



Fundus image analysis for automatic screening of ophthalmic pathologies



Adrián Colomer Granero

PhD Thesis
February, 2018

Departamento de Comunicaciones

Supervisor
Prof. Valery Naranjo Ornedo



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Fundus image analysis for automatic screening of ophthalmic pathologies

February, 2018

DEPARTAMENTO DE COMUNICACIONES

Author: Adrián Colomer Granero

Supervisor: Prof. Valery Naranjo Ornedo

The research derived from this doctoral thesis was funded by the *Ministerio de Economía y Competitividad* of Spain, Project ACRIMA (TIN2013-46751-R). The work of Adrián Colomer Granero to carry out this research and to elaborate this dissertation has been supported by the Spanish Government under the FPI Grant BES-2014-067889. Two visiting research fellowships (EEBB-2016-11276 and EEBB-2017-12567) were also funded by the *Ministerio de Economía y Competitividad* of Spain.

Acknowledgments

Este documento es el resultado de cuatro años de esfuerzo y dedicación que han propiciado mi evolución tanto a nivel profesional como personal. El principal motivo de dicho desarrollo han sido todas esas personas que han formado parte de este camino o que de alguna forma se han cruzado en el mismo. A todos ellos quiero dedicarles unas palabras de agradecimiento.

En primer lugar, quiero agradecer a Valery, ella es la responsable de que yo esté ahora mismo escribiendo estas líneas. Me brindó la oportunidad de mi vida hace ya más de cinco años y me ha ido formando tanto en la vertiente investigadora como docente de la misma forma que una madre educa e inculca los valores esenciales a su hijo. Gracias por todo Valery, juntos cerramos una etapa clave en mi formación y juntos abrimos una nueva cargada de ilusión. Espero poder estar siempre a tu lado.

También quiero agradecer todo el apoyo, energía y momentos divertidos que me han brindado mis compañeros de laboratorio. Gracias Sandra por todo lo que me has enseñado desde el primer momento en el que llegué hasta el día de hoy, no sé qué haría yo sin ti en esa sala. Gracias Fer, Félix, Andrés, Reinier, Ángel, Gabri, Fran P., Rober, Jorge y Adri por todos los consejos aportados y por amenizar como sólo vosotros sabéis mi día a día. También quiero recordar y agradecer a los que decidieron emprender otro rumbo: Javi, Conchi, Pablo, Eliseo, Fran M. y Miguel.

De igual forma quiero mostrar mi gratitud hacia varias personas que han colaborado de una u otra forma en este trabajo. A Jesús Angulo, Débora

Gil y Ramón Baldrich por su supervisión del trabajo desarrollado en las estancias llevadas a cabo durante esta etapa. A Kjersti Engan por aportar sus conocimientos sobre *Dictionary Learning & Sparse Representation* que han culminado en diversas publicaciones. A Rafael Molina por permitirme adentrarme en el fascinante mundo de los Procesos Gaussianos y sus múltiples aplicaciones.

Gracias a mi madre Jacqueline, mi hermano Fran, mis abuelos M^a Carmen y Paco, mi padre Luis y Eva que siempre me han brindado su apoyo incondicional. Como bien sabéis sois fundamentales para mí, por ello ésta tesis también va dedicada a vosotros. No me puedo olvidar de Rosa que ocupa una parte grande de mi corazón. Os quiero.

A Duke y Tom por todas esas horas de compañía desinteresada que me han ofrecido durante sus vidas y en especial en la época de redacción de este manuscrito.

Por último, quiero dedicar especialmente este trabajo al motor de mi vida, a la persona que tiene que lidiar conmigo en los momentos más difíciles, a la persona que siempre me acaba sacando una sonrisa, a la persona que me enseña día tras día lo que es amor, fuerza, lucha, perseverancia e infinidad de cualidades que la definen. Gracias por todo cariño, sin tu apoyo estoy seguro de que esto no hubiera sido posible. Durante estos 10 años hemos abierto y cerrado muchas etapas juntos pero lo más ilusionante son todas aquellas que nos quedan por vivir.

Abstract

In last years, the number of blindness cases has been significantly reduced. Despite this promising news, the World Health Organisation estimates that 80% of visual impairment (285 million cases in 2010) could be avoided if diagnosed and treated early. To accomplish this purpose, eye care services need to be established in primary health and screening campaigns should be a common task in centres with people at risk. However, these solutions entail a high workload for trained experts in the analysis of the anomalous patterns of each eye disease. Therefore, the development of algorithms for automatic screening system plays a vital role in this field.

This thesis focuses on the automatic identification of the retinal damage provoked by two of the most common pathologies in the current society: diabetic retinopathy (DR) and age-related macular degeneration (AMD). Specifically, the final goal of this work is to develop novel methods, based on fundus image description and classification, to characterise the healthy and abnormal tissue in the retina background. In addition, pre-processing algorithms are proposed with the aim of normalising the high variability of fundus images and removing the contribution of some retinal structures that could hinder in the retinal damage detection.

In contrast to the most of the state-of-the-art works in damage detection using fundus images, the methods proposed throughout this manuscript avoid the necessity of lesion segmentation or the candidate map generation before the classification stage. Local binary patterns, granulometric profiles and fractal dimension are locally computed to extract texture, morphological and

roughness information from retinal images. Different combinations of this information feed advanced classification algorithms formulated to optimally discriminate exudates, microaneurysms, haemorrhages and healthy tissues. Through several experiments, the ability of the proposed system to identify DR and AMD signs is validated using different public databases with a large degree of variability and without image exclusion.

Moreover, this thesis covers the basics of the deep learning paradigm. In particular, a novel approach based on convolutional neural networks is explored. The transfer learning technique is applied to fine-tune the most important state-of-the-art CNN architectures. Exudate detection and localisation tasks using neural networks are carried out in the last two experiments of this thesis. An objective comparison between the hand-crafted feature extraction and classification process and the prediction models based on CNNs is established.

The promising results of this PhD thesis and the affordable cost and portability of retinal cameras could facilitate the further incorporation of the developed algorithms in a computer-aided diagnosis (CAD) system to help specialists in the accurate detection of anomalous patterns characteristic of the two diseases under study: DR and AMD.

Resumen

En los últimos años el número de casos de ceguera se ha reducido significativamente. A pesar de este hecho, la Organización Mundial de la Salud estima que un 80% de los casos de pérdida de visión (285 millones en 2010) pueden ser evitados si se diagnostican en sus estadios más tempranos y son tratados de forma efectiva. Para cumplir esta propuesta se pretende que los servicios de atención primaria incluyan un seguimiento oftalmológico de sus pacientes así como fomentar campañas de cribado en centros proclives a reunir personas de alto riesgo. Sin embargo, estas soluciones exigen una alta carga de trabajo de personal experto entrenado en el análisis de los patrones anómalos propios de cada enfermedad. Por lo tanto, el desarrollo de algoritmos para la creación de sistemas de cribado automáticos juega un papel vital en este campo.

La presente tesis persigue la identificación automática del daño retiniano provocado por dos de las patologías más comunes en la sociedad actual: la retinopatía diabética (RD) y la degeneración macular asociada a la edad (DMAE). Concretamente, el objetivo final de este trabajo es el desarrollo de métodos novedosos basados en la extracción de características de la imagen de fondo de ojo y clasificación para discernir entre tejido sano y patológico. Además, en este documento se proponen algoritmos de pre-procesado con el objetivo de normalizar la alta variabilidad existente en las bases de datos públicas de imagen de fondo de ojo y eliminar la contribución de ciertas estructuras retinianas que afectan negativamente en la detección del daño retiniano.

A diferencia de la mayoría de los trabajos existentes en el estado del arte sobre detección de patologías en imagen de fondo de ojo, los métodos propuestos a lo largo de este manuscrito evitan la necesidad de segmentación de las lesiones o la generación de un mapa de candidatos antes de la fase de clasificación. En este trabajo, *Local binary patterns*, perfiles granulométricos y la dimensión fractal se aplican de manera local para extraer información de textura, morfología y tortuosidad de la imagen de fondo de ojo. Posteriormente, esta información se combina de diversos modos formando vectores de características con los que se entrenan avanzados métodos de clasificación formulados para discriminar de manera óptima entre exudados, microaneurismas, hemorragias y tejido sano. Mediante diversos experimentos, se valida la habilidad del sistema propuesto para identificar los signos más comunes de la RD y DMAE. Para ello se emplean bases de datos públicas con un alto grado de variabilidad sin excluir ninguna imagen.

Además, la presente tesis también cubre aspectos básicos del paradigma de *deep learning*. Concretamente, se presenta un novedoso método basado en redes neuronales convolucionales (CNNs). La técnica de transferencia de conocimiento se aplica mediante el *fine-tuning* de las arquitecturas de CNNs más importantes en el estado del arte. La detección y localización de exudados mediante redes neuronales se lleva a cabo en los dos últimos experimentos de esta tesis doctoral. Cabe destacar que los resultados obtenidos mediante la extracción de características “manual” y posterior clasificación se comparan de forma objetiva con las predicciones obtenidas por el mejor modelo basado en CNNs.

Los prometedores resultados obtenidos en esta tesis y el bajo coste y portabilidad de las cámaras de adquisición de imagen de retina podrían facilitar la incorporación de los algoritmos desarrollados en este trabajo en un sistema de cribado automático que ayude a los especialistas en la detección de patrones anómalos característicos de las dos enfermedades bajo estudio: RD y DMAE.

Resum

En els últims anys el nombre de casos de ceguera s'ha reduït significativament. A pesar d'este fet, l'Organització Mundial de la Salut estima que un 80% dels casos de pèrdua de visió (285 milions en 2010) poden ser evitats si es diagnostiquen en els seus estadis més primerencs i són tractats de forma efectiva. Per a complir esta proposta es pretén que els servicis d'atenció primària incloguen un seguiment oftalmològic dels seus pacients així com fomentar campanyes de garbellament en centres regentats per persones d'alt risc. No obstant això, estes solucions exigixen una alta càrrega de treball de personal expert entrenat en l'anàlisi dels patrons anòmals propis de cada malaltia. Per tant, el desenrotllament d'algoritmes per a la creació de sistemes de garbellament automàtics juga un paper vital en este camp.

La present tesi persegueix la identificació automàtica del dany retinià provocat per dos de les patologies més comunes en la societat actual: la retinopatia diabètica (RD) i la degeneració macular associada a l'edat (DMAE) . Concretament, l'objectiu final d'este treball és el desenrotllament de mètodes novedosos basats en l'extracció de característiques de la imatge de fons d'ull i classificació per a discernir entre teixit sa i patològic. A més, en este document es proposen algoritmes de pre- processat amb l'objectiu de normalitzar l'alta variabilitat existent en les bases de dades publiques d'imatge de fons d'ull i eliminar la contribució de certes estructures retinianas que afecten negativament en la detecció del dany retinià.

A diferència de la majoria dels treballs existents en l'estat de l'art sobre detecció de patologies en imatge de fons d'ull, els mètodes proposats al llarg d'este

manuscrit eviten la necessitat de segmentació de les lesions o la generació d'un mapa de candidats abans de la fase de classificació. En este treball, *Local binary patterns*, perfils granulomètrics i la dimensió fractal s'apliquen de manera local per a extraure informació de textura, morfologia i tortuositat de la imatge de fons d'ull. Posteriorment, esta informació es combina de diversos modes formant vectors de característiques amb els que s'entrenen avançats mètodes de classificació formulats per a discriminar de manera òptima entre exsudats, microaneurismes, hemorràgies i teixit sa. Per mitjà de diversos experiments, es valida l'habilitat del sistema proposat per a identificar els signes més comuns de la RD i DMAE. Per a això s'empren bases de dades públiques amb un alt grau de variabilitat sense exlcuir cap imatge.

A més, la present tesi també cobrix aspectes bàsics del paradigma de *deep learning*. Concretament, es presenta un nou mètode basat en xarxes neuronals convolucionales (CNNs) . La tècnica de transferència de coneixement s'aplica per mitjà del *fine-tuning* de les arquitectures de CNNs més importants en l'estat de l'art. La detecció i localització d'exudats per mitjà de xarxes neuronals es du a terme en els dos últims experiments d'esta tesi doctoral. Cal destacar que els resultats obtinguts per mitjà de l'extracció de característiques "manual" i posterior classificació es comparen de forma objectiva amb les prediccions obtingudes pel millor model basat en CNNs.

Els prometedors resultats obtinguts en esta tesi i el baix cost i portabilitat de les cambres d'adquisició d'imatge de retina podrien facilitar la incorporació dels algoritmes disenrotllats en este treball en un sistema de garbellament automàtic que ajude als especialistes en la detecció de patrons anomálos característics de les dos malalties baix estudi: RD i DMAE.

Contents

Abstract	iii
Contents	xiii
1 Introduction	1
1.1 Motivation	3
1.2 Objectives	4
1.3 Main contributions	5
1.4 Framework	6
1.5 Outline	7
2 The retina’s anatomy and ophthalmic pathologies	9
2.1 Anatomical structures	10
2.2 Fundus photographs	12
2.3 Eye diseases and fundus lesions	14
2.3.1 Diabetic Retinopathy	15
2.3.2 Age-related macular degeneration	17
2.3.3 Relevant groups	19
3 Retinal image pre-processing	23
3.1 Introduction	25
3.2 Material	25
3.3 Normalisation of public fundus databases	27
3.3.1 Spatial normalisation	28
3.3.2 Colour transformation	29
3.4 Blood vessel inpainting	32
3.4.1 Sparse representation & dictionary learning	35
3.4.2 Inpainting using Sparse Representation and Dictionary Learning	39
3.4.3 Measuring the performance of blood vessel inpainting	41
3.4.4 Assessment of retinal blood vessel inpainting	44

4	Extraction of fundus visual information for pathological tissue identification	53
4.1	Description of retina background lesions	55
4.1.1	Introduction	55
4.1.2	Local Binary Patterns Variance	57
4.1.3	Fractal analysis	64
4.1.4	Granulometry profile	71
4.2	Machine Learning methods for retinal tissue classification	81
4.2.1	Classification strategy	81
4.2.2	Random Forest	82
4.2.3	Support Vector Machine	84
4.2.4	Gaussian processes for classification	89
5	Deep Learning approach for retinal damage detection	95
5.1	Introduction	97
5.2	CNN architecture	100
5.2.1	Convolutional layer	101
5.2.2	Activation layer	103
5.2.3	Pooling layer	104
5.2.4	Fully connected layer	105
5.3	Training and test stages	107
5.4	Improving CNNs performance	108
5.5	Transfer learning & state of the art in CNN models	111
5.6	CNNs applied to retinal images	115
6	Results and discussion	119
6.1	Material	121
6.2	Experiments	122
6.2.1	Effect of the pre-processing methods	123
6.2.2	Behaviour of texture, shape and roughness descriptors . .	126
6.2.3	Performance of machine learning algorithms	131
6.2.4	Global detection of bright and dark retinal lesions for automatic screening.	136
6.2.5	Pathological retinal tissue identification by CNNs	142
7	Conclusions	147
	Merits	151
	Bibliography	153

List of Figures

2.1	Eye ball anatomy	10
2.2	Vision process	11
2.3	Main parts of the retina	13
2.4	Non-mydratic retinal cameras	14
2.5	Fundus image and its field of view mask	14
2.6	Eye diseases effect	15
2.7	Signs of diabetic retinopathy in fundus images	17
2.8	Grades of diabetic retinopathy	18
2.9	Signs of age-related macular degeneration	19
3.1	Typical inconveniences in fundus databases	28
3.2	Illustration of the capture angle and width of the field of view mask	29
3.3	Chromaticity diagram and parameters extracted to perform the colour transformation	30
3.4	Colour normalisation process	31
3.5	Visual results of the proposed colour normalisation	33
3.6	Retinal vessel inpainting based on SR and DL	40
3.7	An example of blood vessel inpainting	40
3.8	Overview of the proposed sparse-based inpainting	40
3.9	Artificial mask generation	41
3.10	Different neighbourhoods to perform the inpainting	45
3.11	Results of the greyscale image inpainting	46
3.12	Results of colour image inpainting	49
3.13	Blood vessels inpainting comparison	51
4.1	Illustration of the local analysis in fundus images	56
4.2	Binary mask of a fundus image and the final grid to carry out the local analysis	57
4.3	Two neighbourhoods of different radii for LBP computation	59
4.4	Uniform Local Binary Patterns	62

4.5	Local texture analysis	63
4.6	The first four iterations of the Koch snowflake	66
4.7	Fundus image to illustrate the fractal analysis computation . . .	70
4.8	Process to compute the local fractal dimension	70
4.9	Pyramid of openings computed for a fundus image using an isotropic structuring element	77
4.10	Pyramid of closings computed for a fundus image using an isotropic structuring element	78
4.11	Local extraction of the granulometry and anti-granulometry curves	79
4.12	Global pattern spectrum from a fundus patch	79
4.13	Process of creating the machine learning models using a generic dataset	82
4.14	Example of tree classifier	83
4.15	Illustration of the training process in Random Forests	84
4.16	Illustration of the testing process in Random Forests	85
4.17	Representation of the SVM classification process	86
4.18	Internal cross validation overview	89
4.19	Gaussian Processes for Classification	90
5.1	Input image and its corresponding activation map after a convolutional layer	101
5.2	Zero padding representation	102
5.3	Output volume of a convolutional layer when the input is a RGB image	102
5.4	Activation functions	104
5.5	Illustration of the maxpooling operation	105
5.6	Representation of the bottom layers in a CNN architecture . . .	106
5.7	Typical and basic convolutional network architecture	106
5.8	Illustration of the dropout technique	109
5.9	Illustration of the batch normalisation technique	110
5.10	Architecture of VGG16 network	113
5.11	Illustration of the residual block	113
5.12	Representation of the Inception module	114
5.13	Illustration of Xception module	115
5.14	Fundus image to show the imbalanced character of the problem	116
5.15	Pipeline followed to fine-tune the CNN models	117
6.1	Ground truth mask composition of a representative fundus image	122
6.2	General overview of the possible pre-processing methods	124
6.3	Pipeline of the experiment to evaluate the proposed pre-processing methods	125

6.4	ROC curves for the different tests	126
6.5	ROC curves for the different tests	133
6.6	Representation of the lesion detection results	137
6.7	Representative global descriptors	139
6.8	ROC curve of global detection of bright lesions on <i>DIARETDB1</i> database	140
6.9	ROC curve of global detection of dark lesions on <i>DIARETDB1</i> database	141
6.10	Healthy and pathological retinal samples involved in the transfer learning process	143
6.11	Accuracy, sensitivity and specificity distribution of the different CNN models	144
6.12	ROC curves of the fine-tuned CNN architectures in bright lesion detection	145

List of Tables

3.1	Distribution of the <i>E-OPHTHA</i> images according to their resolution	26
3.2	Mean Recovery Error for grey-scale sparse-inpainting method . . .	46
3.3	Mean Recovery Error for RGB-separate inpainting	47
3.4	Mean Recovery Error for RGB-jointly inpainting	48
3.5	Inpainting quality evaluation index for different inpainting methods	50
3.6	Mean opinion score metric for the inpainted images of <i>DRIVE</i> and <i>STARE</i>	50
3.7	Computational time of inpainting one image using different inpainting methods	52
4.1	Uniformity of Local Binary Patterns	61
6.1	Distribution of the <i>E-OPHTHA</i> images according to their resolution	121
6.2	Tunable parameters involved in the different stages of the lesion detection system	124
6.3	Classification performance evaluation of the different pre- processing methods presented	125
6.4	Classification performance evaluation of different morphological feature vectors for exudate detection in <i>E-OPHTHA</i> database .	127
6.5	Classification performance evaluation of different texture and morphological feature vectors for exudate detection in <i>E- OPHTHA</i> database	128
6.6	Classification performance evaluation of different texture mor- phological and roughness feature vectors for exudate detection in <i>E-OPHTHA</i> database	129
6.7	Classification performance evaluation of different morphological feature vectors for dark lesion detection in <i>DIARETDB1</i> database	129
6.8	Classification performance evaluation of different texture and morphological feature vectors for dark lesion detection in <i>DIARETDB1</i> database	130

6.9	Classification performance evaluation of different texture, morphological and roughness feature vectors for dark lesion detection in <i>DIARETDB1</i> database	130
6.10	Classification performance evaluation of different ML algorithms for exudate detection in <i>E-OPHTHA</i> database	132
6.11	Normalized classification performance evaluation of different ML algorithms for exudate detection in <i>E-OPHTHA</i> database	133
6.12	Classification performance evaluation of different ML algorithms for exudate detection in <i>DIARETDB1</i> database	134
6.13	Normalized classification performance evaluation of different ML algorithms for exudate detection in <i>DIARETDB1</i> database	134
6.14	Comparison of exudate detection methods	134
6.15	Classification performance evaluation of different ML algorithms for dark lesion detection in <i>DIARETDB1</i> database	135
6.16	Normalized classification performance evaluation of different ML algorithms for dark lesion detection in <i>DIARETDB1</i> database	135
6.17	Comparison of dark lesion detection methods	136
6.18	Quantitative results of the global bright lesion detection on <i>DIARETDB1</i> database	139
6.19	Quantitative results of the global dark lesion detection on <i>DIARETDB1</i> database	140
6.20	Quantitative results of bright and dark lesions reported by the authors of <i>DIARETDB1</i> database	141
6.21	Layers and optimizers used in the fine-tuning of CNNs	143
6.22	Results of bright lesion detection using fine-tuned CNNs	144

List of Algorithms

4.1	Local texture descriptor computation	64
4.2	Local fractal analysis computation	72
4.3	Computation of the local granulometries	80
4.4	Learning algorithm using GPCs	93

Chapter 1

Introduction

This chapter presents the motivations behind this thesis, its objectives and its main contributions. In addition, it also introduces the thesis framework and the thesis outline.

Contents

1.1	Motivation	3
1.2	Objectives	4
1.3	Main contributions	5
1.4	Framework	6
1.5	Outline	7

1.1 Motivation

According to the last study published by the World Health Organisation (WHO), there exists a major opportunity of changing the lives of a lot of people. The WHO calculates that 285 million people were visually impaired in 2010. Recent studies demonstrate that the number of blindness cases has been significantly reduced in last years (39 million people in 2010)(World Health Organization, 2015). In spite of this promising data, the WHO estimates that 80% of visual impairment could be avoided if diagnosed and treated early. In addition, the WHO points out that 102\$ billions could be saved with appropriate eye care services.

The action plan presented by the World Health Organization (2015) reflects the eye care services need to become an integral part of primary health care. The proposed plan also highlights the importance of the screening campaigns, an effective method for detecting retinal pathologies in their early stage (World Health Organization, 2015). However, these proposals require a large workload of trained experts in the identification of anomalous retinal patterns. This fact along with the population growth, ageing, physical inactivity, tobacco consumption and rising levels of obesity, that increase the population at risk, make these proposals economically unfeasible. Moreover, it must be pointed out that this type of retinal disease diagnosis is highly influenced by the inherent subjectivity of each expert. These arguments suggest the need of developing automatic screening systems that can be used in primary health care reducing the working time to the specialists and providing the location and quantification of the pathological retinal tissue propitiated by retinal diseases.

Automatic screening systems based on digital imaging should be supported by acquisition devices of high resolution to guarantee the successful of the diagnosis in difficult cases. In the retinal image field, fundus photography, optical coherence tomography (OCT) and angiography allow the specialists to make decisions about the patients' care. Treatment decisions are made taking into account the information provided by combining these acquisition techniques. However, fundus cameras are characterised by their portability and a lower cost compared with the OCT and angiography. For this reason, fundus cameras are the most suitable devices to be established in primary health care centres and to be used in screening campaigns. Due to this fact, automatic algorithms able to extract the relevant information from the fundus images with the aim of providing help to clinicians in diagnosis and disease prevention have become increasingly important in recent years.

From an engineering point of view, in the last years, the research in the computer vision field have been focused in new techniques for describing the visual information contained in a digital image. Feature extraction and machine learning methods have been studied in depth and deep learning have been proposed as a promising paradigm for the automatic image description and classification.

The motivation of this PhD thesis is defined by the intersection between the needs in the ophthalmology field and the emergent image processing techniques which allows to create new solutions for helping experts in the early diagnose of two the most common pathologies that provoke retinal damage in the current society: diabetic retinopathy and age-related macular degeneration.

Note that the field of this thesis is currently very active at the international level. The arrival of strong players like Google keeps the automatic screening of eye diseases in a constant advance by applying the last computer vision techniques with the most powerful hardware.

1.2 Objectives

The main goal of this work is the development of new automatic algorithms based on fundus image description and classification methods to characterise the healthy and pathological tissue in the retina background. The methods proposed throughout this manuscript avoid the lesion segmentation step. This fact provides a new point of view in the analysis of fundus photographs and propitiates a series of specific objectives:

- Study of the anatomy and physiology principles of the human eye. Exhaustive analysis of the anomalous patterns manifested in pathological retinas for each disease under study.
- Determination of the different phenomena involved in the acquisition protocol of the fundus photographs. Development and implementation of methods for the normalisation of retinal images acquired in different conditions.
- Evaluation of the retinal structure contribution in the analysis of the abnormal patterns manifested in the retina background. Development of automatic methods for reducing the contributions that can hinder the posterior analysis.

- Study and application of texture, shape and roughness descriptors to the fundus images with the aim of extracting relevant features for feeding advanced classification algorithms.
- Investigate the basics of the deep learning paradigm and apply convolutional neural networks for the automatic feature extraction and classification of anomalous patterns manifested in pathological retinas.
- Comparison of the different models for the classification of healthy and pathological retinal tissue and selection of the optimal one. Creation of an automatic tool that provides the location of the anomalous regions and objective measures for helping in the final diagnose of the pathologies under study.

1.3 Main contributions

This thesis provides novel methods to automatically characterise pathological and healthy retinal tissues by using local image descriptors and machine learning algorithms. One of its main contributions is the combination of different visual information to train predictive models robust against the high variability of fundus images, i.e. without excluding problematic images of the database.

To provide robustness to our models, an in-depth study of the most common factors that introduce variability to the fundus images is carried out. In this way, a new method to normalise retinal images is proposed in this thesis. This approach, based on geometric transformations of the chromatic diagram of the images, is able to translate the retina background colour of a whole database into the “ideal” retinal colour of a reference image (Colomer et al., 2017a).

Not only the variability among different images produce a negative effect in the fundus image description stage and posterior classification. Some of the own retinal structures could hamper in the process of accurately detecting retinal damage. Regarding to this fact, an inpainting method based on sparse representation and dictionary learning techniques is proposed to remove the contribution of the largest retinal structure: the blood vessels (Colomer et al., 2017b).

The main novelty of this thesis focuses on the way in which image description stage is performed. Lesion segmentation and the generation of lesion candidate maps are avoided in this novel approach, in which, texture, morphological and

roughness operators are applied in a local way (Colomer et al., 2018). Another relevant point of this work is the fact that both kinds of retinal lesions (i.e. bright and dark) are successfully detected using the same visual information, in other words, in the most of the state-of-the-art papers, procedures are designed ad-hoc to identify bright or dark retinal damage. Regarding the classification phase, in this manuscript, Gaussian processes for classification are formulated for the first time for the problem under study and a comparison with the classical machine learning algorithms is established.

The last contribution of this thesis to the field of fundus image processing, in general, and the retinal damage detection and localisation tasks, in particular, is a novel approach based on deep learning techniques. Convolutional neural networks (CNNs) are studied in depth with the aim of automatically describing the fundus image background and extract objective features independent to human decisions (i.e. without the need of choosing specific image descriptors). The knowledge about healthy and abnormal retinal tissues is provided to the most important CNN architectures by means of the fine-tuning technique. The results obtained with these models are also compared with the previous methodology proposed in this work (i.e. hand-crafted image descriptors and classification stage).

The promising results of this PhD thesis suggest the inclusion of the best classification models into the first version of a screening system developed in our research group to automatically diagnose the most common eye diseases in the current society. The development of this system was funded by the Spanish government under the project ACRIMA (TIN2013-46751-R).

1.4 Framework

This thesis stands within the framework of the research project entitled “Fundus image processing for automatic screening of ophthalmological diseases - ACRIMA” (TIN2013-46751-R) funded by the *Ministerio de Economía y Competitividad* of Spain. The work of Adrián Colomer Granero to carry out this research and to elaborate this dissertation has been supported by the Spanish Government under the FPI Grant BES-2014-067889.

The main purpose of ACRIMA project is the development of an automatic screening system for the three most significant diseases related to retinal damage: diabetic retinopathy (DR), age-related macular degeneration (ARMD) and glaucoma. Nowadays, there is no screening system as the proposed in the project, which combines the screening of these three diseases. This project was

fruit of the collaboration between the *Instituto de Investigación e Innovación en Bioingeniería (I3B)*¹ and *FISABIO-Oftalmológica Médica*².

In addition, the *Ministerio de Economía y Competitividad* funded Adrián Colomer Granero two visiting research fellowships in which the PhD student learned the principles of advanced image processing and computer vision techniques. In 2016 Adrián Colomer worked three months at the *Centre of Mathematical morphology (CMM)*³ belonging to the *École de Mines ParisTech* under the Grant EEBB-2016-11276. This centre is the place where the mathematical morphology technique was born. The aim of this scholarship was to apply different mathematical morphology descriptors to describe retinal damage. The second mobility period (EEBB-2017-12567) was propitiated by the collaboration between the *Instituto de Investigación e Innovación en Bioingeniería (I3B)* and the *Centre de Visió per Computador (CVC)*⁴ belonging to the *Universitat Autònoma de Barcelona*. The purpose of this period was the study of the deep learning principles in general and convolutional neural networks in particular. The knowledge acquired in both research fellowships has been applied to the aim of this thesis and some of the results reported in Section 6 are part of the work developed during these periods.

1.5 Outline

This thesis is divided in 7 chapters. This chapter has presented the motivations behind the research involved in this thesis, the main and secondary objectives and its main contributions. Finally, it also explains the thesis framework.

Chapter 2 describes the anatomical and physiological aspects of the retina. An explanation about the retinal diseases under study is also included. Finally, in this chapter, the most relevant groups working on the development of algorithms for retinal damage detection are introduced.

In Chapter 3 an exhaustive analysis about the acquisition factors that reduce the quality of fundus images and the physiological factors that increase the variability among databases is carried out. Automatic methods for pre-processing retinal images are presented with the objective of mitigating these problems. In addition, the contribution of blood vessels over the posterior

¹www.i3b.upv.es

²<http://fisabio.san.gva.es>

³<http://cmm.ensmp.fr>

⁴www.cvc.uab.es

retinal damage detection is evaluated and a novel inpainting method is proposed to reduce their effect.

In Chapter 4 the reader will find novel methods for fundus image description. Texture, morphological and roughness descriptors are applied to extract the relevant information from the retina background. In this chapter, advance classification techniques are also formulated for discriminating between healthy and pathological retinal tissue making use of the visual information extracted.

Chapter 5 details the basics of the deep learning paradigm. In particular the theoretical framework about the convolutional neural networks is explained in depth. Moreover, a novel approach to detect abnormal retinal tissue by means of fine-tuning state-of-the-art CNN architectures is presented.

Chapter 6 reports the findings of this thesis. On the one hand, experiments based on retinal image processing and machine learning methods try to demonstrate the ability of different descriptors and the influence of the proposed preprocessing methods for the automatic detection of ophthalmic pathologies. On the other hand and with the same purpose, experiments based on convolutional neural networks are carried out. This section includes the results obtained by means of this paradigm. In addition, a comparison between the “hand-crafted” feature extraction plus classification techniques and the deep learning approach is performed. The results exposed in this chapter are conveniently discussed and compared with similar state-of-the-art works, analysing the weaknesses and strengths of our methods.

General and final conclusions in addition to future prospects are presented in Chapter 7. Next, the scientific publications derived from this thesis are cited.

Chapter 2

The retina's anatomy and ophthalmic pathologies

This chapter presents a brief explanation about the human vision process, the anatomical aspects of the retina, the retinal diseases under study and the most relevant research groups working on fundus damage detection.

Contents

2.1	Anatomical structures	10
2.2	Fundus photographs	12
2.3	Eye diseases and fundus lesions	14
2.3.1	Diabetic Retinopathy	15
2.3.2	Age-related macular degeneration	17
2.3.3	Relevant groups	19

2.1 Anatomical structures

Eyes are the windows to the outside world. This organ is able to harness light from the environment, regulate its intensity through a diaphragm and focus it using an adjustable lens to form a pattern of light. The inner layer of the eye converts this pattern of light into electrochemical impulses to be processed by the brain giving place to an image (Zhu et al., 2012).

The eyeball is composed by three layers or coats. The outer layer is the sclera, with its anterior part, the cornea. The main purpose of this layer is to refract the light and allow it to converge inside the eye. In addition, the cornea filters the most damaging ultraviolet wavelengths and protects the rest of the eye of germs, dust and other harmful matter. The following coat of the organ of sight is the uvea that includes the choroid, ciliary body and iris. In this part of the eyeball, the amount of light passing through the system is regulated and the optimal focal point is established allowing the creation of a clear image. Finally, the inner layer of the eye is the retina (Figure 2.1).

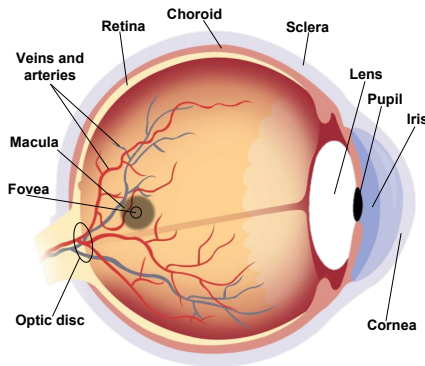


Figure 2.1: Eye ball anatomy. The main structures and layers of the eye are detailed.

The retina catches the light via its photoreceptors and pigmented epithelial cells. The photopigment molecules of these photoreceptors absorb the light, leading to a change in electrical signals. In the human retina there exist around 125 million of photoreceptors: 120 million of rods and 5 million of cones, approximately. Rods are cylindrical, long structures extremely sensitive to light and responsible to transmit shades of grey to the brain. Cones, on the other hand, are thicker, shorter cells which are able to capture fine detail and colour, provided they received enough light. The cones possess three different

types of photopigment and each of them reacts to different ranges of light frequency giving place to the three primary colours (red, blue and green).

The conversion of light energy to electrical impulses (①,② in Figure 2.2) arrives to the ganglion cells which generate an action potential (③) that begins its journey to the brain (④,⑤). Approximately 1.2 million axons of ganglion cells converge to form the optic nerve, which serves as highway to electrical signals towards the brain. The brain transforms the electrical signal into a visual perception which is processed in the visual cortex (Galloway et al., 2016). The whole overview of the process can be observed in Figure 2.2.

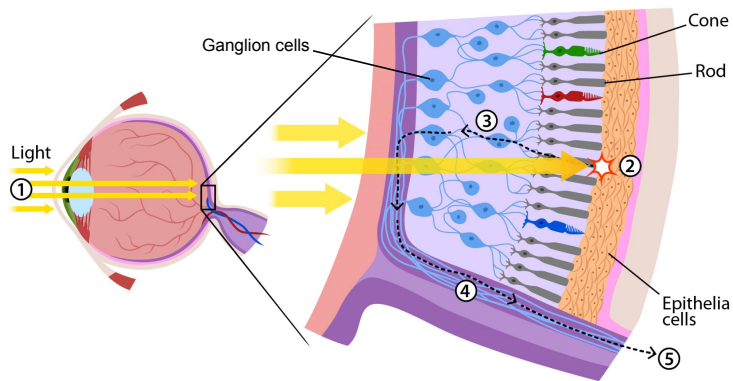


Figure 2.2: Overview of the whole vision process. From the light reception in the retina layer to the electrical signal generation and transfer.¹

The retina is the focus of this work because the ophthalmic pathologies under study are manifested in this layer of the human eye. The anatomical structures of the retina play an important role in the diagnose of some eye diseases. They are detailed below and can be observed in Figure 2.3.

- The optic nerve and optic disk. The optic nerve is the pathway between the retina and the brain visual processing centre. The circular to oval area where the optic nerve crosses through the posterior fundus of the eyeball is called the optic disk. It measures approximately 1.5 mm in diameter and can be seen as a yellow disk in Figure 2.3. Because the optic disk contains no photoreceptors it is also called blind spot. This structure is where the central retinal artery enters and the central retinal vein exits. In the centre of the disk there exists a bright depression called optic cup. The optic cup is not occupied by nerve fibers. The neuroretinal rim is

¹Image adapted from the Arizona state university (www.sols.asu.edu)

the area of the optic disk which contains the neural elements and it is located between the edge of the disk and the optic cup.

- Blood vessels. The retinal blood vessels are responsible to provide oxygen and nutrients to the retina. The retinal vascular network is composed by arteries and veins. The central retinal artery enters the eye through the optic nerve and splits itself into the superior (upper) and inferior (lower) branches. These branches are ramified in smaller vascular paths until they form a network of very thin vessels called capillaries. It is mainly at the capillaries where the blood leaves the oxygen and nutrients to feed the retinal tissue. At the same place, the carbon dioxide and waste products leave the retina through the blood to be expelled. Most of the eye diseases affect to the retinal blood vessels by either blocking these capillaries or causing them to become leaky. The capillaries join up composing branch veins. This veins converge at the optic nerve to form the central retinal vein that dives into the optic nerve on its way to the heart. The main branches of the central retinal artery and vein form four vascular arcades.
- Macula and fovea. The macula and fovea are the most sensitive parts of the retina providing for finest central vision. The macula is a highly pigmented area near the centre of the retina with a diameter of 5 mm and it is located two optic disk diameter temporal to optic disk. The highest concentration of cones is found in the fovea, a small pit located at the centre of the macula with a diameter of approximately 1 mm in the human eye. This high density of cones is responsible for discerning minute details. The cornea and lens focus light onto this small area in order to produce the clearest, most detailed image.

2.2 Fundus photographs

Fundus cameras are simple-to-use revolutionary devices to capture retina information using non-invasive protocols. It is true that in some cases fundus photographs are not sufficient precise to establish a final diagnosis and ophthalmologists require more sophisticated, invasive and expensive methods for determining eye diseases. However, in screening campaigns and primary care centres, retinal images are of vital importance due to the modest cost of the device and the possibility of using non-mydratic fundus cameras avoiding the need of patients' pupil dilation.

Figure 2.4(a) shows a non-mydratic retinal camera from TOPCON manufacturer. This kind of devices capture a specific region of the retinal tissue

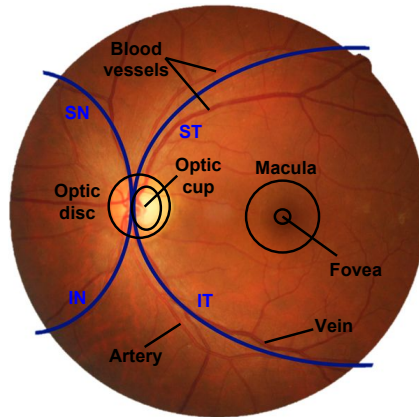


Figure 2.3: Main parts of the retina: blood vessels, optic disc, optic cup, macula and fovea. The vascular arcades formed by blood vessels are marked in blue: superotemporal (ST), inferotemporal (IT), superonasal (SN) and inferonasal(IN).

determined by a field of view (FOV) angle to be selected by the specialist in the acquisition process. Last models of retinal cameras provide high image quality up to 16.2 megapixels and they are equipped with automatic focusing and capture making its use quick and simple. The incorporated filters and internal firmware allow the user to obtain colour and red-free images. Figure 2.4(b) shows a typical image obtained by such simple and portable device. Of course, the specifications and the image quality of this hardware are considerably reduced but its dimensions allow to carry out screening campaigns in nursing homes or special centres.

For the automatic screening of eye diseases, RGB fundus images provide the required visual information to discriminate between healthy and pathological retinas. The fundus images are acquired centred on the macula or centred on the disk depending on the pathology to be diagnosed. The acquisition room must have particular lighting conditions and the person responsible to capture the image must follow the acquisition protocol and a rigorous configuration of the camera. Figure 2.5(a) shows a colour fundus image captured with a non-mydratiac retinal camera and a field of view angle of 40° . As can be observed the retinal tissue is delimited by a circular lens defined by the FOV angle. For the development of algorithms able to automatically detect retinal pathologies, the extraction of the information contained only inside the lens must be guaranteed (Figure 2.5(b)).

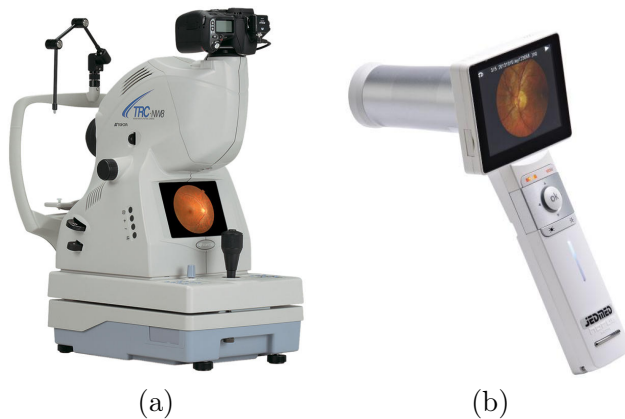


Figure 2.4: Non-mydiatic retinal cameras. (a) TOPCON TRCNW8 camera and (b) JEDMED 40° FOV portable camera.

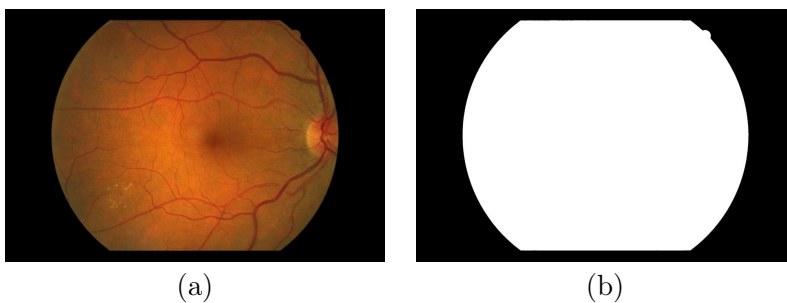


Figure 2.5: (a) Colour retinal image and (b) its field of view mask.

2.3 Eye diseases and fundus lesions

Diabetic retinopathy and age-related macular degeneration are two of the most common diseases in the current society that provoke retinal damage and can cause visual impairment and, even, permanent blindness. These diseases are detectable through fundus inspection (Abramoff et al., 2010; World Health Organization, 2015). In this chapter, the characteristic signs of these eye diseases will be described.

By way of introduction, in Figure 2.6 the effects of DR and AMD on the vision can be observed.

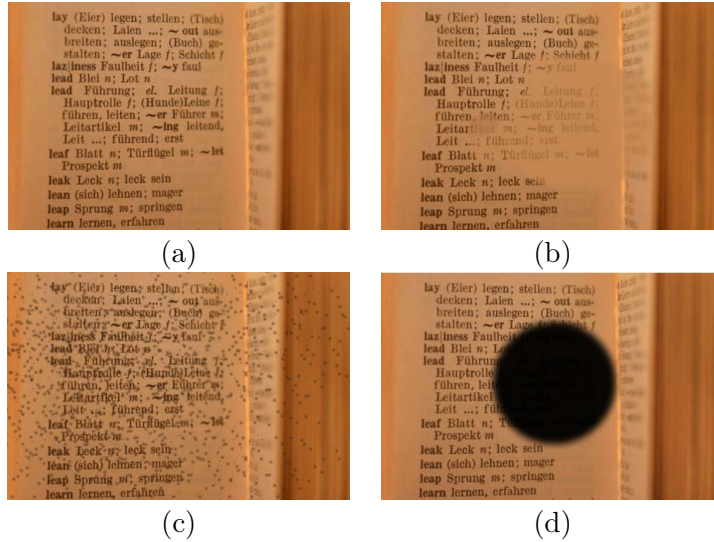


Figure 2.6: Effect of the eye diseases studied in this thesis on the vision: (a) Normal vision, (b) with non-proliferative diabetic retinopathy, (c) with proliferative diabetic retinopathy and (d) with age-related macular degeneration.²

2.3.1 Diabetic Retinopathy

Insulin plays a vital role in the process of food absorption by the human body. Insulin is responsible to transport glucose from bloodstream to muscles, fats and hepatic cells, where it can be accumulated or used as energy source. Diabetic patients suffer from high sugar levels in blood, because the pancreas does not produce sufficient insulin to distribute sugar (diabetes type I) or its body is resistant to the insulin action and the cells of the organism do not assimilate the insulin in a normal way (diabetes type II). So, diabetes is a metabolic and chronic disease characterised by high levels of sugar in blood which can produce serious damage to the eyes, heart and nerves.

Diabetic Retinopathy (DR) is a common consequence of diabetes that causes about five percent of world blindness. In fact, signs of this eye disease are found in around 33% of the diabetic population (World Health Organization, 2015) and 80% of patients who have suffered diabetes for more than ten years are

²Images provided by the American Academy of Ophthalmology (www.aao.org)

affected by DR (Wang et al., 2000). To ensure that DR treatment is applied on time, fundus images from diabetic patients must be examined at least once a year (Fong et al., 2004).

