Automatic Analysis of Retinal Images to Aid in the Diagnosis and Grading of Diabetic Retinopathy

Roberto Romero Oraá

Universidad de Valladolid

Doctoral Program of
Information and Telecommunication Technologies

Doctoral Thesis

Automatic analysis of retinal images to aid in the diagnosis and grading of diabetic retinopathy

Thesis presented by **D. Roberto Romero Oraá**
to apply for the *Ph.D. degree*
from the *University of Valladolid*

Directed by:
Dr. María García Gadañón

2022
Valladolid, Spain

"Shakespeare said, "The eyes are the window to your soul." Little did he realize the truth of his epigram. When Romeo gazed into Juliet's eyes, he may have seen the love that was in her soul; however, an eye doctor sees much more."

Hans E. Grossniklaus

Defense

TÍTULO   Automatic analysis of retinal images to aid in the diagnosis and grading of diabetic retinopathy

AUTOR   D. Roberto Romero Oraá

DIRECTOR   Dr. D. María García Gadañón

DEPARTAMENTO Teoría de la Señal y Comunicaciones e Ingeniería Telemática

Tribunal

PRESIDENTE   Dr. D.

VOCALES   Dr. D.
              Dr. D.
              Dr. D.

SECRETARIO   Dr. D.

acuerda otorgarle la calificación de

En Valladolid, a   de   del

Universidad de **Valladolid**

Escuela Técnica Superior de Ingenieros de Telecomunicación

Dpto. de Teoría de la Señal y Comunicaciones e Ingeniería Telemática

Research Stay for the International Mention

City:	Porto (Portugal)
Faculty:	Universidade do Porto
Institution:	Institute for Systems and Computer Engineering, Technology and Science (INESC TEC)
Research group:	Biomedical Imaging Lab
Dates:	17/01/2022-17/04/2022
Duration:	90 days (3 months)
Supervisor:	Dr. Ana Maria Mendonça

Agradecimientos

Me gustaría agradecer, en primer lugar, a mi directora de Tesis, la Doctora María García Gadañón, la confianza depositada en mí para llevar a cabo esta investigación. La paciencia y el apoyo que ha demostrado me permitió seguir adelante en los momentos más difíciles.

Igualmente, quiero expresar mi gratitud al resto de miembros del Grupo de Ingeniería Biomédica (GIB) con los que he tenido la suerte de coincidir durante mi etapa predoctoral: Roberto, Jesús, Carlos, Daniel, Gonzalo, Javier G., Rebeca, Alejandro, Víctor M., Fernando, Pablo, Saúl, Verónica, Eduardo, Jorge, Adrián, Víctor R., Víctor G., Aarón, Javier O., Marcos, Celia, Sergio, Diego y Selene. Me siento orgulloso de ser parte de este equipo de profesionales a los que admiro.

También quiero agradecer a mi tía, la Doctora María Isabel López Gálvez, del Hospital Clínico Universitario de Valladolid y del Instituto Universitario de Oftalmobiología Aplicada, por animarme a conocer el mundo de la investigación y acompañarme durante toda esta etapa.

I would also like to extend my gratitude to Dr. Ana Maria Mendonça and the rest of members from Biomedical Imaging Lab at the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC) of the Univesrsity of Porto. Despite the pandemic situation, my stay in Portugal was an enriching experience both professionally and personally. I feel privileged to have worked with such outstanding researchers and to have enjoyed the charming city of Porto.

Finalmente, el mayor de mis agradecimientos es para mi familia. Mis padres, hermanos y Rachel han sido los pilares de mi vida y, por eso, cualquiera de mi éxitos también les pertenece.

Abstract

Many eye-related diseases and conditions where the blood circulation or the brain are affected manifest themselves in the retina. This is the case of diabetic retinopathy (DR), the most common complication of diabetes mellitus (DM), with a prevalence above 22% among the diabetic population. DR has become one of the leading causes of preventable blindness in the adult working population. In this condition, the hyperglycemia induced by DM is known to damage the walls of the retinal vessels causing several abnormalities in the retina. During the initial stages, these abnormalities include red lesions (RLs), such as hemorrhages (HEs) and microaneurysms (MAs), and hard exudates (EXs), yellowish-white deposits of lipoproteins and other proteins.

Visual loss cannot be recovered but can be prevented from the early stages of DR, when the treatments are effective. Therefore, early diagnosis is paramount. However, DR may be clinically asymptomatic until the advanced stage, when vision is already affected and treatment may become difficult. For this reason, diabetic patients should undergo regular eye examinations through screening programs. Due to its safety and cost-effectiveness, fundus imaging is the most established retinal imaging modality to conduct these examinations aimed at identifying the DR clinical signs. Traditionally, DR screening programs are run by trained specialists through visual inspection of the retinal images. However, this manual analysis is time consuming and expensive. With the increasing incidence of DM and the limited number of clinicians and sanitary resources, the early detection of DR becomes non-viable. Additionally, there is a certain subjectivity related to diagnosis (around 11% discrepancy between specialists). For all of these reasons, computed-aided diagnosis (CAD) systems are required to assist specialists for a fast, reliable diagnosis, allowing to reduce the workload and the associated costs.

The complexity of DR diagnosis suggests that CAD systems should be divided into several stages, leading to an extremely wide field of research. Although mul-

tiple methods can be found in the literature to carry out each of these stages, they are not exempt from limitations and there is still room for improvement and techniques to explore. In this context, we hypothesize that the application of novel, automatic algorithms for fundus image analysis could contribute to the early diagnosis of DR. Consequently, the main objective of the present Doctoral Thesis is to study, design and develop novel methods based on the automatic analysis of fundus images to aid in the screening, diagnosis, and treatment of DR.

In order to achieve the main goal, we collected 2107 fundus images to build a private database provided by the Instituto de Oftalmobiología Aplicada (IOBA) of the University of Valladolid (Valladolid, Spain) and the Hospital Clínico Universitario de Valladolid (Valladolid, Spain). Additionally, we used five retinal databases publicly available: DRIMDB, DIARETDB1, DRIVE, Messidor and Kaggle. Since all these databases have different characteristics and purposes, they have been used in different stages of the Thesis. This has allowed us to validate the proposed methods with variable fundus images obtained with different protocols, cameras, resolution and quality. The stages of fundus image processing covered in this Thesis are: retinal image quality assessment (RIQA), the location of the optic disc (OD) and the fovea, the segmentation of RLs and EXs, and the DR severity grading.

RIQA was studied with two different approaches. The first approach was based on the combination of novel, global features derived from the spatial and spectral entropy-based quality (SSEQ) and the natural images quality evaluator (NIQE) methods together with sharpness and luminosity measures based on the continuous wavelet transform (CWT) and the hue-saturation-value (HSV) color model. Results achieved 91.46% accuracy, 92.04% sensitivity, and 87.92% specificity using the private database. We developed a second approach aimed at RIQA based on deep learning due to the great potential that this field is showing in computer vision in recent years. A convolutional neural network (CNN) with InceptionResNetV2 architecture was used to detect retinal images with satisfactory quality. Data augmentation and transfer learning were also part of this approach. We achieved 95.29% accuracy with the private database and 99.48% accuracy with the DRIMDB database.

The location of the OD and the fovea was performed using a combination of saliency maps. Spatial relationships between the main anatomical structures of the retina as well as their visual features were considered. The proposed methods were evaluated over the private database and the public databases DRIVE, DIARETDB1 and Messidor. For the OD, we achieved 100% accuracy for all

databases except Messidor (99.50%). As for the fovea location, we also reached 100% accuracy for all databases except Messidor (99.67%).

The joint segmentation of RLs and EXs was accomplished by decomposing the fundus image into layers. Novel indicators, such as the reflective features of the retina and the choroidal vasculature visible in tigroid retinas, were proven useful for the classification of retinal lesions. Results were computed per pixel and per image. Using the private database, 88.34% per-image accuracy (ACC_i), 91.07% per-pixel positive predictive value (PPV_p), and 85.25% per-pixel sensitivity (SE_p) were reached for the detection of RLs. Using the public database DIARETDB1, 90.16% ACC_i, 96.26% PPV_p, and 84.79% SE_p were obtained. As for the detection of EXs, 95.41% ACC_i , 96.01% PPV_p, and 89.42% SE_p were reached with the private database. Using the public database, 91.80% ACC_i, 98.59% PPV_p, and 91.65% SE_p were obtained. An additional method was proposed for the segmentation of RLs based on superpixels. The Entropy Rate Superpixel algorithm was used to segment the potential RL candidates, which were classified using a multilayer perceptron neural network. Evaluating this method with the private database, we obtained 81.43% SE_p, 86.59 PPV_p, 84.04% SE_i, 85.00% SP_i, and 84.45% ACC_i. Using the DIARETDB1 database, we achieved 88.10% SE_p, 93.10% PPV_p, 84.00% SE_i, 88.89% SP_i, and 86.89% ACC_i.

Finally, we proposed an end-to-end deep learning framework for the automatic DR severity grading. The method was based on a novel attention mechanism which performs a separate attention of the dark and the bright structures of the retina. The framework includes data augmentation, transfer learning and fine-tuning. We used the Xception architecture as a feature extractor and the focal loss function to deal with data imbalance. The Kaggle DR detection dataset was used for development and validation. The International Clinical DR Scale was considered, which is made up of 5 DR severity levels. Classification results for all classes achieved 83.70% accuracy and a Quadratic Weighted Kappa of 0.78.

The combination of the different stages developed in this Doctoral Thesis form a complete, automatic DR screening system, contributing to aid in the early detection of DR. In this way, diabetic patients could receive better attention for their ocular health avoiding vision loss. In addition, the workload of specialists could be relieved while healthcare costs are reduced.

Acronyms

AUC	Area Under the Curve
ASF	Alternating Sequential Filtering
CAD	Computer Aided Diagnosis
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DCT	Discrete Cosine Transform
DM	Diabetes Mellitus
DR	Diabetic Retinopathy
DRIMDB	Diabetic Retinopathy IMage DataBase
DRIVE	Digital Retinal Images for Vessel Extraction
ERS	Entropy Rate Superpixel
EROD	Expected Region of the Optic Disc
EX	Exudate
FCBF	Fast Correlation-Based Filter
FCN	Fully Convolutional Network
FOV	Field Of View
GGD	Generalized Gaussian Distribution
HSV	Hue-Saturation-Value
IDF	International Diabetes Federation
JCR	Journal Citation Reports
k-NN	k-nearest neighbor
MLP	MultiLayer Perceptron
MVG	MultiVariate Gaussian
NIQE	Natural Images Quality Evaluator
NR-IQA	No-Reference Image Quality Assessment

NSS	Natural Scenes Statistics
OCT	Optical Coherence Tomography
OD	Optic Disc
PPV	Positive Predictive Value
PSD	Power Spectral Density
QWK	Quadratic Weighted Kappa
RGB	Red-Green-Blue
RIQA	Retinal Image Quality Assessment
RL	Red Lesion
ROC	Receiver Operating Characteristic
SLIC	Simple Linear Iterative Clustering
SLO	Scanning Laser Ophthalmoscopy
SSEQ	Spatial and Spectral Entropy-based Quality
SU	Symmetrical Uncertainty
SVM	Support Vector Machine
t-SNE	t-Distributed Stochastic Neighbor Embedding
VEGF	Vascular Endothelial Growth Factors
XAI	Explainable Artificial Intelligence

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The present Doctoral Thesis focuses on the automatic analysis of fundus images to aid in the diagnosis of diabetic retinopathy (DR). With this purpose, a private retinal database was collected during the course of the Thesis and several image processing methods were studied. As a consequence of this research work, four articles in journals indexed in the Journal Citation Reports (JCR) from the Web of Science™were published between March 2019 and November 2020. Furthermore, an additional article was sent for publication and is currently under review. This set of articles resulting from the current research allows this Doctoral Thesis to be presented as a compendium of publications.

The structure of this introduction is shown below. Section 1.1 justifies the thematic consistency of the papers included in the Doctoral Thesis. A brief context about biomedical engineering and medical image processing is exposed in section 1.2. Section 1.3 introduces the main concepts regarding fundus anatomy while section 1.4 describes the imaging modality with which we have worked: fundus images. In section 1.5, DR is explained covering causes, consequences, diagnosis and treatment. Finally, section 1.6 is devoted to analyze the state of the art in the context of automatic fundus image analysis.

1.1 Compendium of publications: thematic consistency

The retina is the only part in the human body where blood vessels can be directly visualized non-invasively *in vivo*. Many eye-related diseases and conditions where

the blood circulation or the brain are affected manifest themselves in the retina before showing any other visible sign (Abramoff et al., 2010; Patton et al., 2006). Among the mentioned ocular diseases, macular degeneration, glaucoma and DR are included, which are the most important causes of blindness in the developed countries. Imaging the retina allows any of these diseases to be properly detected, diagnosed and managed. For this reason, the ability to image the retina and develop techniques for retinal image analysis is of great interest (Abramoff et al., 2010).

Due to its safety and cost-effectiveness, fundus imaging is the most established retinal imaging modality (Abramoff et al., 2010). This modality is a 2-D representation of the retinal tissues obtained using reflected light. In color fundus photography, a specific category of fundus imaging, image intensities represent the amount of reflected red-green-blue (RGB) wavebands, as determined by the spectral sensitivity of the sensor. This type of image has been widely used for population-based, large scale detection of DR, the condition covered in this Thesis (Abramoff et al., 2010).

Multiple Computer Aided Diagnosis (CAD) systems have been developed to aid in the diagnosis of DR using color fundus images. The complexity of this challenge involves that these systems are usually divided into several stages. First, image quality assessment should be performed to prevent inaccurate diagnosis (Paulus et al., 2010). Second, a preprocessing step is required to prepare the image for the subsequent stages. Third, the identification of retinal landmarks such as the optic disc (OD), the fovea and the retinal vessels as reference coordinates is a prerequisite before systems can achieve more complex tasks regarding the identification of pathological entities (Patton et al., 2006). Next, the segmentation of abnormalities should be accomplished. The main visible lesions of interest characterizing the DR are the red lesions (RLs) and the exudates (EXs). Finally, severity grading is the last stage of automatic detection systems.

The present Doctoral Thesis aims to contribute to this research field by proposing novel methods for each of the stages comprising an automatic DR detection system. With this purpose, a number of image processing techniques were applied over color fundus images. Each paper of the compendium covers one of the mentioned stages and, therefore, they have a clear sequential relationship. This way, the set of publications can be seen as a whole, forming a complete CAD system to aid in DR diagnosis. The thematic consistency between the papers is illustrated in Figure 1.1.

The papers of the compendium are listed below in chronological order. The
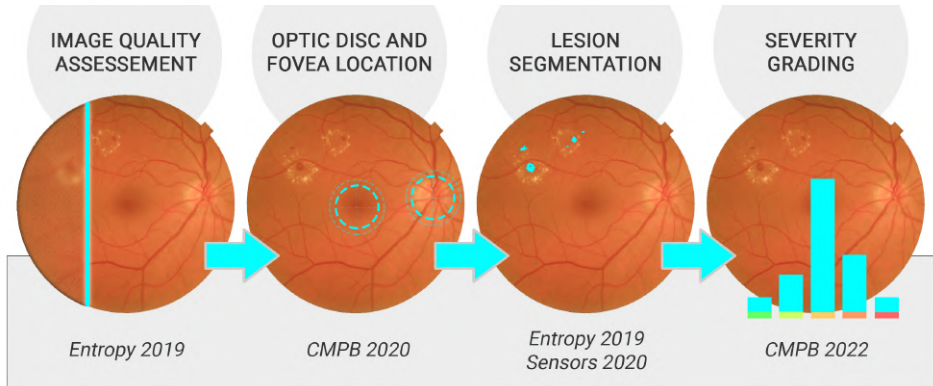
Figure 1.1: Schematic diagram of the methods proposed in the compendium of publications: image quality assessment, optic disc and fovea location, segmentation of red lesions and hard exudates and severity grading.

first paper (Jiménez-García et al., 2019) was focused on the automatic quality assessment of retinal images, which is the first stage in any CAD system (Entropy 2019 in Figure 1.1). For this purpose, we combined features derived from the spatial and spectral entropy-based quality (SSEQ) and the natural images quality evaluator (NIQE) with novel sharpness and luminosity measures based on the continuous wavelet transform (CWT) and the hue-saturation-value (HSV) color model. The second paper (Romero-Oraá et al., 2019) was aimed at developing a method to automatically detect RLs in retinal images, including hemorrhages and microaneurysms. This is part of the third stage in Figure 1.1. For this task, the Entropy Rate Superpixel (ERS) method was applied in combination with a multilayer perceptron (MLP) neural network. In this study, we also proposed a novel preprocessing stage to normalize the image appearance and enhance the retinal structures. In the third paper (Romero-Oraá et al., 2020), the main objective was to develop robust methods for the automatic location of the OD and the fovea in retinal images. This aligns with the second stage of the global CAD system (CMPB 2020 in Figure 1.1). The proposed method was based on the combination of novel saliency maps representing the spatial relationships between some structures of the retina and the visual appearance of the OD and the fovea. The fourth paper of the compendium (Romero-Oraá et al., 2020) was aimed at the automatic detection of RLs and EXs (the third stage in Figure 1.1). As the main contribution, the fundus image was decomposed into various layers, including the lesion candidates, the reflective features of the retina, and the choroidal vasculature visible in tigroid retinas. The last paper written during the Thesis, still under

review, is focused on the severity grading of DR, which completes the whole DR CAD system (CMPB 2022 in Figure 1.1). In this study, we proposed a deep learning framework based on the separate attention for dark and bright structures in the retina. Independent detection of dark lesions and bright lesions facilitates the optimization of the classification task and, consequently, can improve the results.

It is important to note that the methods developed in the studies carried out during the course of this Thesis served as the basis for the subsequent publications. This fact evidences the close relationship and thematic consistency of the papers presented in this compendium of publications. Since the present Doctoral Thesis is presented as a compendium of publications, the original papers are provided in Appendix **??**. The main information about these papers is exposed below, showing titles, authors, abstracts and journals:

Combination of Global Features for the Automatic Quality Assessment of Retinal Images (Jiménez-García et al., 2019).

Jorge Jiménez-García, Roberto Romero-Oraá, María García, María I. López, and Roberto Hornero. *Entropy*, vol. 21 (3), p. 311, March, 2019. Impact factor in 2019: 2.494, Q2 in "PHYSICS, MULTIDISCIPLINARY" (Journal Citation Reports - Web of Science, JCR-WOS).

Diabetic retinopathy (DR) is one of the most common causes of visual loss in developed countries. Computer-aided diagnosis systems aimed at detecting DR can reduce the workload of ophthalmologists in screening programs. Nevertheless, a large number of retinal images cannot be analyzed by physicians and automatic methods due to poor quality. Automatic retinal image quality assessment (RIQA) is needed before image analysis. The purpose of this study was to combine novel generic quality features to develop a RIQA method. Several features were calculated from retinal images to achieve this goal. Features derived from the spatial and spectral entropy-based quality (SSEQ) and the natural images quality evaluator (NIQE) methods were extracted. They were combined with novel sharpness and luminosity measures based on the continuous wavelet transform (CWT) and the hue saturation value (HSV) color model, respectively. A subset of non-redundant features was selected using the fast correlation-based filter (FCBF) method. Subsequently, a multilayer perceptron (MLP) neural network was used to obtain the quality of images from the selected features. Classification results achieved 91.46% accuracy, 92.04% sensitivity, and 87.92% specificity. Results suggest that the proposed RIQA method could be applied in a more

general computer-aided diagnosis system aimed at detecting a variety of retinal pathologies such as DR and age-related macular degeneration.

Entropy Rate Superpixel Classification for Automatic Red Lesion Detection in Fundus Images (Romero-Oraá et al., 2019).

Roberto Romero-Oraá, Jorge Jiménez-García, María García, María I. López, Javier Oraá-Pérez, and Roberto Hornero. *Entropy*, vol. 21 (4), p. 417, 2019. Impact factor in 2019: 2.494, Q2 in "PHYSICS, MULTIDISCIPLINARY" (JCR-WOS).

Diabetic retinopathy (DR) is the main cause of blindness in the working-age population in developed countries. Digital color fundus images can be analyzed to detect lesions for large-scale screening. Thereby, automated systems can be helpful in the diagnosis of this disease. The aim of this study was to develop a method to automatically detect red lesions (RLs) in retinal images, including hemorrhages and microaneurysms. These signs are the earliest indicators of DR. Firstly, we performed a novel preprocessing stage to normalize the inter-image and intra-image appearance and enhance the retinal structures. Secondly, the Entropy Rate Superpixel method was used to segment the potential RL candidates. Then, we reduced superpixel candidates by combining inaccurately fragmented regions within structures. Finally, we classified the superpixels using a multilayer perceptron neural network. The used database contained 564 fundus images. The DB was randomly divided into a training set and a test set. Results on the test set were measured using two different criteria. With a pixel-based criterion, we obtained a sensitivity of 81.43% and a positive predictive value of 86.59%. Using an image-based criterion, we reached 84.04% sensitivity, 85.00% specificity and 84.45% accuracy. The algorithm was also evaluated on the DiaretDB1 database. The proposed method could help specialists in the detection of RLs in diabetic patients.

A robust method for the automatic location of the optic disc and the fovea in fundus images (Romero-Oraá et al., 2020).

Roberto Romero-Oraá, María García, Javier Oraá-Pérez, María I. López, and Roberto Hornero. *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105599, 2020. Impact factor in 2020: 5.428, Q1 in "COMPUTER SCIENCE, THEORY & METHODS", "ENGINEERING, BIOMEDICAL",

"COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS" and "MEDICAL INFORMATICS" (JCR-WOS).

Background and objective. The location of the OD and the fovea is usually crucial in automatic screening systems for diabetic retinopathy. Previous methods aimed at their location often fail when these structures do not have the standard appearance. The purpose of this work is to propose novel, robust methods for the automatic detection of the OD and the fovea. *Methods.* The proposed method comprises a preprocessing stage, a method for retinal background extraction, a vasculature segmentation phase and the computation of various novel saliency maps. The main novelty of this work is the combination of the proposed saliency maps, which represent the spatial relationships between some structures of the retina and the visual appearance of the OD and fovea. Another contribution is the method to extract the retinal background, based on region-growing. *Results.* The proposed methods were evaluated over a proprietary database and three public databases: DRIVE, DiaretDB1 and Messidor. For the OD, we achieved 100% accuracy for all databases except Messidor (99.50%). As for the fovea location, we also reached 100% accuracy for all databases except Messidor (99.67%). *Conclusions.* Our results suggest that the proposed methods are robust and effective to automatically detect the OD and the fovea. This way, they can be useful in automatic screening systems for diabetic retinopathy as well as other retinal diseases.

Effective Fundus Image Decomposition for the Detection of Red Lesions and Hard Exudates to Aid in the Diagnosis of Diabetic Retinopathy (Romero-Oraá et al., 2020).

Diabetic retinopathy (DR) is characterized by the presence of red lesions (RLs), such as microaneurysms and hemorrhages, and bright lesions, such as exudates (EXs). Early DR diagnosis is paramount to prevent serious sight damage. Computer-assisted diagnostic systems are based on the detection of those lesions through the analysis of fundus images. In this paper, a novel method is proposed for the automatic detection of RLs and EXs. As the main contribution, the

fundus image was decomposed into various layers, including the lesion candidates, the reflective features of the retina, and the choroidal vasculature visible in tigroid retinas. We used a proprietary database containing 564 images, randomly divided into a training set and a test set, and the public database DiaretDB1 to verify the robustness of the algorithm. Lesion detection results were computed per pixel and per image. Using the proprietary database, 88.34% per-image accuracy (ACC_i), 91.07% per-pixel positive predictive value (PPV_p), and 85.25% per-pixel sensitivity (SE_p) were reached for the detection of RLs. Using the public database, 90.16% ACC_i, 96.26% PPV_p, and 84.79% SE_p were obtained. As for the detection of EXs, 95.41% ACC_i, 96.01% PPV_p, and 89.42% SE_p were reached with the proprietary database. Using the public database, 91.80% ACC_i, 98.59% PPV_p, and 91.65% SE_p were obtained. The proposed method could be useful to aid in the diagnosis of DR, reducing the workload of specialists and improving the attention to diabetic patients.

1.2 Context: biomedical engineering and medical image processing

Biomedical engineering is an interdisciplinary field involving the design, development, and utilization of materials, devices and techniques for clinical research and use. It can be seen as the link between engineering and medicine, where technology brings new solutions for difficult healthcare problems confronting our society (Bronzino and Peterson, 2017). Biomedical engineers apply electrical, mechanical, chemical, optical, and other engineering principles to understand, modify, or control biologic systems. Likewise, assistance in diagnosing and treating patients is also covered through the design and manufacturing of products for physiological functions monitoring (Bronzino and Peterson, 2017). Therefore, biomedical engineering covers various career areas including, among others, the computer analysis of patient-related data and clinical decision-making (i.e., medical informatics and artificial intelligence) and medical imaging (the graphic display of anatomic detail or physiologic function) (Bronzino and Peterson, 2017). Both of these areas perfectly fit the context of the present Doctoral Thesis.

Since the discovery of the X-ray radiation in 1895, the field of medical imaging has developed into a huge scientific discipline (Ritter et al., 2011). Pathologies can be observed directly rather than inferred from symptoms, allowing specialists

to make accurate diagnoses and provide appropriate treatments (Angenent et al., 2006). Therefore, the recent introduction of advanced imaging techniques in clinical practice has meant a significant improvement in patient care. In the context of medical image processing, the following fundamental classes of algorithms can be found (Bankman, 2000):

- **Enhancement**. This type of algorithms is used to reduce image noise and increase the contrast of structures of interest. In many cases, enhancement improves the quality of the image and facilitates diagnosis.

- **Segmentation**. It consists of delineating structures of interest and discriminating them from the background. In the medical field, segmentation is applied to separate the anatomical structures and tissues.

- **Quantification**. Quantification algorithms are applied to segmented structures to extract the essential diagnostic information such as shape, size, texture, angle, and motion.

- **Registration**. The purpose of the registration is to find the correspondences between homologous points of the images to be registered. In this way, if images of the same area are captured at different moments of time, the evolution of a certain disease can be followed. Additionally, a complete image can be created from other images that cover a small area of a certain body structure.

- **Visualization**. This kind of algorithm is meant to facilitate visual inspection of medical and biological data.

- **Compression, storage, and communication**. Storing medical images requires efficient algorithms for posterior retrieval. When it comes to communication purposes, compression, specialized formats, and standards need to be considered.

Among the available image modalities for retinal imaging, in this Doctoral Thesis color fundus images are processed to aid in the diagnosis of DR. The study and use of proper algorithms has allowed us to build novel methods for every stage of an automatic detection system, including an image quality assessment filter, a preprocessing step, the segmentation of important landmarks and lesions in the retina, and the classification of the DR severity degree. The next sections provide a detailed overview about the eye fundus, fundus images and DR, which is essential

to understand the characteristics of the disease and how it manifests on fundus images.

1.3   The eye fundus

The retina is the layer of the eye that is intended to be visualized through imaging. To understand the visual complications induced by DR, it is important to first introduce the main anatomical structures of a healthy retina which are visible in fundus images.

The human eyeball consists of three primary layers: (1) the outermost supporting layer of the eye, which consists of clear cornea and the opaque sclera; (2) the middle uveal layer of the eye, constituting the central vascular layer of the globe, which encompasses the iris, the ciliary body, and the choroid; and (3) the interior layer of the eye, commonly designated as the retina (Kels et al., 2015). The function of the retina is to transform light into electrical neural impulses to the brain to create visual perception. At the macroscopic level, different elements can be observed in the eye fundus: (Abramoff et al., 2010):

- **Optic disc or papilla**. It is the entry point of the optic nerve into the eyeball. Due to the absence of photoreceptors, it is also known as the blind spot. Under ophthalmological examination, the papilla is seen as a circular or slightly oval disk with an approximate diameter of 1.5 mm and yellow pigment. The OD is also the entry point for the major blood vessels that supply the retina.

- **Blood vessels**. Arteries are responsible for providing oxygen and nutrients to the retina. The central retinal artery enters the eye through the optic nerve and separates into two branches, which in turn diverge to form an extensive network of capillaries. The central retinal vein exits the retina through the OD. The arteries have a light red color, while the veins appear with a more intense red color. Many eye diseases affect the blood vessels, blocking them or making them more fragile.

- **Macula**. This region has a darker appearance than the rest of the retinal tissue due to its high pigmentation and is characterized by the absence of vasculature. It is located in the center of the retina and has an approximate diameter of 5 mm. It is in charge of both the central vision and the vision in detail and in movement.

- **Fovea**. It is a small slit located in the center of the macula with an approximate diameter of 1 mm. It is the area of greatest visual acuity due to the presence of a large number of photoreceptors.

Figure 1.2 shows the main anatomical structures visible on a retina.

