



**Universidade do Minho**  
Escola de Engenharia

Joana Daniela da Silva Pereira

**Local Analysis Strategies  
for Exudate Detection  
in Fundus Images**

Dissertação de Mestrado  
Mestrado Integrado em Engenharia Biomédica  
Ramo de Informática Médica

Trabalho efetuado sob a orientação de  
**Victor Manuel Rodrigues Alves**  
**Valery Naranjo Ornedo**

novembro de 2017

## DECLARAÇÃO

Nome: Joana Daniela da Silva Pereira

Endereço eletrónico: joana-75@hotmail.com

Cartão de Cidadão: 14901093

Título da dissertação: *Local Analysis Strategies for Exudate Detection in Fundus Images*

Orientação: Victor Alves, Valery Naranjo

Ano de Conclusão: 2018

Designação do Mestrado: Mestrado Integrado em Engenharia Biomédica

Área de Especialização: Informática Médica

Escola de Engenharia

DE ACORDO COM A LEGISLAÇÃO EM VIGOR, NÃO É PERMITIDA A REPRODUÇÃO DE QUALQUER PARTE DESTA DISSERTAÇÃO

Universidade do Minho, 24/10/2018

Assinatura: \_\_\_\_\_

## ACKNOWLEDGEMENTS

This document is the result of one year of effort and dedication that has been of intense learning for me, not only on a professional level, but also on a personal one. Getting here became possible not only for all the effort and commitment that I dedicated myself but also for all the people around me who supported me during this journey.

First, I would like to thank my supervisor at University of Minho, Victor Alves, for all his availability, support, and kindness during the development of this dissertation. I am also truly grateful for the help and guidance of my supervisor at Universitat Politècnica de València, Valery Naranjo. A special gratitude goes for Adrián Colomer who was always available in all the phases of this project, despite the project that he also had at hand - his Ph.D. thesis.

I thank my university friends for always being with me throughout this journey, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last five years. I also want to give my heartfelt thanks to my high school friends who have always been a major source of support, making the last few years easier with their friendship.

I would like to offer my special thanks to my soulmate and best friend for being my true and great supporter and for loving me unconditionally during my good and bad times. Thank you for having faith in me and for always sticking by my side.

Finally, I must express my very profound gratitude to my parents and sister for providing me with unfailing emotional support and continuous encouragement throughout my life. I am also grateful to my nieces for making me laugh and forget all my worries even in the worst days. This accomplishment would not have been possible without them and my other family members who have supported me along the way. Thank you.



## ABSTRACT

Diabetic Retinopathy (DR) is a common complication of diabetes, which is among the major causes of vision loss in the world. An early detection of the disease is the key to avoid the patient's blindness. However, at the initial phase of the disease, the vision impairment is not easily perceived by the patient. Therefore, regular follow-up exams are recommended in order to detect anomalous patterns in the patient's retina. Exudates are one of the most prevalent signs during the early stage of DR and, therefore, its early detection is vital to prevent the patient's blindness. However, the manual detection of exudates by experts is laborious and time-consuming. Thus, automated screening techniques for exudate detection have great significance in saving cost, time and labor, allowing the ophthalmologists to make the treatment decision timely. In this sense, one of the main objectives of this thesis is to develop and compare different strategies to locally extract information of fundus images for detecting exudates.

Several methods related to the automatic detection of exudates have been proposed in the literature however, these methods focus their efforts in the segmentation of exudates or require the extraction of features from a lesion candidate map. On the other hand, in the methodologies proposed in this thesis, the characterization of healthy and damaged retinal areas is performed by applying image descriptors in a local way, avoiding the segmentation step and the generation of candidate maps.

A system based on local feature extraction and Support Vector Machine classification is used to develop and compare different strategies for automated detection of exudates. The main novelty of this work is allowing the detection of exudates using non-regular regions to perform the local feature extraction. To accomplish this objective, different methods for generating superpixels are applied to the fundus images of E-OPHTA database and texture and morphological features are extracted for each of the resulting regions. Finally, each region is classified according to healthy and pathological classes, during the classification stage.

The strategies proposed in order to generate superpixels rely on applying the marker-controlled watershed transformation to a spatially regularized gradient. From these strategies, two different types of superpixels are created: c-Waterpixels and m-Waterpixels. In the end, an elaborated comparison between the proposed methods for generating m and c-waterpixels and the state-of-the-art method for generating SLIC superpixels is performed.

Additionally, a system based on Convolutional Neural Networks (CNN) is explored to discriminate between healthy and pathological regions in fundus images. Transfer learning is applied to fine-tune

some of the most important state-of-the-art CNN architectures. Exudates usually represent less than one percent of the total number of pixels that compose the retinal image. This is the reason why, in both the systems presented in this thesis, the fundus images are divided in superpixels and the classification is performed for each of the regions.

Lastly, an exhaustive comparison between the two created systems to automatically detect exudates is performed. In other words, the classification results obtained through the system involving CNNs are compared with the ones obtained by applying the approach based on feature extraction and subsequent classification using machine learning algorithms.

## RESUMO

A Retinopatia Diabética (RD) é uma complicação da diabetes, que está entre as maiores causas de perda de visão no Mundo. A detecção da doença durante a sua fase inicial é a chave para evitar a cegueira do paciente. No entanto, a debilitação da visão é muitas das vezes impercetível para o doente durante esta fase. Por este motivo, é recomendado que pacientes com diabetes realizem exames de acompanhamento regulares de forma a detetar possíveis anomalias nas suas retinas. Os exsudatos são um dos sinais mais prevalentes durante a fase inicial da RD e, deste modo, a sua detecção precoce é vital para prevenir a cegueira do paciente. No entanto, a detecção manual destas lesões é trabalhosa e demorada. As técnicas que permitem a detecção automática de exsudatos desempenham então um papel importante na diminuição dos custos, tempo e trabalho, permitindo que os oftalmologistas atribuam o tratamento a tempo de evitar a cegueira do paciente. Neste sentido, um dos principais objetivos desta tese é o desenvolvimento de diferentes estratégias que extraiam informação localmente de imagens fundus, com o intuito de detetar exsudatos.

Diversos métodos relacionados à detecção automática de exsudatos têm sido propostos na literatura. No entanto, esses métodos baseiam-se na segmentação de exsudatos ou exigem a extração de informação a partir de um mapa de candidatos. Por outro lado, nas metodologias apresentadas neste trabalho, a caracterização do tecido saudável e patológico é feita aplicando descritores de imagem localmente. Um sistema baseado numa extração local de características e posterior classificação é usado para desenvolver e comparar diferentes estratégias de detecção automatizada de exsudatos. O carácter inovador deste trabalho é permitir a detecção de exsudatos recorrendo a regiões não-regulares para realizar a extração local de características. Para atingir este objetivo, diferentes métodos para gerar *superpixels* foram aplicados às imagens da base de dados *E-OPHTA* e características texturais e morfológicas foram extraídas para cada uma das regiões resultantes. Por fim, cada região é classificada em patológica ou saudável durante o processo de classificação. As estratégias propostas para gerar superpixels consistem em aplicar a chamada *watershed transformation* a um gradiente espacialmente regularizado. A partir da referida estratégia, são obtidos dois tipos de superpixels: c-Waterpixels e m-Waterpixels. No final, é realizada uma comparação elaborada entre os métodos propostos para a geração de m e c-Waterpixels e o método estabelecido na literatura para gerar superpixels SLIC.

Para além disso, é ainda explorado um método baseado em Redes Neurais Convolucionais (CNN) para detetar regiões patológicas e saudáveis em imagens da retina. A técnica de *Transfer Learning*

é usada para transferir conhecimento sobre o tecido da retina para alguns dos modelos mais relevantes pré-treinados com a base de dados ImageNet. Normalmente, os exsudatos representam menos de um por cento do número total de pixels que compõem a imagem e este é o motivo pelo qual, em ambos os sistemas apresentados neste trabalho, as imagens são divididas em superpixels e a classificação é feita para cada uma das regiões.

Por fim, uma comparação exaustiva entre os dois sistemas criados para detetar automaticamente exsudatos é realizada. Isto é, os resultados de classificação obtidos através do sistema baseado em CNNs são comparados com os obtidos através da aplicação da abordagem baseada na extração de características e subsequente classificação usando algoritmos de *Machine Learning*.



# TABLE OF CONTENTS

1	Introduction.....	1
1.1	Context and Motivation .....	3
1.2	Objectives.....	4
1.3	Dissertation Structure .....	5
2	Concepts and Technologies .....	7
2.1	Medical Contextualization.....	9
2.1.1	Eye Anatomy.....	9
2.1.2	Diabetic Retinopathy .....	11
2.2	Superpixels.....	12
2.2.1	Definition and Properties .....	13
2.2.2	State-of-the-art Generation Methods.....	14
2.2.3	Superpixels and the Watershed Transformation .....	17
2.3	Image Descriptors .....	18
2.3.1	Local Binary Pattern Variance.....	18
2.3.2	Granulometry.....	22
2.4	Support Vector Machines .....	24
2.5	Deep Learning.....	27
2.5.1	CNN Architecture .....	28
2.5.2	Training and Testing.....	33
2.5.3	Transfer Learning.....	34
3	Material and Methods .....	37
3.1	Material.....	39
3.2	Computation of Waterpixels .....	41
3.2.1	Creation of the Grid and Definition of the Gradient .....	42

3.2.2	Selection of the Markers .....	43
3.2.3	Spatial Regularization of the Gradient and Watershed .....	45
3.3	System based on Traditional Methods .....	47
3.3.1	Introduction.....	47
3.3.2	Feature Extraction .....	49
3.3.3	Classification of the Retinal Tissue .....	50
3.4	System based on Deep Learning .....	52
3.4.1	Introduction.....	52
3.4.2	CNN for Exudate Detection .....	52
4	Results and Discussion .....	56
4.1	Evaluation Metrics.....	58
4.2	System based on traditional methods.....	58
4.2.1	Comparison of the Superpixels Generation Methods.....	58
4.2.2	Influence of the Blood Vessels Removal .....	60
4.3	System based on deep learning.....	62
4.3.1	Comparison of the fine-tuned CNN architectures.....	62
5	Conclusions.....	64
	References.....	68

## LIST OF FIGURES

Figure 1. Exudates and microaneurysm in a fundus image. ....	12
Figure 2. Illustration of Superpixels.....	13
Figure 3. Topographic view of the watershed transformation [31].....	17
Figure 4. Original LBP operator: 3 x 3 neighborhood thresholded by the value of the center pixel. ....	19
Figure 5. Circularly symmetric neighborhood of $P = 8$ and $R = 2$ .....	19
Figure 6. The 36 unique rotation invariant binary patterns that can occur in the circularly symmetric neighbor set of $LBP_8, R_{ri}$ . Black and white circles correspond to bit values of 0 and 1, respectively. The first row contains the 9 “uniform” patterns and the numbers inside them correspond to their unique $LBP_8, R_{riu}$ codes. ....	21
Figure 7. Illustration of the SVM classification process. ....	26
Figure 8. A 3 x 3 filter applied to an input image and the corresponding activation map obtained after the convolution. ....	29
Figure 9. Output volume of a convolutional layer.....	30
Figure 10. Activation functions used in Neural Networks. ....	31
Figure 11. Example of the maxpooling procedure where a filter of size 2 x 2 is applied to the input volume with a stride of 2. ....	32
Figure 12. Original image and corresponding ground-truth. (a) e-opthta EX image, (b) manual annotation of exudates. ....	40
Figure 13. Illustration of m-Waterpixels generation. (a) Original image, (b) Morphological gradient of the original image, (c) Regular grid of hexagonal cells ( $s = 32$ ), (d) Selected markers within the regular grid, (e) Spatially regularized gradient, (f) m-Waterpixels.....	42
Figure 14. Regular grid of hexagonal cells ( $s = 32$ ). ....	43
Figure 15. Morphological gradient of the original image. ....	43
Figure 16. Selected markers within the regular grid in order to generate c-Waterpixels.....	44
Figure 17. Selected markers within the regular grid in order to generate m-Waterpixels. ....	45
Figure 18. Spatially regularized gradient obtained in order to compute c-Waterpixels. ....	46
Figure 19. Spatially regularized gradient obtained in order to compute m-Waterpixels. ....	46
Figure 20. c-Waterpixels. ....	47
Figure 21. m-Waterpixels.....	47
Figure 22. Optic disk detection results. (a) Waterpixels, (b) Optic disk regions. ....	49

Figure 23. m-Waterpixels extracted from the E-OPHTHA database and provided to the CNN. (a) m-Waterpixels belonging to the class "healthy", (b) m-Waterpixels belonging to the class "pathological".  
.....53

Figure 24. Types of superpixels being compared. (a) m-Waterpixels, (b) c-Waterpixels, (c) SLIC Superpixels. ....59

Figure 25. ROC curves for the different superpixels generation methods.....62

## LIST OF TABLES

Table 1. Uniformity of different Local Binary Patterns.....	22
Table 2. Commonly used Kernel Functions.....	27
Table 3. Distribution of the images of e-ophta database according to their resolutions. ....	40
Table 4. Tunable parameters involved in the different stages of the exudate detection system. ....	59
Table 5. AUC, accuracy, sensitivity and specificity related to the exudate detection on the original images (OI) of the E-OPHTA database, using a linear kernel, for each of the region of interest developed: m-waterpixels, c-Waterpixels and SLIC superpixels.....	60
Table 6. AUC, accuracy, sensitivity and specificity related to the exudate detection on the original images (OI) and images without vessels (W/V) for each of the region of interest developed: m-Waterpixels, c-Waterpixels and SLIC superpixels.....	61
Table 7. AUC, accuracy, sensitivity and specificity related to the exudate detection, for each fine-tuned state-of-the-art architecture. ....	63
Table 8. Optimizer used for each each fine-tuned state-of-the-art architecture.....	63



## LIST OF ABBREVIATIONS

### A

AUC Area Under the Curve

### C

CNN Convolutional Neural Network

### D

DL Area Under the Curve

DME Convolutional Neural Network

DNN Deep Neural Network

DR Diabetic Retinopathy

### L

LBP Local Binary Pattern

### M

MA Microaneurysm

ML Machine Learning

MSE Mean Squared Error

### N

NPDR Non-Proliferative Diabetic Retinopathy

### P

PDR Proliferative Diabetic Retinopathy

### R

RBF Radial Basis Function

ReLU Rectified Linear Units

ROC Receiver Operating Characteristic

### S

SE Structuring Element

SLIC Simple Linear Iterative Clustering

SVM Support Vector Machines

### T

TNR True Negative Rate

TPR True Positive Rate

## V

VAR Rotational Invariant Local Variance

## W

WHO World Health Organization



# 1 INTRODUCTION



## 1.1 CONTEXT AND MOTIVATION

Diabetic Retinopathy (DR) is a common complication of diabetes, that is manifested in the retina. In fact, approximately 40% of diabetic patients suffer from DR which is an alarming factor since this disease is among the major causes of vision loss in the world. Although treatment is available, approximately 75 people lose vision each day as a consequence of DR [1]–[3]. This happens because, at the initial phase of the disease, the vision impairment is not easily perceived by the patient. For instance, in order to avoid the patients' blindness, an early medical detection of the disease is required. Actually, the World Health Organization (WHO) estimates that 80% of the visual impairment cases could be avoided through an early diagnosis and treatment [4].

The key to the early recognition of DR is the detection of exudates on the retina. Exudates are one of the most prevalent signs during the early stage of DR. These lesions are formed due to the leakage of blood and its early detection can improve patients' chances to avoid blindness [5]. However, manual detection of exudates by ophthalmologists is laborious and time-consuming. Also, it must be pointed out that the diagnosis is highly influenced by the inherent subjectivity of each expert. These are the reasons why an intensified effort has been undertaken in the last years in order to develop tools able to assist in the diagnosis of DR. Automated screening techniques for exudate detection have great significance in saving cost, time and labor, allowing the ophthalmologists to make the treatment decision timely [6]. In this sense, the main motivation of this thesis consists in developing an automated system able to discriminate between healthy and pathological areas in fundus images, in order to help the experts in the early diagnosis of Diabetic Retinopathy.

In the last years, several researches have been made in the computer vision field in order to describe the visual information contained in digital images. Systems involving feature extraction and machine learning methods, as well as systems based on deep learning techniques, have been proposed for the automatic image description and classification.

Therewith, several methods related to the automatic detection of exudates have also emerged in the literature. In this thesis, two methodologies are presented to allow the classification of the retinal tissue in healthy or pathological. In the first one, image descriptors are applied in a local way and, subsequently, the classification of the retinal tissue is performed by making use of the extracted features and machine learning methods. On the other hand, in the second approach, a system based on Convolutional Neural Networks (CNN) is explored. Transfer learning is applied to fine-tune some of the most important state-of-the-art CNN architectures.

Exudates usually represent less than one percent of the total number of pixels that compose the retinal image. For this reason, in the methodologies presented in this thesis, the fundus images are divided in regions or patches and the classification is performed for each of the regions. Different methods to generate superpixels are presented as a strategy to create non-regular regions.

Therefore, one of the main purposes of this work is to develop and compare different strategies to obtain regions of interest with the purpose of locally describing healthy and pathological retinal areas. Allowing the detection of exudates using non-regular regions is the main novelty of this work.

## 1.2 OBJECTIVES

The main goal of this work is to develop different local analyses strategies for the detection of exudates in fundus images. With this in mind, different methods to obtain regions of interest with the purpose of locally describing healthy and pathological retinal areas are developed and compared. With this it will be possible to understand the influence of using non-regular regions in the development of systems that allow the automatic classification of the retinal tissue. In order to accomplish this goal, several objectives had to be achieved, being the first the computation of waterpixels (the non-regular regions above-mentioned). As it will be detailed later in this thesis, waterpixels are superpixels created by applying the watershed transformation to a spatially regularized gradient.

The extraction of relevant features from the visual information of the images represent the second objective. To accomplish this one, texture and morphological descriptors are studied and applied to the fundus images (the features are extracted for each of the waterpixels).