Vision problems related to diabetic retinopathy are registered in advanced stages of the disease, in which retina changes have progressed to a level where treatment is complicated and costly. For this reason, the early detection of DR is essential to guarantee the maintenance of the vision. Fluorescein angiography and optical coherence tomography (OCT) are the most common methods used by experts for diabetic retinopathy identification. However, these techniques are not appropriate in screening campaigns because angiography is an invasive procedure in which a contrast agent must be injected and OCT is an expensive device only available in the ophthalmology centres. Luckily, the first signs of DR can be noticed using fundus photographs acquired by means of a retinal camera. The affordable price of a retinal camera in comparison with an OCT, the high level of non-invasiveness provided by this device and its usability allows primary health centres and other institutions to early diagnose DR during screening campaigns. This strategy ensures regular screening that facilitates the effectiveness of the administered treatment.

For automatic screening of DR through fundus imaging, systems are based on the detection of specific anomalous patterns or lesions which are reported in Figure 2.7 and detailed below:

- **Microaneurysms (MA).** Small saccular dilatations of capillaries that appears as round spots of dark red colour with sharp edges in the retina background. MA usually measure between 15 and 125 microns and they are propitiated by venous and arterial capillaries. The importance of this lesion relies in the fact that MA are the first identifiable symptom of diabetic retinopathy.
- **Exudates.** Deposits of lipids and proteins in the retina. They are typically bright lesions of yellowish-white colour with prominent and irregular edges. They can appear as isolated lesions or in groups composing ring or star shapes taking as a focal point a mycroaneurism.
- **Cotton wool spot.** This kind of lesion is caused by damage to nerve fibres and are a result of accumulations of axoplasmic material within the nerve fibre layer. It can be seen as superficial retinal infarcts. Cotton wool spots appear as fluffy white patches on the retina and they can be confused with exudates.

- Haemorrhages. Wide accumulation of blood in the retina. Haemorrhages become visible when the DR degree progresses. They indicate an ischemia increase (lack of oxygen) of the retina.

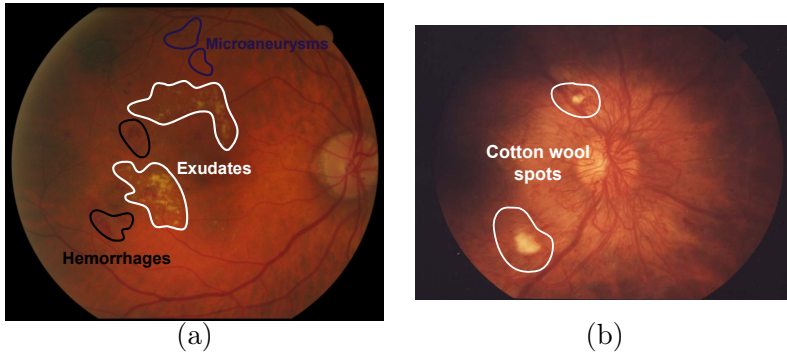


Figure 2.7: Signs of diabetic retinopathy in fundus images: (a) Exudates, microaneurysms, haemorrhages and (b) cotton wool spots can be observed.

There exist two types of DR: Non-Proliferative and Proliferative Diabetic Retinopathy (PDR) (American Academy of Ophthalmology, 2016). Non-Proliferative DR (NPDR) is the earliest stage of the pathology and it is characterised by deposits of extra fluid and small amounts of blood, from the vessels into the retina, exudates and microaneurysms respectively (Figure 2.7). NPDR usually progresses to develop PDR because many blood vessels in the retina are closed hindering the proper blood flow. The retina responds to this fact by growing new blood vessels (this process is called neovascularization). However, these new blood vessels are abnormal and do not provide the correct blood flow to the retina. Their fragility often produce bleeding into the vitreous (haemorrhages) propitiating shades and blocking the vision. PDR can affect both peripheral and central vision, for this reason it is more aggressive than NPDR. Figure 2.8 identifies both types of diabetic retinopathy.

2.3.2 Age-related macular degeneration

As explained in Section 2.1, the macula is made up of millions of light-sensitive cells that provide clear central vision. Age-related macular degeneration (AMD) is a degenerative eye condition which degrades progressively the macula. The loss of central vision, propitiated by AMD, interferes in daily tasks such as appreciate faces, cooking, driving, reading, writing, etc. This eye disease is the leading cause of vision loss among people age 50 and older (American Academy of Ophthalmology, 2015). Its progressions is so slow and

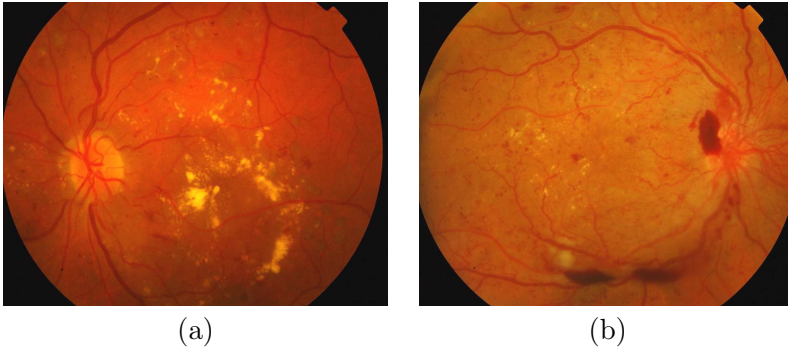


Figure 2.8: Grades of diabetic retinopathy: (a) Non-proliferative DR and (b) Proliferative DR (in which neovascularization is visible).

in the first stages patients only notice small changes in their vision, although in advanced phases AMD can provoke a total loss of central vision. The effects of this pathology can be retarded and mitigated, by means of vitamins and minerals, if an early detection is carried out (American Academy of Ophthalmology, 2015).

Drusen are the accumulation of yellow deposits in the macula (Figure 2.9(a)) and they are the first signs of AMD. The progress of this pathology can be divided in three periods or stages defined in part by the size and number of drusen. The early AMD is characterised by the presence of a reduced number of round and small ($< 63 \mu\text{m}$ in diameter) or medium-sized ($63\text{-}125 \mu\text{m}$ in diameter) drusen with sharp edges (hard drusen). People with early AMD typically do not suffer vision loss and have a low risk of progressing to late AMD after 5 years in either eye (Age-Related Eye Disease Study Research Group, 2001). When drusen edges are not clearly defined and they appear in clusters (soft drusen) the risk of AMD progress is considerable. Intermediate AMD is manifested by means of an extensive number of medium drusen or one or more large drusen ($\geq 125 \mu\text{m}$ in diameter). Pigment changes in the retina can be also noticeable. The progression to late AMD at 5 years of patients of this group is approximately 18%. The most problematic stage of this eye disease is known as advanced or late AMD and is classified into two types. Geographic atrophy or dry AMD is characterised by the gradually damage of the supporting tissue beneath the macula and its light-sensitive cells responsible of transmitting the visual information to the brain (Figure 2.9(b)). On the other hand, in neovascular or wet AMD, abnormal blood vessels grow underneath the retina. These fragile vessels can leak fluid and blood which

may damage the macula generating even disciform scars (Figure 2.9(c)). The damage may be rapid and severe instead of dry AMD. It is possible to suffer both types of AMD in the same eye, and either condition can appear first.

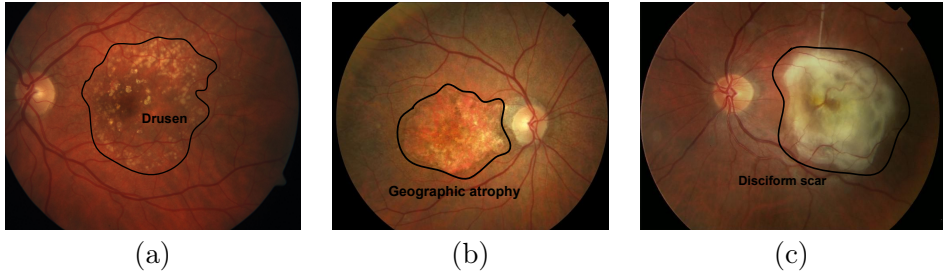


Figure 2.9: Signs of age-related macular degeneration: (a) Drusen (b) Geographic atrophy of the retinal pigment epithelium and (c) disciform scar.

2.3.3 Relevant groups

Some of the most relevant groups that worked or are still working on retinal damage detection for the automatic diagnosis of DR and/or AMD are detailed below:

International groups

TeleOphta³ was a collaborative research project, funded by the National Agency of Research (ANR) through the program TecSan. Its main objective was to develop the tools for fundus image classification, in order to assist in the detection of retinal diseases, especially of diabetic retinopathy. Some research works developed by Decenci ere et al. (2013) and Zhang et al. (2012) and the publication of the E-OPHTHA fundus database were important results of this project.

The department of ophthalmology and visual sciences belonging to the University of Iowa⁴ is directed by one of the most influential researchers in the mission of ending blindness with artificial intelligence, professor Michael D. Abr amoff. This research group is responsible of important scientific publications in this field by using both classical (Abr amoff et al., 2013, 2010; Niemeijer et al., 2010) and novel approaches (Abr amoff et al., 2016; Costa et al., 2017).

³www.cmm.enscm.fr/Projects/TeleOphta/index.html

⁴www.uihealthcare.org

VisionQuest Biomedical⁵ (Albuquerque) focused on diabetic retinopathy screening although they also worked on the detection of other pathologies such as age-related macular degeneration (Agurto et al., 2011a,b). The automatic screening system EYESTAR⁶ was the result of this research project.

The University of Aberdeen (Scotland) has worked on an automated grading of diabetic retinopathy in primary care screening (Goatman et al., 2011; Philip et al., 2007)⁷. The only existing CE-accredited commercial system in terms of analysis for automatic fundus screening was developed by this research group.⁸

The Retinal Image Computing & Understanding group⁹ at the University of Lincoln (United Kingdom) have some publications focusing on the diabetic retinopathy detection (Massey et al., 2009). In addition, they were the coordinators of a European project, named "REVAMMAD - RETinal Vascular Modelling, Measurement and Diagnosis"¹⁰, included in the Seventh Framework Programme.

The Google team¹¹ responsible of machine learning, computer vision and health research areas are focused on diabetic retinopathy detection. Their work (Gulshan et al., 2016) is a key contribution to the state of the art in the automatic identification of retinal diseases based on deep learning techniques.

The collaboration between the University of Paris Descartes, the École Polytechnique de Montréal and the Diagnos inc. company propitiated an important research work for the scientific community, in which a public database for detecting age-related macular degeneration was proposed (Garnier et al., 2014).

The LaTIM¹² is a research team made up of the University Hospital of Brest, the Faculty of Medicine and Telecom Bretagne. The IT2IM group belonging to LaTIM and directed by M. Lamard and G.Quellec among others published some research works focusing on the automatic identification of the common signs of diabetic retinopathy (Quellec et al., 2012, 2011; Zhang et al., 2014).

⁵www.visionquest-bio.com

⁶www.visionquest-bio.com/eyestar-tm.html

⁷www.abdn.ac.uk/heru/research/assessment-of-technologies/projects/the-role-of-automated-grading-of-diabetic-retinopa/

⁸www.shil.co.uk/Products/igradingtm-platform-diabetic-retinopathy-screening-software.html

⁹www.aldiri.info/Default.aspx

¹⁰<http://revammad.blogs.lincoln.ac.uk/revammad/>

¹¹<http://research.google.com/pubs/105170.html>

¹²<http://latim.univ-brest.fr/index.html>

National groups

The University of Huelva worked on an expert system for early automated detection of diabetic retinopathy¹³. At this moment, there are not results of this system but some algorithms for retinal structure segmentation have been published (Aquino et al., 2010; Marin et al., 2011).

The VARPA group¹⁴ at University of A Coruña worked mainly on drusen and dark lesion detection. As result of their work, a web application for diagnosis and disease monitoring known as SIRIUS (Ortega et al., 2010) was developed.

A consortium composed by the *Centro Regional de Servicios Avanzados S.A.*, the Biomedical Engineering Group (BIC) of the university of Valladolid and the *Instituto de Oftalmología Aplicada* was funded to develop automatic algorithms for retinal damage detection under the project entitled “*Análisis automático de imágenes de fondo de ojo como implementación a los sistemas de cribado de la retinografía diabética (A2IFO)*”¹⁵. Roberto Paredes (coordinator of BIC) published previous works on this field (García et al., 2009; Niemeijer et al., 2010; Sánchez et al., 2009).

The signal processing and communication group from the University of Sevilla worked on the automatic detection of DR under the project “*Evaluación de la potencia diagnóstica de la estrategia de fotodetección en la retinopatía diabética con un algoritmo optimizado de detección automatizada*”. Different publications in this field demonstrate the successful achieved during the work (Abbas et al., 2017; Fondón et al., 2012; Jiménez et al., 2010).

The Computer Vision and Behaviour Analysis laboratory (CVBLab)¹⁶ belonging to the *Instituto de Investigación e Innovación en Bioingeniería (I3B)* at the Polytechnic University of Valencia started working on the identification of fundus pathologies seven years ago. During those years, a lot of works related to the automatic detection of DR and AMD have been published (Colomer et al., 2017a,b, 2018; Morales et al., 2017, 2013, 2014). In this way, the research project entitled “Fundus image processing for automatic screening of ophtalmological diseases - ACRIMA” (TIN2013-46751-R) was funded by the *Ministerio de Economía y Competitividad* of Spain.

¹³www.uhu.es/retinopathy/index.php

¹⁴www.varpa.es

¹⁵www.a2ifo.es

¹⁶www.cvblab.webs.upv.es

Retinal image pre-processing

In this chapter, fundus pre-processing methods to reduce the high variability among public databases are presented. In addition, a novel inpainting method is proposed to reduce the contribution of blood vessels in the posterior retinal damage detection.

In this chapter, works previously published, although adapted for this manuscript, are included:

- *Section 3.3 - Normalisation of public fundus databases.*

Colomer, A., Naranjo, V., Angulo, J. (2017). Colour normalisation of fundus images based on geometric transformations applied to their chromatic histogram. In *24th IEEE International Conference on Image Processing (ICIP 2017)* (Article in Press).

- *Section 3.4 - Blood vessel inpainting.*

Colomer, A., Naranjo, V., Engan, K., Skretting, K. (2015). Retinal vessel inpainting using recursive least square dictionary learning algorithm. In *5th International Conference on Image Processing Theory, Tools and Applications (IPTA 2015)*, pp.429-433.

Colomer, A., Naranjo, V., Engan, K., Skretting, K. (2017). Assessment of sparse-based inpainting for retinal vessel removal. In *Signal Processing: Image Communication*, vol.59, pp.73-82.

Contents

3.1	Introduction	25
3.2	Material	25
3.3	Normalisation of public fundus databases	27
3.3.1	Spatial normalisation	28
3.3.2	Colour transformation	29
3.4	Blood vessel inpainting	32
3.4.1	Sparse representation & dictionary learning	35
3.4.2	Inpainting using Sparse Representation and Dictionary Learning	39
3.4.3	Measuring the performance of blood vessel inpainting	41
3.4.4	Assessment of retinal blood vessel inpainting	44

3.1 Introduction

Image pre-processing is a transformation process applied to an original image (i.e. image acquired by any sensor or instrument) with the objective of improving its quality or reducing the noise. In computer vision tasks, the pre-processing stage optimises the original input image of an automatic system in order to improve the effectiveness and robustness of the algorithmic stage developed to solve the problem under study.

In this thesis, the retinal tissue is studied by means of digital fundus images with the aim of identifying abnormal regions or signs of the most common eye diseases. However, some *a priori* inconveniences are present in the raw fundus photographs:

- The lack of rigor in the accomplishment of the image acquisition protocol produces a high variability of the fundus photographs: Non-uniform illumination, poor contrast and artifacts or random brights are the most common undesirable effects in this type of images.
- Some fundus image particularities propitiated by physiological eye properties and artefacts in the acquisition process could affect in the discrimination of retinal lesions due its similarity in appearance.
- Due to the features of some retinal structures, some of them can hinder in the automatic detection of bright and dark lesions. Some pre-processing steps are necessary to avoid these structures in the posterior fundus image analysis stage.

In this chapter, novel pre-processing methods in fundus image are proposed to solve some of the aspects mentioned above, in which, to the best of the author's knowledge, no previous works in the state of the art in retina image exist.

3.2 Material

Three public fundus databases were used to validate the pre-processing methods proposed in this chapter. Next, a description of each of them is detailed.

E-OPHTHA

E-OPHTHA (Decencière et al., 2013) is a database of fundus images especially designed for diabetic retinopathy screening¹. This public database is divided in two subsets depending on the lesion type: exudates and microaneurysms. These lesions are manually annotated by experts and the ground-truth is provided. The bright lesion subset (*E-OPHTHA_exudates*) is composed by forty seven pathological images (Pathological_EX) and thirty five images with no lesion (Healthy_EX). *E-OPHTHA_microaneurysms* subset makes available 148 images with microaneurysms or small haemorrhages (Pathological_MA) and 233 healthy images (Healthy_EX). All the retinal images were acquired with the same field of view, i.e. 40° , and they present different spatial resolutions, as it is reported in Table 3.1.

Image resolution (pix.)	1440 × 960	1504 × 1000	2048 × 1360	2544 × 1696
Pathological_EX	13	2	9	23
Pathological_MA	54	5	17	72
Healthy_EX	17	0	5	13
Healthy_MA	96	39	1	97

Table 3.1: Distribution of the *E-OPHTHA* images according to their resolution.

DRIVE

DRIVE (Digital Retinal Images for Vessel Extraction)(Staal et al., 2004) is a public database established to enable comparative studies on segmentation of blood vessels in retinal images². The database is composed by forty fundus photographs of which thirty-three do not show any sign of diabetic retinopathy and seven show signs of mild early diabetic retinopathy. Retinal images of dimensions 768×584 pixels were acquired using a Canon CR5 non-mydratic 3CCD camera with a 45 degree field of view.

The set of images was divided into training and test subsets. For the training subset of images, a single manual segmentation of the retinal vasculature is provided while for the test cases two experts manually annotated the pixels belonging to the blood vessel network. For each image, the field of view is delineated by a mask image provided in the public database.

¹www.adcis.net/en/Download-Third-Party/E-Ophtha.html

²www.isi.uu.nl/Research/Databases/DRIVE

STARE

STARE (S**TR**uctured Analysis of the Retina) database (Hoover et al., 2000) is a set of 20 images (700×605 pixels) designed for blood vessel segmentation³. For each image, two hand labelled vessel networks are provided by different experts.

3.3 Normalisation of public fundus databases

The fact of having access to wide public databases is really helpful in the process of developing and validating automatic models able to detect or classify pathological patterns extracted from the retinal images. However, the high heterogeneity among fundus images hinders the creation of robust and reliable classification models.

The most evident factor that propitiates a high level of variability among public fundus databases is the image dimensions. These dimensions can be variable even among images belonging to the same database as a consequence to the configuration of the retinal camera. This fact propitiates that the size of the anatomical structures and lesions may not be comparable to each other, hindering in the automatic detection process of the eye diseases under study. In this section the spatial normalisation method used during this PhD thesis will be described.

Other more problematic factors that affect to the heterogeneity of fundus databases are artefacts and noisy images (Figure 3.1(a) and Figure 3.1(b)). They are really common in the databases due to the uncalibrated acquisition process. In addition, non-uniform illumination in retinal images is quite common. This fact is propitiated by the different lighting conditions in the acquisition rooms. Besides the problems in the acquisition process, the retina variability can produce differences in the images such as highlights near the vessels characteristic of young retinas (Figure 3.1(c)) and the fundus colour that depends on age, ethnicity, differences in retina pigmentation and other anatomical human elements (Figure 3.1(d)). All these factors of variability may produce the appearance of false anomalous patterns or the occlusion of true retinal lesions propitiating the confusion of automatic classification models.

The factors described above are important for the success and reliability of the computer aided diagnosis of retinal pathologies. Previously to apply algorithms

³www.cecac.clemson.edu/~ahoover/stare

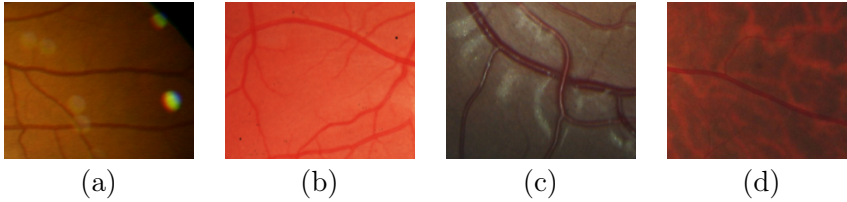


Figure 3.1: Typical inconveniences in fundus databases. (a) Noisy images, (b) Images with artefacts, (c) Highlighted retina, (d) Tessellated images due to different ethnicity.

able to detect the pathology under study, a database normalization is an essential step in order to analyse the images under the same conditions.

Image pre-processing methods applied to fundus images have been studied in the literature. In (Rathinam and Selvarajan, 2013) basic pre-processing techniques such as contrast adjustment, adaptive histogram equalisation, median filtering and average filtering are applied before glaucoma detection. In most of the state-of-the art works (Aliaa et al., 2007; Marrugo and Millán, 2011; Meier et al., 2007; Rathinam and Selvarajan, 2013), the non-uniform illumination is the main correction explored. In (Aliaa et al., 2007) a colour normalisation based on histogram equalisation and shape matching is applied on each RGB component separately with the aim of detecting the retinal anatomy.

Colour information allows the visual discrimination of different lesions and anomalous patterns and it is essential in the automatic detection of eye diseases. Fundus photographs must be colour-normalised to reliably compare images belonging to different patients and coming from different acquisition devices. In this section, a chromatic colour-normalisation involving the three RGB components is presented. In this novel method, different geometric transformations are applied to the chromaticity diagram extracted from the images to be corrected.

3.3.1 *Spatial normalisation*

Fundus images belonging to the same and different databases present distinct resolutions and in some cases retinal structures and lesions are not comparable, for this reason a spatial normalisation is essential to solve this problem. In this work, the spatial normalisation proposed by (Zhang et al., 2012) is applied.

Given that the same capture angle β used in the acquisition process (Figure 3.2(a)), the width of the field of view mask (Figure 3.2(b)) and the pixel size are linearly correlated. Let I_i be a retinal image and D_{ref} a specific field-of-view width of reference, the resizing factor (RF_i) can be computed by means of the following equation:

$$RF_i = \frac{D_{ref}}{D_i} \quad (3.1)$$

where D_i is the field-of-view width of the image to be resized I_i .

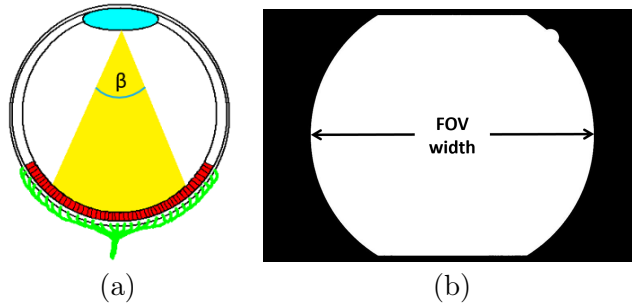


Figure 3.2: (a) Representation of the capture angle β and (b) the width of the field of view mask.

It is worth noting that bicubic interpolation is used for resizing; the output pixel value is a weighted average of pixels in the nearest 4-by-4 neighbourhood.

Applying the computed resizing factor to all the images of the specific database, the retinal structures and the possible lesions can be compared. Note that if the external mask that delimitates the FOV it is not provided in the database, it is easy to extract by thresholding the red channel of the fundus images.

3.3.2 Colour transformation

In this thesis, a novel colour normalisation method based on geometric transformations of the chromatic distribution of fundus images is proposed with the aim of increasing the colour homogeneity of public databases. For this purpose a fundus image acquired under ideal conditions and showing a normal retina background colour (i.e. image without noise, tessellations or brights characteristics of young retinas) should be selected as “reference image”. The objective is to transform the chromatic diagram of the images to be corrected

into the “reference” one by applying several geometric transformations on this diagram.

Let $I_i = (R_i, G_i, B_i)$ be a colour fundus image, we can obtain a luminance normalisation as follows:

$$I_{i_{norm}} = \left(\frac{R_i}{L_i}, \frac{G_i}{L_i}, \frac{B_i}{L_i} \right) = (r_i, g_i, b_i) \quad (3.2)$$

$$\text{with } L_i = R_i + G_i + B_i.$$

The distribution of intensity values in the channels r_i and g_i can be represented in a 2D histogram, the chromaticity diagram or chromatic distribution (CHD). In Figure 3.3 is possible to observe the CHD of an image I_i . The centre $c_i(c r_i, c g_i)$ of this CHD is given by the coordinates of the maximum 2D histogram value. This dominant mode of the chromatic distribution is associated to the colour of the retinal background. Also, we can compute the angle (θ_i) formed by the major axis of the fitted ellipse to the CHD and the horizontal axis (Figure 3.3).

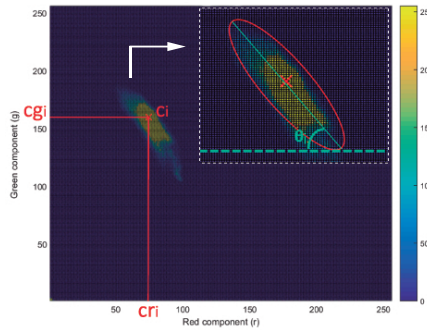


Figure 3.3: Chromaticity diagram from a fundus image and visual description of $c_i(c r_i, c g_i)$ and θ_i parameters.

The colour transformation of the target image I_i , taking as reference the image I_{ref} , is carried out in two steps. In the first one, a rotation of the CHD is applied, according to:

$$\begin{bmatrix} r_i' \\ g_i' \end{bmatrix} = \begin{bmatrix} r_i & g_i \end{bmatrix} \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) \\ \sin(\Delta\theta) & \cos(\Delta\theta) \end{bmatrix} \quad (3.3)$$

where $\Delta\theta$ is the difference between the angle extracted from the target image to be transformed (θ_i) and the angle computed for the reference image (θ_{ref}).

In a second step, a translation of the centre of the CHD is performed following this equation:

$$\begin{bmatrix} r_i'' \\ g_i'' \end{bmatrix} = \begin{bmatrix} r_i' \\ g_i' \end{bmatrix} - \begin{bmatrix} \Delta c_r \\ \Delta c_g \end{bmatrix} \quad (3.4)$$

where $\Delta c_r = cr_i - cr_{ref}$, being cr_i the red coordinate belonging to the centre of the CHD extracted from the image to be corrected and cr_{ref} the red coordinate belonging to the centre of the CHD extracted from the reference image. In the same way, $\Delta c_g = cg_i - cg_{ref}$ where cg indicates the green coordinate of the CHD centres.

After the colour transformation, the components r_i'' and g_i'' must be de-normalised using the following equations:

$$b_i'' = 1 - r_i'' - g_i'' \quad (3.5)$$

$$\tilde{I}_i = (r_i'' L_i, g_i'' L_i, b_i'' L_i) = (\tilde{R}_i, \tilde{G}_i, \tilde{B}_i) \quad (3.6)$$

Finally, the $\tilde{R}_i, \tilde{G}_i, \tilde{B}_i$ components are the output of the colour transformation process detailed graphically in Figure 3.4. In this figure it is possible to observe the rotation and translation operations applied to the CHD extracted from a target image taking into account a reference image.

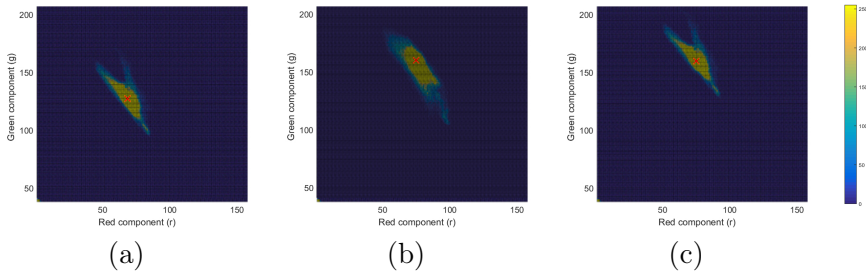


Figure 3.4: Colour normalisation process by linear transformation of chromatic histograms. (a) CHD from the image to be colour-transformed, (b) CHD from the reference image (c) CHD from the resulting image.

As it was mentioned in Section 3.3, the objective of the introduced colour normalisation method is to reduce the high colour variability between images belonging to the same or different databases. The healthy and pathological images from the *E-OPHTHA exudates* database were colour-transformed according to a reference image selected from this database. The image selected as reference was acquired using the appropriate protocol to capture fundus images with good illumination conditions and without special anatomical characteristics (as highlights, tessellations, etc.). Figure 3.5 shows several original images (b-g) and their corresponding colour transformations (h-m) taking as reference the image displayed in Figure 3.5(a).

Apart from these qualitative results, the impact of the proposed colour transformation into the automatic detection of fundus lesions will be quantified in Section 6 by means of a classification system based on texture and shape descriptors and advanced machine learning algorithms.

3.4 Blood vessel inpainting

Blood vessels cover a high percentage of the fundus image hindering the automatic detection of important structures as optic disc, optic cup and macula among others. Vessels are also considered as noise or artefacts that hamper the segmentation of different lesions such as exudates, microaneurysms and drusen among others and also the classification of pathologies based on background textures. A possible procedure to avoid blood vessels is to consider these structures as missing pixels and trying to restore them using the background.

Image inpainting is a technique for restoring missing or damaged areas in a digital image. Image inpainting has been recently used in a lot of applications such as in image restoration from scratches or text overlays, for removal or replacement of selected objects, for disocclusion in image-based tracking in order to recover missing persons or objects in specific frames, etc.

Inpainting methods assume that pixels in the known and unknown areas of the image have the same statistical properties or geometrical structures. Different algorithms exist in the literature but all of them can be grouped in three categories or families.

Diffusion-based inpainting covers a family of methods exploiting the smoothness principle by parametric models or partial differential equations (PDEs) to propagate local structures from the exterior to the interior of the unknown area. These algorithms are designed for completing straight lines and curves,

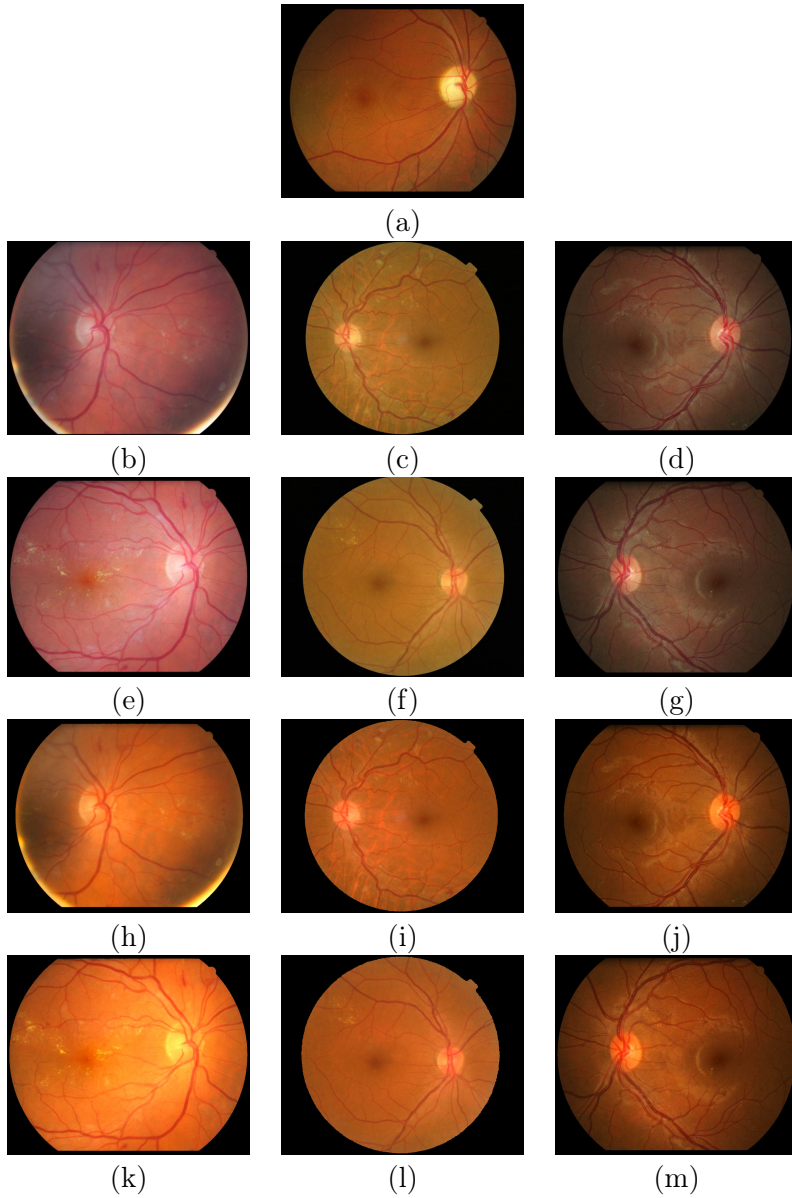


Figure 3.5: Colour normalisation results. (a) Reference image, (b-g) original images and (h-m) colour transformed images.

for processing piecewise smooth images, for propagating strong structures and filling small gaps. Using this kind of methods to fill large gaps produce a blur result (Guillemot and Le Meur, 2014). Different state-of-the-art approaches exist using particular models (linear, nonlinear, isotropic or anisotropic) to favour the propagation in specific directions (Bertalmio et al., 2001; Telea, 2004; Tschumperlé, 2006).

The second category is exemplar-based inpainting that exploits image statistical and self-similarity principles. Using methods belonging at this category, the unknown region is filled by means of copying and stitching the best match pixel or patch from the sample texture. Exemplar-based methods follow two different ways. In pixel-based algorithms the unknown area can be filled copying the central pixel of the most similar patch from the sample texture. The idea of patch-based algorithms is to fill the unknown region in one step by sampling and copying texture patterns (entire patches) from the source (Doria and Radke, 2012; Huang and Hsiao, 2010; Liang et al., 2001). Obviously, patch-based solutions are more computationally efficient than pixel-based approaches, for this reason nowadays exemplar-based methods refer to patch-based methods.

Different researchers focus their efforts on improving and optimising patch-based methods trying to find the most efficient way to search the best match patch. One of them are the sparse-based algorithms (Elad, 2010; Mairal et al., 2008a) where the image is supposed to be sparse in a given domain and known and unknown parts are assumed to share the same sparse representation. The domain can be represented by a dictionary matrix which can be learned using dictionary-learning methods.

The last category, hybrid methods, is based on separating the image information (texture and structure), and thereafter inpaint these components separately using the most appropriate method (diffusion-based inpainting or exemplar-based inpainting).

Another important line of research closely related to image inpainting is the quality assesment of the resulting image (Dang et al., 2013a,b; Voronin et al., 2015). The efforts are focused on looking for an objective measure that automatically predicts the human perception of an inpainted image without human help. When performing inpainting on real world data without a reference image available (non-artificial inpainting), a quality assessment of the inpainting is not straight forward. In this section, two different ways to measure the quality for non-artificial inpainting are proposed and validated in a real application: retinal blood vessel removal.

Different authors (Giachetti et al., 2011; Meier et al., 2007; Morales et al., 2013) have inpainted blood vessels by using diffusion-based algorithms but only as a middle step of their purpose, in other words, without any kind of evaluation of the inpainting method used. For example, Morales et al. (2013) uses inpainting techniques in order to segment the optic disk in fundus images. Inpainting quality evaluation, in this application, is an unsolved problem because the reference image (fundus image without vessels) does not exist. Image inpainting assessment is an essential step for our future purposes because the better the vessels are removed, the more accurate the texture classification or the lesion detection will be.

In this thesis, a sparse-based inpainting method with spread neighbourhoods is proposed to inpaint blood vessels in fundus images and a novel metric to assess inpainting methods is presented.

Three experiments were carried out in order to find the best configuration of the proposed inpainting method applied to retinal vessel removal. The influence of using spread neighbourhoods, the variation of the sparseness factor, the algorithm used in the dictionary learning process and the way to build the RGB dictionaries are some of the key concepts studied in these experiments.

With the aim of validating the inpainting method, the blood vessels of grey-scale and colour retinal images, from *DRIVE* and *STARE* public databases, were inpainted. The resulting images were evaluated using two metrics. Firstly, the traditional Recovery Error is computed on some artificially made vessels added to the image. On the other hand, we also present a novel inpainting quality assessment index based on the idea exposed by Wang et al. (2008). By means of this index, a comparison between our sparse-based method and a diffusion-based inpainting algorithm, via partial differential equations (PDE), is carried out.

3.4.1 Sparse representation & dictionary learning

Sparse representations of signals are meaningful under the assumption that the signal can be represented sparsely in a domain, usually represented by atoms collected as columns in a dictionary matrix. The dictionary D is a matrix $N \times K$, which contains K prototype signals of length N , also referred to as atoms. The model assumes that for any signal x , there exists a sparse linear combination of atoms from D that approximates it well. The approximation of x can be written as:

$$x \approx Dw, \quad \|w\|_0 \ll N \quad (3.7)$$

where w is a vector containing the coefficients and most of the entries in w are zero and the operator $\|\cdot\|_0$ counts the number of non-zero elements in a vector. Typically it is assumed that the dictionary is redundant in describing x .

So, given the dictionary D , the approximation \hat{x} of signal x can be written as $\hat{x} = Dw$, and the representation error or residue can be written as $r = x - \hat{x} = x - Dw$. Most of the entries of w are zero, s is the number of non-zero coefficients, and s/N is the sparseness factor.

A common way to find w , i.e the sparse approximation problem, is solving the following equation:

$$w_{opt} = \underset{w}{\operatorname{argmin}} \{ \|w\|_p + \gamma \|x - Dw\|_2^2 \}, \quad p \in \{0, 1\}. \quad (3.8)$$

The problem with $p = 0$ (to minimise the number of non-zero coefficients) is NP-hard, but an approximate solution can be found by greedy methods. Alternatively the problem can be relaxed by setting $p = 1$ (to minimise the sum of absolute values) providing a convex problem that can be solved (LASSO). Both problem resolutions start with an all zero vector w , which is the solution when γ is close to zero. As γ factor increases the solution is getting more dense.

Dictionary Learning is often formulated as the problem of finding a dictionary such that the approximations of many vectors, the training set, are as good as possible given a sparseness criterion on the coefficients, i.e. allowing only a small number of non-zero coefficients for each approximation. Let X be a matrix containing K signals to be represented.

The dictionary learning problem can be formulated as an optimisation problem with respect to the coefficient matrix W and the dictionary D as:

$$\{D_{opt}, W_{opt}\} = \underset{D, W}{\operatorname{argmin}} \sum_{i=1}^K \|w_i\|_p + \gamma \|X - DW\|^2 \quad (3.9)$$

Different methods have been proposed to solve the previous optimisation problem. Most of them solve the problem with a two-step algorithm:

- Step 1: Calculating the vector of coefficients $W = W_{opt}$, using some method for selecting the optimal vector by fixing a dictionary.
- Step 2: Obtaining D by fixing W . Updating the dictionary in this step depends on the dictionary learning method used so, D and W are randomly initialised.

In this work the Recursive Least Squares Dictionary Learning Algorithm (RLS-DLA) (Skretting and Egan, 2010) was employed. In this method a single training vector x_i or a mini-batch (subset) of X is processed in each iteration. To obtain the equations for developing the algorithm, the solution for the first t training values is assumed and the new solution should be computed when the next training vector is included. Let $X_t = [x_1, x_2, \dots, x_t]$ of size $N \times t$ be the matrix containing the training vectors and $W_t = [w_1, w_2, \dots, w_t]$ of size $K \times t$ be the vector with the coefficients, the least square solution for the dictionary is:

$$D_t = (X_t W_t^T)(W_t W_t^T)^{-1} = B_t C_t \quad (3.10)$$

Imagine now that a new training vector $x = x_{t+1}$ is made available. To facilitate the notation we will omit the subscript of time for the vectors. The corresponding coefficient vector $w = w_{t+1}$ must be found using the current dictionary D_i and a vector selection method (in this work the Online Matching Pursuit (OMP) algorithm is used). The approximation and the reconstruction error for the new vector x are:

$$\hat{x} = D_t w = B_t C_t w, \quad r = x - \hat{x} \quad (3.11)$$

Taking into account the new training vector, a new dictionary can be calculated as $D_{t+1} = B_{t+1} C_{t+1}$, where:

$$B_{t+1} = B_t + x w^T \quad \text{and} \quad C_{t+1}^{-1} = C_t^{-1} + w w^T \quad (3.12)$$

The main improvement in RLS-DLA, compared to Least Square Dictionary Learning Algorithm (LS-DLA) (Egan et al., 2007), is that instead of calculating the least squares solution in each step, the matrix inversion lemma (Woodbury matrix identity) can be used to update C_{t+1} directly by:

$$C_{t+1} = C_t - \alpha u u^T \quad (3.13)$$

and in the same way:

$$D_{t+1} = D_t + \alpha r_t u^T \quad (3.14)$$

being

$$u = C_t w \quad (3.15)$$

and

$$\alpha = \frac{1}{1 + w^T u} \quad (3.16)$$

Introducing an adaptive forgetting factor λ , Equations (3.13) and (3.15) are transformed in:

$$C_{t+1} = (\lambda^{-1} C_t) - \alpha u u^T \quad (3.17)$$

and

$$u = (\lambda^{-1} C_t) w \quad (3.18)$$

The advantage of RLS-DLA compared with the Method of Optimised Directions (MOD) (Engan et al., 1999) and K-Singular Value Decomposition (K-SVD) (Aharon et al., 2006) comes with the flexibility introduced by the forgetting factor λ . The idea is to forget more quickly in the beginning and then forget less as learning proceeds and we get more confidence in the quality of the dictionary. This can be done by starting with λ slightly smaller than one and slowly increasing λ towards 1 as learning progresses. The update scheme and the initial value of λ should be chosen so that the initial dictionary is soon forgotten and convergence is obtained in a reasonable amount of iterations.

3.4.2 Inpainting using Sparse Representation and Dictionary Learning

Sparse representations with learned dictionaries are, among other things, capable of producing state of the art inpainting results (Mairal et al., 2008b). Let x be a $\text{length}(N)$ data vector, for example originated from an image patch (squared or from another neighbourhood), with the pixels stacked as a vector. A dictionary described by a $N \times K$ matrix: $D = [d_1 d_2 \dots d_K]$ being K the number of background patches from the image to be inpainted (only taking into account the blocks without missing pixels). Assume there exists a dictionary that represents the data well and it is computed using Dictionary Learning techniques. This optimised dictionary is composed by the L most representative atoms, being $L = 2N$ (Figure 3.6(a)).

In inpainting there are missing pixels, and the position of the missing pixels are known. Let us define \tilde{x} of length $(N - p)$ where the p missing pixels are removed. The inpainting process (Figure 3.6(b)) consists of:

1. Removing the corresponding rows of D , giving \tilde{D} of size $(N - p) \times L$.
2. Finding w as $\min_w \|\tilde{x} - \tilde{D}w\|$ s.t. $\|w\|_0 < s$.
3. Finding the reconstructed vector $x_r = Dw$ using the full learned dictionary.
4. Finally, the missing pixels in \tilde{x} can be replaced by the values in the corresponding positions of x_r , alternatively all \tilde{x} is replaced by x_r if a denoising of all pixels is needed in addition to filling in the missing pixels.

As can be observed in Figure 3.6(b), the reconstructed vector of each patch (x_r), extracted from the image to be inpainted, is computed. The inpainted patch (x_{inp}) is composed replacing the missing pixels in the original patch (\tilde{x}) by the values in the corresponding positions of the reconstructed patch (x_r). The final inpainted image is formed by the composition between the original image and the inpainted vascular tree (Figure 3.7).

Figure 3.8 shows an overview of the proposed image inpainting method applied to retinal images. In this flow chart it is possible to differentiate the two main stages detailed above: the Dictionary Learning and the image inpainting processes.

The process described above can be implemented defining an image M of the same size as the original image, masking the positions of all the missing pixels

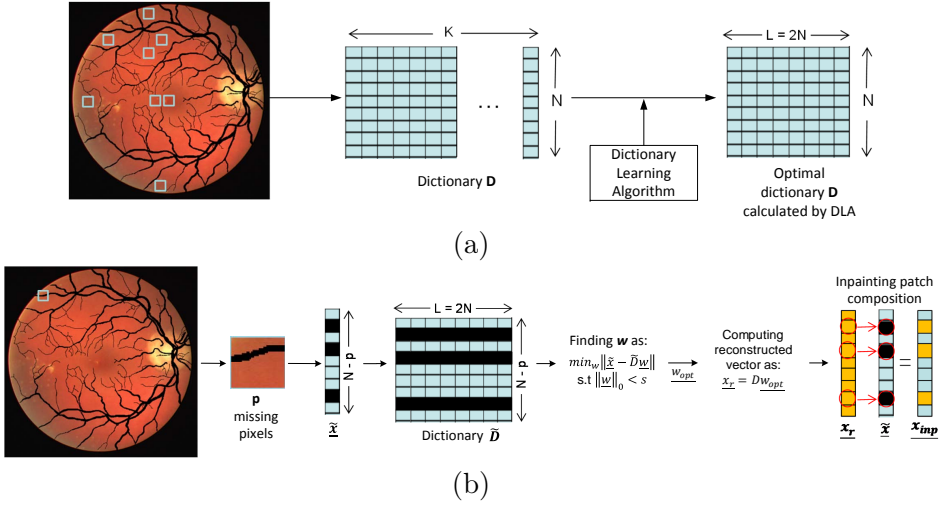


Figure 3.6: Retinal vessel inpainting using Sparse Representation and Dictionary Learning. (a) Dictionary Learning process, (b) Image inpainting process.