1.4   Fundus images

The study of the eye fundus has become a routine and essential examination for the diagnosis and treatment of numerous ocular and systemic pathological processes, producing a rapid development of retinal imaging (Abramoff et al., 2010). Given the great relevance of regular eye examinations, there are several factors that motivate automatic retinal screening. First, clinicians are scarce and costly
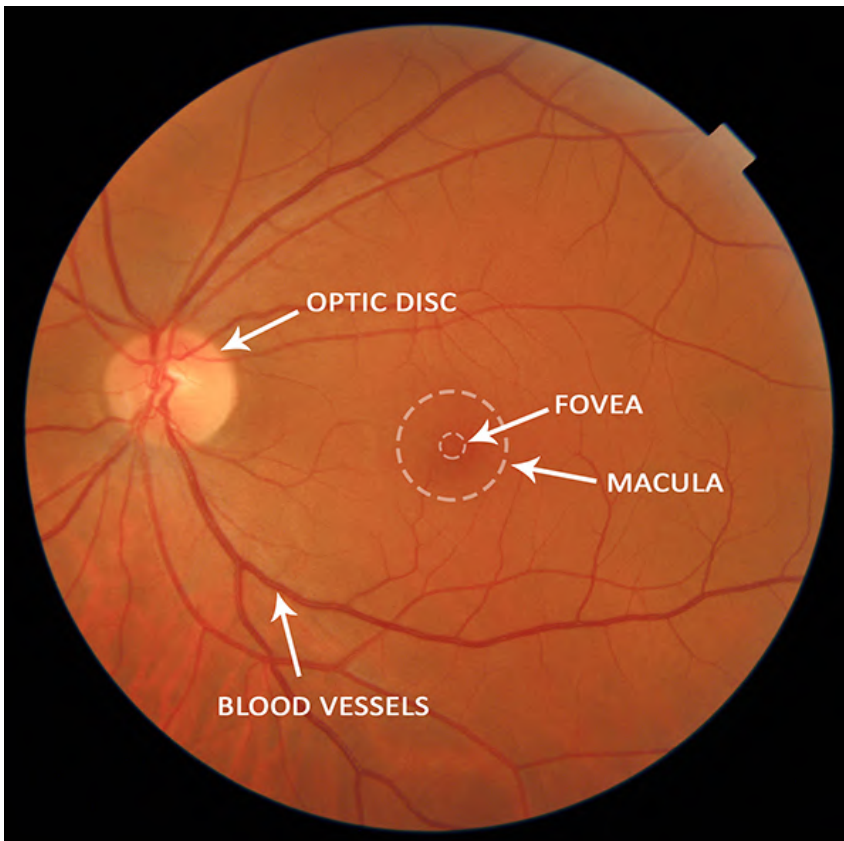
Figure 1.2: Main anatomical structures on a fundus image at a macroscopic level: optic disc, blood vessels, macula and fovea.

experts. The expense associated with the manual inspection of the eye fundus is very high. Therefore, the time devoted to each patient should be optimized. Second, the growing number of patients needing eye examinations and the development of new imaging technology rapidly increases the amount of retinal images to be reviewed. This situation is already exceeding the limit of sanitary resources and capacity of clinicians. Third, there is a certain subjectivity related to the decisions of clinicians, resulting in inter- and intra-observer variability. Therefore, there is an additional need for reliable automatic analysis to increase the accuracy and objectivity of the diagnosis (Abramoff et al., 2010).

Currently, retinal screening is carried out using two different retinal imaging modalities: Optical Coherence Tomography (OCT) imaging and fundus imaging. OCT scans are hard to obtain since OCT cameras are not easily accessible. The high price of these cameras has prevented their spread among primary care screening centers (Sharafeldin et al., 2018). Conversely, fundus cameras can easily be found in many primary care centers. They are easy-to-use and relatively low-cost, which explains the popularity of this imaging modality (Abramoff et al., 2010). Therefore, the use of fundus images is more suitable for screening purposes, since it can take advantage of the currently available sanitary resources.

In the context of fundus imaging for retinal screening, the most obvious application is the early detection of retinal diseases. Important screening programs for the detection of glaucoma and age-related macular degeneration already exist (Abramoff et al., 2010). However, main screening application focuses on early detection of DR (Abramoff et al., 2010). In addition to the screening of retinal diseases, fundus photography allows the detection of cardiovascular risk factors, such as strokes, hypertension or myocardial infarcts.

Fundus images are defined as a 2-D representation of the 3-D retinal semi-transparent tissues projected onto the imaging plane using reflected light. The following modalities/techniques belong to this category (Abramoff et al., 2010):

- **Fundus photography (including so-called red-free photography).** Image intensities represent the amount of reflected light of a specific waveband.

- **Color fundus photography**. Image intensities represent the amount of reflected red, green, and blue wavebands, as determined by the spectral sensitivity of the sensor.

- **Stereo fundus photography**. Intensities represent the amount of reflected light from two or more different view angles for depth resolution.

- **Hyperspectral imaging**. Intensities represent the amount of reflected light of multiple specific wavelength bands.

- **Scanning laser ophthalmoscopy (SLO)**. Intensities represent the amount of reflected single wavelength laser light obtained in a time sequence.

- **Adaptive optics SLO**. Intensities represent the amount of reflected laser light optically corrected by modeling the aberrations in its wavefront.

- **Fluorescein angiography and indocyanine angiography**. Image intensities represent the amounts of emitted photons from the fluorescein or indocyanine green fluorophore that need to be injected into the subject.

In this Doctoral Thesis, color fundus images were used. These images are captured using a non-invasive and painless procedure. Moreover, they are the most accessible retinal imaging modality worldwide and the most widespread imaging modality for DR diagnosis, the condition covered in this Thesis (Mookiah et al., 2013). Additionally, color fundus images stand out for being easy to obtain and multiple public databases can be found in the literature. In order to perform some of our experiments, we built a completely annotated private database composed of fundus images obtained from the clinical environment. The details about this database are exposed in section 3.1. To complete our experiments, various fundus image datasets publicly available were considered: DRIMDB (Sevik et al., 2014), DIARETDB1 (Kauppi et al., 2007), DRIVE (Staal et al., 2004), Messidor (Decencière et al., 2014) and Kaggle (Kaggle, 2015). These databases are described in detail in section 3.2.

1.5   Diabetic retinopathy

Diabetes mellitus (DM) is a serious, long-term condition with a major impact worldwide, being one of the 10 most frequent causes of death in the world (Saeedi et al., 2019). It is a metabolic disorder characterized by hyperglycemia, the presence of elevated blood glucose, resulting from defects in insulin secretion, insulin action, or both. This disease can be divided into 2 types. Type 1 DM is caused by an absolute deficiency of insulin secretion. Type 2 DM, which is much more prevalent, is caused by the combination of resistance to insulin action and an inadequate compensatory insulin secretory response (American Diabetes Association, 2014). The causes of DM are not fully understood, but genetic background, obesity, and sedentary lifestyle all confer increased risk of developing diabetes.

Treatment is primarily through diet changes and the administration of insulin and/or anti-hyperglycemic drugs (Abramoff et al., 2010). The International Diabetes Federation (IDF) estimates the global population with DM to be 463 million and projected it to be 700 million by 2045 (Saeedi et al., 2019). Furthermore, the rapid aging global population, the increasing lifespan of people with DM, and the lifestyle changes are leading to an increased risk for DM (Teo et al., 2021).

The progress of DM is associated with long-term damage, dysfunction, and failure of different organs, especially the eyes, kidneys, nerves, heart, and blood vessels (American Diabetes Association, 2014). The most common of these complications is the DR, one of the leading causes of preventable blindness in the adult working population (Teo et al., 2021). The prevalence of DR is estimated to be 22.27% globally within the DM population. The global number of adults with DR in 2020 was estimated to be 103.12 million. Meanwhile, the prevalence of vision-threatening DR was estimated to be 6.17% with 28.54 million of adults. In addition, the global prevalence of clinically significant macular edema was estimated to be 4.07% within the DM population, with a global population of 18.83 million (Teo et al., 2021).

1.5.1 Pathogenesis

Hyperglycemia, the result of DM, is known to damage the walls of the retinal vessels (Abramoff et al., 2010). This fact causes several abnormalities in the retina, producing the following clinical signs visible in the eye fundus (Mookiah et al., 2013):

- **Microaneurysms** are the earliest visible lesions of DR. They look like round red spots less than 125 μm in size with sharp margins. Microaneurysms are caused by abnormal permeability and/or non-perfusion of retinal blood vessels.

- **Hard exudates** are yellowish-white deposits of lipoproteins and other proteins leaking through abnormal retinal vessels. They have sharp edges and are often arranged in clumps.

- **Soft exudates or cotton-wool spots** are shown as fluffy white lesions and occur due to occlusion of arteriole as a consequence of the ischemia caused by reduced blood flow.

- **Hemorrhages** are red spots with irregular margin and/or uneven density. They occur due to the leakage of weak capillaries.

- **Neovascularization** is the abnormal growth of new blood vessels on the inner surface of the retina. The problem is that these new vessels tend to bleed into vitreous cavity, obscuring the vision.

- **Macular edema** is a relevant sign, characterized by the swelling of the retina, which directly affects the central vision. The permeability of abnormal retinal capillaries causes the leakage of fluid and solutes around the macula.

Figure 1.3 shows the aspect of some of these clinical lesions on a fundus image.

Depending on the progress of the disease, DR can be divided into 2 stages: nonproliferative and proliferative. The earliest visible signs in nonproliferative DR, the initial stage, are microaneurysms and retinal hemorrhages, collectively known
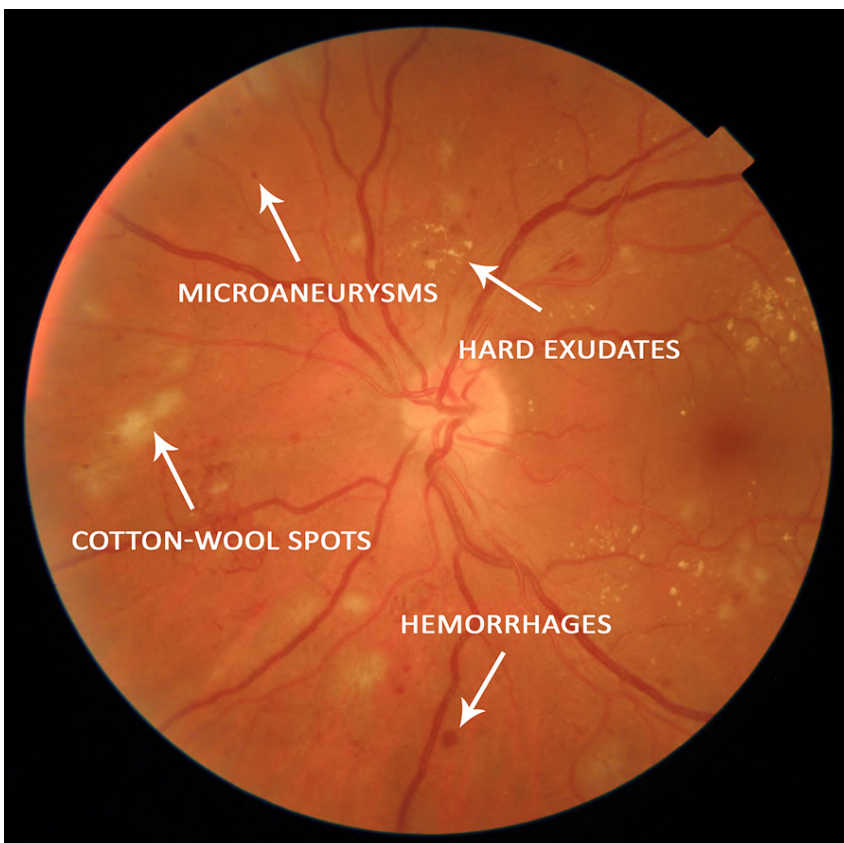
Figure 1.3: Example of a fundus image showing some of the main lesions associated with DR: microaneurysms, hard exudates, cotton-wool spots and hemorrhages.

as RLs. Progressive capillary nonperfusion is accompanied by the development of cotton-wool spots, venous beading, and intraretinal microvascular abnormalities. Proliferative DR, the most advanced stage, occurs with further retinal ischemia and is characterized by the growth of new blood vessels (the above-mentioned neovascularization) and connective tissue on the surface of the retina or the optic nerve head and into the vitreous. The proliferative stage is a consequence of retinal damage during the preceding stage (Engerman, 1989; Mohamed et al., 2007).

1.5.2 Diagnosis

Early stages of DR may be clinically asymptomatic and the disease may not be detected until the advanced stage, when vision is already affected and treatment may become difficult (Mookiah et al., 2013). There is enough scientific evidence that blindness and visual loss in DR patients can be prevented through regular screening and early diagnosis (Abramoff et al., 2010). For this reason, DR patients should undergo periodical examinations where fundus images are analyzed to identify DR clinical signs.

Initially, DR screening programs are run by trained specialists through visual inspection of the retinal images. However, this manual analysis is time consuming and expensive. With the increasing incidence of DM and the limited number of clinicians and sanitary resources, the early detection of DR becomes non-viable (Stolte and Fang, 2020). Additionally, the level of agreement between specialists tends to be moderate (around 11% discrepancy), which implies a certain subjectivity related to diagnosis (Krause et al., 2018). This level of concordance varies depending on the DR severity degree. For all of these reasons, CAD systems are required to assist specialists for a fast, reliable diagnosis, allowing to reduce the workload and the associated costs (Abramoff et al., 2010; Stolte and Fang, 2020).

Clinical decision making requires a standard set of definitions to describe the severity of DR. Practical clinical standard terminology is also useful for communication among medical doctors (Wilkinson et al., 2003). Different treatment recommendations might be followed in different regions around the world. However, an uniform criteria is necessary to classify and treat DR. In this sense, various severity scales have been proposed. The scale introduced in the Early Treatment Diabetic Retinopathy Study (ETDRS) was based on the modified Airlie House classification of DR and is recognized as the "gold standard" for DR severity grading in clinical trials (Diabetic Retinopathy Study Research Group, 1981). Nevertheless, its use in everyday clinical practice has not proven to be easy or practical. The ETDRS

scale has more levels than might be necessary for clinical care and is too complex (Wilkinson et al., 2003). Consequently, several simplified severity scales have been developed in different countries until the arrival of a single standardized practical clinical classification system: the International Clinical DR Scale (Wilkinson et al., 2003). This classification system is a consensus based on the ETDRS and the Wisconsin Epidemiologic Study of DR publications. It should be mentioned that the latest studies aimed at automatic screening of DR use the International Clinical DR Scale. It consists of five stages with increasing risks of DR (Wilkinson et al., 2003):

- **No apparent retinopathy**. No abnormalities.

- **Mild nonproliferative diabetic retinopathy**. Microaneurysms only.

- **Moderate nonproliferative diabetic retinopathy**. More than just microaneurysms but less than severe nonproliferative diabetic retinopathy.

- **Severe nonproliferative diabetic retinopathy**. Any of the following: more than 20 intraretinal hemorrhages in each of 4 quadrants; definite venous beading in 2 quadrants; prominent intraretinal microvascular abnormalities in 1 quadrant and no signs of proliferative retinopathy.

- **Proliferative diabetic retinopathy**. One or more of the following: neovascularization; vitreous/preretinal hemorrhage.

1.5.3   Treatment

The management of DM primarily involves reducing the blood sugar through diet, lifestyle changes and anti-diabetic drugs (American Diabetes Association Professional Practice Committee, 2022). Patients with mild DR do not require any particular treatment other than optimal control of DM and the associated risk factors like hypertension, anemia, and renal failure (Mookiah et al., 2013). Intensive glycemic and blood pressure control, known as primary interventions, can definitely reduce the incidence of DR (Mohamed et al., 2007).

In the advanced stage of DR, secondary interventions, such as laser photocoagulation, may be required to prevent blindness from vitreous hemorrhages and tractional retinal detachment. In most of the treated eyes, stable vision can be maintained once the disease becomes inactive but the patients must be regularly re-examined (Mohamed et al., 2007). Vitrectomy, a type of surgery where the vitreous is replaced with another solution, can prevent blindness in patients with

advanced stages of DR. Both laser photocoagulation and vitrectomy carry a risk of additional vision loss. A specific treatment for macular edema is the anti-Vascular Endothelial Growth Factors (anti-VEGF). VEGF angiogenic activity is inhibited using VEGF protein and the receptor activation is prevented (Mookiah et al., 2013).

Considering the available treatments for DR management, it is clear that early detection of clinical signs associated with DR is essential for effective treatment of the disease and to prevent significant vision loss in diabetic patients.

1.6 State of the art: automatic analysis of retinal images in DR diagnosis

The main contributions of this Doctoral Thesis are related with the automatic analysis of fundus images to aid in the early diagnosis of DR. As previously explained, CAD systems divide the problem into several independent stages: image quality assessment, retinal landmark detection, lesion segmentation, and severity grading. We can find in the literature that previous research has been accomplished for each of those stages. Therefore, the next subsections reflect a comprehensive state of the art revision of the mentioned stages.

1.6.1 Image quality assessment

The quality of the input image is important for an observer or an automated system to be able to interpret the pathological signs. Although fundus cameras have greatly improved in performance and ease of use, there are several factors affecting the image quality, including the experience of the operator, the camera model or the possible ocular pathology of the patient (Paulus et al., 2010). In these situations, the retinal structures and possible lesions related to DR are not clearly visible. For this reason, not all images captured in a clinical setting are of sufficient quality to be analyzed, which could lead to misdiagnosis (Paulus et al., 2010). The most frequent causes of ungradable fundus images are inadequate focus, blurring, or insufficient illumination (Besenczi et al., 2016). Some large-scale studies reported an ungradable image rate of 10–20% due to insufficient quality (Besenczi et al., 2016). Given this context, it seems clear that objective and automatic quality assessment of fundus images should be performed before they are analyzed by automatic systems or human graders. Thus, an image quality assessment stage should be included as the first step in any CAD system (Paulus

et al., 2010). Nonetheless, this stage is also useful as a stand-alone algorithm. The automated and immediate assessment of the quality of fundus images would allow to repeat the capture of the image of insufficient quality in the same consultation avoiding the need to repeat the appointment (Zago et al., 2018).

Numerous methods have been proposed for retinal image quality assessment (RIQA). They can be categorized into structural methods and generic methods (Fleming et al., 2012; Paulus et al., 2010; Pires Dias et al., 2014; Welikala et al., 2016). Structural methods are based on the segmentation of some structures of the retina. Some authors computed the area of the vasculature (Usher et al., 2003) or the density of small vessels around the macula (Fleming et al., 2006). Niemeijer et al. (2006) divided the image into clusters to separate the background from the vessels and the OD. Quality assessment was then performed using a support vector machine (SVM) classifier. In another study, the areas of clearly visible vessels were analyzed by means of a vessel enhancement operator (Fleming et al., 2012). Local vessel density was also studied in different regions of the image (Giancardo et al., 2010). A more recent approach was based on the area, fragmentation, and complexity of the vasculature to train a SVM classifier (Welikala et al., 2016). Unlike structural approaches, generic methods avoid the structure segmentation stage (Pires Dias et al., 2014). Some previous generic methods focused on histogram analysis and edge detection (Lalonde et al., 2001a; Lee and Wang, 1999). In other studies, the width of the image edges Lin et al. (2017a) or the local sharpness and illumination were considered (Bartling et al., 2009). Entropy, texture, and the analysis of blurred areas have also been investigated (Davis et al., 2009; Fasih et al., 2014; Remeseiro et al., 2017; Wang et al., 2016). Anisotropy-related focus measures based on the Rényi entropy and the discrete cosine transform (DCT) have also been studied (Marrugo et al., 2012; Marrugo Hernández et al., 2011). Pires Dias et al. (2014) combined color, focus, contrast, and illumination features and used a MLP neural network. More recent studies focused on the analysis of illumination, homogeneity, saturation, and sharpness using the wavelet transform (Abdel-Hamid et al., 2016, 2017). Wavelet transform was also applied to estimate focus in (Veiga et al., 2014), where Chevyshev moments, and a median filter-based statistical measure were additionally studied. Hybrid methods have also emerged in the context of RIQA as a combination of the generic and structural methods. Some authors complemented a previous clustering approach (Niemeijer et al., 2006) using texture features to improve the results (Paulus et al., 2010). Sevik et al. (2014) segmented the vasculature, the OD, and the macula, and extracted features related to form, texture, and intensity.

The OD was also detected in (Shao et al., 2017) to complement an illumination measure and a general-purpose quality evaluator. Latest studies aimed at RIQA are based on deep learning approaches. In some of them, the pretrained AlexNet architecture (Raj et al., 2019; Saha et al., 2018) was considered while, in others, VGG19, ResNet18, ResNet50 and DenseNet121 were studied (Pan and Yang, 2010; Tennakoon et al., 2016). (Zago et al., 2018) used the convolutional neural network (CNN) Inception v3 and compared training from scratch with the use of transfer learning.

Structural and hybrid methods are limited by segmentation algorithms, which are usually inaccurate and error prone (Paulus et al., 2010; Pires Dias et al., 2014). Conversely, generic methods achieve great performance while being computationally simpler (Bartling et al., 2009; Pires Dias et al., 2014). Recent generic methods combined wavelets and alternative color models (Abdel-Hamid et al., 2016, 2017). However, they are based on simple quality features which frequently differ from the opinion of the human graders (Wang et al., 2016). Therefore, more accurate image quality measurements need to be explored. Previous research applied general-purpose quality metrics to develop generic RIQA methods (Fasih et al., 2014; Wang et al., 2016). In this context, general-purpose no-reference image quality assessment (NR-IQA) methods inspired by the Natural Scenes Statistics (NSS) approach have shown promising results for image quality assessment (Liu et al., 2014; Mittal et al., 2012, 2013; Moorthy and Bovik, 2010). Among the most popular NSS-based NR-IQA methods, NIQE (Mittal et al., 2013) and SSEQ (Liu et al., 2014) have gained relevance in the last years. NIQE and SSEQ quality-aware features are being used to build robust NR-IQA methods in a variety of applications (Li et al., 2017). However, to the best of our knowledge, only one NSS-based NR-IQA method has been successfully applied in the context of fundus image analysis (Shao et al., 2017).

1.6.2 Optic disc and fovea location

The OD and the fovea are the most important anatomical landmarks in fundus images, together with the vascular network. The detection of these structures is usually crucial in CADs (Niemeijer et al., 2009). The OD has the appearance of a bright yellow disc in fundus images. Consequently, it could be easily confused with EXs (Niemeijer et al., 2009). Additionally, the vasculature emerges from the OD and its center is a reference point for the detection of other retinal structures, like the fovea (Al-Bander et al., 2018). For these reasons, the detection of the OD

is highly recommended before further processing (Harangi and Hajdu, 2015). The macula appears darker than the rest of the retinal background (Welfer et al., 2011). The fovea is the central area of the macula and responsible for sharp central vision. For this reason, lesions affecting this region are especially relevant. Therefore, the location of the fovea also plays an important role in CADS (Niemeijer et al., 2009).

The detection of the OD has received considerable attention in the literature. Exclusively locating the OD center allows estimating the OD by means of a circumference, which is enough to differentiate OD and lesions (García et al., 2009; Lalonde et al., 2001b). In some studies, the pixel intensity variation was considered (Sinthanayothin et al., 1999). Other authors applied template matching (Lalonde et al., 2001b) or morphological filtering and watershed transformation (Walter et al., 2002). Several previous methods rely on the vasculature to locate the OD (Chalakkal et al., 2018; Foracchia et al., 2004; Hoover and Goldbaum, 2003; Niemeijer et al., 2009). In the work developed by Niemeijer et al. (2009), a k-nearest neighbor (k-NN) classifier was additionally included. Later, Lu (2011) used a circular transformation while Qureshi et al. (2012) proposed a combination of techniques comprising pyramidal decomposition, edge detection and entropy filter. A spatial weighted graph was defined in (Harangi and Hajdu, 2015) and an illumination correction method was studied by Hsiao et al. (2012). Other studies were based on swarm techniques, including the ant colony technique (Pereira et al., 2013), firefly algorithms (Rahebi and Hardalaç, 2016) and different swarm intelligence algorithms (Abed et al., 2016; Zhang and Lim, 2020). Finally, deep learning techniques have also been applied in some recent studies (Al-Bander et al., 2018; Bhatkalkar et al., 2021; Hasan et al., 2021).

Various methods can also be found in the literature for the automatic location of the fovea. Some authors studied the anatomical position of the fovea relative to the OD and the vasculature (Aquino, 2014; Gegundez-Arias et al., 2013; Welfer et al., 2011). Additionally, some of them applied mathematical morphology (Aquino, 2014; Welfer et al., 2011). Li and Chutatape (2004) modeled the main vessel arcades using a parabolic curve and focused on the low intensity pixels. Feature extraction techniques were applied in some works to find the center of the fovea (Gegundez-Arias et al., 2013; Niemeijer et al., 2007b). Other authors improved the performance of individual algorithms by applying special geometric rules to combine different methods (Qureshi et al., 2012). The method proposed in (Giachetti et al., 2013) was based on the fast radial symmetry transform and the vascular density. The use of machine learning has also been considered over hand-crafted features (Harangi and Hajdu, 2015; Niemeijer et al., 2009; Qureshi

et al., 2012). In recent studies, deep CNNs and heuristic based clustering have achieved a high accuracy (Al-Bander et al., 2018; Bhatkalkar et al., 2021; Hasan et al., 2021).

1.6.3   Detection of red lesions and exudates

The most important clinical signs characteristic of the DR are the RLs, such as microaneurysms and hemorrhages, and the bright lesions, such as EXs. MAs are the earliest visible sign of DR and appear as reddish, small, and circular dots. HEs generally look like bigger red spots with irregular shapes. Finally, EXs appear as yellowish, bright patches of varied shapes and sizes with sharp edges (Kar and Maity, 2018; Mookiah et al., 2013).

Several approaches have been proposed in the last years aimed at the detection of DR-related lesions in fundus images (Xiao et al., 2019). Some methods were aimed at detecting MAs alone and can be divided into four groups (Mookiah et al., 2013): mathematical morphology-based (Abràmoff et al., 2013), region growing-based (Fleming et al., 2006), wavelet-based (Quellec et al., 2008), and hybrid approaches (Lazar and Hajdu, 2013; Wu et al., 2017). Other methods were aimed at exclusively detecting HEs and can be divided into two categories (Mookiah et al., 2013): mathematical morphology (Fleming et al., 2008; Hatanaka et al., 2008) and pixel classification (Zhang and Chutatape, 2005). However, both MAs and HEs usually look similar and tend to be detected together, which is enough to determine the presence of DR (Abramoff et al., 2010). Therefore, algorithms for the joint detection of all RLs are common in the literature. In this context, Seoud et al. (2016) proposed a technique based on dynamic shape features that represent the evolution of the shape during image flooding. Other authors divided the image into superpixels to detect RLs (Romero-Oraá et al., 2019; Zhou et al., 2017a). In the work conducted by Srivastava et al. (2017), several filters were applied on patches of different sizes and multiple kernel learning was used to combine the results. Finally, deep learning architectures have also received attention for the detection of RLs (Abràmoff et al., 2016; Huang et al., 2022; Lam et al., 2018; Orlando et al., 2018).

Numerous approaches have also been proposed to detect EXs. Generally, they can be divided into three groups: clustering-based (Hsu et al., 2001; Osareh et al., 2009; Sopharak et al., 2009); mathematical morphology, thresholding, and region growing-based (Imani and Pourreza, 2016; Sánchez et al., 2009; Walter et al., 2002); and pixel classification-based (Niemeijer et al., 2007a; Sánchez et al., 2008).

Among the most relevant studies, an unsupervised approach based on the ant colony optimization algorithm has been proposed (Pereira et al., 2015). Supervised learning approaches, such as SVM and neural networks, have also been explored (Theera-Umpon et al., 2019). Superpixel segmentation has been implemented by Zhou et al. (2017b). Other authors have employed deep learning techniques for the segmentation of EXs (Adem, 2018; Guo et al., 2019; Huang et al., 2022; Khojasteh et al., 2019; Prentašić and Lončarić, 2016). In this context, Guo et al. (2019) introduced a novel loss function which focuses on the hard-to-classify pixels to deal with class-unbalance and loss-unbalance.

1.6.4 Severity grading

In practice, DR severity grading is performed as a global estimation based on the type and extension of the overall retinal lesions (Araújo et al., 2020). In the end, the decision about the treatment is based on this graduation and, consequently, DR grading is the ultimate goal for CADSs (Araújo et al., 2020).

In the literature, several approaches to automatically detect the DR severity degree can be found (Abramoff et al., 2010; Stolte and Fang, 2020). Traditional methods were based on manually designed features (Li et al., 2021b). In recent years, deep learning methods have achieved better performance and have become the preferred solution for many automatic classification tasks, including DR grading. Unlike traditional methods, deep learning models allow for automatically optimizing the features in an end-to-end manner (Li et al., 2021b). For these reasons, the number of publications that reference DR and deep learning has increased dramatically in the last years (Li et al., 2021b; Stolte and Fang, 2020). However, DR grading is a complex challenge and, consequently, most DR research has generally focused on the binary classification of referable DR. For instance, Gulshan et al. (2016) developed a deep learning algorithm for automated detection of referable DR based on an ensemble of 10 CNNs with Inception-v3 architecture (Szegedy et al., 2016). The method was trained using 128,175 images and was validated over 2 datasets. Similarly, other authors proposed an algorithm composed of 2 CNNs to detect referable DR while detecting the pathological pixels (Quellec et al., 2017). In the study conducted by Abràmoff et al. (2016), a hybrid system was proposed: several deep learning models were used to extract DR related features to be integrated into a classic system. Other authors applied a pretrained fully convolutional network (FCN) to build a weakly-supervised model (Costa et al., 2019). They were able to produce patch-level predictions for DR lesions

and to detect DR while using image labels only. In the literature, some studies aimed at graduating the DR in several stages can be found. These studies use the International Clinical DR Scale described in section 1.5.2. González-Gonzalo et al. (2018) proposed an iterative approach which obtains a refined localization of abnormalities using a VGG16 network architecture. In (Krause et al., 2018), 1,665,151 images were used to train the same CNN architecture exposed in (Gulshan et al., 2016). The method included a cascade of thresholds on the output probabilities to obtain the final classification. In another study, de la Torre et al. (2018) introduced a weighted kappa loss function for multi-class classification proving effectiveness in DR grading. Later, the same authors proposed an interpretable classifier based on a novel pixel-wise score propagation model (de la Torre et al., 2020). Recently, Araújo et al. (2020) published a method that, together with the DR grading, provides a medically interpretable explanation and an uncertainty estimation. For this task, a novel Gaussian-sampling approach based on a Multiple Instance Learning framework was proposed. All mentioned approaches were built upon a certain CNN architecture that extracted a set of features over the whole image. However, some parts of the image are more relevant than others when determining the DR severity. For this reason, attention mechanisms enter the picture allowing learning to focus on the areas of the image most useful for classification and ignoring the less relevant areas (Chen et al., 2016; Vaswani et al., 2017). Based on this mechanism, Wang et al. (2017) proposed a CNN algorithm that mimicked the zoom-in process of a clinician to examine the retinal images. This method could generate attention maps for suspicious regions and predict DR severity degree based on both the whole image and its high-resolution suspicious patches. However, the approach required the images of both eyes, which are not always available.