The third and final objective of the first method proposed is the classification of each region (or waterpixel) according to healthy and pathological classes. To do so, the Support Vector Machines (SVM) classifier is fed with the features extracted during the feature extraction stage.

The objective of the second method proposed is to understand the concepts involved in deep learning techniques and apply convolutional neural networks to perform the automatic classification of each waterpixel.

Finally, the last objective is to establish a comparison between the different strategies presented in this thesis.

## 1.3 DISSERTATION STRUCTURE

This thesis is structured in 5 chapters. This first chapter provides the reader with the context and motivation of the study. The main contributions behind the approaches presented in this thesis are also discussed, along with the objectives of the study.

Thereafter, a theoretical chapter about the concepts and technologies related to this project is presented. To begin a theoretical section about the anatomical structures of the retina and diabetic retinopathy is provided, followed by a section dedicated to the concept of superpixels. A brief definition is presented along with the properties that superpixels should satisfy. A presentation of the existing superpixels generation methods of the state-of-the-art is also made. Finally, a discussion about whether or not the watershed transformation is well suited for superpixel generation is provided. The theoretical background of the image descriptors used to perform the feature extraction as well as about the Support Vector Machines (SVM) classifier is also presented in Chapter 2. The last section of this chapter is dedicated to the description of the most important concepts related to deep learning and, more specifically, to convolutional neural networks.

In Chapter 3, the material and methods used to develop the automated detection systems presented in this work are presented. To begin, a strategy to create superpixels by applying the watershed transformation to a spatially regularized gradient is presented. From the strategy presented in this chapter, two different types of superpixels are created: c-Waterpixels and m-Waterpixels. Subsequently, the formulation necessary to apply the image descriptors, presented in Chapter 2, in the context of this work, i.e., in the extraction of relevant information from the retinal tissue, is provided. Later, an explanation about the classification strategy formulated is also exposed. Finally, the last section of this chapter is dedicated to the explanation of the proposed approach based on fine-tuning state-of-the-art CNN architectures to discriminate between healthy and pathological regions in fundus images.

In Chapter 4, an elaborated comparison between the c-waterpixels, m-waterpixels and SLIC superpixels is performed through the analyses of the classification results obtained for each of the regions. In addition, the effect of using the original images or the same images but without blood vessels is evaluated. To finish, the results obtained through the deep learning approach are also presented along with a comparison between the different methods proposed in this thesis.

Lastly, final conclusions and future work prospects are presented in Chapter 5.



# 2 CONCEPTS AND TECHNOLOGIES





## 2.1 MEDICAL CONTEXTUALIZATION

### 2.1.1 EYE ANATOMY

Some people say that the eyes are the windows to the soul. Various emotions can be observed through the eyes (sadness, joy, confusion, among others), but their main function is to help us process and understand our world, becoming one of the main human senses with more than 70% of all the sensory receptors in the human body [7].

The eye is then an organ with an extremely complex structure, composed essentially by three layers or tunics: outer, middle and inner. The main structures of the referred layers are illustrated in Figure 1.

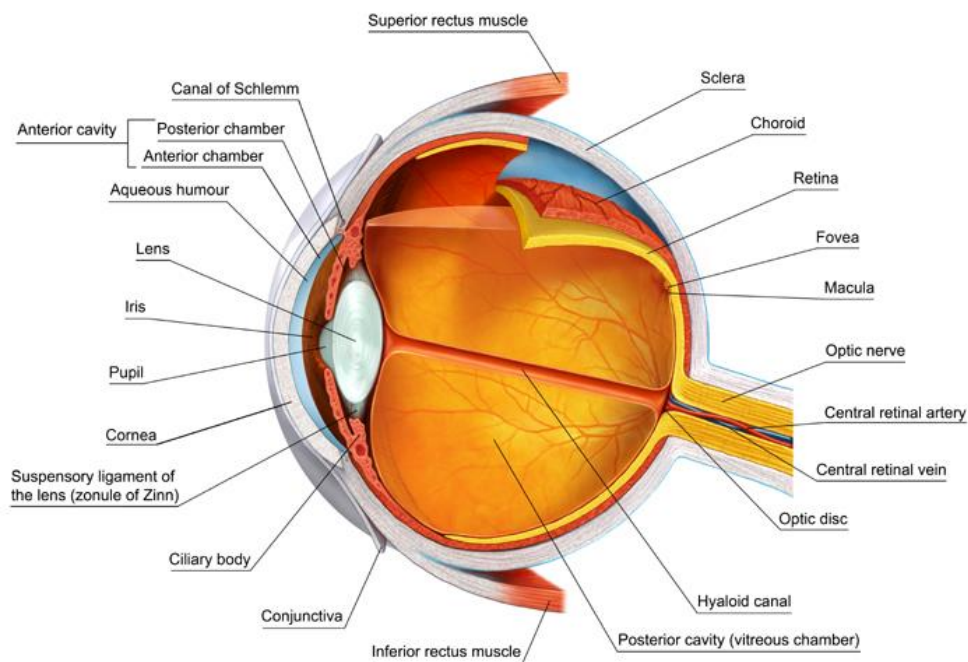


Figure 1. Anatomical structures of the eye [8].

The fibrous outer ocular layer, which acts as eye protection, is composed of the sclera and the cornea. The sclera, formed by connective tissue, constitutes the majority of the external lining of the eye presenting a white and opaque color. The cornea, a transparent fibrous tissue, is the main refractive element in the eye's optical system, allowing light to enter the globe. The sclera and cornea together protect the other components of the eyes from infections and injuries [8], [9].

The vascular middle layer (or uvea) includes the iris, ciliary body and choroid, each having separate functions but being interconnected. The main functions of the ciliary body are the production of liquid humor and the control of the shape of the lens (concave and convex) through the ciliary muscles. The choroid is a thin membrane with many blood vessels that supply the retina with nutrients and oxygen. The iris is an adjustable colored diaphragm present around the pupils, regulating the amount of light entering [8], [9].

Finally, the inner neural layer is the retina, the focus of this work since the studied pathology is manifested in this layer. The retina is a layer approximately 0.25 mm thick that has the function of turning a light stimulus into a nerve stimulus and sending it to the brain so that the images are processed. This neural layer contains numerous photoreceptor cells, about 120 million rods and 6 to 7 million cones as well as several transmission cells. Rods are bipolar photoreceptor cells that are involved in low-light vision, whereas cones are responsible for daylight color vision. The color distinction is based on the number of different cones that the human being has, more concretely three different types of cones that respond to three different frequencies: blue light, green light and red light [10]–[12].

Three different components can be found in the retina: the macula, the fovea and the optic disc. The macula is a specialized area in the center of the retina responsible for clear vision, transmitting 90% of the visual information to the brain and exhibiting a high concentration of cones. The ultra-specialized center of the macula is called the fovea and is the region of maximum visual acuity of the macula. The fovea, responsible for central vision, is a depression in the center of the macula, with a diameter of approximately 1.5 millimeters, where no rods are found but with the highest cones density throughout the retina [10], [11].

The optic disc also serves as the input to the central retinal artery bringing in nutrients and oxygen, and as output to the central retinal vein carrying excretory products from retinal cells. Inside the eye, the central retinal artery spreads into various branches and blood then passes through the thin network of very small vessels called capillaries. After moving through this network of capillaries, blood enters the branches of the retinal veins that come together to form the central retinal vein that drains the blood from the eye.

### 2.1.2 DIABETIC RETINOPATHY

Diabetes is a chronic disease caused by a deficient production of insulin (hormone produced by the pancreas) or by the ineffectiveness of the insulin produced. Both scenarios result in serious damages to the patient's organism since the lack of insulin leads to high concentration of glucose in the blood [13].

The most common diabetes-related health problems can be broadly categorized in two groups: macrovascular and microvascular complications. Macrovascular complications affect generally the heart, brain and legs of the patient, while microvascular complications are related to problems in the eyes, kidneys, feet and nerves [14].

The development of ophthalmic complications is very common in diabetic patients. Diabetic Macular Edema (DME), Ischemic Optic Neuropathy and Glaucoma are only a few examples. The most common and alarming ophthalmic complication of diabetes is Diabetic Retinopathy (DR). Approximately 80% of the patients who have been diabetic for more than 10 years will suffer from DR which is an alarming factor since this disease is among the major causes of vision loss in the world [15]. Although treatment is available, approximately 75 people lose vision each day as a consequence of DR [1]–[3]. This happens because, at the initial phase of the disease, the vision impairment is not easily perceived by the patient. For instance, in order to avoid the patients' blindness, an early detection of the disease is required. Actually, the World Health Organization (WHO) estimates that 80% of the visual impairment cases could be avoided through an early diagnose and treatment [13].

Microaneurysms and exudates are the primary symptoms indicating diabetic retinopathy and, therefore, their early detection is essential to avoid the patient's blindness [16]. Microaneurysms (MA) are small saccular dilatations of capillaries that appear as small, red dots in the superficial retinal layer. MA usually measure between 15 and 125 microns and they can rupture and leak blood into the retina. On the other hand, exudates are deposits of lipids and proteins in the retina which are typically manifested as random whitish/yellowish bright lesions of irregular edges. These lesions are caused by the breakdown of the blood-retina barrier, allowing the leakage of serum proteins, lipids, and protein from the vessels [17]. Microaneurysms and exudates are illustrated in Figure 2.

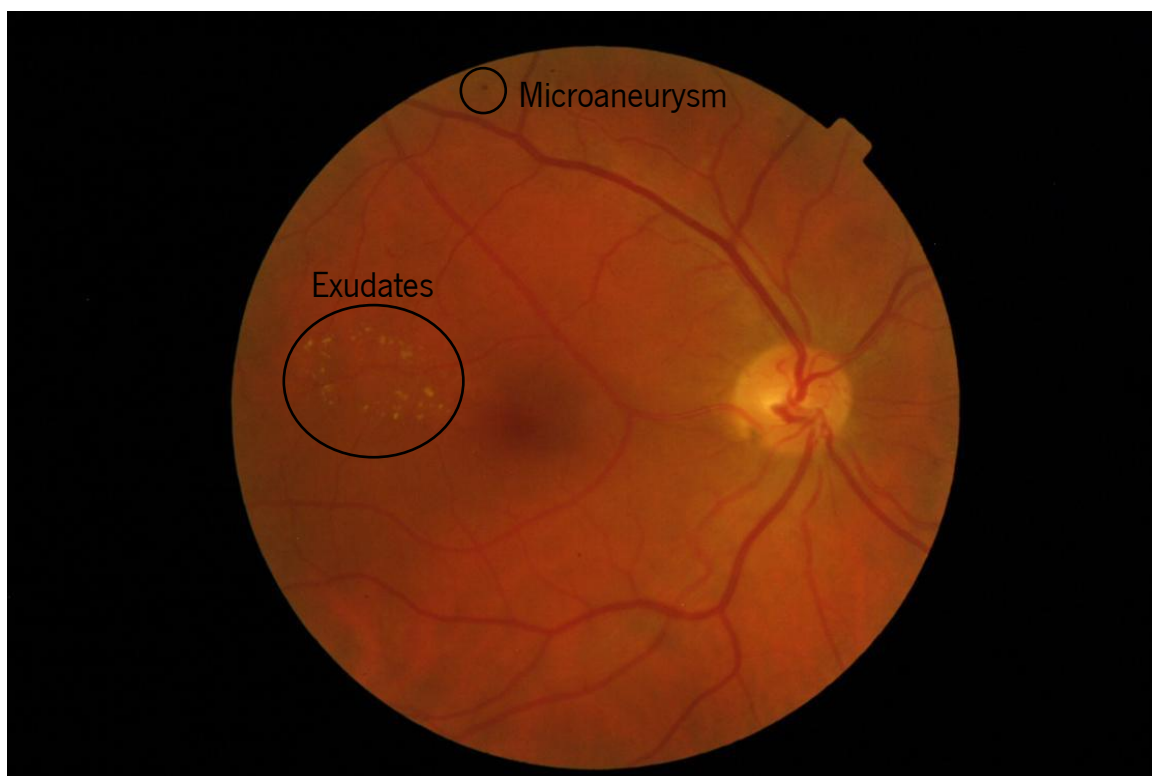


Figure 2. Exudates and microaneurysm in a fundus image.

Diabetic Retinopathy has two main stages: non-proliferative and proliferative. Non-Proliferative Diabetic Retinopathy (NPDR) is the earliest stage of DR where there is no neovascularization (growth of abnormal blood vessels) in the retina. However, NPDR may evolve into Proliferative Diabetic Retinopathy (PDR) which, on the other hand, is characterized by the presence of neovascularization in the retina and may lead to serious visual consequences [18].

NPDR is initially characterized by microaneurysms that may end up leaking into the retina. Exudates, as well as hemorrhages, usually occur with the progression of NPDR [19].

PDR is defined by the growth of abnormal blood vessels and scar tissue on the retina. These ones usually attach to the back surface of the vitreous that will subsequently pull on the scar tissue. This can lead to the bleeding of the blood vessels into the vitreous cavity and consequently result in an immediate loss of the vision [20].

## 2.2 SUPERPIXELS

As previously discussed, different methods to generate superpixels are presented in this thesis as a strategy to create non-regular regions. For this reason, this section is dedicated to the concept of

*superpixels*. A brief definition will be presented along with the properties that superpixels should satisfy. A presentation of the existing superpixels generation methods of the state-of-the-art will also be presented. Finally, the relationship between the watershed transformation and the computation of superpixels will be explored since the strategies to generate superpixels proposed in this work, rely on the marker-controlled watershed transformation.

### 2.2.1 DEFINITION AND PROPERTIES

Superpixels are regions resulting from a low-level segmentation of an image and they are characterized by their ability to capture image redundancy, to provide a convenient primitive from which to compute image features, and to greatly reduce the complexity of subsequent image processing tasks (see Figure 3 for an illustration). For this reason, superpixels are becoming increasingly popular in computer vision applications being frequently used as primitives for further analysis such as detection, segmentation, and classification of objects. The idea is to group together pixels in order to obtain a new semantic unit. As mentioned, the way to achieve such a partition is to perform a low-level segmentation of the image.

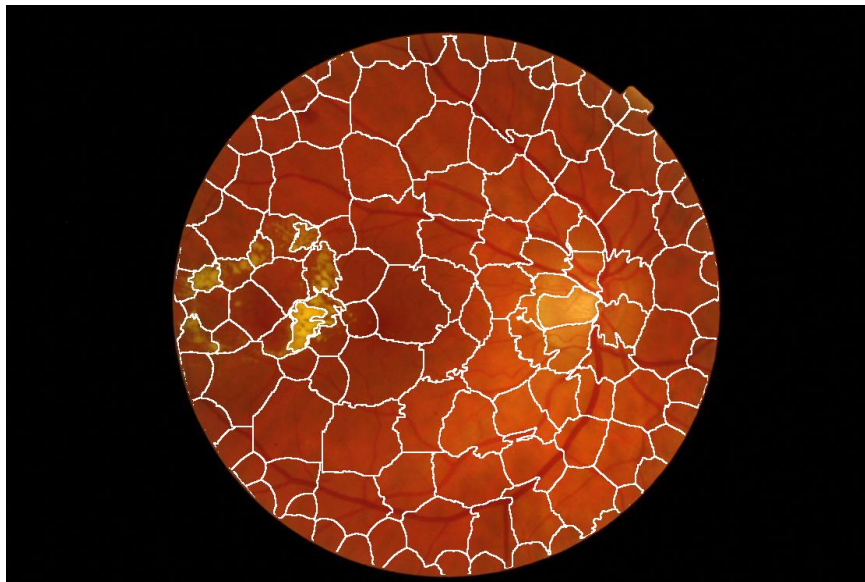


Figure 3. Illustration of Superpixels.

There are many approaches to generating superpixels, each with its own advantages and drawbacks that may be better suited to a particular application. There exist different variants of superpixels'

properties in the abundant literature and, while it is difficult to define what constitutes an ideal approach for all applications, the following properties are generally required: superpixels should adhere well to the object boundaries in the image, i.e., the object boundaries should be included in the superpixels boundaries; the pixels of a given superpixel should present similar colors or gray levels; superpixels of a given image should present similar shapes and sizes, forming a regular pattern on the image; and, finally, superpixels should be fast to compute, memory efficient, and simple to use [21], [22]. In short, superpixels are regions that must present the following properties:

1. Adherence to object boundaries
2. Homogeneity
3. Regularity
4. Easiness of computation

The first two properties - adherence to object boundaries and homogeneity - are required to ensure that a superpixel does not contain pixels belonging to different objects. On the other hand, regularity allows to perform pertinent comparisons of features values for learning. A good generation method must focus on finding a compromise between adherence to object boundaries and regularity since these properties are to a certain level oppositional. Lastly, the easiness of computation is frequently an essential aspect, once superpixels are generally only the first step of a potentially complex and time-consuming workflow [23].

### 2.2.2 STATE-OF-THE-ART GENERATION METHODS

Low-level segmentations have been used for a long time as a first step towards segmentation [23]. However, the term *superpixels* appeared later in the literature, established by X. Ren and J. Malik in [24]. Since then, this approach has generated an increasing interest and various methods to create regular and non-regular superpixels have appeared in the literature. Those can be broadly categorized as: graph-based algorithms; gradient ascent methods; or clustering methods. Of these, the latter has particular interest since the Simple Linear Iterative Clustering (SLIC) algorithm [25], which is an adaptation of  $k$ -means for superpixel generation, will play an important role during the evaluation of this work.

## 1. Graph-Based Algorithms

In this type of algorithms, each pixel is treated as a node in a graph, and edge weight between two nodes are set proportional to the similarity between the pixels. In the end, the superpixel segments are extracted by effectively minimizing a cost function defined on the graph [21].

One of the many graph-based segmentation schemes that have been used to generate superpixels is the Normalized cuts algorithm [26], in which image segmentation is treated as a graph partitioning problem. This algorithm recursively partitions a given graph using contour and texture cues. However, the algorithm is computationally expensive for large images.

