measure has a clear indication about the performance of the algorithm in terms of the number of candidates that the algorithm has successfully detected.

Another useful measure to assess the performance of a microaneurysm detection algorithm is a per-image evaluation. This measure produces a single evaluation for each image depending on whether or not it contained microaneurysms. Since microaneurysms are among the early signs of Diabetic Retinopathy, the existence or absence of a microaneurysm in an image can indicate whether a patient is healthy. This is why an image-based measure could have applications in a clinical application (for the diagnosis of DR).



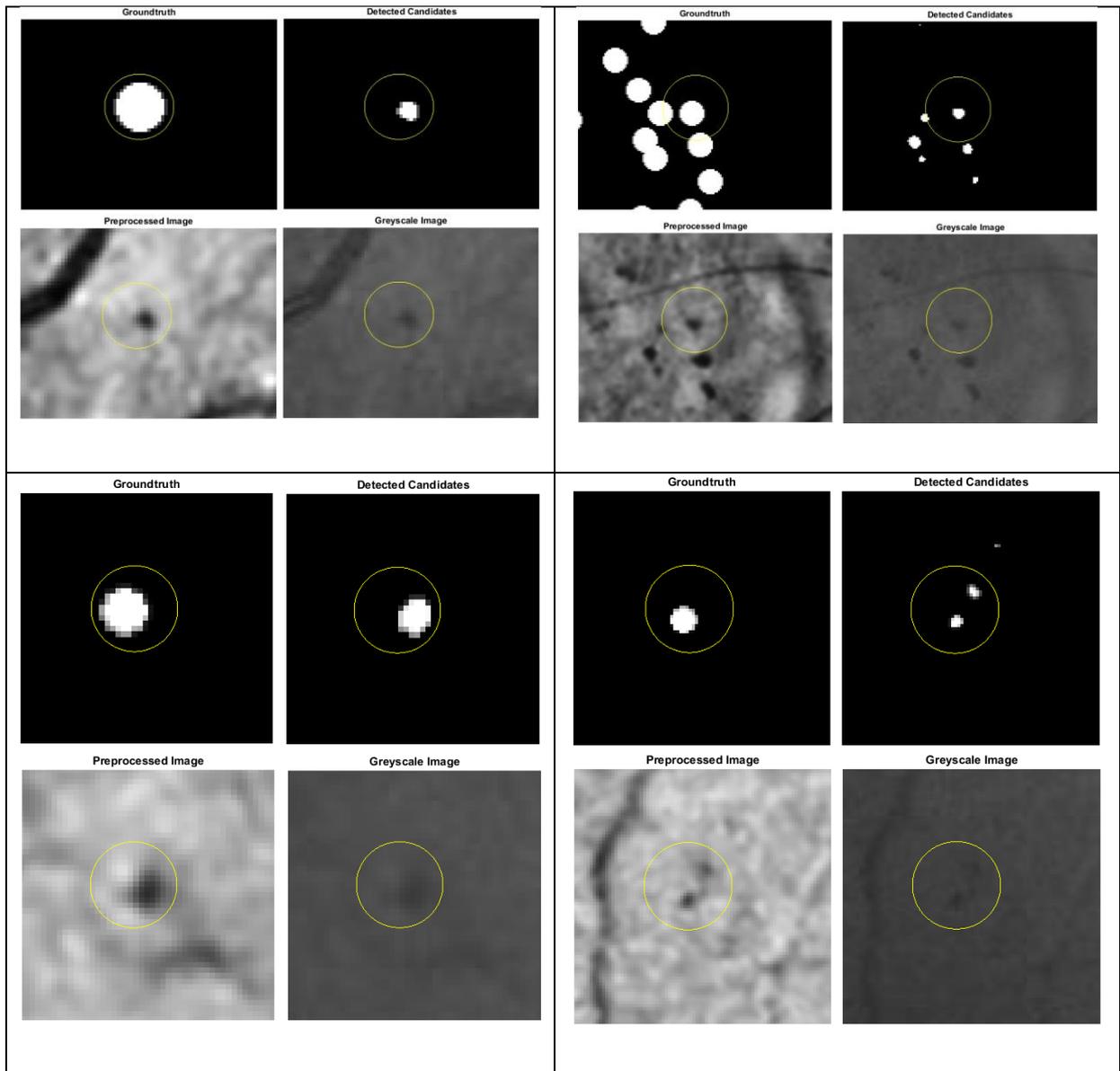


Figure 3.12 Examples of Region growing stage applied to various candidates in the ROC dataset

### 3.2.3 Feature Analysis

Since a large number of features are used in the proposed algorithm a presentation of the discriminative ability of each feature would be insightful to understand the most impactful feature and also for other researchers developing microaneurysm detection algorithms. Figure 3.13 shows the importance of the features in the same order of appearance as in Table 3.1. The features are categorised by type. This performance was measured based on

the trees generated from the tree ensemble classifier. A rough visual analysis of the chart shows that there are some important features for each category of features. The important features which achieved a high importance value are: pre-processed image intensity (an intensity feature), minor axis length, (a shape feature), major axis length (a Fleming feature), 1D Gaussian response perpendicular to maximum (a Gaussian feature), maximum response of the morphological structuring element (a morphological feature). A diverse range of features was detected in each case. This indicates that each category is contributing to the discrimination of features.

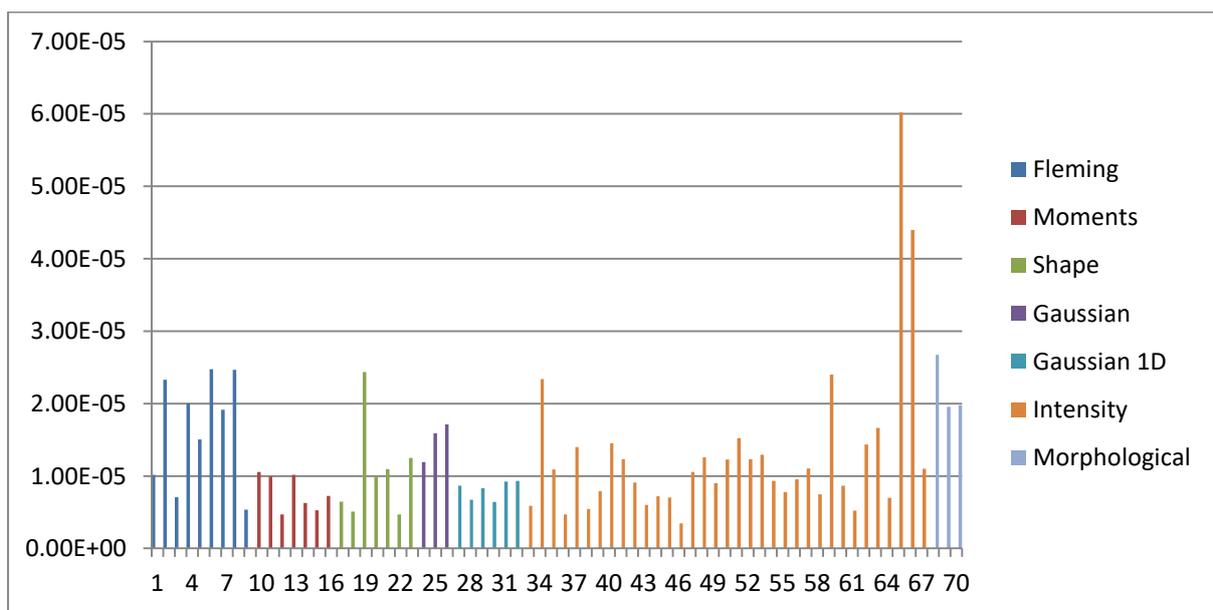


Figure 3.13 Feature importance measured using the MESSIDOR dataset. X-axis represents the feature number. Y-axis represents the feature importance measure.

The extended feature set of 70 features that was used is based on features that have been applied in the literature. The features used by the algorithm are among the most common features that were present in the literature.

However, a question which arises is whether all of the features are important features that contribute to the performance of the Tree ensemble classifier [145]. Some features may not contribute much information to the classifier and hence may be ignored. Experiments were performed to rank the features according to the Predictor Importance. The Predictor Importance for a given attribute is calculated by computing the entropy (or Information Gain) for each tree in the Tree ensemble and then computing the average entropy for each tree. The predictor importance can be computed while building the Tree ensemble model and provides an indication of the importance of features. Figure 3.13 shows the measured predictor importance for the 70 features in Figure 3.13. It is observed that there is varying importance for the features in the dataset.

In general, by visualizing the graph it is observed that there are some important features for each category of features. It is observed that the Gaussian 1D and the Shape features are in general less important than the rest of features (though visual observation). This does not imply that they should be ignored, however, since to decide which features need to be removed a feature selection method should be utilized, and this step is left for future work [46].

The top 5 and least 5 features in terms of discriminative ability are listed in Table 3.6. Some interesting observations can be made based on this table. Firstly, it is observed that intensity features appear twice in the most discriminative list and also twice in the least discriminative list. The intensity features that appear in the most discriminative list are computed from the preprocessing channel, suggesting that computing features from this category will help produce discriminative features. Another observation is that 2 of 9 Fleming features appear as most discriminative. The minor axis length feature also appears

to be in the list of most discriminative features. Interestingly, there is another Major axis length feature that appears in the shape feature category (feature 18). The difference between the major axis length in Fleming (2) and the shape feature list (18) is that the first is measured after fitting a paraboloid to each candidate whereas the latter does not fit a paraboloid. It is observed that feature 18 (major axis length shape feature) is ranked low in the graph whereas feature 2 (major axis length Fleming feature) is among the most discriminative features. This raises the question about whether they are both correlated features which causes only one of them to be given a high importance while the other to be given an understated importance value [146]. If that is the case, then one may utilize this fact and eliminate some of the Fleming features by substituting them with shape features that are more efficient since they are calculated at the pixel level.

A final remark about the least discriminative features is that 3 out of the least 5 discriminative features are moment features, which suggests that the use of moment features does not help in the classification process. The process of experimenting with feature elimination and selecting a smaller set of the 70 features is left for future work and is outside the scope of this thesis.

Table 3.6 The 5 most (a) and least (b) discriminative features according to the bagging feature importance measure.

(a)			(b)		
Feature number	Feature description	Category	Feature number	Feature description	Category
65	The intensity range in the value channel	Intensity	46	Mean candidate intensity in red channel	Intensity
			36	Range in the hue channel	Intensity

66	Intensity range in the preprocessing channel	Intensity
6	Depth of candidate in the preprocessed image	Fleming
19	minor axis length	Shape
2	Major axis length	Fleming

12	3 <sup>rd</sup> moment invariant	Moment
14	5 <sup>th</sup> moment invariant	Moment
15	6 <sup>th</sup> moment invariant	Moment

### 3.2.4 Algorithm Performance

Tests were performed to measure the computational performance on the two datasets. The machine used for the tests was a core i5-4590 @ 3.30GHz CPU with 8GB RAM and an SSD hard drive. The total time taken for each stage over the whole dataset was computed and this was divided by the number of images in the dataset to get the average time spent per stage. The only exceptions are: 1) The ensemble training phase which was timed only for the training set of each dataset stage and 2) the classification step which was only timed for the test images of the dataset.

The average time required by the algorithm for each image in the MESSIDOR dataset was 166 seconds, while the average time per image in the DIARETDB1 dataset was 65 seconds.

The list of times that each stage of the algorithm takes is presented in Table 3.7.

Table 3.7 Average performance of the proposed algorithm on both MESSIDOR and DIARETDB2 datasets. The given times correspond to the average time per image for each dataset (with the exception of the classification time).

Stage	MESSIDOR time (s)	DIARETDB2 time (s)
Preprocessing*	1.77	1.85
Initial Candidates Detection*	5.91	7.42

Region Growing*	51.08	20.66
Features computation*		
Paraboloid fitting	13.06	3.09
Fleming intensity	16.40	5.87
Fleming size	9.94	3.54
Moments	0.087	0.024
Gaussian	2.14	0.76
Gaussian 1D	1.39	0.5917
Intensity	51.92	17.43
Shape	11.34	3.81
Morphological	0.678	0.817
Ensemble training**	0.90	0.025
Classification (per candidate)***	0.000038	0.000049
<b>TOTAL</b>	<b>166.61</b>	<b>65.869</b>
* Averaged over the entire dataset		
** Averaged over the training set only		
*** Averaged over the test set only		

Table 3.7 shows time taken by various parts of the algorithm on both MESSIDOR and DIARETDB2 datasets. It should be noted that the classification time was measured per candidate rather than per image. If it is assumed that each image contains 10 microaneurysms on average then this value needs to be multiplied by 10-folds to get the classification time per image. The proposed algorithm takes close to 3 minutes for each algorithm on average (for MESSIDOR). However, it is noticed that there are three main bottlenecks which constitute around 2.5 minutes (83%) of the computation time: region growing, intensity features computation and Fleming features computation (which constitutes paraboloid fitting, size features and intensity features computation). Eliminating or optimising these steps could cause the algorithm to run in under a minute, and this is left

for future work for a more time-optimized version of the algorithm. Furthermore, the algorithms have run on the images without any resizing performed – resizing could decrease the computation time significantly.

On the other hand, the average computation time for the DIARETDB2 dataset is 65.8s, which was only 40% of the time needed by the MESSIDOR dataset on average per image. The main reason for this discrepancy was that the number of microaneurysms labelled per image in the DIARETDB2 is 2.85 (89 images and 254 labelled candidates) compared to 8 microaneurysms per image for MESSIDOR (32 images and 256 labelled microaneurysms). Intuitively this implies that less initial candidates will be detected per image in the DIARETDB2 dataset, which results in less computation for region growing and feature computation, since this needs to be computed for less candidates on average in DIARETDB2. Despite this discrepancy it is still noted that most of the time (76%) was spent on the aforementioned 3 bottleneck stages for the DIARETDB2 dataset as well.

The performance of the proposed method is compared with the reported performances of other methods. Fleming has proposed several postprocessing methods and on average the reported performance in a dataset of 422 images was between 58 and 100s per image. The time performance of the hybrid classifier used by Akram [7] has been reported. The hybrid classifier required  $1.4 \times 10^{-3}$ s per candidate. One of the advantages of an ensemble classifier was that it was very efficient to train and test. The ensemble classifier requires less than a second for training in both DIARETDB1 and MESSIDOR datasets. Furthermore, classification was also very efficient for decision trees and it takes  $4 \times 10^{-5}$ s on average.

The theoretical computational complexity of several stages of this method will now be analysed. Let the following input sizes be defined:

1. The image input of size  $n \times m$ ,
2. The number of candidates detected in the image,  $c$ . This is image dependant, but some of the steps' complexities depend on this value.
3. The height of the deepest tree in the ensemble classifier,  $h$ . This is length from the root node of the tree until a leaf node.

Based on the input sizes above, the complexity of each step in the algorithm is analysed (in big-oh notation):

1. **Preprocessing:** Preprocessing involves convolutions and arithmetic operations at the pixel level. Therefore, the number of the computations is a factor of the image size  $n \times m$ . Therefore the computational complexity is  $O(n \times m)$
2. **Initial Candidates Detection:** Initial candidates detection involves a morphological bottom-hat, a Gaussian convolution filter and a thresholding operation. Both the convolution filter and the thresholding operation are a function of the number of pixels and therefore their complexity is  $O(n \times m)$ . The morphological operation also depends on the length of the structuring element, which is constant, and therefore its complexity is also  $O(n \times m)$ .
3. **Region Growing:** Region growing involves recursively expanding each candidates' surroundings until a stopping condition is met, or a maximum size is reached. In our case the maximum size is 100 pixels. At each iteration during the region growing, an energy function is computed, which depends on the number of boundary pixels of the candidate at this iteration. This value changes for each iteration. To simplify, it is assumed that the number of pixels is the maximum growth size (100 in our case). The number of iterations will also vary for each candidate. To simplify, it is assumed

that the number of iterations is constant for each candidate as well,  $i$ . This means that the number of operations will be  $100 * i * c$ , and therefore the complexity is  $O(c)$ .

4. **Features Computation:** The complexity for each of the feature categories is as follows:

- a. **Fleming features:** All the Fleming features depend on fitting a paraboloid to each candidate function. This is an optimisation function and the number of iterations will vary depending on the candidate intensity profile and how fast the optimisation algorithm will converge. Further analysis of this complexity is left for future work.
- b. **Moments:** Moment Invariant features are computed from the candidate pixel locations. Their computation involves arithmetic operations on the candidate pixel locations and therefore the computational complexity depends on the number of pixels in the candidate.
- c. **Gaussian:** These features are all based on convolution filters and therefore they are of order  $O(n \times m)$
- d. **Gaussian 1D:** These features are all based on convolution filters and therefore they are of order  $O(n \times m)$
- e. Intensity:  $O(1)$
- f. **Shape:** Computing the shape features involves finding the connected component labels for the candidates. This can be done for each individual candidate. Therefore, the number of computations is dependent on the number of candidate pixels found.
- g. **Morphological:**  $O(n \times m)$

5. **Ensemble Classification:** There are a number of classification trees,  $k$ , each of which has a certain height. To simplify, assume that all the trees have the same height,  $h$ . Every classification step is applied to a potential candidate. A comparison occurs at each node in the tree from the root until a leaf node. Therefore,  $h$  comparisons are made for the deepest tree. Therefore, the total number of comparisons required for the  $k$  trees is  $k \times h$ . Since  $k$  is a constant factor, the complexity is  $O(h)$ .

By combining all information above the complexities of the algorithm is of the order  $O(n \times m + h)$  [147] – which means that the number of computations performed varies linearly with both the image size and the height of the classification tree.

### 3.3 Conclusion

In this chapter, an algorithm that detects microaneurysms from colour fundus images has been proposed. The main contribution of this method was the use of a tree ensemble classifier during the classification stage, which has been very effective in other segmentation applications [5], [128]. Another contribution is the use of a large number of features per microaneurysm candidate for the purpose of classification. The computed features belonged to multiple categories such as shape, Gaussian filter responses, intensity, and morphological responses.

The proposed algorithm applies some preprocessing steps in order to enhance the contrast of microaneurysms and remove uneven illumination. A Gaussian filter followed by simple thresholding was used in order to detect the initial set of candidates. Since vessel cross-sections have an intensity profile similar to microaneurysms, they need to be removed before applying the Gaussian filter. A morphological bottom-hat operation was used in order to remove vessels from the image. A region growing operation was used to enhance

the shapes of the candidates. Finally, 70 features were computed for each candidate and these are used to perform the final classification.

A new groundtruth dataset of the MESSIDOR dataset was introduced. The dataset consisted of microaneurysm labels for 32 images of the MESSIDOR dataset. The MESSIDOR dataset does not contain microaneurysm labels and therefore the groundtruth was completed in order to evaluate the algorithm on a per-candidate basis. The microaneurysm labels were categorized according to the appearance of each microaneurysm. The categories that were labels are: obvious, regular, subtle and close-to-vessel. This was done to assess the performance of the algorithm for candidates that possess varying appearances. The groundtruth dataset has been published and made publicly available online in order to encourage comparison with other techniques in the future.

The algorithm was evaluated on three datasets (DIARETDB1, ROC and MESSIDOR) and compared with another state-of-the-art technique. Results were compared with an implementation of another technique (Fleming), and a variant of it. Visual results have shown that the proposed method's ensemble classifier removes more of the false positives than Fleming's K-nearest-neighbours classifier. However, in some cases the ensemble classifier incorrectly removes some of the true positive candidates. It was also observed that in some cases a candidate that is close to a blood vessel is removed at the early stages during the vessel removal step and as a result this candidate goes undetected.

Quantitative results were presented for the proposed algorithm and compared with Fleming. Per-candidate FROC curves were presented for both MESSIDOR, DIARETDB1 and ROC datasets. In addition, FROC curves were plotted for each category in the MESSIDOR groundtruth (obvious, regular, subtle and close-to-vessel). In addition, results of training the

classifier on the MESSIDOR dataset and evaluation on the DIARETDB1 set were presented. The objective of doing this was to observe whether the proposed algorithm was generalisable and robust to images from different sets. This is important, since in a practical scenario the classifier will only be trained once and then run on images from a variety of patients. The ROC score achieved by the proposed method on the MESSIDOR dataset was 0.3923, compared to 0.2479 for the Fleming algorithm. The ROC score achieved by the proposed method on the DIARETDB1 dataset was 0.2109, compared to 0.1143 for the Fleming algorithm.

One of the advantages of using a tree ensemble classifier is that it can rank features according to their importance during the training phase. The proposed feature set was analysed to understand which of the 70 features are more distinctive for classifying microaneurysms. The features belonged to multiple categories, including: Moment features, shape features, Gaussian features, intensity features and morphological features. Some of the features in the set are similar and therefore are likely to be eliminated. For instance, the major and minor axis lengths of each candidate are computed twice: once at a pixel level and once from the best-fit paraboloid. An open question is whether all four features are necessary or whether two of those can be eliminated from the feature set. According to the feature ranks only the major axis length was ranked highly while the pixel-level feature was not. This suggests that only one of these features is necessary, and there is a chance that they may be correlated. Further feature selection to reduce the size of the feature set of the classifier was left for future work. The presentation of the feature ranking will be helpful for future algorithm designers who wish to include a supervised classification stage.

The computational efficiency of the proposed technique was measured and results were reported. The total time required to run the algorithm was in the range of 60-160seconds. On the other hand, the reported performance of Fleming was between 58 and 100 seconds. Further efficiency can be achieved by eliminating some of the features that need to be computed (particularly the intensity features take a lot of time to compute). Furthermore, the region growing stage can be made more efficient since it is one of the bottlenecks of the proposed method according to the empirical results. Theoretical computational complexity of the proposed method was discussed. The overall computational complexity of the proposed method was found to be  $O(n \times m + h)$ ; where  $n$  and  $m$  are the image width and height respectively,  $h$  is the height of the classification tree. This complexity is in Big-Oh notation, and it means that the number of computations performed on the algorithm grows linearly as a function of both the image size and the height of the classification tree.