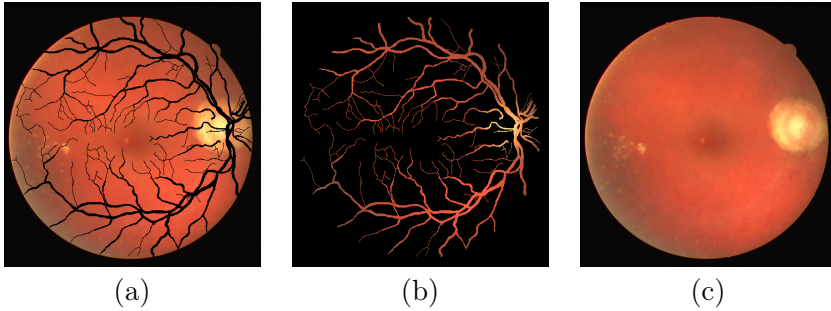


Figure 3.7: (a) Original mask with missing pixels, (b) Inpainting performed on retinal vascular tree (c) Final inpainted image.

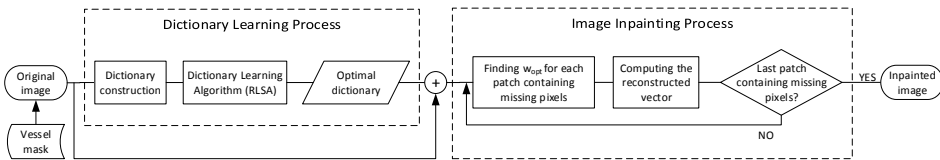


Figure 3.8: Overview of the proposed image inpainting method applied to retinal images.

and using efficient implementations. In that way, the dictionary matrix \tilde{D} does not have to be explicitly defined. The Orthogonal Matching Pursuit algorithm (with mask) (Elad, 2010) was used in these experiments, but that could easily be changed to a L1 norm algorithm, like LASSO (Tibshirani, 1996), if preferred. SPAMS library, SPArse Modeling Software (Mairal et al., 2014), includes an efficient implementation of this algorithm.

3.4.3 Measuring the performance of blood vessel inpainting

The quantification of the image inpainting offers some challenges. Blood vessel removal is a real application in which the “ideal” fundus image without blood vessels does not exist and thus it cannot be compared to the inpainted image. In most state-of-the-art works focusing on inpainting, the assessment is done by comparing the resulting image to a truth image, while doing inpainting on a corrupted version of the truth image. Two different ways to carry out the inpainting quality assessment are presented in the following and they are used to evaluate the results in each experiment of this work.

Recovery Error

A simple method to measure the image quality inpainting is modifying the vessel mask segmented by specialists. In particular, false vessel pixels (artificial mask) are added to the original vessel mask giving place to the missing vessel mask that will be used in the experiments (Figure 3.9).

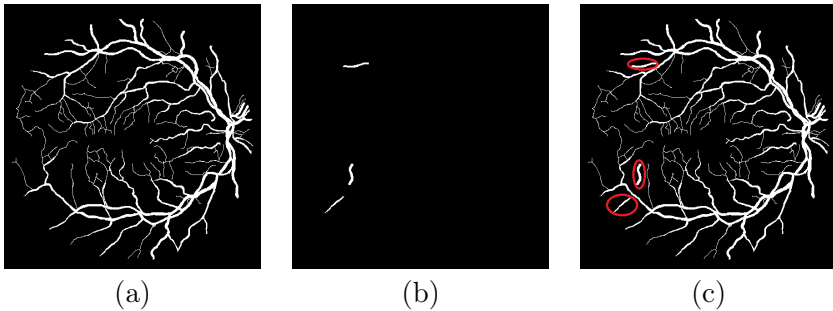


Figure 3.9: Example of mask transformation. (a) Original mask, (b) Artificial mask (c) Missing pixel mask.

Using the Recovery Error (RE) metric it is possible to quantify the quality of the inpainted images. The intensity values of the missing pixels artificially

added to the vessel mask are known in the original image, for this reason, it is possible to extract the error due to the inpainting process by computing the difference between the original pixel and the inpainted one at each additional missing pixel. The average of Recovery Error (RE) along the whole additional missing pixel mask provides a good metric of how good the image inpainting is:

$$\text{RE} = \sqrt{\frac{\sum_{n=1}^{N_c} \sum_{k=1}^M (I_{out}(x_k, y_k, n) - I_{in}(x_k, y_k, n))^2}{M}} \quad (3.19)$$

where I_{out} is the resulting image after inpainting and I_{in} is the original (truth) image, both normalised to the range of 0 to 1 and evaluated at the M pixels (x_k, y_k) of the artificial mask. N_c is the number of components (1 for grey-level images and 3 for RGB). A small RE indicates a high quality inpainting.

Inpainting Quality Evaluation Index

The Recovery Error needs the creation of an artificial mask for each image and also it only measures the quality in the relatively small areas of artificial mask. For these reasons, a novel metric is proposed in order to quantify the quality image inpainting without the use of a reference (truth) image.

We have observed that subjective quality by visual inspection is based on two factors:

- The coherence of the inpainted region and the background, in other words, the low level of correlation between the resulting inpainted region and the same region evaluated at the original image.
- The prominence/smoothness of the resulting edges of the inpainted structure to be removed. The more smooth the edges that joint the inpainted region and the background are, the better the quality of the inpainting.

We propose an automatic Inpainting Quality Evaluation (IQE) index exploiting these two factors, in order to evaluate the resulting images, defined as:

$$\text{IQE} = 1 - [G \times HF \times C] \quad (3.20)$$

where G evaluates the gradient similarity between the original and the inpainted image, HF measures the ratio of high frequencies between the images and C quantifies the image correlation. This dissimilarity index is ranged from 0 to 1 and a high IQE value corresponds to a good inpainting quality.

Human vision is very sensitive to contrast variations or texture changes in images. Image edges separate different objects or structures of the image, thus the gradient can capture well contrast and texture variations. In addition, by estimating a high-frequency parameter, it is possible to extract information about the prominence or smoothness of the contours in the junctions between the inpainted region and its surrounding background. The gradient similarity (G) and high-frequency (HF) terms focus on these ideas and both parameters are computed according to the equations detailed in (Wang et al., 2008).

With the aim of quantifying the coherence of an inpainted region and its neighbourhood, a luminance term is proposed in (Wang et al., 2008). Using this term, only the mean intensities between a particular patch from the inpainted image and the original one can be compared. With this idea, both, pixels with and without inpainting inside the window contributes in the same way to the final value of the metric for the patch. Moreover, in the global metric, both, patches with missing pixels (which suffer inpainting) and patches with no missing pixels, are taken into account modifying the global value of the metric (increasing the value, because the patches non suffering inpainting have the maximum value and contribute overvaluing the metric). For these reasons we propose a new term to quantify how coherent a new inpainted patch is with the background, or how well blood vessels have been transformed into retinal texture. The correlation level between the inpainted image and the original one is computed in each RGB channel. A sliding window, of dimensions $m \times n$, loops the vessel mask. For each window in which at least one pixel corresponds to blood vessel, the correlation term $C(x, y)$ is calculated as:

$$C(X, Y) = \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [X(i, j) - \bar{X}][Y(i, j) - \bar{Y}]}{\sqrt{(\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [X(i, j) - \bar{X}]^2)(\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [Y(i, j) - \bar{Y}]^2)}} \quad (3.21)$$

where $X(i, j)$ is the i^{th}, j^{th} pixel in the sliding window evaluated in the inpainted image and $Y(i, j)$ is the i^{th}, j^{th} pixel in the sliding window evaluated in the original one (image with blood vessels). \bar{X} and \bar{Y} are the average values of the intensity calculated in the inpainted and the original image respectively.

The correlation term $C(A,B)$ between the whole inpainted image A and the original one B is calculated as the average of $C(X,Y)$ terms for each window in which at least one pixel is blood vessel:

$$C(A,B) = \frac{1}{N_w} \sum_{k=0}^{N_w-1} C_k(X,Y) \quad (3.22)$$

where N_w is the number of windows in which at least one pixel is labelled as blood vessel. A low correlation corresponds to a high coherence between the inpainted region and its surrounding background.

3.4.4 Assessment of retinal blood vessel inpainting

Three experiments were carried out. The objective of the first experiment was to assess the influence of the sparseness factor and the neighbourhood. For that, the sparse-based method with RLS-DLA on grey-level images were used. In the second experiment, two methods of dictionary learning were used and compared with the aim of measuring the effect of inpainting each component of a RGB image (separate inpainting) versus to inpaint the three components together (jointly inpainting). Finally, the sparse-based inpainting is compared with a diffusion-based algorithm and the exemplar-based method proposed by Criminisi et al. (2004) on RGB images.

Grey-Level Experiment

In this part of the work, the green channel of a RGB image was chosen to perform the vessel inpainting because it is the component commonly used to detect the lesions (the objective of a computer-aided diagnosis system) (Abràmoff et al., 2013; Walter et al., 2002; Zhang et al., 2014).

The original image was divided in L overlapping patches of size $N = BB_{height} \times BB_{width}$ or other non-squared neighbourhoods. The initial dictionary D was composed by atoms from the training signal, in this context that means that atoms were constructed by means of normalised stacked pixels of patches (without any missing pixels) extracted from the image. The RLS-DLA (Skretting and Engan, 2010) was applied in order to optimise the dictionary D , (size $N \times K$). The new dictionary was composed by the L most relevant atoms being $L = 2N$, so the over-completed learned dictionary is a matrix $N \times L$, which contains L prototype signals of length N .

Inpainted images were obtained using the learned dictionary and the fast implementation of the OMP method (with mask) provided in the SPAMS⁴ library (Mairal et al., 2014).

Different neighbourhoods were tested: squared blocks of sizes 8×8 , 21×3 and 3×21 . The choice of the neighbourhood size is a key decision because is directly related with the size of the anatomical structure to be inpainted and, therefore, with the image resolution. To validate this work, two public databases were used and the different resolutions affect in the neighbourhood size decision. A spatial resizing was performed in the *STARE* images in order to set the same resolution as the *DRIVE* images which present the minimum dimensions. Using these image dimensions (565 x 584 pixels), the thickest blood vessel takes up around five pixels. So, the patch size (in each dimension) should be higher than the size of the biggest vessel, to avoid patches containing only missing pixels. Finally, the patch size was chosen as a trade-off between the sparsity of zeros inside the patch and the computational cost. In this work square and rectangular patches as well as sparser neighbourhoods were tested in order to compare their behaviors.

The inpainting based on sparse representation with squared neighbourhoods works well when the missing pixel areas are small (Mairal et al., 2008a). In our problem, we have continuous, and large, missing pixels areas which are not properly addressed by this kind of neighbourhoods. Our idea was to use more spread neighbourhoods trying to minimise the number of missing pixels in the same patch. So, in addition to the squared neighbourhoods, the patterns shown in Figure 3.10 were tested. These patterns contain 77, 43 and 115 elements respectively. The same neighbourhood was used for learning the dictionary as used later in the inpainting step.

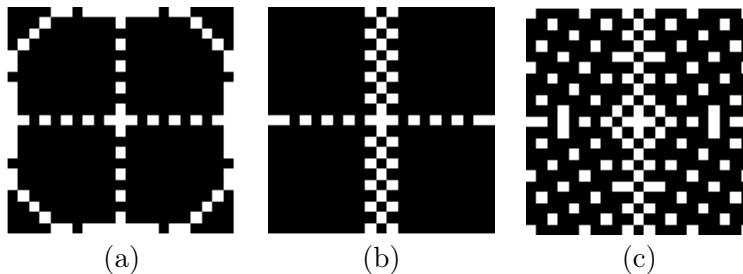


Figure 3.10: Tested neighbourhoods. (a) p1, (b) p2 (c) p3.

⁴spams-devel.gforge.inria.fr

Table 3.2 shows the results of the average recovery error (for the forty images from the *DRIVE* database) in the different tested configurations (8×8 , 21×3 , 3×21 , p1, p2 and p3 neighbourhoods). The tested number of non-zero coefficients of vector w were $s=4$ and $s=10$.

Parameters	8×8	21×3	3×21	p1	p2	p3
$s=4$	0.0386	0.0409	0.0431	0.0286	0.0275	0.0291
$s=10$	0.0390	0.0414	0.0432	0.0283	0.0272	0.0290

Table 3.2: Mean Recovery Error calculated from the resulting images of grey-scale inpainting using RLS-DLA for different values of sparseness factors ($s=4$ and $s=10$) and different neighbourhoods.

As can be observed in Table 3.2, results are slightly worse in the case of squared neighbourhoods. However, in relation to the non-zero coefficients (sparseness factor) the performance is similar using $s = 4$ and $s = 10$.

Figure 3.11 shows an original image and the inpainting result for the best case (neighbourhood p2).

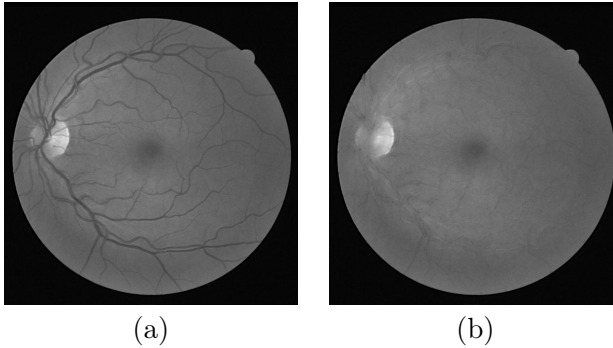


Figure 3.11: Grey-scale inpainting for the *01_test.tif* image from *DRIVE* public database. (a) Original image and (c) Inpainted image. This result was obtained using RLS-DLA with the sparseness factor $s = 10$ and the second spread neighbourhood.

RGB Experiment

This experiment aims to analyse which configuration provides better results: to inpaint each colour component separately or the jointly inpainting of the three components (Mairal et al., 2008b). Another objective in this experiment is to demonstrate if the choice of the dictionary learning algorithm (in order to create the optimal dictionary) influences in the quality of the resulting inpainted image. *DRIVE* database and recovery error were used in this experiment.

- RGB-separate Inpainting

Each component of RGB images were extracted and inpainted following the same method as presented in the grey-level experiment. The resulting RGB image is the composition of the three inpainted images, one for each channel. In this experiment the performance of the inpainting employing the RLS-DLA was compared with the performance of the result using Online Dictionary Learning (ODL) algorithm (Mairal et al., 2009). Table 3.3 shows the results of the mean recovery error (for the whole *DRIVE* dataset) measured on the composed RGB image, for RLS-DL and ODL algorithms.

Parameters		8x8	21x3	3x21	p1	p2	p3
RLS-DL	s=4	0.0578	0.0549	0.0651	0.0496	0.0506	0.0463
	s=10	0.0575	0.0553	0.0645	0.0489	0.0503	0.0453
ODL	s=4	0.0559	0.0577	0.0662	0.0474	0.0518	0.0471
	s=10	0.0567	0.0581	0.0641	0.0473	0.0512	0.0464

Table 3.3: Mean Recovery Error calculated from the resulting images of RGB-separate inpainting using RLS-DL and ODL algorithms for different values of sparseness factors (s=4 and s=10) and different neighbourhoods.

- RGB-jointly Inpainting

In this method an atom of the dictionary is composed by the same patch in each channel, so the learned dictionary D' is a matrix $N' \times L'$ where $N' = 3 N$ and $L' = 2 N'$. In this case the number of non-zero coefficients of vector w' should be tripled, compared to the grey-level experiment, because its length is three times the length of w . The inpainted RGB image is obtained in one step and the recovery error is measured on this image directly. Table 3.4 shows the

results of the mean recovery error for the RGB image using RLS-DL and ODL algorithms.

Parameters		8x8	21x3	3x21	p1	p2	p3
RLS-DL	s=12	0.0377	0.0396	0.0496	0.0225	0.0253	0.0283
	s=30	0.0382	0.0396	0.0499	0.0213	0.0249	0.0281
ODL	s=12	0.0382	0.0396	0.0470	0.0249	0.0274	0.0272
	s=30	0.0387	0.0384	0.0486	0.0242	0.0270	0.0280

Table 3.4: Mean Recovery Error calculated from the resulting images of RGB-jointly inpainting using RLS-DL and ODL algorithms for different values of sparseness factors (s=12 and s=30) and different neighbourhoods.

Comparing Table 3.3 and Table 3.4 an improvement for all neighbourhoods is observed inpainting the three channels in one step, rather than individually. Thus, it is beneficial using one dictionary in which each atom is composed for a patch extracted from all channels. This results evidence that the correlation of the three channels of the image provide more information than the three individual channels.

Figure 3.12 displays the results of two inpainted images using the proposed method (with the best result, i.e with p1 and RLS-DLA). Figure 3.12(a) shows an image that includes exudates and Figure 3.12(d) shows an image with microaneurysms and haemorrhages. As can be observed from the output images, (Figure 3.12(b) and Figure 3.12(e)), the blood vessels are significantly suppressed without any visual degrading of the exudates (Figure 3.12(c) or microaneurysms (Figure 3.12(f)).

Comparison of Inpainting Methods

With the aim of characterising the best inpainting category for our application, a comparison between diffusion-based inpainting, the inpainting method used in the retina by other authors, the exemplar-based method proposed by Criminisi et al. (2004), and the sparse-based inpainting, proposed in this paper, was carried out and the results are presented in this section.

Besides *DRIVE* database, in order to provide a wider comparison, the fundus images from the public database *STARE*, introduced in Section 3.2, were added in this experiment. In order to avoid the task of the artificial mask extraction for each image belonging to *STARE* database, the proposed IQE index was used to quantify the quality of the blood vessel inpainting.

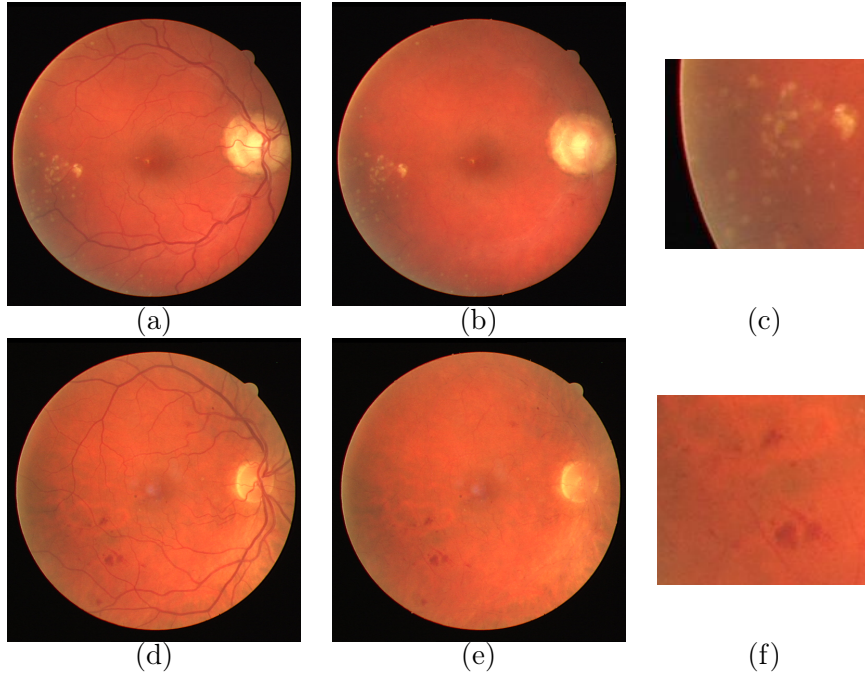


Figure 3.12: RGB-jointly inpainting for two images. Original images (a) *08_test.tif* and (d) *14_test.tif* from *DRIVE* public database. Inpainted images (b) *08_test.tif* (RE = 0.207) and (e) *14_test.tif* (RE = 0.211). The inpainting was performed using RLS-DLA with the sparseness factor $s = 30$ and the first spread neighbourhood. As can be observed, blood vessels are significantly suppressed without any visual degrading of the exudates (c) or microaneurysms (f).

Diffusion-based inpainting category by means of the simplest isotropic diffusion model⁵ (D’Errico, 2004) and a public implementation⁶ of the exemplar-based inpainting method presented by Criminisi et al. (2004) were selected to perform the comparison with the proposed sparse-based inpainting method (Table 3.5). The same neighbourhoods used in the previous experiments were tested for the sparse-based method.

In order to validate the IQE index, the RGB-jointly inpainting and the RGB-separate inpainting were computed. Note that in relation to the non-zero coefficients, $s = 10$ was used for RGB-separate (RGS) inpainting and $s = 30$

⁵<https://es.mathworks.com/matlabcentral/fileexchange/4551-inpaint-nans>

⁶www.github.com/ikuwow/inpainting_criminisi2004

for the RGB-jointly (RGBJ) inpainting. In addition, a Mean Opinion Score (MOS) evaluation of the inpainting results was performed by five experts in digital image processing (Table 3.6). These experts assessed the 60 resulting images according to the following quality scale: Very poor(0), Poor(1), Fair(2), Good(3), and Excellent(4). In the evaluation process, the resulting images were presented in random order to the experts and without any kind of information about the inpainting method used to remove the vessels.

Parameters		8x8	21x3	3x21	p1	p2	p3	DBI	Crim.
DRIVE	RGBS	0.5410	0.5276	0.5189	0.6044	0.5970	0.6170	0.5039	0.6641
	RGBJ	0.5596	0.5461	0.5350	0.6838	0.6194	0.6915		
STARE	RGBS	0.6224	0.6013	0.5897	0.6539	0.6562	0.6524	0.5748	0.6691
	RGBJ	0.6152	0.6094	0.5979	0.6648	0.6572	0.6582		

Table 3.5: IQE index average for the RGB-jointly inpainting ($s = 30$) and the RGB-separate inpainting ($s = 10$) computed from *DRIVE* and *STARE* databases using diffusion-based inpainting (DBI), Criminisi et al. (2004) method and different neighbourhood configurations of sparse-based Inpainting.

Parameters		8x8	21x3	3x21	p1	p2	p3	DBI	Criminisi
DRIVE	RGBS	0.300	0.050	0.075	3.350	1.575	3.625	1.625	2.750
	RGBJ	0.650	0.180	0.220	3.475	2.275	3.700		
STARE	RGBS	0.388	0.117	0.210	2.950	2.100	2.650	1.500	3.100
	RGBJ	0.750	0.330	0.412	2.950	2.200	2.650		

Table 3.6: Mean Opinion Score (MOS) evaluation of the inpainting results obtained from the *DRIVE* and *STARE* databases. The assessment was carried out using the following scale: Very poor(0), Poor(1), Fair(2), Good(3), and Excellent(4).

The novel IQE index allows to quantify the performance of the inpainting result taking into account the result in the whole image, not only in a few pixels artificially added composing an artificial mask. For this reason, comparing the results for *DRIVE* database (Table 3.5) with the results presented in Tables 3.3 and 3.4 different results are observed. In Table 3.4 the best results are obtained using the pattern $p1$, but assessing the inpainting quality by the IQE index the best pattern is found as $p3$ in Table 3.5. We believe that the IQE index provides us a more reliable measure of the quality of the inpainting compared to the Recovery Error because RE only evaluates the quality in the pixels of the artificial mask.

As can be observed in Table 3.5, the IQE index ratifies that the RGB-jointly inpainting of the three channels provides better results than the RGB-

separate inpainting. The results of the MOS evaluation (Table 3.6) shows high correlation with the IQE results, demonstrating the reliability of the proposed metric. Diffusion-based techniques perform well filling small gaps but the results are quite poor when the unknown region to be inpainted is large and blood vessels are a large structure that covers a high percentage of the fundus image. This fact could be the consequence of the humble results observed in Table 3.5 when diffusion-based inpainting is used.

The method proposed by Criminisi et al. (2004) is based on the idea of finding the image patch that maximises the correlation with the region to be inpainted. Table 3.5 shows similar results using this method in comparison with the best configuration of the proposed sparse-based method. Figure 3.13 allows a visual comparison of the resulting images using the inpainting methods compared in this work.

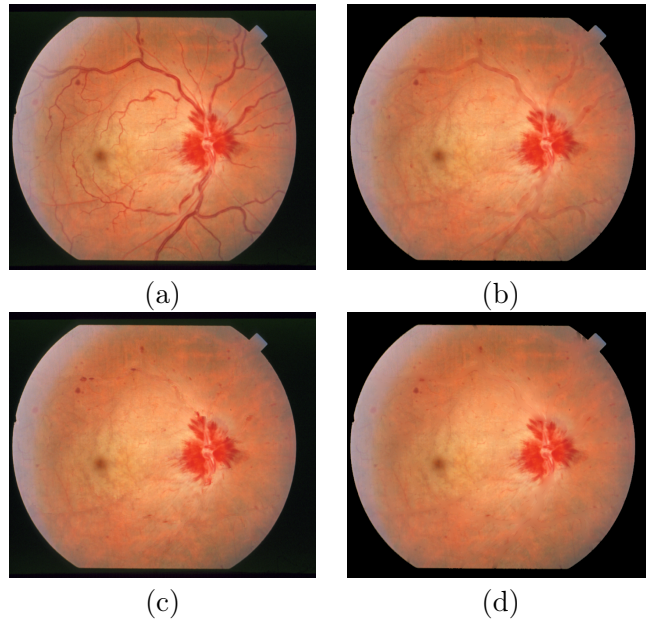


Figure 3.13: Blood vessels inpainting comparison on *im0005.ppm* from *STARE* public database. (a) Original, (b) inpainted image by diffusion-based method (IQE = 0.5322), (c) inpainted image by Criminisi et al. (2004) method (IQE = 0.6631) and (d) RGB-jointly inpainted ($s = 30$) image by sparse-based method using the first pattern (IQE = 0.6886).

Despite the good performance using the Criminisi et al. (2004) method, the public implementation used in this work is really time-consuming. In order to check this fact, Table 3.7 shows the average time taken by the different algorithms compared in this work when blood vessel removal is performed in *DRIVE* and *STARE* databases. Note that the experiment was carried out using a Intel Core i7-860@2.80GHz with 16GB of RAM and MATLAB[®]R2015a as a development environment.

	Sparse-based	Diffusion-based	Criminisi et al. (2004)
DRIVE	7.42 sec.	2.29 sec.	15.23 sec.
STARE	15.18 sec.	4.91 sec.	32.38 sec.

Table 3.7: Averaged time (in sec.) per image consumed for the Sparse-based method, Diffusion-based method and Criminisi et al. (2004) method to perform the blood vessel removal in *DRIVE* and *STARE* databases.

Chapter 4

Extraction of fundus visual information for pathological tissue identification

This chapter presents novel methods to locally describe fundus images. Texture, morphological and roughness information is extracted from the retina background. Moreover, the basis of advanced classification techniques are described because these algorithms will be used to discriminate between healthy and pathological retinal tissue. The performance of these methods applied to the topic of this thesis will be shown in Chapter 6.

In this chapter, works previously published, although adapted for this manuscript, are included:

- *Section 4.1.3 - Normalisation of public fundus databases.*

Colomer, A., Naranjo, V., Janvier, T. (2016). Retinal Diseases Characterisation using Fractal Analysis. In *18th Edition of the Mathematical Modelling in Engineering & Human Behaviour (MMEHB 2016)*, pp,1-4.

Colomer, A., Naranjo, V., Janvier, T., Mossi, J.M. (2018). Evaluation of Fractal Dimension effectiveness for damage detection in retinal

background. In *Journal of Computational and Applied Mathematics* (Article In Press).

In addition, the information contained in Section 4.1.4 of this chapter and its corresponding results (see Section 6) was part of the work developed at the *Centre of Mathematical morphology (CMM)* belonging to the *École de Mines ParisTech* under the Grant EEBB-2016-11276.

Contents

4.1	Description of retina background lesions	55
4.1.1	Introduction	55
4.1.2	Local Binary Patterns Variance	57
4.1.3	Fractal analysis	64
4.1.4	Granulometry profile	71
4.2	Machine Learning methods for retinal tissue classification	81
4.2.1	Classification strategy	81
4.2.2	Random Forest	82
4.2.3	Support Vector Machine	84
4.2.4	Gaussian processes for classification	89

4.1 Description of retina background lesions

4.1.1 Introduction

As it was mentioned in Chapter 1, the main objective of this thesis is to develop novel algorithms that allow the automatic classification of retinal tissue into healthy and pathological (according to the different lesions presented in Section 2.3). In the literature, the most common procedure is to segment these lesions using different methods (Ghafourian and Pourreza, 2012; Walter et al., 2002; Welfer et al., 2010). These approaches present a high false-positive rate at pixel level. This fact is the main motivation of the new perspective proposed in this work, in which, the characterisation of healthy and damaged retinal areas is studied by applying image descriptors in a local way, avoiding the segmentation step. Our approach is different to previous works based on feature extraction and classification. In (Mookiah et al., 2013) a previous segmentation of exudates is required to perform the feature extraction. The most habitual procedure is to extract features from a lesion candidate map generated by different techniques such as mathematical morphology (Sopharak et al., 2008; Walter et al., 2002; Zhang et al., 2014), background subtraction (Giancardo et al., 2012), clustering (Amel et al., 2012) or using banks of filters and applying a low adaptive threshold (Akram et al., 2014a,b) among others. To the best of the author's knowledge, the work proposed by Quéllec et al. (2012) is the unique previous attempt which has some resemblance with the strategy proposed in this thesis. In that work, a local analysis without overlapping based on wavelet features is proposed to automatically detect diabetic retinopathy. Areas under the receiver operating curve of 0.881 and 0.761 were obtained in messidor and e-optha public databases, respectively. In this thesis, robust descriptors extract texture, shape and roughness features from the visual information of retinal images. This methodology does not require the previous segmentation of retinal lesions or the generation of candidate maps, avoiding error sources due to the segmentation process and computational cost.

The lesions induced by diabetic retinopathy or macular degeneration present different sizes according to the stage of the disease (see Section 2.3). In most cases, lesions represent less than one percent of the total number of pixels that compose the retinal image. For this reason, in the feature extraction stage of the proposed methodology, descriptors are locally performed, in other words, the image is divided in patches using a sliding window (see Figure 4.1(a)) and image descriptors are computed for each patch. This procedure can be related to the way in which the human inspects and analyses an image. From an image-processing point of view, the sliding window strategy is the same as a

2D grid or a dense sampling of the image, in which the samples correspond to the central pixel of the window.

In the experiments carried out during this thesis, the window used to process the retinal tissue is a square of canonical size (N_w) with overlap of $(\Delta x, \Delta y)$. Figure 4.1 represents the process detailed above and its equivalence with a dense image sampling.

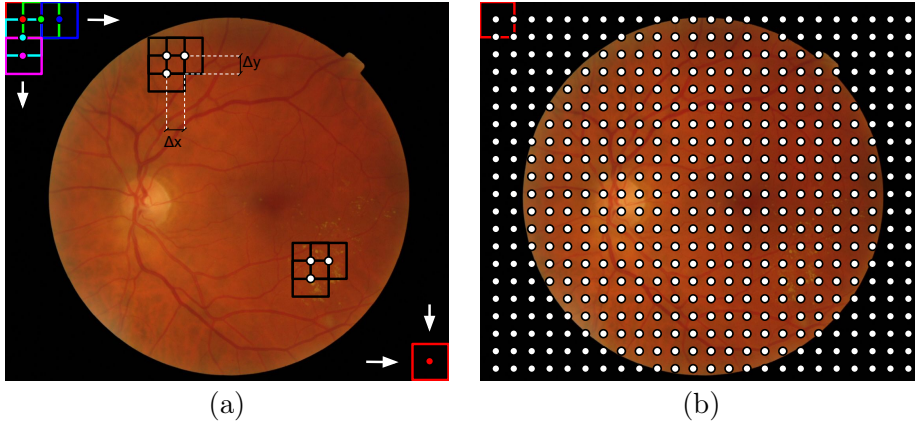


Figure 4.1: Illustration of the local analysis in fundus images: (a) Visual representation of how the sliding window of dimensions $N_w \times N_w$ loops a fundus image and (b) the resulting image grid composed by the centers of the sliding window in each position. This process is equivalent to a dense sampling. For the representation $N_w = 64$ and $(\Delta x, \Delta y) = (32, 32)$ were used.

It is important to note that patches should contain at least one pixel belonging to retinal texture in order to be processed, in other words, patches contained entirely outside the lens are discarded. In addition, patches containing optic disk pixels, obtained by the method proposed in (Morales et al., 2013), are not considered in the process (Figure 4.2).

As it was mentioned above, lesions related to DR and AMD represent less than one percent of the total number of pixels that compose the retinal image. This fact can be observed in Figure 4.1(a), in which the exudates are manifested in a few pixels in the retina background. As the reader may suppose, the patch extraction process will result in an imbalanced dataset. This issue is not taken into account in the image description step, in which a feature vector is computed for each patch of the image. However, it is important to remark this fact because in the classification stage, healthy and pathological datasets

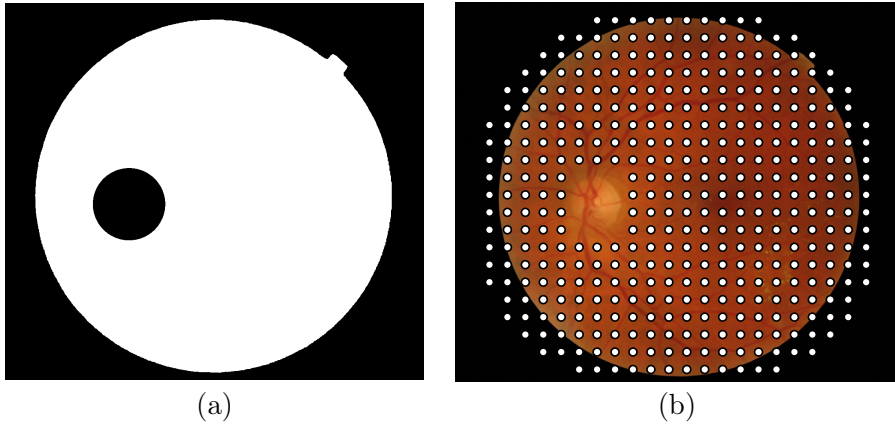


Figure 4.2: (a) Binary mask used to exclude the patches containing optic disk pixels and the patches located out of the field of view and (b) centre of the patches composing the final grid in which image descriptors will be applied.

must be composed by a comparable number of instances. This issue will be discussed deeper in Section 4.2.1 related to the classification process.

Throughout this section, the theoretical background of the proposed image descriptors will be detailed and the formulation necessary to apply them in the extraction of relevant information from the healthy and pathological retinal tissue will be presented.

4.1.2 Local Binary Patterns Variance

Introduction

Local Binary Pattern (LBP) is a powerful grey-scale texture operator introduced by Ojala et al. (2002) that had been given a lot of attention in recent years. It is based on looking at the local variations around each pixel, and assigning labels to different local patterns. Thereafter, the distribution of the labels is evaluated and used as a texture descriptor for classification purposes. The LBP technique is a theoretically and computationally simple approach which is robust in terms of grey-scale variations and which is shown to discriminate a large range of rotated textures efficiently. Taking into account the joint distribution of grey values of a circularly symmetric neighbour set of pixels in a local neighbourhood, the LBP operator is, by definition, invariant against any monotonic transformation of the grey-scale. Rotation invariance

is achieved by recognising that this grey-scale invariant operator incorporates a fixed set of rotation invariant patterns (Ojala et al., 2002).

LBP has been used to successfully describe and classify textures in general (Ahonen et al., 2006; Heikkilä et al., 2009; Liu et al., 2016; Pan et al., 2017; Zhao and Pietikainen, 2007) and also in the field of medical imaging (Kotu et al., 2012; Nanni et al., 2010; Naresh and Kumari, 2015; Nawarathna et al., 2014; Sørensen et al., 2008). In the particular case of fundus images, LBP technique has mostly been used on the automatic retinal vessel detection (Hatami and Goldbaum, 2016; Moin et al., 2010; Zabihi et al., 2010) rather than on the identification and localisation of abnormal retinal tissue. However, some state-of-the-art works focused on the automatic detection of DR and AMD can be found (Garnier et al., 2014; Mookiah et al., 2013; Morales et al., 2017). Garnier et al. (2014) extract the texture information at several scales through wavelet decomposition and LBP with the aim of detecting AMD. Mookiah et al. (2013) extract thirteen features such as the area of exudates, the area of blood vessels, bifurcation points, textures, entropies, etc. in order to detect normal fundus and two DR stages. The texture features were extracted by means of LBP and Laws energy.

The work of Morales et al. (2017) (in which A. Colomer significantly contributed) was the starting point of this thesis chapter. In that work, the goal was to distinguish between DR, AMD and normal fundus images avoiding any previous segmentation stage of retinal lesions. The texture of the retina background was directly analysed by means of different global statistics extracted from the LBP image. Promising results in the field of computer-aid diagnosis systems for retinal disease screening were achieved. However, the exclusion criteria in the database creation were quite restrictive. Images with any kind of abnormality of the explained in Section 3.3 were not finally included in the database due to the high variability introduced, which propitiates a large number of false positive predictions by the system. This fact may be produced because features were extracted globally, in other words, different statistics were computed from the whole LBP image. As it was reported in Section 2.3, the early lesions induced by DR and AMD are manifested in a small amount of the pixels of the image so, extracting information of the global texture can mislead the ability of discrimination of the LBP.

In this section a novel approach to extract the texture information of fundus images by means of LBP is presented. The feature extraction is locally performed avoiding the exclusion of fundus images with artefacts, noise, brights or tessellations to the database. A robust descriptor based on LBP combined with a contrast measure is proposed as unique step to extract the

texture information from the original fundus images. As it was mentioned in Section 4.1.1, this methodology does not require the previous segmentation of retinal lesions or the generation of candidate maps avoiding error sources and computational cost.

Theoretical framework

The basic Local Binary Pattern version consists in assigning a label to each pixel (i, j) of the image taking into account its local neighbourhood as follows:

$$LBP_{P,R}(i, j) = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p, \quad s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (4.1)$$

where P represents the number of samples on the symmetric circular neighbourhood of radius R , g_c is the grey value of pixel (i, j) and g_p the grey value of each neighbour. If the coordinates of the P neighbours defined by the circle of radius R do not fall in the centre of the pixels, the g_p values are estimated by interpolation. In Figure 4.3 two neighbourhoods with $P = 8$ and different radii are shown.

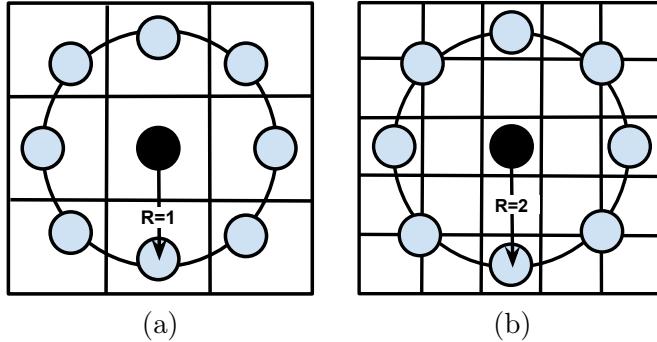


Figure 4.3: Two neighbourhoods of $P = 8$ and different radii for LBP computation. (a) $R = 1$ and $P = 8$ and (b) $R = 2$ and $P = 8$.

As can be deduced from Equation (4.1), the neighbouring pixels are thresholded with respect to the grey value of the central pixel (i, j) denoted by g_c giving place to a binary string or, in other words, a binary pattern. The final $LBP_{P,R}$ label is obtained by converting the binary string into a decimal value. $LBP_{P,R}$ texture operator describes the occurrence of specific patterns in the neighbourhood of each pixel in a P -dimensional histogram.

For example, for constant or uniform regions $LBP_{P,R}$ will be equal to one in all directions. In contrast to as a spot, the texture operator will register high values along all directions. An edge will be described by high $LBP_{P,R}$ values in the gradient direction and zero values along the edge. This basic version of LBP is invariant against monotonic transformation of the grey scale by definition, in other words, if the order of the grey values in the image is not modified, $LBP_{P,R}$ operator will remain constant. However, this version of the operator is sensitive to rotation variations. To solve this problem, the minimum value of all possible pattern rotations is chosen, achieving a rotation invariant pattern.

Using the $LBP_{P,R}$ operator, 2^P different binary patterns can be formed by the P pixels in the neighbour set. However, several binary strings refer to the same pattern under rotation. For example, the binary patterns 11000011 and 1110001 are different rotations of the normalised string 00001111. With the objective of removing this effect, a unique identifier to each rotation invariant local binary pattern is proposed by means of $LBP_{P,R}^{ri}$ defined as follows:

$$LBP_{P,R}^{ri} = \min\{ROR(LBP_{P,R}, t) \mid t = 0, 1, \dots, P - 1\} \quad (4.2)$$

where $ROR(x, t)$ performs t times a circular bit-wise right shift on the P -bit number x .

In the case of $P = 8$, 36 unique rotation invariant local binary patterns can occur and they will be defined by $LBP_{8,R}^{ri}$. However, the ability of discrimination of this operator is not optimal as it was demonstrated in (Pietikäinen et al., 2000). Pietikäinen et al. observed that certain local binary patterns have fundamental properties of texture, describing over 90% of the patterns present in the textures of their experiments. This fact suggested that some patterns provide redundant information and hamper in the discrimination ability of the descriptor. For this reason, the definition of ‘uniform’ patterns was introduced in (Ojala et al., 2002). By means of the uniformity measure $U(LBP_{P,R})$ the number of spatial transitions (bitwise 0/1 changes) of the pattern are achieved. Table 4.1 shows different binary patterns and their corresponding uniformity value. Ojala et al. demonstrated that the patterns with a U value of at most 2 were the most discriminative, and they defined them as ‘uniform’ LBP, proposing the following operator for grey-scale and rotation invariant texture description:

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(LBP_{P,R}) \leq 2 \\ P + 1 & \text{otherwise} \end{cases} \quad (4.3)$$

where

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{P-1} - g_c)| \quad (4.4)$$

LBP patterns	Uniformity
00000000	0
01111111	2
00000011	2
01000111	4
01010101	8

Table 4.1: Uniformity of Local Binary Patterns ($LBP_{8,R}^{ri}$).

In a circular symmetric neighbour set of P pixels, the exact number of ‘uniform’ patterns will be $P+1$, and each of them is labelled in the range of 0 to P (see Figure 4.4). The rest of the patterns are grouped in a ‘non-uniform’ set labelled as $P+1$. $LBP_{P,R}^{riu2}$ has $P + 2$ distinct output values. In the case of $P = 8$, ten different patterns can be described by this advanced operator. As Ojala et al. (2002) demonstrated, the ‘uniform version’ of LBP provides better discrimination in comparison with the previous definitions and the reason comes down to the differences in their statistical properties. In this thesis rotation-invariant uniform LBP ($LBP_{P,R}^{riu2}$) is used to encode the different patterns that can occur in the retina background.

When LBP is used for texture description it is common to additionally include a contrast measure by defining the Rotational Invariant Local Variance (VAR) as:

$$VAR_{P,R}(i, j) = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - \mu)^2, \quad \mu = \frac{1}{P} \sum_{p=0}^{P-1} g_p \quad (4.5)$$

$VAR_{P,R}$ is, by definition, invariant against shifts in grey scale. In addition, $LBP_{P,R}^{riu2}$ and $VAR_{P,R}$ are complementary measures and the combination of

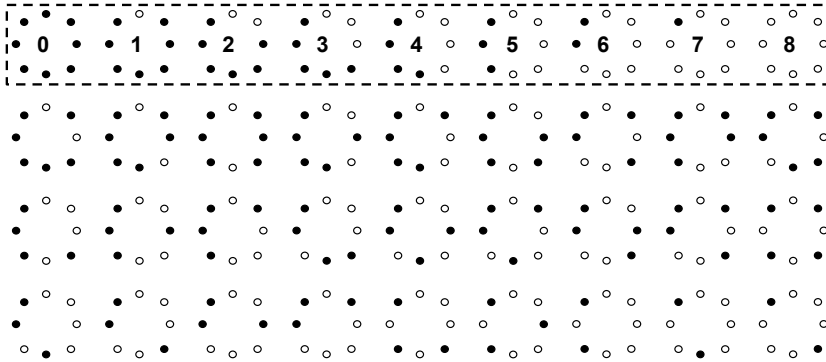


Figure 4.4: The 36 unique rotation invariant binary patterns that can occur in the circularly symmetric neighbour set of $LBP_{8,R}^{ri}$. Black and white circles correspond to bit values of 0 and 1, respectively, in the 8-bit output of the operator. The dotted line delimit the nine “uniform” patterns and the numbers inside them correspond to their unique $LBP_{8,R}^{riu2}$.

both is expected to be a very powerful descriptor for detecting abnormal retinal patterns.