Another graph-based algorithm was presented in [27]. The algorithm presented by P. Felzenszwalb and D. Huttenlocher, performs an agglomerative clustering of pixel nodes on a graph, such that each segment is the shortest spanning tree of the constituent pixels. However, unlike the Normalized Cuts algorithm above presented, this algorithm doesn't offer an explicit control over the number and compactness of the superpixels.

On the other hand, A. Moore et al. [28] presented an algorithm that allows the control of the size, number, and compactness of superpixels. This algorithm generates superpixels by finding optimal vertical (horizontal) paths that cut the image, within vertical (horizontal) strips of pixels, using graph cuts on strips of the image. However, the quality and speed of the output depend on pre-computed boundary maps.

## 2. Gradient-Ascent Algorithms

Starting from an initial rough clustering, gradient ascent methods refine, during each iteration, the clusters from the previous iteration until a convergence criterion is met in order to produce superpixels [21].

D. Comaniciu and P. Meer [29] proposed a mode-seeking procedure for detecting local maxima of a density function. Mean shift is applied to find modes in the color or intensity feature space of an image. The superpixels are composed by the pixels that converge to the same mode. Depending on the input kernel parameters, the obtained superpixels can be larger or smaller but there is no direct control over their number, size, or compactness [21].

Quick-shift [30] also is a mode-seeking algorithm, but in general this method is considerably faster than mean-shift. This algorithm can balance under and over-fragmentation of the clusters by the choice of a real parameter. However, like mean-shift, this algorithm doesn't offer an explicit control over the size and number of superpixels.

A. Levinshtein et al. [31] proposed a geometric-flow-based algorithm to generate superpixels. The algorithm progressively dilates a set of seeds in the image plane to obtain the referred regions. The geometric flow aims to distribute the superpixel in a regular way in the image. The authors called the resulting regions *turbopixels*. The method is among the slowest algorithms examined and offers a poor boundary adherence [21].

Last but not least, the watershed algorithms are also included in this category. These algorithms generally perform gradient ascent starting from local minima in order to generate superpixels. However, the watershed algorithms often present a poor boundary adherence and the resulting superpixels are, generally, irregular in size and shape. L. Vincent and P. Soille [32] proposed a fast version based on queuing of pixels but the approach doesn't offer control over the number and compactness of the resulting superpixels. On the other hand, V. Machairas et al. [23] proposed a strategy to create superpixels by applying the watershed transformation to a spatially regularized gradient. The resulting regions were called *waterpixels*. The method allowed to achieve a tunable trade-off between superpixels regularity and adherence to object boundaries.

This last approach is of particular interest since the strategies to generate superpixels proposed in this thesis, rely on the marker-controlled watershed transformation. For this reason, the relationship between the watershed transformation and the computation of superpixels will be explored in the next section.

### 3. SLIC Algorithm

The Simple Linear Iterative Clustering (SLIC) algorithm is an adaptation of the  $k$ -means clustering applied in a feature space that includes both the spectral and spatial features of each pixel [33]. Compared to the  $k$ -means method, it presents the following improvements [21]:

1. The number of distances calculated in the optimization process is reduced to decrease the complexity to be linear with the number of pixels and independent from the number of superpixels  $k$ .
2. A weighted distance measurement is applied in order to control the size and compactness of the superpixels.

The only parameter of the algorithm is  $k$ , the desired number of approximately equally sized superpixels. The input image is first divided into  $k$  blocks with a cluster center. Then, distance-limited clustering is performed iteratively until the clustering result converges. By limiting the search region, the complexity is reduced and the superpixels have regular sizes [34].



This algorithm presents a considerably high computational speed that is linear with the number of pixels, good boundary recall properties, and is easily generalizable to multiple spectral bands [33].

### 2.2.3 SUPERPIXELS AND THE WATERSHED TRANSFORMATION

The watershed transformation is a powerful tool for image segmentation [35]. It was introduced in image analysis by Beucher and Lantuéjoul [36], and several algorithms for its implementation have been proposed since then.

The watershed transformation considers a two-dimensional image  $f$  as a topographic surface  $S$  and defines the catchment basins of  $f$  and the watershed lines by means of a flooding process. The watershed lines divide the image into catchment basins, so that each basin is associated with one minimum  $m_i(f)$  in the image [37]. Let us imagine that each  $m_i(f)$  of  $S$  is pierced and that  $S$  is plunged into a lake with a constant vertical speed. The water entering through the holes floods the surface  $S$ . During the flooding, to avoid that two or more floods coming from different minima merge, a dam is built on the points of  $S$  where the floods would merge (see Fig. 4). At the end of the process, only the dams, which define the watershed of the function  $f$ , emerge [38] [39].

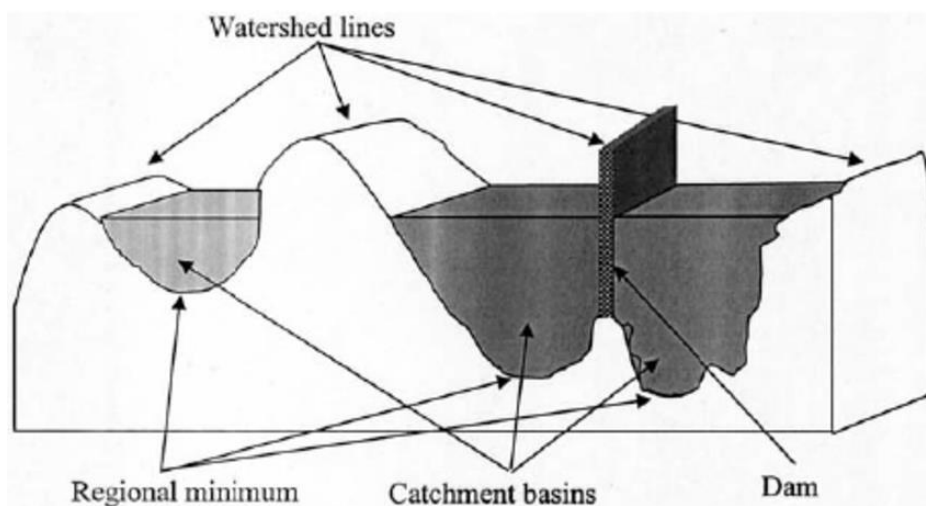


Figure 4. Topographic view of the watershed transformation [39].

The watershed transformation is typically applied to the gradient function of the image. Generally, this gradient has high values on the borders between objects and minima in the homogeneous regions. Therefore, if the crest lines in the gradient image correspond to the boundaries of the objects in the image, the watershed transformation partition the image into meaningful regions [37].

Usually, a watershed segmentation on the gradient image without any additional processing leads to oversegmentation, since every local minimum of the gradient results in a different region [37]. In the methodology presented in this thesis, as it will be discussed in Section 3.2., markers were used in order to reduce the number of local minima, avoiding the oversegmentation.

After this brief explanation about the watershed transformation's functioning, it is possible to conclude that, in principle, the watershed transformation is well suited for the superpixel generation, since:

1. When computed on the image gradient, it produces a good adherence to object boundaries.
2. It allows to control the number and spatial arrangement of the resulting regions through the choice of markers.
3. The connectivity of the resulting regions is guaranteed.

For these reasons, the watershed transformation has been used in many applications, inclusively in 3D applications computationally intensive [40], [41], in order to produce low-level segmentations. Although some previous publications [21], [31] had defended that the watershed transformation wasn't suitable for the generation of superpixels, some latest publications [42]–[44] have showed the opposite.

V. Machairas et al. [23] proposed a strategy to create superpixels by applying the watershed transformation to a spatially regularized gradient to achieve a tunable trade-off between superpixels regularity and adherence to object boundaries. With this approach, the term “waterpixels” was introduced in the literature.

## 2.3 IMAGE DESCRIPTORS

The first system to automatically detect exudates developed in this work is based on performing a local feature extraction from fundus images and subsequently classify the retinal tissue using machine learning algorithms. The theoretical background of the image descriptors used during the feature extraction will be presented in this section.

### 2.3.1 LOCAL BINARY PATTERN VARIANCE

The Local Binary Pattern (LBP) operator was introduced in the literature by T. Ojala et al. in [45] and represents a powerful grey-scale texture operator that summarizes the local structure in an image by

comparing each pixel with its neighborhood. This first LBP operator introduced by the authors used a fixed 3 x 3 neighborhood. The operator labels the pixels of an image by thresholding the neighborhood of each pixel. In other words, the original 3 x 3 neighborhood is thresholded by the value of the center pixel - if the intensity of the center pixel is greater or equal its neighbor, the neighbor takes the value of 1, otherwise, the neighbor is denoted with 0. See Figure 5 for an illustration. As the neighborhood is composed by 8 pixels, a total of 28 different labels can be obtained. Then, the distribution of the labels is studied and used as a texture descriptor for classification purposes.

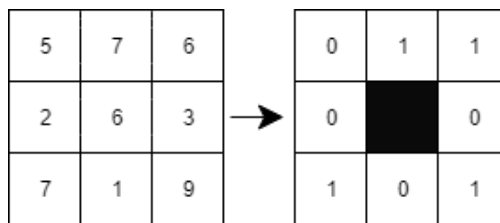


Figure 5. Original LBP operator: 3 x 3 neighborhood thresholded by the value of the center pixel.

Later the operator was extended to use neighborhoods of different sizes. T. Ojala et al. [46] conceived an operator that is invariant against any monotonic transformation of the gray scale. The authors accomplished that by starting from the joint distribution of gray values of a circular symmetric neighborhood set of pixels in a local neighborhood. Rotation invariance is achieved by recognizing that this grey-scale invariant operator incorporates a fixed set of rotation invariant patterns [46]. The operator was conceived for a general case based on a circularly symmetric neighbor set of  $P$  members on a circle of radius  $R$ . See Figure 6 for an illustration of the circularly symmetric neighborhood set ( $P=8, R=2$ ). The parameter  $P$  controls the quantization of the angular space, while the parameter  $R$  determines the spatial resolution of the operator.

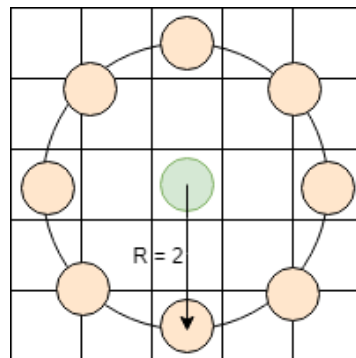


Figure 6. Circularly symmetric neighborhood of  $P = 8$  and  $R = 2$ .

The basic LBP version consists in assigning a label to each pixel  $(i, j)$  of the image, considering its local neighborhood:

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

where  $P$  represents the number of samples on the symmetric circular neighborhood of radius  $R$ ,  $g_c$  is the grey value of the center pixel  $(i, j)$  and  $g_p$  is the grey value of each neighbor. The pixels of the neighborhood are thresholded with respect to  $g_c$  giving rise to a binary pattern. It should be noted that, if the coordinates of the  $P$  neighbors do not fall in the center of the pixels, the  $g_p$  values are calculated by interpolation.

Using the  $LBP_{P,R}$  operator,  $2^P$  different binary patterns can be formed by the  $P$  pixels in the neighbor set. That means that a total of  $2^P$  different output values or labels can be obtained by the  $LBP_{P,R}$  operator. However, rotating a particular binary pattern will naturally results in a different  $LBP_{P,R}$  value since several binary strings may refer to the same pattern. For example, both the binary patterns 10000111 and 11000011 are different rotations of 11100001. Naturally, this doesn't apply to patterns that are entirely composed by 0's or 1's. With the objective of removing the effect of rotation, a unique identifier to each rotation invariant local binary pattern is assigned as follows:

$$LBP_{P,R}^i = \min\{ROR(LBP_{P,R}, t) \mid t = 0, 1, \dots, P-1\} \quad (2)$$

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

Figure 7 represents the 36 unique rotation invariant local binary patterns that can occur in the case of  $P=8$ . It is important to discuss the ability of discrimination of this operator. Certain local binary patterns have fundamental properties of texture, describing over 90% of the patterns present in the textures of their experiments [47]. Thus, it is possible to assume that some patterns may provide redundant information, compromising the ability of discrimination of the operator. To overcome this problem, T. Ojala et al. [46] introduced the 'uniform' patterns by defining a uniformity measure  $U(LBP_{P,R})$ . This measure corresponds to the number of spatial transitions (bitwise 0/1 changes) in the pattern, i.e., if a binary pattern presents two 0/1 transitions, then its corresponding uniformity

$U$  is 2. For example, the patterns 0 and 8 of Figure 7 have  $U = 0$ , while the other seven patterns of the first row have  $U = 2$ . For a better understanding of this concept, see Table 1. T. Ojala et al. designate patterns with a  $U$  value of at most 2 as ‘uniform’ and propose the following operator for gray scale and rotation invariant texture description instead of  $LBP_{P,R}^{ri}$ :

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if } U(LBP_{P,R}) \leq 2 \\ P + 1 & \text{otherwise} \end{cases} \quad (3)$$

where

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_{p-1} - g_c) - s(g_p - g_c)| \quad (4)$$

By definition,  $P+1$  ‘uniform’ binary patterns can occur in a circularly symmetric neighbor set of  $P$  pixels and each of them is labelled in the range of 0 to  $P$ . The rest of the patterns are grouped in a ‘non-uniform’ set labelled as  $P+1$ . In Figure 7, the labels of the ‘uniform’ patterns (first row) are denoted inside the patterns.  $LBP_{P,R}^{riu2}$  has  $P+2$  distinct output values.

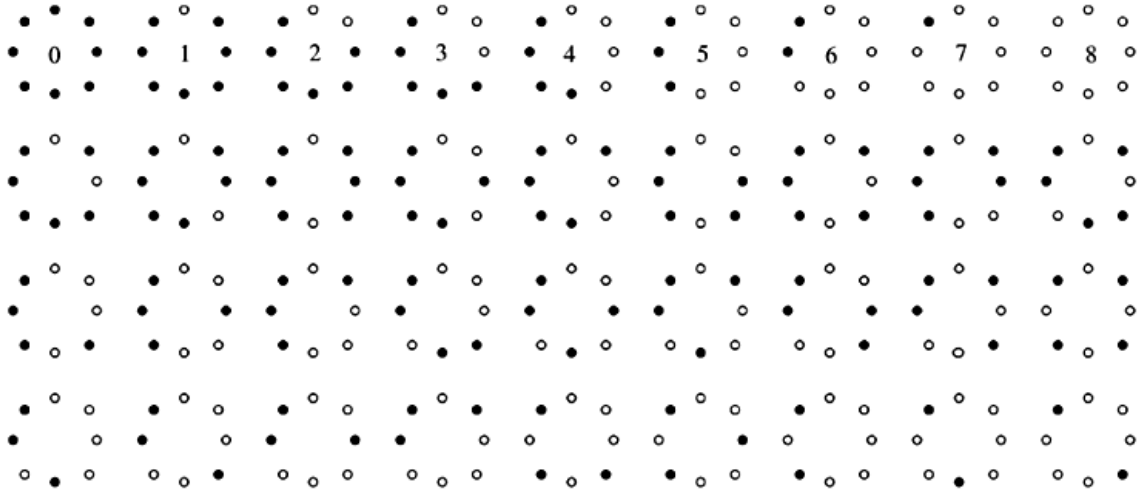


Figure 7. The 36 unique rotation invariant binary patterns that can occur in the circularly symmetric neighbor set of  $LBP_{8,R}^{ri}$ . Black and white circles correspond to bit values of 0 and 1, respectively. The first row contains the 9 “uniform” patterns and the numbers inside them correspond to their unique  $LBP_{8,R}^{riu}$  codes.

Table 1. Uniformity of different Local Binary Patterns.

LBP Pattern	Uniformity
11111111	0
00000100	2
11111100	2
10011101	4
01010101	8

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

$$VAR_{P,R}(i,j) = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - \mu)^2, \quad \mu = \frac{1}{P} \sum_{p=0}^{P-1} g_p \quad (5)$$

$LBP_{P,R}^{riu2}$  and  $VAR_{P,R}$  are complementary measures and the combination of both is expected to be a powerful descriptor for detecting abnormal retinal patterns. For this reason, in this thesis,  $LBP_{P,R}^{riu2}$  and  $VAR_{P,R}$  operators are locally computed for each pixel of the green channel of the retinal image. As a result, LBP and VAR images are obtained. These resulting images are divided into waterpixels regions and normalized histograms are computed for each region combining the information provided by both images, as it will be explored in Section 3.3.2.

### 2.3.2 GRANULOMETRY

Mathematical Morphology is a theory, introduced by G. Matheron and J. Serra [48], that provides useful tools for image processing and analysis. As its name suggests, this theory aims at analyzing the shape and size of objects. As a matter of fact, mathematical morphology is not only a theory but also a powerful image analyses technique that has wide application in texture analysis, image segmentation and object recognition. This technique has become increasingly relevant in the image processing essentially due to its versatility and rigorous mathematical description. The main idea behind mathematical morphology is to analyze geometric information by “probing” an image with a structuring element (mask with given shape and size). This idea, although simple, provided a new approach for processing visual information [49].

Granulometry is one of the most interesting techniques based on mathematical morphology. This technique was initially developed to extract size distribution from binary images and was then extended to function operators to analyze and extract texture features from greyscale images [50]. A morphological granulometry is essentially computed using a serial of morphological openings (closings) with increasing size of the structuring element (SE) in order to obtain a size distribution of the objects in the image. In ordered textures, the obtained size distribution provides shape and size information, otherwise, granulometry extract the degree of granularity. As previously mentioned, the SE is a mask with given shape and size, being, generally, a disk or a line giving place to an isotropic or angular granulometry, respectively. At each opening, the finer details are successively eliminated, and the volume of the input image is reduced. Successive openings create a decreasing sequence of images. The image volume remaining after each opening constitutes a decreasing sequence which eventually reaches zero [51]. This technique can be computed on binary or grey-scale images providing signatures that, when applied to a classifier, can be used to identify several patterns or to distinguish different textures. In the context of this thesis, granulometry is proposed as a descriptor to detect exudates, one of the first signs related to diabetic retinopathy.