After this introduction to the Doctoral Thesis, the organization of the rest of this manuscript is described below. Chapter 2 exposes the hypothesis and the objectives covered in this research work. Chapter 3 describes the databases that were used for the development of the proposed methods. The methodological contributions are shown in chapter 4, separately explaining each of the algorithms proposed for the different stages of the automatic diagnostic system to aid in the diagnosis of DR. Chapter 5 exhibits the results obtained in this research work, which are thoroughly discussed in chapter 6. Finally, chapter 7 describes the main conclusions extracted from the Doctoral Thesis. The content of the manuscript

is completed with the following additional sections: appendix **??** gathers the papers included in the compendium of publications, appendix B shows the scientific achievements reached during the course of the Doctoral Thesis, and appendix C exposes an abstract of the Thesis in Spanish.

Chapter 2

Hypotheses and objectives

The proposal of the present Doctoral Thesis is aimed at contributing to the early diagnosis of DR. To this end, retinal fundus images need to be analyzed by means of advanced image processing techniques.

Section 2.1 contains the main hypothesis that has motivated the development of this Doctoral Thesis, and the particular hypotheses that have guided the individual studies carried out. The main objective as well as the specific goals of this research work are exposed in section 2.2.

2.1   Hypotheses

The rising incidence of DM is causing its associated complications to require a growing attention. Among these complications, DR is already a primary cause of blindness and vision loss globally (Grzybowski et al., 2020). The high prevalence, social impact and sanitary costs associated with this condition corroborates the great relevance of this issue today.

There is enough scientific evidence that most visual loss associated with DR can be prevented through early diagnosis and adequate treatment (Abramoff et al., 2010; Araújo et al., 2020; Grzybowski et al., 2020). However, this condition is asymptomatic in its early stages and tends to remain undetected until an advanced vision-threatening stage (Islam et al., 2020). For this reason, diabetic patients should undergo periodic eye exams via screening programs (Abramoff et al., 2010). In these programs, DR detection is performed by trained specialists through visual inspection of fundus images (Araújo et al., 2020). However, the vast number of diabetic population and the limited resources in personnel and technology make

the number of images to analyze unwieldy, producing a work overload for the clinicians. Consequently, the manual analysis of fundus images is non-viable in practice (Stolte and Fang, 2020). Additionally, there is a well-known discrepancy in DR diagnosis between specialists due to the subjective character of their decisions (Krause et al., 2018). For these reasons, there is a great need for CAD systems aimed at the automatic analysis of fundus images to assist ophthalmologists (Islam et al., 2020). This way, DR diagnosis can be improved in terms of accuracy, speed and objectivity, while reducing the workload of specialists and the health costs (Abramoff et al., 2010).

In this context, we hypothesize that *the development of novel, automatic methods for fundus image analysis could contribute to the early diagnosis of DR*. Furthermore, the combination of such methods could provide a complete CAD system that could be used as a diagnostic tool in clinical practice. As explained before, CAD systems divide the diagnostic process into several stages. In this Doctoral Thesis, we conducted several studies to cover each of these stages. The particular hypotheses that motivated these studies are set out below.

In the first of the studies of the compendium of publications of this Thesis (Jiménez-García et al., 2019), a new RIQA method was proposed. It is important to ensure that input images are of sufficient quality for analysis. Otherwise, the CAD system could lead to misdiagnoses. Given the state of the art for image quality assessment exposed in section 1.6.1, where individual generic features were separately studied, we hypothesized that *combining NSS-based NR-IQA methods with generic features based on sharpness and luminosity can be useful to assess the quality of fundus images.*

An additional RIQA method was developed during the course of the present Thesis, which was presented in an international conference (Romero Oraá et al., 2020). In recent years, deep learning architectures, such as CNNs, have gained importance in image processing and, in particular, in quality assessment of fundus images. In this context, we hypothesized that *the use of transfer learning and fine-tuning with the InceptionResNetV2 architecture could lead to robust results and outperform conventional methods in RIQA.*

The preprocessing stage is a paramount process in fundus images to normalize the inter-image and intra-image appearance as well as to enhance the retinal structures. With the preprocessing method proposed in (Romero-Oraá et al., 2019), we hypothesized that *the sequential application of certain operations would allow us to highlight the lesions (in particular, the RLs), avoid border effects in subsequent stages and achieve inter-image normalization.* This set of operations is: bright

border artifact removal, background extension, illumination and color equalization, denoising and contrast enhancement.

The third paper of the compendium of publications (Romero-Oraá et al., 2020) covers the location of the OD and the fovea, the most important anatomical landmarks of the retina together with the vascular network. These structures serve as reference in the retinal image for further processing. Multiple previous methods were proposed to accomplish this task automatically, as stated in section 1.6.2. However, most of them focus on specific criteria that are insufficient to represent the OD and the fovea in all images, such as the highest variation in intensity of adjacent pixels (Lalonde et al., 2001b; Sinthanayothin et al., 1999). Some studies combine several indicators, like the entropy of the image or the convergence of the main vessels, but the way that those indicators are used is very restrictive. Hence, these algorithms fail when any of these indicators deviates from the standard pattern (Giachetti et al., 2013; Qureshi et al., 2012). We hypothesized that *the combination of novel complementary indicators would allow us to accurately represent the areas where both the OD and the fovea are located. This way, the computation of certain saliency maps would be enough to detect the center locations even for the specific cases where the OD and the fovea do not show a standard appearance.*

The detection of EXs and RLs was jointly studied in (Romero-Oraá et al., 2020). These two types of lesions are characteristic signs of the early stages of DR. This is why its detection is relevant for DR screening purposes. Even though numerous methods can be found in the literature (see section 1.6.3), none of them have individually considered other structures of the retina beyond OD, the fovea, and the vasculature. We hypothesized that *the reflective features of the retina and the choroidal vasculature visible in tigroid retinas could also be useful for the detection of retinal lesions. This way, fundus images could be decomposed into several layers to facilitate the segmentation of EXs and RLs.*

The detection of RLs was additionally covered in (Romero-Oraá et al., 2019). Almost all previous methods considered pixels as the basic unit in the image. However, the superpixel approach, as a group of pixels representing natural entities, is more consistent with human visual cognition and contains less redundancy (Zhou et al., 2017a). This kind of approach has been used through Simple Linear Iterative Clustering (SLIC). However, this technique has important limitations in terms of accuracy and boundary adherence (Xie et al., 2019). In (Romero-Oraá et al., 2019), we hypothesized that *the ERS segmentation algorithm could be useful to segment RLs in fundus images, outperforming the SLIC algorithm.*

Finally, the last study conducted during the course of the present Doctoral Thesis, which is currently under review, was aimed at the ultimate goal in a CAD system: the severity grading. All previous studies aimed at DR severity grading considered the joint detection of all the DR signs at the same time. We hypothesized that *this joint detection makes the classification task harder to optimize than a separate detection of the bright and the dark lesions.*

2.2   Objectives

The main objective of the present Doctoral Thesis is *to study, design and develop novel methods based on the automatic analysis of fundus images to aid in the screening, diagnosis, and treatment of DR.* The proposed methods cover image quality assessment, OD and fovea detection, RL and EX segmentation, and DR severity grading. In order to achieve the main objective, the following specific objectives arise:

  I. To review the state of the art in medical image processing and, in particular, fundus image analysis useful for the diagnosis of DR. This review involves the study of methods for the different stages of a DR CAD system.

 II. To build a private fundus image database with varied quality levels and diverse types of lesions, including DR patients and healthy controls.

III. To select the publicly available databases of fundus images useful for the development and validation of the studies involved in this Doctoral Thesis.

 IV. To implement and optimize the existing image processing techniques studied for every stage of the DR detection process. This goal includes image quality assessment, OD and fovea detection, RL and EX segmentation, and DR severity grading.

  V. To obtain the results that allow evaluating the performance of the proposed methods.

 VI. To discuss and compare the obtained results with those reviewed in the state of the art.

VII. To extract the proper conclusions based on the previous discussion.

VIII. To disseminate the main results and conclusions of the conducted studies in JCR indexed journals and in scientific forums such as national and international conferences.

Chapter 3

Fundus image databases

The fundus image databases that were used during the course of the Doctoral Thesis are described in this chapter. Section 3.1 shows the information related to the private database that was built as a goal of this research. Section 3.2 gathers the set of public databases that were used in one or more of the studies carried out during this Thesis.

3.1   Private database

In order to build a private database, we collected 2107 fundus images from 688 patients provided by the Instituto de Oftalmobiología Aplicada (IOBA) of the University of Valladolid (Valladolid, Spain) and the Hospital Clínico Universitario de Valladolid (Valladolid, Spain). All participants gave their informed consent to participate in the research. The Ethics Committee of the Hospital Clínico Universitario de Valladolid approved the studies conducted using this database in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

All images were captured using the Topcon TRC-NW400 automatic fundus camera (Topcon Medical Systems, Inc., Oakland, NJ, USA) with 45-degree field of view (FOV), and were stored using the 24-bit JPEG format with a size of 1956 × 1934 pixels. Figure 3.1 shows this camera model as well as an example of a color fundus image obtained with it. Images were obtained using the two-field protocol adopted by the National Service Framework for Diabetes in the United Kingdom for DR screening (Department of Health and Social Care (Government of the United Kingdom), 2002). This protocol involves the acquisition of a fovea-
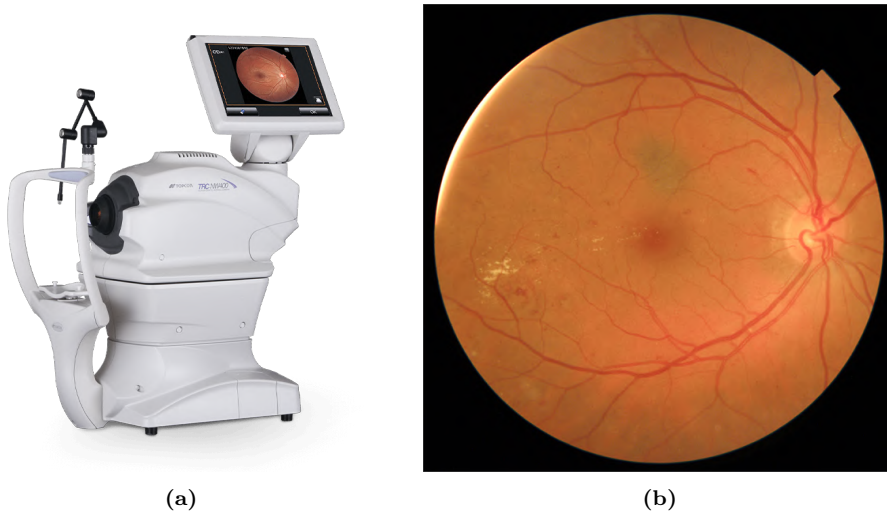
Figure 3.1: (a) Topcon TRC-NW400 automatic retinal camera (Topcon Medical Systems, Inc., Oakland, NJ, USA). (b) Example of a color fundus image obtained with it.

centered image and an OD-centered image per eye. Thereby, four images were captured for every patient (except for those cases where there was some capture-related problem).

Two experienced ophthalmologists determined, for each image, whether they had enough quality to be analyzed or not. Four criteria were considered by the specialists to assess the image quality: (i) The OD edges were well defined, (ii) the blood vessels (especially the main arcades) were well defined, (iii) the retinal parenchyma (i.e., retinal nerve fiber layer) was visible, and (iv) the macula was distinguishable. In general, an image was considered to have an adequate quality when all requirements were met. Based on these criteria, 1810 out of the 2107 images were considered as adequate quality images, while the remaining 297 images had inadequate quality. This manual classification of image quality served as the gold standard for our studies devoted to the automatic assessment of the quality of fundus images (Jiménez-García et al., 2019; Romero Oraá et al., 2020).

Among the fundus images having enough quality to be analyzed, a reduced subset was selected for lesion annotation. Both mentioned ophthalmologists manually drew the outlines of the RLs and EXs on 564 images. To carry out this task, an *ad-hoc* software tool was developed using the proprietary programming language and numeric computing environment Matlab®. While 270 out of 564 fundus images showed DR signs, the remaining 294 lacked any type of lesion. Among the 270

pathological images, 183 showed EXs, 239 showed RLs, and 152 images included both EXs and RLs. The manual annotation of these lesions served us as the gold standard for the study aimed at the detection of RLs (Romero-Oraá et al., 2019), and for the study aimed at the joint detection of RLs and EXs (Romero-Oraá et al., 2020).

Finally, the ophthalmologists manually annotated the OD and fovea centers in the same subset of fundus images selected for lesion annotation. This way, we also had the gold standard coordinates for these two important landmarks. This allowed us to use this database to carry out our study (Romero-Oraá et al., 2020), aimed at the automatic location of the OD and the fovea.

Table 3.1 summarizes the content of the collected private database.

3.2 Public databases

Several public databases of fundus images can be found in the literature, and many of them were created as research material for the automatic diagnosis of DR (Decencière et al., 2014). However, they were created for different purposes. For example, some were aimed at conducting studies on automatic quality assessment, thus including the manual annotation of the quality level of the images (Sevik et al., 2014). Other databases include the annotation of the main DR lesions (RLs and EXs) (Kauppi et al., 2007). There are also databases with the severity degree of each retina (Decencière et al., 2014; Kaggle, 2015). Even other databases include the manual annotation of the vasculature (Staal et al., 2004). This section details all the public databases used in the studies carried out during the course of this Doctoral Thesis. These were selected for being the most widespread databases, which allows direct comparisons to be made with other previous studies.

Table 3.1: Private database summary.

Total	Quality	Annotation	DR
2107	1810 Adequate quality	564 Annotated	270 Pathological
			294 Control
		1246 Non annotated	
	297 Inadequate quality		

3.2.1   DRIMDB

The Diabetic Retinopathy Image Database (DRIMDB) was provided by the Retina Department of Ophthalmology, Medical Faculty, Karadeniz Technical University (Sevik et al., 2014). All images were obtained with a Canon CF-60UVi fundus camera at 60 degree FOV and were stored in JPEG files at 570 × 760 pixels resolution. This database consists of 216 fundus images. An expert identified three quality classes: "good", "bad", and "outlier". The "outlier" class (22 images) covers the nonretinal images that could have been obtained for several reasons, such as wrong focus or patient absence. The "good" class (125 images) is associated with medically suitable image candidates. The rest (69 images) were considered in the "bad" class (Sevik et al., 2014).

   The DRIMDB dataset was used in our study (Romero Oraá et al., 2020), aimed at fundus image quality assessment. However, the outlier class was not employed. Therefore, only 194 images were considered, corresponding to the 125 images of the good class and 69 images of the bad class. It is important to note that this database was only used for validation purposes.

3.2.2   DIARETDB1

The DIARETDB1 database is composed of 89 images captured in the Kuopio University Hospital and divided into a training set (28 images) and a test set (61 images) (Kauppi et al., 2007). They were captured at a 50-degree FOV and had a resolution of 1500 × 1552 pixels. Only one image per eye was captured, which was fovea-centered. In this database, four medical experts annotated any sign of microaneurysms, hemorrhages, and hard and soft EXs. However, it is important to note that the ground truth of the lesions was only roughly annotated using circles, ellipses, and polygons. According to the lesion annotations, 27 images were associated with healthy retinas, 7 images with mild DR, 28 images with moderate and severe non-proliferative DR, and 27 images with proliferative DR (Kauppi et al., 2007).

   We used the DIARETDB1 database in three studies: (Romero-Oraá et al., 2019), (Romero-Oraá et al., 2020), and (Romero-Oraá et al., 2020). The first study was aimed at the detection of RLs, the second one focused on the location of the OD and the fovea, and the third study was aimed at the joint detection of RLs and EXs. For the second study, one of our ophthalmologists had to add to this database the ground truth of the centers of the OD and the fovea. It important to mention that DIARETDB1 was not used for model building in either

of the mentioned studies. Instead, this database was exclusively used to assess the performance of the final methods on a new set of images obtained with different protocols, resolution and quality.

3.2.3 DRIVE

The Digital Retinal Images for Vessel Extraction (DRIVE) (Staal et al., 2004) dataset consists of 40 images obtained from a diabetic screening program conducted in the Netherlands. Images were captured from a Canon CR5 nonmydriatic 3 charge-coupled device at 45 degree FOV, and compressed in 24-bit JPEG format. The size of the images was 768×584 pixels and all of them were fovea-centered. In this dataset, 7 out of the 40 images showed pathological signs.

This database was intended to carry out studies for the automatic segmentation of the vascular network. This is why three observers marked all pixels associated with blood vessels (Staal et al., 2004). However, we used this database to evaluate the performance of our method (Romero-Oraá et al., 2020), aimed at the location of the OD and the fovea. For this task, one of our ophthalmologists annotated the corresponding centers. In our experiments for the detection of the fovea, 3 images were excluded from the study for not showing a visually detectable fovea (Welfer et al., 2011) (see Figure 3.2). In our study, the DRIVE dataset was not used for the development of the method but for validation purposes.

3.2.4 Messidor

The Messidor database comprises 1,200 images captured using a Topcon TRC NW6 retinal camera at a 45-degree FOV (Decencière et al., 2014). They were

(a)	(b)	(c)
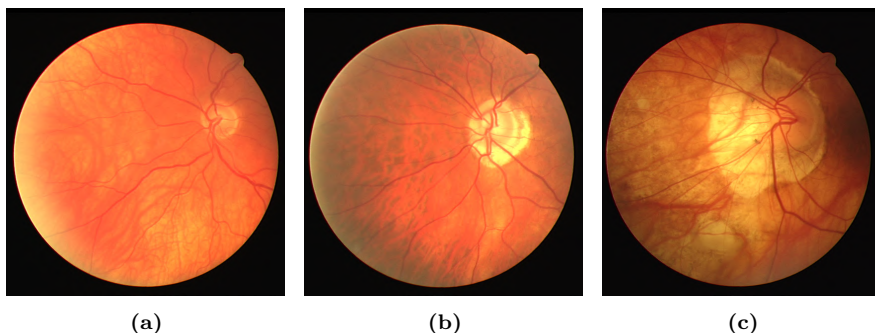
Figure 3.2: Images from DRIVE database that were excluded for the detection of the fovea since this structure is not visually detectable.

stored in 24-bit TIFF format at three different resolutions: 1440 × 960, 2240 ×
1488 and 2304 × 1536 pixels. All the images were fovea-centered. The database
contains a medical diagnosis for each image (severity degree), but no manual an-
notations on the images, such as lesions contours or position (Decencière et al.,
2014).

As for DRIVE, the Messidor database was used in our study designed for OD
and fovea location (Romero-Oraá et al., 2020). Likewise, one of our ophthal-
mologists annotated the corresponding centers of both structures. The Messidor
dataset was only applied to assess the robustness of the proposed method.

3.2.5   Kaggle

This retinal image dataset was provided by EyePACS, a company devoted to DR
screening solutions, for the Diabetic Retinopathy Detection competition published
on Kaggle (Kaggle, 2015), hence the name by which it is known. It is the largest
public DR-related database available with 88,702 images: 35,126 images meant for
training and 53,576 for testing. In this database, a clinician rated the DR severity
degree according to the International Clinical DR Scale (Wilkinson et al., 2003),
which classifies the data into 5 levels: 65,343 images with no DR, 6,205 images
with Mild NPDR, 13,153 images with Moderate NPDR, 2,087 images with Severe
NPDR, and 1,914 images with Proliferative DR. As noted above, the dataset is
highly imbalanced. A fundus image was included from both eyes of every patient.
The images in the dataset come from different models and types of cameras, which
can affect its visual appearance and the output resolution. Additionally, they may
contain artifacts, be out of focus, underexposed, or overexposed. This way, it
represents a real-world scenario where images are affected by noise (Kaggle, 2015).
However, we have noticed that many of the images are not suitable for analysis,
as stated in other studies (Rakhlin, 2018). These cases should not be considered
for diagnosis and, therefore, were discarded using the RIQA algorithm proposed
in this Thesis and described in section 4.1.2. After discarding the poor-quality
retinal images, this dataset contains 52973 fundus images with adequate quality.

The Kaggle database was used in the last of the studies included in the present
Doctoral Thesis, aimed at DR severity grading. In this study, Kaggle was the only
dataset employed because it is the most widely used in the literature and contains a
much larger number of images than other databases. Kaggle was divided into three
groups, preserving the original division after discarding the poor-quality images:
training (18,860 images), validation (2,096 images), and test (32,017 images).

Chapter 4

Methods

This chapter is devoted to explain the different methods that have been developed during the course of this Doctoral Thesis. These methods have been grouped in the following sections according to the stage of image analysis in the diagnosis of DR. Section 4.1 describes the method proposed for the image quality assessment of fundus images as well as a newer alternative based on deep learning. The method corresponding to the preprocessing stage is detailed in section 4.2. The method for the automatic location of the OD and the fovea is presented in section 4.3. Section 4.4 provides two methods aimed at RL segmentation while section 4.5 explains the method for EX segmentation. Finally, section 4.6 shows the proposed approach for DR severity grading.

4.1 Image quality assessment

A large number of retinal images are not suitable for manual nor automatic analysis due to poor quality. The development of RIQA methods is necessary to prevent misdiagnosis (Paulus et al., 2010). In this Doctoral Thesis we have conducted two studies with this purpose. The first one is included in the compendium of publications and is based on novel generic quality features (Jiménez-García et al., 2019). The second study was presented in an international conference and was based on deep learning techniques (Romero Oraá et al., 2020). Both methods are included in this manuscript due to the great relevance of this stage. RIQA is essential to form a complete CAD system. Both methods are described in the following subsections.

4.1.1   Method based on the combination of global features

This method was proposed in (Jiménez-García et al., 2019) as part of the compendium of publications of the present Doctoral Thesis. Our methodology was composed of four stages. First, a preprocessing stage was implemented to adapt the images for subsequent processing. Second, various generic features were extracted for every image. Third, the set of extracted features was reduced using a feature selection algorithm. Finally, an MLP neural network was trained using this reduced subset of features to assess the quality of images.

Preprocessing

The processing of fundus images tends to produce unwanted border effects around the FOV. Moreover, the black pixels outside the FOV should not be considered for analysis. To solve these issues, a preprocessing stage needs to be applied. However, this stage is not required to enhance the retinal structures or normalize the appearance since the purpose of the study was to only assess the image quality. Hence, a specific preprocessing stage was applied.

First, the circular region that represents the FOV was segmented by estimating its diameter and center (García et al., 2010). The FOV diameter was estimated using the intensity profile along one diagonal of the image. Then, the circular Hough transform was used to find the FOV center (García et al., 2010). These two parameters allowed us to generate a circular mask (M_{FOV}) of the FOV. Then, the retinal fundus was extended outside of the FOV to reduce the influence of border effects near the FOV boundaries (Soares et al., 2006). For this task, we applied an iterative algorithm using the original image, I_{ORIG}, and M_{FOV}. For each iteration of the algorithm, the following operations were performed (Soares et al., 2006):

1. The border of M_{FOV} was extended using a dilation operator with a 4-neighborhood diamond-shaped structuring element.

2. The pixel values of I_{ORIG} associated with the dilated border were replaced with the average value of the neighbor pixels inside the dilated mask of the previous iteration.

This workflow was iteratively repeated on every channel of the RGB image until the black pixels outside the FOV were completely replaced. It is important to note that the pixels inside the FOV remained unaltered but the border effects around the FOV were prevented.

Feature extraction based on spatial and spectral entropies

SSEQ is a NR-IQA method based on features related to spatial and spectral entropies computed in small non-overlapped regions (blocks) of the image (Liu et al., 2014). In order to apply this technique, first, three input images were obtained: the green channel of I_{PREP} and 2 down-sampled versions of the same image obtained with using bicubic interpolation. In this study, scale factors 1 (1956 × 1934 pixels), 1/2 (978 × 967 pixels), and 1/3 (652 × 645 pixels) were used to perform a multi-scale analysis (Liu et al., 2014). Next, images were divided into blocks of size $M \times M$ pixels (Liu et al., 2014).

Spatial entropy ($SpacEn$) was then computed for each block as the Shannon's entropy (Liu et al., 2014). Spectral entropy ($SpecEn$) was also computed for each block. To obtain an spectral representation of the image blocks, we applied the 2-D DCT to obtain the normalized power spectral density (PSD) (Liu et al., 2014). $SpecEn$ was then calculated for each block as the Shannon's entropy of the normalized PSD (Liu et al., 2014).

After computing $SpacEn$ and $SpecEn$ in all blocks, values between percentiles 15% and 85% were exclusively considered, based on the original work (Liu et al., 2014). This way, only the central part of each distribution was analyzed and $SpacEn$ and $SpecEn$ were less sensitive to outliers (Liu et al., 2014). Finally, the mean and the skewness of both entropies through all the selected blocks were calculated, obtaining the features $SpacEn_{MEAN}$, $SpacEn_{SKEW}$, $SpecEn_{MEAN}$, and $SpecEn_{SKEW}$ (Liu et al., 2014). This process was repeated for the rescaled images. Hence, 12 features were extracted using the SSEQ method (2 mean values and 2 skewness values for each of the 3 scales). It should be noted that the block size M must be fixed. Although M is not a critical parameter, it has been selected as 1/60 of the image size in previous studies (Liu et al., 2014). Following this approach, we set $M = 32$ pixels according to the size of the images in the training set.

Feature extraction based on naturalness

NIQE is a NR-IQA method based on the comparison of an image with a reference model which represents the characteristics of images of adequate quality (Mittal et al., 2013). Prior knowledge about the potential distortions is not required to build the reference model. Adequate quality images are expected to have similar features to those in the reference model, while the distorted images are expected to show larger differences. Image quality assessment using NIQE was performed by

comparing two NSS models: one that represented the reference model and another that characterized the image to be assessed (Mittal et al., 2013).

To build the reference model, we built a NSS model using the adequate quality fundus images in the training set of our database. For this task, a set of parameters was extracted from the blocks in these images. The process to obtain the reference NSS model is as follows (Mittal et al., 2013):

1. The image I_{PREP} was normalized to obtain the image I_{NORM}: the local mean was subtracted for each pixel and the result was divided by the local standard deviation (Mittal et al., 2013).

2. The image I_{NORM} was divided into blocks of size $P \times P$ pixels. Then, we selected a subset of the blocks that exceeded a minimum local sharpness (Mittal et al., 2013).

3. Each of the selected blocks was subsequently characterized by a zero-mean generalized Gaussian distribution (GGD). Additionally, in each of the selected blocks, the products between adjacent pixels along 4 directions were calculated and characterized by four asymmetric generalized Gaussian distributions (AGGD). The process was repeated with a rescaled version of the same image in order to perform multi-scale analysis (978 × 967 pixels). A total of 36 parameters characterized each block (2 from the GGD and 16 from 4 AGGDs) using 2 scales.

4. Steps 1–3 were repeated for each image used to build the reference model.

5. The parameters from selected blocks in all the images were fitted to a 36-D multivariate Gaussian (MVG) model. The MVG probability distribution is defined as (Mittal et al., 2013):

$$f_X(\boldsymbol{\chi}) = \frac{1}{(2\pi)^{k/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{\chi}-\boldsymbol{\nu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\chi}-\boldsymbol{\nu})}, \tag{4.1}$$

where the vector $\boldsymbol{\nu}$ and the covariance matrix $\boldsymbol{\Sigma}$ define the MVG model (Mittal et al., 2013). In this work, the parameters of the reference NSS model were denoted by $\boldsymbol{\nu_M}$ and $\boldsymbol{\Sigma_M}$.

To build the NSS model that characterizes the image to be assessed, the process was identical except for that fact that only the image under study was considered. Thus, the resulting NSS model is another MVG model defined by the parameters $\boldsymbol{\nu_I}$ and $\boldsymbol{\Sigma_I}$

Once both models were built, we computed the NIQE quality index (Q_{NIQE}) of the image to be assessed as the distance (Mittal et al., 2013):

$$Q_{NIQE} = \sqrt{(\mathbf{v_M} - \mathbf{v_I})^T \left(\frac{\mathbf{\Sigma_M} + \mathbf{\Sigma_I}}{2} \right)^{-1} (\mathbf{v_M} - \mathbf{v_I})}. \qquad (4.2)$$

This metric, which represents the quality of the whole image, is the only feature extracted and passed to the subsequent classifier using the NIQE method. It can be seen that a low value of the NIQE quality index is associated to an image that resembles the reference model. Therefore, this quality index can be considered a measure of the naturalness of an image (Mittal et al., 2013).

During the development of the NSS models, block size P and the threshold T had to be fixed. Previous studies proposed that values of P between 32–160, as well as values of T between 0–1, were adequate for quality assessment (Mittal et al., 2013). For our images, we empirically found that the greater differences between adequate quality and inadequate quality images in the training set were obtained with $P = 64$ pixels and $T = 0.1$, which is consistent with previous studies (Liu et al., 2014).

Feature extraction based on the continuous wavelet transform

Another set of features extracted in this study were derived from the CWT. The 2D CWT decomposes the image I_{PREP} into several representations related with a scale factor s (Antoine et al., 1993). As the mother wavelet, we selected the Mexican hat since it is proven suitable for the detection of sharp edges in medical images (Antoine and Murenzi, 1996; Rangayyan, 2004).

The CWT was applied to the green channel of I_{PREP} at scales $s = 2, 4, 8,$ 16, 32 and 64. These scales were specifically selected to detect the borders of retinal structures (blood vessels, OD, and macula) (Niemeijer et al., 2006, 2007b). Then, the variability of the CWT was calculated using the Shannon's entropy (ENT_{CWT}) (Abdel-Hamid et al., 2016).

It has been observed that the amplitude of the CWT around the vessels and the OD was directly related to the sharpness of their edges (Jiménez-García et al., 2019). Thus, the CWT can be useful to identify if the retinal structures are clearly visible in the images. We assessed edge sharpness by calculating the local variance (Aja-Fernández et al., 2006). For this task, standard deviation filters were applied and the distributions of the obtained local variance maps were analyzed (Aja-Fernández et al., 2006; Gonzalez and Woods, 2009). Since different scales of

the CWT emphasize the borders of objects of different sizes, a circular standard deviation filter with radius s was selected for each scale (Aja-Fernández et al., 2006). Therefore, a circular filter with radius $r = s$ was selected. For each scale, the mean ($MEAN_{CWT}$) and the standard deviation (SD_{CWT}) of the local variance maps were computed inside the FOV. Consequently, the CWT analysis resulted in 18 new features: ENT_{CWT}, $MEAN_{CWT}$, and SD_{CWT} for each of the 6 scales.