LBP for retinal damage detection

Lesions due to diabetic retinopathy vary in size depending on the stage of the disease. In most cases, lesions represent less than one percent of the total number of pixels in the retinal fundus image. For this reason, the texture descriptor is applied locally to obtain relevant features for a large variety of images.

Using the public implementation¹ of the $LBP_{P,R}^{riu2}$ and $VAR_{P,R}$ operators, they are locally computed for each pixel of the green channel of the retinal images. In this channel, the contrast between the lesions and background is maximised (Abràmoff et al., 2013; Walter et al., 2002; Zhang et al., 2014). As a result LBP and VAR images, both of dimensions $M_1 \times M_2$, are obtained. These resulting images are divided into patches by means of a sliding window and normalised histograms are computed for each patch combining the information provided by both images (Figure 4.5). LBP variance (LBPV) histogram (Guo et al., 2010) accumulates the $VAR_{P,R}$ value for each $LBP_{P,R}^{riu2}$ label inside the window according to the following equation:

¹[www.ee.oulu.fi/~\thicksim\\$gyzhao/LBP_Book.htm](http://www.ee.oulu.fi/~\thicksim$gyzhao/LBP_Book.htm)

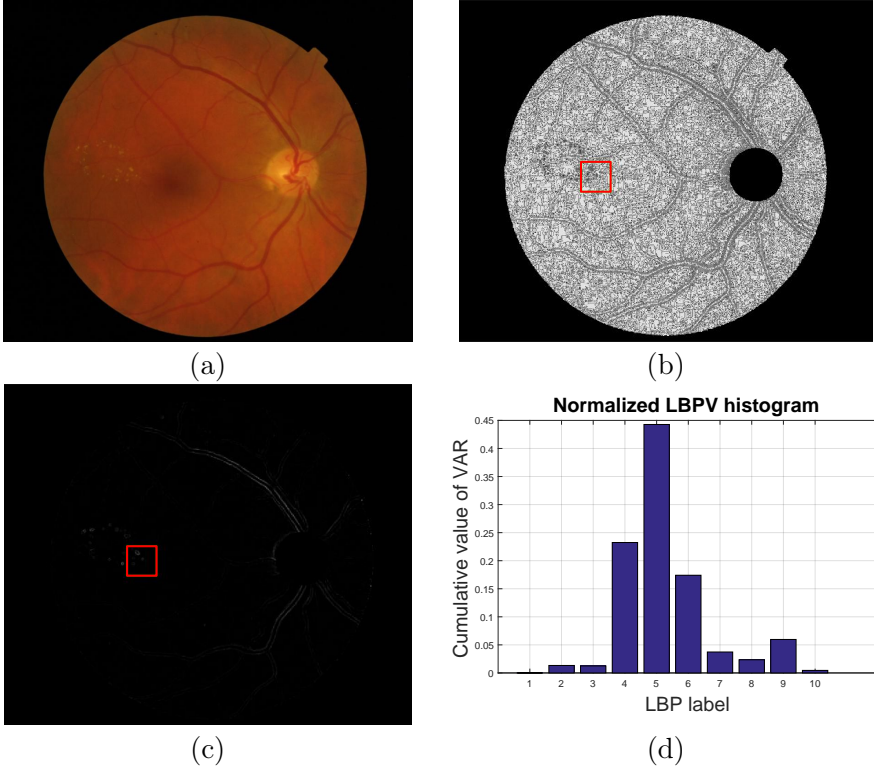


Figure 4.5: Local texture analysis. (a) Original fundus image, (b) LBP image and (c) VAR image with a patch highlighted in red and (d) the LBPV normalised histogram computed for the yellow patch.

$$LBPV_{P,R}(k) = \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} w(LBP_{P,R}(i,j), k), \quad k \in [0, K] \quad (4.6)$$

$$w(LBP_{P,R}(i,j), k) = \begin{cases} VAR_{P,R}(i,j), & LBP_{P,R}(i,j) = k \\ 0, & otherwise. \end{cases} \quad (4.7)$$

where K is the maximal LBP label.

Algorithm 4.1 describes the procedure to locally extract texture information from RGB retinal images.

Algorithm 4.1: Local texture descriptor computation

Data: RGB fundus image $\mathbf{f} = (f_r, f_g, f_b)$ (can be the original retinal image or a pre-processed version as it will be detailed in Section 6.2.1), optic disc mask \mathbf{M}_{od} , external mask \mathbf{M}_{ext} , size of the sliding window \mathbf{N}_w and vectors \mathbf{r} and \mathbf{c} containing the y and x coordinates of the centres of the patches in which the local analysis is performed (see Section 4.1.1).

Result: Feature matrix \mathbf{T} containing one local texture descriptor per row.

Initialization:

P and R neighbourhood parameters

$N_{bins} \leftarrow P + 2;$

$ind \leftarrow 1;$

Green channel selection:

$f_{in} \leftarrow f_g;$

LBP and VAR image extraction:

$f_{LBP} \leftarrow LBP_{P,R}^{riu^2}(f_{in});$

$f_{VAR} \leftarrow VAR_{P,R}(f_{in});$

Mask computing:

$M \leftarrow and(M_{od}, M_{ext})$

Local texture descriptor:

```

for  $i \leftarrow 1$  to  $length(\mathbf{r})$  do
  for  $j \leftarrow 1$  to  $length(\mathbf{c})$  do
     $M^\diamond \leftarrow M^{N_w}(r_i, c_j);$ 
     $f_{LBP}^\diamond \leftarrow f_{LBP}^{N_w}(r_i, c_j);$ 
     $f_{VAR}^\diamond \leftarrow f_{VAR}^{N_w}(r_i, c_j);$ 
    for  $l \leftarrow 1$  to  $length(\mathbf{f}_{LBP}^\diamond)$  do
      if  $M^\diamond(l) > 0$  then
         $h(f_{LBP}^\diamond(l)) \leftarrow h(f_{LBP}^\diamond(l)) + h(f_{VAR}^\diamond(l));$ 
         $ind \leftarrow ind + 1;$ 
      end
     $T_{ind} \leftarrow h/sum(h);$ 
  end
end
end

```

4.1.3 Fractal analysis

Introduction

Fractal objects are characterised by a high degree of complexity but also by self-similarity represented by a repetition of patterns or statistical properties

over different scales. Based on this property of self-similarity it is possible to describe an image by means of a fractal dimension (FD) parameter. The characterisation of grey-level images is described using the potential usefulness of the fractional Brownian motion (fBm) model (Harrar et al., 2013).

Fractals have been recognised as an effective descriptor of complex structures in biology and medicine (Deering and West, 1992). In the literature, fractal analysis has been proposed for characterising biomedical signals and images (Bruno et al., 2008; Dougherty and Henebry, 2001; Jennane et al., 2001; Landini, 2011; Phothisonothai and Nakagawa, 2008). Regarding to retinal fundus images, fractals have only been applied to the retinal blood vessels characterisation. In (Mudigonda et al., 2015), a fractal analysis approach based on the box-counting method was used to quantify vascular patterns in normal and abnormal cases showing neovascularisation (caused by proliferative diabetic retinopathy in advanced stages of the disease). The algorithm was validated in ten images. In (MacGillivray et al., 2007), a blood vessel segmentation method based on fractal analysis was proposed and a comparison of the fractal dimension (extracted from the blood vessels) between hypertension and non-hypertension groups was carried out without significant evidences of discrimination ability. Azemin et al. (2011) proposed a semi-automatic method to compute the fractal dimension of fundus images. They established a comparison between their method and the box-counting method. Finally, the authors applied the proposed algorithm with the aim of testing the impact in the variation of blood vessels due to the angle between the acquisition camera and the optical axis. To the best of the author's knowledge, fractal analysis has not been used previously in the characterisation of retinal lesions.

This section focuses on demonstrating that the fBm model suits well for the characterisation of pathological texture in retinal images and as a novelty we demonstrate how it is possible to describe retinal lesions by means of the fractal dimension.

Theoretical framework

The minimum number of independent variables to describe an object is defined by the Euclidean dimension (E). A point is a 0-dimensional object; a line is 1-dimensional while a plane is 2-dimensional. However, this approach does not extract information about the "roughness" of the object, in other words, Euclidean dimension of a line will always remain constant (E=1) whatever it is straight or crooked.

Two centuries ago different mathematicians such as Koch, Sierpinski and Hausdorff established a geometric definition of fractals demonstrating that shapes and objects have fractional dimension. Fractals objects are built based on a pattern repeated at different scales and can therefore be described as a part of themselves. This property illustrated in Figure 4.6 is known as “self-similarity”.

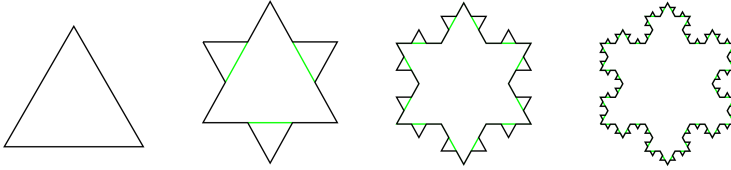


Figure 4.6: The first four iterations of the Koch snowflake.

To characterise such objects, a non-integer dimension is used: the fractal dimension (FD). However, this term does not reflect a single descriptor, as several definitions coexist. The simplest way to apprehend the FD is using the homothety a.k.a. similarity dimension (D_H):

$$D_H = \frac{\ln(N)}{\ln(r^{-1})} \tag{4.8}$$

where N is the number of elementary patterns included in the object and r is their reduction factor. This computation is valid only for a geometrical set with internal homotheties, and can be extended to any set using the box-counting, a.k.a. Minkowski-Bouligand dimension (D_B):

$$D_B = \lim_{\epsilon \rightarrow 0} \frac{\log(N_\epsilon)}{\log(\epsilon^{-1})} \tag{4.9}$$

where N_ϵ is the minimal number of subsets with a size ϵ required to cover the object. This FD reflects the complexity of the object and its construction. In particular, a signal with higher FD is more complex or rough than a signal with lower FD.

Another popular way to define the fractal dimension concept is the so-called *capacity* (D_C). This definition makes reference to the characterisation of fractal patterns or sets by quantifying their complexity as a ratio of the change in detail

to the change in scale. The relation between the length of an object and its roughness is given by:

$$L_\delta = K \cdot \delta^{(1-D_C)} \quad (4.10)$$

where L_δ is the length of the object, δ the measuring scale, K a constant and D_C the object fractal dimension (FD).

Fractional Brownian motion model

In the work presented by Mandelbrot and Van Ness (Mandelbrot and Van Ness, 1968), fractional Brownian motion is defined by its stochastic representation based on the long-range dependence and self-similar behaviour. This representation is governed by a single parameter called the Hurst exponent (H) (Hurst et al., 1965) which is linked to the fractal dimension by $H = E + 1 - FD$, where E is the Euclidean dimension.

The spectral representation of fBm, $B_H(t)$, is given by (Reed et al., 1995):

$$B_H(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{1}{(i\omega)^{H+1/2}} (e^{it\omega} - 1) dB(\omega) \quad (4.11)$$

It is a Gaussian, continuous, centred and non-stationary second-order process. Applying the initial condition $B_H(0) = 0$, its covariance function (ρ) is defined as:

$$\rho(s, t) = E[B_H(s)B_H(t)] = \frac{a^2 V_H}{2} (|t|^{2H} + |s|^{2H} - (t - s)^{2H}) \quad (4.12)$$

for $0 < s \leq t$ where $E[\cdot]$ is the mathematical expectation, a is a constant and V_H a function of H defined as (Barton and Poor, 1988):

$$V_H = \Gamma(1 - 2H) \frac{\cos(\pi H)}{\pi H} \quad (4.13)$$

where Γ is the gamma function.

The self-similarity property of the fBm can be expressed as:

$$B_H(km) \equiv k^H B_H(m); \quad \forall k \text{ and } m > 0 \quad (4.14)$$

As fBm is a non-stationary process, it is more convenient to study its incremental process called fractional Gaussian noise (fGn) defined as:

$$G_m(k) = B_H(k+1) - B_H(k) \quad (4.15)$$

The corresponding autocovariance function $\gamma(\cdot)$ of $G_m(k)$ is given by:

$$\gamma(\tau) = E[G_m(k)G_m(k+\tau)] = \frac{\sigma_m^2}{2|m|^{2H}} (|\tau+m|^{2H} + 2|\tau|^{2H} + |\tau-m|^{2H}) \quad (4.16)$$

where $\sigma_m^2 = a^2 V_H |m|^{2H}$ is the variance of G_m .

For fGn class signals, the probability distribution of a segment of the signal is independent of the segment size and its temporal position in the signal (Eke et al., 2002). This property plays an important role in this work because a local computation of the fractal dimension is performed as detailed in the following subsection.

The Power Spectra Density (PSD) of the fGn can be calculated from (4.16), and when m tends to infinity, a normalised spectrum of the fBm increments can be defined as (Flandrin, 1992):

$$PSD_{G_m}(f) \propto |f|^{1-2H} \quad (4.17)$$

In this thesis, the fGn-based Spectral Estimator (GSE) is used in order to compute the parameter H from the PSD of the fBm increments. Equation (4.17) shows that the PSD follows a law in $|f|^{1-2H}$. In a log-log scale, the PSD function of the fBm increments is a line of slope $1 - 2H$. The H parameter can be estimated by linear regression. Regarding to the GSE implementation, the average spectrum of the fGn is estimated for each patch by the square of the modulus of the Fourier transform computed by the Fast Fourier Transform (FFT) algorithm.

Fractal analysis for retinal damage detection

In this thesis, fractal dimension of 2D image patches is extracted. With the aim of reducing the loss of information, a multi-dimensional extension of the fBm is used by means of the multiparameter Wiener process, or Brownian sheet presented in (Pesquet-Popescu and Lévy-Véhel, 2002). An anisotropic extension of the fGn allows the extraction of different Hurst exponents, one parameter in each of the L directions, gathered in the vector parameter $H = (H_1, H_2, \dots, H_L)$. In particular, the Hurst parameter in $L = 2$ directions (0 and 90 degrees) is computed from the average PSD (in each direction) and stacked as a vector $H = [H^{0^\circ}, H^{90^\circ}]$.

Let X be an image patch of dimensions $M \times N$, for each row m we obtain the 1D signal $d[n] = X[m, n + 1] - X[m, n]$ and compute the discrete Fourier transform given by:

$$D[k] = \sum_{n=0}^{N-1} d[n] e^{-j2\pi kn/N} \quad (4.18)$$

These 1D spectra compound the M rows of a 2D matrix $D'[k, l]$. Finally, the PSD is computed by a pointwise averaging in each column as:

$$PSD[k] = \frac{1}{M} \sum_{l=0}^{M-1} D'[k, l] \quad (4.19)$$

Figure 4.8(a) shows a patch extracted from the retina background of the fundus image displayed in Figure 4.7; the white arrow defines a specific patch row in which the intensity is studied. We can notice the non-stationarity of this intensity signal (Figure 4.8(b)) along the patch width. As recalled above, the increments of the intensity are easier to study due to their stationarity (Figure 4.8(c)). In Figure 4.8(d) it is possible to observe the \overline{PSD}^{0° extracted from the patch under study by means of Equation (4.19).

Finally, in order to extract the H parameter of the patch, a linear regression in a log-log scale for the PSD curve (Figure 4.8(d)) is performed and the Hurst exponent is estimated using Equation (4.17). Linear regression is performed in the frequency interval $f \in [f_{ini}, f_{end}]$ in which the PSD curve, plotted in a log-log scale, is linear. This linear behaviour shows the fractal nature of the tissue under study. The intergroup roughness of image patches is quite different, in

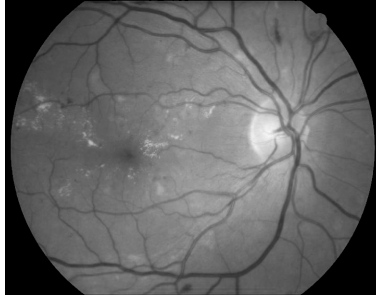


Figure 4.7: Fundus image to illustrate the fractal analysis computation.

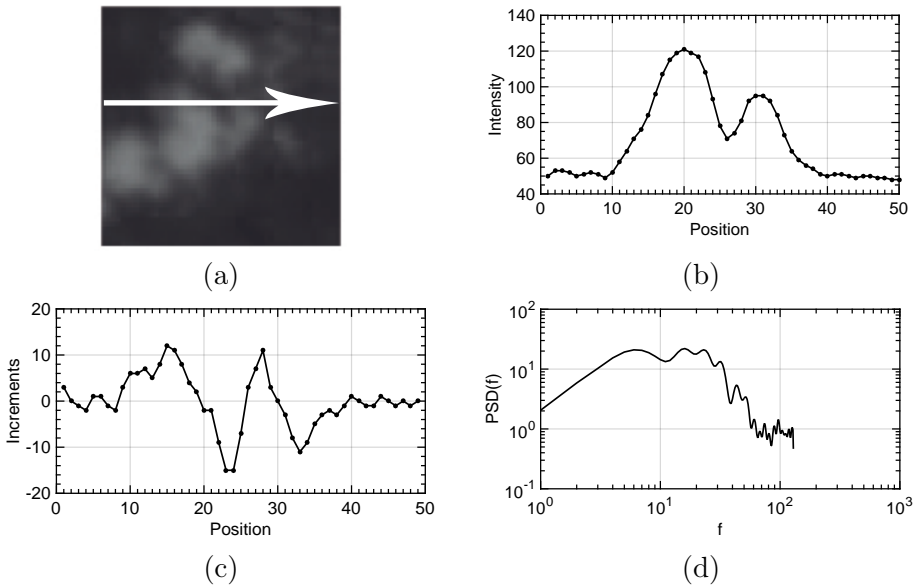


Figure 4.8: Process to compute the local fractal dimension: (a) A pathological ROI extracted from a retinal image, (b) the intensity of each pixel extracted from the ROI according to the arrow, (c) the increments of the row presented and (d) the PSD of the increments of all rows of the ROI.

other words, the linear segment in the log-PSD curve for patches labelled as healthy and pathological is located in a different frequency interval. However, this range of frequencies is quite similar between patches of the same class (i.e. low intragroup deviation). These facts allow the automatic computation of the optimal frequency range for each category.

Two representative subsets of patches containing N_h healthy patches and N_p pathological patches respectively are extracted from the training set. For each patch from both subsets, the optimal frequency range is manually selected $f_i = [f_{ini_i}, f_{end_i}]$ where i ranges between 1 and N_p or N_h depending on the subset. Thanks to the low intragroup deviation, two frequency ranges of reference $f'_h = [f_{hini}, f_{hend}]$ and $f'_p = [f_{pini}, f_{pend}]$ can be computed from each subset of the selected frequencies $\Omega_h = \{f_1, f_2, \dots, f_{N_h}\}$ and $\Omega_p = \{f_1, f_2, \dots, f_{N_p}\}$ as the minimum f_{ini_i} and maximum f_{end_i} respectively. These frequency ranges of reference are used in the automatic computation of the fractal dimension. For each patch, the curvature inside both frequency ranges is calculated. The frequency range in which the log-PSD curvature is minimum is the frequency range selected to carry out the linear regression. Algorithm 4.2 describes the complete process to locally estimate the fractal dimension in $L = 2$ directions from RGB fundus images.

4.1.4 Granulometry profile

Introduction

Since the 1960s, mathematical morphology has become increasingly popular in the image processing field due mainly to its versatility in many applications and rigorous mathematical description. Mathematical morphology was originally designed by Matheron and Serra to solve problems in mineralogy for characterising physical properties of several materials by means of the analysis of their geometrical structure (Matheron, 1967, 1975; Serra, 1983). Their investigations led to a series of novel morphological operators, based on non-linear functions, to extract relevant structures of an image that are useful in the representation of region shape and size.

One of the most interesting techniques based on mathematical morphology is granulometry (Matheron, 1967). Granulometries and some metrics extracted from them can be applied to a wide range of tasks such as segmentation, feature extraction, texture characterisation or size estimation among others. Applying a series of morphological opening (closing) operations with increasing-size structuring elements (SE), size distributions of different objects in an image can be obtained. Typically SE is a disk or a line giving place to an isotropic or angular granulometry, respectively. The obtained size distribution provides shape and size information in ordered textures while in disordered textures, granulometry extract the degree of granularity. This technique can be computed on binary or grey-scale images providing signatures that can be

Algorithm 4.2: Local fractal analysis computation

Data: RGB fundus image $\mathbf{f} = (f_r, f_g, f_b)$ (can be the original retinal image or a pre-processed version as it will be detailed in Section 6.2.1), optic disc mask \mathbf{M}_{od} , external mask \mathbf{M}_{ext} , size of the sliding window \mathbf{N}_w , vectors \mathbf{r} and \mathbf{c} containing the y and x coordinates of the centres of the patches in which the local analysis is performed (see Section 4.1.1).

Result: Hurst exponent matrix \mathbf{H} containing the fractal dimension in two directions per patch.

Initialization:

$ind \leftarrow 1$;

Green channel selection:

$f_{in} \leftarrow f_g$;

Mask computing:

$M \leftarrow and(M_{od}, M_{ext})$

Local fractal analysis:

for $i \leftarrow 1$ **to** $length(\mathbf{r})$ **do**

for $j \leftarrow 1$ **to** $length(\mathbf{c})$ **do**

$M^\circ \leftarrow M^{N_w}(r_i, c_j)$;

$f_{in}^\circ \leftarrow f_{in}^{N_w}(r_i, c_j)$;

if $\sum_x \sum_y M^\circ > 0$ **then**

$PSD^{0^\circ} \leftarrow PSD(f_{in}^\circ)$;

$PSD^{90^\circ} \leftarrow PSD(rotate(f_{in}^\circ, 90))$;

$\overline{PSD}^{0^\circ} \leftarrow mean(PSD^{0^\circ})$;

$\overline{PSD}^{90^\circ} \leftarrow mean(PSD^{90^\circ})$;

$\widehat{PSD}^{0^\circ} \leftarrow polyfit(\log_{10}(\overline{PSD}^{0^\circ}))$;

$\widehat{PSD}^{90^\circ} \leftarrow polyfit(\log_{10}(\overline{PSD}^{90^\circ}))$;

$H_{ind} \leftarrow [-0.5(slope(\widehat{PSD}^{0^\circ}) + 1), -0.5(slope(\widehat{PSD}^{90^\circ}) + 1)/2]$;

$ind \leftarrow ind + 1$;

end

end

used for feeding a classifier able to recognise different patterns or discriminate among several textures.

Texture classification and feature extraction by granulometries have many applications in digital imaging (Khatun et al., 2011a,b; Kupidura and Uwarowa, 2017; Ye and Liu, 2008). In the field of medical imaging, this technique has been used for example in the detection of malaria parasites

(Ruberto et al., 2000), to measure interstitial chronic damage (Colomé-Serra et al., 1992), in the estimation of trabecular thickness in grey-scale images (Moreno et al., 2012), or in the segmentation of microscopic breast cells (Cherni et al., 2009). In the particular case of fundus images, the use of granulometry as texture descriptor is not common but in Agurto et al. (2012) this technique is used with the aim of detecting neovascularization, i.e. the abnormal formation of blood vessels in the retina due to the lack of oxygen. This condition is indicative of late-stage diabetic retinopathy as it was explained in Section 2.3.

It is important to remember the main purpose of this thesis is to create novel algorithms to detect DR and AMD in the first stage of the pathologies. For these reason, in this chapter granulometry is proposed as an original and novel descriptor to detect exudates, haemorrhages and microaneurysms, in other words, the first signs related to the eye diseases manifested in fundus images.

Theoretical framework

Mathematical morphology is a methodology based on the application of lattice theory to spatial objects and operators in continuous or discrete spaces. Let f be a grey-level image, $f \in F(E, T)$ such that $f(x) : E \rightarrow T$ where $x = (x, y) \in E$ is the pixel position and $E \subset \mathbb{Z}^2$. In the case of valued discrete image, $T = t_{min}, t_{min+1}, \dots, t_{max}$ is an ordered set of grey-levels. Typically in 8-bit images $t_{min} = 0$ and $t_{max} = 255$. In the case of binary images, they are usually considered as discrete sets, $X \in P(R)$.

After choosing a fixed set, $B \subseteq E$ according to some a priori knowledge about the geometry of the relevant or irrelevant image structures, also called structuring element (SE), the dilation operator (denoted by $\delta_B(X)$) is defined as:

$$\delta_B(X) = X \oplus \check{B} = \{h \in E, X \cap B_h \neq \emptyset\} = \bigcup_{b \in \check{B}} X_b = \bigcup_{x \in X} \check{B}_x, \quad (4.20)$$

and the erosion operator (denoted by $\epsilon_B(X)$) is given by:

$$\epsilon_B(X) = X \ominus \check{B} = \{h \in E, B_h \subseteq X\} = \bigcap_{b \in \check{B}} X_b. \quad (4.21)$$

where \check{B} is the reflection of B with respect to the origin, i.e., $\check{B} = \{-b \mid b \in B\}$. Note that the SE is a subset defined in the origin $o \in E$, then to each point p of E corresponds the translation mapping o to p , and this translation maps B onto B_p , i.e., $B_p = \{b + p \mid b \in B\}$. In both operations, for a fixed structuring element B , all its translations B_p are considered.

Given a grey-level image $f \in F(E, T)$, its dilation (erosion) by a flat structuring element B is introduced as the dilation (erosion) of each level set $X_t(f)$ by B :

$$\begin{aligned} \delta_B(f)(x) &= \sup\{t_l \mid x \in \delta_B(X_{t_l})\}, t_l \in T \\ \epsilon_B(f)(x) &= \inf\{t_l \mid x \in \epsilon_B(X_{t_l})\}, t_l \in T \end{aligned} \quad (4.22)$$

Dilation and erosion can be combined successively to obtain a new set of operators or basic filters. In the binary case, the opening operation (denoted by $\gamma_B(X)$) is defined as:

$$\begin{aligned} \gamma_B(X) &= (X \ominus \check{B}) \oplus B = \{h \in E \mid \nexists g \in E, B_g \subseteq X \text{ and } h \in B_g\}, \\ \gamma_B(X) &= \bigcup_{h \in E} \{B_h \mid B_h \subseteq X\}. \end{aligned} \quad (4.23)$$

and the closing filter (denoted by $\varphi_B(X)$) is given by:

$$\varphi_B(X) = (X \oplus \check{B}) \ominus B = \{h \in E \mid h \in \check{B}_y \Leftarrow \check{B}_y \cap X \neq \emptyset\}. \quad (4.24)$$

The two elementary operations of grey-level erosion and dilation can be composed together to yield a new set of grey-level operators having desirable feature extractor properties which are given by, the grey-level opening:

$$\gamma_B(f)(x) = (f \circ B)(x) = \delta_B(\epsilon_B(f))(x) \quad (4.25)$$

and the grey-level closing:

$$\varphi_B(f)(x) = (f \bullet B)(x) = \epsilon_B(\delta_B(f))(x) \quad (4.26)$$

Making use of the previously explained operators, a shape descriptor can be defined. Let us first consider an opening $\gamma_i(f)$, applied to an image f with a SE of size i . The opening can be computed as a sequence of an erosion followed

by a dilation. When opening is computed on the image with a SE of increasing size (λ), we obtain a morphological opening pyramid (or granulometry profile) which can be formalized as:

$$\Pi_\gamma(f) = \{\Pi_{\gamma_\lambda} : \Pi_{\gamma_\lambda} = \gamma_\lambda(f), \forall \lambda \in [0, \dots, n_{max}]\} \quad (4.27)$$

where n_{max} represents the maximum size of the structuring element.

Some basic properties can be extracted from the morphological opening pyramid $\Pi_\gamma(f)$:

- γ_0 is the identity mapping; i.e., $\gamma_0(f) = f$
- γ_λ is increasing; i.e., $f < g \Rightarrow \gamma_\lambda(f) \leq \gamma_\lambda(g), \forall \lambda \geq 0$, for every f and g .
- γ_λ is anti-extensive; i.e., $\gamma_\lambda(f) \leq f, \forall \lambda \geq 0$, for every f .
- γ_λ follows the absorption law (Soille, 2003); i.e. $\forall \lambda \geq 0, \forall \mu \geq 0, \gamma_\lambda \gamma_\mu = \gamma_\mu \gamma_\lambda = \gamma_{max(\lambda, \mu)}$

The morphological openings, γ_{nB} , B compact convex, satisfies the four granulometric postulated detailed above. They also satisfy two fundamental properties:

- The γ_{nB} are translation invariant
- The following equation is satisfied $\gamma_n(f) = \gamma_1(\frac{1}{n}f)$. It says that there is a unit sieve γ_1 , and that any other sieve in the process can be evaluated by first scaling the image by the reciprocal of the parameters, filtering by the unit sieve, and then rescaling.

By duality, a closing, $\varphi_i(f)$ is defined as the dilation of an original image f with a SE of size i . In the same way, a morphological closing pyramid is an anti-granulometry profile and can be computed on the image performing repeated closings with a SE of increasing size (λ) defined as:

$$\Pi_\varphi(f) = \{\Pi_{\varphi_\lambda} : \Pi_{\varphi_\lambda} = \varphi_\lambda(f), \forall \lambda \in [0, \dots, n_{max}]\} \quad (4.28)$$

Let $m(f)$ be the Lebesgue measure of a discrete image f : $m(f)$ is the area of f in the binary case (number of pixels) and the volume in the grey scale case (sum of

pixel values). Making use of the morphological pyramids established above, a shape descriptor can be defined. The granulometry curve, or pattern spectrum of f with respect to Γ is defined as the following (normalised) mapping:

$$PS_{\Gamma}(f, n) = PS(f, n) = \frac{m(\Pi_{\gamma_n}(f)) - m(\Pi_{\gamma_{n+1}}(f))}{m(f)}, n \geq 0 \quad (4.29)$$

The pattern spectrum $PS_{\Gamma}(f, n)$ (also called size density of f) maps each size n to some measure of the bright image structures with this size: loss of bright image structures between two successive openings. It is a probability density function (a histogram) in which a large impulse in the pattern spectrum at a given scale indicates the presence of many image structures at that scale.

By duality, the concept of pattern spectrum extends to anti-granulometry curve $PS_{\Phi}(f)$, by closings $\varphi_n(f)$, stacked in the morphological pyramid Π_{φ} :

$$PS_{\Phi}(f, -n) = PS(f, -n) = \frac{m(\Pi_{\varphi_n}(f)) - m(\Pi_{\varphi_{n-1}}(f))}{m(f)}, n \geq 0 \quad (4.30)$$

This spectrum characterises the size of dark image structures. Both granulometry curve and anti-granulometry curve can be appended into a unique curve, with closings versus size on the left side (negative side) and openings versus size on the right side (positive side) of the diagram; i.e.:

$$\{-n, 0, n\} \rightarrow PS(f, n) = \{PS_{\Phi}(f, -n), 0, PS_{\Gamma}(f, n)\}, n \geq 0 \quad (4.31)$$

Granulometry profile of retinal lesions

Granulometry is also proposed in this thesis as a novel descriptor for retinal damage detection. With this objective, Π_{γ} and Π_{φ} are computed from a specific fundus image f following the Equations 4.27 and 4.28 respectively. The morphological pyramids are constructed by performing openings and closings with an increasing SE size defined by a specific step (s) and a maximum value (n_{max}). In Figure 4.9 and Figure 4.9, the different levels of the pyramids Π_{γ} and Π_{φ} computed for a particular fundus image are shown.

From the morphological pyramids, a local description of the shape and size of the retinal texture is performed by computing the pattern spectrum of squared patches. In particular $PS_{\Gamma}^{\circ}(f, n)$ and $PS_{\Phi}^{\circ}(f, n)$ are locally computed

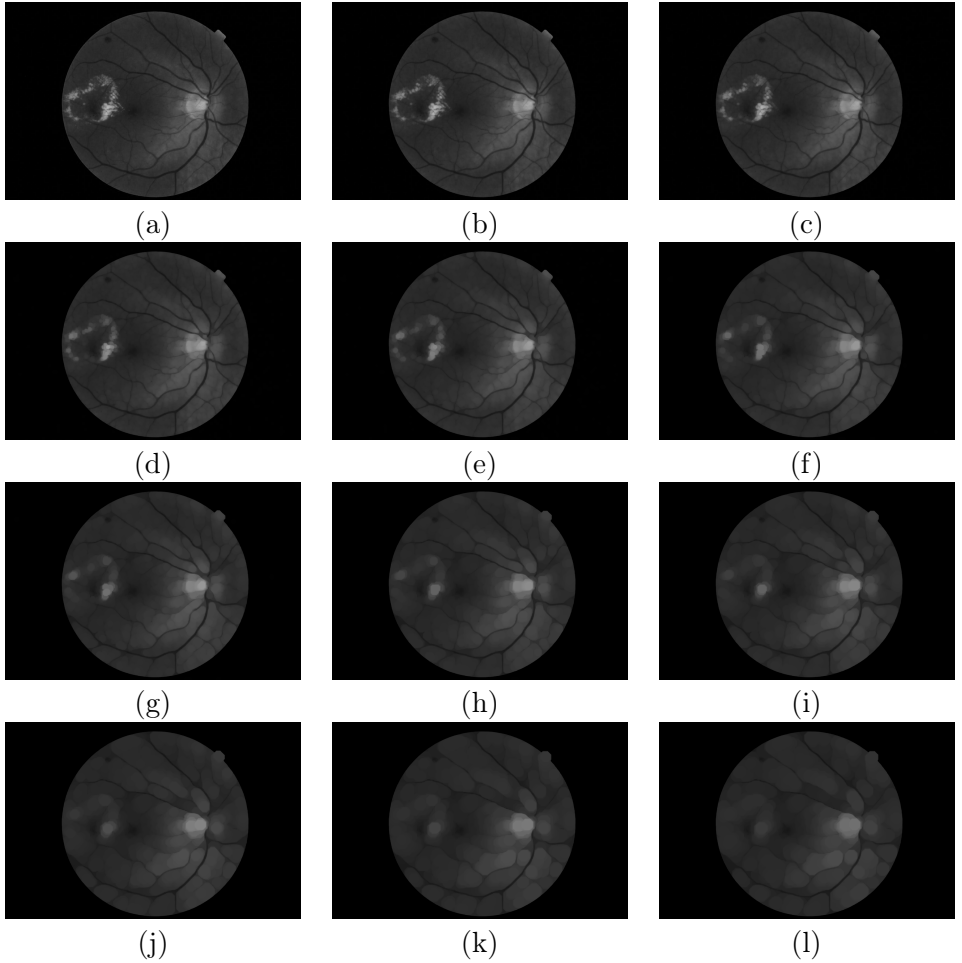


Figure 4.9: Pyramid of openings computed for a fundus image using an isotropic structuring element. It is composed by twelve images determined by using $s = 2$ and $n_{max} = 22$. (a) Original image; i.e. γ_0 is the identity mapping; (b-k) are the results of applying the opening operator with increasing size (according to $s = 2$) and (l) is the last image of the pyramid corresponding to the operation $\gamma_{n_{max}}$.

for each patch of the green channel of fundus images (extracted according to the explanation in Section 4.1.1), following the Equations (4.29) and (4.30) respectively. In addition, the combination of both descriptors takes place to the curve $PS^\diamond(f, n)$ according to the Equation (4.31). Note that the \diamond symbol denotes the locality of the descriptor. In Figure 4.11(a,c), the local extraction

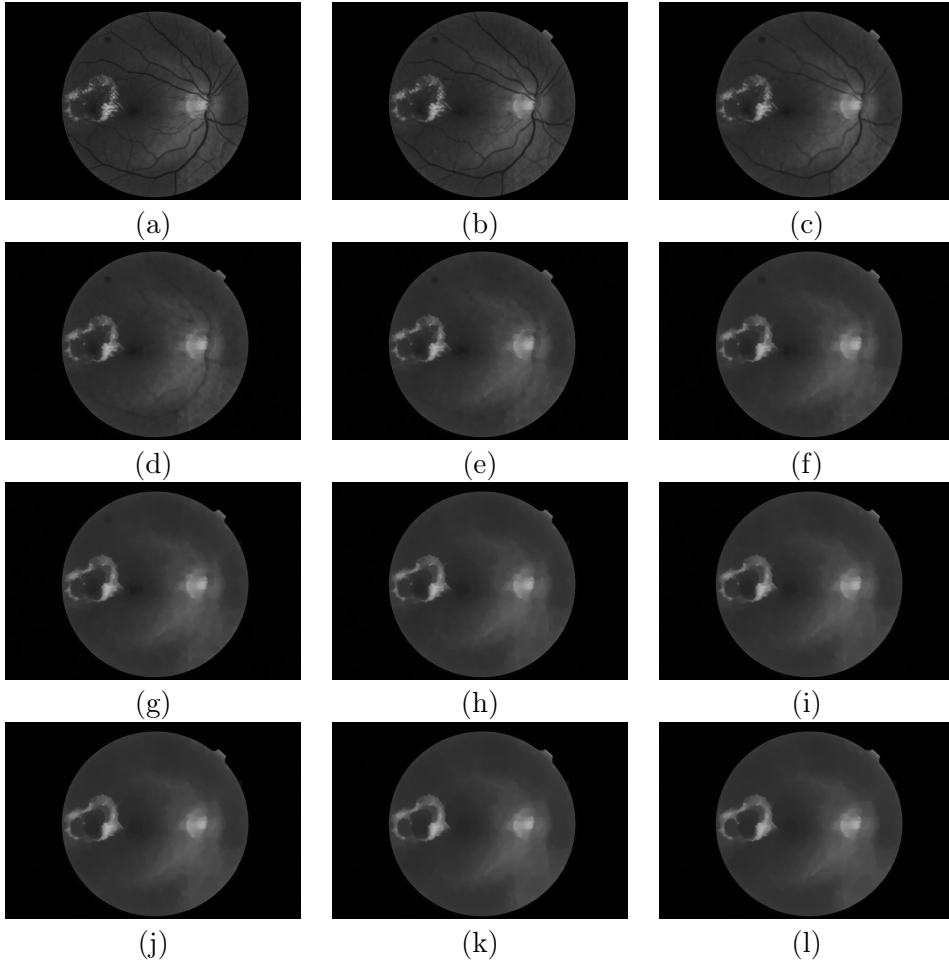


Figure 4.10: Pyramid of closings computed from a fundus image using an isotropic structuring element. It is composed by twelve images determined by using $s = 2$ and $n_{max} = 22$. (a) Original image; i.e. φ_0 is the identity mapping; (b-k) are the results of applying the closing operator with increasing size (according to $s = 2$) and (l) is the last image of the pyramid corresponding to the operation $\varphi_{n_{max}}$.

of the granulometry and anti-granulometry curves from the morphological pyramids Π_γ and Π_φ is represented. $PS_\Gamma^\diamond(f, n)$ and $PS_\Phi^\diamond(f, n)$ descriptors extracted from the patch marked in red in a fundus image are reported in Figure 4.11(b,d). In addition, the whole pattern spectrum $PS^\diamond(f, n)$ is

observed in Figure 4.12. Algorithm 4.3 details the process to compute the local granulometries in RGB fundus images.

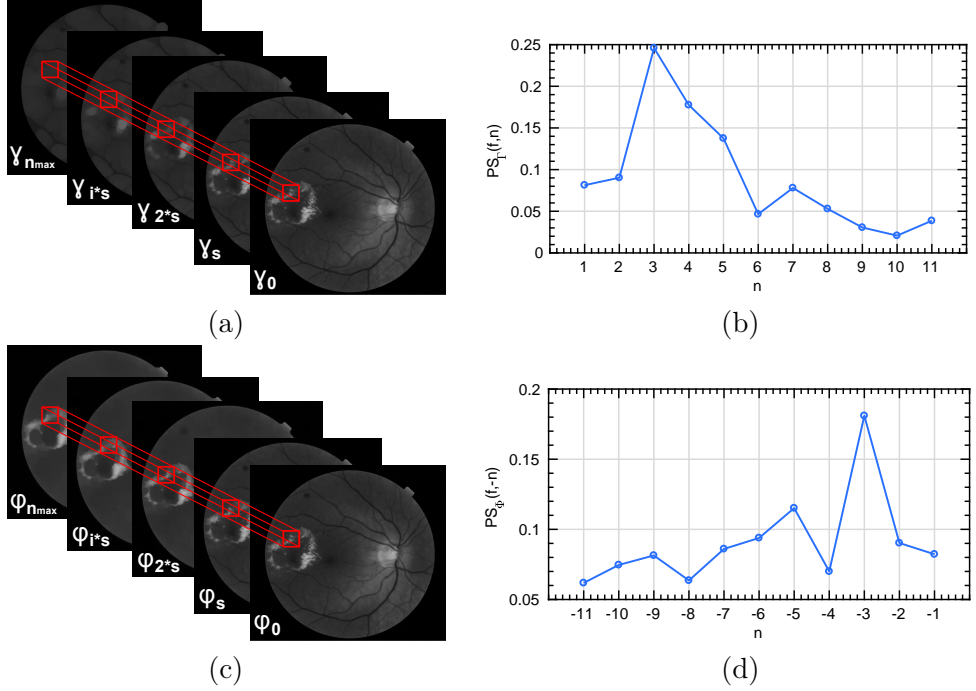


Figure 4.11: (a,c) represent the local extraction of the granulometry and anti-granulometry curves from the morphological pyramids Π_γ and Π_φ . (b,d) show the pattern spectrum computed for the patch marked in red in each case.

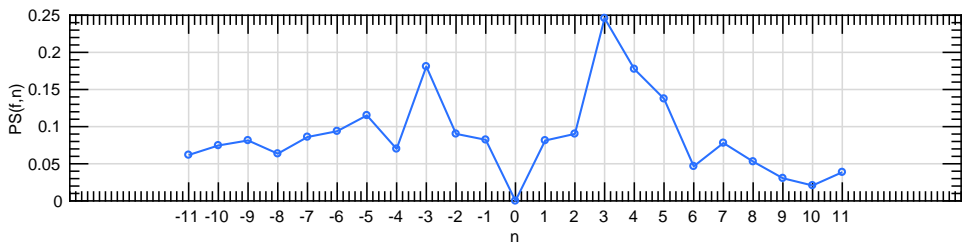


Figure 4.12: A global pattern spectrum extracted by the combination of the granulometric and anti-granulometric profiles.

Algorithm 4.3: Computation of the local granulometries

Data: RGB fundus image $\mathbf{f} = (f_r, f_g, f_b)$ (can be the original retinal image or a pre-processed version as it will be detailed in Section 6.2.1), optic disc mask \mathbf{M}_{od} , external mask \mathbf{M}_{ext} , size of the sliding window \mathbf{N}_w , vectors \mathbf{r} and \mathbf{c} containing the y and x coordinates of the centre of the patches in which the local analysis is performed (see Section 4.1.1).

Result: Feature matrices \mathbf{PS}_Γ and \mathbf{PS}_Φ containing one pattern spectrum per row.

Initialization:

s and n_{max} pyramid parameters

$ind \leftarrow 1$;

Green channel selection:

$f_{in} \leftarrow f_g$;

Morphological pyramid computation:

$\Pi_\gamma \leftarrow \gamma_{0 \rightarrow s \rightarrow n_{max}}(f_{in})$ Openings with increasing SE size;

$\Pi_\varphi \leftarrow \varphi_{0 \rightarrow s \rightarrow n_{max}}(f_{in})$ Closings with increasing SE size;

Mask pyramid computation:

$M \leftarrow and(M_{od}, M_{ext})$

Removing disk and external contribution:

$\Pi_\gamma \leftarrow \Pi_\gamma(M > 0)$;

$\Pi_\varphi \leftarrow \Pi_\varphi(M > 0)$;

Local granulometries:

for $i \leftarrow 1$ **to** $length(\mathbf{r})$ **do**

for $j \leftarrow 1$ **to** $length(\mathbf{c})$ **do**

$\Pi_\gamma^\diamond \leftarrow \Pi_\gamma^{N_w}(r_i, c_j)$;

$\Pi_\varphi^\diamond \leftarrow \Pi_\varphi^{N_w}(r_i, c_j)$;

for $k \leftarrow 1$ **to** $length(\Pi_\gamma^\diamond)$ **do**

$PS_{\gamma_k} \leftarrow \frac{\sum(\Pi_{\gamma_k}) - \sum(\Pi_{\gamma_{k+1}})}{\sum(\Pi_{\gamma_0})}$;

$PS_{\varphi_k} \leftarrow \frac{\sum(\Pi_{\varphi_k}) - \sum(\Pi_{\varphi_{k-1}})}{\sum(\Pi_{\varphi_0})}$;

end

$PS_{\Gamma_{ind}} \leftarrow PS_\gamma$;

$PS_{\Phi_{ind}} \leftarrow PS_\varphi$;

$ind \leftarrow ind + 1$;

end

end

4.2 Machine Learning methods for retinal tissue classification

4.2.1 Classification strategy

In the previous section, novel descriptors in the field of retinal imaging were formulated with the purpose of automatically detecting eye diseases. The proposed approach is based on the local analysis of fundus images, in other words, we propose a patch-based solution to identify the most common causes of vision impairment and blindness by using computer vision techniques. In this way, the next step to the extraction of the visual information is the classification of patches according to healthy and pathological classes. Let be a dataset composed by N retinal images, the first step of the proposed classification pipeline is to perform an external K -fold cross-validation with the objective of removing the bias introduced by a specific division of the dataset into training and test subsets. The external cross-validation also provides robustness to the decision system because all the instances of the database will be used in the model creation and prediction stages. In each fold iteration, a different partition is selected as test subset while the rest of partitions are used to train the model. As it was introduced in Section 4.1.1 and as it can be observed in Figure 4.1(a), pathological areas represent only small regions of the whole image, which results in a very imbalanced dataset when local feature extraction is performed. Training a classifier with an imbalanced dataset can produce overfitting to the majority class (“healthy” in our case) (Tapia et al., 2016). To avoid this problem, we proceed as follows. Let us assume that the number of healthy and pathological samples are M and N respectively where $M \gg N$. Thus, the set of all healthy samples is randomly permuted and partitioned into $T = \text{round}(M/N)$ subsets with the same cardinality as the number of pathological samples. A committee of T classifiers is then learned with training sets formed by joining all pathological training samples and each partition of healthy training samples. In the test stage, testing samples are evaluated for each of the T models and soft majority voting is applied to the output probabilities as final criterion. If the obtained probability is higher than a given threshold δ , the patch is assigned to the class “pathological”. Notice that this is not a bagging procedure (Breiman, 1996), since sampling is not uniform and with replacement. An overview of the whole classification process can be observed in Figure 4.13.