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  $E \subseteq Z^2$ .  $T$  is an ordered set of grey-levels. After the selection of a fixed set  $B \subseteq E$ , the two elementary operations of erosion ( $\epsilon_B(X)$ ) and dilation ( $\delta_B(X)$ ) can be composed together to generate a new set of grey-level operators given by the grey-level opening ( $\gamma_B$ ) and closing ( $\varphi_B$ ):

$$\gamma_B(f)(x) = (f \circ B)(x) = \delta_B(\epsilon_B(f))(x) \quad (6)$$

$$\varphi_B(f)(x) = (f \bullet B)(x) = \epsilon_B(\delta_B(f))(x) \quad (7)$$

Let  $\gamma_i(f)$  be an opening applied to an image  $f$  with a SE of size  $i$ . The opening can be computed by an erosion followed by a dilation. When a series of openings with SE of increasing size ( $\lambda$ ) are sequentially computed on the image, a morphological opening pyramid is obtained, and it can be defined as:

$$\Pi_\gamma(f) = \{\Pi_{\gamma\lambda}: \Pi_{\gamma\lambda} = \gamma_\lambda(f), \quad \forall \lambda \in [0, \dots, n_{max}]\} \quad (8)$$

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

By duality, when a series of closings with SE of increasing size ( $\lambda$ ) are sequentially computed on the image, a morphological closing pyramid  $\Pi_\varphi$  is obtained:

$$\Pi_\varphi(f) = \{\Pi_{\varphi\lambda}: \Pi_{\varphi\lambda} = \varphi_\lambda(f), \forall \lambda \in [0, \dots, n_{max}]\} \quad (9)$$

A shape descriptor can be defined using the morphological pyramids above described. Let  $m(f)$  be the Lebesgue measure of a discrete image  $f$ . The granulometry curve, or pattern spectrum of  $f$  with respect to  $\Gamma$  is defined as:

$$PS_\Gamma(f, n) = PS(f, n) = \frac{m(\Pi_{\gamma n}(f)) - m(\Pi_{\gamma n+1}(f))}{m(f)}, n \geq 0 \quad (10)$$

$PS_\Gamma(f, n)$  maps each size  $n$  to some measure of the bright image structures with this size: loss of bright image structures between two successive openings. Therefore, the pattern spectrum  $PS_\Gamma(f, n)$  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 [52].

By duality, this concept extends to the anti-granulometry curve  $PS_\Phi(f)$ :

$$PS_\Phi(f, -n) = PS(f, -n) = \frac{m(\Pi_{\varphi n}(f)) - m(\Pi_{\varphi n-1}(f))}{m(f)}, n \geq 0 \quad (11)$$

The pattern spectrum  $PS_\Phi(f, -n)$  characterizes the size of dark image structures.

## 2.4 SUPPORT VECTOR MACHINES

The Support Vector Machines (SVM) is one of the most widely used machine learning algorithms. Some of the reasons that made SVMs a successful modeling and prediction tool for a variety of applications are: their ability to learn well with only a very small number of free parameters; their robustness against several types of model violations and outliers; and their computational efficiency compared with several other methods [53].

This classifier generates 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. With this



purpose, SVM maximizes the distance between the hyperplanes defined by the support vectors with the aim of finding the optimal separation between classes [54]. In other words, SVM is a discriminative classifier formally defined by a separating hyperplane. In a two-dimensional space, this hyperplane is a line that divides the plane in two parts - each class lays in a different side. When the linear separation is not possible, they can work in combination with the technique of 'kernels', that automatically realizes a non-linear mapping to a feature space. The hyperplane found by the SVM in feature space corresponds to a non-linear decision boundary in the input space.

Let us consider a training sample  $(x_i, y_i)^N$ ,  $i = 1, \dots, N$ , where  $x_i$  is the input pattern and  $y_i \in \{+1, -1\}^t$  is the corresponding label (desired output). The hyperplane that does the separation of classes can be defined as:

$$w^T x_i + b = 0 \quad (12)$$

where  $w$  is the normal vector to the hyperplane solution and  $b$  is the intersection.

From the proposed hyperplane, the space is separated into two regions: a positive ( $w^T x_i > 0$ ) and a negative ( $w^T x_i < 0$ ) one. The samples belonging to the positive class are contained in the first region, while the samples that correspond to the negative class belong the second region. See Figure 8 for an example. In the referred figure, the green and red samples correspond to the negative and the positive class, respectively. The highlighted samples in the image correspond to the support vectors that define the two hyperplanes ( $h_-$  and  $h_+$ ) parallel to the hyperplane solution. Note that the hyperplane solution is given by the maximization of the distance  $d_+ + d_-$ . The classification function can be expressed as:

$$f(x) = \text{sgn}(w^T x_i + b) \quad (13)$$

The hyperplanes  $h_+$  and  $h_-$  are defined as  $w^T x_i + b = 1$  and  $w^T x_i + b = -1$ , respectively. Consequently, a sample has to satisfy the condition  $w^T x_i + b \geq 1$  to be labelled as positive and the condition  $w^T x_i + b \leq -1$  to be labelled as negative.

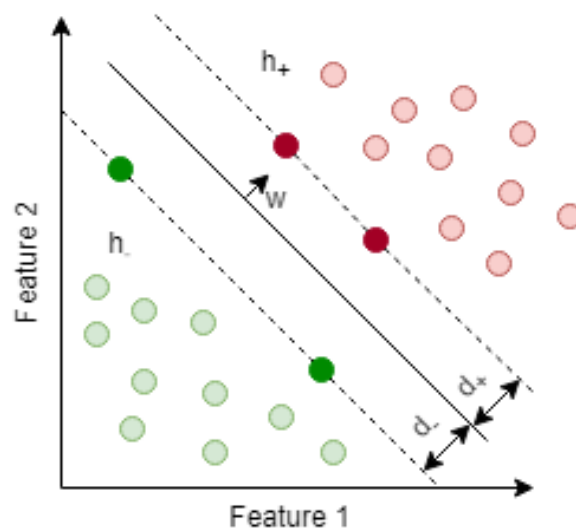


Figure 8. Illustration of the SVM classification process.

The complexity of the hyperplane can be bounded in terms of another quantity called margin [55], which can be computed as the distance between the hyperplanes in which the support vectors are included:

$$\text{margin} = d^+ + d^- = 2 \left( \frac{|wx + b|}{\|w\|} \right) = 2 \frac{1}{\|w\|} \quad (14)$$

The objective of the SVM algorithm is to maximize the margin value, obtained by Equation 14. This can be achieved by minimizing the inverse function:

$$\Phi(w) = \frac{1}{2} \|w\|^2 \quad (15)$$

where  $\|w\|^2 = w^T w$ . The hyperplane solution will be obtained solving the minimization problem and computing the optimal  $w$  and  $b$  parameters.

A linear-separable training set was assumed at the beginning of the explanation. However, the datasets extracted from real applications do not generally meet this particularity. In these cases, the problem can be reformulated as:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (16)$$

subject to

$$y_i(w^T \Phi(x_i) + b) \geq 1 - \xi_i \quad (17)$$

where  $w$  represents the normal vector to the hyperplane;  $\xi_i$  are slack variables responsible for determining the degree of misclassification;  $C$  is the penalty parameter of the error term; and  $\Phi(x)$  is the function responsible for transforming the input vectors  $x_i$  into a higher dimensional space. SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. In cases where no linear separation is possible, SVM can be used in combination with kernels, that automatically realize a non-linear mapping to a feature space. A non-linear function ( $\Phi$ ) is used for mapping non-linear separable data points in the original space  $\sum$ , into a higher dimensional space in which the data is linearly separable. When this hyperplane of separation is projected back to the input space ( $\sum$ ), it has the form of a non-linear curve. The kernel function is defined as  $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$  and the most commonly used kernels are presented in Table 2 [55].

Table 2. Commonly used Kernel Functions.

<i>Kernel</i>	$K(x_i, x_j)$
<i>Linear</i>	$x_i^T \cdot x_j$
<i>Polynomial</i>	$((x_i^T \cdot x_j) + c)^d$
<i>Radial Basis Function (RBF)</i>	$\exp(-\gamma \ x_i - x_j\ ^2), \gamma > 0$
<i>Sigmoid</i>	$\tanh(\gamma(x_i^T \cdot x_j) + c)$

where  $\gamma$ ,  $c$  and  $d$  are kernel parameters.

## 2.5 DEEP LEARNING

Deep learning is a form of Machine Learning (ML) that enables computers to learn from experience by imitating the functioning of the human brain. The recognition task is one of the first skills developed by a person since the birth moment. When children are learning how to recognize a particular animal, for instance, they are not told by their parents that they should measure the contours of its body. They learn by being shown multiple examples and being corrected when they made a wrong guess. Let us imagine that when we are born, our brain provides us with a model that

takes in our sensory inputs and makes a guess about what we are experiencing. If someone confirms that guess, our model is reinforced. On the other hand, if someone says that our guess was wrong, this new information is incorporated in our model. Machine Learning is built on this idea of learning from examples [56].

Deep Learning (DL) emerged when researches in Machine Learning attempted to build models that resemble the structures used by the human brain. The need to create these models occurred while attempting to solve more complex problems, such as object recognition. Currently, deep learning is the focus of most computer vision researches since it not only far surpasses other ML algorithms, but also rival the accuracies achieved by humans [56].

Several types of Deep Neural Networks (DNNs) exist in the DL domain. Each one solves a specific problem in a better way than the others. Convolutional Neural Networks (CNNs) have been successfully used in object recognition tasks. As previously discussed, the recognition task is one of the first skills developed by a person. When we see an image, we are immediately able to identify each of the objects that compose it. This capability is provided by the human visual cortex. Therefore, the recognition task is not easily performed by computers. However, in the last years, CNNs have proved to be a powerful tool in the automatic detection of objects. It is important to remark that CNNs are inspired by the human visual cortex. This one is characterized by cell regions sensitive to specific areas of the visual field, i.e., different cell regions react to different visual stimuli, such as edges or colors. It's the combination of the different responses to a particular stimulus that make the visual perception possible. The neurons of the visual cortex are organized by a columnar architecture [57]. CNN architectures for recognition tasks emerged based on these facts.

In this work, convolutional neural architectures are fine-tuned to detect exudates. Therefore, the following subsections are dedicated to the description of the most important concepts related to CNNs.

### 2.5.1 CNN ARCHITECTURE

Convolutional Neural Networks are composed by several hidden layers. Each of these layers is composed by a set of neurons that are fully connected to all the neurons in the previous layer. The neurons in the same layer, function completely independently and do not share any connections. This subsection is dedicated to exploring the principles of each of the hidden layers that compose a CNN architecture.

1. Convolutional Layer

The convolutional layer is the first layer of the CNN and its input is an image of dimensions  $N \times M \times 3$ . This layer's parameters consist of a set of filters (also referred to as kernels) that will slide (or convolve) across all the areas of the input image as a sliding window with a specific step. The kernel must have the same depth of the input image being its dimensions  $k \times k \times 3$ . The set of pixels involved within the sliding window in each step is called receptive field [58].

The kernels are defined by weights and, as the kernel is sliding across the image, those weights are multiplied by the original pixel values of the image. The mathematical operation between the kernel and the receptive field is called convolution. The above-mentioned multiplications are all summed up in order to obtain a single number (Figure 9). By applying each kernel over all the pixels of the input image we obtain an activation map.

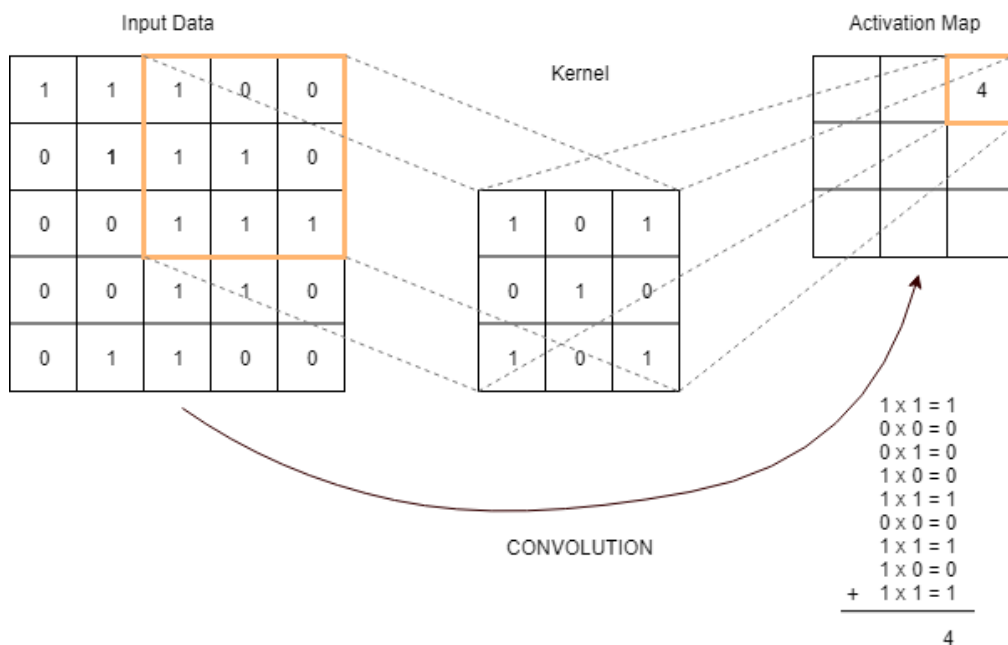


Figure 9. A  $3 \times 3$  filter applied to an input image and the corresponding activation map obtained after the convolution.

As it can be observed in Figure 9, the dimensions of the input image aren't preserved, which will result in a loss of information. In order to preserve the spatial size of the input volume a zero-padding can be performed. Let us imagine that we have an image of dimensions  $N \times M \times 3$ . If we want to preserve these dimensions, it is necessary to perform a zero-padding before the convolution. Remember that  $K$  is the kernel's height and width. In the end, the output of the convolutional layer is an activation map of dimensions  $N \times M \times Z$ , where  $Z$  is the number of kernels (Figure 10).

The kernels weights are the key element in the automatic learning process. Each of the kernels can be thought of as feature identifiers. Intuitively, the network will learn how to highlight edges, curves and other shapes of the image by using the appropriate kernel weights.

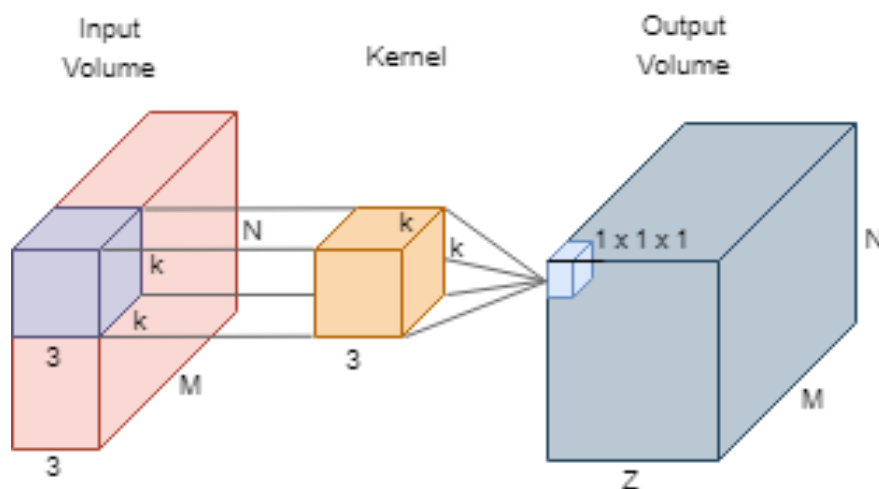


Figure 10. Output volume of a convolutional layer.

## 2. Activation Layer

After each convolutional layer, a nonlinear layer, known as activation layer, is applied. The purpose of this type of hidden layers is to provide nonlinearity to the network in order to improve its robustness. The introduction of the mentioned nonlinearity is important since the network has just been computing linear operations (multiplications and summations) during the convolutional layers. Nonlinear functions, such as hyperbolic tangent or sigmoid (Figure 11), have been used in the past but V. Nair and G. Hinton [59] found out that Rectified Linear Units (ReLU) layers allow the network to train faster without altering the accuracy of the classification. The ReLU layer applies the function  $f(x) = \max(0, x)$  to all of the values in the input volume (output of the previous convolutional layer). In other word, this layer changes all the negative activations to 0 [60].

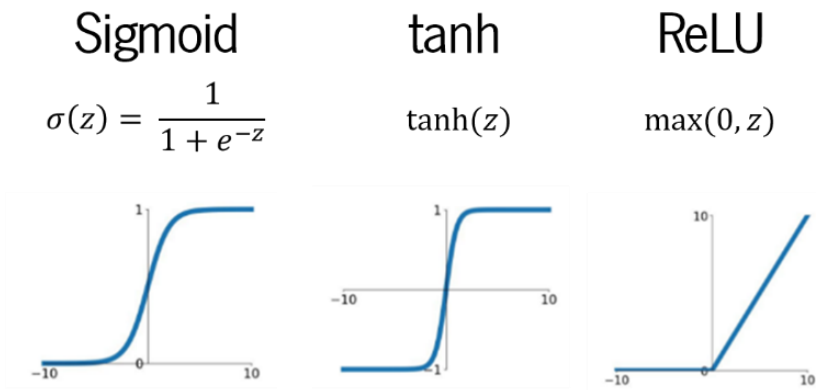


Figure 11. Activation functions used in Neural Networks.

### 3. Pooling Layer

A CNN architecture commonly includes a pooling layer between convolutional layers. This layer progressively reduces the spatial dimensions (width and height) of the input volume. Several methods exist to do so but the most common procedure is known as maxpooling. This technique consists in keeping the value of a specific neighborhood.