Feature extraction based on luminosity

Often, fundus images are unexposed or underexposed. Similarly, images are frequently captured with uneven illumination. In these scenarios, the retinal structures and lesions may not be properly appreciated (Bartling et al., 2009; Shao et al., 2017; Wang et al., 2016). Therefore, luminosity features can be useful to identify poorly illuminated images. When luminosity is to be analyzed, the HSV color model is more appropriate than RGB (Davis et al., 2009; Shao et al., 2017; Wang et al., 2016). The luminosity (V) channel of HSV separates luminosity from color information and is designed to represent the differences between light and dark areas (Zhou et al., 2018). Using this channel, the background of the image can be estimated. First, noise was removed by applying a median filter with a square neighborhood of size 5 × 5 pixels (Gonzalez and Woods, 2009). Then, the background of the images, $B(x, y)$, was extracted using a large Gaussian filter (Foracchia et al., 2005; Gonzalez and Woods, 2009). The standard deviation, σ, of this filter should be large enough to estimate the background and remove the vasculature and the rest of dark structures (Marín et al., 2011). We empirically set $\sigma = 19$ according to the image size in our training set.

In order to obtain the luminosity level of the darkest areas of the image, we computed the percentiles 1% (Lum_1), 5% (Lum_5), 10% (Lum_{10}), 15% (Lum_{15}), and 20% (Lum_{20}) of the pixel values in $B(x, y)$. For those images with a lighter and more uniform background, these percentiles would be higher than for images where darker background areas appear. In order to detect the uneven illumination, the differences between consecutive luminosity percentiles were also calculated (Lum_{5-1}, Lum_{10-5}, Lum_{15-10}, and Lum_{20-15}). These differences can represent intensity variations along the background. The luminosity analysis resulted in 9 additional features.

Feature selection: Fast Correlation-Based Filter

A total number of 40 features were finally extracted for each image. Some of them may be redundant or irrelevant for our specific problem (Guyon and Elisseeff, 2003; Saeys et al., 2007). It is also important to note that a large number of features may lead to overfitting and reduce the performance of the classifier (Saeys et al., 2007). Instead, learning can be achieved more efficiently and effectively with just relevant and non-redundant features (Yu and Liu, 2004). Feature selection algorithms try to overcome these difficulties by obtaining a reduced and optimum subset of features for a certain problem (Saeys et al., 2007).

The fast correlation-based filter (FCBF) feature selection algorithm was used to identify relevant and non-redundant features (Yu and Liu, 2004). This is a classifier-independent method. This way, different classifiers could be explored using the same selection of features. Moreover, the algorithms based on the subset evaluation approach can become very inefficient for high-dimensional data. FCBF performs an efficient, explicit relevance and redundancy analysis which has shown a great effectiveness (Yu and Liu, 2004).

FCBF has two stages. In the first stage, features are ordered according to their relevance. In the second stage, redundant features are removed. FCBF uses symmetrical uncertainty (SU) to assess both relevance and redundancy. It is defined as (Yu and Liu, 2004):

$$SU(X_i|X_j) = 2\frac{H(X_i) - H(X_i|X_j)}{H(X_i) + H(X_j)}, \tag{4.3}$$

where $H(X_i)$ is the Shannon entropy of the feature X_i, and $H(X_i|X_j)$ is the Shannon entropy of the feature X_i after the observation of the feature X_j.

The relevance of feature X_m is defined as the SU between the class C (in this case, image quality) and X_m. In the same way, redundancy is defined as the SU between pairs of features (X_m and X_n). Feature X_m is considered redundant with X_n, and thus removed, if (Yu and Liu, 2004):

$$SU(X_n|C) \geq SU(X_m|C), \text{and } SU(X_m|X_n) \geq SU(X_m|C). \tag{4.4}$$

In order to improve the robustness of the feature selection process, a bootstrapping procedure was implemented (Witten et al., 2016). Instances from the training set were randomly selected using the sampling with replacement technique to form bootstrap replicates (Witten et al., 2016). For each replicate, instances from the training set were sampled with uniform probability (repeated instances

were allowed) until the original training set size was reached (Witten et al., 2016). We formed 1000 bootstrap replicates and applied the FCBF algorithm to each one. Features that were selected on at least half (500) of the runs formed the final optimum subset. Using bootstrapping, the feature selection stage was less dependent on the particularities of training set data (Guyon and Elisseeff, 2003).

Classification: Multilayer Perceptron Neural Network

Once a reduced subset of selected features was obtained, we employed an MLP neural network to classify the images into two categories: adequate quality and inadequate quality. This type of network has been widely used in classification tasks in the field of retinal image processing (García et al., 2009; Marín et al., 2011; Pires Dias et al., 2014). MLPs are feed-forward networks consisting of various fully connected layers of neurons (Bishop, 1995; Haykin, 1999). They map a set of input variables onto a set of output variables using a nonlinear function (Bishop, 1995). An MLP is composed of one input layer, one output layer and, at least, one hidden layer of neurons (Bishop, 1995; Haykin, 1999). Optionally, more hidden layers can be added between the input layer and the output layer (Bishop, 1995). Nevertheless, an MLP with a single hidden layer of neurons is capable of universal approximation (Huang et al., 2000). Therefore, a three-layer MLP network (input, hidden, and output layers) was implemented in this study to solve the classification task (Bishop, 1995). The input layer had a number of neurons equal to the number of selected features. The output layer had only one neuron to perform binary classification (Bishop, 1995; Witten et al., 2016). The number of hidden neurons (N_{HIDDEN}) was experimentally obtained to optimize the classification performance (Bishop, 2006; Witten et al., 2016). Hyperbolic tangent sigmoid activation function (Bishop, 1995) was used in the hidden layer since it has been successfully applied in similar MLP networks from previous studies (García et al., 2009). The logistic sigmoid activation function was selected in the output neuron since it is defined in the range (0–1) and, consequently, MLP outputs can be interpreted as posterior probabilities (Bishop, 1995).

The training process of the MLP was aimed at minimizing an error function. We selected a cross-entropy error function for the minimization process (Bishop, 1995). Additionally, the scaled conjugate gradient was used for optimization in this study, since it generally converges faster than other techniques (Bishop, 1995). It is necessary to note that MLP training may result in overfitting, leading to errors when the network is tested on new data. To overcome this problem, weight decay regularization was implemented during training (Bishop, 2006). This technique

redefines the error function by adding a penalty term that increases as the magnitude of weights increases (Bishop, 2006). The modified error function depends on a regularization parameter (λ) that balances the contribution of the cross-entropy error function and the sum of the squares of the weights (Bishop, 2006). The value of this parameter was also experimentally obtained.

As described in chapter 3, the training set of the private database employed in this study had 148 inadequate quality images, considerably less than the number of adequate quality images (905 images). In order to deal with class imbalance during training, we increased the number of instances corresponding to inadequate quality images using the synthetic minority oversampling technique (SMOTE) method (Chawla et al., 2002). This method creates synthetic instances of the data by combining the k-NN of each sample in the minority class (Chawla et al., 2002). The number of synthetic samples depends on the number of neighbors (k) (Chawla et al., 2002). We set $k = 5$ in the training stage to obtain 740 synthetic minority class training samples. Thus, the training set was finally comprised of 905 adequate quality instances and 888 inadequate quality instances.

4.1.2 Method based on deep learning

The second approach aimed at RIQA was based on deep learning. The proposed method required a minimal preprocessing step to normalize the input images. Next, a CNN was used to separate fundus images of adequate quality from the rest. It is important to highlight the use of data augmentation and transfer learning. Finally, the training procedure is specified.

Preprocessing

As for the method based on the combination of global features, a minimal preprocessing step was required. However, this stage was not intended to inspect the image but to adapt it to the input of the method. For this reason, a specific preprocessing stage was applied for this study. In this stage, all images were resized to 299x299 pixels to fit the CNN input layer later used (the CNN architecture employed in this study was built for these image dimensions) (Zago et al., 2018). If the input image was not square (as in DRIMDB database), this operation changed the aspect ratio. Second, the images were normalized to the interval $[-1, 1]$ by subtracting 0.5 and dividing by 0.5 the $[0, 1]$ intensity image (Zago et al., 2018). It should be noted that the preprocessing stage is independent of the database used, and can be adapted to images of various characteristics.

Data augmentation

Deep neural networks are proven to work better when trained with a large amount of data. In order to increase the number of training samples, we applied online data augmentation. This technique involves generating new random, fake samples from the original data to feed the model in every training batch with new images (Perez and Wang, 2017). In this study, we applied the following simple transformations (Zago et al., 2018):

- Rotations in the range $[-20, +20]$ degrees.

- Vertical and horizontal shifts up to a maximum of 7% of the image width.

- Horizontal and vertical flips.

- Zoom in the scale range $[0.85, 1.15]$.

Transfer learning

Training a deep network from scratch could be very slow. Additionally, when the training data is limited, the model may not converge or may achieve poor results. In these situations, the use of transfer learning is a common practice that has proven to work remarkably well (Pan and Yang, 2010). This technique allows the resolution of a machine learning problem in a particular domain of interest with the knowledge learned from the training data of another domain of interest (Pan and Yang, 2010). In practice, the easiest way to apply transfer learning is the use of pretrained networks, where the model is initialized with a set of pretrained weights. Then, fine-tuning is applied. This operation consists of retraining all or some of the final layers of the CNN with a training set adapted to the specific problem. In this way, the weights are better adapted to the given task (Coyner et al., 2018).

The concept of transfer learning and pretrained networks has already been successfully used in RIQA (Coyner et al., 2018; Saha et al., 2018; Yu et al., 2017; Zago et al., 2018). In this work, we used a CNN pretrained on the images from the project ImageNet (Jia Deng et al., 2009). This database is made up of more than 14 million generic images belonging to more than 20,000 different classes, such as dogs, strawberries and airplanes. We used the pretrained weights of ImageNet because is the largest image database publicly available and it has been extensively tested (Jia Deng et al., 2009).

Convolutional neural network architecture

CNNs are neural networks composed of convolutional layers, as well as others such as pooling and fully connected layers. They are capable of extracting representative features from a large set of raw images in an optimized way (Saha et al., 2018; Zago et al., 2018). In conventional neural networks, each neuron is connected to all the neurons in the next layer. When the network input is an image, that is, a matrix composed of a large number of pixels, these connections involve a huge number of internal parameters, which is not feasible for training. On the contrary, in CNNs each neuron is only connected to a neighboring subset, which allows to significantly reduce the parameters of the network to be trained. This property is known as sparse interactions (Wang et al., 2019). Additionally, CNNs are characterized by weight sharing, where all units in a layer use the same weights and deviations. This feature reduces the network training parameters to a greater extent, which can effectively prevent the network from overfitting and improve the efficiency of network operation (Wang et al., 2019).

In this work, a CNN with InceptionResNetV2 architecture was used, which was adapted to our binary classification problem. This hybrid architecture is a combination of Inception (Szegedy et al., 2015) and ResNet (He et al., 2016). On one hand, the Inception architecture includes filters of multiple sizes operating at the same level. This allows better learning of the information that is distributed in regions of various sizes on the image without excessively increasing the computational cost (Szegedy et al., 2015). The latest version of this network is known as Inception v4 and has a simpler architecture than previous versions (Szegedy et al., 2017). On the other hand, ResNet facilitates the learning process by adding residual connections between the output and input of each Inception module (Sevik et al., 2014). In this way, hybrid InceptionResNet models combine the advantages of both networks, being possible to obtain higher accuracies in earlier epochs (Szegedy et al., 2017). For this reason, the architecture used in this work is InceptionResNetV2, which is 572 layers deep (Szegedy et al., 2017). In order to adapt this architecture to our binary classification problem, we replaced the last 3 layers by 3 fully connected layers of 1024, 512 and 1 neuron, respectively (Zago et al., 2018). The first two had a ReLU activation function (Zago et al., 2018). For the latter, a sigmoid activation function was used, which provides an output between 0 and 1 and can be interpreted as a posterior probability.

Training procedure

All layers of our CNN were fine-tuned for 100 epochs (Coyner et al., 2018). Binary cross-entropy was applied as the loss function and stochastic gradient descent as the optimization algorithm (Coyner et al., 2018; Saha et al., 2018; Zago et al., 2018). The learning rate was set to 0.001 and the momentum value to 0.9 (Chalakkal et al., 2019; Zago et al., 2018). To avoid overfitting in advanced epochs, the learning rate was reduced by a factor of 10 every time the validation error reached a plateau and remained constant (Chalakkal et al., 2019; Zago et al., 2018). In addition, a batch size of 16 images was established (Zago et al., 2018).

4.2 Preprocessing

This section details the preprocessing method proposed in (Romero-Oraá et al., 2019), which is part of the compendium of publications of this Doctoral Thesis. This method deserves special attention because it was employed in the methods for the location of the OD and the fovea and the segmentation of DR lesions, which are presented in the following sections of this chapter.

The appearance of the fundus images relies on the intrinsic features of the patient, such as his skin or iris color, among others (Osareh et al., 2009; Sánchez et al., 2009). Furthermore, local illumination and contrast are often non-uniform within the retinas. Therefore, we can find a wide variability of images in terms of color, luminosity, contrast and quality (Sánchez et al., 2009).

When processing fundus images, the black pixels in the border cause unwanted border effects. For this reason, it is important to separate the FOV from the surrounding black border. This way, only the pixels of the image belonging to the retina are considered. In this study, the FOV was automatically determined from the original image, I_{orig}, prior to preprocessing, by estimating its diameter and center. The diameter, D, was estimated analyzing the intensity profile along one image diagonal in the red component (García et al., 2010). Then, FOV edges were found using a Canny edge detector and the FOV center was estimated with a circular Hough transform (García et al., 2010).

Once the mask of the FOV was obtained, we proceed to intra-image and inter-image normalization. Additionally, retinal landmarks were enhanced, with a special attention on improving the visualization of the lesions. For this task, we performed five sequential operations on I_{orig}, shown in Figure 4.1a:
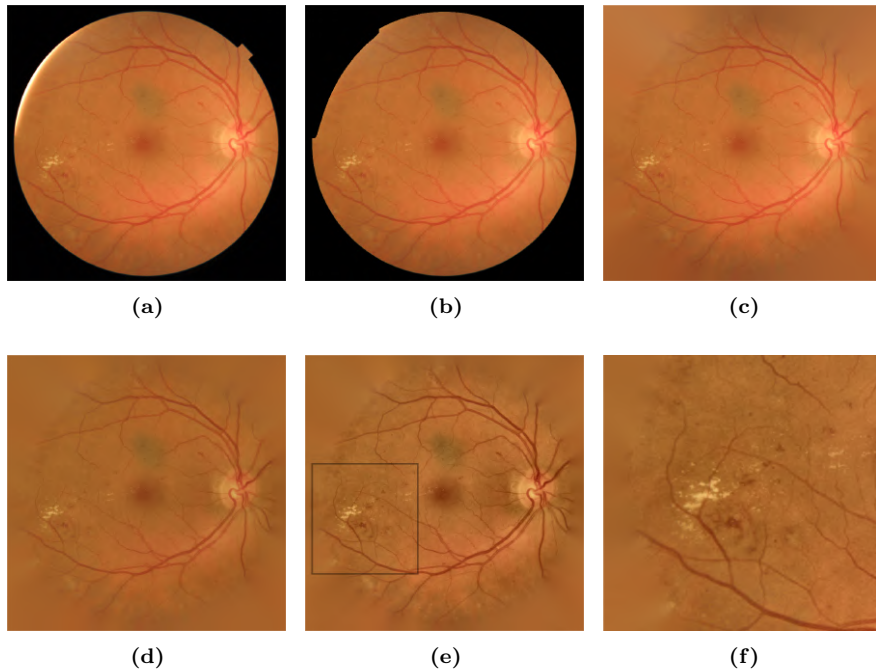
Figure 4.1: Preprocessing stage: (**a**) original image, I_{orig}; (**b**) image after bright border artifact removal, I_{rem}; (**c**) effect of background extension, I_{ext}; (**d**) result after illumination and color equalization, I_{eq}; (**e**) final preprocessed image with contrast enhancement, I_{prep}; (**f**) zoom of the final preprocessed image.

- **Bright border artifact removal.** Some images present excessively bright regions along the FOV border. This is generally due to inadequate illumination during image acquisition. While they do not prevent image visualization, these bright regions may be problematic in the remaining stages of the proposed method due to their color and border features (Zhang et al., 2014). Bright border artifacts were removed using the blue channel, B_{orig}, in the RGB color space (Zhang et al., 2014). First, a mask of bright pixels was computed as $B_{mask} = B_{orig} - B_{mean}$, where B_{mean} is a mean-filtered version of B_{orig} (Zhang et al., 2014). Second, a morphological reconstruction was applied using the FOV contour as the marker and B_{mask} as the mask (Zhang et al., 2014). Then, the binary mask of the bright border artifacts was obtained by applying a threshold and a morphological opening (Zhang et al., 2014). Figure 4.1b shows the resulting image after the artifact removal, I_{rem}, over the original image shown in Figure 4.1a.

- **Background extension.** The strong contrast between the retinal fundus
  and the black pixels outside the FOV may produce border effects in later
  operations (Soares et al., 2006). For each RGB channel of I_{rem}, retinal fundus
  was extended outside of the FOV using an iterative algorithm. Pixels outside
  the aperture were replaced with the mean value of the neighboring pixels
  inside the FOV (Soares et al., 2006), as described in section 4.1.1. The result
  of this operation, I_{ext}, can be seen in Figure 4.1c.

- **Illumination and color equalization.** Retinal images have often non-
  uniform illumination and different color ranges. Equalizing illumination as
  well as overall color is necessary to obtain standardized images. For this
  task, we used the HSV image version of I_{ext}. First, local variations of in-
  tensity within the image were equalized over the intensity channel (I_{ext_I})
  by applying (Hoover and Goldbaum, 2003):

$$I_{eq_I} = I_{ext_I} + \mu - I_{ext_I} * h_{m1}, \tag{4.5}$$

  where h_{m1} is a large mean filter, similar in size to the OD (Seoud et al., 2016),
  and the parameter μ is the average pixel intensity of I_{ext_I} inside the FOV for
  all the images of the training set. This allowed us to normalize the overall
  illumination among images. To normalize hue (I_{ext_H}) and saturation (I_{ext_S})
  channels, we applied:

$$I_{eq_{color}} = I_{in} + \mu_1 - \mu_2, \tag{4.6}$$

  where μ_1 is the average pixel intensity of I_{in} inside the FOV of the images
  of the training set, and μ_2 is the average pixel intensity inside the FOV
  of the input image (I_{ext_H} or I_{ext_S}). Given that I_{ext_H} represents an angle,
  the following equation was required to compute the average hue:

$$\mu_{hue} = \operatorname{atan2} \Big[\sum_{i=1...N} \sin I_{ext_H}(i), \sum_{i=1...N} \cos I_{ext_H}(i) \Big], \tag{4.7}$$

  where N is the total number of pixels in the image. Finally, converting the
  HSI image back to the RGB space, we obtained the image I_{eq} shown in
  Figure 4.1d.

- **Denoising.** This step allowed us to reduce the noise associated with image
  capture and compression. The noise was eliminated by applying an additional
  mean filter over I_{eq} to obtain I_{den} (Seoud et al., 2016). The size of this filter

was very small, empirically set to 3 pixels, with the purpose of preserving the smallest lesions (Seoud et al., 2016).

- **Contrast enhancement.** Some fundus images show poor contrast. In this cases, retinal structures and lesions can not be clearly inspected. The contrast limited adaptive histogram equalization (CLAHE) method was applied to enhance local contrast. This is a histogram processing technique that operates on small regions, highlighting the edges of the DR-related lesions to be later segmented (Rasta et al., 2015). Contrast enhancement was the last step to obtain the final preprocessed image I_{prep}, which can be seen in Figure 4.1e. Figure 4.1f shows an zoomed-in region of the image, where several lesions are present.

The different operations in the preprocessing stage allowed us to highlight the retinal lesions and avoid border effects in subsequent stages. Inter-image normalization was also achieved, as it can be seen in Figure 4.2. Two original images from our private database show different illumination, contrast and color. However, their corresponding preprocessed images have a more similar appearance.

For a better understanding of the complete preprocessing method, the pseudocode in Algorithm 1 is provided:

Algorithm 1: Preprocessing

---
1   $I_{rem} = \text{brightBorderArtifactRemoval}(I_{orig})$;
2   $I_{ext} = \text{backgroundExtension}(I_{rem})$;
3   $I_{ext_{HSI}} = I_{ext} \rightarrow \text{HSI color space}$;
4   $I_{eq_I} = I_{ext_I} + \mu - I_{ext_I} * h_{m1}$;
5   $I_{eq_H} = I_{ext_H} + \mu_{1_H} - \mu_{2_H}$;
6   $I_{eq_S} = I_{ext_S} + \mu_{1_S} - \mu_{2_S}$;
7   $I_{eq} = (I_{eq_I}, I_{eq_H}, I_{eq_S}) \rightarrow \text{RGB color space}$;
8   $I_{den} = \text{meanFilter}(I_{eq})$;
9   $I_{prep} = \text{CLAHE}(I_{den})$;

4.3   Location of the optic disc and the fovea

In this section we describe our algorithm proposed in (Romero-Oraá et al., 2020), aimed at the automatic location of the OD and the fovea. Figure 4.3 shows an overview of the method, which starts by obtaining the preprocessed image, I_{prep}, using the algorithm described in section 4.2. Next, three images were obtained
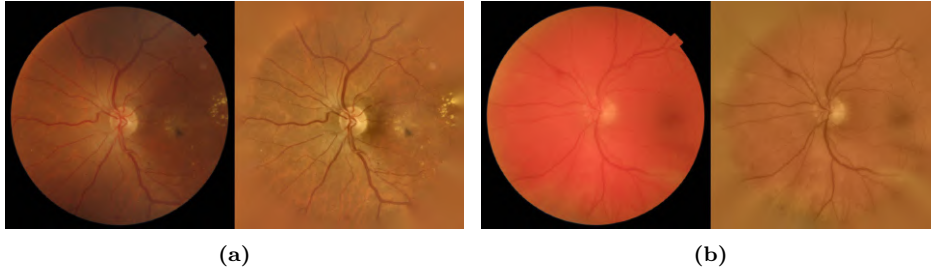
Figure 4.2: Preprocessing for fundus images with different illumination, contrast and color. (**a**) Example 1. (**b**) Example 2.

from the retinal background extraction: I_{bg} is the estimated background of the fundus image after removing both bright and dark structures; I_{bg-bri} is the estimated background after removing the dark structures while preserving the bright ones; and $I_{bg-dark}$ is the estimated background after removing the bright structures while preserving the dark ones. The image M_{vess} is the vascular network
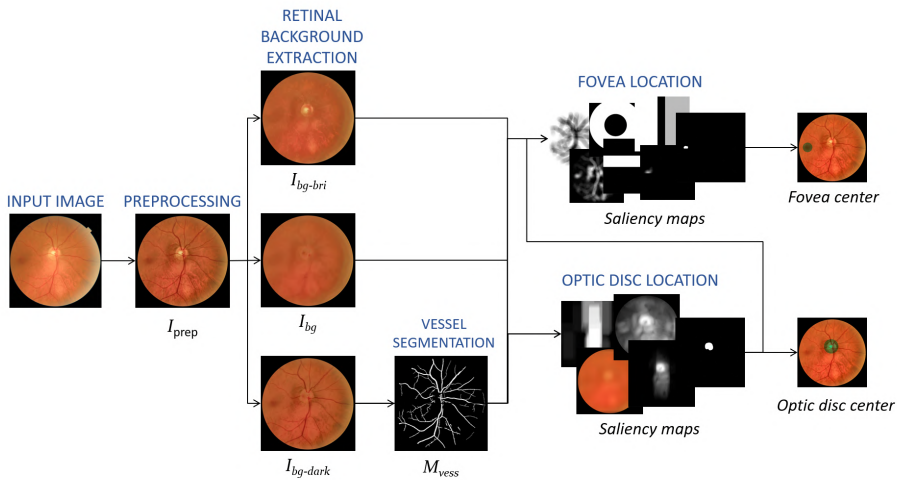
Figure 4.3: Overview of the proposed method. The image I_{prep} is the result of the preprocessing stage. I_{bg} is the estimated background of the fundus image after removing both bright and dark structures. I_{bg-bri} is the estimated background after removing the dark structures while preserving the bright ones. $I_{bg-dark}$ is the estimated background after removing the bright structures while preserving the dark ones. The image M_{vess} corresponds to the vessel segmentation result from $I_{bg-dark}$. For the subsequent detection of the OD and the fovea, various saliency maps were computed using the previous obtained images. Finally, the saliency maps were combined to obtain the final centers.

segmented from $I_{bg-dark}$. For the final detection of the OD and the fovea, various saliency maps were computed using the previous obtained images and were combined to obtain the coordinates of their centers. All these stages are detailed in the following subsections.

4.3.1 Background extraction

Fundus images can be decomposed into a background, referring to the retinal tissue, and a foreground, which covers the retinal structures and the visible lesions (Foracchia et al., 2005). This foreground can be divided into dark and bright pixels. In this stage, the aim was to estimate the background of the fundus image, I_{bg}, as if it was free of any vascular structures or visible lesions. Additionally, we obtained an estimated background where only the bright pixels were preserved (I_{bg-bri}) and another version where only the dark pixels remained ($I_{bg-dark}$).

First, we detected the dark pixels in I_{prep} by applying a multiscale algorithm to obtain the image I_{dark}. For this task, a two-step process was carried out (Romero Oraá et al., 2018):

1. The Alternating Sequential Filtering (ASF) method was applied on I_{prep} to roughly remove all of the dark pixels from the image (mainly blood vessels and RLs) (Zhang et al., 2014). As a result, the image I_{asf} was obtained. Next, a multiscale operation was applied over the three color channels of I_{asf} to obtain $I_{bri_{init}}$, an initial estimate of the bright pixels (Romero Oraá et al., 2018):

$$I_{bri_{init}} = \max_{s} \left(\alpha_s \big(I_{asf} - I_{bg_{asf}(s)} \big) \right), \tag{4.8}$$

where the parameter s represents a scale that depends on D (FOV diameter). In this study, $s = \left\{ \frac{D}{48}, \frac{D}{24}, \frac{D}{12}, \frac{D}{6}, \frac{D}{3} \right\}$. The parameter $I_{bg_{asf}(s)}$ is the background of I_{asf}, estimated with a mean filter of size s. Finally, the parameter α_s was empirically calculated as (Romero Oraá et al., 2018):

$$\alpha_s = 1 - \left(3.84 \frac{s}{D} \right). \tag{4.9}$$

Image $I_{bri_{init}}$ was used in the next step to prevent the pixels surrounding bright structures from being considered as dark regions.

2. Dark pixels in the image were detected (I_{dark}) (Romero Oraá et al., 2018). This image represents the level of darkness of the pixels with respect to the background of the retina. Similarly to the previous step, we calculated I_{dark}

as (Romero Oraá et al., 2018):

$$I_{dark} = \max_s \Big(\alpha_s \big((I_{prep} - I_{bri_{init}}) - I_{bg_{asf}(s)} \big) \Big).$$ (4.10)

The term $I_{prep} - I_{bri}$ prevents false detections of dark pixels at the edges of the bright structures. We considered all three color channels avoiding losing information. This highlighted the color difference of the dark pixels in contrast to the background.

After detecting the dark pixels, we detected the bright pixels in I_{prep} (I_{bri}) by applying an adaption of the same multiscale operation:

$$I_{bri} = \max_s \alpha_s (I_{prep} + I_{dark} - I_{bg_{dark}(s)}).$$ (4.11)

The parameter s represents a scale dependent on D (FOV diameter). It was empirically settled to $s = \left\{ \frac{D}{48}, \frac{D}{24}, \frac{D}{12}, \frac{D}{6}, \frac{D}{3} \right\}$ (Romero-Oraá et al., 2019). The parameter $I_{bg_{dark}(s)}$ is the background of I_{dark}, estimated with a mean filter of size s. Finally, the parameter α_s was defined in equation 4.9.

The next step was to binarize the high intensity pixels (peak pixels) in the images I_{dark} and I_{bri}. The thresholds were empirically set by inspecting multiple images from the training set to guarantee that all peak pixels were selected:

$$M_{dark} = \begin{cases} 0, & I_{dark} < 0.005 \\ 1, & I_{dark} \geq 0.005 \end{cases}.$$ (4.12)

$$M_{bri} = \begin{cases} 0, & I_{bri} < 0.01 \\ 1, & I_{bri} \geq 0.01 \end{cases}.$$ (4.13)

Finally, $I_{bg-dark}$ was calculated. First, we eliminated the pixels in I_{prep} that matched M_{bri} and we filled them using a region-growing algorithm. This algorithm iteratively dilated the image r pixels in order to fill the eliminated pixels with new values. To calculate the value for each new filled pixel, we computed the average value of its surrounding pixels in a neighborhood of $w \times w$ pixels. In this work, r was in the range $[\frac{D}{1000}, \frac{D}{8}]$ and w in the range $[\frac{D}{1000}, \frac{D}{4}]$. Both r and w grew by an empirical factor of 1.3 in each iteration. The increase in the value of r allowed us to accelerate the algorithm significantly. The increase of w allowed us to soften the background estimate as the growing region progressed. Dilation was performed iteratively until all the eliminated pixels had been filled.

The image I_{bg-bri} was obtained in a similar way. We eliminated the pixels in I_{prep} that matched M_{dark} and we filled them by applying the region-growing algorithm described above.

Finally, the image I_{bg} was obtained following the same idea. We eliminated the pixels in I_{prep} that matched both M_{dark} and M_{bri}. Then, we filled them by applying the region-growing algorithm previously explained.

An example of all three images obtained in the background extraction stage is exposed in Figure 4.4. In $I_{bg-dark}$, the bright structures are smoothly removed from I_{prep} while preserving the dark structures. The image I_{bg-bri} also adequately estimates the retinal background ignoring the dark structures while preserving the bright structures. Image I_{bg} represents the background of the fundus image removing both bright and dark structures. Obtaining these images was useful for the computation of subsequent saliency maps. The image $I_{bg-dark}$ was also useful for blood vessel segmentation.