In the next chapter, the proposed algorithm will be extended to work with a spatial pair of images. Given a spatial pair of images from the same eye (one optic-disc centred and another macula centred) a collective segmentation the microaneurysms will be presented which makes use of the extra information from both images.

# 4

## **MICROANEURYSM DETECTION USING SPATIAL INFORMATION**

---

In many Diabetic Eye Screening Programmes, including the UK National Health Service Diabetic Eye Screening Programme (NHS DESP), at least 2 views of the retina are captured including both optic disc centred view and the fovea centred images. These images overlap together and thus have common MAs that appear in both views (with variability in contrast). Despite the availability of both views, the majority of algorithms that have been proposed in the literature have only considered the information contained in a single image.

In this chapter, a combination of spatial information from two retinal images for microaneurysm (MA) detection will be performed. As discussed earlier, Diabetic Retinopathy (DR) is one of the most common causes of blindness among working-age adults [148]. Signs of DR can be detected from images of the retina which are captured using a fundus camera. Microaneurysms (MAs) are one of the early signs of DR. Several algorithms for automated MA detection from single 45-degree fundus images have been proposed in the literature. However, in many Diabetic Eye Screening Programmes, including the UK

National Health Service Diabetic Eye Screening Programme (NHS DESP) [149], at least 2 views of the retina are captured including both optic disc centred view and the fovea centred images (Figure 4.1). These images overlap together and thus have common MAs that appear in both views (with variability in contrast). Despite the availability of both views, the algorithms that have been proposed have only considered the information contained in a single image. In this chapter, an increase in detection accuracy is achieved by fusing the information from two views of the retina.

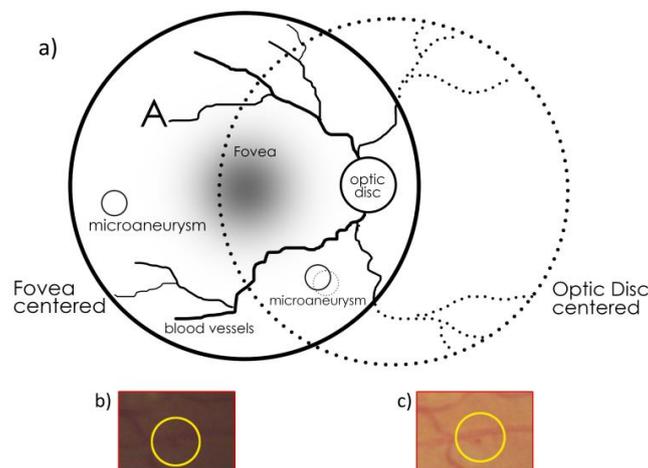


Figure 4.1. a) A conceptual diagram of the two spatial views of the retina (optic disc and fovea centered). The patches in b) and c) demonstrate how the same microaneurysm in different views of the retina can have a varying level of contrast.

In the previous chapter, a new algorithm for microaneurysm detection has been presented. The algorithm has used a Gaussian filter to detect candidate microaneurysms from an image. A tree ensemble classifier was used to classify and detect the likelihood of a candidate being a true candidate or a spurious one. This algorithm has relied on features from a single image to perform this classification. In this chapter, the combination of classifier results from spatial information in retinal images will be explored. Figure 4.1

illustrates the concept of a spatial pair of images that contain an overlapping region. The methodology of the algorithm will be explained in the next section. Experimental results will be presented in Section 4.2. Finally, the Chapter will be concluded in Section 4.3.

## **4.1 Methodology**

Some of the multi-image analysis techniques have been covered in the literature review chapter (Section 2.4). Connor [25] performed vessel segmentation on a series of fundus images and measured both vessel tortuosity and width in these images. This was done in order to find a correlation between these measures and some signs including Diabetic Retinopathy. Arpenik [124] used fractal analysis to distinguish between normal and abnormal vascular structures in a human retina. Patterson [125] developed a statistical approach for quantifying change in the optic nerve head topography using a Heidelberg Retinal Tomograph (HRT). This was done for measuring disease progression in glaucoma patients. Artes [126] reported on the temporal relationship between visual field and optic disc changes in glaucoma patients. Bursell [127] investigated the difference in blood flow changes between insulin-dependent diabetes mellitus (IDDM) patients compared to healthy patients in video fluorescein angiography. Narasimha [123] used longitudinal change analysis to detect non-vascular anomalies such as exudates and microaneurysms. A Bayesian classifier is used to detect changes in image colour. A “redness” increase indicates the appearance of microaneurysms. Similarly, an increase in white or yellow indicates the appearance of exudates. While the problem of analysing “change” and “progression” of disease has been studied in the literature, to the best of our knowledge, the combination of a spatial pair of retinal images for the improvement of detection of MAs has not yet been explored.

In this section, a novel method for combining information from two views of the retina (optic disc centred and fovea centred) will be presented. The technique will be evaluated using a dataset of spatial image pairs. Figure 4.2 illustrates the overview of the proposed method. A way to compare MA candidate detection using the combined image versus the 2 singular images from the same patient was needed.

The method proposed in the previous chapter has been adapted to work on 2 images of different spatial views. As shown in Figure 4.2, the 2 images were analysed using the algorithm proposed in Chapter 3 as normal. This produces a set of 2 scores (scores<sup>(1)</sup> and scores<sup>(2)</sup>). The intention is to combine these 2 scores together. However, before fusing the scores a way to find correspondences between candidates is needed. Therefore, the images need to be aligned first (Image Registration) and then a match between corresponding candidates should be found (Candidate Matching). Each of the matched candidate's scores can then be combined to produce a single set of scores for both images. This combination should increase the confidence in some true MA candidates, and will hence improve the final algorithm after the final scores are thresholded with an operating point  $\alpha$ . In the following sections the Image Registration, Candidates, Matching and Fusion of Scores stages are described in greater detail.

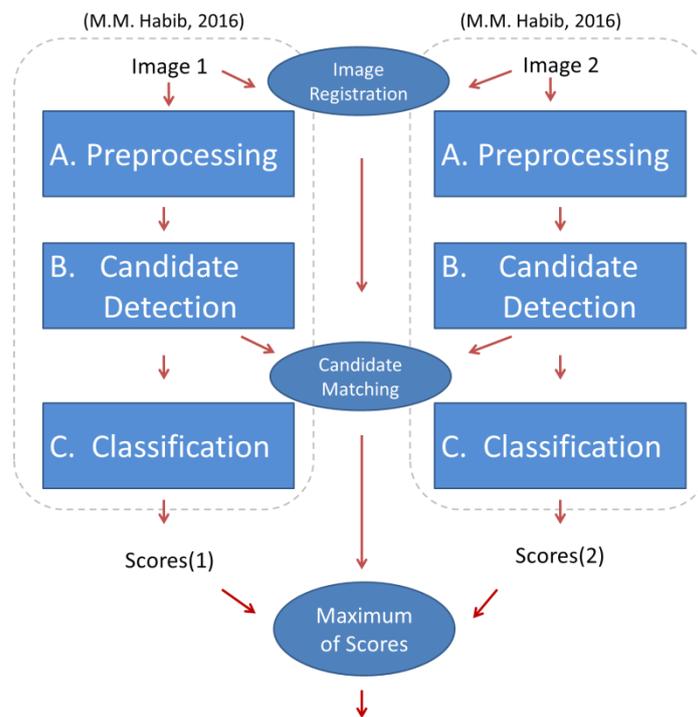


Figure 4.2. An overview of the proposed methodology.

#### 4.1.1 Image registration

Image registration is the process of aligning two images so that their corresponding pixels lie in the same space. One image is known as the *reference image* and the other image (known as the *moving image*) is transformed to be aligned to this reference image. A global transformation model was used which means that a single transformation was applied across the entire moving image. This has the advantage of simplicity and efficiency, but may not be as accurate as localised registration techniques. Since the goal of this chapter was to introduce a proof of concept with regards to combining spatial information during microaneurysm detection, the accuracy achieved was sufficient for this purpose. More accurate registration techniques will be investigated in future work.

A manual registration technique was employed where corresponding ‘control points’ are selected from each image. These control points are used to solve the global transformation model equations and find the transformation parameters (Figure 4.3). Corresponding points were annotated in each pair of images (as specified by Table 4.2).

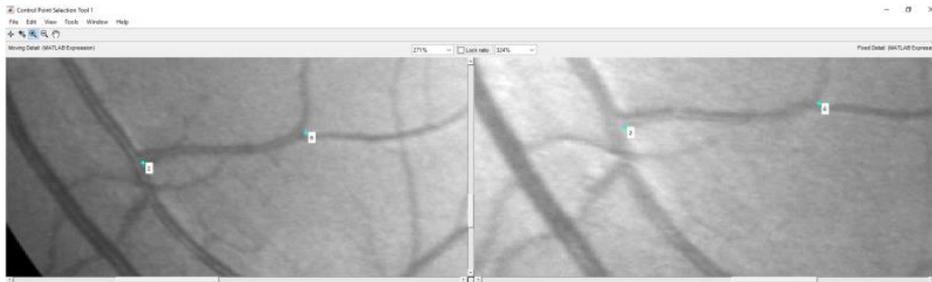


Figure 4.3. The manual control point selection process for the image registration phase.

A similarity model is selected as the global transformation model for alignment (justification for this was provided in Section 4.2.2). After the control points were selected, the transformation model was estimated by solving for the unknowns given the point correspondences. Finally, the moving image is warped onto the reference image to produce the final alignment. Once the images were aligned together, correspondences can be found between candidates detected in each image. This will be discussed next.

#### 4.1.2 Candidate Matching

Once both images were aligned, the candidates detected from both images lie in the same coordinate space and hence can be matched by their location. As will be indicated by the non-zero CEM values in Figure 4.11, it is expected that there will be some shifts in the aligned images. This will result in some of the MA candidates not being perfectly aligned in

the registered image pair. In order to account for some inaccuracies in the registration, the following method was used to find matches between 2 candidates:

Given two aligned images  $I_1$  and  $I_2$ , each candidate  $R$  detected in  $I_1$  needs had to be matched to one of the candidates detected in  $I_2$ . The centre of candidate  $R$  was found and a circular search region with radius  $r$  around  $R$  was defined. A match was made with the candidate in  $I_2$  whose center lay closest to  $R$ . If no candidate in  $I_2$  was found in this region, no match will be made. This procedure was repeated for all candidates in  $I_1$ . In our case  $r$  was defined to be 15 pixels which is twice the size of an average candidate in our dataset (Figure 4.4). This offers more tolerance to account for potential inaccuracies in the image registration.

More formally, let  $P \in \{p_1, p_2, \dots, p_n\}$  be the set of candidates detected from the image  $I_1$  and  $Q \in \{q_1, q_2, \dots, q_m\}$  be the set of candidates detected from image  $I_2$ . Our goal was to find a set of correspondences  $C = \{\{p_{r_1}, q_{s_1}\}, \{p_{r_2}, q_{s_2}\}, \dots, \{p_{r_l}, q_{s_l}\}\}$  (where  $r_i \in [1..n]$  and  $s_i \in [1..m]$ ) which represent the correspondences between these candidates. Note that some candidates would not have any correspondences and in this case, they would not be a member of any set pair in  $C$ . In practice either  $P$  or  $Q$  were chosen as a 'reference set' and matches were found from the other set. For instance, if we picked  $P$  as a reference then for each  $p_i$  we found a corresponding match from  $Q$  and added it to  $C$  if any existed. But we would not do vice versa –  $Q$  would not be used as a reference, and this was done for consistency – the goal is to find a one-to-one relationship between the matched candidates.

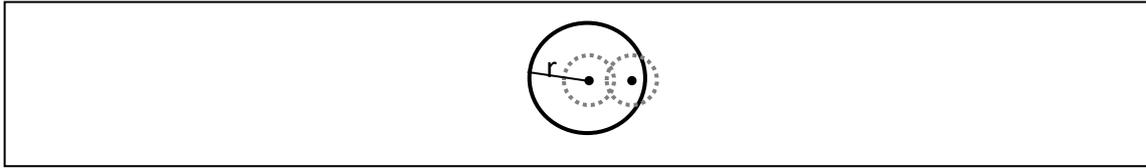


Figure 4.4. An illustration of the tolerance added while matching 2 candidates. The dashed circles represent two candidates from two views of the retina. It is assumed that they are misaligned due to registration inaccuracy. The solid circle represents a tolerance region around the first candidate and using this method a match is made.

Figure 4.5 shows examples of correspondences found after following the procedure above. In each cell of the figure, the first row(a, b) shows a colour image pair while the second row in the figure (c,d) represents the green channel extracted from each image in the first row. Note that the candidates are matched from the image on the right column (b, d) to the image on the left column (a, c). The annotation numbers represent the matches from P to Q (A visual representation of C). Candidates annotated with “-1” in the right image represent a candidate that has no correspondence in the other image (no match found in C). The blue circle in the figure represents a true candidate. In Figure 4.5(1) a match has been found between the true candidate in (b) and its corresponding candidate in (a). Furthermore, it was observed that the candidate has a much higher contrast in the right image than it does on the left one. The MA candidate was still visible in the left image but was much more subtle. Nevertheless, a combination of information from both candidates will give us higher confidence that is a true candidate. Figure 4.5(2) shows a vessel bifurcation being falsely detected in both views of the retina. This is one of the challenges of the candidate matching method that needs to be overcome. Figure 4.5(3) Shows a faded vessel segment that is falsely detected in both views. Figure 4.5(4) Shows a true candidate detected in both views.

A variability in contrast can be observed and hence the fusion of scores will be beneficial in this case.

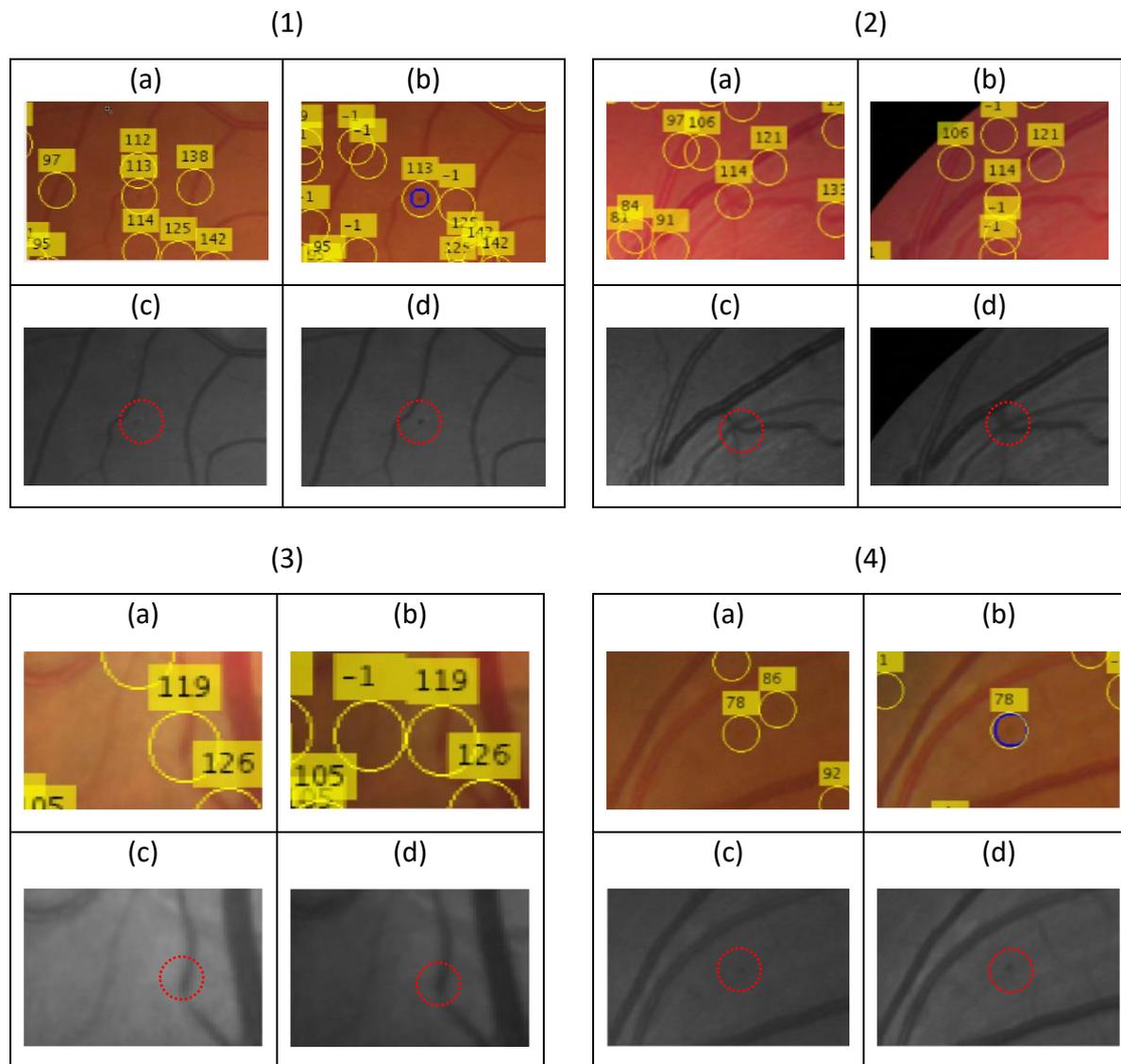


Figure 4.5. An example of the candidate detection result. In each cel: a) A colour image from an optic disc centered image. b) The fovea centered view of (a). (c) The green channel image of (a). d) the green channel image of (b). The numeric annotations on (b) represent the result of the matching operation with a). '-1' represents no match. 1) Candidate #113 is a true candidate and it has been correctly matched in both images (b). The candidate has variable contrasts in both images as can be seen in (c) and (d). The matching will hence improve the confidence regarding this candidate.2) A vessel bifurcation point being falsely detected in both views. 3) A faded vessel segment being falsely detected in both views. 4) A true candidate being correctly matched in both views. The candidates more subtle in the left column.

In fact, human retinal graders often switch between both views of the retina when they have suspicions regarding one candidate. The existence of signs in both images would give graders more confidence to classify an abnormality as a microaneurysm. The process of matching in the proposed method attempts to replicate this type of human behaviour.

#### 4.1.3 Scores fusion

Given that correspondences have been established between candidates and that each classifier has produced scores for each candidate, the final step is to find a fused set of scores that represent a combination of information from both images. Suppose that there are two matched candidates  $a$  and  $b$  from images  $I_1$  and  $I_2$  respectively. Furthermore, assume that  $I_2$  is the reference image –i.e. the current requirement is to classify the candidates in  $I_2$ . Define the function fuse as follows (Equation (4.1)):

$$\text{fuse}(a, b) = \begin{cases} \max(a, b) & \text{if } \beta_1 < b < \beta_2 \\ b, & \text{otherwise} \end{cases} \quad (4.1)$$

Where  $\beta_1$  and  $\beta_2$  are algorithm parameters specified between 0 and 1. In other words, given 2 matching candidates, the maximum of both their scores is taken only if  $b$  lies between the two threshold parameters ( $\beta_1$  and  $\beta_2$ ). The parameters  $\beta_1$  and  $\beta_2$  are used to limit the number of false candidates that get their scores maximized in the final set. This is because maximization of scores should only be done for true candidates. The fusing process is also illustrated in Figure 4.6.

It should be noted that performing fusing the scores using a mean or weighted mean operation is the most sensible in this case if we consider that the two scores belong to independent signals and the objective is to maximise the signal-to-noise ratio [62]. However, it was observed empirically that a majority of the detected candidates belong to

noise, and performing a mean in this case would decrease confidence in some of the true candidates. Therefore, performing a max operation would avoid this since in most cases the score of the true candidate is the higher among both scores.

If a candidate in a reference image has no match in the other image then its score is simply copied over to the fused set (Figure 4.6). Once the fused score set is computed a final threshold can be performed at an operating point  $\alpha$  to find the final set of classifications as described in Section 4.1.

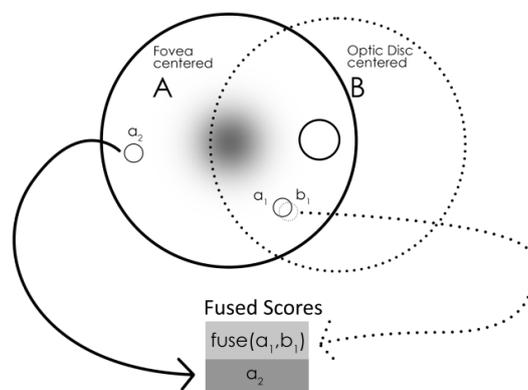


Figure 4.6. The method employed for the scores fusion phase.  $\beta_1$  and  $\beta_2$  are parameters set for the model.

## 4.2 Experimental Evaluation

This section reports on the evaluation results of the algorithm after performing experiments on a dataset which consists of spatial image pairs. The results of this chapter are also reported in [150].

### 4.2.1 Dataset

The dataset used for evaluation consists of 40 patients imaged as a part of UK NHS DESP. The set consists of retinal images for 40 patients. 16 images are available for each patient:

Fovea and macular centred images for 2 eyes over 2 years. In other words, there are 4 images per patient per year, and there are 8 images per patient. The total amount of images is 320 images, 160 images for each year. Figure 4.7 illustrates the set of images that are available for each patient. All the images were captured between 2012 and 2013. Research Governance approval was obtained. Images were pseudonymized, and no change in the clinical pathway occurred.