In short, the pooling layer operates on every depth slice of the input and resizes it spatially, generally using the maxpooling procedure. Typically, filters of size 2 x 2 are applied to the input volume with a stride of 2, decreasing the weights of the filters by 75%. Note that the filter is applied in every depth slice along both width and height and thus the depth dimension remains the same [61]. Figure 12 illustrates the above described procedure.

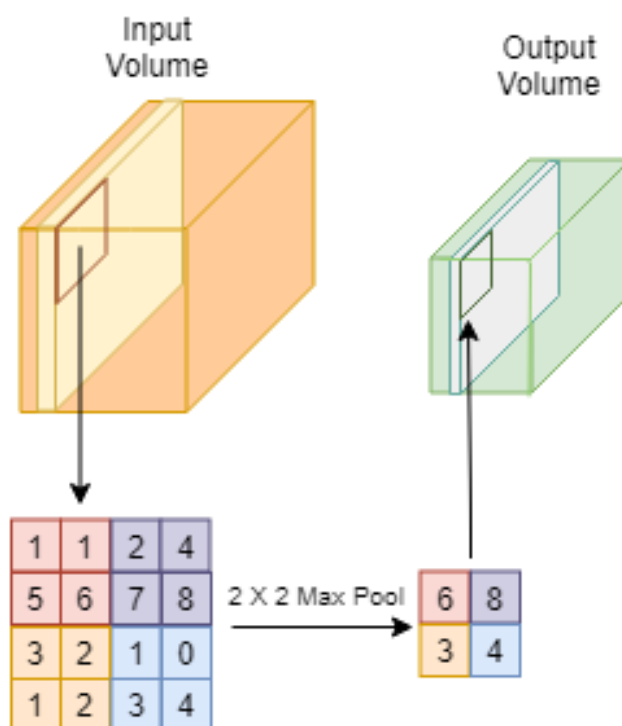


Figure 12. Example of the maxpooling procedure where a filter of size 2 x 2 is applied to the input volume with a stride of 2.

As it can be observed in Figure 12, the spatial dimension of the input volume is drastically reduced. This is good for two reasons: the computational cost is reduced since the amount of weights is reduced by 75%; and, the overfitting is controlled. These are the main purposes of including pooling layers in a CNN architecture.

#### 4. Fully Connected Layer

The purpose of a fully connected layer is to convert an input volume (the output of a convolutional, activation or pooling layer) into an N dimensional vector where N is the number of classes. The output vector gives the probability of any of the classes to be true.

To do so, this layer applies the 1D convolution to the flatten volume of the last 2D filter, reducing the number of activations to the number of classes ( $N_c$ ). Thereafter, it is possible to map the activations into the final probabilities by using an appropriate activation function such as softmax. This function normalizes the outputs of the previous layer to be between zero and one and divide each so that they sum up to one. Therefore, the output of this function gives the probability of any of the classes to be true.



## 2.5.2 TRAINING AND TESTING

When children are born, their minds are unable to distinguish between a dog and a cat. The same happens when a CNN is created, the kernels weights are randomly initialized. The low-level filters do not know how to look for an edge or a curve and the high-level filters don't have any knowledge about colors or about a particular object on the image. However, and as it was already discussed, as they grow older, children acquire knowledge thanks to what they are learned by their parents and other people. A common way of teaching children is to show them images and tell them what's on it. This idea of showing an image and offer the answer (corresponding label), is the principle of the training stage of a CNN.

The overall training process of a CNN can be divided in the following steps [58]:

### 1. Forward Propagation

In this step, a training image is provided to the network and goes through it, crossing all its layers. Remember that the kernels weights are randomized when the CNN is created. For this reason, the CNN can not assign a correct prediction to the input image and the output will not give preference to any class in particular, i.e., it will return the same probability for all the classes.

### 2. Loss Function

A CNN with random weights cannot make any reasonable prediction in the first iteration, but a label is provided to the system. Therefore, the CNN will learn by updating the weights of its filters according to the loss function. The loss function can be defined in several ways but the most common one is using the Mean Squared Error (MSE):

$$E_{total} = \sum \frac{1}{2} (target - output)^2 \quad (18)$$

### 3. Backpropagation

At each iteration of the training stage, the goal is to minimize the loss function in order to obtain a predicted label (output of the CNN) equal to the training label provided to the system. To do so, it is first necessary to find out which weights most directly contributed to the loss registered by the loss function. Additionally, the optimal way to adjust these weights must be established in order to

decrease the loss. This step is known as backpropagation and it mathematically consists on the computation of the derivative of the loss function with respect to the weights of the filters ( $dL/dW$ ).

#### 4. Weight Update

Afterward, the weights of the filters must be updated in order to make them change in the opposite direction of the gradient:

$$w = w_i - \eta \frac{dL}{dW} E_{total} \quad (19)$$

where  $w$  are the updated weights,  $w_i$  are the initial weights and  $\eta$  is the learning rate. The learning rate is a parameter that must be chosen by the network designer. A high learning rate propitiates larger jumps in the updating process which will probably result in a decreasing of the time needed for the model to converge on an optimal set of weights. However, it is important to have in mind that a learning rate that has too high values can result in too large jumps and therefore, never reach the global minimum of the loss function (where weights are optimal).

It is important to note that the above described steps are part of the same training iteration and that the process will be repeated for each set of the training images (also known as batch) during all the epochs. The number of batches and their size, as well as the number of epochs are parameters chosen by the network designer.

The creation of the model is concluded when the last batch of the training set is processed in the last epoch and the weights are updated. Finally, to understand how well the model predicts, the same must be evaluated using the testing set (which can not contain any image used to perform the training).

### 2.5.3 TRANSFER LEARNING

Due to the availability of large-scale annotated datasets, remarkable progresses have been made in the image recognition field. The most popular image dataset used in computer vision is ImageNet [62]. This dataset offers more than 14 million images belonging to more than 1000 classes. The CNN models trained upon this database serve as support for significantly improving many object detection and image segmentation problems using other datasets [63]. This happens since training a Convolutional Network (ConvNet) from scratch using a small dataset significantly affects the

ConvNet's ability to generalize, resulting in overfitting. This is very important in the medical field where the reduced amount of data is the biggest limitation when applying CNNs, since expert annotation is expensive and time-consuming, and the diseases (e.g., lesions) are scarce in the datasets [64]. In practice, it is very rare to have a dataset of sufficient size to train a ConvNet from scratch. Instead, it is common to pretrain a ConvNet on a very large dataset (e.g. ImageNet) and then use the ConvNet either as an initialization or a fixed feature extractor for the task of interest [61].

Transfer Learning has then emerged as a powerful technique to overcome the problem of the small amount of data. This technique consists in creating a model to solve a specific classification problem by using a pretrained model. In other words, transfer learning uses the knowledge acquire by the model trained with large image datasets (e.g. ImageNet), to reduce the amount of data necessary to create the new model.

In this work, a transfer learning strategy called fine-tuning was used in order to overcome the above-mentioned problems that may result from training a CNN from scratch using a small dataset. The first step involved in the process of fine-tuning is the transfer of the weights from a pre-trained network to the network we wish to train. Except for the last fully connected layer that is commonly replaced with a new fully connected layer that has as many neurons as the number of classes in the new target application [64]. For example, in the particular study of this work we have two classes ("pathological" and "healthy") therefore, the new fully connected layer has two neurons.

It is possible to fine-tune all the layers of the CNN or to fine-tune only the higher-level layers, keeping the earlier layers fixed. This procedure is commonly used since the earlier layers contain more generic features (e.g. edges or curves) that are usually useful for several applications. On the other hand, the higher-level layers become more specific to the details of the classes contained in the original dataset [61]. The most popular state-of-the-art CNN architectures have been trained using the ImageNet dataset and the resulting weights are publicly available.



# 3 MATERIAL AND METHODS



### 3.1 MATERIAL

The database chosen to validate the methods proposed in this thesis was the E-OPHTHA public database [65]. This database of fundus images was specifically designed for the screening of diabetic retinopathy and it results from the extractions of the examinations gathered during years 2008 and 2009 through the OPHDIAT network (a telemedical network to screen for diabetic retinopathy). To be more exact, 25.702 examinations were extracted, each one containing at least four images (two per eye). Additionally, contextual data is also provided for each examination, such as the patient's age, the number of years of diabetic retinopathy and the diabetic retinopathy level for each eye, among others.

In order to provide correct and accurate information to the research community, 500 examinations were chosen and two supplementary reading were performed for each of them. Thus, three expert opinions will be available for each examination. Additionally, in order to simplify the development of lesion detection methods, the lesions (exudates and microaneurysms) were manually annotated on some images by an ophthalmologist. These annotations were later confirmed by a second expert.

The resulting database is then divided in two subsets: e-ophtha EX and e-ophtha MA. The first one is composed by 47 images containing 12.278 exudates (Pathological\_EX) and 35 images without lesions (Healthy\_EX). On the other hand, e-ophtha MA subset provides 148 images containing 1306 microaneurysms (Pathological\_MA) and 233 images without lesions (Healthy\_MA). As previously mentioned, both exudates and microaneurysms, were manually annotated by experts and the ground-truth is also provided along with the pathological images. Figure 13 shows an image of the e-ophtha database and the corresponding ground-truth mask. It is important to mention that several images among the healthy ones contain structures, such as reflections and anatomical structures, which can easily mislead lesion detection methods.

All the retinal images of this database were obtained using a field of view angle of 40°. The spatial resolutions of the images are reported in Table 3.

It is important to note that only the e-ophtha EX subset was used to develop the systems presented in this work.

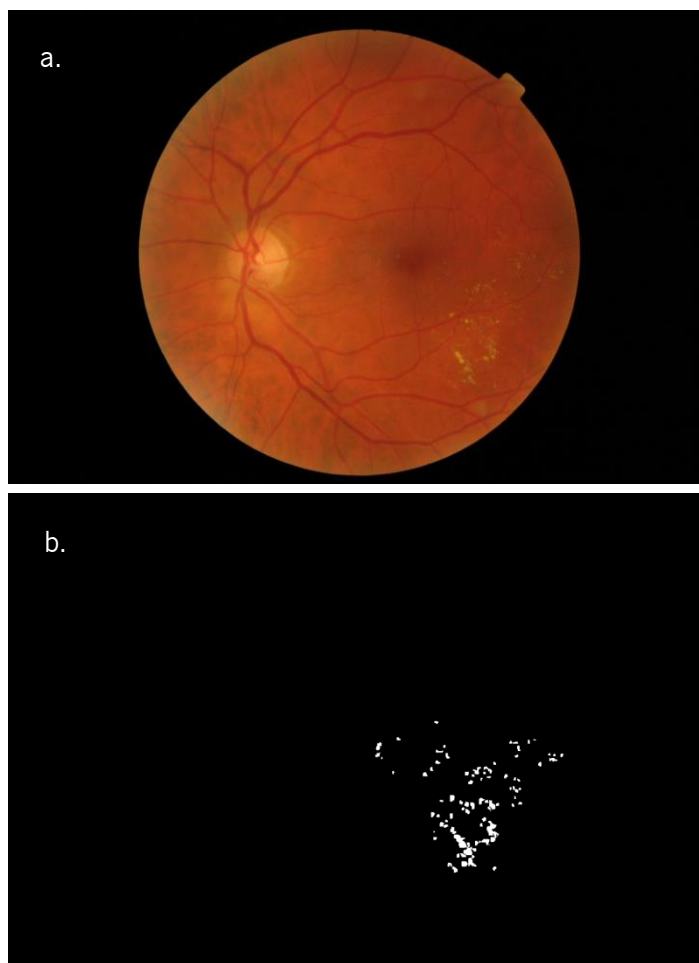


Figure 13. Original image and corresponding ground-truth mask. (a) e-optha EX image, (b) exudates mask.

Table 3. Distribution of the images of e-optha database according to their resolutions.

<i>Image Resolution</i>	<i>1440 x 960</i>	<i>1504 x 1000</i>	<i>2048 x 1360</i>	<i>2544 x 1696</i>
<i>Pathological_EX</i>	13	2	9	23
<i>Pathological_MA</i>	54	5	17	72
<i>Healthy_EX</i>	17	0	5	13
<i>Healthy_MA</i>	96	39	1	97



## 3.2 COMPUTATION OF WATERPIXELS

There are two main stages associated with the computation of waterpixels and with most watershed-based segmentation methods: the definition of the markers - from which the flooding starts - and the definition of the gradient, which corresponds to the image to be flooded. The watershed transformation enables to obtain a low-level segmentation which, alone, already shows good adherence to object boundaries. In this work, the waterpixel-generation method was designed in order to enforce the regularity of the resulting regions while keeping a reasonable trade-off between boundary adherence and regularity, therefore, it involved the following steps:

1. Creation of a grid of regular cells.
2. Computation of the morphological gradient of the image.
3. Selection of one marker per cell of the grid defined in 1.
4. Computation of a spatially regularized gradient to guarantee a compromise between boundary adherence and regularity.
5. The watershed transformation is performed on the spatially regularized gradient defined in 4, starting the flooding from the markers defined in 3.

These steps are developed in the following subsections and illustrated in Figure 14.

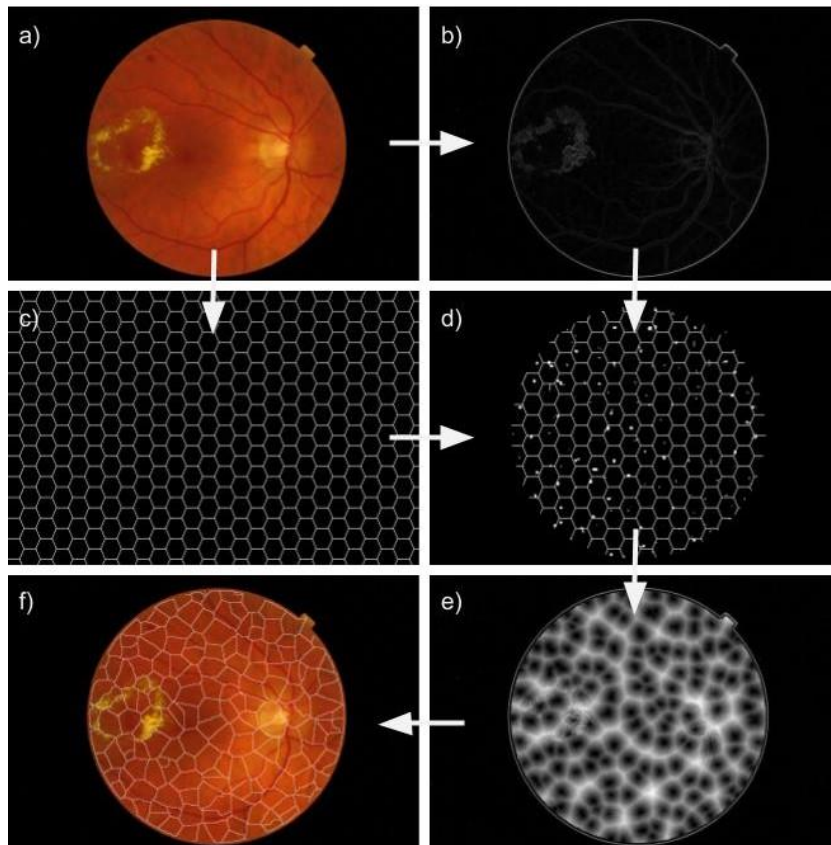


Figure 14. Illustration of m-Waterpixels generation. (a) Original image, (b) Morphological gradient of the original image, (c) Regular grid of hexagonal cells ( $s = 32$ ), (d) Selected markers within the regular grid, (e) Spatially regularized gradient, (f) m-Waterpixels.

### 3.2.1 CREATION OF THE GRID AND DEFINITION OF THE GRADIENT

The first step consists in computing a grid of regular cells  $C_i$ . A grid of hexagons is created with the size of the input image (see Figure 15). Note that the hexagon side  $s$  is a tunable parameter. The smaller  $s$  is, the smaller the resulting waterpixels and the higher the number of resulting regions. The distance between adjacent hexagon centers is denoted as  $\sigma$  and it plays a normalization role in the waterpixel computation process.

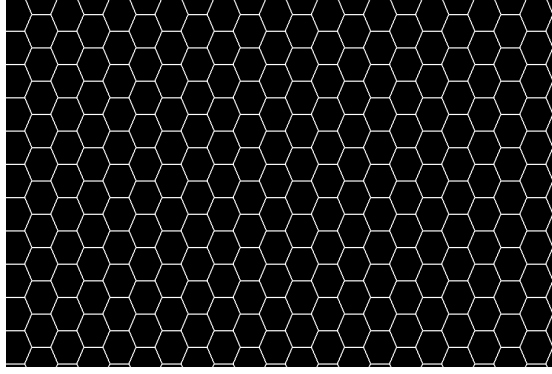


Figure 15. Regular grid of hexagonal cells ( $s = 32$ ).

At this stage, the computation of the morphological gradient  $g$  from the input image  $f : D$  (where  $D$  is a rectangular subset of  $\mathbb{Z}^2$ ) is performed (see Figure 16 for an illustration). The morphological gradient can be defined as:

$$g(f) = \delta_B(f) - \varepsilon_B(f) \quad (20)$$

where  $\delta$  and  $\varepsilon$  are, respectively, the operators dilation and erosion and  $B$  is a unitary structuring element.

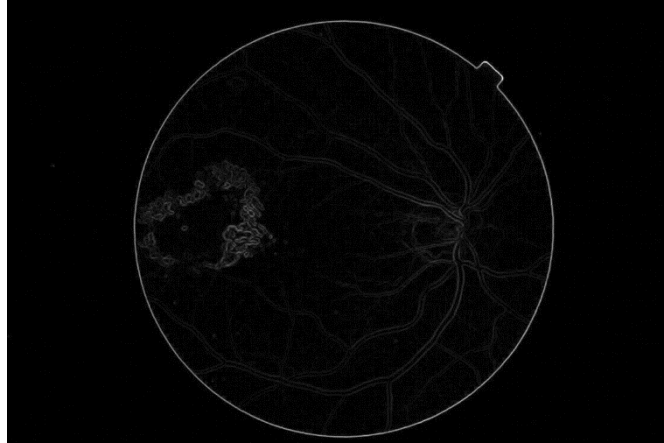


Figure 16. Morphological gradient of the original image.