In this thesis, classical and novel machine learning methods in the field of automatic retinal damage detection are used to perform the classification step. The theoretical basis of these algorithms are detailed in the following lines of this section.

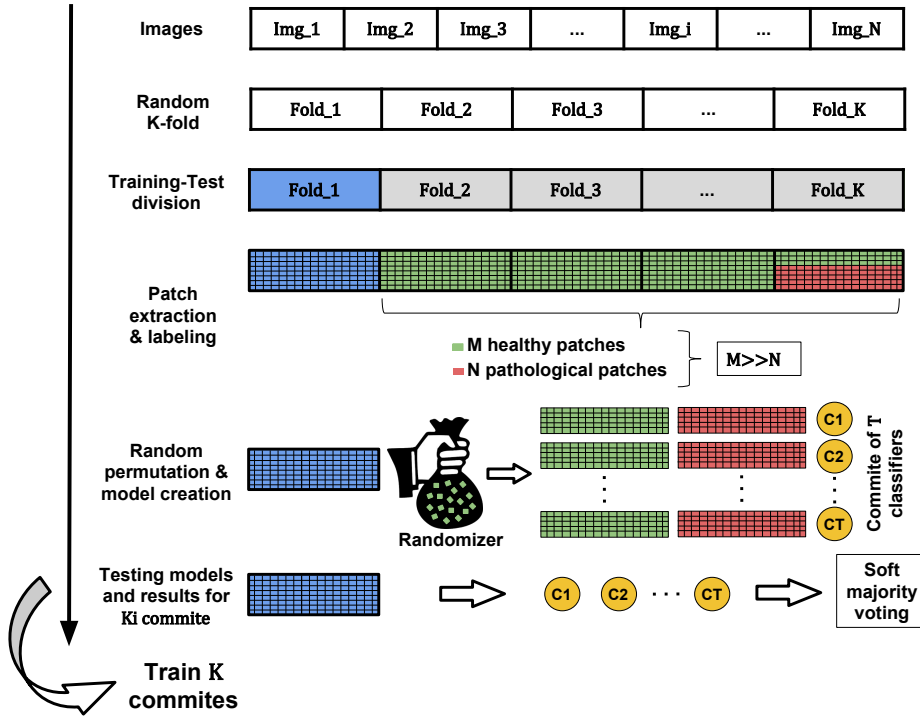


Figure 4.13: Process of creating the machine learning models using a generic dataset. Green and red samples are used in the creation of the model while blue instances refer to the samples used in the testing stage.

4.2.2 Random Forest

Decision trees conform a well-known category of machine learning methods. The name of these algorithms comes from the tree structure they have, in which the branches define conditional bifurcations based on the attributes of each instance. In the same way, the leaves of the trees indicate the class assigned to the object that follow a specific path (Figure 4.14). This subset of methods composes an important category of the supervised learning (Witten et al., 2011).

Additionally, bootstrap aggregation (bagging) (Breiman, 1996) is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms, reducing the variance and avoiding the overfitting. The essential idea of this technique is to average many noisy

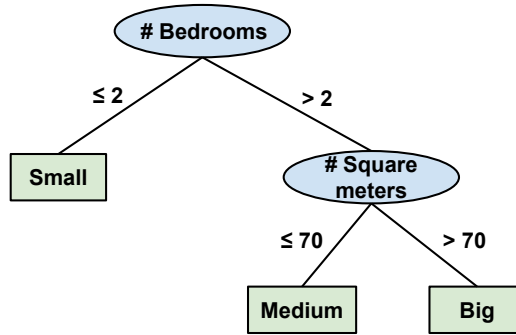


Figure 4.14: Example of a simple tree classifier in which two features participate in the classification process.

but approximately unbiased models, and hence reduce the variance. Trees are ideal candidates for bagging, since they can capture complex interaction structures in the data, and if they are grown sufficiently deep, trees have relatively low bias.

Random Forests are a combination of tree-structured predictors $\{h(\mathbf{x}, \Theta_k), k = 1, \dots, K\}$ so that each tree depends on the training set and values of a random vector, $\{\Theta_k\}$, independently sampled from the past random vectors $\{\Theta_1, \dots, \Theta_{k-1}\}$ and with the same distribution for all trees in the forest (Breiman, 2001).

Being N the number of cases in the training set, different samples of these N cases are taken at random but with replacement for creating K different predictors or trees (Figure 4.15), with training sets T_1, \dots, T_K . Suppose M be the number of input variables or features of each case. Let m be the number of features to be analysed by one node ($m < M$), which are randomly selected from the group of M input variables, typically values for m are \sqrt{M} or even as low as 1. The method combines the bagging approach and random feature selection in order to build a collection of decision trees with controlled variance. The use of random feature subsets ensures low correlation between the generated trees.

Let focus on the tree k with a training set T_k , for each node of the tree:

1. Select m variables at random from the M features or input variables.
2. Pick the best variable/split-point among the m .
3. Split the node into two daughter nodes.

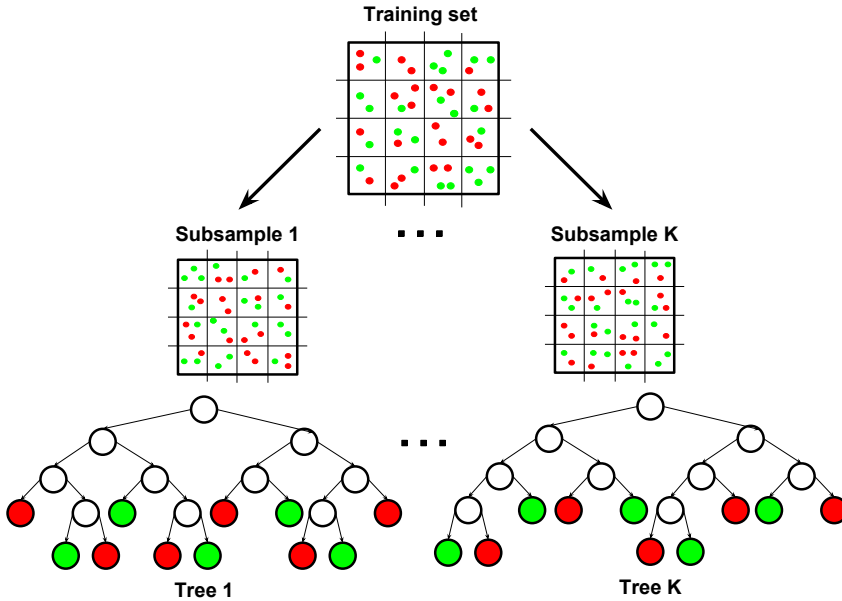


Figure 4.15: Training stage of the Random Forest algorithm, in which, tree predictors are created taking random samples of the training set.

To predict a new instance (testing data X), it is pushed down the tree, following the algorithm described above. Then, the label assigned to the instance will be the label corresponding to the terminal node of the tree. This process is iterated by the K predictors or trees and the final classification label will be obtained by majority voting of the trees. Figure 4.16 illustrates the process.

4.2.3 Support Vector Machine

One of the classification algorithms widely used in the scientific community is the Support Vector Machine (SVM). This technique belongs to the family of linear predictors and discriminatory models. SVM is used in high-dimensional feature spaces. This fact propitiates a high complexity and computational cost in the training stage; two factors to be taken into account for choosing SVM as classification method.

The objective of SVM is to produce a classification model, based on the information of features and labels of the training set, able to predict the correct class of new samples belonging to the test set. For this purpose, Support Vector

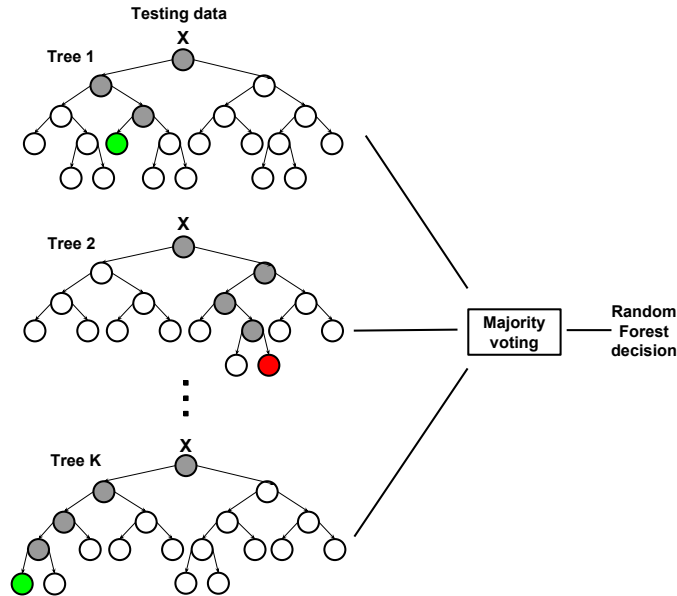


Figure 4.16: Testing procedure for a new sample X as input to the Random Forest.

Machine builds a hyperplane of separation in the input space maximising the distance (a.k.a. margin) with respect to the supported vectors of the different classes (Chang and Lin, 2011).

Given a linear-separable training set of N instance-label pairs (\mathbf{x}_i, y_i) , $i = 1, \dots, N$, where $\mathbf{x}_i \in R^n$ and a label $y_i \in \{+1, -1\}^t$ is assigned to each vector \mathbf{x}_i , the hyperplane solution can be expressed as:

$$\mathbf{w}^T \mathbf{x}_i + b = 0 \quad (4.32)$$

where \mathbf{w} is the normal vector to the hyperplane solution and the coefficient b is the intersection.

From the proposed hyperplane, the space can be separated into a positive and a negative regions. The region that satisfies $\mathbf{w}^T \mathbf{x}_i > 0$ corresponds to the samples of the positive class while the region that satisfies $\mathbf{w}^T \mathbf{x}_i < 0$ corresponds to the negative instances. In Figure 4.17 an example dataset is represented. Red instances represent the positive class while green samples represent the negative one. The support vectors (highlighted samples) define

two hyperplanes parallel to the hyperplane solution given by the maximization of the distance $d_+ + d_-$. The classification function can be defined as:

$$f(x) = \text{sgn}(\mathbf{w}^T \mathbf{x}_i + b) \quad (4.33)$$

and considering the sign of applying the hyperplane solution, the predicted class will be obtained.

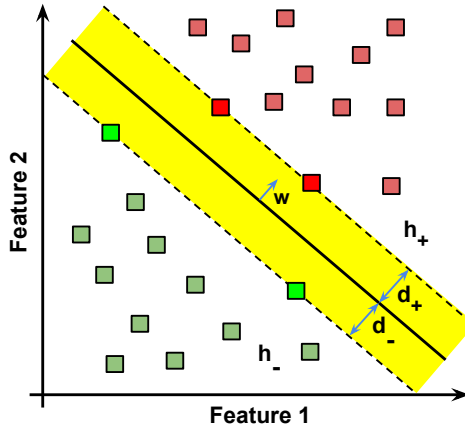


Figure 4.17: Representation of the SVM classification process. h_+ and h_- are the hyperplanes defined by the support vectors for the positive (red) and negative (green) class, respectively. The maximization of the distance $d_+ + d_-$ (a.k.a margin) gives place to the hyperplane solution.

The hyperplanes corresponding to the support vectors, parallels to the hyperplane solution, are defined as $\mathbf{w}^T \mathbf{x}_i + b = 1$ for the hyperplane of the positive class (denoted as h_+ in Figure 4.17) and $\mathbf{w}^T \mathbf{x}_i + b = -1$ for the hyperplane of the negative class (h_-). Therefore, a sample to be labelled as positive has to satisfy the condition $\mathbf{w}^T \mathbf{x}_i + b \geq 1$ while a sample to be labelled as negative has to satisfy $\mathbf{w}^T \mathbf{x}_i + b \leq -1$.

Considering that the positive samples are labelled as +1 and the negative samples as -1 the last conditions can be jointed into a unique condition of classification:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad (4.34)$$

The margin can be computed as the distance between the hyperplanes in which the support vectors are included:

$$\text{margin} = d^+ + d^- = 2 \left(\frac{|\mathbf{w}\mathbf{x} + b|}{\|\mathbf{w}\|} \right) = 2 \frac{1}{\|\mathbf{w}\|} \quad (4.35)$$

The objective of SVM algorithm is to maximise the margin value. Taking into account that to maximise Equation (4.35) is equivalent to minimise the inverse function given by:

$$\Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 \quad (4.36)$$

being $\|\mathbf{w}\|^2 = \mathbf{w}^T \mathbf{w}$ and subject to the classification condition exposed in Equation (4.34). The hyperplane solution will be obtained solving the minimization problem and computing the optimal \mathbf{w} and b parameters.

At the beginning of the explanation a linear-separable training set was assumed. However, the most of datasets extracted from real applications do not meet this particularity. In these cases, slack variables must be introduced to relax the conditions of the optimisation problem allowing for wrong classifications. The formulation of the problem can be redefined as:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \\ \text{subject to} \quad & y_i(\mathbf{w}^T \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \end{aligned} \quad (4.37)$$

where \mathbf{w} denotes the normal vector to the hyperplane, ξ_i are lack variables which measure the degree of misclassification of the data, $b/\|\mathbf{w}\|$ is the distance between the origin and the hyperplane, $C > 0$ is the penalty parameter of the error term and Φ is a function that transforms the input vectors \mathbf{x}_i into a higher dimensional space. SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space.

Apart from the relaxation of the conditions, in some cases, more complex decision boundaries (i.e. not a simple line) are required to improve the classification results. In the kernel trick technique, a specific non-linear function (Φ) is used for mapping non-linear separable data points in the original space Σ , into a higher dimensional space in which the data is linearly separable. When this hyperplane of separation is projected back to the input space (Σ), it

has the form of a non-linear curve. The new feature space can be of unlimited dimension and thus, the hyperplane separating the data can be very complex. The function $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$ is known as kernel function and the basic ones are the following:

- linear: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$
- polynomial: $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d, \gamma > 0$
- radial basis function (RBF): $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \gamma > 0$
- sigmoid: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$

where γ , r and d are kernel parameters.

In different experiments of this work (see Chapter 6), SVM classifiers with linear and RBF kernels were involved. RBF kernel non-linearly maps samples into a higher dimensional space so it, unlike the linear kernel, can deal with datasets in which the relation between the class labels and features is nonlinear. The linear kernel is a specific case of RBF (Keerthi and Lin, 2003) since the linear kernel presents the same performance as the RBF kernel with a particular couple of values for the parameters C and γ . RBF kernel presents a reduced number of hyperparameters in comparison with a polynomial kernel for example, and it is highly suitable when the number of attributes is not very large (Chang and Lin, 2011). For these reasons RBF kernel is employed in this thesis when SVM algorithm is applied. Classification experiments using a linear kernel is also performed in order to establish a baseline method for comparison purposes.

As it was mentioned above, RBF kernel is characterised by two parameters (C, γ) . The parameter C trades off miss-classification of training examples against simplicity of the decision surface, while the parameter γ defines how far the influence of a single training example reaches. γ can be seen as the inverse of the radius of influence of samples selected by the model as support vectors. For a given problem, the best C and γ parameters are unknown; consequently some kind of model selection (parameter search) must be performed. The objective is to identify the optimal (C, γ) using the training set, so that the classifier can accurately predict unknown data (i.e. testing data). v -fold cross-validation is used to accomplish this goal. In this procedure, each instance of the whole training set is predicted and the cross-validation accuracy is computed taking into account the data correctly classified in the v folds. Note that cross-validation can prevent the overfitting problem.

Internal cross-validation was used in the experiments involving SVM with RBF kernel in order to “grid search” the optimal parameters. Several pairs of (C, γ) values were tried and the one with the highest cross-validation accuracy was selected. Exponential growing sequences of C and γ conform a good method to identify the optimal parameters. $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$ and $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$ were the sequences used in the SVM classification of healthy and pathological tissue. In terms of computational efficiency it is important to note that the “grid-search” is easily parallelized because each (C, γ) is independent. In addition, the complete grid-search is performed in two stages. First, a coarse grid delimitates the region in which the optimal parameters can be found. Afterwards, a fine grid search is responsible for accurately finding C and γ . The flow chart of Figure 4.18 illustrates the procedure followed in this PhD thesis for optimising the RBF-SVM parameters to accurately discriminate between healthy and pathological retinal tissues.

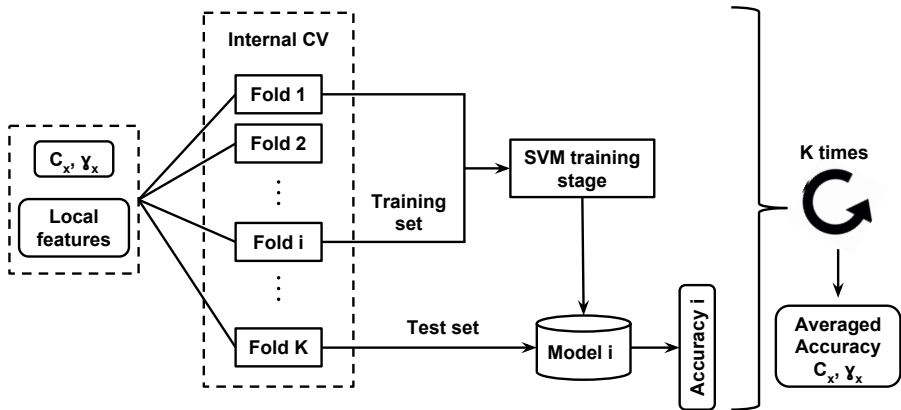


Figure 4.18: Internal cross-validation overview. Grid search of the RBF-SVM optimal parameters (C, γ) . For each possible pair (C, γ) an internal cross-validation with $k = 8$ folds is carried out. The (C, γ) values that maximise the classification accuracy are chosen as optimal parameters.

4.2.4 Gaussian processes for classification

Gaussian processes for classification purposes (GPCs) is a probabilistic approach which models the posterior probabilities of the test instances for a new input vector, given a training dataset. These probabilities must lie in the interval $(0,1)$, whereas a model based on Gaussian Processes make predictions that lie on the entire real axis. However, Gaussian Processes can be easily adapted to classification problems by mapping their output value using the

appropriate non linear activation function (Bishop, 2010). In the case of the binary classification problem in which the vector of labels is given by $y \in 0, 1$. If we define a Gaussian process over a function $f(x)$ and then transform the function using a logistic sigmoid $z = \sigma(f(x))$, then we will obtain a non-Gaussian stochastic process over functions $z(x)$ where $z \in (0, 1)$ (Bishop, 2010). In Figure 4.19(a), a sample from a Gaussian process prior over the function $f(x)$ is shown. In addition, the transformation of this sample by means of the logistic sigmoid function $\sigma(t) = (1 + \exp(-t))^{-1}$ can be observed in Figure 4.19(b). This function allows to obtain the probability class $\pi(x) = \sigma(f(x))$.

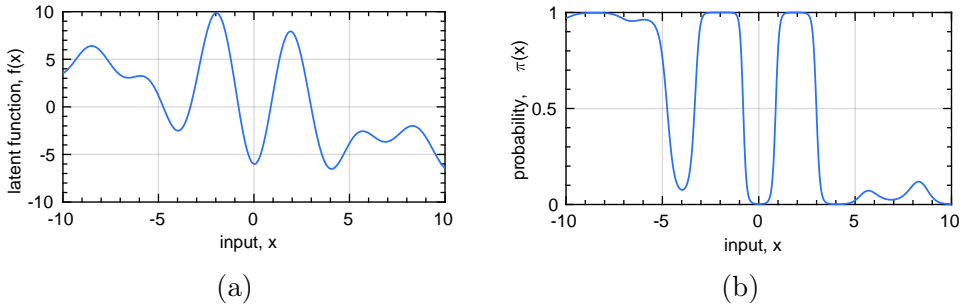


Figure 4.19: Gaussian Processes for Classification: (a) Sample from a Gaussian process prior over functions $f(x)$ and (b) the result of transforming this sample into logits using the logistic sigmoid function $\sigma(t)$.

Let $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^t$ be the matrix containing the samples, where each $\mathbf{x}_i \in \mathbb{R}^K$ is the descriptor for the i -th patch, calculated as described in Section 4.1.1, N is the number of patches used for training, and K is the length of the feature vector. Let $\mathbf{y} \in \{0, 1\}^N$ be the vector of labels, where 1 refers to “pathological patches” and 0 refers to “healthy patches”.

To model the relationship between samples and labels, the GPC formulation (Rasmussen and Williams, 2006) introduces a latent variable $\mathbf{f} \in \mathbb{R}^N$, which is in fact a prior on the functions over the feature space. This latent variable has the following Gaussian prior distribution:

$$p(\mathbf{f}|\Omega) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{K}), \quad (4.38)$$

where \mathbf{K} is the covariance matrix, which depends on a set of parameters Ω to be estimated and has the form $\mathbf{K}_{ij} = \mathbf{k}(\mathbf{x}_i, \mathbf{x}_j)$ where $\mathbf{k}(\cdot, \cdot)$ is a kernel function (Rasmussen and Williams, 2006). \mathbf{K} is the way in which GPs exploit the correlation between samples, since each \mathbf{K}_{ij} is calculated by evaluating a kernel function (an inner product in a transformed vector space) on the samples

\mathbf{x}_i and \mathbf{x}_j . Although \mathbf{K} depends on \mathbf{X} , we have removed it from dependencies in $p(\mathbf{f}|\Omega)$ for simplicity.

The relationship between \mathbf{f} and the labels \mathbf{y} is modeled by the so-called logistic likelihood function:

$$p(\mathbf{y}|\mathbf{f}) = \prod_{i=1}^N p(y_i|f_i) = \prod_{i=1}^N \left(\frac{1}{1 + e^{-f_i}} \right)^{y_i} \left(\frac{1}{1 + e^{f_i}} \right)^{1-y_i}. \quad (4.39)$$

Notice that for large positive values of f_i , $p(y_i = 1|f_i) \approx 1$, and for negative values of f_i with large absolute values $p(y_i = 1|f_i) \approx 0$.

The joint distribution of \mathbf{y} , \mathbf{f} , and Ω is:

$$p(\mathbf{y}, \mathbf{f}, \Omega) = p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\Omega)p(\Omega). \quad (4.40)$$

where we have used a flat improper prior for Ω .

Since $p(\mathbf{f}, \Omega|\mathbf{y})$ can not be calculated, we resort to the mean field approximation (Parisi, 1998) and solve the problem as:

$$\hat{q}(\mathbf{f}), \hat{q}(\Omega) = \arg \min_{q(\mathbf{f}), q(\Omega)} \int q(\mathbf{f})q(\Omega) \ln \frac{q(\mathbf{f})q(\Omega)}{p(\mathbf{y}, \mathbf{f}, \Omega)} d\mathbf{f}d\Omega, \quad (4.41)$$

where $\hat{q}(\Omega)$ is restricted to be a degenerate distribution, to finally obtain $p(\mathbf{f}, \Omega|\mathbf{y}) \approx \hat{q}(\mathbf{f})\hat{q}(\Omega)$, see (Bishop, 2010) for details.

Unfortunately Equation (4.41) can not be evaluated due to the functional form of the likelihood function in Equation (4.39). To alleviate this problem we use the lower bound provided in (Bishop, 2010):

$$\begin{aligned} \log p(\mathbf{y}|\mathbf{f}) &\geq \log \mathbf{H}(\mathbf{y}, \mathbf{f}, \boldsymbol{\xi}) = \\ &\mathbf{f}^t \left(\mathbf{y} - \frac{1}{2} \mathbf{1} \right) - \mathbf{f}^t \boldsymbol{\Lambda} \mathbf{f} + \boldsymbol{\xi}^t \boldsymbol{\Lambda} \boldsymbol{\xi} + \frac{1}{2} \boldsymbol{\xi}^t \mathbf{1} - \sum_{i=1}^N \log(1 + e^{\xi_i}), \end{aligned} \quad (4.42)$$

where $\boldsymbol{\xi} = (\xi_1, \dots, \xi_N)^t$ is a set of nonnegative parameters to be estimated, $\boldsymbol{\Lambda} = \text{diag}(\lambda(\xi_1), \dots, \lambda(\xi_N))$, with $\lambda(\xi) = \frac{1}{2\xi} \left(\frac{1}{1+e^{-\xi}} - \frac{1}{2} \right)$, and $\mathbf{1}$ is a column vector with all components equal to 1.

The joint distribution in Equation (4.40) can then be lower bounded as:

$$p(\mathbf{y}, \mathbf{f}, \Omega) \geq \mathbf{M}(\mathbf{y}, \mathbf{f}, \Omega, \boldsymbol{\xi}) = \mathbf{H}(\mathbf{y}, \mathbf{f}, \boldsymbol{\xi})p(\mathbf{f}|\Omega)p(\Omega), \quad (4.43)$$

which produces:

$$\begin{aligned} & \int q(\mathbf{f})q(\Omega) \ln \frac{q(\mathbf{f})q(\Omega)}{p(\mathbf{y}, \mathbf{f}, \Omega)} d\mathbf{f}d\Omega \\ & \leq \min_{\boldsymbol{\xi} \geq \mathbf{0}} \int q(\mathbf{f})q(\Omega) \log \frac{q(\mathbf{f})q(\Omega)}{\mathbf{M}(\mathbf{y}, \mathbf{f}, \Omega, \boldsymbol{\xi})} d\mathbf{f}. \end{aligned} \quad (4.44)$$

Let $\boldsymbol{\xi}^k$ be the current value of $\boldsymbol{\xi}$, then:

$$\begin{aligned} & \min_{q(\mathbf{f})} \int q(\mathbf{f}) \log \frac{q(\mathbf{f})}{\mathbf{M}(\mathbf{y}, \mathbf{f}, \Omega, \boldsymbol{\xi}^k)} d\mathbf{f} = \text{const} \\ & - \ln \mathcal{N}\left(\frac{1}{2}(\mathbf{y} - \frac{1}{2}\mathbf{1})|\mathbf{0}, \frac{1}{2}\boldsymbol{\Lambda}^k + \boldsymbol{\Lambda}^k \mathbf{K}_\Omega \boldsymbol{\Lambda}^k\right) \end{aligned} \quad (4.45)$$

which produces Ω^k , the current estimate of the vector where $\hat{q}(\Omega)$ is degenerate:

$$\Omega^k = \arg \max_{\Omega} \mathcal{N}\left(\frac{1}{2}(\mathbf{y} - \frac{1}{2}\mathbf{1})|\mathbf{0}, \frac{1}{2}\boldsymbol{\Lambda}^k + \boldsymbol{\Lambda}^k \mathbf{K}_\Omega \boldsymbol{\Lambda}^k\right). \quad (4.46)$$

Fixing $\boldsymbol{\xi}$ to $\boldsymbol{\xi}^k$ and Ω to Ω^k we obtain:

$$q^k(\mathbf{f}) = \mathcal{N}(\langle \mathbf{f} \rangle_k, \boldsymbol{\Sigma}_k) \quad (4.47)$$

where $\langle \mathbf{f} \rangle_k = \boldsymbol{\Sigma}_k(\mathbf{y} - \frac{1}{2}\mathbf{1})$ and $\boldsymbol{\Sigma}_k = (\mathbf{K}_{\Omega^k}^{-1} + 2\boldsymbol{\Lambda}^k)^{-1}$. Finally, solving:

$$\boldsymbol{\xi}^{k+1} = \max_{\boldsymbol{\xi}} \langle \ln \mathbf{M}(\mathbf{y}, \mathbf{f}, \Omega^k, \boldsymbol{\xi}) \rangle_{q^k(\mathbf{f})} \quad (4.48)$$

we obtain:

$$\xi_i^{k+1} = \sqrt{\langle f_i \rangle^2 + \boldsymbol{\Sigma}_k(i, i)}. \quad (4.49)$$

In Algorithm 4.4 the estimation procedure is summarised.

Given a new patch \mathbf{x}_* we would like to calculate:

Algorithm 4.4: Learning algorithm using GPCs

Require: Initial $\boldsymbol{\xi}^1 = \mathbf{1}$ and $k = 1$.

- 1: **repeat**
 - 2: Calculate Ω^k using $\boldsymbol{\xi}^k$ in Equation (4.46).
 - 3: Calculate $q^k(\mathbf{f})$ using Ω^k and $\boldsymbol{\xi}^k$ in Equation (4.47) .
 - 4: Calculate $\boldsymbol{\xi}^{k+1}$ using $q^k(\mathbf{f})$ and Ω^k in Equation (4.49).
 - 5: $k = k + 1$
 - 6: **until** convergence
-

$$\begin{aligned} p(y_*|\mathbf{y}) &= \int_{f_*} p(y_*|f_*) \left(\int_{\mathbf{f}} p(f_*|\mathbf{f})p(\mathbf{f}|\mathbf{y})d\mathbf{f} \right) df_* \\ &\approx \int_{f_*} p(y_*|f_*) \left(\int_{\mathbf{f}} p(f_*|\mathbf{f})\hat{q}(\mathbf{f})d\mathbf{f} \right) df_*, \end{aligned} \quad (4.50)$$

where the approximated posterior $\hat{q}(\mathbf{f})$ has been calculated, at convergence, in Algorithm 4.4, and $p(y_*|f_*)$ is given by Equation (4.39). The conditional distribution $p(f_*|\mathbf{f})$ can be calculated from Equation (4.38) as:

$$p(f_*|\mathbf{f}) = \mathcal{N}(f_*|\mathbf{h}^t\mathbf{K}^{-1}\mathbf{f}, c - \mathbf{h}^t\mathbf{K}^{-1}\mathbf{h}), \quad (4.51)$$

with $\mathbf{h} = (\mathbf{k}(\mathbf{x}_*, \mathbf{x}_1), \dots, \mathbf{k}(\mathbf{x}_*, \mathbf{x}_N))^t$, $c = \mathbf{k}(\mathbf{x}_*, \mathbf{x}_*)$.

The probability that patch \mathbf{x}_* belongs to class ‘‘pathological’’ can be written then as:

$$p(y_* = 1|\mathbf{y}) \approx \int \frac{1}{1 + e^{-f_*}} \mathcal{N}(f_*|m(f_*), v^2(f_*))df_*, \quad (4.52)$$

with $m(f_*) = \mathbf{h}^t\mathbf{K}^{-1}\langle\mathbf{f}\rangle$ and $v^2(f_*) = c - \mathbf{h}^t(\mathbf{K} + (2\boldsymbol{\Lambda})^{-1})^{-1}\mathbf{h}$. The integral in Equation (4.52) is approximated as in (Bishop, 2010), to obtain:

$$p(y_* = 1|\mathbf{y}) \approx \left(1 + \exp \left\{ -m(f_*)(1 + \frac{1}{8}\pi v^2(f_*))^{-1/2} \right\} \right)^{-1}. \quad (4.53)$$

Deep Learning approach for retinal damage detection

This chapter explains the basics of the deep learning paradigm. In particular, the theoretical framework around convolutional neural networks is detailed. A novel approach to detect abnormal retinal tissue, fine-tuning well-known CNN architectures, is also presented.

The information contained in this chapter and some of its corresponding results (see Section 6) was part of the work developed at the *Centre de Visió per Computador (CVC)* belonging to the *Universitat Autònoma de Barcelona* under the Grant EEBB-2017-12567.

Contents

5.1	Introduction	97
5.2	CNN architecture	100
5.2.1	Convolutional layer	101
5.2.2	Activation layer	103
5.2.3	Pooling layer	104
5.2.4	Fully connected layer	105
5.3	Training and test stages	107
5.4	Improving CNNs performance	108
5.5	Transfer learning & state of the art in CNN models	111
5.6	CNNs applied to retinal images	115

5.1 Introduction

During a representative part of this PhD thesis, several image descriptors (detailed at Section 4.1) have been studied with the objective of transforming the visual information perceived in an image into numerical data (a.k.a. hand-crafted feature extraction). Making use of the extracted features, classification algorithms are applied in order to detect retinal damage manifested in the fundus images and caused by ophthalmic pathologies. The success and reliability of a system based on these methods is conditioned to the descriptor used to extract the most relevant image information. The aforementioned descriptors can be considered the universal image feature extraction methods that are applicable to most computer vision tasks. They are not dedicated to solve a specific problem and are highly dependent on the optimal selection of its parameters for each application. The experience of the researcher managing this kind of algorithms and the help of medical specialists are key in the selection of the most appropriate descriptors that best define what we want to detect or classify. This stage is known as feature engineering and it is characterised by a subjective character. This fact motivates the study of a novel technique able to automatically extract the most relevant features of an image without the need of determining the optimal descriptors by the researchers. Currently, deep learning is the focus of most computer vision researches and in this chapter, a novel approach to detect retinal pathologies using convolutional neural networks (CNNs) is explored.

The recognition task is one of the first skills naturally developed without difficulty by a person since the birth moment. This fact is propitiated by the human visual cortex. When we see an image or simply the environment that surrounds us, we are immediately able to characterise the scene, as well as unconsciously identify and classify each of the objects that compose it. These recognition and characterisation tasks are not easy to be performed by machines. However, in the last years CNNs emerged as a powerful tool in the computer vision field to make the automatic identification of well-defined tasks possible. Convolutional neural networks are inspired by the aforementioned human visual cortex. It is characterised by cell regions sensitive to specific areas of the visual field. As it was demonstrated in (Hubel and Wiesel, 1962), different cell regions react to distinct visual stimuli such as vertical and horizontal edges, blobs, colours, etc. Hubel and Wiesel also discovered that neurons related to the visual cortex are organised by a columnar architecture. The combination of different responses to a given stimulus from different cell regions makes possible the visual perception.

Following the principles detailed above, LeNet model (Lecun et al., 1995) emerged as the first convolutional neural network architecture for automatic recognition tasks. A stack of convolutions and pooling operations (i.e. hidden layers) were utilised for feature extraction and spatial sub-sampling purposes, respectively. In 2012, these ideas were refined into the AlexNet architecture (Krizhevsky et al., 2012), in which, convolution operations were repeated multiple times in-between pooling operations with the aim of learning richer features at every spatial scale. The following years, the efforts of the researchers focused on increasing the depth of this style of network. These efforts have been highly encouraged by the yearly ImageNet Large Scale Visual Recognition Challenge (ILSVRC)¹, in which, the objective is to propose new CNN architectures able to reduce the classification error of a large labelled dataset composed by natural images (a.k.a. ImageNet² dataset).

The main power of the CNN models relies on their ability to extract highly discriminating features at multiple levels of abstraction. These network architectures are designed to extract automatic features from raw image pixels. However, training a CNN from scratch is not a trivial task. The training stage of CNN models requires a huge amount of labelled data and high-performance computing systems, such as GPUs or large scale distributed clusters (Dean et al., 2012). The first of both premises is well-known as the bottleneck of most medical applications due to the tedious and time-consuming image annotation task. Fine-tuning technique was born to allow the creation of predictive models, based on CNN, when only reduced datasets are available. Fine-tuning procedure consists in re-training network architectures, which have been previously trained using a large labelled dataset (e.g. ImageNet), using our reduced set of images (see Section 5.5 for more details).

Recent works show how state-of-the-art CNN models, pre-trained using general computer vision datasets, are useful for creating decision models in the field of medical imaging. In (Shin et al., 2016), a deeper evaluation of the fine-tuning technique is performed in thoraco-abdominal lymph node detection and interstitial lung disease classification. An exhaustive comparison between training a CNN from scratch (when large medical datasets are available) and re-training deep convolutional neural networks is established in (Tajbakhsh et al., 2016) by means of four different medical imaging applications. Classification, detection and segmentation tasks are carried out in radiologic, cardiac and gastroenterologic image datasets to perform the comparison. Kumar et al. (2017) propose the ensembling of several fine-tuned CNN models to

¹www.image-net.org/challenges/LSVRC

²www.image-net.org/download-images

discriminate among thirty different medical image modalities (using the ImageCLEF2016³(García Seco de Herrera et al., 2016) public dataset) taking into advantage the prior knowledge of the pre-trained models in natural image datasets.

Regarding to retinal images and CNNs, the first approach based on deep learning techniques was motivated by the diabetic retinopathy detection challenge promoted by Kaggle⁴. Moreover, the work presented by Gulshan et al. (2016) propitiated an important milestone in the field of DR identification using deep neural networks. In particular, they trained a deep convolutional neural network using a retrospective development dataset of 128175 retinal images graded by 3 to 7 experts. However, the solutions proposed using this image-labelled datasets require complex neural network structures and can be seen as black boxes because no information is provided about the retinal areas which cause the eye disease (Pratt et al., 2016). In this way, Wang and Yang (2017) proposed a novel CNN architecture and computed a regression activation map to visualize the possible pathological areas in the fundus image. It is worth to note that in several activation maps presented as exemples some of the suspicious areas do not present any of the verified signs of diabetic retinopathy.

In the particular case of the topic of this thesis, a CNN must be able to discriminate between healthy retinal regions and regions containing signs of the eye diseases under study, simulating the subconsciously process carried out in the human visual cortex. Different attempts have been performed in the literature in this sense. For example, in (Prentašić and Lončarić, 2016) a home-made convolutional neural network calculates the pixel probability of belonging to exudate or non-exudate classes. The authors propose a CNN architecture consisting of four convolutional layers and four max-pooling layers. The kernel size decreases as it is going deeper in the network from 6×6 to 2×2 . After the last pooling layer, two fully-connected layers make possible the classification in two classes (the two neurons of the last fully-connected layer). This CNN is trained from scratch for ten hours using patch-images extracted from a subset of DRiDB database (Prentašić et al., 2013). A sensitivity, positive predictive and F-Score values of 0.77 are reported from the testing subset. Another work which aims to automatically detect bright pathological tissue in fundus images was proposed by Yu et al. (2017). In this paper, a convolutional neural network with convolutional blocks of 3×3 is trained by using patches of size 64×64 pix. Promigin results are reported using *E-OPHTHA* database.

³www.imageclef.org/2016/medical

⁴www.kaggle.com/c/diabetic-retinopathy-detection

In particular, an accuracy, sensitivity and specificity of 91%, 88% and 96% respectively. The automatic detection of microaneurysms by deep neural nets is explored in (Haloi, 2015). In this work, a five-layer architecture composed by three convolutional layers each followed by max-pooling is presented. The convolutional layers (with 5×5 kernels) are followed by one fully connected hidden layer and the softmax classification layer is fully connected with two neurons for microaneurysms and non-microaneurysms discrimination. Notice that dropout technique (Srivastava et al., 2014) is used to avoid overfitting. From the pixel-level ground truth annotated by experts, an imbalanced dataset of patches is constructed using three public databases. Comparable results with the classical approaches of the state of the art are reported. In (Grinsven et al., 2016) a method to improve and speed-up the CNN training for medical image analysis tasks is proposed. The authors focus on the detection of hemorrhages in color fundus images. In this work, a neural network composed by five convolutional layers of 32 filters of 3×3 is trained using patches of 41×41 pixels. They obtained areas under the receiver operating characteristic curve (AUC) of 0.894 and 0.972 on Kaggle and Messidor datasets. Another important research work in this direction was presented by Yang et al. (2017). In this paper, the authors propose an automatic diabetic retinopathy analysis algorithm based on two-stage deep convolutional neural networks (DCNN). The method can point out the location and type of lesions in the fundus images, as well as giving the severity grades of DR. Two different architectures compose the local and global neural networks. Both of them uses 3×3 filter kernels but the global network is deeper than the local one. An AUC value of 0.9687 suggests promising results for the identification of the different lesions manifested in the retinal background of a patient that suffers DR.

A different approach is followed in this thesis. In particular, well-known convolutional neural architectures are fine-tuned to detect and localise bright retinal lesions. The rest of the chapter details the most important concepts related to convolutional neural networks and describes the proposed procedure to discriminate between pathological and healthy retinal tissue.

5.2 CNN architecture

A computer processes an RGB image as a three-channel array whose elements describe an intensity value (in the range from 0 to 255) for each of the colour components. A convolutional network requires an input image of dimensions $N \times M \times 3$ and the output provides the probability that the input image belongs to a certain class or category. Apart of the input and output, CNNs

are composed by multiple linear and non-linear hidden layers: convolutional layers, pooling layers, fully connected layers and normalisation layers. The aim of the designed architecture is to extract the visual information from the image. In the following subsections, the principles of the different layers that compose a CNN architecture are detailed.

5.2.1 Convolutional layer

The convolutional layer gives the name to the CNN and it is the basis of deep learning applied to digital imaging (Nielsen, 2015). The convolutional layer is characterised by the use of Z filters, also known as kernels, of dimensions $K \times K \times 3$ (in the case of RGB images). Each filter of the convolutional layer is moving along the image as a sliding window with a specific step or stride. The set of pixels involved within the sliding window in each step is known as receptive field. The mathematical operation between the kernel and the receptive field is the convolution (Figure 5.1).

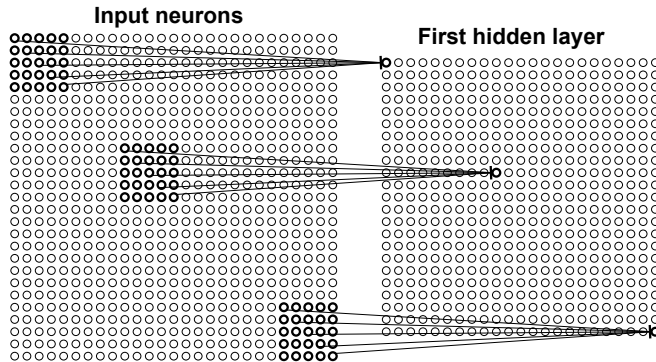


Figure 5.1: An input image and its corresponding activation map obtained after a convolutional layer (i.e. a hidden layer). Highlighted pixels correspond to the receptive field in different positions of the kernel. Note that a 2D representation is shown to provide a better understanding but the image and the filter have three dimensions.

Each of the Z filters in a convolutional layer is defined by a series of weights (involved in the convolution operation). These weights are the key element in the automatic learning process as it will be detailed later. The result of applying each filter over all the pixels of the input image is a feature or activation map. As can be seen in Figure 5.1, the dimensions of the input image are reduced by the convolutional layer with the consequent loss of

information. This dimensionality reduction will be determined by the size of the filter and it can be alleviated by a previous zero padding. If we want to preserve the dimensions of an image $N \times M \times 3$, a zero padding image of dimensions $N + (K - 1) \times M + (K - 1) \times 3$ (where K refers to the filter height and width) is computed previously to perform the convolutions (Figure 5.2). Thus, the output of the convolutional layer is an activation map of dimensions $N \times M \times Z$ in which the response of the input pixels to the filter is registered (Figure 5.3).

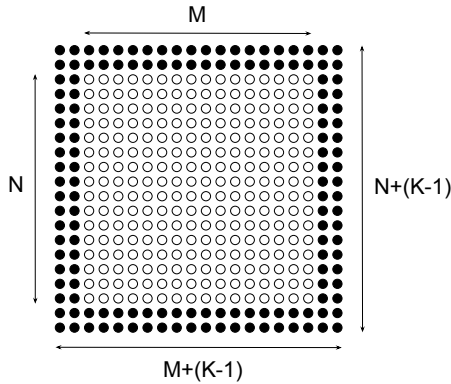


Figure 5.2: Zero padding representation. Note that black circles corresponds to pixels of zero value.

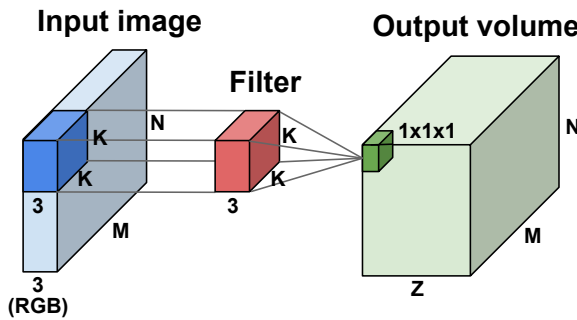


Figure 5.3: Output volume of a convolutional layer when the input is a RGB image.