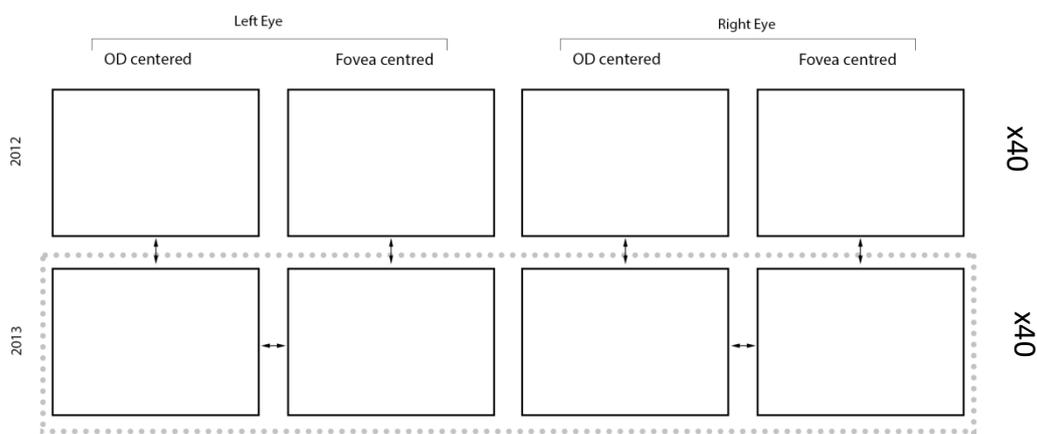


Figure 4.7 The images and views available for each patient in the longitudinal dataset. Key: OD centred - Optic disc centred image, Arrows: Images that were registered together, dotted line: Images that were groundtruthed for microaneurysms.

Table 4.1 shows a summary of the various image resolutions in the dataset. While the images were of different resolution, they all had the same aspect ratio of 1.5. Moreover, all the images had the same field of view (45 degrees). A professional grader (RW) was asked to ground truth the microaneurysms that were available in the images for the year 2013. Only the images of 2013 were ground truthed since the experiments have focused on the detection of microaneurysms in a single year. The main idea of this dataset was to provide a proof-of-concept that relying on features from multiple images will improve the performance of a retinal imaging algorithm. This could be either images from multiple views

(optic disc and fovea-centered), or images captured over time (same-view images captured a year apart). This concept, in essence, replicates what graders do in real-life situations where they refer to another view of the retina, or an image from the previous year, when they suspect the presence of a microaneurysm (or any other abnormality). The framework proposed could be extended to other features on retinal images (such as exudates or new vessels).

The professional grader was asked to highlight circular shapes around each microaneurysm [47] that were visible. These annotations were used to create the binary groundtruths that were used to assess the images. Furthermore, the grader was asked to use prior knowledge about multiple views (Fovea centered and OD-centred images) while performing the grading. In particular, the fovea region in the optic disk is usually out of focus and MAs are not clearly visible in this region of the image (they usually appear subtle). However, they are more visible in the fovea region of the fovea-centred image. Graders would normally refer to the OD-centered image to verify an MA candidate that they suspect exists in the fovea-centered image (since the MA would probably appear subtle in the other image). The same concept can also be applied to longitudinal images, where an MA appearing in a retinal image from a later date can be cross-validated from an image from an earlier date.

Hence, as explained above, the grader was asked to mark even those candidates that can be cross-validated from other images, even if they had appeared as subtle candidates. Note that many of these examples of subtle MAs would have been ignored by the grader if he were simply looking at a single image. However, these were specifically marked in order to demonstrate proof-of-concept examples regarding MA candidates that appear in multiple images.

Table 4.1 The distribution of image resolutions in the longitudinal dataset.

Image Resolution	MA Count (Year 2012)	Count (Year 2013)	Total count
3888 x 2592	32	24	56
2592 x 1728	12	48	60
3872 x 2592	0	12	12
4752 x 3168	44	32	76
4288 x 2848	8	20	28
3504 x 2336	48	24	72
2544 x 1696	16	0	16
<b>TOTAL</b>	<b>160</b>	<b>160</b>	<b>320</b>

For the sake of the spatial fusion experiments, only the images from the year 2013 were considered (160 images). An illustration of the images that were used for evaluation is shown in Figure 4.8. In the next section, the results of performing experiments on this dataset will be presented and discussed.

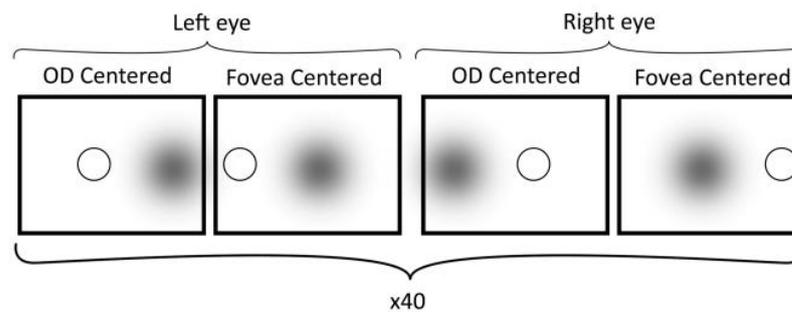


Figure 4.8. An illustration of the types of images available in the spatial dataset. OD centered – optic disc centered. Only images from 2013 were considered for the spatial experiment evaluation.

#### 4.2.2 Image Registration

Based on the literature summarised above, four transformation models were evaluated: Similarity [151], Affine [84], [152], Polynomial [102], [117] and RADIC [119], [153].

In the similarity model, 5 parameters need to be estimated (4 unknowns). These are 2 for translation, 2 for scaling and 1 for rotation. Each control point provides 2 correspondences (x and y) and therefore defines 2 equations. Since there are 5 unknown parameters, at least 3 control points are needed to solve the similarity equation defined in Equation (4.2):

$$\begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_{x_1} & 0 & t_{x_1} \\ 0 & s_{x_2} & t_{x_2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \quad (4.2)$$

The affine model requires 6 parameters need to be estimated since it adds an extra degree of freedom for shearing in addition to rotation. Therefore 3 points are needed:

$$\begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \\ 1 \end{bmatrix} = \begin{bmatrix} s_{x_1} & h_{x_1} & t_{x_1} \\ h_{x_2} & s_{x_2} & t_{x_2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \quad (4.3)$$

The RADIC model applies an affine model with an extra optimisation step in order to correct for radial distortions. The radial correction step is defined in Equation (4.4):

$$v = u \times (1 + k \cdot r^2) \quad (4.4)$$

Where  $u$  is the polar coordinate in the original space and  $v$  is the transformed coordinate,  $r$  is distance to the coordinate centre and  $k$  is a constant between  $-1$  and  $1$  that determines the amount and type of radial distortion that needs to be optimised [153].

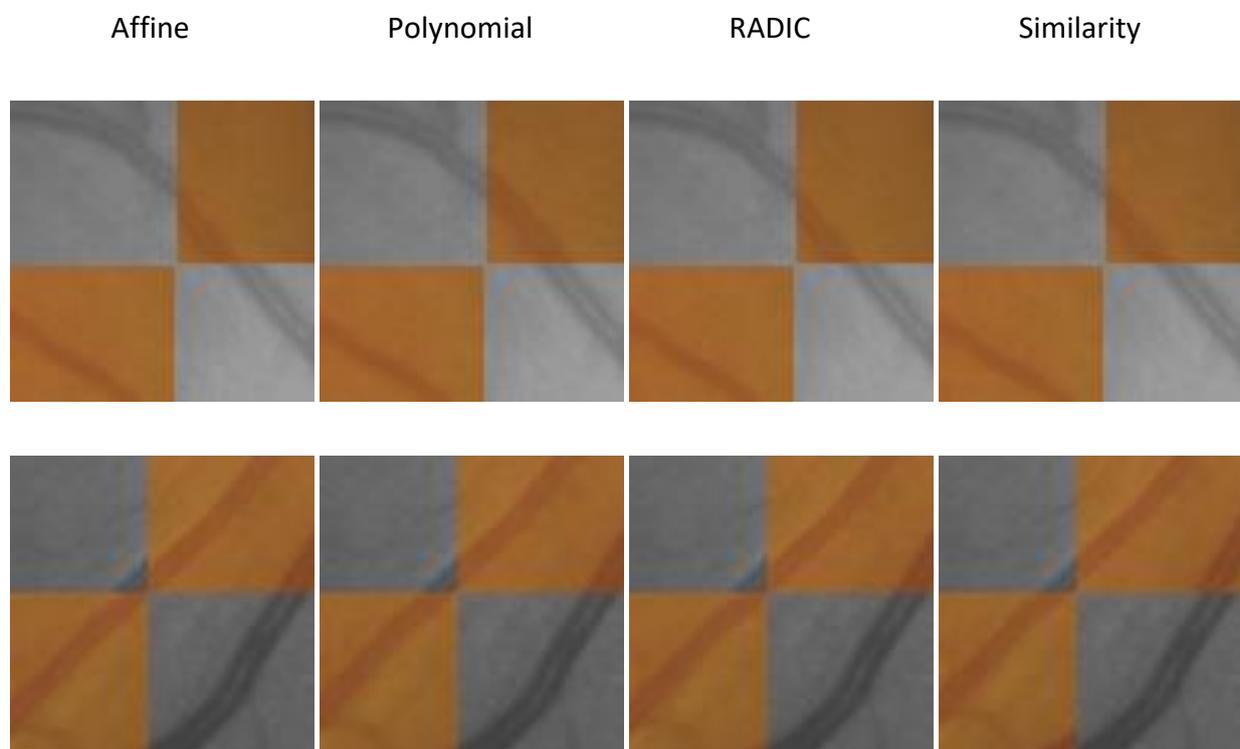
A polynomial model with degree 2 is estimated from the coordinates. 12 unknowns need to be found, therefore 6 control points are needed:

$$\begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \\ 1 \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & A_3 & A_4 & A_5 & A_6 \\ B_1 & B_2 & B_3 & B_4 & B_5 & B_6 \end{bmatrix} \begin{bmatrix} 1 \\ x_1^2 \\ x_2^2 \\ x_1 \\ x_2 \\ x_1 \times x_2 \end{bmatrix} \quad (4.5)$$

Where  $A_i$  and  $b_i; i \in [1..6]$ , are the 12 parameters that need to be estimated for the transformation.

The parameters of each transformation model were estimated using the six control points manually selected on each pair of images. These control points were selected on each image's vessel cross sections since it was easiest to identify corresponding points at these areas. Figure 4.9 shows samples of checkerboard patches selected at random from the registered image pairs. In general, it was difficult to identify the most accurately registered model by visual observation of the patches alone since there was an observed discrepancy in performance across rows.

None of the transformation models perfectly aligns the vessels in all four patches. Hence, a more objective method for selecting the model was needed.



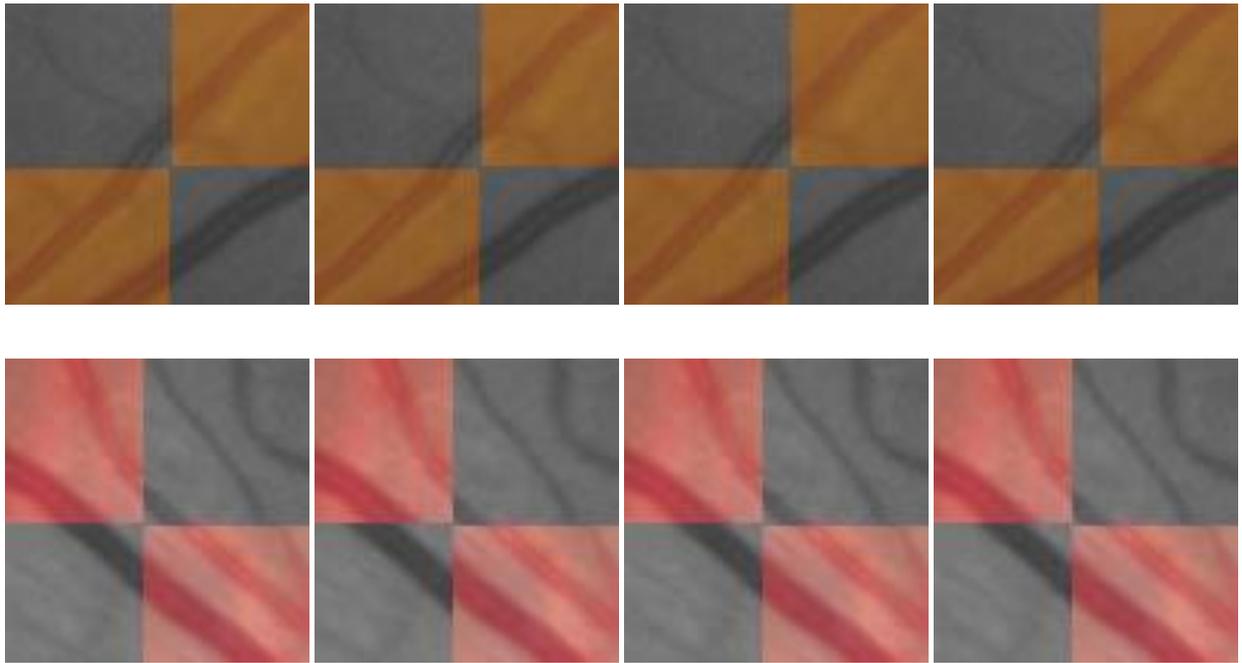


Figure 4.9. Sample checkerboard patches showing registration of multiple transformation models.

Table 4.2 The number of control points needed for each transformation model

Transformation model	No. of control points needed
Similarity	3
Affine	3
Polynomial	6
RADIC	3

As shown in Figure 4.9 it is difficult to decide which transformation model achieves best performance. Therefore, a more objective measure of registration performance is needed. During the registration, there was a reference image  $I_r$  and a moving image  $I_m$  which is transformed to be in the coordinate space of  $I_r$ . After the image was transformed to the coordinate space of  $I_r$  an overlapping region was defined as all the pixel coordinates where both  $I_r$  and  $I_m$  exist (overlap).

The Centreline Error Measure (CEM) [153] quantifies the mean of the minimum distance between each pixel along the centreline of the reference image and the closest pixel in the moving image. Given a set of coordinates in the reference image that lie on its vessel centreline (Figure 4.10) and are on the overlapping region of the two images:  $V=\{v_1, v_2, \dots, v_N; v_i \in (x, y)\}$ . Similarly, let  $U=\{u_1, u_2, \dots, u_L; u_j \in (x, y)\}$  be the set of points on the moving image that lie on its vessel centreline and belong to the overlapping region. Let  $t(p)$  be a transformation that transforms a point  $p$  from the moving image space to the reference space. The centreline error metric for a transformation  $t(p)$  on the moving image is computed as follows:

$$CEM = \frac{1}{L} \sum_{i=0}^{L-1} |M(\mathbf{u}_i) - t(\mathbf{u}_i)| \quad (4.6)$$

$$M(\mathbf{u}_i) = \underset{j=1..N}{\operatorname{argmin}} d(\mathbf{v}_j, t(\mathbf{u}_i)) \quad (4.7)$$

Where  $d(\mathbf{x}, \mathbf{y})$  represents the Euclidean distance between coordinates  $\mathbf{x}$  and  $\mathbf{y}$ . Therefore, the CEM calculates the average distance between each point on the reference image and the nearest point on the registered moving image (in the overlapping region).

A box and whisker plot of CEM values for each registration model in the same view was plotted in Figure 4.11. This plot is helpful to summarize the data since it shows the median value and the spread of the values (upper-quartile, lower quartile and the highest and lowest value). Based on this plot it was observed that while the polynomial model contains some of the lowest CEM values (highest accuracy) compared to other models, the distribution of its values also contains the highest spread (as can be seen by the height of the polynomial 'box', as well as the highest and lowest values). This can also be seen by the number of outliers in the polynomial plot. This high variance in values is also expressed by

the standard deviation of the values as can be seen in Table 4.3. This suggests the undesirable “instability” of the polynomial model. It is observed that the lowest standard deviation values are exhibited by both the affine and the similarity models (with the similarity model having a slightly lower standard deviation). Both the mean and standard deviations of these models were similar to each other which makes their performance comparable. The similarity model was selected since its values exhibited the lowest standard deviation. Some additional experiments to assess the performance of the Radial Distortion Correction (RADIC) Model are discussed below.

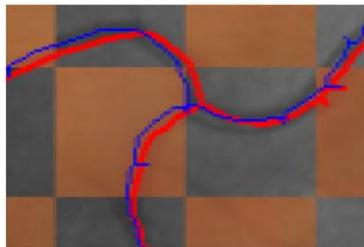


Figure 4.10 An example of overlapping vessel centerlines for computing Centerline error after registration. The distance between each red pixel and the nearest blue pixel is computed and the average of all distances is a measure of alignment accuracy.

Table 4.3 Centerline Error Metric (CEM) mean and standard deviation values for various transformation models.

	Similarity	Affine	Polynomial	RADIC
Mean	4.40	4.50	4.74	4.49
Standard Deviation	1.51	1.536	3.82	1.536

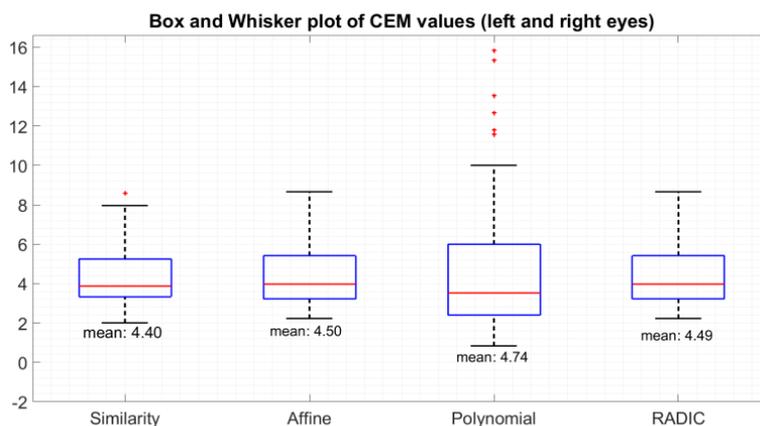


Figure 4.11. A Box and Whisker plot of the Centerline Error Measure (CEM) values.

As shown by the CEM values in Table 4.3, the accuracy of registrations for both Affine and RADIC were very similar. However, it is expected that the RADIC model should offer a better accuracy than the affine model as a result of the radial distortion correction step [153].

In order to affirm the results above, experiments were performed on fovea and optic nerve centred colour retinal images collected from 20 patients with Type 2 diabetes mellitus (with consent), attending medical retinal clinics over a 6 months period. Approval for image collection and analysis was obtained from the Institutional Review Board of the Moorfields Eye Hospital, and adhered to the tenets set forth in the Declaration of Helsinki. The images are anonymised and sample images are shown in Figure 4.12.

The images were captured with a TOPCON TRC 50IX camera. Pupils of the participants were dilated prior to image capture. The field of view (FOV) was 45 degrees.

The following procedure was used to compare several transformation models:

- 1- Image masks and centrelines were generated from the raw images. Visual inspection was used to ensure that the centrelines were accurate. Figure 4.12 shows an example image with these steps applied.

- 2- For each patient, control points in each pair of raw images were selected. This allows the registration of each pair of images for a given patient.
- 3- For each image pair, the following registrations were performed using the manually selected control points:
  - a) Similarity registration (least squares method)
  - b) Affine registration (least squares method)
  - c) 3b + radial correction
- 4- The results of the registration were visualised and inspected to ensure that the control points were accurate.
- 5- The Centreline error metric (CEM) was calculated for all the registrations in point 3. The centreline error metric is the average distance between each centreline pixel in the reference image and the closest centreline pixel in the target image

Point 3(c) above corresponds to the Radial Distortion Correction Model[153]; which includes an affine transformation, followed by a radial correction step. The registration pairs are selected such that each older image is registered with all the newer ones. For example, if we have images 1, 2, and 3 sorted oldest to newest, the registration pairs will be: (1,2), (1,3), (2,3). In the next section, the results of performing the experiments will be presented and discussed.

Figure 4.13 shows a sample of the results for a set of images from one participant. Visual inspection of the results shown in Figure 4.13 indicates that there is a slight improvement in the registration performance as a result of applying the radial correction (3c). The same pattern is observed for the rest of the 20 image sets as summarised in Table 4.4. A set of results relating to one set of images has been excluded from Table 1 (DM0013,

DM00016) because of the poor quality of the segmentation and centrelines generated from this set of images. This was due to significant abnormalities that were present on the retina). An overall average improvement of 1.5% in registration performance of the application of the radial correction (3c) is seen in comparison to the affine method (3b). This suggests that the radial distortion correction step is not adding a significant improvement in accuracy and therefore may be omitted in favour of faster computation speed.