4.3.2 Blood vessel segmentation

The vasculature is an important reference landmark in fundus images. The orientation and distribution of the blood vessels over the retina is useful to detect the OD and the fovea (Niemeijer et al., 2009). The aim of this stage was to detect the pixels associated with the vascular network in the fundus image, M_{vess}. For this task, we developed an improved version of (Mendonça et al., 2006). The original method in (Mendonça et al., 2006) was based on the extraction of vessel centerlines and the application of region growing and morphological filters. When using the original method, the edges of the exudates and other bright regions were frequently detected as blood vessel segments (Figure 4.5b). In order to decrease false detections, we used the image $I_{bg-dark}$, which lacks bright structures, as input image for he method. This allowed us to improve the segmentation notably, as shown in Figure 4.5c.

4.3.3 Optic disc location

The bright appearance of the OD often hinders the segmentation of certain abnormalities, such as EXs in a DR context (Niemeijer et al., 2009). Therefore, the location of the OD in this Doctoral Thesis is useful for the subsequent detection of EXs exposed in section 4.5. Moreover, monitoring other retinal conditions such as the glaucoma relies on the evolution of the OD area (Niemeijer et al., 2009). Thus, OD detection may have additional applications by itself. For these reasons,
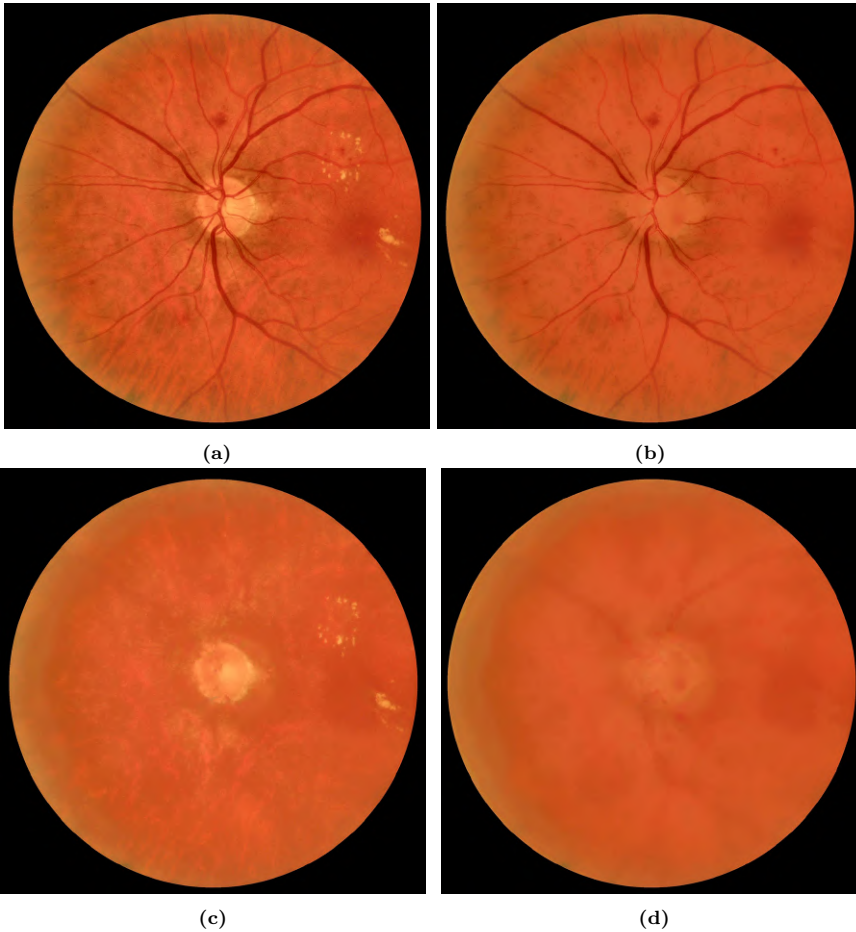
Figure 4.4: Background extraction stage. (a) Preprocessed image within the FOV. (b) Estimated retinal background preserving dark structures ($I_{bg-dark}$). (c) Estimated retinal background preserving bright structures (I_{bg-bri}). (d) Estimated retinal background removing both bright and dark structures (I_{bg}).

the location of the OD may be useful for further image analysis. In this stage, the OD center was automatically detected and the OD was modeled as a circle with radius $R_{OD} = \frac{D}{12}$ pixels (Hsiao et al., 2012).

Since the main blood vessels emerge almost vertically from the OD, this structure is always located around the area where more vertical vessels are found (Sánchez et al., 2008). In order to represent this region, we first applied a morphological opening over the vessel mask (M_{vess}) using a linear morphological operator with vertical orientation. The size of this operator was empirically set to $\frac{D}{50}$ pixels
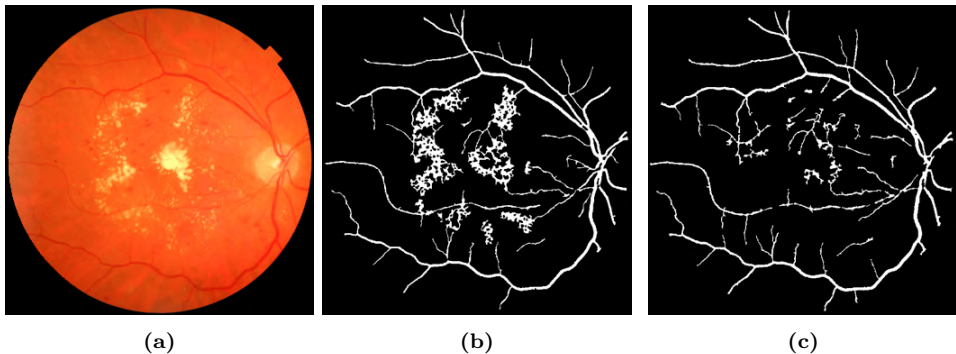
Figure 4.5: Blood vessel segmentation stage. (a) Original image. (b) Segmented mask using I_{prep} as input. The edges of the bright lesions are incorrectly detected as blood vessel segments. (c) Segmented mask using $I_{bg-dark}$ as input. The edges of the bright lesions are not detected as part of the vasculature.

with the purpose of selecting the main segments of the vertical vessels. Next, a rectangular shaped spatial filter was applied. A width of R_{OD} pixels and a height of $4R_{OD}$ pixels were empirically determined for this filter to ensure that the main arcades were covered. The resulting image (I_{vvr}) can be seen in Figure 4.6b. The next step was based on template matching. We computed the correlation between the image I_{bg-bri} and a circle with radius R_{OD} used as template, obtaining the color image $I_{corr-OD}$ (Figure 4.6c). After this, we noticed that the red channel of I_{prep} frequently showed a high local standard deviation where the OD was located. Therefore, we also applied a local standard deviation filter using a disk element with radius R_{OD} over the red channel of I_{prep} to obtain the image I_{std} (Figure 4.6d). This radius was empirically chosen to cover the area of the OD.

In order to combine the previous saliency maps, we computed the following probability map:

$$I_{OD} = I_{vvr} \cdot I_{corr-OD}^{G} \cdot I_{corr-OD}^{B} \cdot I_{std}, \qquad (4.14)$$

where the symbol "·" represents the element-wise multiplication. The parameters $I_{corr-OD}^{G}$ and $I_{corr-OD}^{B}$ correspond to the green and blue channels of the image $I_{corr-OD}$, respectively (Figure 4.6e).

Finally, we binarized the 1%-top pixel values in I_{OD} (see Figure 4.6f). This percentage was empirically obtained. In most cases, these pixels formed a single connected component. However, when more than one connected component was detected, we selected the one with the largest area. The definitive center of the OD was then the centroid of that region (Figure 4.6g).

(a)   **(b)**   **(c)**

(d)   **(e)**   **(f)**   **(g)**
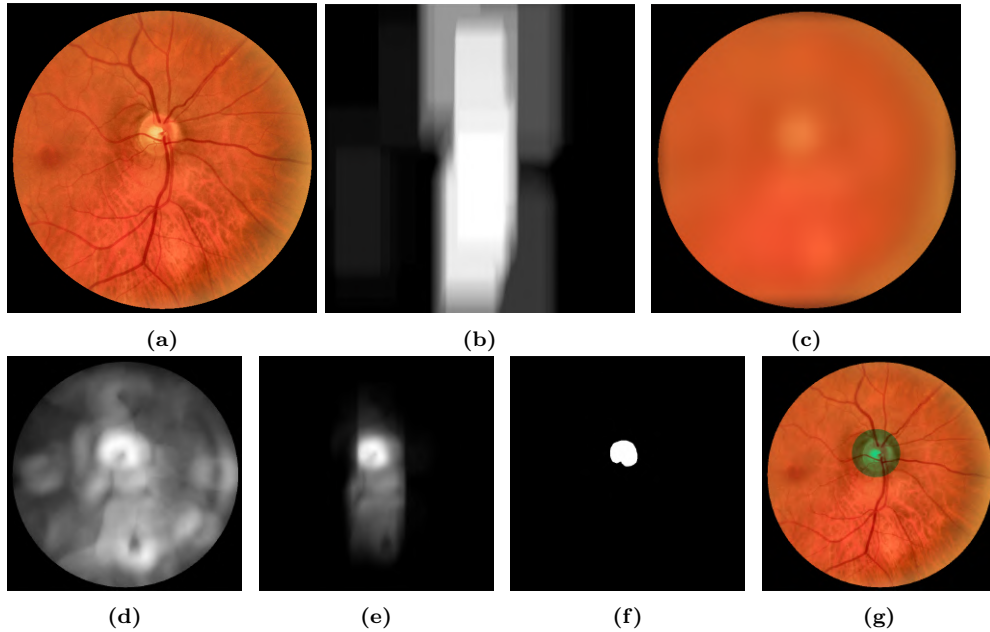
Figure 4.6: Saliency maps for the detection of the OD. (a) Preprocessed image I_{prep}. (b) Image I_{vvr}, representing the area around the vertical blood vessels. (c) Image $I_{corr-OD}$, representing the correlation between the image I_{bg-bri} and a circle. (d) Image I_{std}, the local standard deviation over the red channel of I_{prep}. (e) Image I_{OD} as a combination of the previous saliency maps. (f) Brightest pixels in I_{OD}. (g) Centroid of the region I_{OD}: final location of the OD. The complete OD was approximated by a circle.

4.3.4   Fovea location

The appearance of the fovea is dark with respect to the retinal background, which may hinder the detection of RLs. Furthermore, the lesions near the fovea are especially critical (Niemeijer et al., 2009). For these reasons, fovea location is useful for the diagnosis of DR. In this method, the center of the fovea was automatically detected and was used for the subsequent detection of RLs explained in section 4.4. The macular area, whose center is the fovea, is characterized by the absence of blood vessels. For this reason, we started detecting the areas without vaculature in the fundus. For this task, we first dilated the vessel segmentation (M_{vess}) using a disk with a small radius of $\frac{R_{OD}}{50}$ pixels to make it slightly thicker. Second, we applied a disk-shaped spatial filter with a radius of $\frac{2}{3}R_{OD}$ pixels. This size allowed us to cover the vicinity of the vasculature without reaching the fovea. Third, the complement of the processed image was computed to obtain the image I_{vn}, which

represents the avascular areas of the image (Figure 4.7a).

The next step was based on template matching (Gonzalez and Woods, 2009). Since the fovea is located at the center of a darker area, we first subtracted the images I_{bg} and $I_{bg-dark}$ to get the dark pixels in $I_{bg-dark}$. Second, we calculated the correlation between the image resulting from this subtraction and a disk with radius $\frac{R_{OD}}{2}$, used as the template. This way, we estimated that the radius of the fovea is approximately half the radius of the OD. Then, we obtained the image I_{corr-f} as the maximum, computed pixel by pixel, of the red and green channels of the correlation image (Figure 4.7b).

Next, we considered the fact that the fovea is always located at an approximately constant distance from the OD (Niemeijer et al., 2009). After analyzing the training dataset, we noticed that the fovea is always located between three and seven times R_{OD} from the OD. Therefore, we built a ring centered at the OD center, with the minor radius $3R_{OD}$ pixels and the major radius $7R_{OD}$ pixels (M_{ring}), which can be observed in Figure 4.7c.
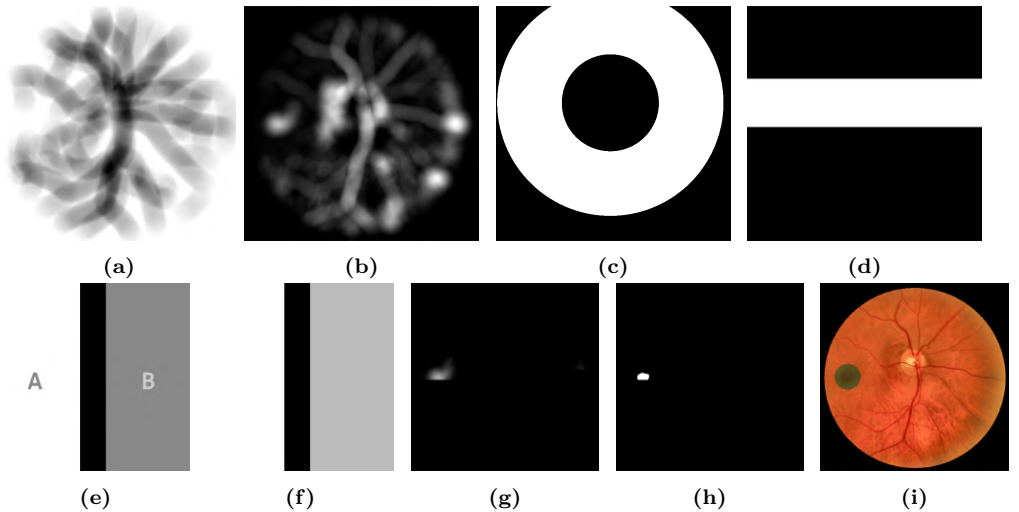
Figure 4.7: Saliency maps for the detection of the fovea. (a) Image I_{vn}, where the white pixels represent the avascular areas. (b) Image I_{corr-f}, the correlation of $(I_{bg} - I_{bg-dark})$ with a disk. (c) Image M_{ring}, representing a ring centered at the OD center. (d) Image M_{band}, representing a horizontal band vertically aligned with the OD center. (e) Image I_{den-1}, blood vessel density within M_{band}. (f) Image I_{den-2}, blood vessel density outside M_{band}. (g) Image I_{FOVEA}, combination of the previous saliency maps. (h) Brightest pixels in I_{FOVEA}. (i) Centroid of the region I_{FOVEA}: final location of the fovea. The complete fovea was approximated by a circle.

The next saliency map was obtained assuming that the fovea is always approximately located at the same vertical level as the OD in the fundus images. Accordingly, we built another mask, M_{band}, covering a horizontal band with $3R_{OD}$ pixels high and vertically aligned with the center of the OD (Figure 4.7d). This estimated height was set by analyzing the training dataset.

The subsequent step of the fovea location method allowed us to detect the temporal side (where the fovea is), based on the density of blood vessels. For this task, the image was divided into two regions, $R = A, B$, with a vertical separation in the center of the OD (Figure 4.7e). The pixel values for both regions were set as the blood vessel density within M_{band}, computed as:

$$VD_{in-band_R} = \frac{M_{vess} \times M_{in-band_R}}{Area_R}, R = A, B, \quad\quad (4.15)$$

leting $M_{in-band_R}$ be the pixels of M_{band} in the region R and $Area_R$ the area of the region R. In this way, we obtained the image I_{den-1} (Figure 4.7e). We then computed the blood vessel density outside M_{band} as:

$$VD_{out-band_R} = \frac{M_{vess} \times M_{out-band_R}}{Area_R}, R = A, B, \quad\quad (4.16)$$

being $M_{out-band_R}$ the pixels outside M_{band} in the region R. For each part of the divided image (A/B), we assigned all of its pixels the value $VD_{out-band_R}$. Thus, the image I_{den-2} was obtained (Figure 4.7f).

Finally, all of the previous complementary information was combined as:

$$I_{FOVEA} = I_{vn} \cdot I_{corr-f} \cdot M_{ring} \cdot M_{band} \cdot I_{den-1} \cdot I_{den-2}, \quad\quad (4.17)$$

where the symbol "·" refers to the element-wise multiplication. The image I_{FOVEA} can be seen as a probability map (see Figure 4.7g). We selected the 0.2% of the pixels in I_{FOVEA} with the highest value (see Figure 4.7h). This percentage was empirically obtained. In most cases, those pixels formed a single connected component. However, when more than one connected component was detected, we selected the biggest one. The centroid of that region was considered the center of the fovea (see Figure 4.7i).

4.4   Detection of red lesions

RLs, such as microaneurysms and hemorrhages, are among the first signs of DR. Consequently, RL detection is an important step in the automated screening of DR (Niemeijer et al., 2005). In the present Doctoral Thesis, two different methods were proposed for RL segmentation as part of the compendium of publications. The first method (Romero-Oraá et al., 2019) was based on superpixel classification and is described in section 4.4.1, while the second method (Romero-Oraá et al., 2020), which offers an improved performance, was based on the decomposition of the fundus image into layers and is detailed in section 4.4.2.

4.4.1   Method based on the classification of superpixels

The proposed method was published in (Romero-Oraá et al., 2019) and comprises three sequential stages, as shown in Figure 4.8. First, the image was preprocessed using the method described in section 4.2. Second, we segmented the candidate regions to be RLs using superpixel segmentation. Third, we used an MLP neural network to classify the true RLs. These last two stages are explained in the following subsections.

Candidate Segmentation

The aim of this stage was to segment the candidate regions to be RLs. Since these lesions appear as dark regions in contrast with the background, every dark region in the fundus image was considered an RL candidate in this work (Niemeijer et al., 2005). Then, the segmentation of the dark regions was performed in three steps. First, the dark pixels in the image were detected. Second, all pixels of the image were grouped in superpixels. Finally, the candidate regions were computed as the generated superpixels belonging to the detected dark pixels.

● **Dark Pixel Detection**

The detection of the dark pixels was carried out using the multiscale algorithm proposed for retinal background extraction in section 4.3.1. The resulting image (I_{dark}) highlighted the dark structures of the retina, such as the vasculature, the fovea and the RLs, as can be seen in Figure 4.9b.
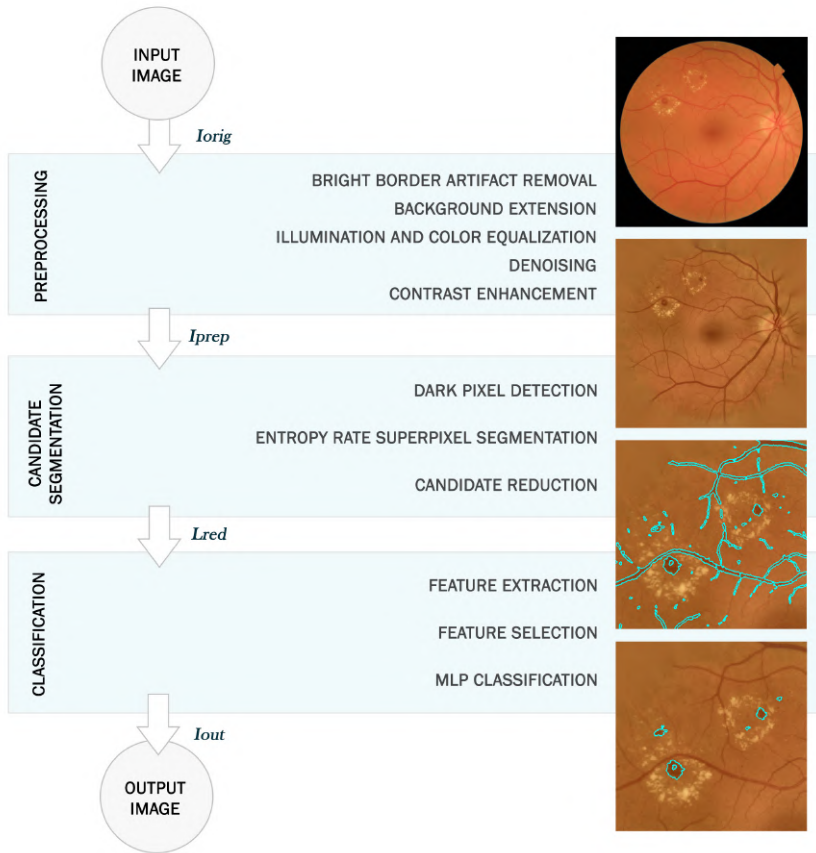
Figure 4.8: Block diagram of the proposed approach based on superpixel classification to detect red lesions.

.

• **Entropy Rate Superpixel Segmentation**

The next step was to group the pixels of I_{dark} into superpixels that separated the different elements of the image. A superpixel is a perceptually uniform region in the image. It groups pixels with similar color and texture and adapts to the image borders (Zhou et al., 2017a). Since it represents natural entities in the image, the superpixel is an appropriate region from which to extract features (Zhou et al., 2017a). In addition, it reduces the complexity of subsequent image processing tasks and improves the computational efficiency (Achanta et al., 2012; Liu et al., 2011).

With this purpose, we applied the ERS method, which is regarded as a clustering problem solved using graph partitioning (Liu et al., 2011). First, an undirected
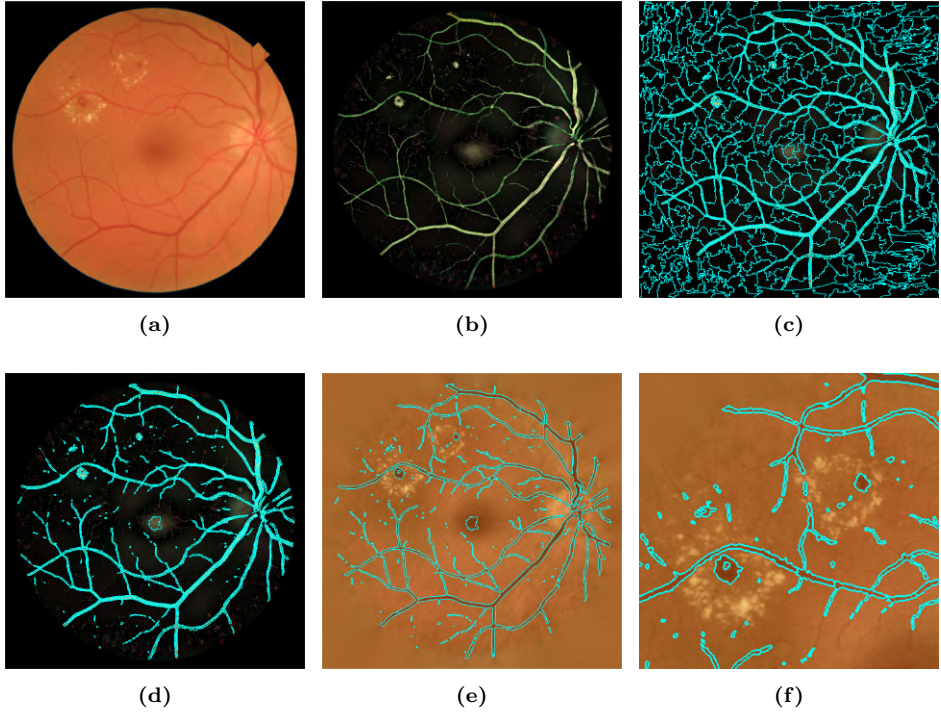
Figure 4.9: Candidate segmentation. (**a**) original image with some RLs; (**b**) dark pixel detection computed using the multiscale algorithm; (**c**) segmented superpixels using the Entropy Rate Superpixel method; (**d**) reduced candidates on I_{dark}; (**e**) combined, reduced candidates shown over I_{prep}; (**f**) zoom of combined, reduced candidates shown over I_{prep}.

graph was constructed mapping its vertices ($V = \{v_1, v_2, ...\}$) to the pixels of I_{dark}. The weights (w) of the edges ($E = \{e_1, e_2, ...\}$) of the graph were the pairwise similarities between vertices. They were calculated using a Gaussian kernel as (Liu et al., 2011):

$$w_{i,j} = e^{-\frac{d(v_i, v_j)^2}{2\sigma}}, \qquad (4.18)$$

where σ is the user-defined Gaussian kernel bandwidth and $d(v_i, v_j)$ is the intensity difference between pixels v_i and v_j multiplied by the spatial distance between them. Since the graph was undirected, the edge weights were symmetrical ($w_{i,j} = w_{j,i}$) (Liu et al., 2011). In addition, every vertex of the graph had a self loop. Then, a subset of edges $A \subseteq E$ was selected such that the resulting graph contained exactly K connected disjoint subgraphs $\{S_1, S_2, ..., S_K\}$. These subgraphs must also satisfy that $S_i \cap S_j = \varnothing$ for $i \neq j$ and $\cup_i S_i = V$ (Liu et al., 2011). For the non-selected edges, the edge weight of the self loop of the associated vertices was

increased so that the total incident weight for each vertex remained constant (Liu et al., 2011).

Thus, the graph partitioning problem results in the selection of a subset of edges A that represents the superpixels in the image. The sequential selection of edges was based on the random walk of a particle through the graph (Cover and Thomas, 1991). First, the particle is positioned at a random node of the graph. Second, the particle randomly moves from node to node forming a sequence of vertices (Cover and Thomas, 1991). At every step, the next vertex is chosen from among the nodes connected to the last vertex of the sequence with a probability proportional to the weight of the edge (Cover and Thomas, 1991). In this way, we searched for the graph topology that maximized the following objective function (Cover and Thomas, 1991):

$$\max_{A} H(A) + \gamma B(A), \tag{4.19}$$

subject to $A \subseteq E$ and $N_A \geq K$, being N_A the number of connected vertices in the graph. The parameter $H(A)$ is the entropy rate, defined as (Liu et al., 2011):

$$H(A) = -\sum_{i} \mu_i \sum_{j} p_{i,j}(A) log(p_{i,j}(A)), \tag{4.20}$$

where μ_i is (Liu et al., 2011):

$$\mu_i = \frac{w_i}{\sum_{i=1}^{|V|} w_i}, \tag{4.21}$$

and $p_{i,j}$ is the transition probability (Liu et al., 2011):

$$p_{i,j}(A) = \begin{cases} \frac{w_{i,j}}{w_i}, & \text{if } i \neq j \text{ and } e_{i,j} \in A, \\ 0, & \text{if } i \neq j \text{ and } e_{i,j} \notin A, \\ 1 - \frac{\sum_{j:e_{i,j} \in A} w_{i,j}}{w_i}, & \text{if } i = j, \end{cases} \tag{4.22}$$

where w_i is the sum of incident weights to the vertex v_i.

Additionally, the parameter $B(A)$ in equation 4.19 is called the balancing function and is defined as (Liu et al., 2011):

$$B(A) = -\sum_{j} p_{Z_A}(i) log(p_{Z_A}(i)) - N_A, \tag{4.23}$$

where p_{Z_A} is a probability distribution defined as (Liu et al., 2011):

$$p_{Z_A}(i) = \frac{|S_i|}{|V|}, i = \{1, ..., N_A\}. \tag{4.24}$$

The remaining parameter in equation 4.19 is the weight of the balancing term, $\gamma \geq 0$, given by (Liu et al., 2011):

$$\gamma = \beta K \lambda, \tag{4.25}$$

where β is defined as (Liu et al., 2011):

$$\beta = \frac{\max_{e_{i,j}} H(e_{i,j}) - H(\varnothing)}{\max_{e_{i,j}} B(e_{i,j}) - B(\varnothing)}. \tag{4.26}$$

In this equation, the difference $H(e_{i,j}) - H(\varnothing)$ refers to the entropy rate increase when including a single edge into the graph. Likewise, the difference $B(e_{i,j}) - B(\varnothing)$ refers to the balancing term increase when including a single edge into the graph. Therefore, β is the ratio of the maximal entropy rate increase and the maximal balancing term increase upon including a single edge into the graph.