### 3.2.2 SELECTION OF THE MARKERS

The selection of markers allows us to control the number and regularity of the resulting waterpixels: the number of markers is equal to the number of waterpixels in the final partition; to obtain regions

which are similar in shapes and sizes, it is necessary to select the markers in a way that they are regularly spaced out over the image. Taking this into account, a unique marker per cell is selected to obtain total control over the number of waterpixels, and a strong impact on their size and shape. During this procedure, it is necessary to have in mind that the ideal is to find a marker that enables to obtain the best performance in terms of boundary adherence and regularity. Therefore, two approaches are tested: selecting the centers of the grid cells as markers (c-Waterpixels) and selecting one minimum of the gradient per cell as markers (m-Waterpixels). In the second approach, each cell  $C_i$  of the grid defines a region of interest where the content of  $g$  is analyzed to select a unique marker, taking into account the following considerations:

- if there exist more than one minimum of  $g$  inside  $C_i$ , the one with the highest surface extinction value [23] is selected.
- if there is no minimum of  $g$  inside  $C_i$ , the center of the cell in question ( $C_i$ ) is defined as marker.
- if there is a unique minimum of  $g$  inside  $C_i$ , this minimum is obviously selected as marker.

It should be noted that the decision to add a marker for the cell in the second case is made in order to keep regularity. Figures 17 and 18 illustrate the markers selected in order to compute c-Waterpixels and m-Waterpixels, respectively.

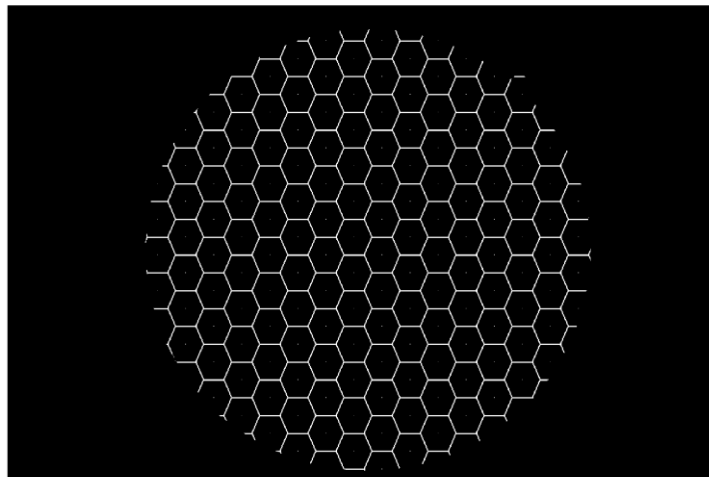


Figure 17. Selected markers within the regular grid in order to generate c-Waterpixels.

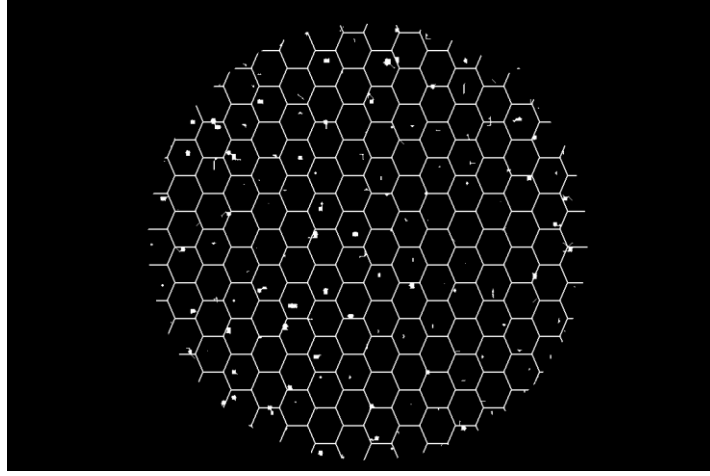


Figure 18. Selected markers within the regular grid in order to generate m-Waterpixels.

### 3.2.3 SPATIAL REGULARIZATION OF THE GRADIENT AND WATERSHED

A spatially regularized gradient  $g_{reg}$  is computed to guarantee a compromise between boundary adherence and regularity:

$$g_{reg} = g + kd_Q \quad (21)$$

where  $g$  is the gradient of the image and  $k$  is the spatial regularization parameter, which, in this work, is set to 1. Let  $Q = \{q_i\}_{1 \leq i \leq N}$  be a set of  $N$  connected components of the image  $f$ . For all  $p \in D$ , the distance  $d_Q$  is defined as:

$$\forall p \in D, d_Q(p) = \frac{2}{\sigma} \min d(p, q_i) \quad (22)$$

where  $\sigma$  is the grid step defined in the subsection 3.2.1. Figures 19 and 20 represent the spatially regularized gradient obtained in order to compute c and m-Waterpixels, respectively.

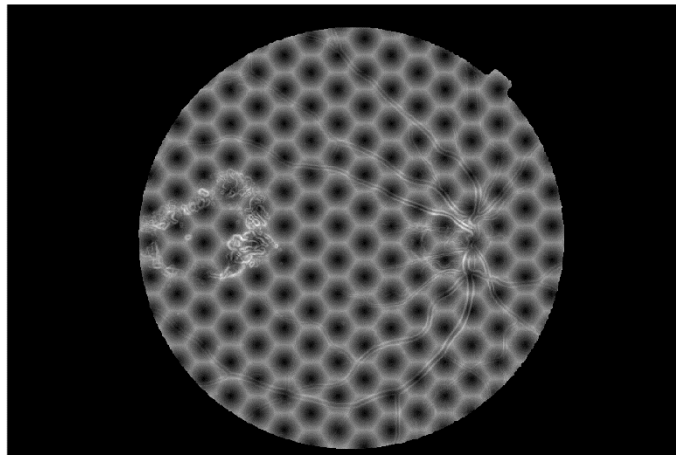


Figure 19. Spatially regularized gradient obtained in order to compute c-Waterpixels.

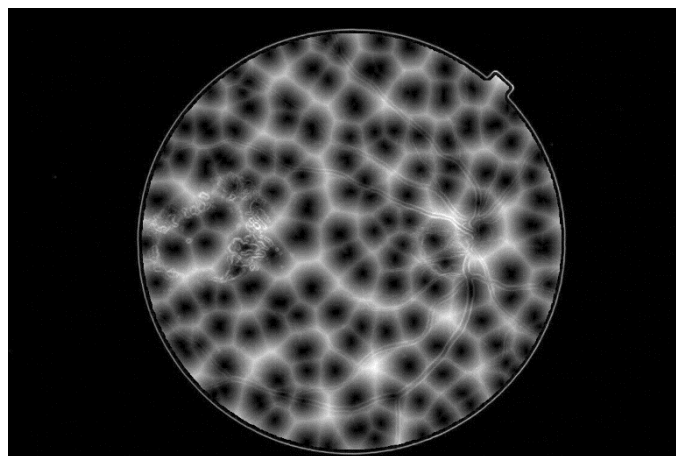


Figure 20. Spatially regularized gradient obtained in order to compute m-Waterpixels.

Finally, the watershed transformation is performed on the spatially regularized gradient  $g_{reg}$ , starting the flooding from the markers. This allowed the partition of the image into waterpixels regions. Figures 21 and 22 show the resulting c and m-Waterpixels, respectively.

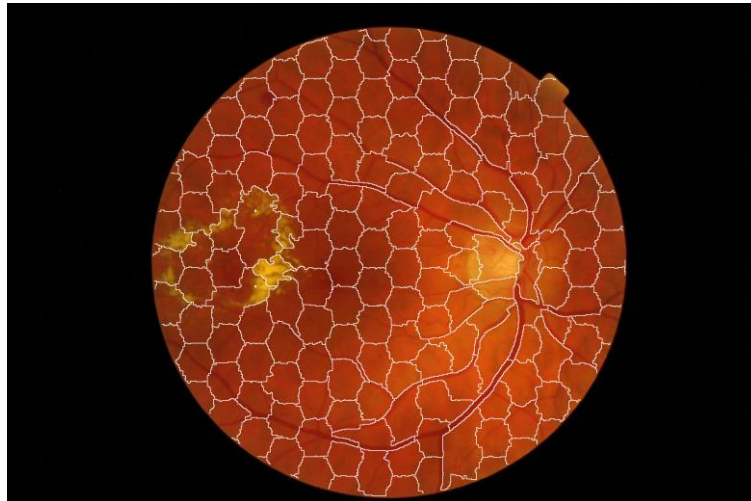


Figure 21. c-Waterpixels.

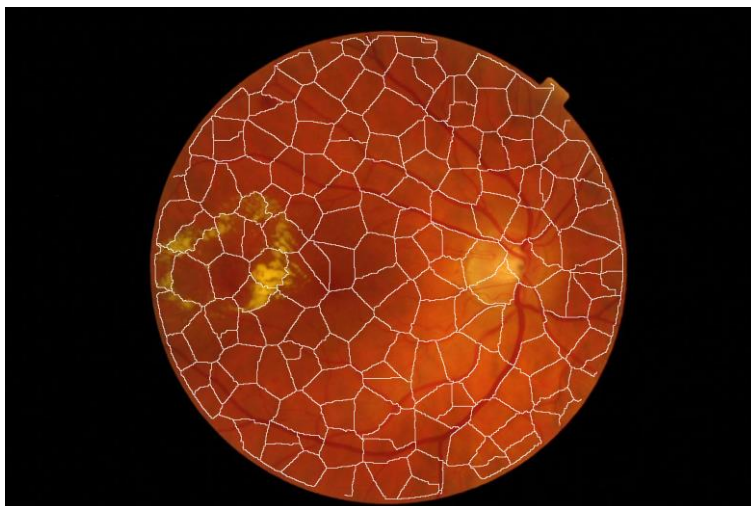


Figure 22. m-Waterpixels.

### 3.3 SYSTEM BASED ON TRADITIONAL METHODS

#### 3.3.1 INTRODUCTION

Several methods related to the classification of the retinal tissue have been proposed in the literature, these can be grouped in: thresholding-based [66], [67], region growing-based [5], [68] and morphological-based classification [69]. These methods focus their efforts in the exudates segmentation, however, these approaches are characterized by presenting a high false-positive rate. For this reason, in the methodology proposed in this thesis the segmentation step is avoided and the

characterization of healthy and pathological tissue is performed by applying image descriptors in a local way. Methods involving feature extraction and classification of the retinal tissue have been studied in the literature. The most common procedure requires the extraction of features from a lesion candidate map generated by different techniques, such as: mathematical morphology [6], [67]; background subtraction [70]; clustering [71]; or using banks of filters and applying a low adaptive threshold [72], [73]. The strategies proposed in this thesis do not require the previous segmentation of exudates or the generation of candidate maps. Texture and morphological features are extracted from the fundus images through powerful descriptors.

It is important to note that, through the method presented in the previous section, the image is divided in non-regular regions (waterpixels) before the feature extraction step since the exudates usually represent less than one percent of the total number of pixels that compose the retinal image. Thus, during the feature extraction, texture and morphological features vectors are extracted for each region through LBP and granulometry, respectively. Finally, during the classification stage, each region is classified in “healthy” or “pathological” using the SVM classifier.

The regions containing optic disk pixels are not considered during the feature extraction stage. To detect the optic disk, the method proposed in [74] is used. Figure 23 illustrates the results obtained by applying this method.



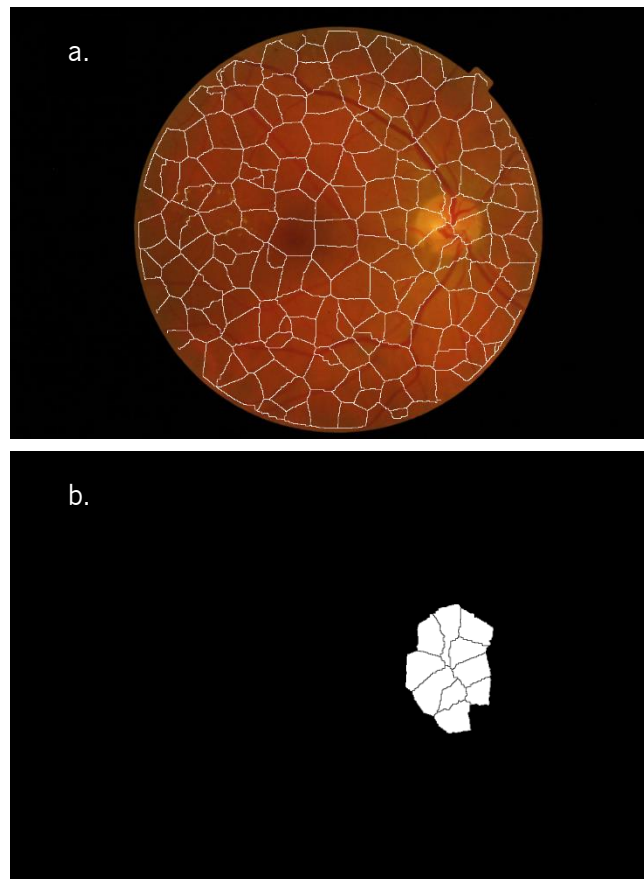


Figure 23. Optic disk detection results. (a) Waterpixels, (b) Optic disk regions.

### 3.3.2 FEATURE EXTRACTION

$LBP_{P,R}^{riu2}$  and  $VAR_{P,R}$  operators are locally computed for each pixel of the green channel of the retinal image, using the public implementation<sup>1</sup> of the operators. It should be noted that the green channel was selected since it provides a better contrast between the lesions and the background [75]. LBP and VAR images of dimensions  $M_1 \times M_2$  are obtained, as a result. These resulting images are divided into waterpixels regions and normalized histograms are computed for each region combining the information provided by both images. It is worth noting that the texture descriptor is applied locally since usually exudates represent less than one percent of the total number of pixels in the retinal image.

LBP variance (LBPV) histogram [76] accumulates the  $VAR_{P,R}$  value for each  $LBP_{P,R}^{riu2}$  label inside the waterpixel:

<sup>1</sup> [http://www.ee.oulu.fi/~gyzhao/LBP\\_Book.htm](http://www.ee.oulu.fi/~gyzhao/LBP_Book.htm)

$$LBPV_{P,R}(k) = \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} w(LBP_{P,R}(i,j), k), \quad k \in [0, K] \quad (23)$$

where

$$w(LBP_{P,R}(i,j), k) = \begin{cases} VAR_{P,R}(i,j), & \text{if } LBP_{P,R}(i,j) = k \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

where  $K$  is the maximal LBP label.

In the end, feature vectors of ten  $(P+2)$  components are locally extracted from the green channel of the fundus images.

Additionally, granulometry is used in this work by applying a series of morphological opening (closing) operations with increasing size structuring elements (SE) defined by a specific step ( $s = 2$ ) and a maximum value ( $n_{max} = 22$ ), to obtain a local description of the shape and size of the retinal exudates. As explained, SE is a disk or a line giving place to an isotropic or angular granulometry, respectively. In the end, four feature vectors are obtained: two regarding the isotropic granulometry (opening and closing operations) and two regarding the angular granulometry. Thus, these pattern spectrums are combined giving place to the morphological feature vector composed by 44 elements. Note that angular granulometry is computed in the directions  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$ .

### 3.3.3 CLASSIFICATION OF THE RETINAL TISSUE

After the feature extraction step, the SVM classifier is used to classify each region in healthy or pathological. However, before applying this classifier, some other steps were involved in the creation of the machine learning method used to classify the retinal tissue.

To begin, an external  $K$ -fold cross validation is performed to provide robustness to the decision system and to guarantee that all the instances are used in the creation of the model and in the prediction stage. A different partition is selected as test subset while the rest of the partitions are used to train the model.

As it was previously discussed, exudates represent only a small percentage of the total number of pixels that compose the retinal image, which results in a very unbalanced dataset when the feature

extraction is performed. Training the classifier with this dataset can result in an overfitting to the class “healthy”.

To avoid this problem, the set of all healthy samples is randomly permuted and partitioned into  $T = \text{round}\left(\frac{M}{N}\right)$  subsets, being  $M$  and  $N$  the number of healthy and pathological samples, respectively. After that, a set of  $T$  classifiers is trained using all pathological training samples and each partition of healthy training samples. Finally, during the test stage, testing samples are evaluated for each of the  $T$  models and soft majority voting is applied to the output probabilities as the final criterion. If the obtained probability is higher than a given threshold  $\delta$ , the region is assigned to the class “pathological”.

The classification of the regions according to healthy and pathological classes was then performed using the Support Vector Machines (SVM) classifier. Radial Basis Function (RBF) kernel is employed when the SVM algorithm is applied. RBF kernel non-linearly maps samples into a higher dimensional space so, unlike the linear kernel, it can deal with datasets in which the relation between the class labels and features is nonlinear. The referred kernel is characterized by two parameters ( $C, \gamma$ ). The parameter  $C$  trades off misclassification of training examples against simplicity of the decision surface while the parameter  $\gamma$  defines how far the influence of a single training example reaches. The linear kernel is a specific case of RFB since it presents the same performance as the RBF kernel with a particular couple of values for the parameters  $C$  and  $\gamma$ . It should be noted that the RBF kernel is employed in this thesis because it is highly suitable when the number of attributes is not very large [54].

For a given problem, the best  $C$  and  $\gamma$  parameters are unknown, and the objective is to identify the optimal  $(C, \gamma)$  using the training set, so that the classifier can accurately predict unknown data (i.e. testing data). Internal cross-validation is used in the experiments involving SVM with RBF kernel in order to “grid search” the optimal parameters. Several pairs of  $(C, \gamma)$  values are tried and the one with the highest cross-validation accuracy is selected.  $C = 2^{-5}, 2^{-3}, \dots, 2^{15}$  and  $\gamma = 2^{-15}, 2^{-13}, \dots, 2^3$  are the sequences used in the SVM classification of healthy and pathological tissue.