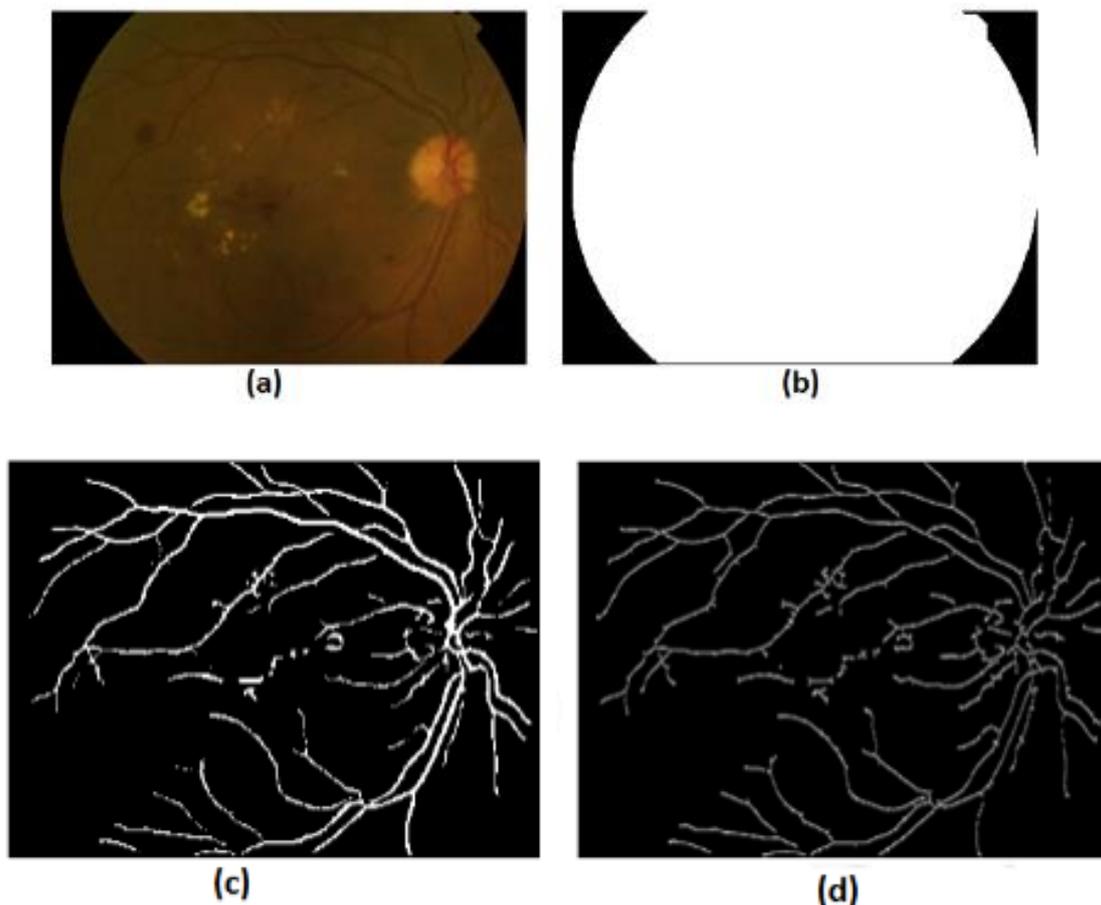


Figure 4.12 (a) Retinal image (b) Extraction of retina mask (c) vessel segments and (d) vessel centrelines (c).

It is important to emphasise that the images in our dataset have a large overlap of greater than 80%. In addition, since the images in the dataset were mostly macula centred, they were acquired from a similar viewpoint. This means that the radial distortion during image capture should be similar for most image pairs [153]. The results in Table 4.4 also show that

the similarity model performs well in many cases and sometimes slightly outperforms the affine model. This may be due to the non-existence of shearing distortion in the image dataset, making the similarity model useful in this case. This result supports the selection of the similarity model for registration which was performed according to the evaluation results earlier in this section.

An important remark is that these the radial correction is applied to only one of the images. However, since the radial distortion is a non-linear transformation, it needs to be corrected for each image independently [154]. This may explain the non-significance of the radial correction stage in the obtained results

Table 4.4 Confidence interval for each of the 20 image sets (mean  $\pm$  standard deviation). The third column represents the % improvement of 3c relative to 3b (RADIC compared to affine).

Image ID	Similarity (3a) CEM (pixels)	Affine (3b) CEM (pixels)	RADIC CEM improvement compared to Affine (3b) $\frac{ 3b - 3c }{3b} \times 100\%$
DM0001	14.51 $\pm$ 3.13	14.58 $\pm$ 3.38	0.81 $\pm$ 0.17
DM0002	8.64 $\pm$ 3.76	9.12 $\pm$ 3.39	1.21 $\pm$ 0.49
DM0003	8.76 $\pm$ 2.99	8.79 $\pm$ 2.94	2.06 $\pm$ 1.14
DM0004	4.64 $\pm$ 1.29	4.68 $\pm$ 1.37	6.44 $\pm$ 3.15
DM0005_OS	12.59 $\pm$ 4.44	12.65 $\pm$ 4.24	0.8 $\pm$ 0.55
DM0005_OD	12.61 $\pm$ 0.61	12.65 $\pm$ 0.77	0.52 $\pm$ 0.31
DM0006	8.16 $\pm$ 1.34	8.28 $\pm$ 1.3	1.07 $\pm$ 0.27
DM0007	7.53 $\pm$ 4.88	7.48 $\pm$ 4.94	3.33 $\pm$ 2.71
DM0008	8.26 $\pm$ 0.74	8.07 $\pm$ 0.89	1.03 $\pm$ 0.15
DM0009	14.55 $\pm$ 6.93	14.82 $\pm$ 6.97	0.79 $\pm$ 0.36
DM0010	11.32 $\pm$ 3.87	11.4 $\pm$ 3.6	0.93 $\pm$ 0.38
DM0011	6.01 $\pm$ 0.99	6.16 $\pm$ 1	3.14 $\pm$ 1.81
DM0012	35.89 $\pm$ 19.03	35.73 $\pm$ 19.06	0.11 $\pm$ 0.09
DM0014	7.12 $\pm$ 5.16	7.03 $\pm$ 5.06	0.9 $\pm$ 0.53
DM0015	7.28 $\pm$ 1.2	7.46 $\pm$ 1.05	1.56 $\pm$ 0.66
DM0017	65.4 $\pm$ 35.56	63.8 $\pm$ 33.91	0 $\pm$ 0
DM0018	11.95 $\pm$ 5.55	12.02 $\pm$ 5.45	1.03 $\pm$ 0.77

DM0019	9.26 ± 1.26	9.34 ± 1.25	0.06 ± 0.04
DM0020	8.76 ± 0.36	9.07 ± 0.47	1.92 ± 2.51
AVERAGE	13.855 ± 5.425	13.849 ± 5.317	1.457 ± 0.847

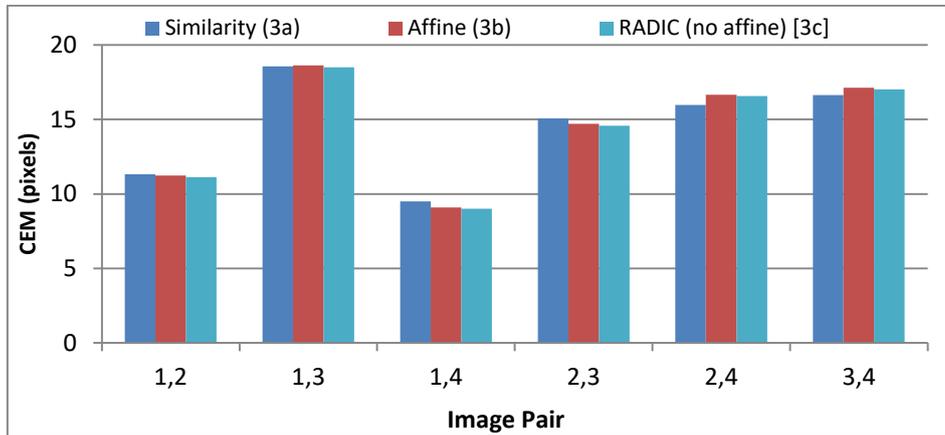


Figure 4.13 CEM values for image DM0001, where 3(a), 3(b) and 3(c) refer to the methodology steps mentioned above.

The effectiveness of radial correction for retinal image registration was investigated using longitudinal images from 20 patients with Type 2 diabetes mellitus. In order to assess the contribution of the *radial correction* step, it was applied separately and the CEM values were measured and compared with other lower-order models (similarity and affine). The results show that in the case of large overlaps and small FOV, it may be sufficient to use the affine or the similarity model, hence saving the computational time needed to optimise the extra radial correction parameters.

Experimental results show that although radial correction offers an improvement, it was small compared to the initial affine alignment accuracy. Importantly, the dataset used for the experiment had image pairs with large overlap (>80%). This suggests that if the image pairs have a large overlap, the radial correction step is less significant and it may be simpler to use the affine model or similarity model to achieve accurate enough registrations if

computational time and simplicity of method are important. Furthermore, the experiments performed earlier in this section on the spatial set of multi-view image pairs also affirm this result (on a smaller overlap).

### 4.2.3 Quantitative Results

The method for fusing the scores has been discussed in Section 4.1.3. In order to evaluate the effectiveness of this method a baseline method is needed for comparison. The baseline used in these experiments was the original proposed method that detects the candidates from a single image [130]. The goal was to provide a direct comparison between the performance when spatial information was accounted for and when it was not accounted for. The method used to compare [130] with the proposed method of this paper was as follows:

- 1- A tree ensemble model was generated using the training set. This was the same algorithm documented in Chapter 3. Candidates were generated for each image and then 70 features were computed for each candidate. The groundtruth was used to label the candidates and both features with labels are used to train the ensemble model. 80 images were used for training (Figure 4.7, 2013 only).
- 2- Features were generated from the validation set and the ensemble model (decision tree ensemble from point 1) was used to assign scores to each candidate in the 80 images of the validation set.
- 3- In the case of the original image that used single images, an FROC (Free-Receiver operating) curve [61] was generated using these scores alone. This is the solid curve in Figure 4.14.

- 4- To incorporate spatial information, each image pair (optic disc centred and fovea centred) had their corresponding candidate scores fused as explained in the methodology section. The fused scores were then evaluated *collectively* for each image pair and this was used to generate the FROC curve. This is the dotted curve in Figure 4.14.

The parameters for scores fusion were  $\beta_1 = 0.4$  and  $\beta_2 = 0.95$  (Equation (4.1)). Automating the process of selecting these parameters will require further analysis on a larger dataset and is left for future work. A slight increase in accuracy was observed after spatial information is incorporated. This increase can be captured quantitatively using the ROC score measure [47]. This score measures the sensitivity values at various x-axis intervals. The measured ROC score shows an increase of 0.02 after adding spatial information, which is a 2% increase. This increase can be explained intuitively as follows: Some candidates appear very subtle in the optic disc centred image, especially at the edge towards around the fovea region (Figure 4.5(1) and Figure 4.5(4)). This is because the retina is spherical and would get distorted during image acquisition. This distortion would affect the candidates towards the edge of the image. But these same candidates would appear more clearly in the Fovea centred image. This is because the same candidates now lie in the centre of the captured image, and hence their appearance will be obvious in the image. If an MA candidate appears more obvious visually, it is expected that the classifier should produce a higher score for this candidate. If the same candidate appears more subtle in another view of the retina, the classifier will assign a lower score to that candidate. However, when the scores are fused, the maximum score of both candidates is computed, and this will result in a higher score for

the more subtle MA candidate that was originally given a lower score. This is why the FROC curve for the fused candidates shows a better performance.

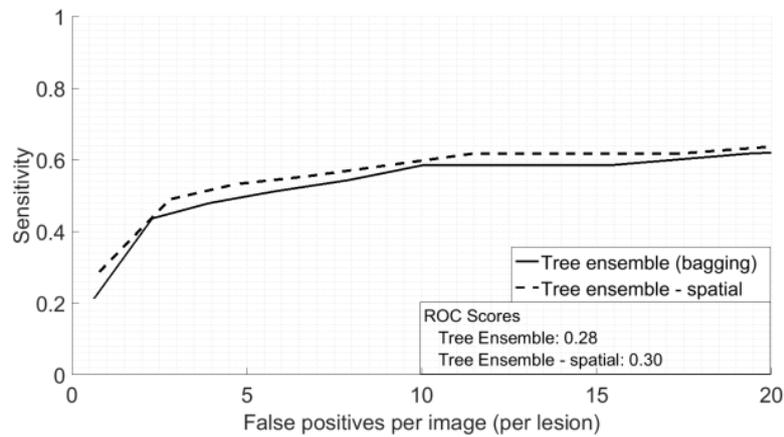


Figure 4.14 A comparison of performance between the technique applied to single images (solid) and after incorporating spatial information (dashed)

Figure 4.15 illustrates this intuitive concept by showing a plot of the corresponding pairs of scores. The figure shows a correspondence of scores for the test set where “score 1” and “score 2” on the axes refer to the methodology scores in Figure 4.2. Furthermore, true candidates are labelled in blue while false detections are labelled in orange. Let us assume Score 2 is the reference image and that the candidate scores are being matched to score 1. Since the set parameters are  $\beta_1 = 0.4$  and  $\beta_2 = 0.95$  the main interest is in the cross section of these values from the score 2 axis. By observing the extreme right region of the graph it is observed that some true candidates exist in this cross-sectional region. These receive higher scores along the x-axis. Hence, the maximum of both scores in the fused set will improve their scores and this will result in a higher FROC curve.

In order to place this improvement into context, Table 4.5 lists a summary of the number of candidates that could benefit the information combination. According to the table, 18% of the candidates contain correspondences in the training set, whereas 11% of the candidates contain correspondences in the test set. This means a maximum improvement of 11% is possible from this information combination. The observed 2% improvement is therefore an 18% improvement of the amount possible (11%).

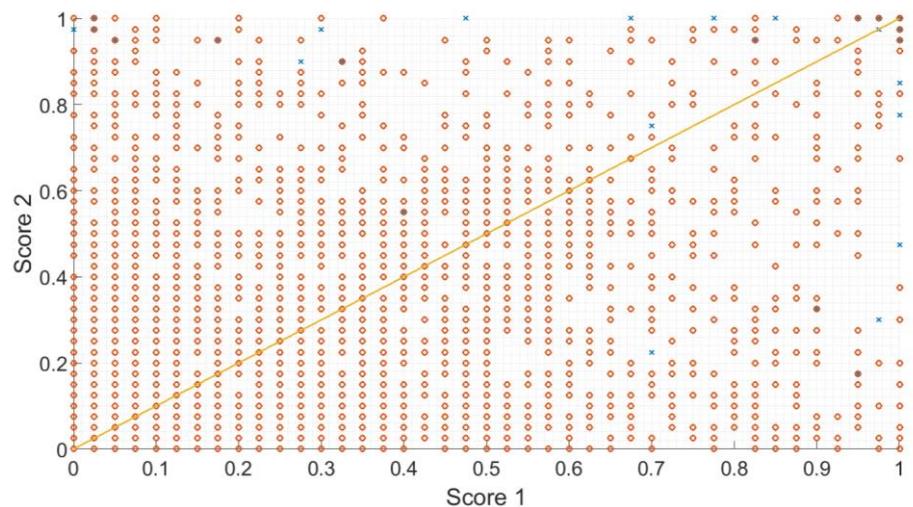


Figure 4.15 The scores from each corresponding spatial pair of images (these are the same scores as illustrated in Figure 4.2). True candidates are labelled in blue circles while false detections are labelled in orange crosses. The orange line represents the linear line for the equation  $x=y$ .

Table 4.5 Summary of the groundtruth candidates that contain matches in the corresponding spatial pair.

	Total Groundtruth candidates	Matched groundtruth candidates
Training set	112	21
Test set	96	11

The additional processing involved in the work described in this chapter is the time for Image Registration, Candidates Matching and Score Fusion. These are the blue ellipse stages in Figure 4.2. The image registration is dependent on the algorithm of choice. In this work

since it is based on manual control points the overhead is the control point selection and the time to solve a series of equations. An automated registration method may require an optimisation step and hence more time will be required for this step [106]. Given a spatial image pair, let  $c_1$  be the number of candidates detected in one image, and  $c_2$  be the number of candidates detected in the other image. The candidates matching step requires the calculation of the distance between each pair of candidates detected in both images and hence has a complexity of  $O(c_1 \times c_2)$  (using Big O notation). The Score Fusion phase will fuse the scores together from one image ( $c_1$  fusions) and then fuse the scores from the other image ( $m$  fusions) and therefore will have a complexity of  $O(c_1 + c_2)$ . In practice, on a core i5-4590 @ 3.30GHz CPU with 8GB RAM and an SSD hard drive, the time per image for each of the three stages was as follows:

1. Image Registration: 0.02 s
2. Candidates Matching: 1.30s
3. Scores fusion: 0.001s.

The sum of the above timings is 1.321s. The total time for the test dataset was computed and then the average value per image pair was found. The average time per pair for the entire process was 4.141s. This means that the average overhead is about 32% of the time required per image (1.321s out of 4.141s). This does not take into account the time for manual control point selection, however, automated registration will be implemented in future work. Methods for optimising the speeds of the other stages are also left for future work.

### 4.3 Conclusion

In this chapter, a novel algorithm that combines spatial information from two views of the same retina for the purpose of microaneurysm (MA) detection is proposed. As seen in the literature review chapter, a lot of the published work has been based on the detection of MAs from single retinal images. A new microaneurysm detection algorithm has been introduced in the previous chapter. In this chapter, the problem has been redefined to assume that two views of the retina are available and the objective is to use the information from both to detect the microaneurysms in both images simultaneously. The clinical application of the work done in this chapter would be its incorporation into diabetic eye screening programmes, thereby assisting the National Health Service in the detection of diabetic retinopathy. Screening in England has a recommendation of at least 2 images per eye so the assumption that two views are available would be valid in this context. Early Treatment Diabetic Retinopathy Study (ETDRS) grading [155] is based on a categorical variable. Accurate counts of abnormalities could lead to a more refined grading/risk prediction using a continuous variable, or more accurate estimates of MA turnover that may better predict progression of disease.

This chapter proposed a method for aligning the images, detecting candidates, matching candidates from both views and then combining the information from both views to perform microaneurysm detection from both views simultaneously. A semi-automatic registration method was used to align the images. Control points were manually selected from a private dataset of spatial image pairs. Six control points were selected in each image based on matches from the vessel bifurcation points. Four transformation models were assessed: Similarity, affine, radial distortion correction (RADIC) and polynomial.

Transformation parameters were estimated using the six control points and the images were aligned. The accuracy of the alignment was measured using the vessel centrelines in each image. The metric used to quantify the accuracy of registration is known as the centreline error metric (CEM). Based on this metric the similarity transformation model was selected since it was the most stable model and had the lowest CEM value (similarity – 4.4, affine – 4.5, polynomial – 4.74, RADIC – 4.49). Following alignment a method for candidates matching was introduced which considers potential misalignments that may occur during registration. The method adds a tolerance value during matching so that candidates are matched even if there is a slight shift between the aligned images. Finally, each pair of matched candidates is classified using by the scores of the candidates. The scores for each candidate are found using the ensemble classifier. The method for fusing scores takes the maximum of both scores if one of the scores lies between two thresholds ( $\beta_1$  and  $\beta_2$ ).

The proposed method was evaluated on a Diabetic Eye Screening Programme dataset of 160 images that contained 207 microaneurysms. The combination of information from multiple images was shown to increase the performance in comparison to the ensemble algorithm applied to a single image. An account of the computational overhead required for the combination of information was presented. Some of the limitations of the work in this chapter are: 1) The computational overhead required for the additional steps of the proposed method, 2) The registration process is semi-automatic and requires manual control points selection. 3) The registration alignment is not accurate enough. As a future work, the scores fusion stage can be further improved to reduce more false positives and increase true positive detections. An improvement in performance compared to detection from an algorithm that relies on a single image (Chapter 3) is shown as an increase of 2%

ROC score, hence demonstrating the potential of this method. A threshold parameter needs to be selected manually to decide which candidates should be fused with correspondences with the other image (based on the classification score). The results of this chapter are also reported in [150].

In the next chapter, another dimension of information combination in terms of temporal images will be explored. These will be images captured for the same patient over time. Information about the progression of the MA candidates and the possibility to achieve more accurate detection results using this additional information will be explored.

# 5

## MICROANEURYSM DETECTION USING TEMPORAL INFORMATION

---

Diabetes patients have images of their retinas captured regularly. This means that multiple retinal images will be available for the patients that have their image captured over multiple intervals of time. These images will be of the same view and therefore they will have a large overlap between each image. Aligning these images is expected to be less challenging due to the larger overlap between them compared to spatial image pairs. Also, more correspondences between candidates can be found due to the larger overlap. The difference, however, is that microaneurysms may have developed during the time interval between consecutive imaging appointments.

In this chapter, a similar technique to that described in the previous chapter will be applied to a temporal set of images. The problem being addressed in this chapter is: given an image captured in the current time interval, is it possible to incorporate information from another image captured at a past time interval in order to improve the detection accuracy of the current image?

While there are similarities between the spatial and temporal aspects of the problem, there are also several differences that need to be highlighted. Firstly, there will be greater variation in the appearance of the microaneurysm since it develops overtime. The difference in variation overtime. This variation will depend on the health progression of the patient. For some patients, a microaneurysm may develop into a haemorrhage, in which case it will no longer be classified as a microaneurysm. Furthermore, progression of diabetic retinopathy can cause an increase in the number of microaneurysms present [156].

The second difference is that the overlap between images is also quite large. In the case of temporal images of the same view the overlap will be almost 100%. While this exposes opportunities to match candidates together, it is also likely to cause more false matches with noise.

These differences may require a modification to the framework to adapt to the new requirements. The proposed method of fusing candidate scores may not be sufficient in this case, and it may be required to explore a better way to assign likelihoods to matched candidates. The methodology section will cover the techniques used to align and combine information from the temporal images, while highlighting the main differences from the experiments performed on the spatial techniques. The discussion section will present an analysis of the experimental results. Final concluding remarks will be mentioned in section 5.3.

## **5.1 Methodology**

The method applied to temporal images is similar to the method applied to spatial images. However, combining information from temporal images differs from combining spatial information as follows:

- 1- The appearance of microaneurysms in temporal images could have changed. Some microaneurysms might have disappeared. Others could have developed further. Others may have stayed the same. This is not the case in spatial images.
- 2- The overlap in temporal images is much larger than a spatial image pair. The images are almost fully overlapped.
- 3- Different kinds of distortions may have occurred in a temporal image pair, since the camera types may have differed and also the lenses used.

The change in appearance of microaneurysms can be a challenging aspect when comparing two candidates in the temporal case. For example, a candidate may have not existed in the past but appeared in the current image. Comparison with candidates in the previous image could lead to matching the candidate with another false positive or noise. This is an undesirable effect that should be avoided, since it could cause more false positives to appear in the future image.

The dataset used for evaluation is derived from the same dataset that was described in Section 3.2.1. The following subsections will cover the phases of image registration and candidate matching. A discussion of a suitable technique for the fusion of scores in temporal images will also be presented. Concluding remarks will be mentioned in section 5.3.