The objective of the first convolutional layers is to highlight edges, curves and particular shapes of the image using the appropriate filter weights in each case. As the reader can imagine, in order to classify an image according to a certain category or to predict if an object appears in an image, the network must extract features of a higher level of abstraction. Consequently, a deep CNN architecture must be defined, in other words, the network has to be composed of several convolutional blocks and the output of a convolutional layer will be the input of the following layer. This depth character of the network is essential in the automatic image description process because the extracted features describe more precisely what is desired to classify as you go deeper throughout the network. After a training process, the filters of the last layers of the network will register a high activation to objects for a certain colour, texture or shape (Zeiler and Fergus, 2014).

5.2.2 Activation layer

A CNN architecture is formed by other hidden layers interspersed between the convolutional layers. The general purpose of these new layers is to provide non-linearities and preserve dimensionality with the aim of improving the robustness of the network. Note that these aspects help to approximate very complex decision boundaries.

After each convolutional layer, a non-linear layer (or activation layer) is applied by convention. The purpose of this layer is to introduce non-linearities to a system, since the only operations performed in the convolutional layers have been multiplications and sums. In the past, nonlinear functions such as the hyperbolic tangent (Figure 5.4(a)) or sigmoid (Figure 5.4(b)) were used until Geoffrey Hinton presented the rectified linear unit (ReLU) layer (Nair and Hinton, 2010)(Figure 5.4(c)). This layer enables a faster training stage of the neural network due to the computational efficiency achieved with this function without altering the accuracy in classification. The ReLU layer applies the function $f(x) = \max(0, x)$ to all values of the input volume (output of the convolutional layer), i.e. this layer changes all negative activations to zero.

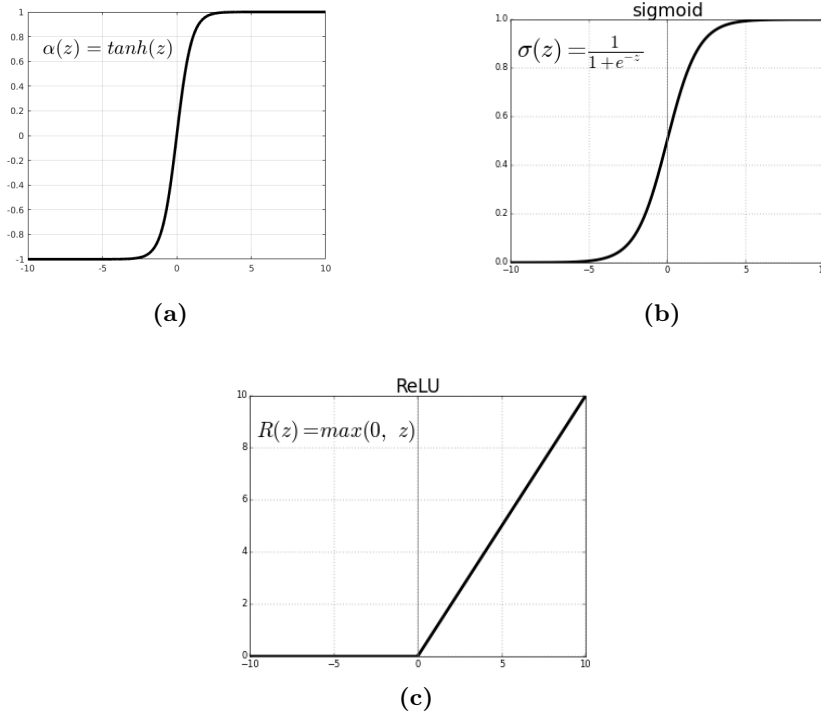


Figure 5.4: Activation functions. (a) hyperbolic tangent, (b) sigmoid and (c) rectified linear unit.

5.2.3 Pooling layer

After the activation layers discussed above, a traditional CNN architecture includes a pooling layer. This layer performs a downsampling operation along the spatial dimensions (width, height). Several methods to reduce the dimensionality of the input data exist. The most common procedures are to keep the maximum value of a specific neighbourhood (maxpooling), the average, (average pooling) or the result of the L2-norm computation (L2-norm pooling). Typical parameters in these layers are filter size $K=2$ and stride=2. Figure 5.5 shows an example of this process using the maxpooling method. The intuitive reason for including this type of layer is because the absolute location of the features does not provide relevant information to the system and the location with respect to other features is computed instead. For this reason, these layers reduce progressively the spatial size, decreasing the weights of the filters by 75% with the consequent reduction of the computational cost.

Another reason to use these layers is their ability to help in the prevention of overfitting.

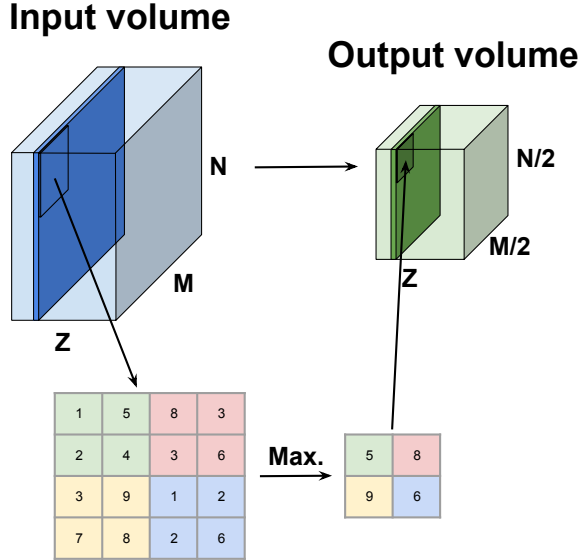


Figure 5.5: Example of pooling technique using the maximum operation (maxpooling). Typical parameters are selected ($K=2$ and $\text{stride}=2$).

5.2.4 Fully connected layer

The bottom layers of a CNN architecture are responsible for converting a volume of data (the output of a convolutional or pooling layer) into a vector containing the same number of elements as the number of classes. This layer is known as fully connected and it applies the 1D convolution to the flattened volume of the last 2D filter, reducing the number of activations to the number of classes (N_c). Subsequently, using an appropriate activation function it is possible to map the activations into the final probabilities. In Figure 5.6, a flattened volume of dimensions $z \times m \times n$ is convolved by a $1 \times (z * m * n)$ dimensional filter resulting an output vector of N_c activations. After the convolution, the softmax activation function is used to obtain the class probability. This function squashes the outputs of each unit to be between zero and one:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{n=1}^{N_c} e^{z_n}} \quad (5.1)$$

where N_c is the total number of classes. Softmax function also divides each output such that the total sum of the outputs is equal to one:

$$\sum \sigma(z)_j = 1 \tag{5.2}$$

Finally, the output of the softmax function is equivalent to a categorical probability distribution, which gives the probability that any of the classes are true.

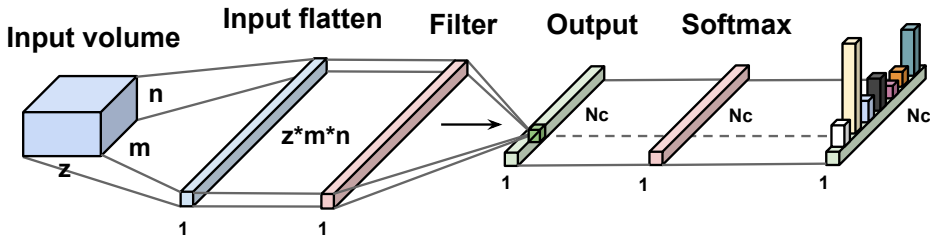


Figure 5.6: Representation of the bottom layers in a CNN architecture. Fully-connected layers and the softmax activation function allow the classification of the input image taking into account the extracted information along the network.

After analyzing the different layers that compose a CNN, Figure 5.7 shows a typical and basic convolutional network architecture (LeCun et al., 1998) containing the layers explained at this section. It is worth noting that normalisation layers will be explained in Section 5.4 due to some previous concepts are required for the understanding.

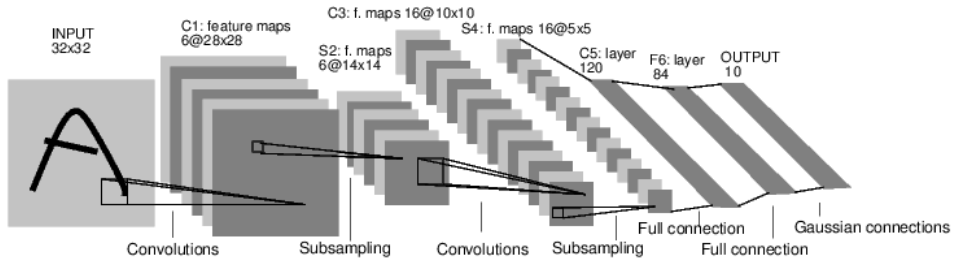


Figure 5.7: Typical and basic convolutional network architecture.⁵

⁵Image provided by LeCun et al. (1998)

5.3 Training and test stages

At the moment a human is born, his mind is totally empty, in other words, at that moment, a person is unable to know what, for example, a dog or a bird is. The same thing happens in the CNNs, when an architecture of a convolutional network is designed and an instance is generated, the network does not have any type of knowledge, i.e. the weights of its filters are totally randomized. Low-level filters do not know what an edge or a curve is, and high-level filters do not have knowledge about colours or particular objects of an image. As the human grows, he acquires knowledge thanks to parents and teachers. A common way of teaching in low levels is to show an image containing the target to be learned and offering the answer. For example, I can show an image containing a yorkshire terrier to my baby and say to him ‘This is a dog’. This idea of showing an image and offer the answer (via a tag, for example) is the basis of the training stage of a neural network. Note that this process requires a very large training set of images with its corresponding label of group or class.

The training process of a neural network can be divided into four steps: forward propagation, loss function, backpropagation and weight updating. In the first iteration of the forward propagation stage, an image enters to the neural network and goes through the network crossing all its layers. At the outputs of the CNN (the outputs of the fully connected layer) we will obtain that the input image belongs to every class with the same probability. A CNN with random weights in the filters of their convolutional layers cannot assign a correct prediction to the input image in the first iteration, but a tag containing the solution (ground truth) is available and provided to the system. So, the CNN learns by updating the weights of its filters according to the loss function and the backpropagation step. The loss function can be defined in various ways but commonly it is defined using the mean squared error:

$$E_{total} = \sum \frac{1}{2}(target - output)^2 \quad (5.3)$$

At each iteration of the training stage, the objective is to minimise the loss function. From a mathematical point of view, we have to solve a minimisation problem in which the aim is to find the optimal weights for the filters of the convolutional layers. To perform this, the contribution of each weight of the filter to the losses registered by the error function must be determined, as well as the optimal way to adjust these weights must be established so that the loss function presents a decreasing profile. This step is known as backpropagation

and is based on the computation of the derivative of the loss function with respect to the weights of the filters and their posterior updating. Note that, weights vary in the opposite direction to the gradient and they are updated by means of:

$$w = w_i - \eta \frac{dL}{dW} E_{total} \quad (5.4)$$

where w_i are the initial weights, w are the updated weights and η is the learning rate. The learning rate is a parameter of the network design. High learning rate values propitiate large jumps in the updating process. This fact may take less time for the model to converge on an optimal set of weights. However, excessive learning rate values could result in too large jumps and never reach the global minimum of the error function, where weights are optimal.

It should be noticed that the four steps of the training process detailed above are part of the same iteration. This process is repeated for each subset of the training images (also known as batch). In addition, the total number of batches, determined by the number of training images and the batch size, is evaluated a specific number of iterations (or epochs). When the last batch of training samples is processed in the last epoch and the weights are updated, the stage of model creation is concluded. As last step and to verify how well the generated model predicts, it must be evaluated using the test set. It is important to guarantee that any image belonging to the test set was not used in the training stage.

5.4 Improving CNNs performance

Another technique related with the amount of images available in our dataset and currently used in the creation of CNN models is known as data augmentation (Wong et al., 2016). This technique consists of artificially enlarging small datasets by means of geometric and intensity transformations. If a particular pixel of an image A stacked as a vector is shifted one position resulting in an image B , for humans A and B are practically identical, but for a computer these images are totally different. This fact is exploited by the technique of data augmentation. By means of rotation, translation, scaling, random cropping and colour transformations among others, new artificial samples are generated. Of course, the synthetic images automatically created belongs to the same class as the original image from which they were generated.

By applying this method it is possible to create more robust models and avoid overfitting enlarging the training set.

Deep neural networks are characterised by a large number of parameters and overfitting can be caused by the high amount of degrees of freedom if not sufficient training data are available. In addition, these deep networks are slow to use, making it difficult to deal with overfitting. To address the overfitting problem, Srivastava et al. (2014) proposed the dropout technique. The key idea is to randomly drop units or neurons (along with their connections) from the neural network during the training (Figure 5.8).

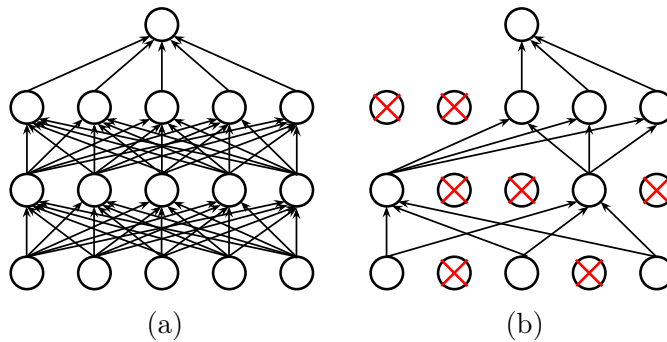


Figure 5.8: Dropout technique: (a) A standard neural network with two hidden layers and (b) An example of a thinned net produced by applying dropout to the network on the left. Crossed units have been dropped.

This technique prevents units from co-adapting too much to the training instances, reducing the degrees of freedom to be fixed. During the training stage, dropout samples from an exponential number of different “thinned” networks. At test time, it is easy to approximate the effect of averaging the predictions of all these thinned networks by simply using a single unthinned network that has smaller weights. This significantly reduces overfitting and gives major improvements over other regularization methods (Srivastava et al., 2014).

Last technique to improve the classification performance using CNNs is Batch Normalisation (BN). The training stage of deep neural networks is difficult due to the distribution of each layer inputs changes during training, because the parameters of previous layers change. This fact slows down the training by requiring lower learning rates and careful parameter initialization. This phenomenon is known as *internal covariate shift* and Ioffe and Szegedy (2015) presented batch normalisation to solve it. The authors proposed a

normalisation of each *training mini-batch* as a part of the model architecture. The idea is to decorrelate, centre and scale the output of one layer to be the input of the next layer. In particular, the normalisation is performed before the activation function of the next layer (Figure 5.9), and it is given by:

$$BN(x) = \alpha \hat{x} + \beta \tag{5.5}$$

$$\hat{x} = \frac{x - avg_{batch}(x)}{std_{batch}(x) + \epsilon} \tag{5.6}$$

where α and β are learnable parameters (two degrees of freedom per neuron) and ϵ guarantees the numerical stability of the operation. From Equations (5.5) and (5.6) can be extracted a key point of BN. If from the training stage the optimal parameters result $\alpha = std_{batch}(x)$ and $\beta = avg_{batch}(x)$, the normalised data will be identical to the input data (i.e. $BN(x) = x$).

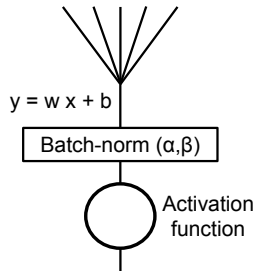


Figure 5.9: Batch normalisation stage is performed before applying the activation function as an additional layer.

Batch Normalisation allows the use of much higher learning rates propitiating the time reduction in the training stage and being less careful in inicialization. It also acts as regularizer, in some cases eliminating the need of dropout technique. It is worth noting that BN can be instantiated as a regular layer in the architecture design.

5.5 Transfer learning & state of the art in CNN models

Convolutional neural networks are becoming more and more popular among some of today's most powerful companies such as Google, Facebook or Amazon. The amount of data generated daily by this kind of companies allows them to create accurate, robust and reliable prediction models. However, in the problem dealt in this thesis in particular and in the field of medical imaging in general, the amount of data is the biggest limitation when applying CNNs. To overcome this problem, a powerful technique called transfer learning has emerged (Bengio, 2012; Yosinski et al., 2014). Transfer learning is the process by which, from a pre-trained network model, a specific model can be created to solve a particular classification problem by using a reduced amount of data. This method takes into advantage the wide knowledge acquired by the original model trained with large image databases in order to reduce the amount of necessary data to create specific solutions. ImageNet is a set of data that contains around 14 million images belonging to about 1000 classes or categories. The most popular state-of-the-art network architectures have been trained using ImageNet dataset and the resulting filter weights of each network are publicly available.

Fine-tuning a CNN is a process based on transfer learning, which basically consists in two phases: 1) the weight initialisation of the convolutional layers using the weights obtained from pre-training the same CNN architecture using ImageNet database and 2) the modification of the top-model layers and the classification function responsible to discriminate among classes. All layers are initialised to the pre-trained weights, except the last FC layer, whose neurons depend on the number of classes of the new task. This fact makes possible the "knowledge" transference from natural images to each particular application.

After the weight initialisation, the CNN can be fine-tuned in a layer-wise manner. It means, start fine-tuning only the last layer, then include layers until fine-tuned all the layers in the network. Depending on the number of fine-tuned layers, this procedure can be called "shallow-tuning" if only the last few layers are fine-tuned and "deep-tuning" if all the layers are fine-tuned (Tajbakhsh et al., 2016).

In summary, this technique allows to create a robust and effective model for a specific problem using reduced datasets and avoiding the need of having millions of images to train a CNN from scratch. In this thesis, five state-of-the-art CNN architectures (pre-trained in ImageNet) are fine-tuned in order to create a robust model, based on deep learning techniques, able to discriminate

between healthy and damaged retinal tissues (see Section 6.2.5). In the following lines these architectures are detailed.

VGG16 and VGG19

After the successful convNet proposed by Krizhevsky et al. (2012) in the ILSVRC-2012, a great number of attempts were proposed to improve the AlexNet architecture in order to achieve better accuracy (Sermanet et al., 2013; Zeiler and Fergus, 2014). In this way, Simonyan and Zisserman (2014) analysed an important aspect of CNN architecture design: its depth. Rather than using relatively large receptive fields in the first convolutional layers (e.g. 11×11 with stride 4 in (Krizhevsky et al., 2012), or 7×7 with stride 2 in (Sermanet et al., 2013; Zeiler and Fergus, 2014)), they used very small 3×3 receptive fields throughout the whole net, which are convolved with the input at every pixel (stride 1). Stacking two or three 3×3 they obtain effective receptive fields of 5×5 and 7×7 , respectively.

Simonyan and Zisserman (2014) discovered two main advantages of using stacks of convNets with small receptive fields instead of a single layer with bigger kernels. On one hand, they incorporate three non-linear rectification layers instead a single one, which makes the decision function more discriminating. On the other hand, the number of kernel weights notably decreases. The authors evaluate the performance of six different CNN architectures with the same basis. 3×3 convolutional filters stacked on top of each other in increasing depth and reducing the volume size by max pooling. At the end, three fully-connected (FC) layers, two of 4096 and one of 1000 neurons (one per class) are followed by a softmax classifier (Figure 5.10). The difference among the six proposed architectures was basically the depth of each network. The obtained classification results, using ImageNet dataset, demonstrated that network depth is a key factor for the classification accuracy. VGG16 and VGG19 were the basis of their ImageNet 2014 submission, with which they secured the first and the second places in the localisation and classification tracks.

Microsoft ResNet

Microsoft ResNet is the CNN architecture proposed by the Microsoft Research Team of Asia (MSRA) to the ImageNet 2015 competition. With this architecture, the MSRA not only won the competition, but also set new records in classification, detection, and localisation.

⁶Image provided by the Computer Science University of Toronto (www.cs.toronto.edu)

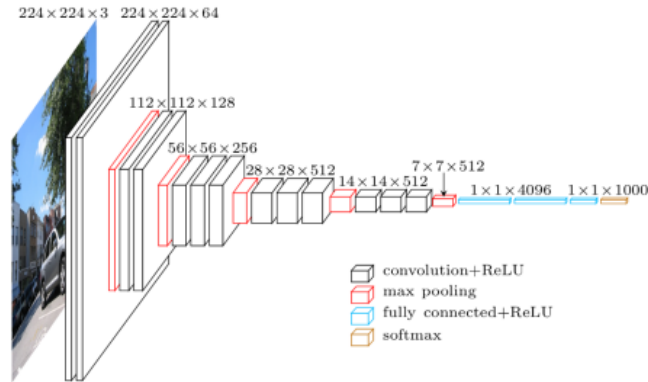


Figure 5.10: Architecture of VGG16 network.⁶

ResNet architecture is based on a novel concept: a residual learning block (Figure 5.11). He et al. (2016) proposed a residual mapping in order to solve the degradation problem exposed in deep networks. In particular, when network depth increases, accuracy gets saturated and then degrades rapidly.

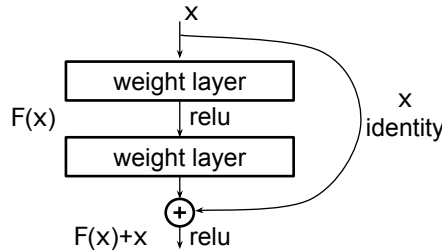


Figure 5.11: Illustration of the residual block proposed by He et al. (2016), in which x is the input and $F(x) + x$ the output.

The authors demonstrated that stacked convolutional layers approximate better a residual function $F(x) := H(x) - x$ than directly the desired underlying mapping $H(x)$. This theoretical approach was successfully applied using shortcut “connections” giving place to five residual networks of 18, 34, 50, 101 and 152 layers. They are composed by convolutional layers of 3×3 receptive field and a stride of 2. The networks end with a global average pooling layer and a 1000-way fully-connected layer with the softmax activation function. ResNets are characterised by fewer filters and lower complexity than VGG

models. This “exotic architecture” notably reduced the classification error in the ImageNet dataset (top 5-error) achieved by sequential networks.

GoogLeNet

The ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14) was highlighted by the novel CNN module introduced by Szegedy et al. (2015): the Inception module. This new kind of “layer” made possible the first place in the object detection track. Inception module behaves as a multi-level feature extractor due to the concatenation of filters of different sizes and dimensions into a single new filter.

Based on the Inception module (Figure 5.12), Szegedy et al. (2015) developed an architecture significantly more complex and deeper than all previous CNN architectures. In the first version, GoogLeNet architecture is composed by Inception filters that combine 1×1 , 3×3 and 5×5 convolutions in each module. The network is 22 layers deep when counting only layers with parameters (or 27 layers if we also count pooling layers). Instead of using fully-connected layers before the classifier, the authors used an average pooling layer with the dropout technique (remaining the 40% of the neurons). The initial architecture of this CNN was called GoogLeNet or InceptionV1. However, after some improvements, InceptionV4 is currently available.

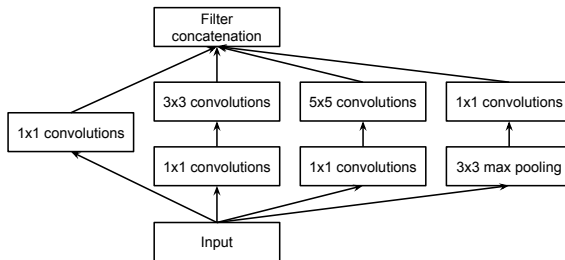


Figure 5.12: Representation of the Inception module proposed by Szegedy et al. (2015).

Xception

Xception stands for Extreme Inception and is the name of the architecture proposed by Chollet (2016). The fundamental hypothesis behind this architecture is that spatial correlations (width and height of the convolutional filters) and cross-channel correlations (third dimension of the convolutional filters) are sufficiently decoupled to be mapped separately. Based on this premise, Chollet (2016) proposed a convolution performed independently over

each channel of an input, followed by a pointwise convolution (i.e. a 1×1 convolution) projecting the channels output by depthwise convolution onto a new channel space. Xception is an extension of the Inception architecture, in which, the standard Inception modules are replaced by depthwise separable convolutions (Figure 5.13).

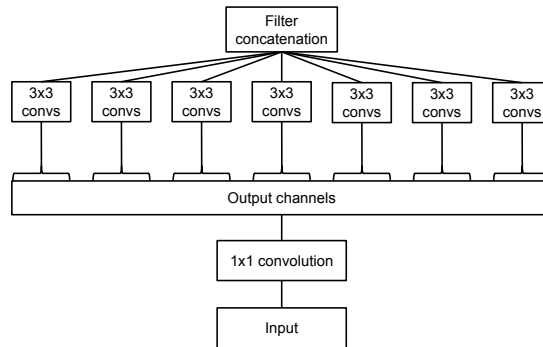


Figure 5.13: An “extreme” version of the Inception module (Xception) with one spatial convolution per output channel. Note that numbers 1×1 and 3×3 are the convolution block dimensions.

The Xception architecture has 36 convolutional layers of 3×3 receptive field and a stride of 2, forming the feature extraction base of the network. These layers are structured into 14 modules, all of which have linear residual connections around them, except for the first and last modules. Fully connected layers are optional depending on the application and logistic regression is used for classification. It is worth noting that a dropout layer (remaining the 50% of the neurons) before the logistic regression layer is included. Xception architecture slightly outperformed InceptionV3 on the ImageNet dataset. Since the Xception design has the same number of parameters as InceptionV3, the performance gains are not due to increase capacity but rather to a more efficient use of model parameters (Chollet, 2016).

5.6 CNNs applied to retinal images

As it was detailed in Sections 4.1.1 and 4.2.1 and can be observed in Figure 5.14, exudates, microaneurisms and haemorrhages are manifested in a few amount of pixels of the whole retinal image. For this reason, a patch-wise CNN approach is proposed with the aim of discriminating between healthy

and pathological regions or patches. This kind of analysis present several advantages related to the deep learning concepts exposed throughout this chapter. First of all, training and test dataset will be larger than image-based datasets because from a fundus image, thousand of patches are extracted (depending on the sliding window size and the overlapping). This fact is key for constructing robust models and avoid overfitting. Additionally, following this methodology, information about the spatial localisation of lesions is available. This information can be strategically used for further purposes.



Figure 5.14: Fundus image in which the imbalanced character of the extracted data can be easily appreciated.

Figure 5.15 shows an overview of the methodology used to fine-tune the most representative state-of-the-art CNN models detailed in Section 5.5.

Using a fundus database composed by N images, it must be divided in two datasets: training and test. The first dataset is responsible for training the neural network while the second allows to objectively evaluate the model performance. It is worth noting that any of the fundus images used for creating the model cannot be part of the testing dataset. To augment the robustness of our approach, the leave-one-out cross-validation technique is utilised to create both datasets. The N images of the database are divided in K folds or partitions. $K - 1$ of these partitions are used to fine-tune a CNN architecture while the remaining fold of images is employed in the test stage. This procedure is carried out K times, leaving a different partition each time for testing. When the image datasets are created, the following step is the patch extraction from all the images of each dataset. This process is performed by means of a sliding window of size (N_w) with overlap of $(\Delta x, \Delta y)$ over the image. The image containing the pixel-annotations by the specialists (ground-truth) is also looped by the sliding window and a “pathological” or “healthy” label is established to each extracted patch. When patch-datasets of

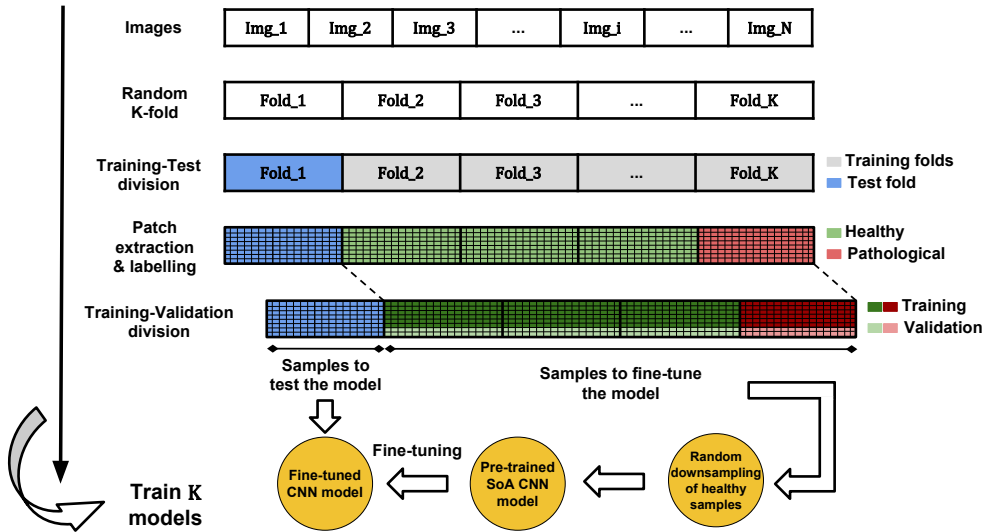


Figure 5.15: Pipeline followed in this thesis to fine-tune CNN models with the aim of discriminating between healthy and pathological retinal areas.

training and test are constructed, the next step consists in splitting the training dataset in two subsets: training and validation. The validation dataset is part of the training stage and it is very useful to find the optimal hyperparameters that define the network (detailed throughout the chapter). Thus, the patches belonging to the $K - 1$ training folds are divided into $X\%$ for training and $Y\%$ for validation, creating the final patch-database. It is important to remark the imbalanced character of the obtained datasets due to the minor region of the fundus images with damage. To solve this problem, we take all the pathological patches as positive class and the same amount of “healthy” instances randomly sampled and without repetition.

Results and discussion

This chapter presents the findings of this PhD thesis. Different experiments analyse the effect of the pre-processing methods, the contribution of each image descriptor and the behaviour of the different classification algorithms presented in this thesis. In addition, a comparison between the classical computer vision methods and the proposed approach based on deep learning techniques is established.

Contents

6.1	Material	121
6.2	Experiments	122
6.2.1	Effect of the pre-processing methods	123
6.2.2	Behaviour of texture, shape and roughness descriptors	126
6.2.3	Performance of machine learning algorithms	131
6.2.4	Global detection of bright and dark retinal lesions for automatic screening.	136
6.2.5	Pathological retinal tissue identification by CNNs	142

6.1 Material

Two public fundus databases were used to validate the methods proposed throughout this thesis. Next, a description of both fundus databases is detailed.

E-OPHTHA

E-OPHTHA (Decencière et al., 2013) is a database of fundus images especially designed for diabetic retinopathy screening¹. This public database is divided in two subsets depending on the lesion type: exudates and microaneurysms. These lesions are manually annotated by experts and the ground-truth is provided. The exudates subset (*E-OPHTHA_EX*) is composed by forty seven pathological images (*Pathological_EX*) and thirty five images with no lesion (*Healthy_EX*). *E-OPHTHA_MA* subset makes available 148 images with microaneurysms or small haemorrhages (*Pathological_MA*) and 233 healthy images (*Healthy_EX*). All the retinal images were acquired with the same field of view angle, i.e. 40°, and they present different spatial resolutions as it is reported in Table 6.1.

Image resolution (pix.)	1440 × 960	1504 × 1000	2048 × 1360	2544 × 1696
Pathological_EX	13	2	9	23
Pathological_MA	54	5	17	72
Healthy_EX	17	0	5	13
Healthy_MA	96	39	1	97

Table 6.1: Distribution of the *E-OPHTHA* images according to their resolution.

DIARETDB1

DIARETDB1 public database (Kauppi et al., 2007) consists of 89 colour fundus images, of which 84 contain at least mild nonproliferative signs of DR and five are considered as normal by medical experts². In particular, 41 fundus images show bright lesions (exudates) and 45 contain microaneurysms or haemorrhages. Images were captured using the same 50 degree field-of-view digital fundus camera and all images present a resolution of 1500 × 1152 pixels. The 89 images of this database were annotated by four experts. These annotations are provided as a probability map. The procedure proposed by the authors of *DIARETDB1* to generate the ground truth images from the probability map, is followed in this work. Each pixel marked as exudate by 3 of the 4 experts is considered as exudate pixel in the binary mask (Figure

¹www.adcis.net/en/Download-Third-Party/E-Ophtha.html

²www.it.lut.fi/project/imageret/diaretdb1

6.1). For the creation of the bright lesions masks, hard and soft exudates were taken into account. Analogously, to create the masks containing dark lesions, haemorrhages and microaneurysms annotations were put together.

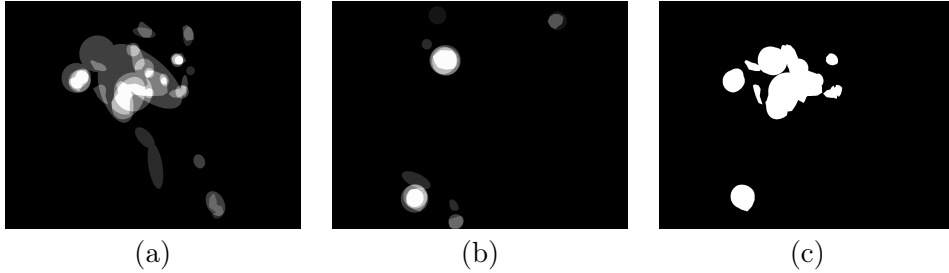


Figure 6.1: Ground truth mask composition from the annotations of the *image20.png* image by the four specialists. (a) Hard exudates annotations, (b) soft exudates annotations and (c) computed binary mask.

6.2 Experiments

In this section, the results obtained for six experiments are reported. The discrimination between healthy and damaged retinal tissues is the common objective of all the experiments. However, each experiment aims to study the behaviour of the different methods exposed throughout this PhD thesis. First of all, the effect of the pre-processing methods for retinal images presented in chapter 3 is quantified. In addition, the contribution of each image descriptor and the performance of the classification methods detailed in chapter 4 is studied. Moreover, a novel approach based on convolutional neural networks (see chapter 5) is explored for detecting and locating the pathological retinal tissue. Finally, the resulting models, embedded in an automatic home-made system for diabetic retinopathy diagnosis, are compared. It is worth noting that the experiments were validated using the public fundus databases explained in the previous section.

Regarding to the hardware specifications used for the experiments carried out during this PhD thesis, an Intel Core i7-4770S@3.10GHz processor with 16GB of RAM and Ubuntu 16.04 LTS as operating system was used. For the experiments 6.2.1, 6.2.2, 6.2.3 and 6.2.4, MATLAB[®]R2016a was used. The training stage of convolutional neural networks requires a graphics processing unit (GPU) that allows the parallel computation thanks to the amount of available cores. For the experiment 6.2.5, a NVIDIA GeForce[®]GTX1080 with 8GB of GDDR5 RAM was installed and NVIDIA CUDA[®]Deep Neural

Network library (cuDNN) was configured in the aforementioned computer. In addition, OPENCV libraries with CUDA support were included in the development environment. Python2.7 was chosen as programming language and PyCharm 2017.1 as IDE. It is worth noting that Keras framework (Chollet et al., 2015) with TensorFlow as backend was used as high-level neural networks API.

Common metrics in the field of machine learning are used to evaluate the classification performance in each experiment. Accuracy (Acc.) is determined by the sum of correct predictions divided by the total number of cases. It shows how correct a diagnostic test identifies and excludes a given condition. Sensitivity (Sens.) or true positive rate (TPR) and Specificity (Spec.) or true negative rate (TNR) measure the proportion of positive and negative cases which are correctly identified as such, respectively. These metrics are highly relevant in the medical field. The positive and negative predicted values (PPV and NPV respectively) measure the proportion of positive (negative) correct predictions with respect to the total number of positive (negative) predictions determined by the classification system.

The receiver operating curve or ROC curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. The ROC curve is created by plotting the TPR against the 1-TNR or the false positive rate (FPR) at various threshold settings. The enclosing area of this curve is known as area under ROC curve (AUC) and it is an extended way to measure the predictive modelling accuracy (Huang and Ling, 2005). This measure allows to establish fair comparisons where there is a strong imbalance between classes (Prati et al., 2011). Based on (Mandrekar, 2010), a diagnostic test in the medical field with an AUC value of 0.5 suggest no discrimination. If the AUC value ranges from 0.7 to 0.8 the test is considered acceptable, if the test presents an AUC value included in the interval 0.8-0.9 it is considered excellent, and more than 0.9 is considered outstanding.

6.2.1 Effect of the pre-processing methods

In this experiment, the effect of the pre-processing methods presented in this thesis is studied. In particular, the colour transformation of fundus images (explained in Section 3.3.2) and the retinal vessel removal method (detailed in Section 3.4) are evaluated. For this purpose, the procedure explained in Section 4.2.1 and graphically represented in Figure 4.13 was used to train different classification models able to automatically identify pathological areas in fundus images. LBP variance histogram was used as image descriptor and SMV (with

linear kernel) as classification algorithm. Four tests using *E-OPHTHA exudates* as image database and combining the pre-processing steps before the feature extraction stage were carried out. The original images (OI) and the colour-transformed images (CTI) applying the sparse-based inpainting technique to remove the blood vessels (without vessels - WOV) and avoiding this step (with vessels - WV) were the datasets utilised for creating each committee of SVM classifiers (Figure 6.2).

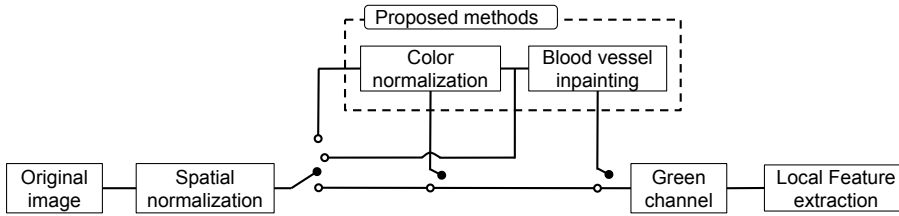


Figure 6.2: Overview of the possible combinations of input images to the local feature extraction stage.

The forty-seven images of *E-OPHTHA exudates* database were divided in $K = 5$ partitions. External cross-validation, using the “leave-one-out” technique, allowed us to carry out a fair validation of the proposed method. The parameters involved in the process were set according to the Table 6.2. Feature vectors of ten ($P+2$) components were locally extracted from the green channel of the fundus images and the patch-dataset was balanced (see Figure 4.13). Nine SVM models compose the decision committee to classify each testing instance. Notice that the optimal patch size and the overlap (Δx , Δy) used to extract the local features were obtained through internal cross-validation with eight folds. The pipeline of the experiment is summarised in Figure 6.3.

	Spatial Norm.	Local analysis	LBP	Classification
Parameters	$D_{ref}=700$	$N_w=64$ $\Delta x=\Delta y=32$	$R=1$ $P=8$	Linear kernel $T=9$

Table 6.2: Tunable parameters involved in the different stages of the lesion detection system. Values obtained after the cross-validation procedure.

Table 6.3 shows the classification results using the aforementioned models, trained using the original images (OI) and the colour-transformed images with (CTI-WV) and without (CTI-WOV) blood vessels. The accuracy, sensitivity, specificity and area under ROC Curve (AUC) values are reported. It is important to remark that the K -fold procedure and the random permutation

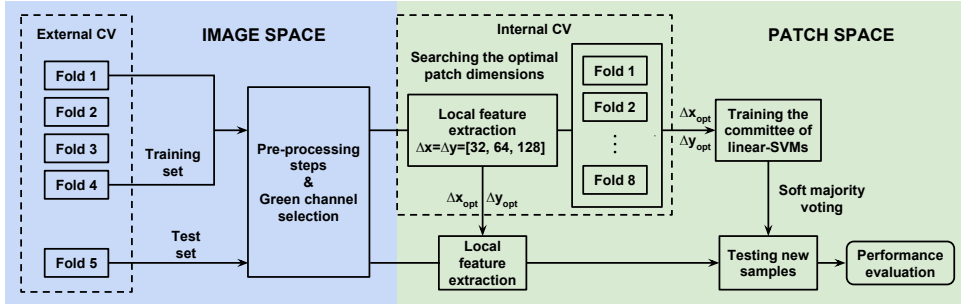


Figure 6.3: Pipeline of the experiment carried out to evaluate the pre-processing algorithms proposed in this PhD thesis.

for balancing the data was performed using a fix seed allowing fair comparison among the different pre-processing methods.

	Accuracy	Sensitivity	Specificity	AUC
OI-WV	0.7325±0.0255	0.6490±0.1106	0.7444±0.0472	0.7724±0.0458
CTI-WV	0.7404±0.0305	0.6439±0.1188	0.7547±0.0539	0.7805±0.0449
OI-WOV	0.8205±0.0320	0.7389±0.1034	0.8331±0.0534	0.8684±0.0435
CTI-WOV	0.8228±0.0379	0.7352±0.1163	0.8350±0.0634	0.8690±0.0379

Table 6.3: AUC, accuracy, sensitivity and specificity related to the exudate detection for *E-OPHTHA* database taking into account different inputs: original image (OI-WV), colour-transformed image (CTI-WV), original image without vessels (OI-WOV) and colour-transformed image without vessels (CTI-WOV).

As can be observed in Table 6.3, the exudate detection is more accurate when pre-processing methods are applied to the original images. In particular, the accuracy increases 8-9% when sparse-based inpainting is applied to remove blood vessels. Regarding the colour transformation technique, results show a slight improvement (less than 1% in each metric) when the normalization is performed. However, this small improvement does not reflect the high colour variability reduction demonstrated in the visual results shown in Figure 3.5. This fact could be produced by the loss of information when the green component of the RGB image is isolated to perform the feature extraction stage. Another possible hypothesis is that colour normalisation does not matter for bright lesion detection, because they have similar colour for most patients.

The averaged ROC curves for each test are plotted in Figure 6.4. As can be observed, these curves corroborate the outperforming of CTI-WOV with

respect to the rest of the explored pre-processing configurations for both, high and low false positives rates.

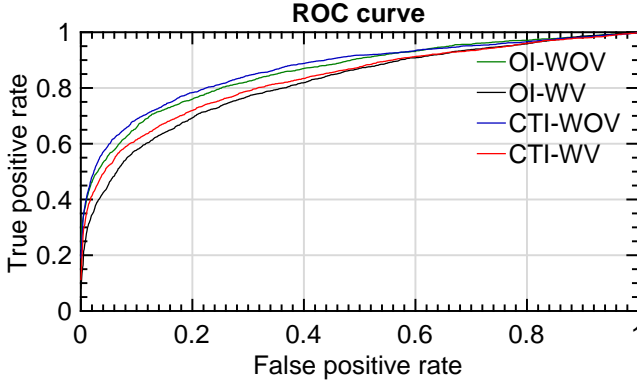


Figure 6.4: ROC curves for the different tests.

This experiment demonstrates the improvement of pre-processing methods for the posterior bright lesion detection on fundus images. Next experiments, will be performed using the best configuration of pre-processing algorithms previously to the feature extraction step, i.e. spatial and colour normalization database and blood vessel removal.

6.2.2 Behaviour of texture, shape and roughness descriptors

In this second experiment, the contribution of the three image description methods detailed in Section 4.1 is analysed. Local binary patterns, granulometric profiles and fractal dimension extract texture, morphological and roughness information from the fundus image, respectively. Different tests to explore which combination of features better describes bright and dark retinal lesions were performed.

The same process as described in the previous experiment was carried out. However, in this experiment the type of input images is constant, i.e. the retinal blood vessels of the spatial and colour-normalized images of fundus databases were removed by the sparse-based inpainting technique before applying the different image descriptors.

In a first set of tests, the behaviour of the morphological descriptors to discriminate bright lesions is assessed. For this purpose, *E-OPHTHA exudates* database was used and four morphological pyramids were computed from

the green component of each pre-processed image: two pyramids of openings using both an isotropic (Π_{γ_B}) and an angular (Π_{γ_L}) structuring element and analogously two pyramids of closings (Π_{φ_B} , Π_{φ_L}). These morphological pyramids were calculated using an increasing SE size defined by a step ($s = 2$) and a maximum value ($n_{max} = 22$). These parameters were optimised making use of the ground truth of the training image dataset. Note that angular granulometry was computed in directions 0° , 45° , 90° and 135° . From the morphological pyramids, a local description of the shape and size of the retinal texture was performed by computing the pattern spectrum of squared patches (PS°) as detailed in Section 4.1.4 and represented in Figure 4.11. The classification results obtained for different configurations of the feature vector are reported in Table 6.4. For simplicity in the notation, the PS° symbol has been omitted. Remark that feature vectors composed by more than one operator are generated by concatenation of their pattern spectra.