## 3.4 SYSTEM BASED ON DEEP LEARNING

### 3.4.1 INTRODUCTION

In the previously described system, image descriptors were used in order to extract features from the visual information of retinal images. Subsequently, using the extracted features, a classification algorithm was applied in order to identify exudates. However, the success and reliability of this type of detection system is conditioned by the descriptors used to perform the feature extraction. The aforementioned descriptors are usually not dedicated to solving a specific problem, being applicable to most computer vision tasks and, as such, each application is highly dependent on the optimal selection of the descriptors parameters.

The choice of suitable descriptors is dependent on the researcher's experience and the help of medical specialists may also be fundamental to understand which descriptors may best define the pathology to detect or classify. Therefore, it is possible to say that this task is characterized by a subjective character.

The above-mentioned facts motivated the study of a different technique able to automatically extract relevant features of the image without the need of determining the optimal descriptors by the researchers. This new approach will be explored in the following subsection and it focus on the detection of exudates using Convolutional Neural Networks (CNNs).

### 3.4.2 CNN FOR EXUDATE DETECTION

As in the previous system, the images of the E-OPHTHA database are divided in irregular regions, through the methodology presented in Section 3.2. Additionally, to the earlier mentioned reasons, the division of the original images is also important to allow the acquisition of a significantly larger dataset since hundreds of regions are obtained from each of the 47 images of the E-OPHTHA database. This point is extremely decisive for the obtainment of a robust model that isn't overfitting. Remember that the number of resulting waterpixels depends on the tunable parameter  $s$ , as it was explained in Section 3.2.1. The smaller  $s$  is, the smaller the resulting waterpixels and the higher the number of resulting regions. In the experiments presented in this dissertation, the parameter  $s$  was set to 32.

Before the computation of the waterpixels, the 47 images of the E-OPHTHA database are divided in two subsets: one for training (containing 80% of the images) and the other one for testing (containing 20% of the images). As its name suggests, the first one is responsible for training the neural network while the testing subset is used to evaluate the performance of the model, i.e., how well the model discriminates between healthy and damaged retinal tissue. It is relevant to mention that the testing set does not contain any image used to perform the training.

Subsequently, the training dataset is divided in training (80%) and validation (20%). The validation set is used during the training to provide an unbiased evaluation of the model while finding the optimal model hyperparameters.

After the creation of the three datasets (training, validation and testing), the images are divided in waterpixels, through the methodology detailed in Section 3.2. Additionally, the images containing the medical specialists' annotations, i.e., the ground-truth, are looped in order to obtain the corresponding label ("healthy" or "pathological") for each waterpixel. As expected, the number of healthy waterpixels is way larger than the number of pathological ones, resulting in very unbalanced datasets. To overcome this problem, all the pathological regions were used, and the same number of healthy regions was randomly selected. Figure 24 shows some of the healthy and pathological m-waterpixels provided to the CNN.

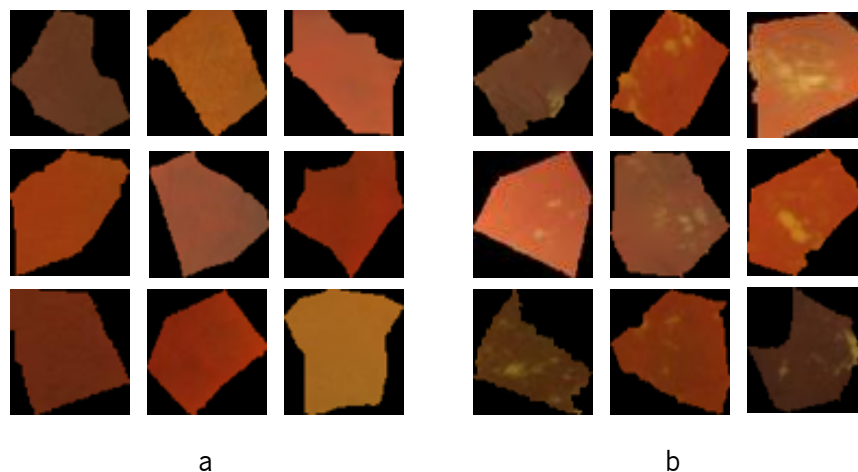


Figure 24. m-Waterpixels extracted from the E-OPHTHA database and provided to the CNN. (a) m-Waterpixels belonging to the class "healthy", (b) m-Waterpixels belonging to the class "pathological".

Shortly, the final dataset is composed by three main subsets - training, validation and testing – and each of these subsets contain a healthy and a pathological dataset. This final dataset is then provided to the CNN architecture to be fine-tuned in order to create a system able to classify the retinal tissue

in healthy or pathological. As previously discussed, the main obstacle in medical applications is the reduced amount of labelled data since the pixel-wise annotation is expensive and time-consuming. Therefore, fine-tuning is applied in order to transfer the knowledge of healthy and pathological retinal tissue to some of the most relevant state-of-the-art models pre-trained on the ImageNet dataset. More precisely, InceptionV3 [77], Xception [78], ResNet50 [79], VGG16 and VGG19 [80] architectures, as well as their corresponding weights, are obtained through the Keras Applications module.

It is important to mention that the weights from the pre-trained network trained with the ImageNet database are transferred to the CNN architecture, except for the last fully connected layer. This one is replaced with a new fully connected layer that has as many neurons as the number of classes in our application, which is two – “healthy” and “pathological”. After the initialization of the CNN with the publicly available weights, the neural network was shallow-tuned, i.e., the weights related to the first layers were frozen while the last layers were set to “learnable” (that means that only the last layers are fine-tuned). Before the training process, data augmentation was performed by rotating, flipping and scaling the waterpixels.



# 4 RESULTS AND DISCUSSION





## 4.1 EVALUATION METRICS

In order to evaluate the classification performance of the experiments, the following metrics, which are commonly used in machine learning and of extreme relevance in the medical field, were used: accuracy, sensitivity and specificity. Accuracy shows how correct a diagnostic test identifies and excludes a given condition. It is calculated as the sum of correct predictions divided by the total number of cases. Sensitivity, also known as True Positive Rate (TPR), measures the proportion of positive cases that are correctly identified as such. On the other hand, Specificity or True Negative Rate (TNR) measures the proportion of negative cases that are correctly recognized.

In addition to the above-mentioned metrics, the area under the Receiver Operating Characteristic (ROC) curve (AUC) was also studied in order to evaluate the performance of the proposed methods. The ROC curve is a plot of the TPR against the False Positive Rate (1-TNR). Thus, AUC is an effective way to summarize the overall diagnostic accuracy of the test. The greater the AUC, the more accurate the test - a perfect test has an area under the ROC curve of 1. A diagnostic test in the medical field with an AUC of 0.5 suggests no ability to diagnose patients. On the other hand, an AUC between 0.7 and 0.8 is considered acceptable, between 0.8 and 0.9 is considered excellent, and bigger than 0.9 is considered outstanding [81].

## 4.2 SYSTEM BASED ON TRADITIONAL METHODS

The experiments presented in this section are performed using the first system proposed in this work, i.e, the system based on traditional methods: feature extraction and posterior classification of the retinal tissue using machine learning algorithms. The referred experiments focus on establishing a comparison between the two proposed strategies for generating waterpixels (c-Waterpixels and m-Waterpixels) and the state-of-the-art SLIC superpixels.

### 4.2.1 COMPARISON OF THE SUPERPIXELS GENERATION METHODS

Two different strategies to obtain waterpixels (c-Waterpixels and m-Waterpixels) were presented in this work. In this first experiment, these regions are used to perform a local feature extraction, and, subsequently, the procedure explained in Section 3.3.3. is used to train a classification model able to classify the retinal tissue of each region. With the purpose of comparing the strategies presented

in this work with another state-of-the-art method, the same procedure is applied using the SLIC algorithm to generate superpixels. In Figure 25 the three different types of superpixels being compared in this experiment are illustrated. It must be highlighted that, to allow a fair comparison between the different strategies, the K-fold procedure and the random permutation for balancing the data, used during the classification stage as explained in Section 3.3.3., are performed using a fix seed.

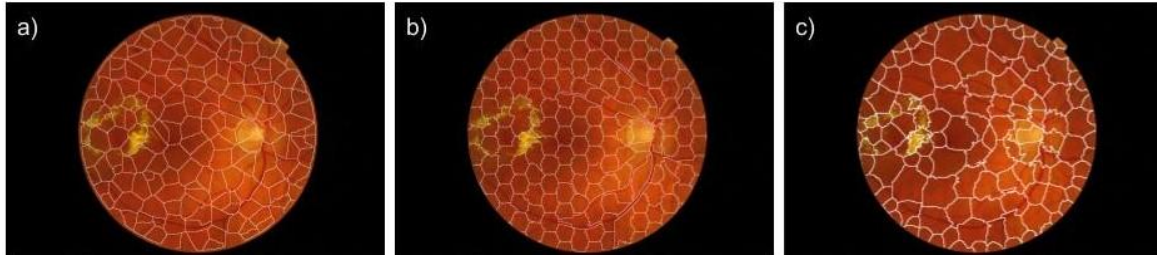


Figure 25. Types of superpixels being compared. (a) m-Waterpixels, (b) c-Waterpixels, (c) SLIC Superpixels.

The tunable parameters involved in the different stages of the exudate detection system were set according to the Table 4.

Table 4. Tunable parameters involved in the different stages of the exudate detection system.

	<i>Computation of Waterpixels</i>		<i>Feature Extraction</i>		<i>Classification</i>
	m-Waterixels	c-Waterpixels	LBP	Granulometry	SVM
<i>Parameters</i>	$s = 32$	$s = 32$	$R = 1$	$s = 2$	RBF Kernel
			$P = 8$	$n_{max} = 22$	

Table 5 contains the Area Under ROC Curve (AUC), accuracy, sensitivity and specificity (as well as the standard deviation associated with each of these metrics) of the resulting classification processes, taking into account the ground-truth provided by the ophthalmologists. The results of accuracy, sensitivity and specificity were obtained using a decision threshold of  $\delta = 0.5$ .

Table 5. AUC, accuracy, sensitivity and specificity related to the exudate detection on the original images (OI) of the E-OPHTA database, using a linear kernel, for each of the region of interest developed: m-waterpixels, c-Waterpixels and SLIC superpixels.

	<i>m-Waterpixels</i>	<i>c-Waterpixels</i>	<i>SLIC Superpixels</i>
<i>AUC</i>	$0.7999 \pm 0.0317$	$0.7937 \pm 0.0325$	$0.7987 \pm 0.0311$
<i>Accuracy</i>	$0.6514 \pm 0.1093$	$0.6650 \pm 0.0870$	$0.7635 \pm 0.0238$
<i>Sensitivity</i>	$0.7779 \pm 0.0987$	$0.7651 \pm 0.0477$	$0.6567 \pm 0.0824$
<i>Specificity</i>	$0.6363 \pm 0.1295$	$0.6546 \pm 0.0990$	$0.7736 \pm 0.0305$

During the evaluation of the different techniques it is important to remember the importance of obtaining a good trade-off between the sensitivity and specificity. Additionally, in the context of this thesis, sensitivity plays a major role since this measure represents the ability of our system to detect exudates. Having that in mind, and through the analysis of Table 5, it is possible to conclude that the waterpixels approaches outperform the SLIC [25] method when it comes to generate the most appropriate regions to detect exudates.

#### 4.2.2 INFLUENCE OF THE BLOOD VESSELS REMOVAL

In this second experiment, the impact of applying the inpainting technique<sup>2</sup> in order to remove the blood vessels from the fundus images is evaluated. With this in mind, and as in the previous experiment, texture and morphological features are extracted for each of the irregular regions through LBP and granulometry, respectively, and SVM with RBF kernel is applied as classification algorithm, following the classification strategy detailed in Section 3.3.3. It should be noted that the retinal vessel removal is performed before the local feature extraction.

Several tests using the original images (OI) of the E-OPHTA\_EX database and the same images but without the blood vessels (W/V) are performed for each of the strategies to generate irregular regions to perform the local feature extraction. Once again, to allow a fair comparison between the different strategies, the K-fold procedure and the random permutation for balancing the data, used during the classification stage, are performed using a fix seed.

Table 6 contains the Area Under ROC Curve (AUC), accuracy, sensitivity and specificity (as well as the standard deviation associated with each of these metrics) of the resulting classification processes,

<sup>2</sup> [https://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint\\_nans](https://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint_nans)

taking into account the ground-truth provided by the ophthalmologists. Once again, the results of accuracy, sensitivity and specificity, presented on Table 6, were obtained using a decision threshold of  $\delta = 0.5$ .

Table 6. AUC, accuracy, sensitivity and specificity related to the exudate detection on the original images (OI) and images without vessels (W/V) for each of the region of interest developed: m-Waterpixels, c-Waterpixels and SLIC superpixels.

	<i>m-Waterpixels</i>		<i>c-Waterpixels</i>		<i>SLIC Superpixels</i>	
	OI	W/V	OI	W/V	OI	W/V
<i>AUC</i>	0.7998 ±	0.8287 ±	0.7937 ±	0.8277 ±	0.7987 ±	0.8169 ±
	0.0317	0.0246	0.0325	0.0199	0.0311	0.0299
<i>Accuracy</i>	0.6514 ±	0.7585 ±	0.6650 ±	0.7545 ±	0.7635 ±	0.8037 ±
	0.1093	0.0676	0.0870	0.0679	0.0238	0.0222
<i>Sensitivity</i>	0.7779 ±	0.7216 ±	0.7651 ±	0.7496 ±	0.6567 ±	0.6656
	0.0987	0.0834	0.0477	0.0295	0.0824	±0.0634
<i>Specificity</i>	0.6363 ±	0.7595 ±	0.6546 ±	0.7539 ±	0.7736 ±	0.8175 ±
	0.1295	0.0789	0.0990	0.0757	0.0305	0.0273

As it can be observed through the analysis of Table 6, the values of accuracy improve, for each of regions of interested developed, when the inpainting technique is applied in order to remove the blood vessels from the original images.

The ROC curves for each of the tests performed with the images without vessels are represented in Figure 26. As it can be observed, the c-Waterpixels outperform the other methods for lower false positive rates while the m-waterpixels exceed the c-Waterpixels and SLIC superpixels for highest false positives rates.

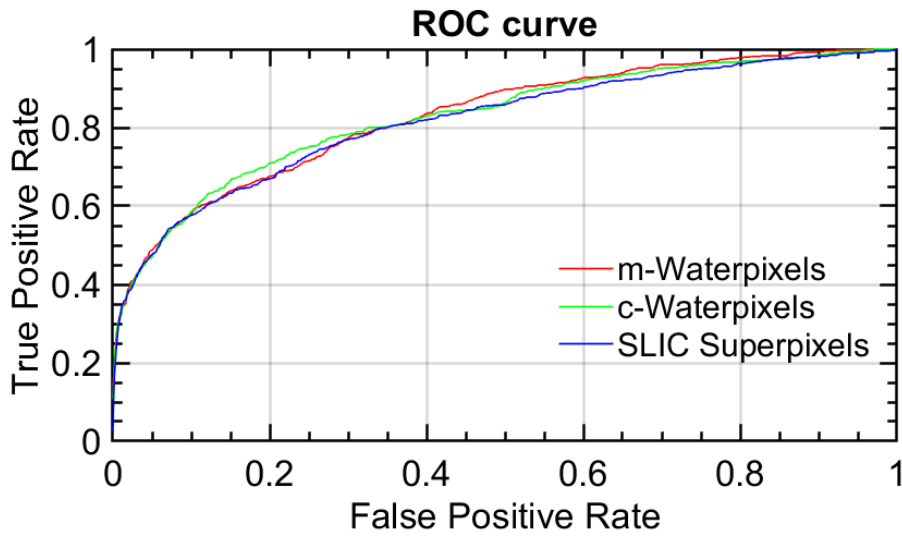


Figure 26. ROC curves for the different superpixels generation methods.

### 4.3 SYSTEM BASED ON DEEP LEARNING

The experiments presented in this section are performed using the second system proposed in this work, i.e, the system based on deep learning techniques. The referred experiments focus on establishing a comparison between the different fine-tuned CNN architectures used.

#### 4.3.1 COMPARISON OF THE FINE-TUNED CNN ARCHITECTURES

In this experiment, the inpainting technique is applied in order to remove the blood vessels of the images of the E-OPHTHA database. Subsequently, m-Waterpixels are computed (through the methodology presented in Section 3.2.) and provided to the CNN architecture to be fine-tuned along with their corresponding labels (“healthy” or “pathological”). Fine-tuning is then applied in order to transfer the knowledge of healthy and pathological retinal tissue to some of the most relevant state-of-the-art models pre-trained on the ImageNet dataset. More precisely, InceptionV3, Xception, ResNet50, VGG16 and VGG19 architectures, as well as their corresponding weights, are obtained through the Keras Applications module and fine-tuned, following the strategy detailed in Section 3.4.2.

Table 7 contains the accuracy, sensitivity and specificity (as well as the standard deviation associated with each of these metrics) obtained for each fine-tuned model. Table 8 illustrates the optimizer used for each of the models, in order to obtain the results presented in Table 7. Several optimizers were

tested for each of the models, but the ones indicated in Table 8 were the ones that allowed better results. Additionally, several tests were performed with different numbers of epochs, batch sizes and learning rates. But the results exhibit in Table 7 were obtained using 100 epochs, a batch size equal to 16 and a learning rate of  $1 \times 10^{-4}$  to perform the training process.

Table 7. AUC, accuracy, sensitivity and specificity related to the exudate detection, for each fine-tuned state-of-the-art architecture.