### **5.1.1 Image Registration**

As discussed in Section 4.1.1, Image registration algorithms that have been proposed for Retinal images have been summarized in Section 2.2. Early work on retinal image registration [99], [84], [85], [100],[86],[101] has used a lower-order transformation model (similarity or affine model) to perform the registration. Since the eye is approximately spherical in shape, a higher order quadratic model is more appropriate to estimate the

distortion caused during the process of retinal imaging (intuitively) [102]. However, since the retina is almost planar on small fields of view (such as fundus photographs, where the field of view captured is around 45 degrees of the retina), these registration methods produce modest errors in such scenarios [102]. The majority of the proposed feature-based methods have relied on vessel bifurcation and crossover points.

The temporal images were registered in the same way spatial images were registered (Section 4.1.1). Each image from 2013 was registered to the corresponding image from 2012 (details about the dataset are described in Section 5.2.2). 6 markers were manually inserted at the bifurcation points. Inserting the markers was a more simple task for the longitudinal images due to the larger overlap in these image and so finding a suitable bifurcation or crossover point was a more straightforward process.

With respect to the temporal images contained in the dataset there were 160 pairs of images to be registered, in contrast to 80 registration pairs in the spatial case (In the spatial case there was one registration performed per eye, corresponding to the 2 views of each eye. Since there are 2 eyes x 40 patients, 80 registrations were required in total. In the temporal case, each image was to be registered to the previous year image of the same view. In total there are 40 patients x 2 eyes x 2 views per eye, and therefore the number of registrations for the temporal case was 160). similarity transformation model parameters are solved using the six control point pairs in each image to perform the registration. Details about the selection of the model, and control points are provided in Section 5.2.2.

### **5.1.2 Candidate Matching**

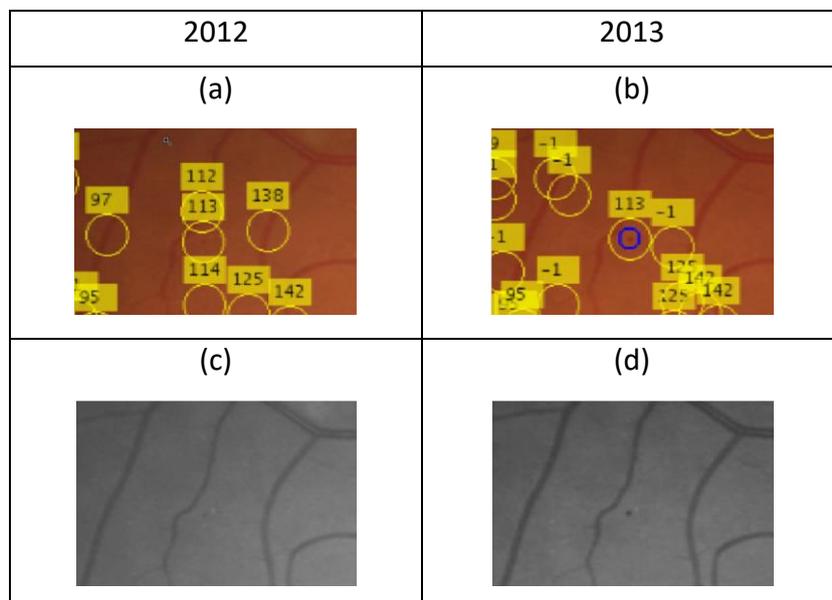
A similar procedure to that of the spatial technique was used for the Candidate Matching phase (Section 4.1.2). After the images are aligned they will both belong to the same

coordinate space. In other words, the coordinates on the transformed image will correspond to the coordinates on the base image. As will be seen in Table 5.1, some inaccuracy of registration is expected and therefore the candidates will not be perfectly aligned. As will be seen in Section 5.2.2 there may be a shift in the range of  $4.65 \pm 2.93$ px in the coordinates which will cause slight misalignments for corresponding candidates. In order to account for these inaccuracies a tolerance is added when searching for a corresponding candidate. This is the same tolerance as in the spatial case since the registration accuracies are similar for both cases (comparing Table 4.3 and Table 5.1).

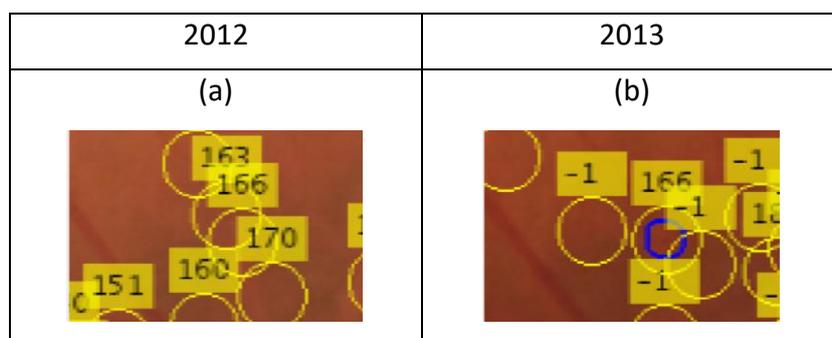
Figure 5.1 shows examples of candidate matches that have been found during this phase. In each subfigure the first row in each subfigure (a, b) shows a colour image pair while the second row (c, d) in the represents the green channel extracted from each image in the first row. Candidates are matched from the image on the right column (b, d) to the image on the left column (a, c). The annotation numbers (yellow labels) represent the matches from the right image to the left image. Candidates annotated with “-1” in the right image represent a candidate that has no correspondence in the other image. The blue circle in the figure represents a true candidate. In Figure 5.1(1) there is a candidate that appears well in 2013 that was not visible in 2012 (#113). This particular candidate is more visible in 2013 and therefore there will not be a big advantage from looking back to the previous candidate in this case. Figure 5.1(2) shows a similar example of a microaneurysm that has further developed over a year (#166), but in this case the microaneurysm is more subtle even in 2013. There may be an advantage of looking back to the 2012 image, since the classifier confidence may not be very high in this case and hence finding a corresponding candidate in the past will increase our confidence after the classifier scores are fused. Figure 5.1(3)

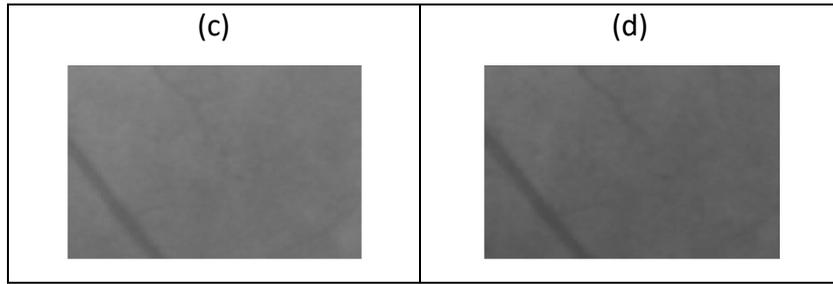
shows an example of a candidate that was not detected in the future image and therefore it will go undetected even after fusion. It was not possible to find an example of a microaneurysm that has been faded out in a future image within the dataset (clearly visible in 2012, but subtle in 2013). Many of the dataset images were manually inspected to search for such examples but none were found. It is likely that such an example does not exist in this dataset.

(1)



(2)





(3)

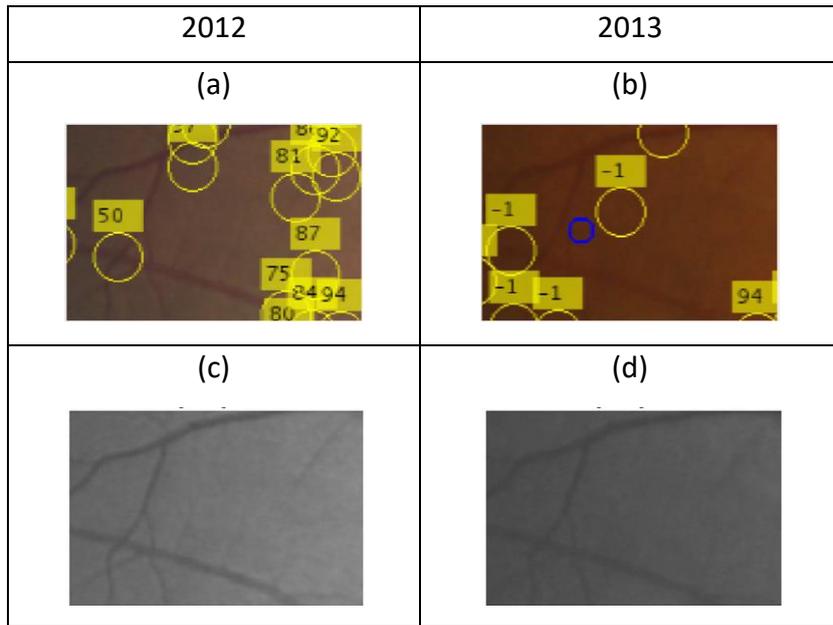


Figure 5.1 Examples of candidates matching from 2013 to 2012. The yellow circles represent the search region around the centre of a detected candidate. The labels on the top-right image represent which label in the top-left image was matched. A label of “-1” in the top-right image means that there was no match from the image in 2012. A blue circle represents a groundtruthed image. 1) A candidate that is visible and detected in 2013 appears with less contrast in 2012 2) A candidate with similar appearance is detected in both images 3) A subtle candidate that is close to a vessel is not detected in both years and therefore will not benefit from the proposed approach.

### 5.1.3 Combining feature vectors

In the previous chapter, the concept of candidate matching followed by fusion of scores was explored. This methodology may not produce desirable results in the temporal case due to the differences discussed in the introduction section. In order to tackle this problem, an alternative approach is explored to detect the patterns between the features of matched

candidates more effectively. The evaluation section will apply both the scores fusion and feature extension methods in order to compare the performances of both frameworks on the temporal set.

Rather than combining the scores after the ensemble classification step of each image pair, a different approach would be to combine the feature vectors together before performing the classification. Figure 5.2 shows the overview of the stages if feature combination was performed. After the candidates in the image pair are matched, the feature vectors for each candidate are computed and corresponding feature vectors are extended to form a 140-dimensional feature vector (70 features from each candidate).

These features are used to both train the model and classify the candidates. As a result, only one classifier is required and at the same time it will depend on features generated from 2 corresponding candidates. Figure 5.3 illustrates this idea. The illustration displays two different views of the retina, but the same concept is applicable to similar views of the retina as is the case with the temporal dataset (in fact it will be more effective if the overlap between images increases). An important question is: What needs to be done when the candidate has no matches in the other image.

The feature vector can be extended if a candidate is matched, but this will leave unmatched candidates with 70 missing features in their corresponding feature vectors (since there is no corresponding candidate from the previous year). In order to account for this, the missing data needs to be handled prior to classification. Several approaches have been proposed in the literature for the handling of missing data in machine learning models. Some of these approaches are specific to certain classifiers such as a modification of an Support Vector Machine (SVM) kernels [157], while others are based on principled approaches for

interpolation such as the Expectation Maximisation (EM) algorithm [158]. Some methods proposed for Random Forests (which are similar to tree ensembles) include replacing each missing variable with the class median if it is non-categorical, and with the most frequent class value if it is categorical [159].

It was empirically observed that a majority of the detected candidates will have no matches in the previous years. For example, in the temporal dataset, 83 percent of the candidates had no matches from the previous year. In other words, 83 percent of the candidates will have half of their feature vectors missing. It is also likely that when a true candidate pair is found, the two candidates' feature values will lie in close proximity (especially true for the spatial case). In light of this, the proposed idea is to extend the feature vector by duplicating the same candidate features again. Another idea is to train 2 models: 1 for matched candidates and the other for candidates with no matches. In the following we explore the idea of extending the feature vectors by feature duplication.

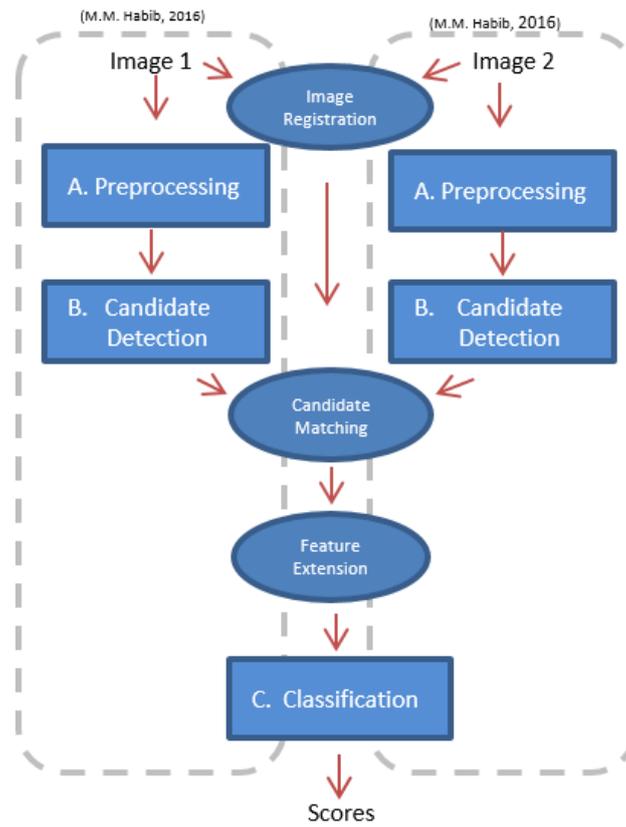


Figure 5.2 Proposed framework for combining features prior classification.

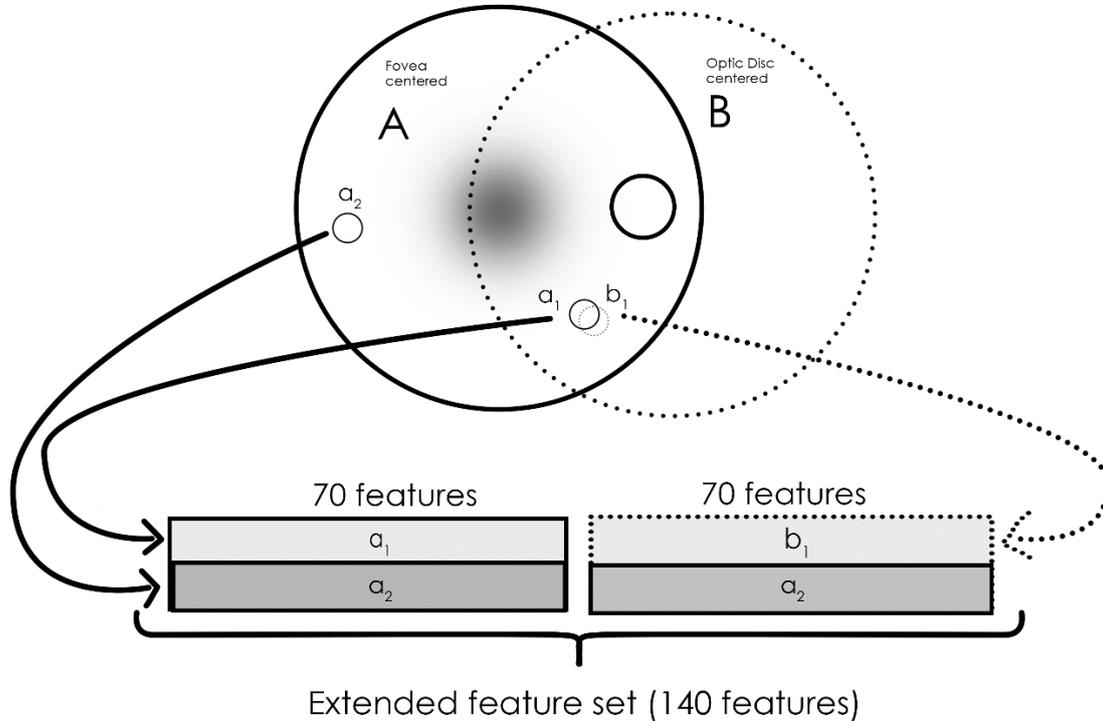


Figure 5.3 An illustration of combining features vector ("Feature Extension" in Figure 5.2) by extending matched candidates ( $a_1, b_1$ ) and duplicating the feature vectors of candidates that have no matches ( $a_2$ ).

## 5.2 Experimental Evaluation

### 5.2.1 Dataset

The same dataset used in Section 4.2.1 was relied on for evaluation purposes, however, the temporal images were considered in this case. As illustrated in Figure 5.4, the dataset used for evaluation consists of 40 patients imaged as a part of UK NHS DESP. The set consists of retinal images for 40 patients. 16 images are available for each patient: Fovea and macular centred images for 2 eyes over 2 years. Hence there are 4 images per patient per year, and there are 8 images per patient. The total amount of images is 320 images (160 images per year).

In the temporal case, the objective is to detect the microaneurysm captured in 2013 provided that information from the 2012 image is available. The basic performance is measured by using the ensemble algorithm to detect candidates in the 2013 alone. The incorporation of temporal information is performed by aligning the 2013 and 2012 image pair and then incorporating the 2012 information in the detection process. The objective is to improve the accuracy of the 2013 image after taking into consideration the image from 2012. The improvement in accuracy is measured by comparing the performance with the ensemble algorithm on 2013 alone. The reliance on the 2012 images and the split between training and testing is illustrated in Figure 5.4. The left side of the figure shows the training phase where the first 80 images are used to generate the training features and generate the ensemble model. The second right half of the figure shows the validation set where candidates are detected from the 2013 set and classified using the classifier. The baseline performance is measured using the 2013 images alone. The performance of the temporal information method is measured by matching the candidates from 2013 to 2012 and then

fusing the classifier score of each candidate pair together. The same ensemble model is used to produce the classifier scores in 2012 as well as 2013.

2013 (x160)	(training set, 20 patients) 1 .. 80	(validation set, 20 patients) 81 .. 160
2012 (x160)	This part of the dataset is ignored	Candidates detected here and matched to 2013 set for validation

Figure 5.4 An illustration of the temporal dataset and how it was used to evaluate the temporal framework.

## 5.2.2 Image Registration

In the literature related to retinal image registration, four transformation models were proposed: Similarity [149], Affine [84], [150], Polynomial [102], [117] and RADIC [119], [151]. The performance of each transformation model was assessed on the temporal dataset of image pairs (Section 5.2.1).

The parameters of each transformation model were estimated using the six control points manually selected on each pair of images. These control points were picked on each image's vessel cross sections since it was easiest to identify corresponding points at these areas. Figure 5.5 shows samples of checkerboard patches selected at random from the registered image pairs. In general, it was difficult to identify the most accurately registered model by

visual observation of the patches alone since there was an observed discrepancy in performance across rows.

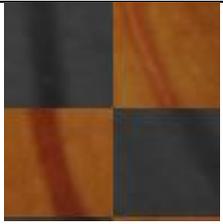
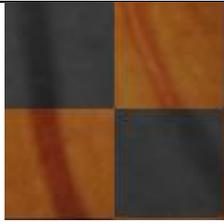
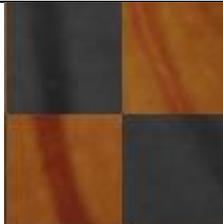
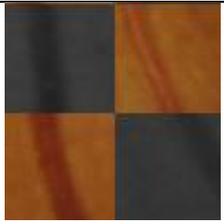
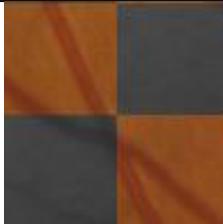
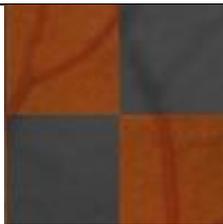
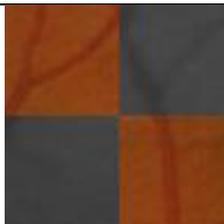
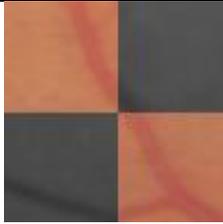
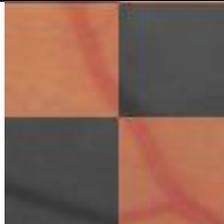
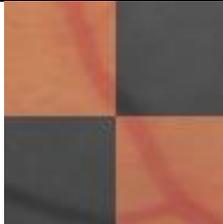
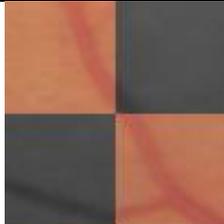
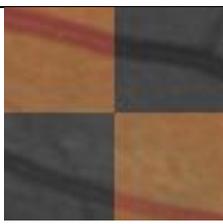
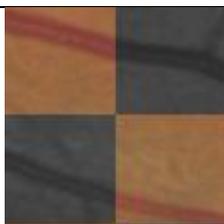
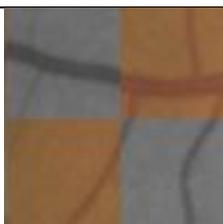
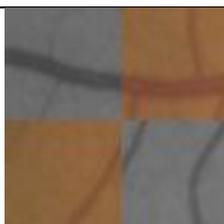
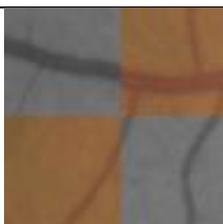
Affine	Polynomial	RADIC	Similarity
			
			
			
			
			
			

Figure 5.5 Examples of checkerboard patches which show other registration results for different transformation models.

In order to provide a more objective assessment of the results the centreline error metric (CEM) for each pair of registrations were computed. Table 5.1 shows the mean and standard deviations for the CEM values for the 4 transformation models. It is observed that the similarity and the affine models achieve similar CEM values on average. Another observation is that the values are close to the evaluation results of the spatial registration evaluation (Table 4.2). Except for the quadratic model, the rest of the transformation models were at most 0.6 pixels between spatial and temporal results on average. This observation is concluded by comparing the mean values in Table 4.3 and Table 5.1. The quadratic model was 1 pixel apart compared to the spatial accuracy results. Figure 5.6 shows the box and whisker plots of the CEM values for all 160 images. The outliers in the box and whisker plot have been ignored since it was visually observed that the outliers were due to inaccurate image registrations Figure 5.7 shows an example of inaccurate vessel segmentation. It is observed that parts of the vessel segments have not been detected and this would affect the CEM value. Inaccurate vessel segmentation occurs due to poor image quality and illumination which causes some vessel segments to be left undetected. The box and whisker plots verify that the distribution of CEM values for both Similarity and Affine models are lower on average and also less scattered than their Polynomial and RADIC counterparts (lower standard deviation). Since the similarity model only requires 3 control points and is slightly more efficient to solve, the similarity model is chosen as the global transformation model for these images.