It should be noted that the parameter $H(A)$ favors the formation of compact and homogeneous clusters. It makes the superpixels adjust to the edges of the structures, resulting in the division of images on perceptual boundaries (Liu et al., 2011). However, in most cases, it induces several superpixels to overlap with a single object. To overcome this issue, $B(A)$ is introduced. This parameter encourages the formation of clusters (superpixels) with similar sizes (Liu et al., 2011). Another parameter that should be taken into account is K, which favors the balancing term (Liu et al., 2011). Additionally, β compensates for the magnitude difference $H(A)$ and $B(A)$ (Liu et al., 2011). Finally, λ is a user-specified constant and should be empirically determined (Liu et al., 2011). Applying ERS to fundus images requires properly adjusting the parameters K, λ and σ. K should be large enough to isolate every structure, paying special attention to the smallest RLs. Meanwhile, the parameter σ should be selected to ensure that the superpixels adjust robustly to the edges of the objects in the image. λ should compensate the fragmentation effect to avoid over-segmentation within retinal structures. In this study, these parameters were experimentally adjusted using the images in the training set. The result of this stage is an image map, L, of K labels that identify the superpixels that compose the image. Figure 4.10 provides the superpixel segmentation results with different values for each parameter. As shown in
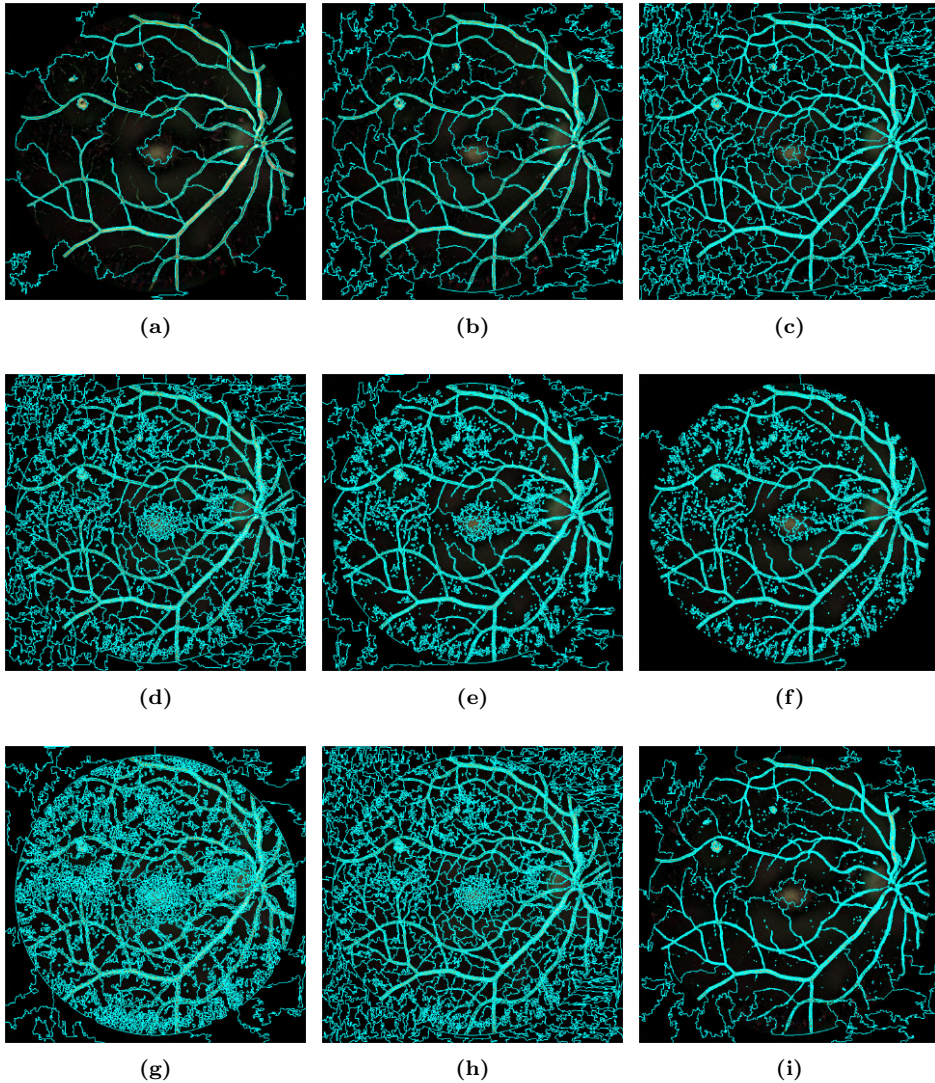
Figure 4.10: Entropy Rate Superpixel Segmentation with different parameter values.
(**a**) $K = 100$, $\lambda = 0.1$ and $\sigma = 5$; (**b**) $K = 500$, $\lambda = 0.1$ and $\sigma = 5$; (**c**) $K = 2000$, $\lambda = 0.1$ and $\sigma = 5$; (**d**) $K = 2000$, $\lambda = 0.8$ and $\sigma = 2$; (**e**) $K = 2000$, $\lambda = 0.01$ and $\sigma = 2$; (**f**) $K = 2000$, $\lambda = 0.001$ and $\sigma = 2$; (**g**) $K = 2000$, $\lambda = 0.1$ and $\sigma = 0.5$; (**h**) $K = 2000$, $\lambda = 0.1$ and $\sigma = 2$; (**i**) $K = 2000$, $\lambda = 0.1$ and $\sigma = 5$.

Figure 4.10a, we properly separated the large HEs with $K = 100$. However, we missed all the MAs. With $K = 500$, as shown in Figure 4.10b, we did not detect the smallest MAs yet. Therefore, a larger K should be selected to detect all the RLs, including MAs. By visual inspection of our results for different values of K (see Figure 4.10a–c), we chose $K = 2000$ as adequate to detect every RL. Regarding the value of λ, we encountered that a large value of this parameter prevented sizeable structures to be over-segmented in tiny regions (see Figure 4.10d). On the contrary, the smallest MAs were more easily detected when λ was small (see Figure 4.10e,f). Therefore, we chose $\lambda = 0.08$ as an optimal commitment value. Regarding σ, we found that, when this parameter was too large, the smallest MAs could not be adequately segmented (see Figure 4.10i). On the other hand, if σ was too small, the superpixels did not adjust accurately to the edges of the structures. We experimentally found that $\sigma = 2$ provided a good compromise between both situations (see Figure 4.10h).

In conclusion, to properly separate the retinal RLs, we considered the following parameter values: $K = 2000$, $\lambda = 0.08$ and $\sigma = 2$. The result of this stage on the example image can be seen in Figure 4.9c.

• Candidate Reduction

All superpixels in which the image was divided were initially considered RL candidates. However, most of them did not correspond to dark regions. Instead, these superpixels covered black pixels in I_{dark} and should be removed as RL candidates. Reducing the number of lesion candidates simplifies the classification task. With this purpose, we eliminated the superpixels in L that were not dark, thus obtaining L_{red}. For this task, the average value of the pixels within each superpixel was computed on the green channel of the image I_{dark} and those that did not exceed a threshold equal to 0.3 were removed (Romero Oraá et al., 2018). This threshold was empirically obtained considering that lower values corresponded to non-RL structures. An accurate value for this parameter was not crucial since our objective was to coarsely discard the superpixels in the image whose pixel intensities did not correspond to dark structures. Figure 4.9d shows L_{red} over the I_{dark} image.

Finally, neighboring superpixels with a similar color were combined in an iterative process. To quantify color similarity among the superpixels, we calculated the average value of the pixels contained in each superpixel in the image I_{dark}. Then, the Euclidean distances between the average values were measured over the CIELAB color space (Romero Oraá et al., 2018). This color space was designed to approximate human vision, representing the nonlinear response of the eye. There-

fore, the relative perceptual difference between any two colors defined in CIELAB can be approximated taking the Euclidean distance between them (Jain, 1989). In this study, the maximum color distance for two neighboring superpixels to be combined was settled to 0.24 during training. We empirically noticed that using higher values for this parameter may combine superpixels belonging to different retinal structures. Figure 4.9e shows the definitive RL candidates after this stage.

The following pseudocode summarizes the proposed segmentation technique:

Algorithm 2: Candidate segmentation technique

---

1 $I_{asf} = \text{ASF}(I_{prep})$;

2 $I_{bri} = 0$;

3 **for** $s = \left\{ \frac{D}{48}, \frac{D}{24}, \frac{D}{12}, \frac{D}{6}, \frac{D}{3} \right\}$ **do**

4 $\quad$ $I_{bg_{asf}}(s) = \text{meanFilter}(I_{asf}, \text{size}=s)$;

5 $\quad$ $\alpha_s = 1 - \left(3.84 \frac{s}{D} \right)$;

6 $\quad$ $I_{bri} = \max \left(I_{bri}, \alpha_s(I_{asf} - I_{bg_{asf}}(s)) \right)$;

7 **end**

8 $I_{dark} = 0$;

9 **for** $s = \left\{ \frac{D}{48}, \frac{D}{24}, \frac{D}{12}, \frac{D}{6}, \frac{D}{3} \right\}$ **do**

10 $\quad$ $I_{bg_{asf}}(s) = \text{meanFilter}(I_{asf}, \text{size}=s)$;

11 $\quad$ $\alpha_s = 1 - \left(3.84 \frac{s}{D} \right)$;

12 $\quad$ $I_{dark} = \max \left(I_{dark}, \alpha_s((I_{prep} - I_{bri}) - I_{bg_{asf}}(s)) \right)$;

13 **end**

14 $L = \text{entropyRateSuperpixelSegmentation}(\ I_{dark},\ K,\ \lambda,\ \sigma\ )$;

15 $G_{dark} = \text{greenChannel}(I_{dark})$;

16 $L_{red} = L$ where $G_{dark} >$ threshold;

17 $L_{cand} = \text{colorSimilarityGrouping}(\ L_{red},\ \text{color\_distance}\ )$;

Classification

Once the RL superpixel candidates were obtained, we used an MLP to separate the true RLs from non-RL candidate superpixels. This type of NN has been used in previous studies for the automatic detection of RLs (García et al., 2008, 2010; Usher et al., 2004). The classification stage comprises three procedures:

• **Feature Extraction**

Superpixel candidates were required to be differentiated among RLs and any other dark regions in the image, such as blood vessel segments. Even though the detection of the vasculature presented in section 4.3.2 could serve to previously remove the candidates corresponding to blood vessel fragments, in this study we decided to maximize the number of true RLs reaching the classification stage by avoiding the possible errors derived from this phase. For each superpixel, a set of 39 features was extracted to represent the visual characteristics of RLs. These features include measures related to the shape of the region, the value of the pixels inside the region and the distance to the OD and the fovea. The extracted features are specified in Table 4.1.

• **Feature Selection**

Some of the extracted features may be correlated or irrelevant for the given problem. Therefore, an automated feature selection stage was implemented to filter the optimal features. Reducing the number of features decreases classification errors and simplifies the structure of the classifier (Bishop, 1995). In this work, we applied the FCBF method described in section 4.1.1 (Yu and Liu, 2004). It is a classifier-independent feature selection technique based on the SU to find the most relevant and non-redundant features (Yu and Liu, 2004). The features selected from this stage are indicated in the last column of Table 4.1.

• **MultiLayer Perceptron Neural Network**

Supervised learning was applied in order to separate true RLs from non-RLs. A three-layer MLP neural network was trained using the features previously selected. This type of classifier was explained in section 4.1.1. The input layer consisted of as many neurons as features were selected by the FCBF method. The output layer was formed by a single neuron since we tried to resolve a dichotomous problem. The number of hidden units, n_{hid}, was experimentally settled during the training stage. The activation function used in the hidden layer was the hyperbolic tangent sigmoid ($tanh$) (Bishop, 1995), since it has been successfully applied in similar MLP networks from previous studies (García et al., 2009). The logistic sigmoid was used as the activation function in the output neuron. Since it is defined in the range [0-1], we could interpret the output of the network as a posterior probability (Bishop, 1995; Haykin, 1999).

Table 4.1: Extracted features for the RL segmentation method based on superpixels. The last column indicate the selected features.

Feature Number	Description	Selected
1	Area of the region.	1
2	Width of the bounding box (smallest rectangle containing the region).	-
3	Heigh of the bounding box.	-
4	Area of the smallest convex hull (smallest convex polygon that can contain the region).	-
5	Eccentricity of the ellipse that has the same second-moments as the region.	5
6	Number of holes in the region.	6
7	Ratio of pixels in the region to pixels in the total bounding box.	7
8	Length of the major axis of the ellipse that with same normalized second central moments as the region.	8
9	Length of the minor axis of the ellipse that with same normalized second central moments as the region.	-
10	Distance around the boundary of the region (perimeter length).	-
11	Proportion of the pixels in the convex hull that are also in the region (solidity).	11
12–14	Mean of the pixels inside the region computed in the RGB channels of the image I_{prep}.	12,13
15–17	Median of the pixels inside the region computed in the RGB channels of the image I_{prep}.	-
18–20	Standard deviation of the pixels inside the region computed in the RGB channels of the image I_{prep}.	19
21–23	Entropy of the pixels inside the region computed in the RGB channels of the image I_{prep}.	21
24–26	Mean of the pixels inside the region computed in the RGB channels of the image I_{dark}.	-
27–29	Median of the pixels inside the region computed in the RGB channels of the image I_{dark}.	27,28
30–32	Standard deviation of the pixels inside the region computed in the RGB channels of the image I_{dark}.	-
33–35	Entropy of the pixels inside the region computed in the RGB channels of the image I_{dark}.	-
36	Mean of the pixels calculated in the border of the region applying Prewitt operator in the image I_{prep} (Gonzalez and Woods, 2009).	36
37	Mean of the pixels inside the region calculated in the result of applying multiscale line operator filters (Nguyen et al., 2013).	37
38	Distance to the center of the optic disc, calculated using (Romero Oraá et al., 2016).	38
39	Distance to the center of the fovea, calculated using (Romero Oraá et al., 2016).	39

We chose the cross-entropy error function since it simplifies the optimization process when the output layer activation function is the logistic sigmoid (García et al., 2010). Regarding the learning function used to update the weight and bias values, we applied the scaled conjugate gradient backpropagation method. This algorithm has been shown considerably faster than other supervised learning algorithms (Moller, 1993).

In order to avoid overfitting and improve generalization, we applied weight decay regularization by penalizing large weights during the learning process (Bishop, 1995). The regularization parameter, λ, was experimentally determined during training.

4.4.2 Method based on the decomposition of the fundus image into layers

This method was published in (Romero-Oraá et al., 2020) and was aimed at the joint detection of RLs and EXs. However, this section focuses on the first type of lesion. Figure 4.11 shows an overview of the proposed algorithm. First, we applied the preprocessing stage explained in section 4.2 to obtain the image I_{prep}, a normalized version of the input image where the retinal structures are enhanced. Second, we obtained an estimation of the retinal background, the segmentation of the vasculature, and the location of the OD and the fovea using the algorithms previously described in section 4.3. The rest of the stages are detailed in the next subsections: the image was decomposed into several layers, multiple RL candidates were segmented, various features were extracted using the obtained layers, feature selection was performed and, finally, a classification stage was conducted.

Red lesion candidate segmentation

This stage was aimed at processing the fundus image to segment the potential RL candidate regions and obtain additional data for the later detection of RLs. For this task, the image was decomposed into several layers. Each layer represented a different structure of the retina and provided useful information for the detection of retinal lesions. Since RLs are shown as dark regions, we calculated the complement of the subtraction $I_{bg-dark} - I_{bg}$ in order to select the dark pixels (I_{dark}). In this image, the color difference between the dark pixels and the background was highlighted, while leaving the rest of the pixels black (see Figure 4.12a).

The preprocessing stage in section 4.2 allowed reducing inter- and intra-image variability. This way, the color of the visible structures in I_{dark} looked always
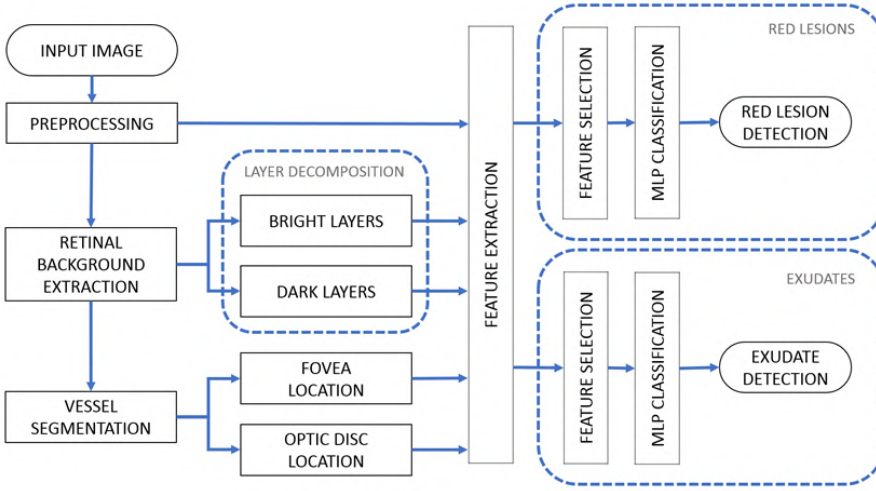
Figure 4.11: Diagram of the proposed method. (1) Preprocessing. (2) Retinal background extraction. (3) Vessel segmentation, optic disc location, and fovea location. (4) Layer decomposition. (5) Feature extraction and selection. (6) Multilayer Multilayer perceptron (MLP) classification

.

(a)   (b)   (c)   (d)
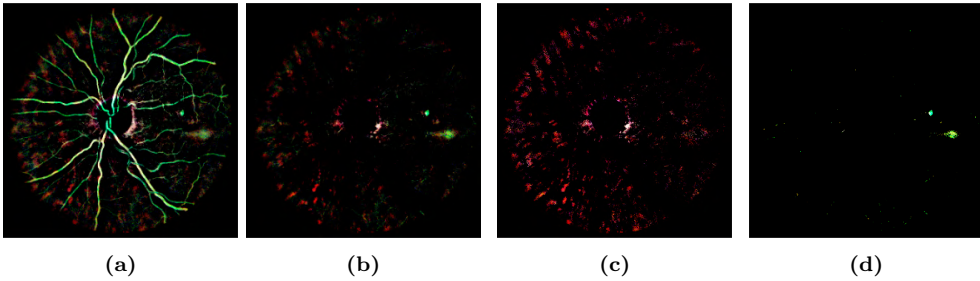
Figure 4.12: Red lesion candidate segmentation. (a) Image I_{dark}. (b) Image I_{dark-2}. (c) Image $L_{chor-dark}$. (d) Image $L_{rl-cand}$. These images are shown with enhanced contrast for an easier readability.

consistent. Therefore, they could be directly separated using color and spatial information. The blood vessels, the choroidal vasculature, and the rest of dark pixels, which are RL candidates, are the structures of interest in I_{dark} that should be taken into account for RL detection.

As expected, the color of the blood vessels in I_{dark} is very similar to the color of RLs. Thus, we eliminated the pixels belonging to the vasculature using the vessel segmentation (M_{vess}) obtained in section 4.3.2. The result, I_{dark-2}, can be seen in Figure 4.12b. There is another type of structure that is easily distinguishable

in I_{dark}: the underlying choroidal vasculature visible in tigroid retinas (Lyu et al., 2017). This is caused by the lack of pigments in the retinal pigment epithelium and is common in aged or myopic patients (Lyu et al., 2017). It has pink tones in I_{dark} and can hinder the detection of RLs due to the high contrast it shows against the background. Separating the choroidal vasculature is useful to classify the RLs in the image, since images featuring very marked choroidal vessels tend to present false positives (Romero-Oraá et al., 2019). For this reason, the next step was to separate the layer corresponding to the choroidal vessels in I_{dark-2}. For this task, we used the color information in the HSV color space. This color space was designed to approximate the eye perception and is useful to replicate human interpretation (Gonzalez and Woods, 2009). It decouples the brightness component from the color-carrying information. Using HSV, we analyzed the pixels belonging to the choroidal vessels using the training set. This way, we empirically selected the ranges of values that represented those pixels for each channel. In this work, these ranges were determined to be $H = [0.75, 0.1]$, $S = [0.2, 1.0]$, and $V = [0.05, 1.0]$. Then, we segmented the pixels in I_{dark-2} that were between the selected ranges, obtaining the image $L_{chor-dark}$, which represented the layer of choroidal vessels (see Figure 4.12c). In the same way, we selected the HSV color ranges associated with RLs. For this task, we analyzed all the pixels belonging to RLs in the images of the training set annotated by the ophthalmologist. The selected ranges in this work were $H = [0.1, 0.45]$, $S = [0.1, 1.0]$, and $V = [0.2, 1.0]$. Then, we segmented the pixels in I_{dark-2} that were between the selected ranges, obtaining the layer $L_{rl-cand}$, associated to potential RLs (see Figure 4.12d). Finally, the mask of potential RL candidates was obtained binarizing the layer $L_{rl-cand}$ in order to obtain the candidate binary mask $M_{rl-cand}$. The obtained layers of interest in this phase, $L_{chor-dark}$ (Figure 4.12c) and $L_{rl-cand}$ (Figure 4.12d), were also useful for lesion candidate classification in later stages.

Red lesion classification

Once the RL candidates were obtained, we used an MLP neural network to classify the RLs from the rest of candidate regions. This type of neural network has been used in previous studies for the automatic detection of RLs, including our method based on superpixel classification (García et al., 2008; Romero-Oraá et al., 2019). This stage comprises three steps:

1. **Feature extraction**. For each region candidate in $M_{lr-cand}$, a set of 100 features was extracted mostly using the decomposed layers previously obtained. This way, we computed from these layers measures related to the shape of the region, the pixel values inside and around the region, and the distance to the OD and the fovea. These whole set of features is specified in Table 4.2.

2. **Feature selection**. Reducing the number of features to a set of relevant, low-correlated ones decreases classification errors and simplifies the structure of the classifier (Bishop, 1995). For this task, the FCBF method was applied (Yu and Liu, 2004). FCBF is classifier-independent and computes SU to find the most relevant and non-redundant features for a certain problem (Yu and Liu, 2004), as explained in section 4.1.1. Among the original 100 features, only 24 were selected and are specified in Table 4.2. Features of different nature were selected, including shape, distance, intensity, and variability around the candidates in different layers.

3. **Multilayer Perceptron Neural Network**. The chosen classifier was an MLP neural network (see section 4.1.1). The number of neurons in the input layer was the number of selected features. We used a single neuron in the output layer, since our problem was dichotomous. The number of hidden units was experimentally optimized during the training stage. The activation function used in the hidden layer was the hyperbolic tangent sigmoid (tanh) (Bishop, 1995) since it has been successfully applied in similar MLP networks from previous studies (García et al., 2009). The logistic sigmoid was used as the activation function in the output neuron. We used the scaled conjugate gradient backpropagation method as the learning function. The cross-entropy was the chosen error function to minimize during training (Bishop, 1995; Romero-Oraá et al., 2019). In addition, we used the regularization parameter weight decay, experimentally optimized during training, to avoid overfitting and improve generalization (Bishop, 1995).

4.5   Detection of hard exudates

Based on our previous experience with the detection of RLs, the approach based on the decomposition of the fundus image into layers outperformed the classification of superpixels. For this reason, although the concept of superpixel is very interesting,

Table 4.2: Extracted features for lesion classification.

Feature Number	Description	Selected for red lesion detection	Selected for exudate detection
1	Area of the region	-	-
2	Width of the bounding box (smallest rectangle containing the region)	-	-
3	Height of the bounding box	-	-
4	Area of the smallest convex hull (smallest convex polygon that can contain the region)	-	-
5	Eccentricity of the ellipse that has the same second moments as the region	5	5
6	Number of holes in the region	-	-
7	Ratio of pixels in region to pixels in the total bounding box	-	7
8	Length of the major axis of the ellipse that has the same normalized second central moments as the region	-	-
9	Length of the minor axis of the ellipse that hast the same normalized second central moments as the region	-	-
10	Distance around the boundary of the region (perimeter length)	-	-
11	Proportion of the pixels in the convex hull that are also in the region (solidity)	11	-
12–14	Mean of the pixels inside the region computed in the Red-Green-Blue (RGB) channels of the image I_{prep}	13	-
15–17	Median of the pixels inside the region computed in the RGB channels of the image I_{prep}	-	17
18–20	Standard deviation of the pixels inside the region computed in the RGB channels of the image I_{prep}	18, 19	18–20
21–23	Entropy of the pixels inside the region computed in the RGB channels of the image I_{prep}	22, 23	21–23
24–26	Mean of the pixels inside the region computed in the Hue-Saturation-Value (HSV) channels of the image $L_{rl-cand}/L_{ex-cand}$	24, 26	26
27–29	Median of the pixels inside the region computed in the HSV channels of the image $L_{rl-cand}/L_{ex-cand}$	28, 29	27, 29
30–32	Standard deviation of the pixels inside the region computed in the HSV channels of the image $L_{rl-cand}/L_{ex-cand}$	32	30, 32
33–35	Entropy of the pixels inside the region computed in the HSV channels of the image $L_{rl-cand}/L_{ex-cand}$	35	34, 35
36–38	Mean of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{rl-cand}$	-	-
39–41	Median of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{rl-cand}$	-	-
42–44	Standard deviation of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{rl-cand}$	44	42
45–47	Entropy of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{rl-cand}$	-	-
48–50	Mean of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{chor-dark}$	-	-
51–53	Median of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{chor-dark}$	-	-
54–56	Standard deviation of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{chor-dark}$	-	-
57–59	Entropy of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{chor-dark}$	59	57
60–62	Mean of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{chor-bri}$	-	62
63–65	Median of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{chor-bri}$	63–65	64, 65
66–68	Standard deviation of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{chor-bri}$	66	-
69–71	Entropy of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{chor-bri}$	-	-
72–74	Mean of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{ex-cand}$	-	73, 74
75–77	Median of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{ex-cand}$	-	-
78–80	Standard deviation of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{ex-cand}$	-	78–80
81–83	Entropy of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image $L_{ex-cand}$	-	83
84–86	Mean of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image L_{bm}	-	-
87–89	Median of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image L_{bm}	-	-
90–81	Standard deviation of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image L_{bm}	90	91
93–95	Entropy of the pixels inside a circle with radius R_{DO} centered on the region computed in the HSV channels of the image L_{bm}	-	93
96	Mean of all the pixels the V channel of the image L_{bm}	96	96
97	Mean of the pixels calculated in the border of the region applying Prewitt operator in the image I_{prep}	97	97
98	Mean of the pixels inside the region calculated in the result of applying multiscale line operator filters	98	98
99	Distance to the center of the optic disc	-	99
100	Distance to the center of the fovea	100	100

we decided to guide the detection of EXs towards the decomposition of the fundus image into layers.

As mentioned before, the study published in (Romero-Oraá et al., 2020) and included in the compendium of publications of this Doctoral Thesis was aimed at the joint detection of RLs and EXs. This section focuses on the part devoted to the detection of EXs. However, we refer again to Figure 4.11 for an overview of the whole algorithm. As indicated in section 4.4.2, the preprocessing stage, the estimation of the retinal background, the segmentation of the vasculature, and the location of the OD and the fovea were explained in sections 4.2 and 4.3. The rest of stages required for EX segmentation are detailed in the next subsections, following the same steps as RL detection: the image was decomposed into several layers, multiple EX candidates were segmented, various features were extracted using the obtained layers, feature selection was performed and, finally, a classification stage was conducted.

Exudate candidate segmentation

EXs appear as bright regions in contrast with the background. Therefore, we subtracted the image I_{bg} from the I_{bg-bri} image in order to select the bright pixels. In the obtained image, I_{bri}, the color difference of the bright pixels with respect to the retinal background was enhanced, while the rest of the pixels remained black (see Figure 4.13a).

Following the same idea as for the segmentation of RLs, I_{bri} was divided into different layers of information. The proposed method relied in the idea that the color of the pixels in I_{bri} for each type of structure is constant. This way, we separated the different structures of interest in I_{bri} using color and spatial information.

(a)   (b)   (c)   (d)
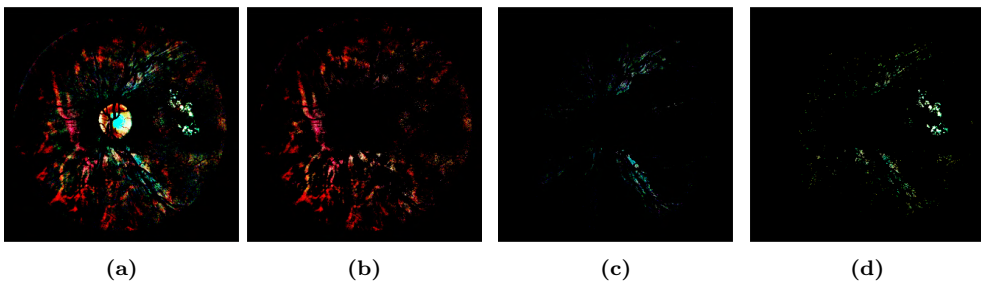
Figure 4.13: Exudate candidate segmentation. (a) Image I_{bri}. (b) Image $L_{chor-bri}$. (c) Image L_{bm}. (d) Image $L_{ex-cand}$. These images are shown with enhanced contrast for an easier readability.

The HSV color space was also employed here (Gonzalez and Woods, 2009). The structures of interest in I_{bri} are the choroidal vasculature, the reflective features, and the EX candidates.

In tigroid retinas, the prominent choroid pigmentation contrasts with the attenuated pigment epithelium showing a spotted or striped appearance (Lyu et al., 2017). This particular appearance, which reflects as red tones in I_{bri} (Figure 4.13a), may make nearby EXs less prominent and, consequently, affect the performance for EX detection. We analyzed the pixels in I_{bri} belonging to choroidal vessels in tigroid retinas using the training set. We empirically selected the ranges in HSV representing those pixels. In this study, those ranges were $H = [0.75, 0.15]$, $S = [0, 1.0]$, and $V = [0, 1.0]$. Then, the layer of choroidal vessels, $L_{chor-bri}$, was obtained segmenting the pixels in I_{bri} that were between the selected ranges (Figure 4.13b). Other structures visible in I_{bri} were the reflective features caused by the nerve fiber layer (Giancardo et al., 2009). They are very common when it comes to retinas in young patients and cannot be considered as abnormalities. Most of these marks are concentrated along the widest vessels (Giancardo et al., 2009) and tend to be green and blue in I_{bri} (Figure 4.13a). Due to their color, they can be confused with EXs. Therefore, separating the reflective features is useful to classify the true EXs. To separate reflective features from lesions, we selected the pixels associated with reflective features using the HSV color space in the training set. The selected ranges were $H = [0.25, 0.85]$, $S = [0, 1.0]$, and $V = [0, 1.0]$, obtaining the image $I_b m1$. Next, we selected the pixels surrounding the main vessels in the vascular network. For this task, we first performed a morphological opening over the image M_{vess} to roughly remove the thin vessels. A disk of radius $R_{OD}/10$ pixels was applied. Second, a morphological dilation was performed using a disk of radius $D/60$ pixels, obtaining the image I_{bm2}. Finally, the reflective features layer, L_{bm}, was obtained by multiplying I_{bm1} and I_{bm2} to select the bright marks surrounding the vasculature (Figure 4.13c). Using the same idea, the layer of potential EXs was also extracted from I_{bri}. We used the HSV color space and, for each component, selected the ranges of values among which the EXs were. For this task, we analyzed all the pixels belonging to EXs in the images of the training set annotated by the ophthalmologist. The selected ranges were $H = [0.15, 0.45]$, $S = [0.1, 1.0]$, and $V = [0.1, 1.0]$. Thus, we obtained the layer $L_e x - cand$ (Figure 4.13d). Finally, to obtain the binary mask of potential EX candidates, $M_{ex-cand}$, we binarized $L_{ex-cand}$. The obtained layers of interest in this phase, $L_{chor-bri}$ (Figure 4.13b), L_{bm} (Figure 4.13c), and $L_{ex-cand}$ (Figure 4.13d), were also useful for candidate classification in later stages.

Exudate classification

After detecting all EX candidate regions, an MLP neural network was used to differentiate the true EXs from the rest. This type of network has also been used in previous studies for the automatic detection of EXs (García et al., 2009). This classification stage comprises three steps:

1. **Feature extraction**. For each region candidate in $M_{ex-cand}$, a set of features was extracted using the layers obtained in the previous stages. It should be noted that the same set of 100 features used for RL candidate classification were also used in the case of EXs. This is adequate to simplify the method and evaluate the suitability of the features depending on the type of lesion. The extracted features are collected in Table 4.2.