	<i>InceptionV3</i>	<i>Xception</i>	<i>ResNet50</i>	<i>VGG16</i>	<i>VGG19</i>
<i>Accuracy</i>	0.9254	0.9104	0.8881	0.7985	0.7090
<i>Sensitivity</i>	0.9701	0.9254	0.9701	0.8209	0.7910
<i>Specificity</i>	0.9646	0.9231	0.9643	0.8125	0.7500

Table 8. Optimizer used for each each fine-tuned state-of-the-art architecture.

<i>Model</i>	<i>Optimizer</i>
<i>InceptionV3</i>	Nesterov Adam
<i>Xception</i>	Nesterov Adam
<i>ResNet50</i>	RMSprop
<i>VGG16</i>	Stochastic Gradient Descent
<i>VGG19</i>	Stochastic Gradient Descent

Through the analyses of Table 7, it is possible to conclude that, generally, the second approach presented in this thesis outperforms the first one. In other words, the approach involving CNNs presents considerably better results than the one based on feature extraction and posterior classification using machine learning algorithms.

In addition, from all the tested CNN architectures, InceptionV3 is the best candidate to solve the problem under study. If the objective is to maximise the sensitivity, in other words, the ability of the model for recognizing abnormal tissue, the most appropriate models are InceptionV3 and ResNet50.

# 5 CONCLUSIONS





## CONCLUSIONS

The main objective of this work was to develop and compare different local analysis strategies for exudate detection in fundus images. Therefore, different strategies to obtain regions of interest were developed with the purpose of locally describing healthy and pathological retinal areas.

Contrary to most approaches found in the literature, in the methodologies proposed in this thesis, the characterization of healthy and damaged retinal areas was performed by applying image descriptors in a local way, avoiding the segmentation step and the generation of candidate maps.

A system based on local feature extraction and Support Vector Machine classification was used to develop and compare different strategies for the automated detection of exudates. The main novelty is allowing the detection of exudates using non-regular regions to perform the local feature extraction. To accomplish this objective, different strategies for generating superpixels were applied to the fundus images of the E-OPHTA database. Those strategies relied on applying the marker-controlled watershed transformation to a spatially regularized gradient. From these strategies, two different types of superpixels were created: c-Waterpixels and m-Waterpixels.

Local binary patterns and granulometric profiles were used to extract texture and morphological features from each waterpixel. Subsequently, the SVM classifier was fed with the features extracted in order to classify the retinal tissue of each region into healthy or pathological. This process was performed using the c-waterpixels, m-waterpixels and the state-of-the-art SLIC superpixels to partition the original images of the E-OPHTA database.

Through the analyses of the classification results obtained using the SVM classifier, it was possible to conclude that the watershed-based approaches lead to a better detection of pathological areas. In other words, the waterpixels approaches presented in this thesis outperform the state-of-the-art SLIC method when it comes to generate the most appropriate regions to detect exudates.

The above described process was performed an additional time but using the images without blood vessels. This experiment allowed to conclude that applying the inpainting technique to remove the blood vessels is essential to obtain a more accurate detection of exudates.

Additionally, a system based on Convolutional Neural Networks (CNN) was explored to discriminate between healthy and pathological regions in fundus images. Transfer learning was applied to fine-tune some of the most important state-of-the-art CNN architectures. In this approach, the fundus images were divided in m-waterpixels and the classification was performed for each of the regions.

Through the analyses of the classification results obtained with this approach it was possible to conclude that, from all the tested CNN architectures, InceptionV3 is the best candidate to solve the problem under study. Comparing the classification results obtained through the system involving

CNNs and the ones obtained applying the approach based on feature extraction and subsequent classification using machine learning algorithms, it was possible to conclude that the CNN approach outperforms the first system presented to automatically detect exudates.

In both the systems presented in this thesis, the characterization of healthy and pathological retinal tissue was performed in a local way, i.e., the classification models were trained using partitions of the image (superpixels). This fact provides the computer-aided diagnosis system with a high precision in the lesion identification.

Despite the positive results obtained, there is still space to investigate in this field. To begin, future work will allow the detection of other lesions related to DR, such as microaneurysms and hemorrhages. As the amount of data is extremely decisive for the obtainment of a robust model that isn't overfitting, the systems developed will be trained using a larger database. Additionally, the system based on deep learning will be performed using c-waterpixels and SLIC superpixels in order to understand which of the different strategies to obtain regions of interest allows a more accurate detection of exudates. Still regarding the deep learning approach, changing and study the parameters like the number of epochs, the batch size or even the amount of clinical information can also be interesting. Regarding to the method based on image description and classification algorithms, as future work, other advanced classification algorithms will be tested in order to improve the image-wise detection of exudates. Additional information will also be extracted from the retinal background in order to increase the precision in the identification of DR lesions. Finally, an approach to classify both lesions and anatomical elements is a further direction since the methodologies proposed require the previous segmentation of the blood vessels and the optic disk, which is not error-free.

# REFERENCES



## REFERENCES

- [1] D. Sidibé, I. Sadek, and F. Mériaudeau, "Discrimination of retinal images containing bright lesions using sparse coded features and SVM," *Comput. Biol. Med.*, vol. 62, pp. 175–184, 2015.
- [2] W. Zhou, C. Wu, Y. Yi, and W. Du, "Automatic Detection of Exudates in Digital Color Fundus Images Using Superpixel Multi-Feature Classification," *IEEE Access*, vol. 5, no. 1, pp. 17077–17088, 2017.
- [3] B. Antal and A. Hajdu, "An ensemble-based system for automatic screening of diabetic retinopathy," *Knowledge-Based Syst.*, vol. 60, pp. 20–27, 2014.
- [4] "WHO | Universal eye health: a global action plan 2014–2019," *WHO*, 2017.
- [5] C. Sinthanayothin *et al.*, "Automated detection of diabetic retinopathy on digital fundus images," *Diabet Med*, vol. 19, no. 2, pp. 105–112, 2002.
- [6] T. Walter, J. Klein, P. Massin, and A. Erginay, "A Contribution of Image Processing to the Diagnosis of Diabetic Retinopathy – Detection of Exudates in Color Fundus Images of the Human Retina," vol. 21, no. 10, pp. 1236–1243, 2002.
- [7] K. Rogers and Britannica Educational Publishing., *The Eye: the Physiology of Human Perception*. Britannica Educational Pub, 2011.
- [8] K. M. Meek, "The Cornea and Sclera," in *Collagen*, Boston, MA: Springer US, 2008, pp. 359–396.
- [9] Abraham Metz, *The Anatomy and Histology of the Human Eye*. .
- [10] W. W. Campbell and R. N. DeJong, *DeJong's the neurologic examination*. .
- [11] L. Remington and D. Goodwin, *Clinical Anatomy and Physiology of the Visual System*, 3rd ed. 2012.
- [12] C. VanPutte and A. Russo, *Anatomia e Fisiologia de Seeley*, 10th ed. 2016.
- [13] "WHO | Diabetes mellitus," *WHO*, 2010.
- [14] "Diabetes - long-term effects - Better Health Channel." [Online]. Available: <https://www.betterhealth.vic.gov.au/health/conditionsandtreatments/diabetes-long-term-effects>. [Accessed: 21-Oct-2018].
- [15] Huan Wang, Wynne Hsu, Kheng Guan Goh, and Mong Li Lee, "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, pp. 181–186.
- [16] C. Sundhar and D. Archana, "Automatic Screening of Fundus Images for Detection of Diabetic Retinopathy," *Int. J. Commun. Comput. Technol.*, 2014.

- [17] J. L. Morrison, L. A. Hodgson, L. L. Lim, and S. Al-Qureshi, "Diabetic retinopathy in pregnancy: a review," *Clin. Experiment. Ophthalmol.*, vol. 44, no. 4, pp. 321–334, May 2016.
- [18] "Diabetic Retinopathy for Medical Students. Classification." [Online]. Available: <https://webeye.ophth.uiowa.edu/eyeforum/tutorials/diabetic-retinopathy-med-students/Classification.htm>. [Accessed: 21-Oct-2018].
- [19] "Diabetes symptoms: Non-proliferative retinopathy information | myVMC." [Online]. Available: <https://www.myvmc.com/diseases/diabetic-eye-disease-non-proliferative-diabetic-retinopathy/>. [Accessed: 21-Oct-2018].
- [20] "Diabetic Retinopathy, Non-Proliferative (NPDR) and Proliferative (PDR) – Cumberland Valley Retina Consultants." [Online]. Available: <http://www.retinacare.net/diabetic-retinopathy-non-proliferativ-and-proliferative/>. [Accessed: 21-Oct-2018].
- [21] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süssstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, 2012.
- [22] J. Jiang, J. Ma, C. Chen, Z. Wang, L. Wang, and S. Member, "SuperPCA: A Superpixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery," 2016.
- [23] H. A. L. Id, "Waterpixels and their application to image segmentation learning Vaïa Machairas To cite this version : HAL Id : tel-01537814 THÈSE DE DOCTORAT de l' Université de recherche Paris Sciences et Lettres Soutenue par Vaïa MACHAIRAS," 2017.
- [24] X. Ren and J. Malik, "Learning a classification model for segmentation," *Comput. Vision, Int. Conf.*, vol. 1, no. c, pp. 10–17, 2003.
- [25] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC Superpixels," *EPFL Tech. Rep. 149300*, no. June, p. 15, 2010.
- [26] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.
- [27] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, 2004.
- [28] A. P. Moore, S. J. D. Prince, J. Warrell, U. Mohammed, and G. Jones, "Superpixel lattices," *26th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR*, 2008.
- [29] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.
- [30] A. Vedaldi and S. Soatto, "Quick Shift and Kernel Methods for Mode Seeking," in *Computer*

## REFERENCES

- Vision – ECCV 2008*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 705–718.
- [31] A. Levinstein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, “TurboPixels: Fast superpixels using geometric flows,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, 2009.
- [32] L. Vincent and P. Soille, “Watersheds in digital spaces: an efficient algorithm based on immersion simulations.,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 583–598, 1991.
- [33] X. Zhang, S. Chew, Z. Xu, and N. Cahill, “SLIC superpixels for efficient graph-based dimensionality reduction of hyperspectral imagery.,” *Algorithms Technol. Multispectral, Hyperspectral, Ultraspectral Imag.*, vol. 9472, no. International Society for Optics and Photonics, 2015.
- [34] C. Hsu, “Efficient image segmentation algorithm using SLIC superpixels and boundary-focused region merging,” *2013 9th Int. Conf. Information, Commun. Signal Process.*, pp. 1–5, 2013.
- [35] S. Beucher, “The Watershed Transformation Applied to Image Segmentation,” *Proc. 10th Pfefferkorn Conf. Signal Image Process. Microsc. Microanal.*, no. July 2000, pp. 299–314, 1992.
- [36] S. Beucher and C. Lantuejoul, “Use of watersheds in contour detection,” *Int. Work. Image Process. Real-Time Edge Motion Detect. Estim.*, 1979.
- [37] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, “Segmentation and classification of hyperspectral images using watershed transformation,” *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, 2010.
- [38] S. Beucher and F. Meyer, “The morphological approach to segmentation: the watershed transformation,” *Math. Morphol. Image Process.*, pp. 433–481, 1993.
- [39] J. Massich, “Lesion segmentation in breast sonography,” 2008.
- [40] J. Stawiaski and E. Decenci, “Interactive Liver Tumor Segmentation Using Graph-cuts and Watershed,” *Liver*, no. December, 2008.
- [41] B. Andres, U. Köthe, M. Helmstaedter, W. Denk, and F. Hamprecht, “Segmentation of SBFSEM volume data of neural tissue by hierarchical classification,” *Pattern Recognit.*, pp. 142–152, 2008.
- [42] T. Walter, I. Curie, and U. Inerm, “WATERPIXELS: SUPERPIXELS BASED ON THE WATERSHED TRANSFORMATION V. Machairas, E. Decenci `ere Centre for Mathematical Morphology MINES ParisTech Fontainebleau, France Centre for Computational Biology,



- MINES ParisTech , Fontainebleau , France,” pp. 4343–4347, 2014.
- [43] P. Neubert and P. Protzel, “Compact watershed and preemptive SLIC: On improving trade-offs of superpixel segmentation algorithms,” *Proc. - Int. Conf. Pattern Recognit.*, pp. 996–1001, 2014.
- [44] Z. Hu, “WATERSHED SUPERPIXEL Shenzhen Key Laboratory of Spatial Smart Sensing and Service , School of Computer Science , Wuhan University , P . R . China,” *Int. Conf. Image Process.*, pp. 19–23, 2015.
- [45] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.
- [46] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.
- [47] M. Pietikäinen, T. Ojala, and Z. Xu, “Rotation-invariant texture classification using feature distributions,” *Pattern Recognit.*, vol. 33, no. 1, pp. 43–52, 2000.
- [48] J. Serra, “Image Analysis and Mathematical Morphology.,” 1983.
- [49] J. Goutsias, L. Vincent, and D. Bloomberg, *Mathematical Morphology and Its Applications to Image and Signal Processing*. 2006.
- [50] T. Deng and Y. Chen, “A Morphological Approach for Granulometry with Application to Image Denoising,” *Fuzzy Inf. Eng.*, 2007.
- [51] M. Khatun, A. Gray, and S. Marshall, “Morphological granulometry for classification of evolving and ordered texture images,” *Eur. Signal Process. Conf.*, no. Eusipco, pp. 759–763, 2011.
- [52] R. Ueke and E. S. Us, “Morphological Quantification of Aortic Calcification,” pp. 81–89, 2003.
- [53] I. Steinwart and A. Christmann, *Support Vector Machines*. 2008.
- [54] C.-C. Chang and C.-J. Lin, “Libsvm,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
- [55] A. Shmilovici, “Support Vector Machines,” in *Data Mining and Knowledge Discovery Handbook*, New York: Springer-Verlag, 2005, pp. 257–276.
- [56] N. Buduma and N. Locascio, *Fundamentals of deep learning: Designing next-generation machine intelligence algorithms*. 2017.
- [57] D. Hubel and T. Wiesel, “Receptive fields, binocular interaction and functional architecture in

## REFERENCES

- the cat's visual cortex," *J. Physiol.*, 1962.
- [58] "A Beginner's Guide To Understanding Convolutional Neural Networks – Adit Deshpande – CS Undergrad at UCLA ('19)." [Online]. Available: <https://adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/>. [Accessed: 24-Oct-2018].
- [59] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proc. 27th Int. Conf. Mach. Learn.*, no. 3, pp. 807–814, 2010.
- [60] A. Overview *et al.*, "Convolutional Neural Networks ( CNNs / ConvNets )."
- [61] "CS231n Convolutional Neural Networks for Visual Recognition." [Online]. Available: <http://cs231n.github.io/transfer-learning/>. [Accessed: 24-Oct-2018].
- [62] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li, "ImageNet: A large-scale hierarchical image database.," *IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 248–255, 2009.
- [63] L. Lu *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection : CNN Architectures , Dataset Characteristics and Transfer Learning Deep Convolutional Neural Networks for Computer-Aided Detection : CNN Architectures , Dataset Characteristics and Transfer," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [64] N. Tajbakhsh *et al.*, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [65] E. Decencière *et al.*, "TeleOphta: Machine learning and image processing methods for teleophthalmology," *Irbm*, vol. 34, no. 2, pp. 196–203, 2013.
- [66] S. Ali *et al.*, "Statistical atlas based exudate segmentation," *Comput. Med. Imaging Graph.*, vol. 37, no. 5–6, pp. 358–368, 2013.
- [67] X. Zhang *et al.*, "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Med. Image Anal.*, vol. 18, no. 7, pp. 1026–1043, 2014.
- [68] H. Li and O. Chutatape, "Automated Feature Extraction in Color Retinal Images by a Model Based Approach," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 246–254, 2004.
- [69] D. Welfer, J. Scharcanski, and D. R. Marinho, "A coarse-to-fine strategy for automatically detecting exudates in color eye fundus images," *Comput. Med. Imaging Graph.*, vol. 34, no. 3, pp. 228–235, 2010.
- [70] L. Giancardo *et al.*, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Med. Image Anal.*, vol. 16, no. 1, pp. 216–226, 2012.
- [71] F. Amel, M. Mohammed, and B. Abdelhafid, "Improvement of the Hard Exudates Detection

- Method Used For Computer- Aided Diagnosis of Diabetic Retinopathy," *Int. J. Image, Graph. Signal Process.*, vol. 4, no. 4, pp. 19–27, 2012.
- [72] M. Usman Akram, S. Khalid, A. Tariq, S. A. Khan, and F. Azam, "Detection and classification of retinal lesions for grading of diabetic retinopathy," *Comput. Biol. Med.*, vol. 45, no. 1, pp. 161–171, 2014.
- [73] M. U. Akram, A. Tariq, S. A. Khan, and M. Y. Javed, "Automated detection of exudates and macula for grading of diabetic macular edema," *Comput. Methods Programs Biomed.*, vol. 114, no. 2, pp. 141–152, 2014.
- [74] S. Morales, V. Naranjo, U. Angulo, and M. Alcaniz, "Automatic detection of optic disc based on PCA and mathematical morphology," *IEEE Trans. Med. Imaging*, vol. 32, no. 4, pp. 786–796, 2013.
- [75] M. D. Abramoff *et al.*, "Automated analysis of retinal images for detection of referable diabetic retinopathy," *JAMA Ophthalmol.*, vol. 131, pp. 351–357, 2013.
- [76] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recognit.*, vol. 43, no. 3, pp. 706–719, 2010.
- [77] C. Szegedy *et al.*, "Going Deeper With Convolutions." pp. 1–9, 2015.
- [78] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," Oct. 2016.
- [79] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition." pp. 770–778, 2016.
- [80] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. Curran Associates Inc., pp. 1097–1105, 2012.
- [81] J. N. Mandrekar, "Receiver operating characteristic curve in diagnostic test assessment," *J. Thorac. Oncol.*, vol. 5, no. 9, pp. 1315–1316, 2010.