Table 5.1 Centerline Error Metric (CEM) mean and standard deviation values for various transformation models

	Similarity	Affine	Polynomial	RADIC
Mean	4.65	4.66	5.81	4.90
Standard deviation	2.93	2.94	3.87	3.23

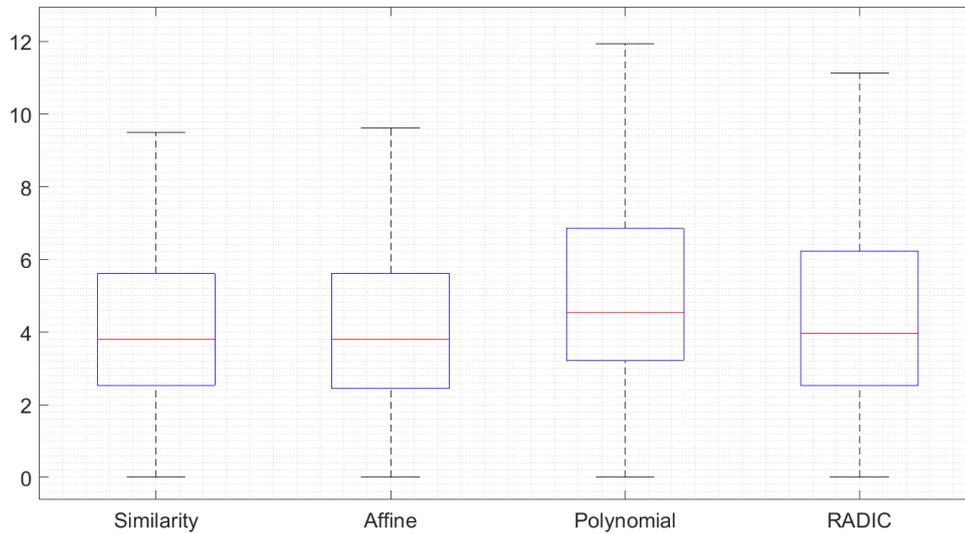


Figure 5.6 Box and whisker plot of the CEM values for 160 registrations of temporal images. Outliers are not shown in the graph since it was visually verified that they are due to an inaccurate vessel segmentation while computing the CEM values (Figure 5.7).

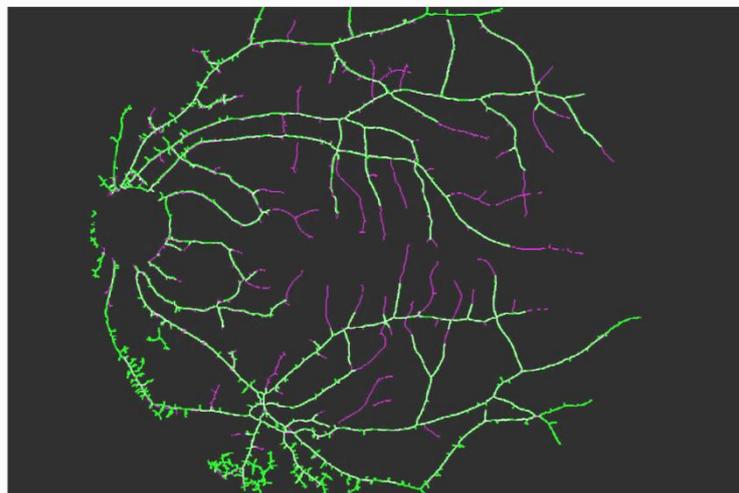


Figure 5.7 An example of an outlier results for the CEM values. It can be observed that the high value of CEM is caused due to inaccurate automatic segmentation. The purple pixels represent vessels that are not aligned with the corresponding vessels in the other pair. While the alignment is very good in this case, the CEM values will be high as a result of inaccurate segmentation.

### 5.2.3 Quantitative Results

In the case of combining spatial information the method to combine information from candidates was clear: computing the maximum score of both candidates. As intuitively explained earlier, candidates in different spatial images are more visible in one image than the other and therefore taking the maximum score will increase our confidence, and also the accuracy achieved by the algorithm. More importantly, in spatial images if one candidate is relatively clear, the corresponding candidate will also appear somewhat clear. This is not the case for temporal images where there is more diversity in the appearance of candidates from one image to another. This is due to the development that occurs to microaneurysms over time, and the multiple possible variations that could occur in the temporal case. In this section, the possibilities for various ways to fuse together scores from temporal candidates will be explored. Furthermore, other ways to improve detection accuracy in the temporal case will also be discussed. An alternative idea for the framework will be explored in Section 5.2.4.

Figure 5.8 shows a plot of corresponding scores of the candidates after performing the matching step. The parameters used during the scores fusion stage were  $\beta_1 = 0.8$  and  $\beta_2 = 1.0$  (Equation (4.1)). In this plot true candidates are represented by blue crosses while false positives are represented by orange circles. As discussed earlier, for the maximum of scores to work, blue circles need to exist at the right side of the  $x = y$  line. Furthermore, these blue circles should be separable from the other false positives, since the goal is to increase the confidence in true positives while keeping the false positives unaffected. As can be observed in the plot, there are a handful of true positives that appear in the upper-right of the graph (blue circles). However, they cannot be separated from the false positives that

are surrounding them using a straightforward thresholding technique. These cases consist of candidates that are subtle in 2013, but were clear in the 2012 image. Figure 5.9 shows the FROC curves for both the ensemble classifier (solid line) and the inclusion of temporal information (dotted line). The FROC curve shows that there is not an improvement in ROC scores (the ROC score without using temporal information is 0.1786 and the comparative ROC score that includes temporal information is 0.163). The explanation above related to the scores of the candidates explains why this is the case. Figure 5.9 shows example patches of a variety of cases to demonstrate which cases are causing the framework to see a lack of improvement. The first row in each cell represents the colour image in 2012 and 2013 respectively. The second row represents the grey channel of the colour image. In Figure 5.9(1) and Figure 5.9(2) a true candidate has been matched to a previous candidate and therefore the scores will be correctly fused. In Figure 5.9(3), due to a mismatch of registration a candidate has been misaligned and a neighboring candidate was matched with the 2012 candidate. This can cause a false positive to be matched to a true positive which will affect the results negatively after the scores are fused. In Figure 5.9(4) a false positive has been correctly matched with another false positive with a higher score. The fusion of false positive scores is one of the main reasons for lack of accuracy improvement as a result of this method.

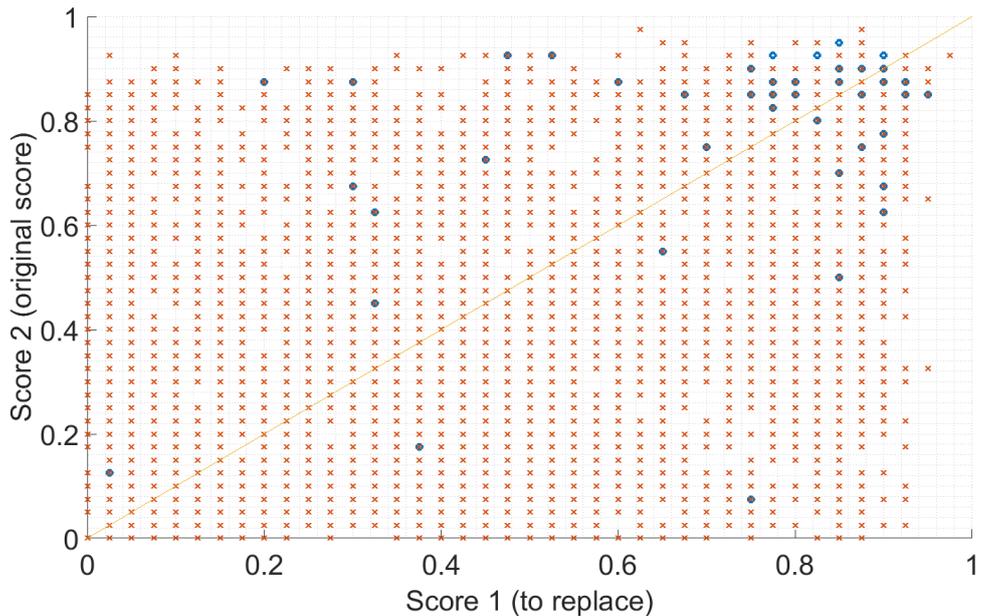


Figure 5.8 scatter plot of the probabilities for the spatial image pairs. Score 2 represents the score of the candidates being classified (2013 image) while Score 1 axis represents the scores that are matched to score 2 (2012 image). Based on the plot above, the threshold has been chosen to be between 0.18 and 0.3.

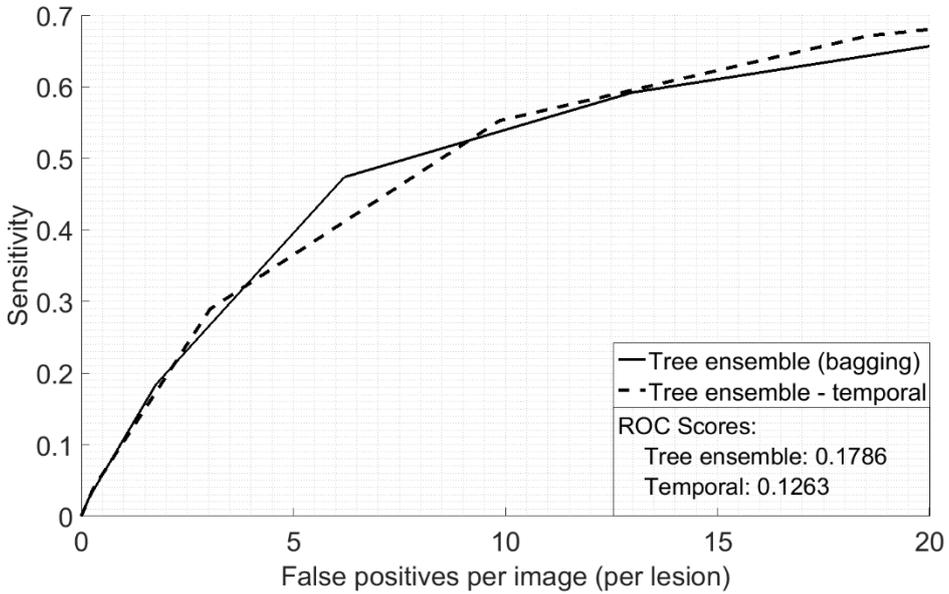


Figure 5.9 FROC plots for the original ensemble applied to a single image. The other FROC curve is applied to the image of interest and then matched with the nearest corresponding candidate. The maximum of both scores is taken if the original candidate score (score 2) lies between the range of 0.18-0.3

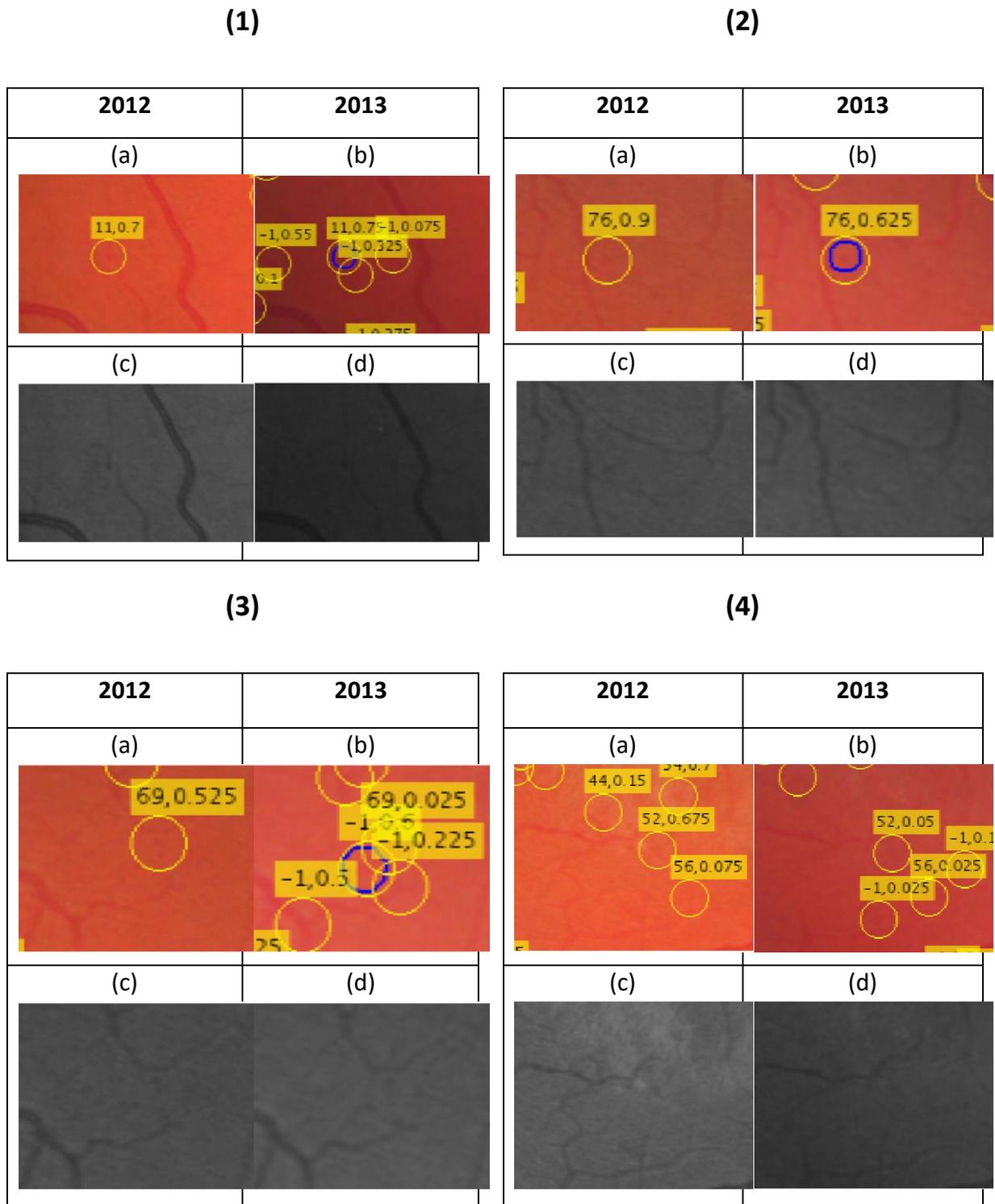


Figure 5.10 An illustration of several patches after candidate matching and assigning scores. In each patch of 2012 (a) the annotations correspond to the candidate number followed by the classifier score. In each annotation of (b) the first number corresponds to the matched candidate from 2012 ('-1' if no match exists) followed by the classifier score of the candidate). In (1) and (2) a true candidate has been matched to a previous candidate and therefore the scores will be correctly fused. In (3), due to a mismatch of registration a candidate has been misaligned and a neighboring candidate was matched with the 2012 candidate. This can cause a false positive to be matched to a true positive which will affect the results negatively after the scores are fused. In (4) a false positive has been correctly matched with another false positive with a higher score. This is the main reason for the lack of improvement of this method.

#### 5.2.4 Combining Feature Vectors

Experiments were performed to assess the performance of an ensemble model trained on an extended feature set compared to an ensemble model trained with features from a single image. The experiments were performed using the same dataset introduced in Section 5.2.1. The ensemble algorithm from Chapter 3 was trained on the 2013 training set and evaluated using the 2013 validation set. The performance from the ensemble algorithm was used as a baseline accuracy.

For generating the extended feature set, each candidate from the training set in 2013 was matched to a corresponding candidate from the 2012 image of the same view. If a matching candidate was found, the feature vector was extended using the computed features from the matched candidate. If no match is found, the feature vector was duplicated. Let  $f_c$  be the feature vector of the candidate from 2013, and  $f_m$  be the matched feature from 2012 (if one exists). The extended feature  $f_e$  is defined as in Equation (5.1):

$$f_e = \begin{cases} [f_c & f_m]; & \text{if a match exists} \\ [f_c & f_c]; & \text{if no match exists} \end{cases} \quad (5.1)$$

Once the extended feature set was generated it was used to train the model. After training, the same process was repeated to generate the extended feature set from the validation set and then classify each candidate. This produces another accuracy set which was compared with the baseline accuracy. Figure 5.11 shows the result of performing the aforementioned experiments. There is a slight improvement observed after feature duplication. The value of the ROC Score increases from 0.2966 to 0.2994. Note that due to the randomness of the ensemble classifier, there is an observed variability in these measures for every run.

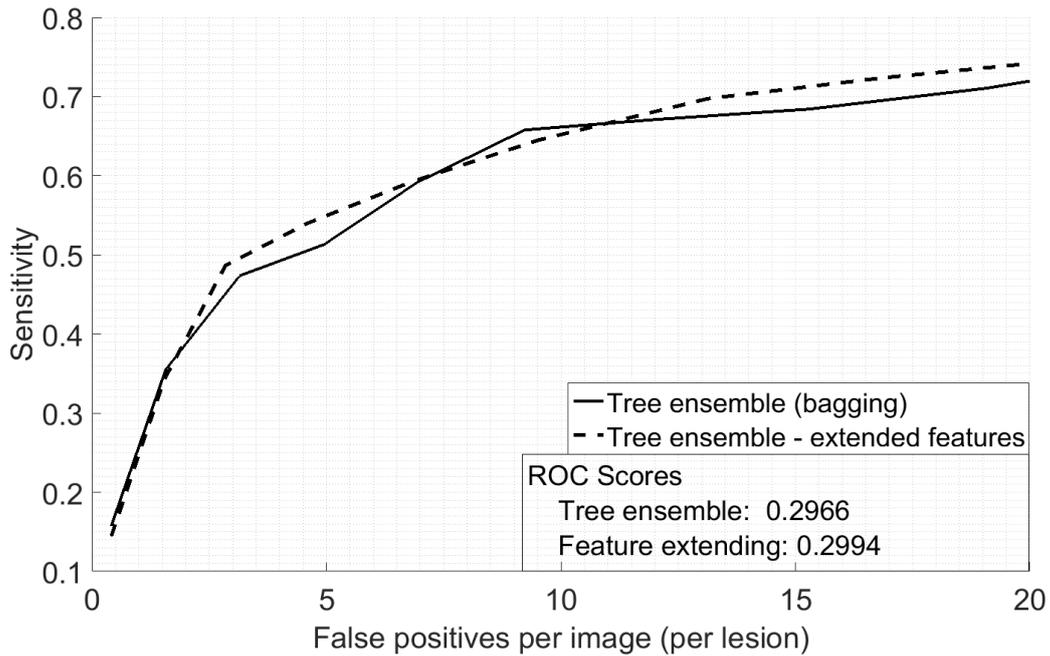


Figure 5.11 FROC curve for the feature extension on the temporal dimension compared with the proposed ensemble algorithm

Extending the feature vector has shown that it is a promising approach for combining the information at the classifier level. One of its drawbacks is that it is difficult to interpret the results after combining the feature vectors and as a result it is difficult to decide on more effective ways to combine the feature vectors. It is not understood whether combining the feature vectors has assisted in eliminating more false positives, detecting more true positives, or both. Furthermore, there are other proposed methods in the literature that deal with missing data and can be employed [157], [158]. Furthermore, the way that the features have been combined could be further improved by including a subset of features from the previous images. This can be selected based on a feature ranking approach after the full feature vector is trained. In summary, the feature extension approach has shown promising results but there are many factors that still need to be explored in order to optimize this approach for combining information.

### 5.3 Conclusion

This chapter has extended the ideas introduced in Chapter 4 (spatial information combination) to consider temporal combination of information from candidates. The combination of information from temporal images has the potential to improve the accuracy of detection while at the same time providing doctors with insights about the progression of the disease over time. However, some challenges need to be overcome, such as the wide variability of microaneurysm appearance across time. Microaneurysms can develop over time and therefore their appearance in the images will differ.

A semi-automatic registration method was used to align the images. Control points were manually selected from a private dataset of temporal image pairs. Six control points were selected in each image based on matches from the vessel bifurcation points. Each image pair consisted of an image from 2013 and a corresponding image from 2012. The images were of the same field of view (optic-disc centred or macula centred). Four transformation models were assessed: Similarity, affine, radial distortion correction (RADIC) and polynomial. Transformation parameters were estimated using the six control points and the images were aligned. The accuracy of the alignment was measured using the vessel centrelines in each image. The metric used to quantify the accuracy of registration is known as the centreline error metric (CEM). Based on this metric the similarity transformation model was selected since it was the most stable model and had the lowest CEM value (similarity – 4.65, affine – 4.66, polynomial – 5.81, RADIC – 4.90). After the alignment, the candidates from one image need to be matched to the candidates from another image. A method for candidate matching was introduced which takes into account potential misalignments that may occur during registration. The method adds a tolerance value

during matching so that candidates are matched even if there is a slight shift between the aligned images. Finally, each pair of matched candidates is classified using by fusing the scores of the candidates. The scores for each candidate were found using the ensemble classifier. The method for fusing scores takes the maximum of both scores if one of the scores lies between two thresholds ( $\beta_1$  and  $\beta_2$ ). Unlike the spatial case, there was no improvement in the temporal case after the fusion of scores. Visually it was observed that this was due to false positives being matched with other false positives in the previous years which lead to an increase in their score after taking the maximum. An alternative approach for combining information has been proposed. This includes the extension of the feature vectors for each pair of matched candidates. Preliminary experiments showed that extending the feature vectors requires further investigation but is a promising direction to pursue.