2. **Feature selection**. As for the RL classification, we used the FCBF method to select a reduced number of features. The 34 selected features in this stage are specified in Table 4.2. Again, features of a different nature were selected, including shape, distance, intensity, and variability around the candidate in different layers.

3. **Multilayer Perceptron Neural Network**. In this step, EX candidates were classified using an MLP with the same configuration as the one used for RL candidate classification 4.4.2. However, a joint classification for both types of lesions would not be possible in our approach, since the selected features were different. Moreover, it is interesting to separately classify the lesions from a clinical point of view. They often appear at different times and have implications in determining the severity of the disease (Abramoff et al., 2010). In this case, the number of neurons in the input layer was the number of selected features in the previous step. The number of neurons in the hidden layer and regularization parameter were also experimentally optimized during the training process.

4.6   Diabetic retinopathy severity grading

In this section, we explain the proposed method for DR severity grading, which is based on a deep learning framework. It is important to mention that the International Clinical DR Scale was considered. This is the standardized practical clinical classification system used in the latest studies in the literature aimed at automatic screening of DR. The International Clinical DR Scale specifies 5 DR

stages (see section 1.5.2). Therefore, the severity grading challenge consisted of a 5-class classification problem.

For a better understanding of the proposed method, the diagram in Figure 4.14 is provided. Initially, the input image was evaluated in terms of quality using the algorithm previously described in section 4.1.2. Only those images with enough quality for analysis underwent the DR grading stage. Then, we applied a minimal preprocessing stage, which is detailed in section 4.6.1, to prepare the image for the CNN architecture. Next, we decomposed the preprocessed image into 2 new images, as described in section 4.6.2: the first image (I_{bri}) represented the bright structures of the retina and the second one (I_{dark}) represented the dark ones. These 2 new images, together with the original preprocessed image, were the inputs of the CNN architecture. Each of these inputs were processed by a pretrained FCN allowing optimal feature extraction. Next, the features extracted from the preprocessed image (feature matrix M_{pre}) were combined, separately, with the features extracted from the other images (features matrices M_{bri} and M_{dark}) using an attention mechanism. This mechanism, detailed in section 4.6.5, selects the relevant elements from the matrices. Finally, the proposed architecture included a set of fully connected layers to classify the image into the different severity degrees based on the complete set of extracted features (section 4.6.6).

In addition to the deep architecture, the use of data augmentation and transfer learning, described in sections 4.6.3 and 4.6.4, should be highlighted. Finally, the training procedure is detailed in section 4.6.7.

4.6.1 Preprocessing

Deep learning approaches are designed to process raw images. Hence, a complex preprocessing stage, like the algorithm exposed in section 4.2, is not required (Saha et al., 2018; Zago et al., 2018). Instead, in this work we applied a minimal preprocessing stage before using the CNN architecture. First, input color images were cropped around the FOV by trimming the black external background borders, which makes the retinal diameter equal to the width of the image (Araújo et al., 2020; de la Torre et al., 2020). This step allowed us to remove useless information for the classification task. Then, they were resized to a resolution of 640 × 640 pixels. As proved in (de la Torre et al., 2020), the optimal size for DR grading is a retina diameter equal to 640 pixels. Finally, we normalized the pixel values to the interval [-1, 1] for a better training process (Goodfellow et al., 2016), obtaining the preprocessed image (I_{prep}).
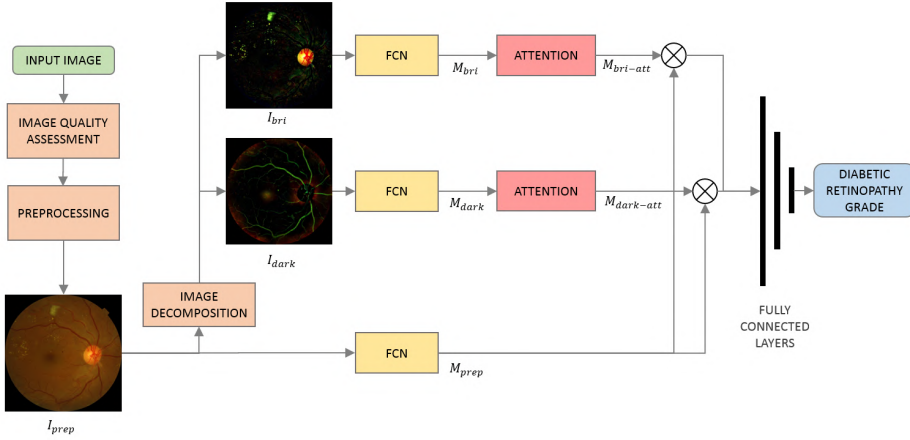
Figure 4.14: Overview of the proposed method. First, we applied an image quality assessment stage to check if the input image was gradable. Second, a minimal preprocessing step was performed (I_{prep}). Then, we decomposed the resulting image to obtain the images I_{bri} and I_{dark}. These images, together with I_{prep}, were processed using a pretrained Fully Convolutional Network (FCN). The resulting feature matrices M_{bri} and M_{dark} underwent an attention mechanism, producing the feature matrices $M_{bri-att}$ and $M_{dark-att}$. Next, the extracted features from I_{prep} and M_{prep} were separately multiplied by $M_{bri-att}$ and $M_{bri-att}$, and then combined. Finally, the final classification was performed using a set of 3 fully connected layers

.

4.6.2   Image Decomposition to Separate the Dark and the Bright Regions

The main purpose of this stage is to separate the dark and the bright pixels in the fundus images. This allowed us to obtain 2 different versions of the image: the first version (I_{bri}) preserved the bright regions of the retina while removing the dark ones. Conversely, the second version (I_{dark}) highlighted the dark regions in the fundus image while ignoring the bright areas. These two new images served us to feed the CNN. In this way, we were able to separate attention from dark structures (e.g., red lesions) from bright structures (e.g., hard exudates). In order to perform the image decomposition, we applied the multi-scale algorithm defined in sections 4.4.2 and 4.5. Figures 4.12a and 4.13a show an example of the images obtained in this stage. In I_{bri}, the color difference of the bright pixels with respect to the retinal background was enhanced, while the rest of pixels remained black. In the same way, the color difference between the dark pixels and the background was highlighted in I_{dark}, while leaving the rest of the pixels black.

4.6.3 Data Augmentation

Having a larger database to train offers better results and increases the generalization ability of the method (Zago et al., 2018). In this study, we applied the technique data augmentation, introduced in section 4.1.2, to increase the number of training samples. This technique acts as a regularizer and helps reduce overfitting when training a machine learning model. Fake images were obtained applying the following simple transformations (Perez and Wang, 2017): rotations in the range $[-50, +50]$ degrees, zoom in the scale range $[-0.1, +0.1]$ and horizontal and vertical flips. The proposed transformations were chosen to produce realistic results.

4.6.4 Feature Extraction with Transfer Learning

In the proposed method, we applied a FCN as a feature extractor for all 3 input images (I_{prep}, I_{bri} and I_{dark}). The architecture selected for this network was Xception (eXtreme version of Inception) (Chollet, 2017). This architecture replaces the Inception modules with depthwise separable convolutions. Xception outperforms many other architectures such as VGG16, ResNet152 or InceptionV3 on large image classifications datasets (Chollet, 2017).

Feature extraction was performed using transfer learning (see section 4.1.2). This technique has already been successfully applied for DR grading (Costa et al., 2019; Wang et al., 2017). In this work, we used an FCN pretrained on the images from the project ImageNet (Jia Deng et al., 2009).

4.6.5 Attention Mechanism

The extracted features for I_{prep}, I_{bri} and I_{dark} are the matrices M_{prep}, M_{bri} and M_{dark}, respectively, all of them having dimensions $14 \times 14 \times 1556$. This is the output size of the Xception backbone for the input size $640 \times 640 \times 3$. The advantage of using an FCN is that the extracted features preserve the spatial information, which allows interpreting the mentioned matrices as a set of 14×14 image patches with 1556 features each. This way, we can focus on those features associated with the spatial locations of the image (patches) that are more relevant for our classification task. In this case, attention mechanisms produce a weight map which assigns a factor to the features based on their relevance (Chen et al., 2016). In this work, we applied an attention model for the matrices M_{bri} and M_{dark}, which, in turn, were computed from the input images I_{bri} and I_{dark}. This

approach allowed us to produce separate attention maps for bright lesions and red lesions. Additionally, these maps were useful to reduce the impact of unneeded lesion information for identifying different DR grades.

The architecture of the proposed attention mechanism can be seen in Figure 4.15. It takes a feature matrix as input and outputs an attention map. It is composed of a set of 4 Conv2D layers with kernel (1×1). The first 3 layers included a ReLU activation function. The last one had a sigmoid activation function, producing an output between 0 and 1, which is the weight of every patch based on their relevance for the classification task.

Once both attention maps were computed, they were applied over M_{prep} using the element-wise multiplication on every channel (layer in the last dimension), obtaining the feature matrices $M_{bri-att}$ and $M_{dark-att}$. Finally, these matrices were concatenated on channel dimension (depth direction) as the input to the fully connected layers.

4.6.6   Fully connected layers

The last part of the proposed architecture takes the combined features obtained with the attention mechanism as inputs and is responsible for the final DR classification. It is composed of 3 fully connected layers. The first 2 layers had 1024 and 512 neurons, respectively, and a ReLU activation function (Goodfellow et al., 2016). Additionally, they included a L2 regularization penalty with a factor of 0.005. The last layer had 5 neurons, one for each DR severity degree. The activation function for this layer was softmax, which represents the output probability of every class to be predicted (in our study, each severity degree) (Goodfellow et al., 2016).
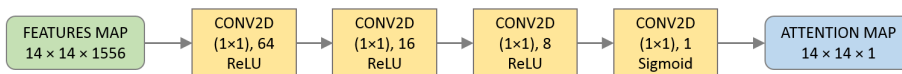
Figure 4.15: Architecture of the attention mechanism.

4.6.7 Training procedure

As shown in section 3.2.5, the database used in this study contains a greater number of images associated with one severity degree than others. In particular, the class corresponding to no DR is represented by many more images than the rest of classes. In order to deal with class imbalance, we used the focal loss error function (Lin et al., 2017b). This approach is a modification of the standard cross entropy loss function used for training. It down-weighs the loss assigned to well-classified examples. The focal loss function introduces 2 hyperparameters: the focusing parameter γ, that controls the strength of the modulating term, and the weighing factor α, which balances the importance of positive/negative examples. Focal loss is defined as (Lin et al., 2017b):

$$FL(p_t) = -\alpha(1 - p_t)^\gamma \log p_t, \tag{4.27}$$

where p_t is the probability of the ground truth class. In this work, γ was set to 2 and α was set to 1 (Lin et al., 2017b).

The model was trained for 100 epochs using a batch size of 8 images. We used the stochastic gradient descent as the optimization algorithm with a momentum of 0.9 (González-Gonzalo et al., 2018) and learning rate of 0.005. To avoid overfitting in advanced epochs, the learning rate was reduced by a factor of 10 when the validation error reached a plateau. Additionally, to avoid exploding gradients, the L2 norm of the gradient vector was limited to 1 (Goodfellow et al., 2016).

4.6.8 Ablation studies

In order to evaluate the effectiveness of the proposed architecture, we performed 2 additional ablation studies:

1. We built another architecture with no attention mechanism, directly feeding the fully connected layers with the matrix M_{prep}. The rest of the hyperparameters were identical to the proposed method. This approach can be seen as a standard deep classification network.

2. We modified the attention mechanism such that it was applied to the matrix M_{prep}, which has joint information about the bright and dark pixels. The rest of the hyperparameters were identical to the proposed method.

Chapter 5

Results

This chapter is dedicated to presenting the results obtained in the studies conducted during the course of this Doctoral Thesis. The next sections are organized according to the hypotheses outlined in section 2.1. This way, sections 5.1.1 and 5.1.2 gather the results for the RIQA methods (Jiménez-García et al., 2019; Romero Oraá et al., 2020). The results associated with the location of de OD and the fovea (Romero-Oraá et al., 2020) are shown in section 5.2. The results for both of the proposed methods aimed at RL detection (Romero-Oraá et al., 2019, 2020) are exposed in section 5.3 while the results of the method aimed at EX detection (Romero-Oraá et al., 2020) are exposed in section 5.4. Finally, we present the results for the DR severity grading method in section 5.5.

5.1 Image quality assessment

5.1.1 Method based on the combination of global features

Performance evaluation

An MLP neural network was used to separate the images with adequate quality and inadequate quality. Classification performance was evaluated in terms of sensitivity (Se), specificity (Sp), accuracy (Acc), positive predictive value (PPV), and F-Score (F_1) (Fawcett, 2006). Additionally, we obtained the receiver operating characteristic (ROC) curve and computed the area under the curve (AUC) to measure the robustness of the classifier (Fawcett, 2006). This value is expected to be close to 1 when the achieved results were robust (Fawcett, 2006).

Feature selection results

The FCBF algorithm was applied over 1000 bootstrap replicates of the features extracted from the training set. Features selected on, at least, 500 runs of the bootstrap method were finally selected as optimum features for the classification task. A total of 10 features formed the reduced subset. It should be noted that features from all proposed feature extraction methods were selected: six features from SSEQ ($SpacEn_{SKEW}$, $SpecEn_{MEAN}$, and $SpecEn_{SKEW}$ in the scales 1 and 3), the NIQE index $QNIQE$, one feature from the CWT analysis (ENT_{CWT} at $s = 4$), and two features from the luminosity analysis (Lum_{5-1} and Lum_{15-10}).

Multilayer perceptron optimization

In order to optimize the parameters of the network, we experimented with values of N_{HIDDEN} ranging from 1 to 100 and values of λ ranging from 0 (no regularization) to 0.9 in steps of 0.1. For each combination, we applied 10-fold cross validation and averaged the validation accuracy across the 10 iterations (Witten et al., 2016). Figure 5.1 shows the obtained performance in terms of Acc for different combinations of N_{HIDDEN} and λ. Maximum Acc was reached with $N_{HIDDEN} = 21$ neurons and $\lambda = 0.1$. Therefore, this combination of hyper-parameters was used to train the final model. Once the hyper-parameters N_{HIDDEN} and λ were fixed, the MLP network was trained using the complete training set.

Finally, the obtained ROC curve can be seen in Figure 5.2 and the AUC showed a value of 0.9803. The optimum point of operation was selected during training as the point in the ROC curve closest to (0,1).
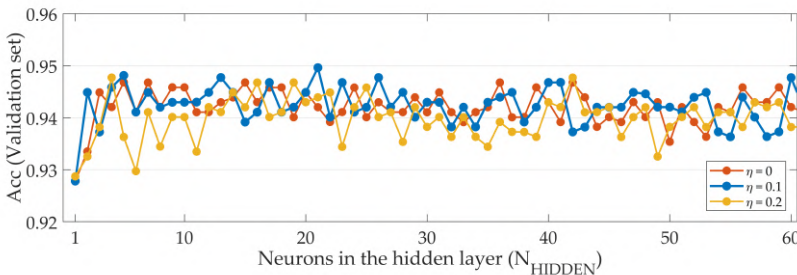
Figure 5.1: Some accuracy curves obtained for multilayer perceptron (MLP) optimization during the training stage.
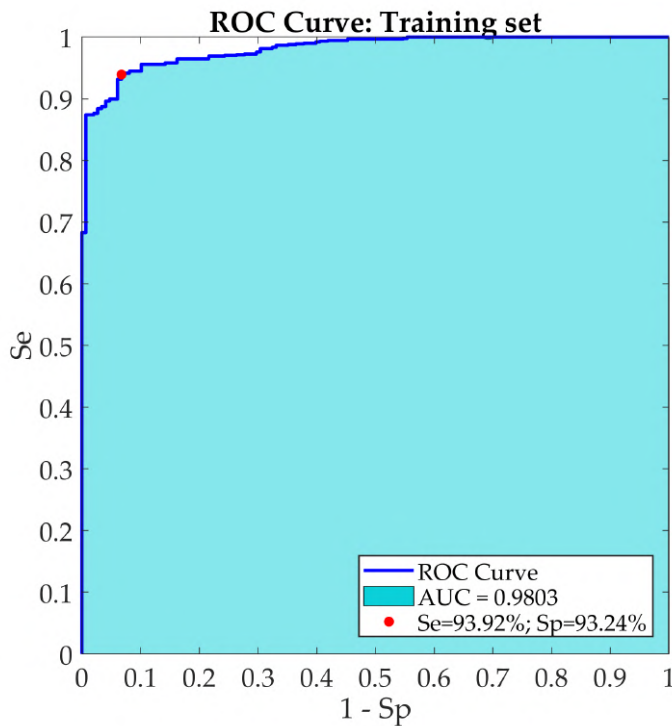
Figure 5.2: Receiver operating characteristic (ROC) curve for retinal image quality assessment based on the combination of global features.

Classification results on the test set

Once all the parameters of the classification stage were fixed, the final results of the proposed method were obtained using the test set of the private database (1,054 images). Table 5.1 summarizes the results for the test set in terms of Se, Sp, Acc. Additional results showed $PPV = 97.88\%$, $F1 = 0.9487\%$, and $AUC = 0.9803$.

Table 5.1: Results for fundus image quality assessment using the proposed methods.

Method	Database	Se (%)	Sp (%)	Acc (%)
Global features	Private	92.04	87.92	91.46
Deep learning	Private	96.82	91.00	95.29
Deep learning	DRIMDB	99.20	100.00	99.48

5.1.2   Method based on deep learning

The CNN training was carried out with the 2,348 images from the private database training set. A Workstation with an Intel Xeon CPU E5-1620 v4 @ 3.5GHz × 8 processor, 32GB of RAM and two NVIDIA TITAN X (Pascal) graphics cards was used. CNN took 128 minutes to train.

Once the CNN was trained using the training set of the private database, the method was evaluated on the test set of the private database (425 images) and on all the images of the public database DRIMDB (194 images) (Sevik et al., 2014). Classification performance was evaluated in terms of sensitivity (Se), specificity (Sp), and accuracy (Acc). Table 5.1 collects these results.

5.2   Location of the optic disc and the fovea

The methods for the OD and fovea location were evaluated over the test set of the private dataset (283 images) and the public databases DIARETDB1 (89 images), DRIVE (40 images) and Messidor (1200 images). Accuracy was computed comparing the automatically detected center with the center annotated by the ophthalmologist as (Harangi and Hajdu, 2015):

$$Acc = \frac{C}{N},\tag{5.1}$$

where C is the number of correct detections and N is the total number of tested images. We considered a correct detection when the Euclidean distance between both points (distance error) was lower than a certain threshold (DE_{max}). Otherwise, the detection was considered incorrect (Welfer et al., 2011). As in other studies for both OD and fovea location, we considered $DE_{max} = R_{OD}$ pixels (Hsiao et al., 2012; Medhi and Dandapat, 2015; Niemeijer et al., 2009). This way, we can establish a direct comparison with the rest of the methods in literature. The results of the proposed methods for the OD and fovea location are shown in Table 5.2.

Additionally, we calculated the overlapping area between the ground truth and the estimated OD. Given that the edges of the fovea are undefined, this measure was not calculated for the detected fovea. Since the ophthalmologist only provided us with the ground truth center, both the ground truth and the estimated OD were modeled as a circle with radius $\frac{D}{12}$ pixels (Hsiao et al., 2012). Then, we computed the overlapping ratio between both circles as (Chalakkal et al., 2018; Lupaşcu

Table 5.2: Results for the location of the optic disc and the fovea.

Database	Number of test images	Optic disc location accuracy	Fovea location accuracy
Proprietary	283	100%	100%
DiaretDB1	89	100%	100%
DRIVE	40*	100%	100%
Messidor	1200	99.5%	99.67%

* 3 images did not show the fovea and were discarded for the fovea detection method (Kauppi et al., 2007; Welfer et al., 2011).

et al., 2008; Morales et al., 2013; Salazar-Gonzalez et al., 2011):

$$S = \frac{Area(T \cap D)}{Area(T \cup D)}, \tag{5.2}$$

where T is the ground truth circle and D is the estimated OD circle obtained using the proposed method. For all the images in test set of the proprietary database, the average overlapping ratio was 0.826.

5.3 Detection of red lesions

5.3.1 Method based on the classification of superpixels

Performance Assessment

The optimal parameters of the proposed algorithm were determined using the training set of our private database, composed of 281 fundus images. The number of hidden neurons and the regularization parameter value were experimentally adjusted. We varied the value of those free parameters using 10-fold-cross-validation. We chose the parameter values that maximized the average accuracy (ACC_{NN}) over the validation set from the first fold, keeping the balance between the average sensitivity (SE_{NN}) and the average specificity (SP_{NN}).

Once the parameters of the algorithm were optimized and the classifier was trained, we evaluated the performance of our approach using the test set of our private database (283 images) and the test set of the DIARETDB1 database (61 images). We obtained the results on both databases in terms of two different criteria (García et al., 2010):

- **Pixel-based criterion**: a lesion was correctly identified when, at least, one of its pixels was detected as a lesion. Additionally, we considered all

the pixels belonging to a correctly detected lesion to be correct. Based on this criterion, we calculated the positive predictive value (PPV_p) and the sensitivity (SE_p) (Osareh, 2004).

- **Image-based criterion**: frequently, the detection of very small regions correspond to noisy spots. These are clinically non-significant but may be detected as lesions in some images belonging to healthy retinas. Following the same approach used in previous studies (García et al., 2010; Osareh, 2004), we considered an image pathological when, at least, 30 pixels in the whole image were detected as lesions. This corresponds to a very small fraction of the pixels of the image (less than 0.00001%), similar in size to that of a microaneurysm. Based on the image-based criterion, the average sensitivity (SE_i), specificity (SP_i) and accuracy (ACC_i) over the test set were calculated.

Training

- **Feature selection**

  All 39 features described in section 4.4.1 were studied in terms of relevancy and redundancy applying the FCBF algorithm. A total of 54,233 superpixels were extracted from the training set, of which only 1,783 corresponded to RLs. To balance the two classes considered during training, we also randomly selected another 1,783 superpixels that did not correspond to RLs. Thus, a vector of 39 features was calculated for each of the 3,566 considered superpixels. It is important to note that feature vectors where normalized (mean = 0, standard deviation = 1) to improve the classification results (Bishop, 1995).

  After applying FCBF, the 16 features indicated in Table 4.1 were selected. It should be noted that the reduced feature set contained features related to shape, pixel intensities or the distance to the retinal structures. Hence, the selected features take into account the different distinctive characteristics of RLs.

- **MLP Configuration**

  We experimentally determined the number of neurons in the hidden layer (n_{hid}) and the value of the regularization parameter λ. The number of hidden neurons was analyzed in the range $n_{hid} = [1 : 1 : 100]$. The regularization parameter values were varied in the range $\lambda = [0 : 0.1 : 1]$. For

each combination of n_{hid} and λ, we obtained ACC_{NN}, SE_{NN} and SP_{NN}. The values of ACC_{NN} are shown in Figure 5.3. It can be seen that ACC_{NN} stabilizes from $n_{hid} = 30$ when $\lambda < 0.6$. For larger values of λ, ACC_{NN} decreases. Thus, we selected $n_{hid} = 30$ and $\lambda = 0.6$. At this operating point, we obtained $ACC_{NN} = 94.50\%$, $SE_{NN} = 93.24\%$ and $SP_{NN} = 95.76\%$. Subsequently, we trained an MLP with $n_{hid} = 30$ and $\lambda = 0.6$ using the complete training set.

Results on the test sets

After parameter setting and final training of the classifier, the method was applied on the images of the test sets of the private database and the public database DI-ARETDB1. The results of the automatic method were compared with the annotations of the ophthalmologist in terms of a lesion-based criterion and an image-based criterion (García et al., 2010). The results are summarized in Table 5.3. The performance of the proposed method is illustrated in Figure 5.4.
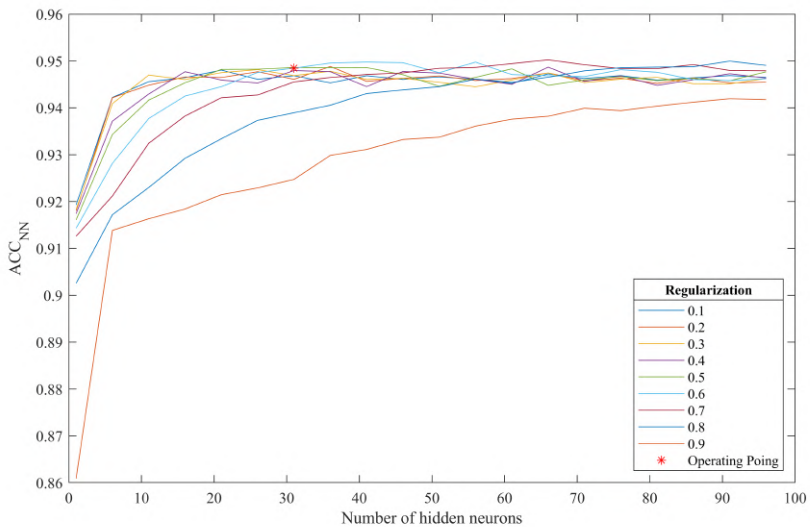
Figure 5.3: Average accuracy for RL classification over the validation set obtained during MLP training for varying the number of hidden neurons and the regularization parameter.

Table 5.3: Results for the detection of red lesions.

Method	Database	Pixel-Based Criterion		Image-Based Criterion		
		SE_p	PPV_p	SE_i	SP_i	ACC_i
Superpixels	Private	81.43	86.59	84.04	85.00	84.45
Superpixels	DIARETDB1	88.10	93.10	84.00	88.89	86.89
Decomposition	Private	82.25	91.07	85.00	90.80	88.34
Decomposition	DIARETDB1	84.79	96.25	88.00	91.67	90.16

(a)   (b)
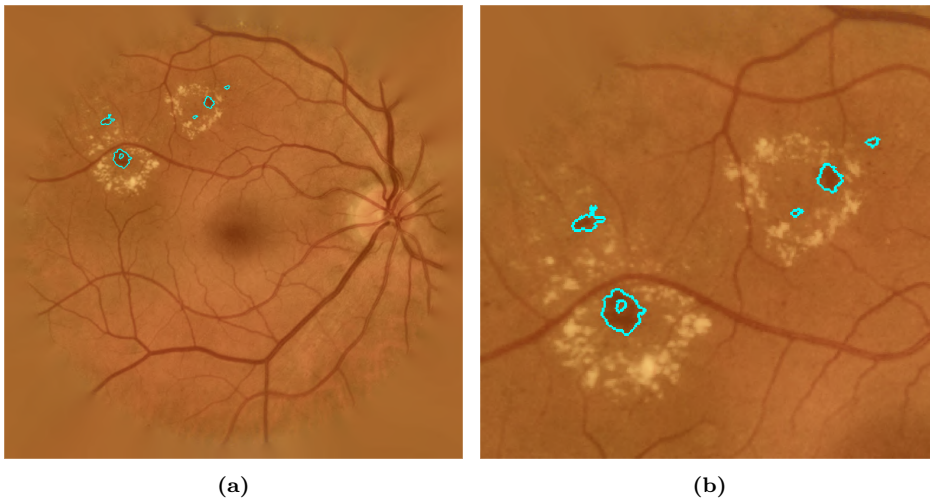
Figure 5.4: Detected red lesions after classification stage. (**a**) example of the private database; (**b**) zoom in previous example.

5.3.2   Method based on the decomposition of the fundus image into layers

All optimal values for the parameters of the proposed method were obtained using the 281 images from the training set of the proprietary database. We obtained the final results using the test set of the proprietary database (283 images) and the test set of the DiaretDB1 database (61 images).

We extracted 4,889 RL candidates from the training set. Only 2,029 of them were true RLs. We randomly selected another 2,029 non-RL regions to balance the two classes. The extracted features over the RL candidates were normalized (mean = 0, standard deviation = 1) to improve the classification results (Bishop, 1995).

MLP configuration on the training set

We experimented with the number of hidden neurons in the range $[1:1:100]$. The regularization parameter values were varied in the range $[0:0.1:1]$. For this task, we applied 10-fold-cross-validation exclusively using the training set of the private database. This technique is a powerful preventative measure against overfitting. The chosen values for those parameters were 51 and 0.5, respectively, since they maximized the average accuracy over the validation test (see Figure 5.5).

Red lesion detection on the test sets

The results for RL detection using both the private database and the public database can be seen in Table 5.3. These results show the performance of the complete algorithm and were obtained according to the criteria defined in section 5.3.1: lesion-based criterion and image-based criterion.
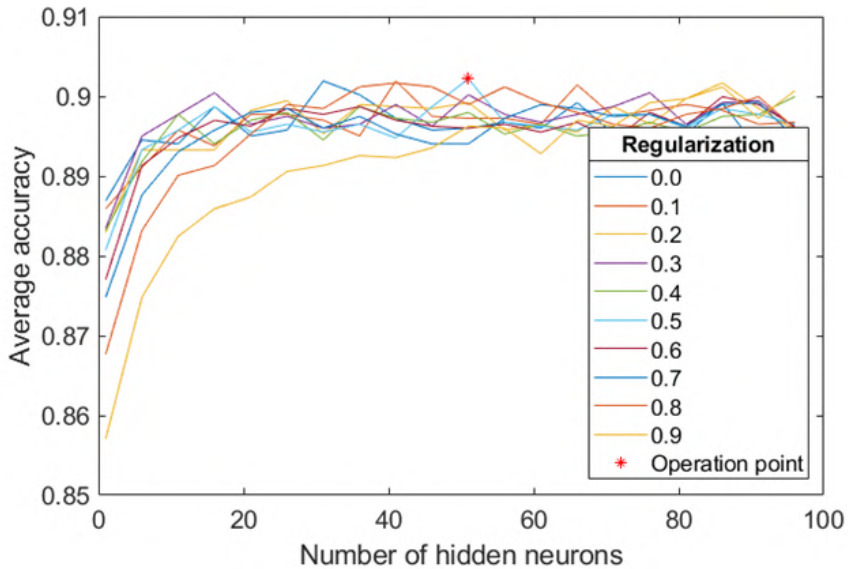
Figure 5.5: Average accuracy for RL classification over the validation set obtained during MLP training for varying the number of hidden neurons and the regularization parameter.