	Accuracy	Sensitivity	Specificity	AUC
γ_B	0.6405±0.0572	0.6137±0.1378	0.6489±0.0813	0.6701±0.0617
γ_L	0.6684±0.0353	0.6009±0.1449	0.6759±0.0521	0.6907±0.0601
φ_B	0.7014±0.0247	0.6396±0.1420	0.7090±0.0426	0.7337±0.0777
φ_L	0.5794±0.0588	0.5105±0.0887	0.5914±0.0823	0.5783±0.0366
$\gamma_B\varphi_B$	0.7554±0.0161	0.6592±0.1179	0.7698±0.0292	0.7877±0.0524
$\gamma_B\varphi_L$	0.6550±0.0528	0.6223±0.1243	0.6636±0.0745	0.6946±0.0546
$\gamma_B\gamma_L$	0.7321±0.0208	0.6394±0.0839	0.7465±0.0271	0.7543±0.0461
$\gamma_L\varphi_B$	0.7112±0.0262	0.6446±0.1344	0.7194±0.0451	0.7482±0.0684
$\gamma_L\varphi_L$	0.6685±0.0320	0.5935±0.1363	0.6774±0.0499	0.6882±0.0564
$\varphi_B\varphi_L$	0.7014±0.0256	0.6346±0.1376	0.7099±0.0432	0.7349±0.0749
$\gamma_B\gamma_L\varphi_B\varphi_L$	0.7620±0.0165	0.6648±0.1124	0.7762±0.0300	0.7924±0.0493

Table 6.4: AUC, accuracy, sensitivity and specificity related to the exudate detection for *E-OPHTHA exudates* database taking into account different morphological input vectors to the SVM classifier.

As can be interpreted from the results, when the feature vector is composed by a single pattern spectrum (eleven features), isotropic granulometries perform better the bright lesion detection than the linear ones. Intuitively, we could think that openings should describe better this kind of lesions, however results show the contrary behaviour. The reason of this fact, could be that bright lesions always are surrounded by dark regions and changes registered by the PS° , when these areas are removed in the morphological pyramid, may be more discriminative. When a combination of two granulometric patterns is used to feed the SVM classifier, an improvement in the performance classification

can be observed, specifically when the granulometry and anti-granulometry curves using an isotropic SE are collated in a unique feature vector. The best description of the bright lesions taking into account only size and shape local information is given when all morphological descriptors are used ($\gamma_B\gamma_L\varphi_B\varphi_L$). Accuracy and specificity values higher than 76%, a sensitivity of 66% and an AUC value around 0.8 are reported as best classification results in Table 6.4. These results suggest the need of extracting another kind of information able to support and strengthen the performance of the morphological descriptors.

According to the conclusions extracted from the first experiment, texture information was computed by means of the local binary patterns variance histogram detailed in Section 4.1.2. Several tests combining shape and texture descriptors were carried out and the obtained results can be observed in Table 6.5.

	Accuracy	Sensitivity	Specificity	AUC
LBPV	0.8205±0.0320	0.7389±0.1034	0.8331±0.0534	0.8684±0.0435
LBPV-φ_B	0.8369±0.0336	0.7603±0.0897	0.8481±0.0519	0.8834±0.0397
LBPV-$\gamma_B\varphi_B$	0.8445±0.0276	0.7657±0.0894	0.8561±0.0449	0.8872±0.0384
LBPV-$\gamma_B\gamma_L$	0.8447±0.0228	0.7496±0.1079	0.8591±0.0432	0.8803±0.0409
LBPV-$\gamma_B\gamma_L\varphi_B\varphi_L$	0.8533±0.0245	0.7721±0.0857	0.8651±0.0399	0.8948±0.0351

Table 6.5: AUC, accuracy, sensitivity and specificity related to the exudate detection for *E-OPHTHA exudates* database taking into account different input vectors to the SVM classifier. In particular they are composed by texture and shape descriptors.

As can be observed in the table exposed above, texture information is key in the ability of discriminating between regions containing exudates and non-damaged areas. Moreover, the best classification results are obtained when texture and shape/size information are combined. Improvements around 10% in all the evaluation metrics are registered in comparison with the previous test, in which, only morphological information was taken into account.

Finally, two more tests were performed including the fractal dimension descriptor to the best previous configuration. In addition, the behaviour of LBPV-H is computed as baseline to demonstrate that morphological descriptors notably improve the classification performance. From Tables 6.5 and 6.6, it is possible to observe that roughness information slightly outperforms the previous results reducing the standard deviation among folds and giving place to an accuracy higher than 85%, a sensitivity value of 77.44%, a specificity of 86.69% and an AUC value next to 0.9. The combination of the three image descriptors studied in depth in this PhD thesis composes a

promising and robust feature vector to detect bright damaged areas in the retina background.

	Accuracy	Sensitivity	Specificity	AUC
LBPV-H	0.8284±0.0314	0.7405±0.1025	0.8418±0.0525	0.8704±0.0405
LBPV-H-$\gamma_B\gamma_L\varphi_B\varphi_L$	0.8552±0.0236	0.7744±0.0856	0.8669±0.0375	0.8967±0.0345

Table 6.6: AUC, accuracy, sensitivity and specificity related to the exudate detection for *E-OPHTHA exudates* database taking into account different input vectors to the SVM classifier. The contribution of fractal dimension descriptor is evaluated.

In the same way, the behaviour of the previous descriptors was analysed in the dark-lesion detection task (i.e. microaneurysms and haemorrhages identification). *DIARETDB1* was utilised for this purpose and the same tests as for bright-lesion detection were performed. Note that parameters relative to spatial normalization, LBP and type of SVM kernel remained as Table 6.2 showed. However, the fact of looking for smaller lesions and using a different database propitiated a smaller optimal patch size of $N_w = 32$ with $\Delta x = \Delta y = 16$ of overlapping and eleven SVM models in the committee. In Table 6.7 the results related to the classification performance using morphological feature vectors can be observed.

	Accuracy	Sensitivity	Specificity	AUC
γ_B	0.7016±0.0244	0.6438±0.1100	0.7094±0.0394	0.7384±0.0449
γ_L	0.7417±0.0200	0.6202±0.1206	0.7572±0.0330	0.7590±0.0486
φ_B	0.6966±0.0292	0.6163±0.1014	0.7064±0.0390	0.7127±0.0533
φ_L	0.6868±0.0398	0.6335±0.0856	0.6939±0.0442	0.7128±0.0533
$\gamma_B\varphi_B$	0.7187±0.0224	0.6520±0.1083	0.7275±0.0368	0.7548±0.0424
$\gamma_B\varphi_L$	0.7343±0.0184	0.6581±0.1031	0.7444±0.0314	0.7660±0.0458
$\gamma_B\gamma_L$	0.7393±0.0224	0.6565±0.1181	0.7502±0.0384	0.7748±0.0423
$\gamma_L\varphi_B$	0.7408±0.0204	0.6182±0.1199	0.7565±0.0336	0.7558±0.0477
$\gamma_L\varphi_L$	0.7404±0.0219	0.6281±0.1098	0.7549±0.0293	0.7534±0.0519
$\varphi_B\varphi_L$	0.7245±0.0277	0.6293±0.1006	0.7365±0.0329	0.7344±0.0520
$\gamma_B\gamma_L\varphi_B\varphi_L$	0.7523±0.0185	0.6637±0.1087	0.7639±0.0327	0.7836±0.0421

Table 6.7: AUC, accuracy, sensitivity and specificity related to the dark lesion detection (microaneurysms and haemorrhages) on *DIARETDB1* database taking into account different morphological input vectors to the SVM classifier.

In this case, from the previous table we can observe that openings extract more information about the reddish lesions. Using isotropic granulometries, the ability of correctly detecting these lesions is superior than using the linear

ones. However, the great amount of false positives makes that specificity falls off. Despite this observation, the best morphological feature vector is composed by the isotropic and linear granulometric curves like the bright-lesion detection case. Finally, several tests incorporating texture and roughness information were performed and the results related to the classifier performance evaluation are reported in Tables 6.8 and 6.9.

	Accuracy	Sensitivity	Specificity	AUC
LBPV	0.7016±0.0142	0.6958±0.0736	0.7313±0.0234	0.7664±0.0397
LBPV-γ_L	0.7083±0.0119	0.6964±0.0777	0.7088±0.0206	0.7706±0.0413
LBPV-$\gamma_B\varphi_L$	0.7332±0.0088	0.6976±0.0856	0.7368±0.0187	0.7722±0.0187
LBPV-$\gamma_B\gamma_L$	0.7262±0.0114	0.6998±0.0872	0.7286±0.0220	0.7866±0.0392
LBPV-$\gamma_B\gamma_L\varphi_B\varphi_L$	0.7368±0.0091	0.7009±0.0843	0.7407±0.0193	0.7954±0.0379

Table 6.8: AUC, accuracy, sensitivity and specificity related to the dark lesion detection (microaneurysms and haemorrhages) on *DIARETDB1* database, taking into account different input vectors to the SVM classifier. In particular, they are composed by texture and shape descriptors.

	Accuracy	Sensitivity	Specificity	AUC
LBPV-H	0.7161±0.0135	0.6883±0.0824	0.7191±0.0229	0.7748±0.0396
LBPV-H-$\gamma_B\gamma_L\varphi_B\varphi_L$	0.7404±0.0134	0.6969±0.0865	0.7462±0.0245	0.7975±0.0367

Table 6.9: AUC, accuracy, sensitivity and specificity related to the dark lesion detection (microaneurysms and haemorrhages) on *DIARETDB1* database taking into account different input vectors to the SVM classifier. The contribution of fractal dimension descriptor is evaluated.

From the last tables, evaluation metrics reveal that texture and morphological information extract the majority of the relevant knowledge for describing dark retinal lesions. Regarding to fractal dimension, in this case, it improves the classification less than 0.5% and sensitivity suffers a slightly decrement. All this propitiates an accuracy and specificity values around 75%, a sensitivity of 70% and an AUC value next to the 0.8.

Comparing the classification results obtained for detecting the two types of retinal lesions several conclusions can be extracted:

- The performance of morphological descriptors is notable in the detection of both kind of lesions. Similar results are obtained in this context as can be observed in Tables 6.4 and 6.7.
- LBP variance histogram has a better ability of describing bright than dark lesions (see Tables 6.5 and 6.8). For bright lesions an improvement around

10% in all the measures is registered while in dark-lesion detection the overall performance remains constant. An increase of the sensitivity value propitiates a decrease of the specificity parameter. AUC, the metric that provides a general overview, improves 0.012 when texture information is included.

- Regarding the fractal dimension descriptor, its behaviour is similar in both cases. The addition of roughness information slightly improves the classification of retinal lesions and helps to reduce the standard deviation among folds (see Tables 6.6 and 6.9).
- In general, results show that the anomalous dark pattern description task is more difficult to successfully perform than the bright lesion detection. The small size of microaneurysms (15-125 microns) makes difficult that image descriptors can catch the information related to all these lesions, even using them locally. This fact can increase the misclassification rate. In addition, the method proposed by Morales et al. (2014) was used to segment the blood vessels network and in some difficult cases it produces under/over segmentation. When this occurs, an error source introduced by the vessel inpainting step can negatively affect to the classification system. The elimination of microaneurysms due to the over-segmentation and remaining vessels due to the under-segmentation can hinder in the success of the classifier.

6.2.3 Performance of machine learning algorithms

The third experiment aims to analyse the behaviour of the different classification algorithms exposed in Section 4.2. The best configuration of database pre-processing was used in this, i.e. spatial and colour-normalized images, in which, vessels were removed using the sparse-based inpainting method proposed in this thesis. In addition, the most dicriminative feature vector (according to Section 6.2.2), composed by texture, morphological and roughness information, was chosen as input to the classification stage, in which, different machine learning methods were tested. Before presenting the obtained reuslts for each classification method, some practical aspects related to the them must be noticed:

- The embedded implementation of Random Forests in MATLAB (the `TreeBagger` class)³, based on (Breiman, 2001), was used to construct a committee of nine random forests models, each one composed by 50

³<http://es.mathworks.com/help/stats/treebagger.html>

trees. A RF committee per fold or partition was computed and its performance was evaluated. The final results were obtained by averaging the evaluation metrics of the K committees (one per fold).

- The library for support vector machines (libSVM)⁴ introduced in (Chang and Lin, 2011) was used to construct a committee of nine linear or RBF-SVM models. A committee per fold or partition was computed and its performance was evaluated. The final results were obtained by averaging the evaluation metrics of of the K committees (one per fold). It is worth noting that in the case of RBF-SVM the “cost” (C) and “length-scale” (γ) optimal parameters were found by grid search (in two stages: coarse and fine grids).
- Gaussian Processes for Classification (GPCs) were used as classification algorithm by means of the public implementation GPLM v3.6 for MATLAB⁵. The same procedure as described in the previous cases was followed using GPCs. Variational Bayes inference and logistic likelihood function for the classification were selected to learn the GP models.

Remember that the K -fold procedure and the random permutation for balancing the data (see the procedure represented in Figure 4.13) were performed using a fix seed to allow fair comparisons among the different classification methods. Table 6.10 shows the results related to the classification performance using each Machine Learning method.

	Accuracy	Sensitivity	Specificity	AUC
Random Forests	0.9508±0.0084	0.4785±0.1015	0.9921±0.0043	0.9256±0.0173
Linear-SVM	0.8552±0.0236	0.7744±0.0856	0.8669±0.0375	0.8967±0.0345
RBF-SVM	0.8796±0.0229	0.8118±0.0618	0.8851±0.0296	0.9240±0.0161
Gaussian Processes	0.8762±0.0206	0.8348±0.0650	0.8795±0.0266	0.9353±0.0174

Table 6.10: AUC, accuracy, sensitivity and specificity related to the bright lesion detection on *E-OPHTHA exudates* database for each classification method studied in this thesis.

The accuracy, sensitivity and specificity results presented above were obtaining using a decision threshold $\delta = 0.5$ after the soft majority voting algorithm (see Section 4.2.1). As can be observed in the previous table, different machine learning algorithms maximise specific evaluation metrics, difficulting an overall analysis of their behaviour. AUC values provide information about the global performance of each classifier because the ROC curve is calculated

⁴www.csie.ntu.edu.tw/~cjlin/libsvm

⁵www.gaussianprocess.org/gplm/code/matlab/doc

by evaluating a sweep of decision thresholds ranging from zero to one (see Figure 6.5). However in the medical field, sensitivity and specificity are key parameters for measuring the goodness of proposed methods. For this reason, from the ROC curve of each fold, an optimal decision threshold δ for each ML method was obtained according to the best trade-off between sensitivity and specificity. The same procedure used to optimise the meta-parameters involved in the feature extraction and classification stages was used to optimise the threshold δ . In particular, it was optimised by using the validation dataset for each partition of the external K -fold cross-validation. The obtained results can be observed in Table 6.11

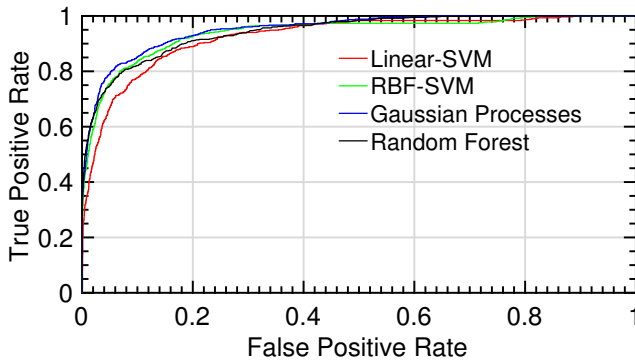


Figure 6.5: ROC curves for the different tests.

	Accuracy	Sensitivity	Specificity	δ
Random Forests	0.8410±0.0181	0.8418±0.0178	0.8411±0.0181	0.8999±0.0214
Linear-SVM	0.8242±0.0279	0.8243±0.0284	0.8242±0.0278	0.5771±0.0958
RBF-SVM	0.8529±0.0190	0.8531±0.0193	0.8529±0.0190	0.5835±0.1012
Gaussian Processes	0.8581±0.0221	0.8579±0.0221	0.8579±0.0222	0.5348±0.0747

Table 6.11: AUC, accuracy, sensitivity and specificity related to the bright lesion detection on *E-OPHTHA exudates* database. Results were optimised to the best sensitivity/specificity trade-off.

In order to evaluate the robustness of the proposed approach, our bright lesion detection methodology was applied using the forty seven images containing exudates of *DIARETDB1* database. This test allowed us to perform a deep validation of the presented approach and to establish an exhaustive comparison with the state-of-the-art works related to the automatic exudate detection. Table 6.12 shows the obtained results using a decision threshold $\delta = 0.5$ while Table 6.13 shows the results maximizing the trade-off between sensitivity and

specificity by searching the optimal decision threshold δ for each classification method.

	Accuracy	Sensitivity	Specificity	AUC
Random Forests	0.9370±0.0155	0.5096±0.0478	0.9889±0.0056	0.8852±0.0314
Linear-SVM	0.8702±0.0217	0.7396±0.0528	0.8849±0.0260	0.8879±0.0301
RBF-SVM	0.8834±0.0311	0.0.7509±0.0489	0.8903±0.0211	0.8901±0.291
Gaussian Processes	0.8703±0.0189	0.7705±0.0500	0.8814±0.0224	0.8971±0.0286

Table 6.12: AUC, accuracy, sensitivity and specificity related to the bright lesion detection on *DIARETDB1* database for each classification method studied in this thesis.

	Accuracy	Sensitivity	Specificity	δ
Random Forests	0.8037±0.0319	0.8027±0.0324	0.8038±0.0319	0.9006±0.0146
Linear-SVM	0.8112±0.0316	0.8107±0.0315	0.8112±0.0316	0.6207±0.0440
RBF-SVM	0.8166±0.0322	0.8154±0.0321	0.8155±0.0322	0.6334±0.0401
Gaussian Processes	0.8184±0.0324	0.8183±0.0324	0.8184±0.0324	0.6023±0.0518

Table 6.13: AUC, accuracy, sensitivity and specificity related to the bright lesion detection on *DIARETDB1* database. Results were optimised to the best sensitivity-specificity trade off.

As it was mentioned above, thanks to *DIARETDB1* public database, the proposed method for exudate detection can be compared with some state-of-the-art works, in which, the detection/segmentation of these lesions is performed by using classical techniques such as filtering and mathematical morphology. Table 6.14 summarises the exudate detection results achieved by the proposed method and by other works of the literature.

Methods	Sensitivity	Specificity	PPV
Sopharak et al. (2008)	0.4348	0.9931	0.2548
Walter et al. (2002)	0.6600	0.9864	0.1945
Welfer et al. (2010)	0.7048	0.9884	0.2132
Ghafourian and Pourreza (2012)	0.7828	-	-
Proposed method	0.8184±0.0324	0.8183±0.0324	0.4373±0.1374

Table 6.14: Comparison of exudate detection methods for the 47 retinal images with exudates of *DIARETDB1* database.

The proposed method presents the best trade-off between sensitivity-specificity improving the representative methods involved in the comparison. In (Sopharak et al., 2008) and (Walter et al., 2002), poor sensitivity values are achieved. This fact is quite relevant taking into account that this parameter

is directly related with the quality of predicting positive instances. Sensitivity value is the unique measure provided by Ghafourian and Pourreza (2012) about the performance of its method. Taking into account only one parameter is not possible to establish the goodness of that method. The algorithm introduced by Welfer et al. (2010) provides interesting results. However, analysing together the three parameters of Table 6.14, is possible to observe that our method outperforms the Welfers' one. It is important to note that if AUC parameter would have been included in the works involved in the comparison a more exhaustive analogy would have enabled.

In the same way, the classification methods studied in this thesis were utilised to discriminate between healthy and dark-damaged areas (i.e. patches containing haemorrhages and/or microaneurysms). Table 6.15 reports the obtained results using a decision threshold $\delta = 0.5$ while Table 6.16 shows the results maximizing the trad-off between sensitivity and specificity by searching the optimal decision threshold δ for each classification method.

	Accuracy	Sensitivity	Specificity	AUC
Random Forests	0.9016±0.0311	0.1718±0.0800	0.9818±0.0035	0.8150±0.0356
Linear-SVM	0.7404±0.0134	0.6969±0.0865	0.7462±0.0245	0.7975±0.0367
RBF-SVM	0.7603±0.0223	0.7299±0.0821	0.7576±0.0311	0.8209±0.344
Gaussian Processes	0.7612±0.0244	0.7489±0.0844	0.7630±0.0350	0.8344±0.0330

Table 6.15: AUC, accuracy, sensitivity and specificity related to the dark lesion detection (microaneurysms and haemorrhages) on *DIARETDB1* database for each classification method studied in this thesis.

	Accuracy	Sensitivity	Specificity	δ
Random Forests	0.7397±0.0328	0.7391±0.0320	0.7398±0.0329	0.8645±0.0329
Linear-SVM	0.7262±0.0329	0.7261±0.0324	0.7262±0.0330	0.5170±0.0591
RBF-SVM	0.7498±0.0277	0.7375±0.0301	0.7374±0.0302	0.5554±0.0601
Gaussian Processes	0.7562±0.0290	0.7561±0.0289	0.7562±0.0290	0.5036±0.0712

Table 6.16: AUC, accuracy, sensitivity and specificity related to the dark lesion detection (microaneurysms and haemorrhages) on *DIARETDB1* database. Results were optimised to the best sensitivity-specificity trade off.

The optimal configuration of the system for dark lesion detection (i.e. the same feature vector and classification algorithm as for bright lesion detection) is also compared with some state-of-the-art works, in which, the detection/segmentation of these lesions is performed by using classical techniques such as filtering and mathematical morphology. Table 6.17

summarises the dark lesion detection results achieved by the proposed method and by other works of the literature.

Methods	Sensitivity	Specificity	AUC
Rocha et al. (2012)	0.9000	0.6000	0.7640
Roychowdhury et al. (2012)	0.7550	0.9373	0.8663
Ashraf et al. (2014)	0.6763	0.6278	0.6500
Junior and Welfer (2013)	0.8769	0.9244	-
Proposed method	0.7561±0.0301	0.7562±0.0290	0.8344±0.0330

Table 6.17: Comparison of dark lesion detection methods for the 45 retinal images with microaneurysms or haemorrhages of *DIARETDB1* database.

The proposed method presents comparable results with the representative methods involved in the comparison. It is worth noting that in the state-of-the-art papers, procedures are designed ad-hoc to identify microaneurysms and haemorrhages. However, in this work, a generic feature vector and classification algorithm have been proposed for describing both kind of retinal lesions: bright and dark.

Throughout these three experiments, the best configuration of a home-made automatic system for detecting abnormal retinal tissue was studied in depth. Colour-normalized fundus images in which blood vessels were removed by using the proposed sparse-based inpainting method maximised the classification accuracy of a committee of GP models trained using texture, morphological and roughness features. It is important to highlight the ability of this methodology to successfully describe both bright and dark lesions using the same feature vector. This fact provides our detection system a high level of robustness and simplicity. In addition, the local analysis in which is based the image description and classification not only makes unnecessary the stage of segmentation or candidate generation, but it also provides an accurate location of the damaged retinal area (Figure 6.6).

6.2.4 *Global detection of bright and dark retinal lesions for automatic screening.*

As it was mentioned in Section 1.2, the main goal of this PhD thesis is the development of new algorithms, based on image descriptors and machine learning algorithms, to characterise healthy and pathological tissues in the retina background, avoiding segmentation steps. Previous experiments focus on finding the best configuration of pre-processing algorithms, local

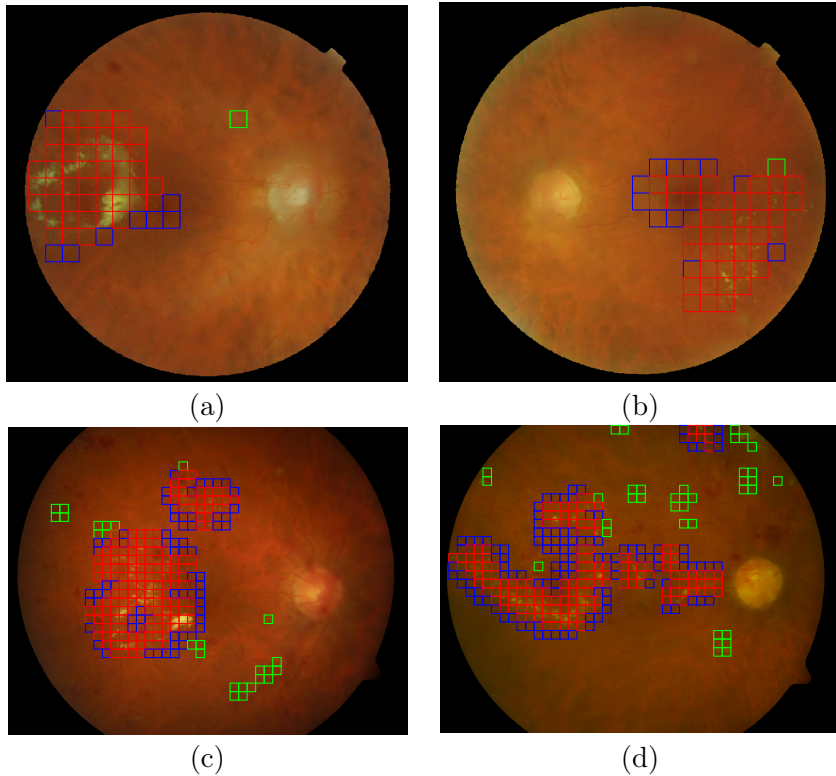


Figure 6.6: Automatic lesion detection in four retinal images using the best configuration of feature vector and gaussian processes for classification. (a,b) Two representative images (DS000U30.jpg and DS0009NU.jpg) from the *E-OPHTHA exudates* database and (c,d) two representative image (image014.png and image016.png) from *DIARETDB1* database. Red squares indicate the true positives, green squares the false positives and the blue squares reveal the false negative detections.

descriptors and classification methods to accomplish the aforementioned goal. However, from a clinical point of view an image-wise detection of the lesions is highly important for the automatic diagnosis of the most common eye diseases. Following this premise, the purpose of this experiment relies in the discrimination among images containing bright lesions, images containing red lesions and images without any signs of pathology.

Thanks to the external cross-validation procedure, all the images of the database were tested (see Figure 4.13) when previous experiments were performed. The prediction using the generated classification models consists

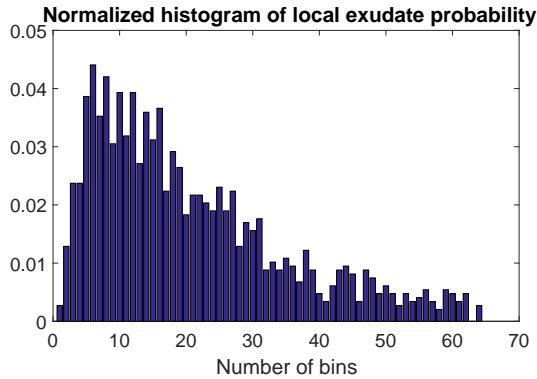
of two probability values for each patch of the image, i.e. the probability p_P of belonging to the pathological class and the probability $p_H = 1 - p_P$ of belonging to the healthy class. So, using these local probabilities, a global descriptor was computed for each image. A probability histogram is proposed as naive global descriptor to detect damaged retinal images.

The probability values for each patch of the 89 images of *DIARETDB1* database were used in this experiment in order to detect damaged images. The imbalanced character of the database (see Section 6.1), makes impossible the automatic detection of diabetic retinopathy. Taking into account this limitation, in this experiment we tried to distinguish among images containing bright lesions (41 images) and images containing dark lesions (45 images) by using the local information extracted from the previous experiments.

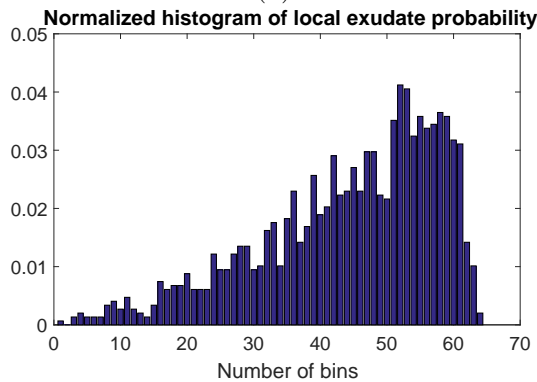
As it was mentioned above, a probability histogram per image was computed. Figure 6.7 show two 64-bin normalized histograms calculated using the local probability values of belonging to the exudate class (with the optimal configuration reported in Table 6.13). In particular, Figure 6.7(a) shows the normalized histogram of probabilities for a healthy image (or image containing only dark lesions) while in 6.7(b) the same feature vector extracted for an image containing bright lesions can be observed.

Previous figures highlight the different behaviour of the proposed global image descriptor depending on the presence or absence of bright lesion. In order to quantify the goodness of this methodology, different classifiers were trained using as feature vector the probability histogram. Specifically, histograms with different number of bins (i.e 16, 32, 64 and 128) were extracted and principal component analysis (PCA)(Jolliffe, 1986) explaining a 95% of variance was selected with the objective of finding the optimal feature vector. For this purpose, a 10-fold cross-validation was carried out, avoiding also biased results. All the ML methods of the classification learner app introduced on MATLAB[®]R2015a were tested (i.e. decision trees, discriminant analysis, logistic regression classifiers, SVM with different kernels and Nearest Neighbour classifiers). In Table 6.18 and Figure 6.8, the output parameters of the best classification process are reported. A feature vector composed for the fifteen principal components extracted from the 64-bin histogram feeding a SVM classifier with linear kernel provided the optimal discrimination between images containing any sign of bright lesion and images without exudate presence.

Regarding to the dark lesions (microaneurysms and haemorrhages), the same test was carried out. The normalized histograms were computed from the patch probabilities extracted in the previous experiment (see Table 6.16) and



(a)



(b)

Figure 6.7: Representative global descriptors: (a) Histogram of the local probability of containing bright lesion for a healthy image (or image containing only dark lesions) and (b) the same feature vector extracted from an image containing exudates.

TP	TN	FP	FN	Acc.	Sens.	Spec.
32	43	5	9	0.8427	0.7805	0.8958

Table 6.18: True positives (TP), true negatives (TN), false positives (FP), false negatives (FN), accuracy (Acc.), sensitivity (Sens.) and specificity (Spec.) related to the global bright lesion detection on *DIARETDB1* database.

a new ground truth vector was generated according to the expert annotations. Table 6.19 reports the performance evaluation parameters related to the best classification process. A set of bagged trees using a feature vector composed

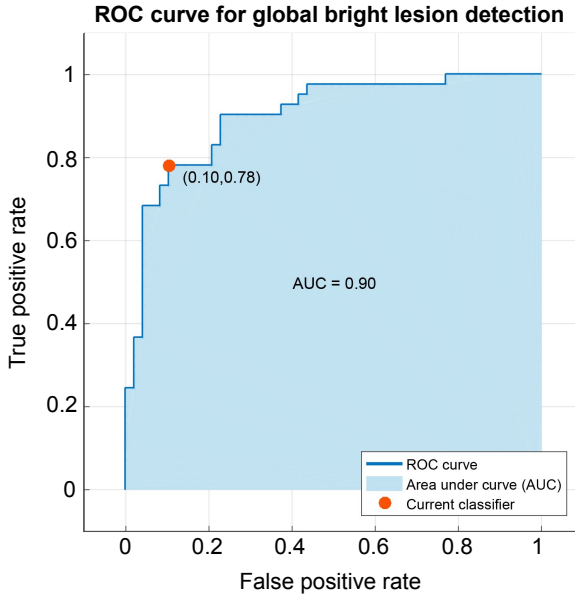


Figure 6.8: ROC curve for the global detection of images containing exudates on *DIARETDB1* database.

by the 18 principal components from the 64-bin histogram propitiated the best discrimination between images containing any microaneurysm or haemorrhage and images without any dark lesion. Figure 6.9 reports the ROC curve (with the AUC value) of the global dark lesion detection task.

TP	TN	FP	FN	Acc.	Sens.	Spec.
33	38	6	12	0.7977	0.7333	0.8636

Table 6.19: True positives (TN), true negatives (TN), false positives (FP), false negatives (FN), accuracy (Acc.), sensitivity (Sens.) and specificity (Spec.) related to the global dark lesion discrimination on *DIARETDB1* database.

Comparison with the majority of the state-of-the art works cannot be established due to those works focused on the identification of diabetic retinopathy while here the purpose is the discrimination between dark and bright lesions. Despite of this limitation, in the work presented by Kauppi et al. (2007), sensitivity and specificity values for both kind of lesions were reported and are shown in Table 6.20. Comparing these results with Tables 6.18 and 6.19, a better performance of the proposed methodology can be observed.

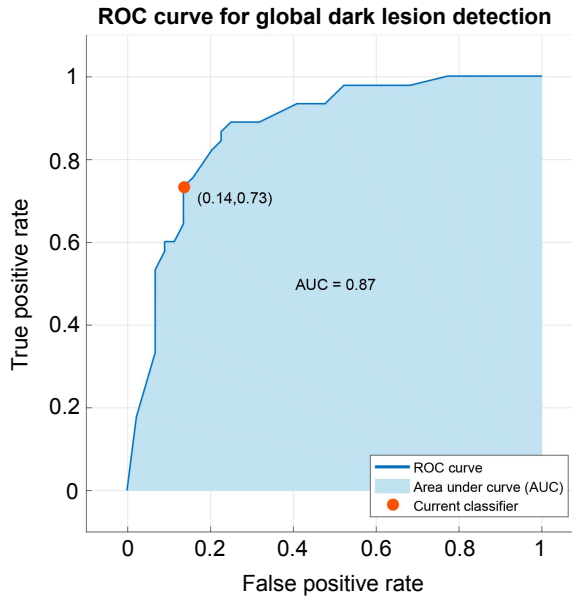


Figure 6.9: ROC curve for the global detection of images containing microaneurysms or haemorrhages on *DIARETDB1* database.

Type of lesion	Sensitivity	Specificity
Bright lesions	0.6525	0.8768
Dark lesions	0.6759	0.6869

Table 6.20: Sensitivity and specificity reported by Kauppi et al. (2007) related to the global bright and dark lesion discrimination on *DIARETDB1* database.

The results of both tests suggest the powerful of the local description analysis studied in this PhD thesis. Making use of this information, a naive descriptor was constructed for detecting retinal damage in the global fundus images. The normalized histogram of the local lesion probabilities is able to distinguish between images containing both kind of lesions (i.e. bright and dark). Due to the imbalanced character of *DIARETDB1* and the nature of *E-OPHTHA* database (different images labelled for each kind of lesion) a diabetic retinopathy detection test was not possible to perform. However, in future work, global histograms for detecting bright and dark lesions will be collated to distinguish between images containing any early sign of DR and healthy images. In addition, more robust global descriptors will be explored taking

into account the spatial information provided by the local analysis carried out in the previous experiments.

6.2.5 Pathological retinal tissue identification by CNNs

In this experiment, a different approach for detecting retinal lesions is presented. In particular, convolutional neural networks were used to automatically describe abnormal retinal tissue making use of the fine-tuning technique. As it was detailed in Section 5.5, the reduced amount of labelled data is the main bottleneck in medical applications due to the tedious and time-consuming pixel-wise annotation task. For this reason, the fine-tuning methodology was performed to transfer the knowledge of healthy and damaged retinal tissue to the most important state-of-the-art models pre-trained on the ImageNet dataset. In particular, VGG16 and VGG19 (Krizhevsky et al., 2012), InceptionV3 (Szegedy et al., 2015), Xception (Chollet, 2016) and ResNet50 (He et al., 2016) architectures and their corresponding weights were obtained by means of the Keras Applications module and fine-tuned using the *E-OPHTHA exudates* database.

In the same way as our previous experiments, external cross-validation using five folds was employed to divide the 47 original images (i.e without any kind of pre-processing) of *E-OPHTHA exudates* database in training and test datasets, avoiding the bias introduced by a specific separation. One model per fold or partition was obtained following the pipeline detailed in Section 5.6 and represented in Figure 5.15. It is important to note that patches were extracted following the process represented in Figure 4.2(b), using a squared sliding window of size $N_w = 64$ with an overlap of $\Delta x = \Delta y = 32$ pixels (Figure 6.10). The total training patches obtained for each fold were divided into 80% for training and 20% for validation and samples were balanced. Note that before the training process, data augmentation was performed by rotating, flipping and scaling the original patches.

After the data extraction process, the architecture to be fine-tuned was loaded from Keras applications module. Note that the top model was not imported because a replacement of the nodes in the last fully connected layer was necessary (i.e. from one thousand classes of ImageNet database to our two possible states: healthy, pathological). The “knowledge” of the pre-trained networks from natural non-medical images was transferred to all layers except our last fully connected by initialising the CNN with the publicly available weights. After this initialisation, the neural network was shallow-tuned. It means that weights related to the first layers were frozen while the last-layers

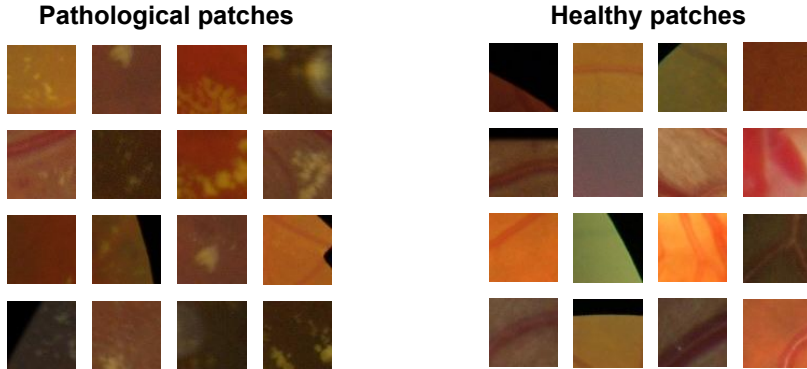


Figure 6.10: Representative patches of pathological and healthy classes extracted from the *E-OPHTHA* database.

weights (top-model layers included) were set to “learnable”. The training process was performed during 150 epochs using a learning rate of 1×10^{-4} and a batch size equals to 16. It is important to remark that the layer from which the fine-tuning is started and the optimiser used to solve the backpropagation process are different for each CNN architecture. Table 6.21 shows the layer from which the training process started (itself included) and the optimiser used for each fine-tuned model.

CNN architecture	Layer name	Optimizer
VGG16	‘block5_conv1’	Stochastic Gradient Descent
VGG19	‘block5_conv1’	Stochastic Gradient Descent
ResNet50	‘res4c_branch2a’	RMSprop
InceptionV3	‘conv2d_57’	Nesterov adam
Xception	‘block14_sepconv1’	Nesterov adam

Table 6.21: Name of the layer from which the fine-tuning process starts depending on the CNN architecture.

The same metrics as previous experiments were used to quantify the performance of the classifier (i.e. accuracy, sensitivity, specificity and AUC). In Figure 6.11, these metrics can be graphically observed for each fine-tuned CNN model. In particular, a boxplot showing the distribution of each metric according to each fold ($K = 5$ in our case) for each fine-tuned model. It is worth noting that the model labelled as HCFE-ML (hand-crafted feature extraction and machine learning) is the best model obtained from the previous

experiments. In particular, it is the classification model obtained by using texture, shape and roughness information for feeding a committee of GPC classifiers. Table 6.22 numerically summarises the average values of each metric along the five partitions and Figure 6.11 reports the ROC curves for each CNN architecture.

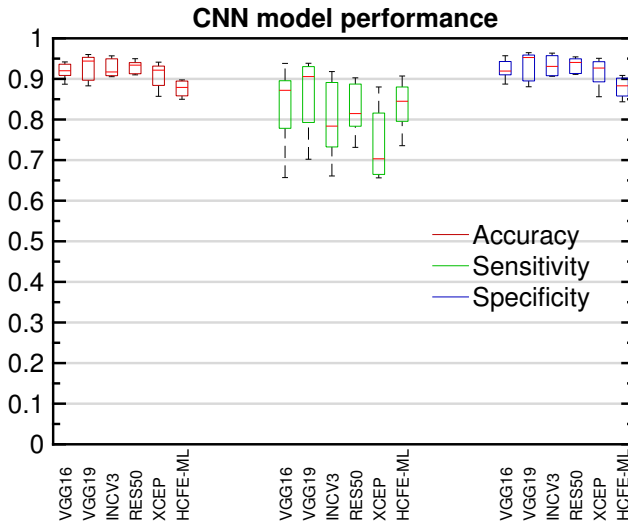


Figure 6.11: Graphic representation of the distribution of accuracy, sensitivity and specificity along the five image folds depending on the tested CNN models.

Model	Accuracy	Sensitivity	Specificity	AUC
VGG16	0.9198	0.8334	0.9239	0.9451
VGG19	0.9272	0.8594	0.9309	0.9539
ResNet50	0.9288	0.8263	0.9336	0.9359
InceptionV3	0.9270	0.8002	0.9329	0.9250
Xception	0.9082	0.7403	0.9156	0.8894
HCFE-ML	0.8762	0.8348	0.8795	0.9353

Table 6.22: AUC, accuracy, sensitivity and specificity related to the bright lesion detection on *E-OPHTHA exudates* database for each fine-tuned state-of-the-art architecture.

From the presented results, several conclusions can be extracted. First of all, in general, the models based on CNNs outperforms the classical methodology of hand-crafted feature extraction and classification for the task of detecting

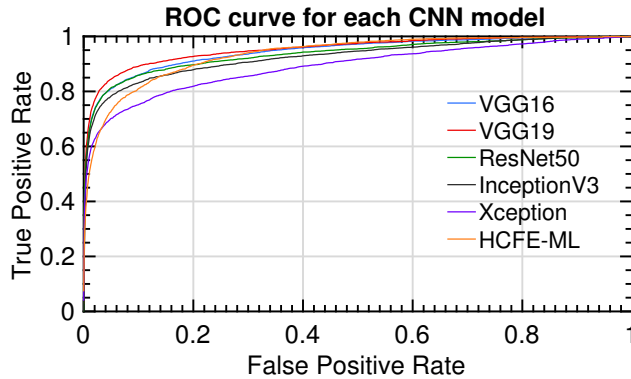


Figure 6.12: ROC curves for each fine-tuned CNN architecture averaged along the five image partitions. Remember that HCFE-ML refers to the best configuration of the system based on hand-crafted image descriptors and machine learning algorithms for the classification.

retinal bright lesions. In addition, from all the tested CNN architectures, VGG19 and ResNet50 are the best candidates to solve the problem under study. If the objective is to maximise the sensitivity, in other words, the ability of the model for recognising abnormal tissue, the most appropriate model is the VGG19. In fact, this model shows the best overall performance as its value of $AUC = 0.9539$ suggests. It is worth highlighting that the CNN approach does not require any kind of pre-processing of the fundus images. This fact propitiates an interesting time reduction in the testing process.

Chapter 7

Conclusions

As was indicated in Chapter 1, the main objective of this thesis was to develop and implement new automatic algorithms based on fundus image description and classification methods to characterise the healthy and pathological tissues in the retina background avoiding the lesion segmentation. As the majority of medical image modalities, fundus photographs present a huge intra/inter database variability due to the non-accurate acquisition process and physiological phenomena. For this reason, in the first part of this manuscript (Chapter 3) novel fundus pre-processing methods were presented. On the one hand, the colour of the retina background was explored and a colour transformation based on geometric transformations was proposed for normalising fundus images. To the best of the author's knowledge, in the literature did not exist methods to reduce the high colour heterogeneity among fundus images. On the other hand, a new inpainting algorithm based on sparse representation and dictionary learning techniques was introduced for removing the retinal blood vessels. These pre-processing methods are the first stage of the whole pipeline presented in this thesis to identify retinal regions containing abnormalities. In this way, three image descriptors were explained and formulated to the problem under study (see Chapter 4). Local binary patterns, fractal dimension and granulometric profiles were locally used to extract texture, roughness and morphological information from the retina background. Feature vectors composed by this complementary information

feed different committees of classifiers learned by both classical and novel machine learning algorithms. This new approach in which, unlike the most of the literature, the creation of a lesion candidates map is not necessary, was validated using two public fundus databases without any image exclusion criteria. Throughout four of the six experiments reported in Chapter 6, the best configuration of a home-made system was analysed.

In the first four experiments different findings were discovered. First of all, the effect of the pre-processing methods presented in this thesis was evaluated through the bright lesion detection. Results showed that removing the blood vessels using the proposed sparse-based inpainting method benefit the miss-classification of exudate detection. The contribution of large vessels in some patches containing small lesions could affect in the learning process of our classification models if only texture information is used. Regarding to the colour normalisation procedure, quantitative results suggested that a slightly improvement in the lesion detection is obtained when the colour-transformation is performed before training the classifiers. However, the improvement is minimum compared with the great quality improvement of the visual results reported in Section 3.3.1. This fact could be produced because the feature extraction stage is performed only in the green channel (in which lesions and background present the maximum contrast). Secondly, several tests using different combinations of feature vectors for the classification of retinal background areas were performed. In this way, according to the results obtained, the combination of texture, morphological and roughness information provide to the system the best discrimination ability for detecting both bright and dark lesions. To the best of the author's knowledge, there are no previous studies in which, using the same approach, bright and dark lesions were detected with remarkable precision. Regarding to the classification stage, in this thesis the use of Gaussian processes for classification was introduced in the fundus image field. A noticeable outperforming of GPCs over the classical random forest and SVM approaches to the abnormal retinal tissue classification was found. Finally, the global image diagnose from the available local information was addressed. For this purpose, a simple histogram of the probabilities was proposed as global image descriptor. Some tests using naive classification algorithms demonstrate the power of this methodology by achieving accuracy values higher than 0.8 in the discrimination of images containing both kind of retinal lesions.