This chapter applied a similar methodology introduced in Chapter 4 to a temporal dataset of images. The methodology section has covered the techniques that were used to align and combine information from temporal images. This chapter has highlighted the main differences between this chapter and the methodology introduced in Chapter 4. The scores of candidates were analysed and discussed. Some ideas regarding the combination of candidates were also presented in this section.

# 6

## **SUMMARY AND CONCLUSIONS**

---

This chapter will summarize the contributions and results of this work and present the conclusions of this work. The current trends and future directions of the field will be discussed.

### **6.1 Summary of Results**

The contributions of this work are as follows:

- 1- Development of a new MA detection algorithm that uses an ensemble classifier to detect candidates from a *single* colour fundus retinal image. The ensemble classifier gave us a flexible framework to extend the analysis from a single image to more than one.
- 2- Development of a novel framework (based on the ensemble algorithm in 1) for combining information from image pairs based on a pair of retinal images that contain an overlap.

The framework developed in 2 was used to combine information both spatially and temporally. A summary of each of the main contributions for each point is presented below.

### **6.1.1 Microaneurysm Detection Using an Ensemble Classifier**

A new method for microaneurysm detection in colour fundus images has been proposed. The proposed algorithm consists of three phases: 1) preprocessing 2) MA Candidate Detection and 3) MA Candidate Classification. The purpose of the preprocessing stage is the removal of non-uniform illumination using a background subtraction technique. Preprocessing is applied to the green channel of the image. A 3x3 median filter is used to remove salt-and-pepper noise. Histogram equalization is used for contrast enhancement. Finally, shade correction is performed by subtracting a background estimate from the original image. During the Candidates Detection phase an initial set of candidates was found from the image. The first step was to remove blood vessels since the bifurcation points cause many false detections. A morphological bottom-hat operation is used to remove blood vessels, while keeping the microaneurysm structures present. A Gaussian matched filter is used to enhance the responses of microaneurysm candidates. Simple thresholding followed by a region growing approach (to enhance the candidates' shapes) was used to find the initial set of candidates. A tree ensemble classifier was used to filter out the initial set of candidates and reduce the number of false positives. 70 features were extracted from each candidate and these were used by a trained classifier model to predict the likelihood of the candidate being a true detection (probability between 0 and 1). The final set of candidates was predicted by thresholding this score at a given operating point.

The proposed method was evaluated on three datasets (DIARETDB1, ROC dataset and a subset of the MESSIDOR dataset) and has shown improvement in performance when compared to another state of the art method. For lesion-based performance, the proposed method has achieved Retinopathy Online Challenge (ROC) scores of 0.3923, 0.2109 and

0.1523 in the MESSIDOR, DIARETDB1 and ROC datasets respectively. In contrast, the comparative technique, Fleming, achieved ROC scores of 0.2479, 0.1143 and 0.0373 respectively. Therefore, the proposed method achieves higher scores for MESSIDOR and DIARETDB1, but a lower score for the ROC dataset.

For pixel-based performance, the proposed method has achieved Areas Under Curves (AUCs) of 0.1971, 0.1133 and 0.0775 in the MESSIDOR, DIARETDB1 and ROC datasets respectively. In contrast, Fleming achieved AUCs of 0.1704, 0.1207 and 0.0717 respectively. Therefore, the proposed method achieves higher AUCs in case of the MESSIDOR and ROC datasets, while the comparative technique achieves a higher AUC in the DIARETDB1 set.

### **6.1.2 A Framework for Combining Information from Retinal Images**

A novel framework for combining information from multiple retinal images has been proposed. The majority of the proposed literature has addressed the problem of MA detection from single retinal images. It is argued that multiple spatial or temporal views of the retina are usually captured for each eye but the existing algorithms are not designed to make use of these multiple views. In England, screening programs have a recommendation of at least 2 images per eye. In addition, temporal retinal images for each patient would also be available since these diabetes patients attend regular eye screening programs.

The proposed framework includes alignment of the images followed by combining the information from both images to perform detection in both images simultaneously. A semi-automatic registration method was used to align the images. Control points were manually selected from a private dataset of spatial image pairs. A global similarity transformation model was selected based on experiments performed on a longitudinal dataset which

contains both spatial and temporal views of the retina. The unknowns of the transformation model were solved using the control points that were selected manually.

Once the images were aligned, candidates were detected on each image. The proposed ensemble algorithm was used to detect the candidates in each image in the registered pair. A method for candidates matching was introduced which takes into account potential misalignments that may occur during the registration process. The method builds in a tolerance value during matching so that candidates are matched even if there is a misalignment between them. Finally, each pair of matched candidates is classified by fusing the scores of the corresponding candidate pair. The scores for each candidate were found using an ensemble classifier. The method for fusing scores computed the maximum of both scores if one of the scores was found to be within a certain range of values. This is to ensure that there is enough confidence in one of the candidates before fusing its score with the other candidate (and limits the number of false positives that are incorrectly matched).

The proposed method was evaluated on a Diabetic Retinopathy Eye Screening Programme dataset of 320 images that contained 207 microaneurysms. The dataset consists of images for 40 patients captured over 2 time intervals (1 year apart). For each patient, 4 images were captured (2 eyes x 2 spatial views – optic disc centred and macula centered). The images were captured once in 2012 and once in 2013. This dataset was used to evaluate the proposed framework in both the spatial and temporal dimensions. For both dimensions, the ensemble algorithm that operates on single images (Section 6.1.1) was used as a baseline for comparison.

In the spatial case, the combination of information from multiple images was shown to increase the accuracy by 2% in comparison to the ensemble algorithm applied to a single image.

In the temporal case, there was no improvement after incorporating information from the temporal dimension. This has been explained by: 1) The variability of appearance of microaneurysms. Since the images were captured one year apart, developments will have occurred to the microaneurysms 2) Variability of noise along time due to various factors (change of camera, different imaging conditions, different clinician etc.). This variability could cause more noise to be falsely matched together and prevent an improvement in accuracy. Visual examples of where candidates have been incorrectly matched were illustrated.

Visually it was observed that this occurred due to false positives being matched with other false positives in the previous years which lead to an increase in their score after taking the maximum. An alternative approach for combining information has been proposed. This includes the extension of the feature vectors for each pair of matched candidates. Preliminary experiments showed that extending the feature vectors is a promising approach that could be developed further.

The limitations of the proposed framework include: 1) the computational overhead required for the additional steps of the proposed method, 2) the registration process is semi-automatic and requires manual control points selection. 3) the registration alignment is limited with respect to accuracy. Some sources of computational overhead by the proposed framework are: 1) alignment of the images, 2) matching the candidates, 3) fusion of scores.

## 6.2 Conclusion

The human retina is a window that reveals important information about retinal, ophthalmic and systematic diseases such as diabetes and hypertension. Several indicators manifest themselves as early signs of diabetic retinopathy. These signs include microaneurysms, haemorrhages and exudates, and can be visualised in fundus images compared to other state-of-the-art techniques. The early detection of these signs is important from a clinical perspective in order to suggest appropriate treatment for diabetes patients. This work improved the detection accuracy of microaneurysms in colour fundus images. Microaneurysms appear at varying contrasts and some can be very subtle which makes them similar in appearance to image noise. Proposed algorithms have not taken into consideration that multiple fundus images are available for patients. These include spatial (multiple views of the same retina) and temporal (multiple images captured at different time intervals) images. A novel algorithm for microaneurysm detection was proposed which uses an ensemble classifier and 70 features to perform classification of an initial set of detected candidates.

A framework was developed based on the ensemble algorithm which is capable of combining information from multiple images. This framework was used to make use of spatial information from 2 views of the retina (the optic-disc centred view and the macular-centred view). These views are commonly captured and consist of overlapping regions. Graders often rely on both views when performing diagnosis of patients. It is shown that an improvement in results can be achieved by combining the information from multiple views. The proposed algorithm extends the ensemble method by performing the detection using information from two images at once. A spatial image pair is first registered so that the

individual features are aligned together. Following alignment, the proposed ensemble algorithm is used to detect potential candidates in each image pair. Each candidate is assigned a probability score by the tree ensemble classifier. Each candidate in one image is matched to the corresponding candidate in the aligned image pair. The scores of each matched pair are fused if a certain criterion is satisfied, and a final score is produced for all of the candidates. Performing a fusion of scores in this way has increased the accuracy of detection by 2% based on the evaluation results.

The proposed framework was also used to combine information from a temporal pair of images (multiple images of the retina that are captured at different time intervals). It was observed that in the temporal case the microaneurysms change in appearance and develop over time. Some microaneurysms may get more visible and other could get more faint. Due to these factors there was no observed improvement after using the framework. An alternative approach for combining information have been proposed. This includes the extension of the feature vectors for each pair of matched candidates. Preliminary experiments showed that extending the feature vectors requires further investigation but is a promising direction to pursue.

### **6.3 Future Work**

This work has explored avenues for combining information from multiple images in order to improve the accuracy of microaneurysm detection in colour fundus images. Both the spatial and temporal dimensions of the problem have been explored. The future directions can be divided into spatial, temporal, and other ideas. These will be explored below.

### **6.3.1 Spatial Dimension**

The spatial algorithm proposed has shown an increase of 2% in the accuracy of microaneurysm detection. As a future work, the scores fusion stage can be further improved to reduce more false positives and increase true positive detections. The selection of the threshold parameter can be automated. The use of an automated registration algorithm will save the time consumed for manual control point selection.

On the other hand, other ideas for the combination of information can be explored. While the proposed algorithm has combined the scores produced by the classifier, other approaches to combine information prior to the classification phase can be explored. For instance, can the feature set be extended with features from both matched candidates together? Can the initial candidates from both images be detected simultaneously with an extended version of the Gaussian filter? These are avenues that can be explored and can potentially help reach higher detection accuracy.

### **6.3.2 Temporal Dimension**

The combination of information from temporal images has proven to be more challenging than spatial images due to the variability of appearance in microaneurysms as they develop overtime. However there is still a lot of potential information to be gained from the combination of information from multiple detected candidates as it will reveal information about the progression of individual microaneurysms overtime. Furthermore, the combination of information in this way can alert graders not only to the locations of microaneurysms, but also which ones are developing further (into haemorrhages or exudates), and which ones are disappearing overtime. A study of further ways to

incorporate information from temporal images is planned. The ideas and directions explored here are similar to the ones proposed in the previous section.

### **6.3.3 Other Avenues**

In the introduction chapter, it was seen that there are many modalities that exist to capture imagery from the human retina. These include fundus images, adaptive optics imaging, PEEK retina, Optic Coherence tomography (OCT) and fluorescence imagery. Each imaging modality has a specific clinical application and reveals a particular aspect about the retina. Chances are that a patient will have many of these modalities captured and the combination of information across modalities is an avenue that can reveal insights to doctors about certain abnormalities, and potentially also improve the accuracy of symptom detection. Future research will attempt to explore combining information from multiple modalities for the purpose of improving the accuracy of detection as well as visualising abnormalities across multiple modalities.

## REFERENCES

- [1] “Facts About Diabetic Eye Disease | National Eye Institute.” [Online]. Available: <https://nei.nih.gov/health/diabetic/retinopathy>. [Accessed: 11-Apr-2017].
- [2] M. M. Fraz, “Retinal image segmentation and quantification of vessel width in non-standard retinal datasets,” Ph.D., Kingston University, 2013.
- [3] E. Decencière *et al.*, “FEEDBACK ON A PUBLICLY DISTRIBUTED IMAGE DATABASE: THE MESSIDOR DATABASE,” *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, Aug. 2014.
- [4] N. Patton *et al.*, “Retinal image analysis: concepts, applications and potential,” *Progress in retinal and eye research*, vol. 25, no. 1, pp. 99–127, 2006.
- [5] M. M. Fraz *et al.*, “An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 9, pp. 2538–2548, Sep. 2012.
- [6] A. Sopharak, K. T. Nwe, Y. A. Moe, M. N. Dailey, B. Uyyanonvara, and S. Barman, “Automatic exudate detection with a support vector machine classifier,” 2008.
- [7] M. U. Akram, S. Khalid, and S. A. Khan, “Identification and classification of microaneurysms for early detection of diabetic retinopathy,” *Pattern Recognition*, vol. 46, no. 1, pp. 107–116, Jan. 2013.
- [8] R. A. Welikala, “Automated detection of proliferative diabetic retinopathy from retinal images,” phd, Kingston University, 2014.
- [9] M. D. Abramoff, M. K. Garvin, and M. Sonka, “Retinal imaging and image analysis,” *Biomedical Engineering, IEEE Reviews in*, vol. 3, pp. 169–208, 2010.
- [10] R. Bernardes, P. Serranho, and C. Lobo, “Digital Ocular Fundus Imaging: A Review,” *Ophthalmologica*, vol. 226, no. 4, pp. 161–181, 2011.
- [11] “Grading Diabetic Retinopathy from Stereoscopic Color Fundus Photographs—An Extension of the Modified Airlie House Classification,” *Ophthalmology*, vol. 98, no. 5, pp. 786–806, May 1991.
- [12] M. A. Pérez, B. B. Bruce, N. J. Newman, and V. Biousse, “The use of retinal photography in non-ophthalmic settings and its potential for neurology,” *Neurologist*, vol. 18, no. 6, pp. 350–355, Nov. 2012.
- [13] S. Dithmar and F. G. Holz, *Fluorescence angiography in ophthalmology*. Springer Science & Business Media, 2008.
- [14] C. M. Klais, M. D. Ober, and L. A. Yannuzzi, “Indocyanine Green Angiography: General Aspects and Interpretation,” in *Retinal Angiography and Optical Coherence Tomography*, J. F. A. M. FACS, Ed. Springer New York, 2009, pp. 43–59.
- [15] J. OLSON, “Stereo Atlas of Fluorescein and Indocyanine Green Angiography.,” *Br J Ophthalmol*, vol. 84, no. 7, p. 805, Jul. 2000.
- [16] D. J. Mordant *et al.*, “Spectral imaging of the retina,” *Eye (Lond)*, vol. 25, no. 3, pp. 309–320, Mar. 2011.

- [17] I. Alabboud *et al.*, “New spectral imaging techniques for blood oximetry in the retina,” in *Novel Optical Instrumentation for Biomedical Applications III (2007)*, paper 6631\_22, 2007, p. 6631\_22.
- [18] K. Ghasemi Falavarjani, K. Wang, J. Khadamy, and S. R. Sadda, “Ultra-wide-field imaging in diabetic retinopathy; an overview,” *J Curr Ophthalmol*, vol. 28, no. 2, pp. 57–60, Apr. 2016.
- [19] G. A. Williams, I. U. Scott, J. A. Haller, A. M. Maguire, D. Marcus, and H. R. McDonald, “Single-field fundus photography for diabetic retinopathy screening: a report by the American Academy of Ophthalmology,” *Ophthalmology*, vol. 111, no. 5, pp. 1055–1062, May 2004.
- [20] O. Pomerantzeff, “Equator plus camera,” *Investigative Ophthalmology*, vol. 14, no. 5, pp. 401–406, 1975.
- [21] P. S. Silva, J. D. Cavallerano, J. K. Sun, A. Z. Soliman, L. M. Aiello, and L. P. Aiello, “Peripheral Lesions Identified by Mydriatic Ultrawide Field Imaging: Distribution and Potential Impact on Diabetic Retinopathy Severity,” *Ophthalmology*, vol. 120, no. 12, pp. 2587–2595, Dec. 2013.
- [22] S. J. Talks, V. Manjunath, D. H. W. Steel, T. Peto, and R. Taylor, “New vessels detected on wide-field imaging compared to two-field and seven-field imaging: implications for diabetic retinopathy screening image analysis,” *British Journal of Ophthalmology*, p. bjophthalmol-2015-306719, Aug. 2015.
- [23] M. M. Wessel, N. Nair, G. D. Aaker, J. R. Ehrlich, D. J. D’Amico, and S. Kiss, “Peripheral retinal ischaemia, as evaluated by ultra-widfield fluorescein angiography, is associated with diabetic macular oedema,” *British Journal of Ophthalmology*, vol. 96, no. 5, pp. 694–698, May 2012.
- [24] P. S. Silva *et al.*, “Peripheral Lesions Identified on Ultrawide Field Imaging Predict Increased Risk of Diabetic Retinopathy Progression over 4 Years,” *Ophthalmology*, vol. 122, no. 5, pp. 949–956, May 2015.
- [25] C. Heneghan, J. Flynn, M. O’Keefe, and M. Cahill, “Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis,” *Medical image analysis*, vol. 6, no. 4, pp. 407–429, 2002.
- [26] D. A. Sim *et al.*, “Patterns of Peripheral Retinal and Central Macula Ischemia in Diabetic Retinopathy as Evaluated by Ultra-widfield Fluorescein Angiography,” *American Journal of Ophthalmology*, vol. 158, no. 1, p. 144–153.e1, Jul. 2014.
- [27] A. Bastawrous, “Increasing access to eye care ... there’s an app for that. Peek: smartphone technology for eye health,” *Int J Epidemiol*, vol. 45, no. 4, pp. 1040–1043, Aug. 2016.
- [28] P. Godara, A. M. Dubis, A. Roorda, J. L. Duncan, and J. Carroll, “Adaptive Optics Retinal Imaging: Emerging Clinical Applications,” *Optom Vis Sci*, vol. 87, no. 12, pp. 930–941, Dec. 2010.
- [29] P. Dodson, *Oxford diabetes library: diabetic retinopathy*. Oxford University Press, 2008.
- [30] H. R. Taylor and J. E. Keeffe, “World blindness: a 21st century perspective,” *British Journal of Ophthalmology*, vol. 85, no. 3, pp. 261–266, Mar. 2001.

- [31] M. M. Habib, R. A. Welikala, A. Hoppe, C. G. Owen, A. R. Rudnicka, and S. A. Barman, "Detection of microaneurysms in retinal images using an ensemble classifier," *Informatics in Medicine Unlocked*, vol. 9, no. Supplement C, pp. 44–57, Jan. 2017.
- [32] T. A. Ciulla, A. G. Amador, and B. Zinman, "Diabetic Retinopathy and Diabetic Macular Edema," *Diabetes Care*, vol. 26, no. 9, pp. 2653–2664, Sep. 2003.
- [33] I. Kocur and S. Resnikoff, "Visual impairment and blindness in Europe and their prevention," *British Journal of Ophthalmology*, vol. 86, no. 7, pp. 716–722, Jul. 2002.
- [34] D. S. Fong *et al.*, "Retinopathy in Diabetes," *Diabetes Care*, vol. 27, no. suppl 1, pp. s84–s87, Jan. 2004.
- [35] T. Bek and M. Erlandsen, "Visual prognosis after panretinal photocoagulation for proliferative diabetic retinopathy," *Acta Ophthalmologica Scandinavica*, vol. 84, no. 1, pp. 16–20, Feb. 2006.
- [36] S. J. Lee *et al.*, "Examination compliance and screening for diabetic retinopathy: a 2-year follow-up study," *Clinical & Experimental Ophthalmology*, vol. 28, no. 3, pp. 149–152, Jun. 2000.
- [37] "Treatment Techniques and Clinical Guidelines for Photocoagulation of Diabetic Macular Edema," *Ophthalmology*, vol. 94, no. 7, pp. 761–774, Jul. 1987.
- [38] M. D. José Pedro De La Cruz, M. D. José Antonio González-Correa, M. D. Ana Guerrero, and M. D. Felipe Sánchez de la Cuesta, "Pharmacological approach to diabetic retinopathy," *Diabetes Metab. Res. Rev.*, vol. 20, no. 2, pp. 91–113, Mar. 2004.
- [39] D. C. Klonoff and D. M. Schwartz, "An economic analysis of interventions for diabetes.," *Diabetes care*, vol. 23, no. 3, pp. 390–404, 2000.
- [40] J. T. Gillow and J. A. M. Gray, "The National Screening Committee review of diabetic retinopathy screening," *Eye*, vol. 15, no. 1, pp. 1–2, Jan. 2001.
- [41] J. M. G. Wilson, G. Jungner, and others, "Principles and practice of screening for disease.," *Principles and practice of screening for disease.*, no. 34, 1968.
- [42] U. N. S. Committee and others, "Essential elements in developing a diabetic retinopathy screening programme," *Workbook*, vol. 4, pp. 1–79, 2007.
- [43] R. N. Frank, "Diabetic retinopathy," *Progress in Retinal and Eye Research*, vol. 14, no. 2, pp. 361–392, 1995.
- [44] J. J. Kanski and B. Bowling, *Clinical ophthalmology: a systematic approach*. Elsevier Health Sciences, 2011.
- [45] T. Kauppi *et al.*, "The DIARETDB1 Diabetic Retinopathy Database and Evaluation Protocol.," in *BMVC*, 2007, pp. 1–10.
- [46] T. Kauppi *et al.*, "Constructing Benchmark Databases and Protocols for Medical Image Analysis: Diabetic Retinopathy," *Computational and Mathematical Methods in Medicine*, vol. 2013, p. e368514, Jun. 2013.
- [47] M. Niemeijer *et al.*, "Retinopathy Online Challenge: Automatic Detection of Microaneurysms in Digital Color Fundus Photographs," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 185–195, Jan. 2010.