In the last part of this thesis, the deep learning paradigm was explored by means of the convolutional neural networks. The most remarkable state-of-the-art CNN architectures were fine-tuned using retinal background patches

with the objective of detecting bright lesions. In general models based on CNNs improved the subjective hand-craft feature extraction and classification process for the task of detecting exudates. In particular, the most conventional networks (i.e. VGG) demonstrated the best performance in the detection task by achieving an AUC value higher than 0.9539.

The affordable cost and the portability of retinal cameras would facilitate the further incorporation of the methods presented in this thesis in a computer-aided diagnosis (CAD) system to discriminate between normal and pathological retinas taking into account DR and AMD. Both approaches presented in this work (i.e. hand-craft feature extraction plus classification and fine-tuned convolutional neural networks) were formulated in a locally way, in other words, classification models were trained using retinal background patches. This fact would provide to the CAD system high precision in the lesion identification and robustness against fundus images with tessellations, noise, random dust spots and brights characteristics from young retinas. On the other hand, the approaches of this PhD thesis are characterised by the simplicity and low computational costs thanks of avoiding lesion segmentation. When a new image will be acquired only with a simple prediction using the learned models, clinicians will received assistance to the interpretation of fundus images.

In future works, four main research lines will be developed. First, the retinal structure responsible to the central vision, the macula will be segmented. Using this segmentation and the localisation of the lesions provided by any of the approaches proposed in this thesis diabetic retinopathy and age-related macular degeneration will be discriminated. Secondly, regarding to the method based on image description and classification algorithms, the detection of the dark lesions must be improved. Additional information should be extracted from the retinal background in order to increase the precision in the identification of microaneurysms and haemorrhages. For example, colour information will be studied by exploring different colour spaces. Additionally, the global identification of abnormal retinal tissue will be improved. Information of the lesion localisation could be introduced composing a more discriminative feature vector than the simple histogram of probabilities. Moreover, more advanced classification algorithms will be tested in order to improve the image-wise detection of eye diseases. Finally and focusing on the proposed deep learning approach, engineered features will be combined with the automatic features extracted by the CNNs to determine if an improvement in the results would be possible. In addition, CNN architectures will be fine-tuned to automatically discriminate dark lesions. Furthermore, image-wise CNN models will be learned from scratch using a dataset created as a

combination of all the existing public fundus databases and data augmentation technique. As final reflection, the methodology proposed in this thesis requires the previous detection of the vascular tree and the optic disk. So even if no lesion segmentation is required, the entire workflow still relies on hand-crafted segmentation, which is also not error-free. An end-to-end approach to classify both lesions and anatomical elements is a further direction in which the candidate will focus his efforts.

Merits

Journal papers

- Morales, S., Engan, K., Naranjo, V., **Colomer, A.** (2015). Retinal Disease Screening through Local Binary Patterns. In *IEEE Journal of Biomedical and Health Informatics*, vol.21, no.1, pp.184-192.
- Colomer, A.**, Naranjo, V., Engan, K., Skretting, K. (2017). Assessment of sparse-based inpainting for retinal vessel removal. In *Signal Processing: Image Communication*, vol.59, pp.73-82.
- Colomer, A.**, Naranjo, V., Janvier, T., Mossi, J.M. (2018). Evaluation of Fractal Dimension effectiveness for damage detection in retinal background. In *Journal of Computational and Applied Mathematics* (Article In Press).

International conferences

- Morales, S.**, Engan, K., Naranjo, V., **Colomer, A.** (2015). Detection of Diabetic Retinopathy and Age-related Macular Degeneration from Fundus Images through Local Binary Patterns and Random Forests. In *IEEE International Conference on Image Processing (ICIP 2015)*, pp.4838-4842.
- Morales, S., Naranjo, V., **Colomer, A.**, Alcañiz, M. (2015). Significant point characterization in fundus images. In *5th International Conference on*

Image Processing Theory, Tools and Applications (IPTA 2015), pp.331-336.

Colomer, A., Naranjo, V., Engan, K., Skretting, K. (2015). Retinal vessel inpainting using recursive least square dictionary learning algorithm. In *5th International Conference on Image Processing Theory, Tools and Applications (IPTA 2015)*, pp.429-433.

Colomer, A., Naranjo, V., Janvier, T. (2016). Retinal Diseases Characterization using Fractal Analysis. In *18th Edition of the Mathematical Modelling in Engineering & Human Behaviour (MMEHB 2016)*, pp,1-4.

Colomer, A., Naranjo, V., Angulo, J. (2017). Colour normalization of fundus images based on geometric transformations applied to their chromatic histogram. In *24th IEEE International Conference on Image Processing (ICIP 2017)* (Article in Press).

National conferences

Rodrigo, B., Morales, S., Naranjo, V., **Colomer, A.**, Alcañiz, M. (2014). Detección de glaucoma mediante la combinación de la relación copa/disco y la regla ISNT. In *XXXII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2014)*.

Colomer, A., Torres, S., Naranjo, V., Morales, S. (2016). Segmentación de la fovea mediante morfología matemática en imágenes de fondo de ojo. In *XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2016)*, pp.379-382.

Bibliography

- Abbas, Q., Fondon, I., Sarmiento, A., Jiménez, S., and Alemany, P. (2017). “Automatic recognition of severity level for diagnosis of diabetic retinopathy using deep visual features”. In: *Medical & Biological Engineering & Computing* 55.11, pp. 1959–1974 (cit. on p. 21).
- Abràmoff, M. D., Folk, J. C., Han, D. P., and al., et (2013). “Automated analysis of retinal images for detection of referable diabetic retinopathy”. In: *JAMA Ophthalmology* 131.3, pp. 351–357 (cit. on pp. 19, 44, 62).
- Abràmoff, M. D., Garvin, M. K., and Sonka, M. (2010). “Retinal Imaging and Image Analysis”. In: *IEEE Reviews in Biomedical Engineering* 3, pp. 169–208 (cit. on pp. 14, 19).
- Abràmoff, M. D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J. C., and Niemeijer, M. (2016). “Improved Automated Detection of Diabetic Retinopathy on a Publicly Available Dataset Through Integration of Deep Learning”. In: *Investigative Ophthalmology Visual Science* 57.13, p. 5200 (cit. on p. 19).
- Age-Related Eye Disease Study Research Group (2001). “A Randomized, Placebo-Controlled, Clinical Trial of High-Dose Supplementation With Vitamins C and E, Beta Carotene, and Zinc for Age-Related Macular Degeneration and Vision Loss: AREDS Report No. 8”. In: *Archives of Ophthalmology* 119.10, pp. 1417–1436 (cit. on p. 18).

- Agurto, C., Barriga, E. S., Murray, V., Nemeth, S., Crammer, R., Bauman, W., Zamora, G., Pattichis, M. S., and Soliz, P. (2011a). “Automatic Detection of Diabetic Retinopathy and Age-Related Macular Degeneration in Digital Fundus Images”. In: *Investigative Ophthalmology & Visual Science* 52.8, pp. 5862–5871 (cit. on p. 20).
- Agurto, C., Barriga, E. S., Zamora, G., Murray, V., Murillo, S., Yu, H., Bauman, W. C., and Soliz, P. (2011b). “Automatic Screening Of Eye Diseases Using 3-field Fundus Photographs”. In: *Investigative Ophthalmology & Visual Science* 52.6, p. 1342 (cit. on p. 20).
- Agurto, C., Yu, H., Murray, V., Pattichis, M. S., Barriga, S., Bauman, W., and Soliz, P. (2012). “Detection of neovascularization in the optic disc using an AM-FM representation, granulometry, and vessel segmentation”. In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4946–4949 (cit. on p. 73).
- Aharon, M., Elad, M., and Bruckstein, A. (2006). “The K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation”. In: *Trans. Sig. Proc.* 54.11, pp. 4311–4322 (cit. on p. 38).
- Ahonen, T., Hadid, A., and Pietikainen, M. (2006). “Face Description with Local Binary Patterns: Application to Face Recognition”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.12, pp. 2037–2041 (cit. on p. 58).
- Akram, M. U., Khalid, S., Tariq, A., Khan, S. A., and Azam, F. (2014a). “Detection and classification of retinal lesions for grading of diabetic retinopathy”. In: *Computers in Biology and Medicine* 45, pp. 161–171 (cit. on p. 55).
- Akram, M. U., Tariq, A., Khan, S. A., and Javed, M. Y. (2014b). “Automated detection of exudates and macula for grading of diabetic macular edema”. In: *Computer Methods and Programs in Biomedicine* 114.2, pp. 141–152 (cit. on p. 55).
- Aliaa, Y., Atef, G., and Amr, G. (2007). “A Comparative Evaluation of Preprocessing Methods for Automatic Detection of Retinal Anatomy”. In: *The 5th International Conference on Informatics and Systems (INFOS2007)*, pp. 24–30 (cit. on p. 28).

- Amel, F., Mohammed, M., and Abdelhafid, B. (2012). “Improvement of the Hard Exudates Detection Method Used For Computer- Aided Diagnosis of Diabetic Retinopathy”. In: *International Journal of Image, Graphics and Signal Processing* 4.4, pp. 19–27 (cit. on p. 55).
- American Academy of Ophthalmology (2015). *Preferred Practice Pattern® Guidelines. Age-Related Macular Degeneration*. Available at: www.aao.org/ppp. San Francisco, CA (cit. on pp. 17, 18).
- (2016). *Preferred Practice Pattern® Guidelines. Diabetic Retinopathy*. Available at: www.aao.org/ppp. San Francisco, CA (cit. on p. 17).
- Aquino, A., Gegundez-Arias, M.E., and Marin, D. (2010). “Detecting the Optic Disc Boundary in Digital Fundus Images Using Morphological, Edge Detection, and Feature Extraction Techniques”. In: *Medical Imaging, IEEE Transactions on* 29.11, pp. 1860–1869 (cit. on p. 21).
- Ashraf, M. N., Habib, Z., and Hussain, M. (2014). “Texture Feature Analysis of Digital Fundus Images for Early Detection of Diabetic Retinopathy”. In: *2014 11th International Conference on Computer Graphics, Imaging and Visualization*, pp. 57–62 (cit. on p. 136).
- Azemin, M. Z. C., Kumar, D. K., Wong, T. Y., Kawasaki, R., Mitchell, P., and Wang, J. J. (2011). “Robust Methodology for Fractal Analysis of the Retinal Vasculature”. In: *IEEE Transactions on Medical Imaging* 30.2, pp. 243–250 (cit. on p. 65).
- Barton, R. J. and Poor, H. V. (1988). “Signal detection in fractional Gaussian noise”. In: *IEEE Transactions on Information Theory* 34.5, pp. 943–959 (cit. on p. 67).
- Bengio, Y. (2012). “Deep Learning of Representations for Unsupervised and Transfer Learning”. In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*. Vol. 27. Proceedings of Machine Learning Research. Bellevue, Washington, USA: PMLR, pp. 17–36 (cit. on p. 111).
- Bertalmio, M., Bertozzi, A. L., and Sapiro, G. (2001). “Navier-Stokes, fluid dynamics, and image and video inpainting”. In: *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pp. 355–362 (cit. on p. 34).

- Bishop, C. (2010). *Pattern Recognition and Machine Learning*. Inglés. Edición: 1st ed. 2006. Corr. 2nd printing 2011. New York: Springer (cit. on pp. 90, 91, 93).
- Breiman, L. (1996). “Bagging Predictors”. In: *Machine Learning* 24.2, pp. 123–140 (cit. on pp. 81, 82).
- (2001). “Random Forests”. In: *Machine Learning* 45.1, pp. 5–32 (cit. on pp. 83, 131).
- Bruno, O. M., Oliveira-Plotze, R. de, Falvo, M., and Castro, M. de (2008). “Fractal Dimension Applied to Plant Identification”. In: *Inf. Sci.* 178.12, pp. 2722–2733 (cit. on p. 65).
- Chang, C. C. and Lin, C. J. (2011). “LIBSVM: A Library for Support Vector Machines”. In: *ACM Trans. Intell. Syst. Technol.* 2.3, 27:1–27:27 (cit. on pp. 85, 88, 132).
- Cherni, M. A., Sayadi, M., and Fnaiech, F. (2009). “Using granulometry and watershed for breast tumor cells segmentation”. In: *2009 3rd International Conference on Signals, Circuits and Systems (SCS)*, pp. 1–5 (cit. on p. 73).
- Chollet, F. et al. (2015). *Keras*. <https://github.com/fchollet/keras> (cit. on p. 123).
- Chollet, F. (2016). “Xception: Deep Learning with Depthwise Separable Convolutions”. In: *CoRR* abs/1610.02357 (cit. on pp. 114, 115, 142).
- Colomer, A., Naranjo, V., and Angulo, J. (2017a). “Colour normalization of fundus images based on geometric transformations applied to their chromatic histogram”. In: *2017 IEEE International Conference on Image Processing (ICIP)*. Article in Press (cit. on pp. 5, 21).
- Colomer, A., Naranjo, V., Engan, K., and Skretting, K. (2017b). “Assessment of sparse-based inpainting for retinal vessel removal”. In: *Signal Processing: Image Communication* 59.Supplement C, pp. 73–82 (cit. on pp. 5, 21).
- Colomer, A., Naranjo, V., Janvier, T., and Mossi, J. M. (2018). “Evaluation of Fractal Dimension effectiveness for damage detection in retinal

- background”. In: *Journal of Computational and Applied Mathematics*. Article in Press (cit. on pp. 6, 21).
- Colomé-Serra, M. F., Serón, D., Moreso, F., Carrera, M., Griñó, J. M., and Serra, J. (1992). “Image analysis: Utility of grey level granulometry to measure renal interstitial chronic damage”. In: *1992 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Vol. 5, pp. 1934–1935 (cit. on p. 73).
- Costa, P., Galdran, A., Meyer, M. I., Niemeijer, M., Abràmoff, M., Mendonça, A. M., and Campilho, A. (2017). “End-to-end Adversarial Retinal Image Synthesis”. In: *IEEE Transactions on Medical Imaging* PP.99, pp. 1–1 (cit. on p. 19).
- Criminisi, A., Pérez, P., and Toyama, K. (2004). “Region filling and object removal by exemplar-based image inpainting”. In: *IEEE Transactions on Image Processing* 13, pp. 1200–1212 (cit. on pp. 44, 48–52).
- Dang, T. T., Beghdadi, A., and Larabi, M. C. (2013a). “Perceptual quality assessment for color image inpainting”. In: *IEEE International Conference on Image Processing*, pp. 398–402 (cit. on p. 34).
- (2013b). “Visual coherence metric for evaluation of color image restoration”. In: *IEEE Colour and Visual Computing Symposium*, pp. 1–6 (cit. on p. 34).
- Dean, J. et al. (2012). “Large Scale Distributed Deep Networks”. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. NIPS’12. Lake Tahoe, Nevada: Curran Associates Inc., pp. 1223–1231 (cit. on p. 98).
- Decencière, E., Cazuguel, G., Zhang, X., Thibault, G., et al. (2013). “TeleOphta: Machine learning and image processing methods for teleophthalmology”. In: *IRBM* 34.2, pp. 196–203 (cit. on pp. 19, 26, 121).
- Deering, W. and West, B. J. (1992). “Fractal physiology”. In: *IEEE Engineering in Medicine and Biology Magazine* 11.2, pp. 40–46 (cit. on p. 65).

- D'Errico, J. (2004). *Inpainting nans*. <http://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint-nans>. Last accessed on 12th January 2016 (cit. on p. 49).
- Doria, D. and Radke, R. J. (2012). “Filling large holes in LiDAR data by inpainting depth gradients”. In: *CVPR Workshops*. IEEE, pp. 65–72 (cit. on p. 34).
- Dougherty, G. and Henebry, G. M. (2001). “Fractal signature and lacunarity in the measurement of the texture of trabecular bone in clinical CT images.” In: *Medical Engineering and Physics* 23.6, pp. 369–380 (cit. on p. 65).
- Eke, A., Herman, P., Kocsis, L., and Kozak, L. R. (2002). “Fractal characterization of complexity in temporal physiological signals”. In: *Physiological Measurement* 23.1, R1 (cit. on p. 68).
- Elad, M. (2010). *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer (cit. on pp. 34, 41).
- Engan, K., Aase, S. O., and Hakon-Husoy, J. (1999). “Method of Optimal Directions for Frame Design”. In: *Proceedings of the Acoustics, Speech, and Signal Processing, 1999. On 1999 IEEE International Conference - Volume 05*. ICASSP '99. IEEE Computer Society, pp. 2443–2446 (cit. on p. 38).
- Engan, K., Skretting, K., and Hakon-Husøy, J. (2007). “Family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation”. In: *Digital Signal Processing* 17.1, pp. 32–49 (cit. on p. 37).
- Flandrin, P. (1992). “Wavelet analysis and synthesis of fractional Brownian motion”. In: *IEEE Transactions on Information Theory* 38.2, pp. 910–917 (cit. on p. 68).
- Fondón, I., Núñez, F., Tirado, M., Jiménez, S., Alemany, P., Abbas, Q., Serrano, C., and Acha, B. (2012). “Automatic Cup-to-Disc Ratio Estimation Using Active Contours and Color Clustering in Fundus Images for Glaucoma Diagnosis”. In: *Image Analysis and Recognition: 9th International Conference, ICIAR 2012, Aveiro, Portugal, June 25-*

- 27, 2012. *Proceedings, Part II*. Springer Berlin Heidelberg, pp. 390–399 (cit. on p. 21).
- Fong, D. S., Aiello, L. P., Ferris, F. L., and Klein, R. (2004). “Diabetic Retinopathy”. In: *Diabetes Care* 27.10, pp. 2540–2553 (cit. on p. 16).
- Galloway, N. R., Amoaku, Winfried M. K., Galloway, P. H., and Browning, A. C. (2016). “Basic Anatomy and Physiology of the Eye”. In: *Common Eye Diseases and their Management*. Springer International Publishing, pp. 7–16 (cit. on p. 11).
- García Seco de Herrera, A., Schaer, R., Bromuri, S., and Müller, H. (2016). “Overview of the ImageCLEF 2016 medical task”. In: *Working Notes of CLEF 2016 (Cross Language Evaluation Forum)*. Évora, Portugal (cit. on p. 99).
- García, M., Sánchez, C. I., López, M. I., Abásolo, D., and Hornero, R. (2009). “Neural network based detection of hard exudates in retinal images”. In: *Computer Methods and Programs in Biomedicine* 93.1, pp. 9–19 (cit. on p. 21).
- Garnier, M., Hurtut, T., Tahar, H. B., and Cheriet, F. (2014). “Automatic multiresolution age-related macular degeneration detection from fundus images”. In: vol. 9035, pp. 9035–9042 (cit. on pp. 20, 58).
- Ghafourian, M. and Pourreza, H. (2012). “Localization of Hard Exudates in Retinal Fundus Image by Mathematical Morphology Operations”. In: *2nd International eConference on Computer and Knowledge Engineering*, pp. 185–189 (cit. on pp. 55, 134, 135).
- Giachetti, A., Chin, K. S., Trucco, E., Cobb, C., and Wilson, P. J. (2011). “Multiresolution localization and segmentation of the optical disc in fundus images using inpainted background and vessel information”. In: *Image Processing, IEEE International Conference on*, pp. 2145–2148 (cit. on p. 35).
- Giancardo, L., Meriaudeau, F., Karnowski, T. P., Li, Y., Garg, S., Tobin, K. W., and Chaum, E. (2012). “Exudate-based diabetic macular edema detection in fundus images using publicly available datasets”. In: *Medical Image Analysis* 16.1, pp. 216–226 (cit. on p. 55).

- Goatman, K., Charnley, A., Webster, L., and Nussey, S. (2011). “Assessment of Automated Disease Detection in Diabetic Retinopathy Screening Using Two-Field Photography”. In: *PLoS ONE* 6.12, e27524 (cit. on p. 20).
- Grinsven, M. J. J. P. van, Ginneken, B. van, Hoyng, C. B., Theelen, T., and Sánchez, C. I. (2016). “Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images”. In: *IEEE Transactions on Medical Imaging* 35.5, pp. 1273–1284 (cit. on p. 100).
- Guillemot, C. and Le Meur, O. (2014). “Image inpainting: Overview and recent advances”. In: *IEEE Signal Processing Magazine* 31.1, pp. 127–144 (cit. on p. 34).
- Gulshan, V., Peng, L., Coram, M., and al., et (2016). “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs”. In: *JAMA* 316.22, pp. 2402–2410 (cit. on pp. 20, 99).
- Guo, Z., Zhang, L., and Zhang, D. (2010). “Rotation invariant texture classification using {LBP} variance (LBPV) with global matching”. In: *Pattern Recognition* 43.3, pp. 706–719 (cit. on p. 62).
- Haloi, M. (2015). “Improved Microaneurysm Detection using Deep Neural Networks”. In: *CoRR* abs/1505.04424 (cit. on p. 100).
- Harrar, K., Hamami, L., Lespessailles, E., and Jennane, R. (2013). “Piecewise Whittle estimator for trabecular bone radiograph characterization”. In: *Biomedical Signal Processing and Control* 8.6, pp. 657–666 (cit. on p. 65).
- Hatami, N. and Goldbaum, M. H. (2016). “Automatic Identification of Retinal Arteries and Veins in Fundus Images using Local Binary Patterns”. In: *CoRR* abs/1605.00763 (cit. on p. 58).
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep Residual Learning for Image Recognition”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778 (cit. on pp. 113, 142).

- Heikkilä, M., Pietikäinen, M., and Schmid, C. (2009). “Description of Interest Regions with Local Binary Patterns”. In: *Pattern Recogn.* 42.3, pp. 425–436. ISSN: 0031-3203 (cit. on p. 58).
- Hoover, A., Kouznetsova, V., and Goldbaum, M. (2000). “Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response”. In: *IEEE Transactions on Medical Imaging* 19, pp. 203–210 (cit. on p. 27).
- Huang, H. Y. and Hsiao, C. N. (2010). “A patch-based image inpainting based on structure consistence”. In: *Computer Symposium (ICS), 2010 International*. IEEE, pp. 165–170 (cit. on p. 34).
- Huang, J. and Ling, C. X. (2005). “Using AUC and accuracy in evaluating learning algorithms”. In: *IEEE Transactions on Knowledge and Data Engineering* 17.3, pp. 299–310 (cit. on p. 123).
- Hubel, D. H. and Wiesel, T. N. (1962). “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”. In: *J Physiol.* 160.1, pp. 106–154.2 (cit. on p. 97).
- Hurst, H. E., Black, R. P., and Simaika, Y. M. (1965). *Long-term storage : an experimental study / by H.E. Hurst, R.P. Black, Y.M. Simaika*. English. Constable London, 145 p. : (cit. on p. 67).
- Ioffe, S. and Szegedy, C. (2015). “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: *CoRR* abs/1502.03167 (cit. on p. 109).
- Jennane, R., Ohley, W. J., Majumdar, S., and Lemineur, G. (2001). “Fractal Analysis of Bone X-Ray Tomographic Microscopy Projections”. In: *IEEE Trans. Med. Imaging* 20.5, pp. 443–449 (cit. on p. 65).
- Jiménez, S., Alemany, P., Fondón, I., Foncubierta, A., Acha, B., and Serrano, C. (2010). “Automatic detection of vessels in color fundus images”. In: *Archivos de la Sociedad Española de Oftalmología (English Edition)* 85.3, pp. 103–109 (cit. on p. 21).

- Jolliffe, I. T. (1986). “Principal Component Analysis and Factor Analysis”. In: *Principal component analysis*. Springer, pp. 115–128 (cit. on p. 138).
- Junior, S. B. and Welfer, D. (2013). “Automatic Detection of Microaneurysms and Hemorrhages in Color Eye Fundus Images”. In: *International Journal of Computer Science & Information Technology* 5.5, pp. 21–37 (cit. on p. 136).
- Kauppi, T., Kalesnykiene, V., Kamarainen, J. K., Sorri, L. Lensu I., Raninen, A., Voutilainen, R., et al. (2007). “DIARETDB1 diabetic retinopathy database and evaluation protocol”. In: *Proc. Medical Image Understanding and Analysis* (cit. on pp. 121, 140, 141).
- Keerthi, S. S. and Lin, C. J. (2003). “Asymptotic Behaviors of Support Vector Machines with Gaussian Kernel”. In: *Neural Computation* 15.7, pp. 1667–1689 (cit. on p. 88).
- Khatun, M., Gray, A., and Marshall, S. (2011a). “Classification of Ordered Texture Images Using Regression Modelling and Granulometric Features”. In: *2011 Irish Machine Vision and Image Processing Conference*, pp. 64–69 (cit. on p. 72).
- (2011b). “Morphological granulometry for classification of evolving and ordered texture images”. In: *2011 19th European Signal Processing Conference*, pp. 759–763 (cit. on p. 72).
- Kotu, L. P., Engan, K., Eftestøl, T., Woie, L., Ørn, S., and Katsaggelos, A. K. (2012). “Local Binary Patterns used on Cardiac MRI to classify high and low risk patient groups”. In: *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pp. 2586–2590 (cit. on p. 58).
- Krizhevsky, A., Sutskever, I., and Hinton, G. E (2012). “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., pp. 1097–1105 (cit. on pp. 98, 112, 142).
- Kumar, A., Kim, J., Lyndon, D., Fulham, M., and Feng, D. (2017). “An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification”. In: *IEEE Journal of Biomedical and Health Informatics* 21.1, pp. 31–40 (cit. on p. 98).

- Kupidura, P. and Uwarowa, I. (2017). “The comparison of GLCM and granulometry for distinction of different classes of urban area”. In: *2017 Joint Urban Remote Sensing Event (JURSE)*, pp. 1–4 (cit. on p. 72).
- Landini, G. (2011). “Fractals in microscopy”. In: *Journal of Microscopy* 241.1, pp. 1–8 (cit. on p. 65).
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). “Gradient-Based Learning Applied to Document Recognition”. In: *Proceedings of the IEEE*. Vol. 86. 11, pp. 2278–2324 (cit. on p. 106).
- Lecun, Y. et al. (1995). “Learning Algorithms For Classification: A Comparison On Handwritten Digit Recognition”. In: *Neural Networks: The Statistical Mechanics Perspective*. World Scientific, pp. 261–276 (cit. on p. 98).
- Liang, L., Liu, C., Xu, Y. Q., Guo, Baining, and Shum, H. Y. (2001). “Real-time Texture Synthesis by Patch-based Sampling”. In: *ACM Trans. Graph.* 20.3, pp. 127–150 (cit. on p. 34).
- Liu, L., Lao, S., Fieguth, P. W., Guo, Y., Wang, X., and Pietikäinen, M. (2016). “Median Robust Extended Local Binary Pattern for Texture Classification”. In: *IEEE Transactions on Image Processing* 25.3, pp. 1368–1381 (cit. on p. 58).
- MacGillivray, T. J., Patton, N., Doubal, F. N., Graham, C., and Wardlaw, J. M. (2007). “Fractal analysis of the retinal vascular network in fundus images”. In: *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6455–6458 (cit. on p. 65).
- Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2009). “Online Dictionary Learning for Sparse Coding”. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, pp. 689–696 (cit. on p. 47).
- Mairal, J., Bach, F., Ponce, J., Sapiro, G., Jenatton, R., and Obozinski, G. (2014). *SPAMS v2.5 (SParse Modeling Software)*. <http://spams-devel.gforge.inria.fr/index.html> (cit. on pp. 41, 45).

- Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. (2008a). “Discriminative learned dictionaries for local image analysis.” In: *CVPR*. IEEE Computer Society (cit. on pp. 34, 45).
- Mairal, J., Elad, M., and Sapiro, G. (2008b). “Sparse representation for color image restoration”. In: *Image Processing, IEEE Transactions on* 17.1, pp. 53–69 (cit. on pp. 39, 47).
- Mandelbrot, B. B. and Van Ness, J. W. (1968). “Fractional Brownian Motions, Fractional Noises and Applications”. In: *SIAM Review* 10.4, pp. 422–437 (cit. on p. 67).
- Mandrekar, J. N. (2010). “Receiver Operating Characteristic Curve in Diagnostic Test Assessment”. In: *Journal of Thoracic Oncology* 5.9, pp. 1315–1316 (cit. on p. 123).
- Marin, D., Aquino, A., Gegundez-Arias, M.E., and Bravo, J.M. (2011). “A New Supervised Method for Blood Vessel Segmentation in Retinal Images by Using Gray-Level and Moment Invariants-Based Features”. In: *Medical Imaging, IEEE Transactions on* 30.1, pp. 146–158 (cit. on p. 21).
- Marrugo, A. and Millán, M. San (2011). “Retinal image analysis: preprocessing and feature extraction”. In: *Journal of Physics: Conference Series* 274.1, p. 012039 (cit. on p. 28).
- Massey, E. M., Lowell, J., Hunter, A., and Steel, D. (2009). “Lesion Boundary Segmentation using Level Set Methods.” In: *VISAPP - Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*. Ed. by Alpesh Ranchordas and Helder Araújo. INSTICC Press, pp. 245–249 (cit. on p. 20).
- Matheron, G. (1967). *Elements pour une théorie des milieux poreux*. Paris: Masson (cit. on p. 71).
- (1975). *Random sets and integral geometry*. Wiley series in probability and mathematical statistics: Probability and mathematical statistics. Wiley (cit. on p. 71).

- Meier, J., Bock, R., Michelson, G., Nyúl, L. G., and Hornegger, J. (2007). “Effects of Preprocessing Eye Fundus Images on Appearance Based Glaucoma Classification”. In: *Computer Analysis of Images and Patterns: 12th International Conference, CAIP 2007, Vienna, Austria, August 27-29, 2007. Proceedings*, pp. 165–172 (cit. on pp. 28, 35).
- Moin, M. S., Tavakoli, H. R., and Broumandnia, A. (2010). “A new retinal vessel segmentation method using preprocessed Gabor and local binary patterns”. In: *2010 6th Iranian Conference on Machine Vision and Image Processing*, pp. 1–6 (cit. on p. 58).
- Mookiah, M. R. K., Acharya, U. R., Martis, R. J., Chua, C. K., Lim, C. M., Ng, E. Y. K., and Laude, A. (2013). “Evolutionary algorithm based classifier parameter tuning for automatic diabetic retinopathy grading: A hybrid feature extraction approach”. In: *Knowledge-Based Systems* 39, pp. 9–22 (cit. on pp. 55, 58).
- Morales, S., Engan, K., Naranjo, V., and Colomer, A. (2017). “Retinal Disease Screening Through Local Binary Patterns”. In: *IEEE Journal of Biomedical and Health Informatics* 21.1, pp. 184–192 (cit. on pp. 21, 58).
- Morales, S., Naranjo, V., Angulo, J., and Alcañiz, M. (2013). “Automatic Detection of Optic Disc Based on PCA and Mathematical Morphology”. In: *Medical Imaging, IEEE Transactions on* 32.4, pp. 786–796 (cit. on pp. 21, 35, 56).
- Morales, S., Naranjo, V., Navea, A., and Alcañiz, M. (2014). “Computer-Aided Diagnosis Software for Hypertensive Risk Determination Through Fundus Image Processing”. In: *IEEE Journal of Biomedical and Health Informatics* 18.6, pp. 1757–1763 (cit. on pp. 21, 131).
- Moreno, R., Borga, M., and Smedby, Ö. (2012). “Estimation of trabecular thickness in gray-scale images through granulometric analysis”. In: vol. 8314, pp. 8314–8323 (cit. on p. 73).
- Mudigonda, S., Oloumi, F., Katta, K. M., and Rangayyan, R. M. (2015). “Fractal analysis of neovascularization due to diabetic retinopathy in retinal fundus images”. In: *E-Health and Bioengineering Conference (EHB), 2015*, pp. 1–4 (cit. on p. 65).

- Nair, V. and Hinton, G. E. (2010). “Rectified Linear Units Improve Restricted Boltzmann Machines”. In: *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. Omnipress, pp. 807–814 (cit. on p. 103).
- Nanni, L., Lumini, A., and Brahnam, S. (2010). “Local binary patterns variants as texture descriptors for medical image analysis”. In: *Artificial Intelligence in Medicine* 49.2, pp. 117–125 (cit. on p. 58).
- Naresh, S. and Kumari, S. V. (2015). “Article: Breast Cancer Detection using Local Binary Patterns”. In: *International Journal of Computer Applications* 123.16. Published by Foundation of Computer Science (FCS), NY, USA, pp. 6–9 (cit. on p. 58).
- Nawarathna, R., Oh, J., Muthukudage, J., Tavanapong, W., Wong, J., Groen, P. C. de, and Tang, S. J. (2014). “Abnormal image detection in endoscopy videos using a filter bank and local binary patterns”. In: *Neurocomputing* 144, pp. 70–91 (cit. on p. 58).
- Nielsen, M. A. (2015). *Neural networks and deep learning*. <http://neuralnetworksanddeeplearning.com/>. Last accessed on 31th August 2017 (cit. on p. 101).
- Niemeijer, M. et al. (2010). “Retinopathy Online Challenge: Automatic Detection of Microaneurysms in Digital Color Fundus Photographs”. In: *IEEE Transactions on Medical Imaging* 29.1, pp. 185–195 (cit. on pp. 19, 21).
- Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.7, pp. 971–987 (cit. on pp. 57, 58, 60, 61).
- Ortega, M., Barreira, N., Novo, J., Penedo, M.G., Pose-Reino, A., and Gómez-Ulla, F. (2010). “Sirius: A web-based system for retinal image analysis”. In: *International Journal of Medical Informatics* 79.10, pp. 722–732 (cit. on p. 21).

- Pan, Z., Wu, X., Li, Z., and Zhou, Z. (2017). “Local Adaptive Binary Patterns Using Diamond Sampling Structure for Texture Classification”. In: *IEEE Signal Processing Letters* 24.6, pp. 828–832 (cit. on p. 58).
- Parisi, G. (1998). *Statistical Field Theory*. Inglés. Edición: New edition. Reading, Mass: Perseus Books (cit. on p. 91).
- Pesquet-Popescu, B. and Lévy-Véhel, J. (2002). “Stochastic Fractal Models for Image Processing”. In: *IEEE Signal Processing Magazine* 19.5, pp. 48–62 (cit. on p. 69).
- Philip, S., Fleming, A. D., Goatman, K. A., Fonseca, S., Mcnamee, P., Scotland, G. S., Prescott, G. J., Sharp, P. F., and Olson, J. A. (2007). “The efficacy of automated “disease/no disease” grading for diabetic retinopathy in a systematic screening programme”. In: *British Journal of Ophthalmology* 91.11, pp. 1512–1517 (cit. on p. 20).
- Phothisonothai, M. and Nakagawa, M. (2008). “EEG signal classification method based on fractal features and neural network”. In: *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 3880–3883 (cit. on p. 65).
- Pietikäinen, M., Ojala, T., and Xu, Z. (2000). “Rotation-invariant texture classification using feature distributions”. In: *Pattern Recognition* 33.1, pp. 43–52 (cit. on p. 60).
- Prati, R. C., Batista, G. E., and Monard, M. C. (2011). “A Survey on Graphical Methods for Classification Predictive Performance Evaluation”. In: *IEEE Transactions on Knowledge and Data Engineering* 23.11, pp. 1601–1618 (cit. on p. 123).
- Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., and Zheng, Y. (2016). “Convolutional Neural Networks for Diabetic Retinopathy”. In: *Procedia Computer Science* 90.Supplement C. 20th Conference on Medical Image Understanding and Analysis (MIUA 2016), pp. 200–205 (cit. on p. 99).
- Prentašić, Pavle and Lončarić, Sven (2016). “Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion”. In: *Computer Methods and Programs in Biomedicine* 137, pp. 281–292 (cit. on p. 99).

- Prentašić, P. et al. (2013). “Diabetic retinopathy image database (DRiDB): A new database for diabetic retinopathy screening programs research”. In: *2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 711–716 (cit. on p. 99).
- Quelleg, G., Lamard, M., Abràmoff, M. D., Decencièrè, E., Lay, B., Erginay, A., Cochener, B., and Cazuguel, G. (2012). “A multiple-instance learning framework for diabetic retinopathy screening”. In: *Medical Image Analysis* 16.6, pp. 1228–1240 (cit. on p. 20, 55).
- Quelleg, G., Lamard, M., Cazuguel, G., Bekri, L., Daccache, W., Roux, C., and Cochener, B. (2011). “Automated Assessment of Diabetic Retinopathy Severity Using Content-Based Image Retrieval in Multimodal Fundus Photographs”. In: *Investigative Ophthalmology Visual Science* 52.11, p. 8342 (cit. on p. 20).
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Adaptive computation and machine learning. Cambridge, Mass: MIT Press (cit. on p. 90).
- Rathinam, S. and Selvarajan, S. (2013). “Comparison of Image Processing Techniques on Fundus Images for Early Diagnosis of Glaucoma”. In: *International Journal of Scientific & Engineering Research* 4.12, pp. 1368–1372 (cit. on p. 28).
- Reed, I. S., Lee, P. C., and Truong, T. K. (1995). “Spectral representation of fractional Brownian motion in n dimensions and its properties”. In: *IEEE Transactions on Information Theory* 41.5, pp. 1439–1451 (cit. on p. 67).
- Rocha, A., Carvalho, T., Jelinek, H. F., Goldenstein, S., and Wainer, J. (2012). “Points of Interest and Visual Dictionaries for Automatic Retinal Lesion Detection”. In: *IEEE Transactions on Biomedical Engineering* 59.8, pp. 2244–2253 (cit. on p. 136).
- Roychowdhury, S., Koozekanani, D. D., and Parhi, K. K. (2012). “Screening fundus images for diabetic retinopathy”. In: *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pp. 1641–1645 (cit. on p. 136).

- Ruberto, C. Di, Dempster, A., Khan, S., and Jarra, B. (2000). “Automatic thresholding of infected blood images using granulometry and regional extrema”. In: *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*. Vol. 3, 441–444 vol.3 (cit. on p. 73).
- Sánchez, C. I., García, M., Mayo, A., López, M. I., and Hornero, R. (2009). “Retinal image analysis based on mixture models to detect hard exudates”. In: *Medical Image Analysis* 13.4, pp. 650–658 (cit. on p. 21).
- Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., and LeCun, Y. (2013). “OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks”. In: *CoRR* abs/1312.6229 (cit. on p. 112).
- Serra, J. (1983). *Image Analysis and Mathematical Morphology*. Orlando, FL, USA: Academic Press, Inc. (cit. on p. 71).
- Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R. M. (2016). “Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning”. In: *IEEE Transactions on Medical Imaging* 35.5, pp. 1285–1298 (cit. on p. 98).
- Simonyan, K. and Zisserman, A. (2014). “Two-Stream Convolutional Networks for Action Recognition in Videos”. In: *CoRR* abs/1406.2199 (cit. on p. 112).
- Skretting, K. and Engan, K. (2010). “Recursive least squares dictionary learning algorithm”. In: *Signal Processing, IEEE Transactions on* 58 (4), pp. 2121–2130 (cit. on pp. 37, 44).
- Soille, P. (2003). *Morphological Image Analysis: Principles and Applications*. 2nd ed. Secaucus, NJ, USA: Springer-Verlag New York, Inc. (cit. on p. 75).
- Sopharak, A., Uyyanonvara, B., Barman, S., and Williamson, T. H. (2008). “Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods”. In: *Computerized Medical Imaging and Graphics* 32.8, pp. 720–727 (cit. on pp. 55, 134).

- Sørensen, L., Shaker, S. B., and Bruijne, M. de (2008). “Texture Classification in Lung CT Using Local Binary Patterns”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008: 11th International Conference, New York, NY, USA, September 6-10, 2008, Proceedings, Part I*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 934–941 (cit. on p. 58).
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *J. Mach. Learn. Res.* 15.1, pp. 1929–1958 (cit. on pp. 100, 109).
- Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., and Ginneken, B. van (2004). “Ridge-based vessel segmentation in color images of the retina.” In: *IEEE Trans. Med. Imaging* 23.4, pp. 501–509 (cit. on p. 26).
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). “Going deeper with convolutions”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9 (cit. on pp. 114, 142).
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J. (2016). “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?” In: *IEEE Transactions on Medical Imaging* 35.5, pp. 1299–1312 (cit. on pp. 98, 111).
- Tapia, S. L., Molina, R., and Blanca, N. P. de la (2016). “Detection and localization of objects in Passive Millimeter Wave Images”. In: *2016 24th European Signal Processing Conference (EUSIPCO)*, pp. 2101–2105 (cit. on p. 81).
- Telea, A. (2004). “An Image Inpainting Technique Based on the Fast Marching Method.” In: *J. Graphics, GPU, & Game Tools* 9.1, pp. 23–34 (cit. on p. 34).
- Tibshirani, R. (1996). “Regression shrinkage and selection via the lasso”. In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288 (cit. on p. 41).

- Tschumperlé, D. (2006). “Fast Anisotropic Smoothing of Multi-Valued Images using Curvature-Preserving PDE’s”. In: *International Journal of Computer Vision* 68.1, pp. 65–82 (cit. on p. 34).
- Voronin, V., Marchuk, V., Semenishchev, E., Maslennikov, S., and Svirin, I. (2015). “Inpainted image quality assessment based on machine learning”. In: *Conference on Computer Graphics, Visualization and Computer Vision* (cit. on p. 34).
- Walter, T., Klein, J. C., Massin, P., and Erginay, A. (2002). “A Contribution of Image Processing to the Diagnosis of Diabetic Retinopathy - Detection of Exudates in Color Fundus Images of the Human Retina.” In: *IEEE Trans. Med. Imaging* 21.10, pp. 1236–1243 (cit. on pp. 44, 55, 62, 134).
- Wang, H., Hsu, W., Goh, K. G., and Lee, M. L. (2000). “An effective approach to detect lesions in color retinal images”. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*. Vol. 2, 181–186 vol.2 (cit. on p. 16).
- Wang, S., Li, H., Zhu, X., and Li, P. (2008). “An Evaluation Index Based on Parameter Weight for Image Inpainting Quality”. In: *Young Computer Scientists, 2008. The 9th International Conference for. IEEE*, pp. 786–790 (cit. on pp. 35, 43).
- Wang, Z. and Yang, J. (2017). “Diabetic Retinopathy Detection via Deep Convolutional Networks for Discriminative Localization and Visual Explanation”. In: *CoRR* abs/1703.10757 (cit. on p. 99).
- Welfer, D., Scharcanski, J., and Marinho, D. R. (2010). “A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images.” In: *Computerized Medical Imaging and Graphics* 34.3, pp. 228–235 (cit. on pp. 55, 134, 135).
- Witten, I. H., Frank, E., and Hall, M. A. (2011). “Front Matter”. In: *Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)*. Third Edition. The Morgan Kaufmann Series in Data Management Systems. Boston: Morgan Kaufmann, pp. i–iii (cit. on p. 82).
- Wong, S. C., Gatt, A., Stamatescu, V., and McDonnell, M. D. (2016). “Understanding Data Augmentation for Classification: When to Warp?” In:

2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp. 1–6 (cit. on p. 108).

World Health Organization (2015). *Universal eye health: a global action plan 2014-2019*. Tech. rep. (cit. on pp. 3, 14, 15).

Yang, Y., Li, T., Li, W., Wu, H., Fan, W., and Zhang, W. (2017). “Lesion Detection and Grading of Diabetic Retinopathy via Two-Stages Deep Convolutional Neural Networks”. In: *Medical Image Computing and Computer-Assisted Intervention MICCAI 2017*. Cham: Springer International Publishing, pp. 533–540 (cit. on p. 100).

Ye, G. and Liu, C. (2008). “Research on Granulometry Approach to Measuring Microbubbles Size Distribution”. In: *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–4 (cit. on p. 72).

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). “How Transferable Are Features in Deep Neural Networks?” In: *Proceedings of the 27th International Conference on Neural Information Processing Systems. NIPS’14*. Montreal, Canada: MIT Press, pp. 3320–3328 (cit. on p. 111).

Yu, S., Xiao, D., and Kanagasigam, Y. (2017). “Exudate detection for diabetic retinopathy with convolutional neural networks”. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1744–1747 (cit. on p. 99).

Zabihi, S. M., Delgir, M., and Pourreza, H. R. (2010). “Retinal vessel segmentation using color image morphology and local binary patterns”. In: *2010 6th Iranian Conference on Machine Vision and Image Processing*, pp. 1–5 (cit. on p. 58).

Zeiler, M. D. and Fergus, R. (2014). “Visualizing and Understanding Convolutional Networks”. In: *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I*. Cham: Springer International Publishing, pp. 818–833 (cit. on pp. 103, 112).

Zhang, X., Thibault, G., Decencière, E., Marcotegui, B., and al., et (2014). “Exudate detection in color retinal images for mass screening of diabetic

- retinopathy”. In: *Medical Image Analysis* 18.7, pp. 1026–1043 (cit. on pp. 20, 44, 55, 62).
- Zhang, X., Thibault, G., Decencière, E., et al. (2012). “Spatial normalization of eye fundus images”. In: *ISBI 2012 : 9th IEEE International Symposium on Biomedical Imaging*. Barcelone, Spain: IEEE (cit. on pp. 19, 28).
- Zhao, G. and Pietikainen, M. (2007). “Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.6, pp. 915–928 (cit. on p. 58).
- Zhu, J., Zhang, E., and Rio-Tsonis, K. Del (2012). “Eye Anatomy”. In: *eLS*. John Wiley & Sons, Ltd (cit. on p. 10).