- [48] A. Mizutani, C. Muramatsu, Y. Hatanaka, S. Suemori, T. Hara, and H. Fujita, “Automated microaneurysm detection method based on double ring filter in retinal fundus images,” 2009, p. 72601N.
- [49] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, K. W. Tobin Jr, and E. Chaum, “Microaneurysm detection with radon transform-based classification on retina images,” in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 5939–5942.
- [50] “Measuring Retinal Vessel Tortuosity in 10-Year-Old Children: Validation of the Computer-Assisted Image Analysis of the Retina (CAIAR) Program | IOVS | ARVO Journals.” [Online]. Available: <http://iovs.arvojournals.org/article.aspx?articleid=2126623>. [Accessed: 31-Jul-2017].
- [51] A. D. Hoover, V. Kouznetsova, and M. Goldbaum, “Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response,” *IEEE Transactions on Medical Imaging*, vol. 19, no. 3, pp. 203–210, Mar. 2000.
- [52] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
- [53] F. Damian, *Aria online, retinal image archive*. 2006.
- [54] “Available Datasets - Dr Hossein Rabbani,” 2015. [Online]. Available: <https://sites.google.com/site/hosseinrabbanikhorasgani/datasets-1>.
- [55] I. Lazar and A. Hajdu, “Microaneurysm detection in retinal images using a rotating cross-section based model,” in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2011, pp. 1405–1409.
- [56] I. Lazar and A. Hajdu, “Retinal Microaneurysm Detection Through Local Rotating Cross-Section Profile Analysis,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 2, pp. 400–407, Feb. 2013.
- [57] S. B. Júnior and D. Welfer, “Automatic Detection of Microaneurysms and Hemorrhages in Color Eye Fundus Images,” *International Journal of Computer Science and Information Technology*, vol. 5, no. 5, pp. 21–37, Oct. 2013.
- [58] Q. Li, R. Lu, S. Miao, and J. You, “Detection of microaneurysms in color retinal images using multi-orientation sum of matched filter,” in *Proc. of the 3rd International Conference on Multimedia Technology*, 2013.
- [59] L. Giancardo, F. Mériaudeau, T. P. Karnowski, K. W. Tobin, Y. Li, and E. Chaum, “Microaneurysms detection with the radon cliff operator in retinal fundus images,” in *SPIE Medical Imaging*, 2010, p. 76230U–76230U.
- [60] C. I. Sánchez, R. Hornero, A. Mayo, and M. García, “Mixture model-based clustering and logistic regression for automatic detection of microaneurysms in retinal images,” in *SPIE medical imaging*, 2009, vol. 7260, p. 72601M–72601M–8.
- [61] T. Spencer, J. A. Olson, K. C. McHardy, P. F. Sharp, and 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*, vol. 29, no. 4, pp. 284–302, Aug. 1996.

- [62] E. R. Davies, *Computer and machine vision: theory, algorithms, practicalities*, 4th ed. Amsterdam ; Boston: Elsevier, 2012.
- [63] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "A fully automated comparative microaneurysm digital detection system," *Eye*, vol. 11, pp. 622–628, 1997.
- [64] J. H. Hipwell, F. Strachan, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool," *Diabetic Medicine*, vol. 17, no. 8, pp. 588–594, Aug. 2000.
- [65] L. Streeter and M. J. Cree, "Microaneurysm detection in colour fundus images," *Image Vision Comput. New Zealand*, pp. 280–284, 2003.
- [66] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1223–1232, Sep. 2006.
- [67] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, "Detection of microaneurysms using multi-scale correlation coefficients," *Pattern Recognition*, vol. 43, no. 6, pp. 2237–2248, Jun. 2010.
- [68] J. Wu, J. Xin, L. Hong, J. You, and N. Zheng, "New hierarchical approach for microaneurysms detection with matched filter and machine learning," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 4322–4325.
- [69] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal Wavelet Transform for the Detection of Microaneurysms in Retina Photographs," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1230–1241, Sep. 2008.
- [70] C. Sinthanayothin *et al.*, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Medicine*, vol. 19, no. 2, pp. 105–112, Feb. 2002.
- [71] S. Abdelazeem, "Micro-aneurysm detection using vessels removal and circular hough transform," in *Radio Science Conference, 2002.(NRSC 2002). Proceedings of the Nineteenth National*, 2002, pp. 421–426.
- [72] T. Inoue, Y. Hatanaka, S. Okumura, C. Muramatsu, and H. Fujita, "Automated microaneurysm detection method based on eigenvalue analysis using hessian matrix in retinal fundus images," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013, pp. 5873–5876.
- [73] R. Srivastava, D. W. Wong, L. Duan, J. Liu, and T. Y. Wong, "Red lesion detection in retinal fundus images using Frangi-based filters," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 5663–5666.
- [74] M. Niemeijer, B. van Ginneken, J. Staal, M. S. A. Suttorp-Schulten, and M. D. Abramoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Transactions on Medical Imaging*, vol. 24, no. 5, pp. 584–592, May 2005.
- [75] A. Sopharak, B. Uyyanonvara, and S. Barman, "Automatic microaneurysm detection from non-dilated diabetic retinopathy retinal images using mathematical morphology methods," *IAENG International Journal of Computer Science*, vol. 38, no. 3, pp. 295–301, 2011.

- [76] A. Rocha, T. Carvalho, H. F. Jelinek, S. Goldenstein, and J. Wainer, “Points of Interest and Visual Dictionaries for Automatic Retinal Lesion Detection,” *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2244–2253, Aug. 2012.
- [77] A. Sopharak, B. Uyyanonvara, and S. Barman, “Simple hybrid method for fine microaneurysm detection from non-dilated diabetic retinopathy retinal images,” *Computerized Medical Imaging and Graphics*, vol. 37, no. 5, pp. 394–402, 2013.
- [78] K. M. Adal, D. Sidibé, S. Ali, E. Chaum, T. P. Karnowski, and F. Mériaudeau, “Automated detection of microaneurysms using scale-adapted blob analysis and semi-supervised learning,” *Computer Methods and Programs in Biomedicine*, vol. 114, no. 1, pp. 1–10, Apr. 2014.
- [79] K. Ram, G. D. Joshi, and J. Sivaswamy, “A Successive Clutter-Rejection-Based Approach for Early Detection of Diabetic Retinopathy,” *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 664–673, Mar. 2011.
- [80] R. Rosas-Romero, J. Martínez-Carballido, J. Hernández-Capistrán, and L. J. Uribe-Valencia, “A method to assist in the diagnosis of early diabetic retinopathy: Image processing applied to detection of microaneurysms in fundus images,” *Computerized Medical Imaging and Graphics*, vol. 44, pp. 41–53, Sep. 2015.
- [81] M. Haloi, “Improved Microaneurysm Detection using Deep Neural Networks,” *arXiv preprint arXiv:1505.04424*, 2015.
- [82] M. García, M. I. López, D. Álvarez, and R. Hornero, “Assessment of four neural network based classifiers to automatically detect red lesions in retinal images,” *Medical Engineering & Physics*, vol. 32, no. 10, pp. 1085–1093, Dec. 2010.
- [83] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.
- [84] G. K. Matsopoulos, N. A. Mouravliansky, K. K. Delibasis, and K. S. Nikita, “Automatic retinal image registration scheme using global optimization techniques,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 3, no. 1, pp. 47–60, 1999.
- [85] N. Ritter, R. Owens, J. Cooper, R. H. Eikelboom, and P. P. Van Saarloos, “Registration of stereo and temporal images of the retina,” *Medical Imaging, IEEE Transactions on*, vol. 18, no. 5, pp. 404–418, 1999.
- [86] D. Lloret, J. Serrat, A. M. López, A. Soler, and J. J. Villaneuva, “Retinal image registration using creases as anatomical landmarks,” in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 2000, vol. 3, pp. 203–206.
- [87] A. V. Cideciyan, “Registration of ocular fundus images: an algorithm using cross-correlation of triple invariant image descriptors,” *IEEE Engineering in Medicine and Biology Magazine*, vol. 14, no. 1, pp. 52–58, Jan. 1995.
- [88] L. Ballerini, “Temporal matched filters for integration of ocular fundus images,” in *Proceedings of 13th International Conference on Digital Signal Processing*, 1997, vol. 2, pp. 1161–1164 vol.2.
- [89] M. Skokan, A. Skoupy, and J. Jan, “Registration of multimodal images of retina,” in *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual*

*Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint, 2002, vol. 2, pp. 1094–1096.*

- [90] B. Zitova and J. Flusser, “Image registration methods: a survey,” *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.
- [91] F. L. Ferris, S. L. Fine, and L. Hyman, “Age-Related Macular Degeneration and Blindness due to Neovascular Maculopathy,” *Arch Ophthalmol*, vol. 102, no. 11, pp. 1640–1642, Nov. 1984.
- [92] “Microaneurysms in the development of diabetic retinopathy (UKPDS 42) | SpringerLink.” [Online]. Available: <https://link.springer.com/article/10.1007/s001250051278>. [Accessed: 26-Jul-2017].
- [93] “Observations Following Laser Treatment for Choroidal Neovascularization | JAMA Ophthalmology | The JAMA Network.” [Online]. Available: <http://jamanetwork.com/journals/jamaophthalmology/article-abstract/637540>. [Accessed: 26-Jul-2017].
- [94] C. Tm, F. Wr, and G. Mh, “Digital overlay of fluorescein angiograms and fundus images for treatment of subretinal neovascularization,” *Retina*, vol. 12, no. 2, pp. 118–126, 1992.
- [95] C. L. Tsai, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, “Covariance-driven retinal image registration initialized from small sets of landmark correspondences,” in *Proceedings IEEE International Symposium on Biomedical Imaging*, 2002, pp. 333–336.
- [96] S. Lee, M. D. Abramoff, and J. M. Reinhardt, “Retinal atlas statistics from color fundus images,” in *SPIE Medical Imaging*, 2010, pp. 762310–762310.
- [97] S. Ali *et al.*, “Statistical atlas based exudate segmentation,” *Computerized Medical Imaging and Graphics*, vol. 37, no. 5, pp. 358–368, 2013.
- [98] T. Chanwimaluang, G. Fan, and S. R. Fransen, “Hybrid retinal image registration,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 10, no. 1, pp. 129–142, 2006.
- [99] W. E. Hart and M. H. Goldbaum, “Registering retinal images using automatically selected control point pairs,” in *Proceedings of 1st International Conference on Image Processing*, 1994, vol. 3, pp. 576–580 vol.3.
- [100] J. Klein, “A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform,” *IEEE Transactions on Medical Imaging*.
- [101] C. Heneghan, P. Maguire, N. Ryan, and P. de Chazal, “Retinal image registration using control points,” in *Proceedings IEEE International Symposium on Biomedical Imaging*, 2002, pp. 349–352.
- [102] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, “A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 3, pp. 347–364, 2002.
- [103] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, “A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: mosaicing

- the curved human retina,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 412–419, Mar. 2002.
- [104] E.-H. Zhang, Y. Zhang, and T.-X. Zhang, “Automatic retinal image registration based on blood vessels feature point,” in *Proceedings. International Conference on Machine Learning and Cybernetics*, 2002, vol. 4, pp. 2010–2015 vol.4.
- [105] H. Shen, C. V. Stewart, B. Roysam, G. Lin, and H. L. Tanenbaum, “Frame-rate spatial referencing based on invariant indexing and alignment with application to online retinal image registration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 3, pp. 379–384, Mar. 2003.
- [106] C. V. Stewart, C.-L. Tsai, and B. Roysam, “The dual-bootstrap iterative closest point algorithm with application to retinal image registration,” *Medical Imaging, IEEE Transactions on*, vol. 22, no. 11, pp. 1379–1394, 2003.
- [107] Z. Ghassabi, J. Shanbehzadeh, A. Sedaghat, and E. Fatemizadeh, “An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors,” *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, pp. 1–16, 2013.
- [108] P. S. Reel, L. S. Dooley, K. C. P. Wong, and A. Börner, “Enhanced retinal image registration accuracy using expectation maximisation and variable bin-sized mutual information,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 6632–6636.
- [109] D. W. K. Wong, J. A. Lee, B. H. Lee, G. Xu, A. Laude, and T. H. Lim, “An Automated System for Multi-model Retinal Image Registration using Novel Low-dimensional Features,” *ARVO Annual Meeting Abstract (2016)*, May 2016.
- [110] Y. Xian, Y. Dai, C. Gao, and R. Du, “Dual-wavelength retinal image registration based on vessel segmentation and optic disc detection,” 2016, vol. 9684, pp. 968417–968417–10.
- [111] F. Guo, X. Zhao, B. Zou, and Y. Liang, “Automatic Retinal Image Registration Using Blood Vessel Segmentation and SIFT Feature,” *Int. J. Patt. Recogn. Artif. Intell.*, vol. 31, no. 11, p. 1757006, Apr. 2017.
- [112] C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, and A. A. Argyros, “Retinal image registration under the assumption of a spherical eye,” *Computerized Medical Imaging and Graphics*, vol. 55, pp. 95–105, Jan. 2017.
- [113] S. Lee, J. M. Reinhardt, P. C. Cattin, and M. D. Abramoff, “Objective and expert-independent validation of retinal image registration algorithms by a projective imaging distortion model,” *Medical image analysis*, vol. 14, no. 4, pp. 539–549, 2010.
- [114] Y. Chen and G. Medioni, “Object modelling by registration of multiple range images,” *Image and Vision Computing*, vol. 10, no. 3, pp. 145–155, Apr. 1992.
- [115] P. S. Silva, J. D. Cavallerano, J. K. Sun, J. Noble, L. M. Aiello, and L. P. Aiello, “Nonmydriatic Ultrawide Field Retinal Imaging Compared With Dilated Standard 7-Field 35-mm Photography and Retinal Specialist Examination for Evaluation of Diabetic Retinopathy,” *American Journal of Ophthalmology*, vol. 154, no. 3, p. 549–559.e2, Sep. 2012.

- [116] F. Laliberte, L. Gagnon, and Y. Sheng, "Registration and fusion of retinal images-an evaluation study," *IEEE Transactions on Medical Imaging*, vol. 22, no. 5, pp. 661–673, May 2003.
- [117] P. C. Cattin, H. Bay, L. Van Gool, and G. Székely, "Retina mosaicing using local features," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pp. 185–192, 2006.
- [118] J. Chen, J. Tian, N. Lee, J. Zheng, R. T. Smith, and A. F. Laine, "A partial intensity invariant feature descriptor for multimodal retinal image registration," *IEEE transactions on bio-medical engineering*, vol. 57, no. 7, p. 1707, 2010.
- [119] S. Lee, M. D. Abramoff, and J. M. Reinhardt, "Retinal image mosaicing using the radial distortion correction model," *Medical Imaging*, pp. 691435–691435, 2008.
- [120] F. Zana and J. C. Klein, "A multimodal registration algorithm of eye fundus images using vessels detection and Hough transform," *IEEE Transactions on Medical Imaging*, vol. 18, no. 5, pp. 419–428, 1999.
- [121] M. M. Fraz *et al.*, "Blood vessel segmentation methodologies in retinal images–A survey," *Computer methods and programs in biomedicine*, vol. 108, no. 1, pp. 407–433, 2012.
- [122] I. N. McRitchie, P. M. Hart, and R. J. Winder, "Image registration and subtraction for the visualization of change in diabetic retinopathy screening," *Computerized Medical Imaging and Graphics*, vol. 30, no. 3, pp. 139–145, 2006.
- [123] H. Narasimha-Iyer *et al.*, "Robust Detection and Classification of Longitudinal Changes in Color Retinal Fundus Images for Monitoring Diabetic Retinopathy," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 6, pp. 1084–1098, Jun. 2006.
- [124] A. Avakian *et al.*, "Fractal analysis of region-based vascular change in the normal and non-proliferative diabetic retina," *Current eye research*, vol. 24, no. 4, pp. 274–280, 2002.
- [125] A. J. Patterson, D. F. Garway-Heath, N. G. Strouthidis, and D. P. Crabb, "A new statistical approach for quantifying change in series of retinal and optic nerve head topography images," *Investigative ophthalmology & visual science*, vol. 46, no. 5, pp. 1659–1667, 2005.
- [126] P. H. Artes and B. C. Chauhan, "Longitudinal changes in the visual field and optic disc in glaucoma," *Progress in retinal and eye research*, vol. 24, no. 3, pp. 333–354, 2005.
- [127] S.-E. Bursell, A. C. Clermont, B. T. Kinsley, D. C. Simonson, L. M. Aiello, and H. A. Wolpert, "Retinal blood flow changes in patients with insulin-dependent diabetes mellitus and no diabetic retinopathy.," *Investigative ophthalmology & visual science*, vol. 37, no. 5, pp. 886–897, 1996.
- [128] R. A. Welikala, M. M. Fraz, T. H. Williamson, and S. A. Barman, "The automated detection of proliferative diabetic retinopathy using dual ensemble classification," *International Journal of Diagnostic Imaging*, vol. 2, no. 2, p. 72, Jun. 2015.
- [129] M. M. Habib, R. A. Welikala, A. Hoppe, C. G. Owen, A. R. Rudnicka, and S. A. Barman, "Detection of microaneurysms in retinal images using an ensemble classifier," *Informatics in Medicine Unlocked*.

- [130] M. Habib, R. A. Welikala, A. Hoppe, C. G. Owen, A. R. Rudnicka, and S. A. Barman, "Microaneurysm Detection in Retinal Images Using an Ensemble Classifier," presented at the The Sixth International Conference on Image Processing Theory, Tools and Applications, Oulu, Finland, 2016.
- [131] S. M. Pizer *et al.*, "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, Sep. 1987.
- [132] A. Sopharak, B. Uyyanonvara, S. Barman, and T. Williamson, "Automatic microaneurysm detection from non-dilated diabetic retinopathy retinal images," in *Proceedings of the World Congress on Engineering*, 2011, vol. 2, pp. 6–8.
- [133] I. Lazar, R. J. Qureshi, and A. Hajdu, "A novel approach for the automatic detection of microaneurysms in retinal images," in *Emerging Technologies (ICET), 2010 6th International Conference on*, 2010, pp. 193–197.
- [134] A. Sopharak, B. Uyyanonvara, S. Barman, and T. H. Williamson, "Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods," *Computerized medical imaging and graphics*, vol. 32, no. 8, pp. 720–727, 2008.
- [135] M. M. Fraz *et al.*, "An approach to localize the retinal blood vessels using bit planes and centerline detection," *Computer Methods and Programs in Biomedicine*, vol. 108, no. 2, pp. 600–616, Nov. 2012.
- [136] O. CREDIT, "Bootstrap-inspired techniques in computational intelligence," *IEEE signal processing magazine*, vol. 1053, no. 5888/07, 2007.
- [137] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, vol. 6, no. 3, pp. 21–45, Third 2006.
- [138] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical Ensemble of Global and Local Classifiers for Face Recognition," *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1885–1896, Aug. 2009.
- [139] N. C. Oza and K. Tumer, "Classifier ensembles: Select real-world applications," *Information Fusion*, vol. 9, no. 1, pp. 4–20, Jan. 2008.
- [140] T. M. Mitchell, *Machine Learning*, 1st ed. New York, NY, USA: McGraw-Hill, Inc., 1997.
- [141] L. Breiman, "Bagging predictors," *Mach Learn*, vol. 24, no. 2, pp. 123–140, Aug. 1996.
- [142] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 103. New York, NY: Springer New York, 2013.
- [143] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962.
- [144] M. M. Fraz, R. A. Welikala, A. R. Rudnicka, C. G. Owen, D. P. Strachan, and S. A. Barman, "QUARTZ: Quantitative Analysis of Retinal Vessel Topology and size – An automated system for quantification of retinal vessels morphology," *Expert Systems with Applications*, vol. 42, no. 20, pp. 7221–7234, Nov. 2015.
- [145] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.

- [146] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, “Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products,” *Chemometrics and Intelligent Laboratory Systems*, vol. 83, no. 2, pp. 83–90, Sep. 2006.
- [147] C. E. Leiserson, R. L. Rivest, C. Stein, and T. H. Cormen, *Introduction to algorithms*. The MIT press Cambridge, 2001.
- [148] G. Liew, M. Michaelides, and C. Bunce, “A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010,” *BMJ Open*, vol. 4, no. 2, p. e004015, Feb. 2014.
- [149] “Diabetic Eye Screening Feature Based Grading Forms.” [Online]. Available: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/402295/Feature\\_Based\\_Grading\\_Forms\\_V1\\_4\\_1Nov12\\_SSG.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/402295/Feature_Based_Grading_Forms_V1_4_1Nov12_SSG.pdf). [Accessed: 02-Apr-2017].
- [150] M. Habib *et al.*, “Incorporating spatial information for microaneurysm detection in retinal images,” *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, no. 3, pp. 642–649, 2017.
- [151] S. S. Patankar and J. V. Kulkarni, “Directional gradient based registration of retinal images,” in *Computational Intelligence & Computing Research (ICCIC), 2012 IEEE International Conference on*, 2012, pp. 1–5.
- [152] E. P. Ong *et al.*, “A robust multimodal retina image registration approach,” 2014.
- [153] S. Lee, M. D. Abramoff, and J. M. Reinhardt, “Feature-based pairwise retinal image registration by radial distortion correction,” *Medical Imaging*, pp. 651220–651220, 2007.
- [154] G. P. Stein, “Lens distortion calibration using point correspondences,” in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 1997, pp. 602–608.
- [155] B. Antal and A. Hajdu, “An ensemble-based system for microaneurysm detection and diabetic retinopathy grading,” *IEEE transactions on biomedical engineering*, vol. 59, no. 6, pp. 1720–1726, 2012.
- [156] R. Klein, S. M. Meuer, S. E. Moss, and B. E. K. Klein, “Retinal Microaneurysm Counts and 10-Year Progression of Diabetic Retinopathy,” *Arch Ophthalmol*, vol. 113, no. 11, pp. 1386–1391, Nov. 1995.
- [157] A. J. Smola, S. V. N. Vishwanathan, and T. Hofmann, “Kernel Methods for Missing Variables.,” in *AISTATS*, 2005.
- [158] Z. Ghahramani and M. I. Jordan, “Learning from Incomplete Data,” Jan. 1995.
- [159] “Random forests - classification description.” [Online]. Available: [https://www.stat.berkeley.edu/~breiman/RandomForests/cc\\_home.htm#missing1](https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#missing1). [Accessed: 21-Aug-2017].