5.4   Detection of hard exudates

Parameter optimization was performed using the 281 images from the training set of the proprietary database. We obtained the final results using the test set of the private database (283 images) and the test set of the DIARETDB1 database (61 images).

We extracted 4,782 EX candidates from the training set. Only 2,072 of them were true EX. Thus, we randomly selected another 2,072 non-EX regions to balance the two classes. The extracted features over the EX candidates were also normalized to improve the classification results (mean = 0, standard deviation = 1) (Bishop, 1995).

MLP configuration on the training set

Like the RL classification, we varied the number of hidden neurons in the range $[1 : 1 : 100]$ and the regularization parameter values in the range $[0 : 0.1 : 1]$. We applied 10-fold-cross-validation exclusively using the training set of the private database allowing to control overfitting. The chosen value for those parameters was 55 and 0.4, respectively, since they maximized the average accuracy over the validation test (see Figure 5.6).

Exudate detection on the test sets

The results for EX detection using both the proprietary database and the public database are presented in Table 5.4. These results show the performance of the complete algorithm and were obtained by applying the criteria in section 5.3.1: lesion-based criterion and image-based criterion.

Table 5.4: Results for the detection of exudates.

| Database | Pixel-Based Criterion | | Image-Based Criterion | | |
	SE_p	PPV_p	SE_i	SP_i	ACC_i
Proprietary	89.42	96.01	88.04	98.95	95.41
DiaretDB1	91.65	98.59	95.00	90.24	91.80
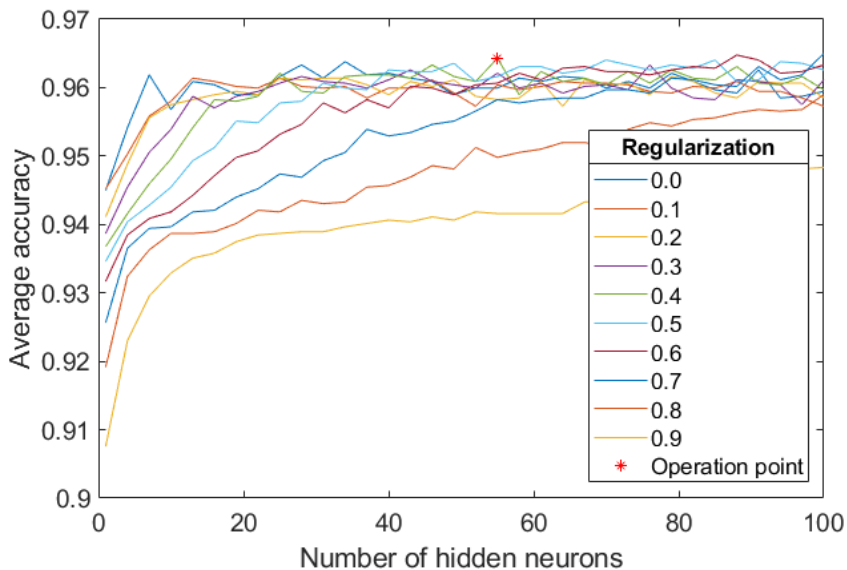
Figure 5.6: Average accuracy for EX classification over the validation set obtained during MLP training for varying the number of hidden neurons and the regularization parameter.

5.5 Diabetic retinopathy severity grading

In the final stage of this PhD Thesis, we dealt with the DR grading multiclass classification problem. We evaluated the proposed method on the test set of 32,017 images from the EyePACS for the Diabetic Retinopathy Detection dataset. As described in section 3.2.5, the dataset was highly imbalanced. For this reason, the proper evaluation of the performance of the proposed method requires an adequate metric. The Cohen's Kappa (Cohen, 1960) is one of the most commonly used statistics to test inter-rate reliability. However, it does not take into account the degree of disagreement (Cohen, 1968). When the categories are ordered, it is preferable to use the Weighted Kappa since it allows disagreements to be weighted differently (Cohen, 1968). This is the case of DR severity degrees, where each class can be seen as an evolution of the previous one. Three matrices are involved in the calculation of the Weighted Kappa: the matrix of observed scores (O), the matrix of expected scores based on chance agreement (E), and the weight matrix (ω). The metric can range from -1 to +1 and is defined as (Cohen, 1968):

$$\kappa = 1 - \frac{\Sigma_{i,j}\omega_{i,j}O_{i,j}}{\Sigma_{i,j}\omega_{i,j}E_{i,j}}, \tag{5.3}$$

where $\omega_{i,j}$, $O_{i,j}$ and $E_{i,j}$ are elements in the weight, observed, and expected matrices, respectively. Quadratic Weighted Kappa (QWK) is the particular case with quadratic weighting (Cohen, 1968):

$$\omega_{i,j} = \frac{(i-j)^1}{(N-1)^2},\tag{5.4}$$

where N is the number of classes. This is the most common metric for DR grading (Araújo et al., 2020; Costa et al., 2019; González-Gonzalo et al., 2018; Krause et al., 2018; Wang et al., 2017). For this reason, we computed QWK, allowing us to compare our results with those in other studies. The QWK achieved with the proposed method was 0.78 on the test set, which corresponds to 83.7% accuracy. When it comes to the first ablation study, where no attention mechanism was applied, the network did not converge adequately due to the class imbalance. The model learnt to classify every sample as the most dominant class (no DR). The second ablation study, where the attention mechanism was applied but the information about the dark and the bright pixels was not separated, showed QWK = 0.76.

However, when dealing with class unbalance, QWK is dominated by the most representative classes, which is the class 0 (No DR) in our dataset. This is why the confusion matrix is also important to evaluate the results. Figure 5.7 shows the confusion matrix obtained with our method. The color intensity of each matrix element is a visual representation of the relative frequency: the ratio of images classified as the predicted class (element value) to the total number of images of the actual class (row sum).
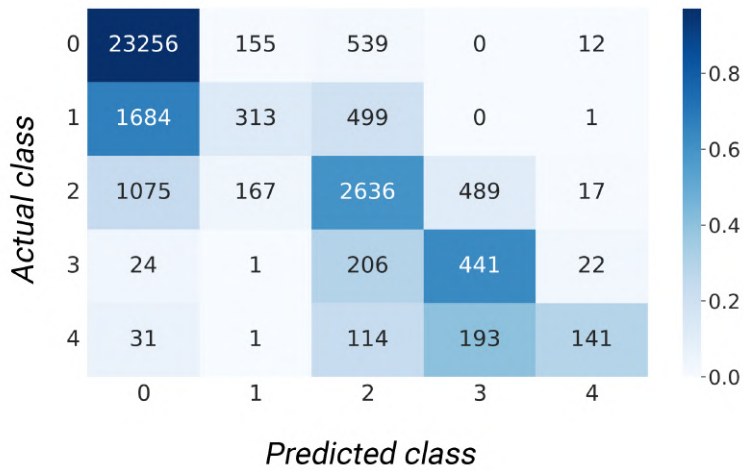
Figure 5.7: Confusion matrix for DR severity grading. The color intensity of each matrix element is a visual representation of the relative frequency: the ratio of images classified as the predicted class (element value) to the total number of images of the actual class (row sum).

Chapter 6

Discussion

In the present Doctoral Thesis, several automatic methods have been developed to build a CAD system aimed at the early diagnosis and severity grading of DR. The motivation for the conducted studies has been guided by the hypotheses displayed in section 2.1. In this sense, we believed that the development of novel, automatic methods for fundus image analysis could contribute to this research field and the limitations of the previous studies could be overcome. Firstly, fundus image quality assessment was achieved with two different approaches, discussed in section 6.1. The results for the first approach, based on the combination of various global features, were outperformed by the second approach, which was based on deep learning. Secondly, the automatic location of the OD and the fovea centers was performed using a set of novel saliency maps. The results of the proposed method were robust and effective, as discussed in section 6.2. Thirdly, the joint segmentation of RLs and EXs was accomplished by decomposing the fundus image into layers. Novel indicators, such as the reflective features of the retina and the choroidal vasculature visible in tigroid retinas, were proven useful for the classification of retinal lesions, as discussed in sections 6.3 and 6.4. An alternative method was also developed for the detection of RLs based on superpixel classification. The technique ERS allowed a precise segmentation of the lesions (section 6.3.2). Finally, the severity grading of DR was addressed with a deep learning framework which performs separate attention for the bright and the dark pixels in the retinal image. Section 6.5) is devoted to discuss the results of this proposed method in detail.

The discussion of the aforementioned studies is exhibited in the next sections of this chapter following the sequence of hypothesis and results previously pre-

sented. At the end of this chapter, the main limitations of this Doctoral Thesis
are presented.

6.1   Image quality assessment

6.1.1   Method based on the combination of global features

In this study, a novel RIQA method was proposed. It was based on generic features
extracted from two different general-purpose NR-IQA methods, the CWT, and
the luminosity of images using the HSV color model. Features were selected and
subsequently classified using the FCBF algorithm and a MLP neural network,
respectively. Results were obtained on a database of 2107 fundus images, reaching
$Se = 92.04\%$, $Sp = 87.92\%$, $Acc = 91.46\%$, $PPV = 97.88\%$, and $F_1 = 0.9487$ on
the test set.

Preprocessing

A preprocessing step was included to enhance the images in our database and
to improve the results in the subsequent processing stages. This preprocessing
method has been successfully applied in previous studies for retinal vessels seg-
mentation (Fathi and Naghsh-Nilchi, 2013; Soares et al., 2006). However, in pre-
vious studies, the FOV extension did not cover the whole area of the image (Fathi
and Naghsh-Nilchi, 2013; Soares et al., 2006). In this work, we propose a novel
modification of this FOV extension algorithm in order to cover the whole image.
This approach has advantages for subsequent retinal image processing algorithms.
One of them is that we could employ SSEQ and NIQE methods to analyze all the
areas in the image before the block selection step (Liu et al., 2014; Mittal et al.,
2013). Furthermore, the preprocessing method prevented border effects when the
CWT and the Gaussian filter employed in luminosity analysis were applied. The
absence of border effects increased the robustness of the features analyzed in this
study.

Feature extraction and selection

After preprocessing, four different feature extraction methods were applied to char-
acterize the images. A total of 40 features were extracted using SSEQ, NIQE,
CWT, and luminosity in the HSV color model. Feature selection was subsequently
applied to discard redundant or irrelevant features using the FCBF algorithm (Yu

and Liu, 2004). Besides, we combined FCBF with bootstrapping to improve the robustness of this stage. To the best of our knowledge, this feature selection approach has not been previously used in RIQA methods. A reduced set of 10 features was finally selected and included features of the different feature extraction approaches analyzed in this study, which demonstrates that the proposed features are complementary. The general-purpose NR-IQA methods used in this work were found useful to characterize the quality of the images. Moreover, the CWT approach was found appropriate to assess the sharpness of the retinal blood vessels and the OD. Finally, the HSV color model was useful to assess the luminosity of images.

The complete set of features included 12 features from the SSEQ method. To the best of our knowledge, SSEQ method has not been previously applied in RIQA. Measurements of $Spac_{En}$ and $Spec_{En}$ through the blocks of the images provided relevant information, especially in scales 1 and 3. Specifically, measures of $SpacEn_{SKEW}$, $SpecEn_{MEAN}$, and $SpecEn_{SKEW}$ for these scales were selected by FCBF to form the final feature subset. However, we observed that $SpacEn_{MEAN}$ was not selected. This may indicate that $SpacEn_{MEAN}$ was found redundant with respect to other features. Regarding the SSEQ method, we found that $SpacEn_{SKEW}$ and both $SpecEn$-derived features can be useful to distinguish images of adequate quality from inadequate quality ones. This result is consistent with other studies that successfully employed entropy-based measurements in the context of RIQA (Abdel-Hamid et al., 2016; Davis et al., 2009; Marrugo Hernández et al., 2011), biomedical signal processing (Bachiller et al., 2015; Poza et al., 2007), and strategic decision making (Rodger, 2019).

A naturalness feature derived from the NIQE method was also considered. The quality feature $QNIQE$ is relevant for quality assessment since it was selected in all the bootstrap runs of the FCBF algorithm, together with $SpecEn_{SKEW}$. Therefore, $QNIQE$ is complementary to all other features proposed in this study. Our results correlate with previous studies (Shao et al., 2017), where the NIQE index was used in combination with illumination and structural-based methods for RIQA.

Another set of features was extracted using the CWT. This method has been previously used to perform retinal vessel segmentation tasks and to discriminate healthy and diseased retinal images (Fathi and Naghsh-Nilchi, 2013; Koh et al., 2017; Soares et al., 2006). We calculated the CWT representations for retinal images at six scales using the Mexican hat mother wavelet. This wavelet was useful to identify sharp edges at varying scales, which are indicative of sharp images.

Representations for $s = 4$–16 showed stronger responses along thin-to-thick blood vessels. Sharp edges were found mainly in the OD and blood vessels and were associated with stronger CWT responses. Previous works that used the wavelet transform also stated the usefulness of this technique to assess the sharpness on fundus images (Abdel-Hamid et al., 2016, 2017; Bartling et al., 2009). In our study, sharpness of CWT representations was assessed with three novel features based on Shannon's entropy and local variance, showing promising results. Characterization of the CWT representations was comprehensive, but these features were found redundant and only one of them (ENT_{CWT} at $s = 4$) was finally selected using the FCBF algorithm.

Finally, luminosity features were also included in the analysis. This type of feature has been previously used to identify poorly illuminated fundus images. Luminosity of retinal images was widely studied using different color models, such as RGB, YUV, or L*a*b* (Davis et al., 2009; Shao et al., 2017; Wang et al., 2016). In this study, we have found that the HSV color model is also useful to characterize poorly illuminated fundus images. The HSV color model allowed us to represent the luminosity of retinal images independently of their color. The color of retinal images is closely associated to the physical features of the patients (such as skin or iris color) and also to the acquisition process (Abdel-Hamid et al., 2016; Zhou et al., 2018). Therefore, separating the luminosity and chromatic information of the image is useful to study brightness-related features. We observed that the luminosity channel of the HSV color model performed better than illumination components of YUV and L*a*b* color models for the images in our database. More specifically, we found that the differences between light and dark areas were more prominent in the V color channel of HSV. This finding can be due to the larger separation of the color information in the hue and saturation channels from the luminosity (V) (Zhou et al., 2018). Therefore, the V channel was less dependent on the color of retinal images. Previous studies have also employed the HSV color model in order to assess retinal image quality (Abdel-Hamid et al., 2016). In that study, color information was more relevant and was extracted from a modified saturation channel. Alternatively, we found relevant features derived from luminosity contained in retinal images. We analyzed several features related to luminosity percentiles and their differences extracted from the V channel. Two of these features ($Lum5 - 1$ and $Lum15 - 10$) were selected by the FCBF algorithm, which indicates the relevance of luminosity features for RIQA.

Results

Results on the test set achieved $Acc = 91.46\%$, $Se = 92.04\%$, $Sp = 87.92\%$, $PPV = 97.88\%$, $F_1 = 0.9487$, and $AUC = 0.9803$. Since the number of adequate quality images was much greater than the number of inadequate quality images, synthetic samples of the minority class were obtained using SMOTE. Despite this class imbalance, this technique allowed us to obtain almost balanced Se and Sp values, which means that our approach was able to model the particularities of both good quality and bad quality images. Besides, the vast majority of images were correctly classified, while PPV and F_1 were also high. The latter measure represents the tradeoff between a high detection rate and high probability of correct detection.

Our results showed 18 false positives and 72 false negatives out of the 1054 images in the test set. Our RIQA method failed to classify images that did not perfectly fit the ideal characteristics of adequate or inadequate quality images. Misclassifications were frequent when images did not have a perfect focus or when a dark but partially sharp image was analyzed. It is important to note that, if this RIQA method was included as the first stage of a more general retinal image analysis algorithm, inadequate quality images misclassified using the proposed method would be further processed. This issue may influence the results of the subsequent image analysis algorithms. Conversely, when an adequate quality image is misclassified, the image would be rejected by the RIQA algorithm. In clinical settings, this can be inconvenient because the photographer would need to capture an adequate-quality fundus image again. However, this issue is not likely to have an important influence over the results of a more general retinal image analysis method on the image.

Some examples of misclassified images in our study are shown in Figure 6.1. In Figure 6.1a, the OD and the main arcades are blurred. However, the vessels and some bright retinal lesions (EXs) are reasonably sharp in other areas of the image. In the case of Figure 6.1b, the image has some dark areas, but the rest of the image is sharp. These two images were incorrectly classified as adequate quality images. The cases shown in Figures 6.1c and 6.1d are examples of false negatives. Both images appear slightly blurred due to poor focus or artifacts, respectively. Thus, the automatic method considered them as inadequate-quality images. However, human graders considered that they had enough quality to be analyzed. These examples reveal that quality assessment of fundus images is a challenging task, that may be influenced by the subjectivity or experience of human graders (Wang

(a)   **(b)**   **(c)**   **(d)**
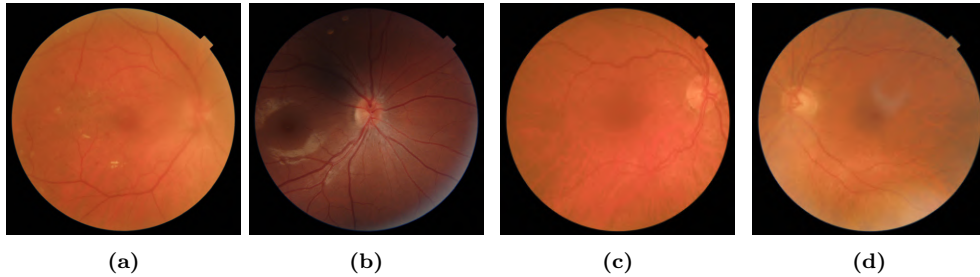
Figure 6.1: Examples of original images in our database that were misclassified using the proposed method based on the combination of global features; (a,b): false positives; (c,d): false negatives.

et al., 2016).

Our final results were similar to those in previous studies. However, comparisons should be made with caution since results are generally measured on different databases and with varying metrics. We evaluated the results of the proposed method using the test set of our database, formed by 1054 images. The majority of studies presented their results using Se and Sp and, in some cases, Acc. In this work, we also included PPV and F_1 in order to better assess the performance of the proposed method. However, these measures are not commonly used in this context. Structural and generic methods achieved Se and Sp around 90%, and Acc was over 90% in most cases. Among structural methods, Fleming et al. (2012) reached $Se = 92.60\%$ and $Sp = 90.00\%$ with 98 images. Other authors validated their method using 400 images, reaching $Se = 95.33\%$ and $Sp = 91.13\%$ (Welikala et al., 2016). Wang et al. (2016) achieved $Se = 87.45\%$ and $Sp = 91.66\%$ using 536 images, while Abdel-Hamid et al. (2017) reached $F_1 = 0.8780$ using 190 images. Other authors also combined the NIQE index with structural and illumination features, reaching $Acc = 93.60\%$ with 194 images. The best performance among generic methods was achieved using the MLP neural network (Pires Dias et al., 2014). Results reached $Se = 99.49\%$ and $Sp = 99.76\%$ in a database formed by 848 images Pires Dias et al. (2014). Hybrid approaches also showed remarkable results, although they are more complex than generic methods (Paulus et al., 2010; Sevik et al., 2014). Paulus et al. (2010) employed 301 images, achieving $Acc = 91.70\%$. In other studies, F_1 reached 0.9960 using 194 images (Sevik et al., 2014). Deep learning-based methods have been previously used for RIQA, achieving excellent performances. Among them, the work developed by Saha et al. (2018) should be remarked, since it achieved a perfect classification over 3425 images. However, it should be noted that only 123 (3.6%) of these images corresponded to inadequate

quality images. Therefore, class imbalance should be taken into account when analyzing these results.

Novel generic features showing great complementarity were found. We achieved similar results compared to other generic methods, suggesting that our novel features form a viable alternative. In fact, the proposed features could be taken into account as a base for a more advanced RIQA solution. Furthermore, the proposed features are fully interpretable, which facilitates clinical viability.

6.1.2 Method based on deep learning

In this work, a method based on transfer learning was proposed to evaluate the quality of fundus images. As a main contribution, we used a CNN with InceptionResNetV2 architecture that, to the best of our knowledge, has never been used for this task. Likewise, the use of fine-tuning with this architecture was validated. The method has been developed exclusively using fundus images from a private database. However, high accuracy was obtained with different images from a public database, which demonstrates its robustness.

The method was evaluated on a test set of our private database (425 images) and the DRIMDB database (194). The results of Table 5.1.2 indicate that the proposed method is a simple and effective way to classify fundus images according to their quality level. However, the results obtained on the DRIMDB database are superior to those obtained on the private database. This may be because the DRIMDB database has more separation between classes. In general, there are clear visual differences between good quality and poor quality images in the DRIMDB database. However, those differences are less evident in the private database.

Our results are similar to those obtained in previous studies, as shown in Table 6.1. The most direct comparison can be established with the work of Zago et al. (2018) and Chalakkal et al. (2019), where the public database DRIMDB (Sevik et al., 2014) was also used. The results obtained in this work with the InceptionResNetV2 architecture on DRIMDB exceed those obtained in (Zago et al., 2018) and (Chalakkal et al., 2019), as can be seen in Table 6.1. In other works, different databases have been used, which influences the results obtained and makes direct comparison between studies difficult. It is worth highlighting those works where the use of CNNs has been combined with additional stages that notably increased accuracy (Coyner et al., 2018; Yu et al., 2017). It also highlights the work of Saha et al. (2018), where 100% accuracy was obtained. However, they ruled out images of ambiguous quality, which are the most error-prone, for the study. When com-

Table 6.1: Comparison of some deep learning methods for image quality assessment.

Study	Database	CNN	Accuracy
Yu et al. (2017)	Kaggle	Alexnet	95.42%
Saha et al. (2018)	Kaggle	Alexnet	100%
Zago et al. (2018)	DRIMDB	Inception v3	98.55%
Chalakkal et al. (2019)	DRIMDB	Inception v1	95.29%
Proposed method	**Private**	**InceptionResNetV2**	**95.29%**
Proposed method	**DRIMDB**	**InceptionResNetV2**	**99.48%**

(a)   (b)
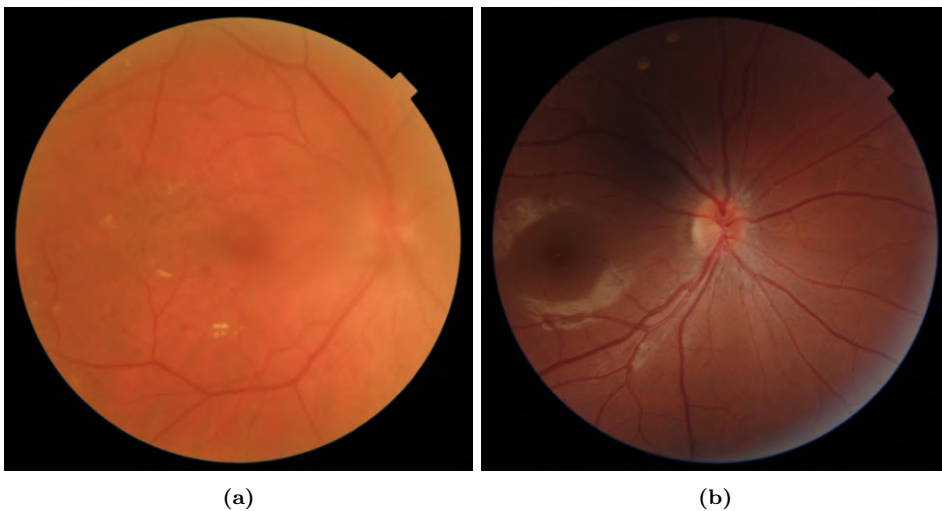
Figure 6.2: Examples of original images in our database that were misclassified using the proposed method based on deep learning; (a): false positive; (b): false negative.

paring our previous method based on hand-crafted features (section 5.1.1), which achieved $Se = 92.04\%$, $Sp = 87.92\%$ and $Acc = 91.46\%$, with our approach based on deep learning, we can claim that the latter provides better results using the same private database and the same metrics ($Se = 96.82\%$, $Sp = 91.00\%$ and $Acc = 95.29\%$).

Despite the high accuracy of the results obtained, some classification errors occur. An example can be seen in Figure 6.2a, considered by the expert to have poor quality. In this image, the structures are well appreciated in part of the image, but not in all of it. This may be the reason for the discrepancy between the expert and the automatic method. Figure 6.2b shows a fundus image classified as of good quality by the expert and in which the system has failed. As it can be seen, it is a dark image, which may be the cause of the error.

It is necessary to mention that the method proposed in this study can be adapted to different databases, since none of the processing steps depends on the specific characteristics of the images. In addition, it is easy to apply to new images once the model has been trained. Therefore, the implementation of the algorithm would be straightforward in the clinical setting. However, the lack of explanatory capability, inherent in deep networks, could hinder the viability in real scenarios. Given this limitation, the application of explainable artificial intelligence (XAI) techniques would be highly useful. Finally, the comparison between the two RIQA methods proposed in this Doctoral Thesis reveals the great impact of deep learning, which is definitely revolutionizing computer vision in medical imaging.

6.2 Location of the optic disc and the fovea

In this study, we proposed automatic methods for OD and fovea location. The initial preprocessing stage, described in section 4.2, was useful for inter-image and intra-image normalization, achieving consistent outputs for the operations applied on every image. The method for the retinal background extraction is an important contribution in this study. The brightest and darkest pixels were replaced with the estimated retinal background, obtaining a realistic result. The retinal background extraction could also be valuable in further studies related to fundus image processing. In fact, using this algorithm, we improved the vascular network segmentation, preventing the edges of the EXs and other bright regions from being wrongly detected as blood vessel segments. The methods to locate the OD and the fovea formed the main stages of the proposed algorithm. We introduced novel saliency maps based on the spatial relationships between the main anatomical structures of the retina and their characteristic visual appearance. These maps provided a great generalization ability.

The proposed methods for the OD and fovea location were assessed using four retinal image databases. The test set of the private database was used to evaluate the performance and the public databases DIARETDB1, DRIVE and Messidor allowed us to evaluate the robustness and generalization ability of the proposed method. It should be noted that all the images included in the databases used in this study were obtained based on a specific capture protocol. According to this protocol, all images contained the vasculature, the OD and the fovea within the FOV. Only 3 images from the DRIVE database did not show a visible fovea and were discarded for the fovea detection method. In addition, both the OD and the fovea were in the expected region of the OD defined in (Sevik et al., 2014). It is

also important to mention that all the parameters of the method are defined as a fraction of D. Therefore, they are relative to the FOV size and the method is independent to the image resolution.

The results for the OD location method showed 100% accuracy for all databases, except for Messidor. In this database we achieved 99.5% accuracy, which means that only 6 detections out of 1200 images were incorrect. These results are in line with those in other studies, as shown in Table 6.2. The results for the DRIVE database are the same as in other studies (100% accuracy). As for DIARETDB1, only two previous studies reported 100% accuracy (Abdullah et al., 2016; Abed et al., 2016). Our method outperforms the location accuracy obtained in all the previous methods for the Messidor database, except for (Giachetti et al., 2013) and (Yu et al., 2015). However, no results were provided for DRIVE and DIARETDB1 databases in (Giachetti et al., 2013). Meanwhile, the method proposed in (Yu et al., 2015) obtained a lower accuracy than our method for the DIARETDB1 database (99.88%).

It should be noted that our method failed to locate the OD in certain images.

Table 6.2: Results for the automatic location of the optic disc.

Study	DRIVE	DIARETDB1	Messidor
Sinthanayothin et al. (1999)	60.00	-	-
Walter et al. (2002)	80.00	-	-
Welfer et al. (2011)	100	97.50	-
Aquino (2014)	-	-	99.00
Lu and Lim (2011)	97.50	98.88	-
Lu (2011)	-	-	98.77
Hsiao et al. (2012)	100	-	-
Qureshi et al. (2012)	100	97.79	-
Yu et al. (2012)	-	-	99.00
Pereira et al. (2013)	100	93.25	-
Giachetti et al. (2013)	-	-	99.67
Basit and Fraz (2015)	100	98.88	-
Yu et al. (2015)	100	99.88	99.67
Harangi and Hajdu (2015)	100	98.88	98.33
Rahebi and Hardalaç (2016)	100	94.38	-
Abed et al. (2016)	100	100	-
Abdullah et al. (2016)	100	100	99.25
Díaz-Pernil et al. (2016)	97.50	97.75	-
Alshayeji et al. (2017)	100	97.75	-
Chalakkal et al. (2018)	100	97.70	98.60
Al-Bander et al. (2018)	-	-	97.00
Proposed method	**100**	**100**	**99.50**

In Figure 6.3a there is a bright structure located near the OD area whose shape and size are easily confused with the OD. In Figure 6.3b we have a very blurred image. Finally, in Figure 6.3c we have a deformed OD, which extends over a large area. However, our method succeeded in locating the OD in other complicated cases. In Figures 6.4a to 6.4e the OD lacks the brightness characteristic. In Figures 6.4f to 6.4i four poor quality images are shown, where the edge of the OD was not well defined and some vessels were uncertain. Our method also worked well with choroidal and retinal thickness showing a large bright peripapillary region (see Figures 6.4f to 6.4g) or images presenting prominent exudates (see Figure 6.4h).

The results for the fovea location method also showed 100% accuracy for all databases, except for Messidor (99.66%). These results are in line with those in other studies, as shown in Table 6.3. We excluded three images that did not present visually detectable fovea in DRIVE database, as in (Welfer et al., 2011). All the foveas in the remaining 37 images were correctly detected (100% accuracy). The same accuracy was obtained by GeethaRamani and Balasubramanian (2018) but using only 35 out of 40 images. This means that they discarded 2 additional images whose fovea was probably hard to detect with their method. The accuracy obtained by Qureshi et al. (2012) was lower using the DRIVE database (91.73%). As for the DIARETDB1 database, to the best of our knowledge, our method obtained the highest accuracy in the literature, correctly locating the foveas in all DIARETDB1 images. Considering the Messidor database, our method also outperformed the accuracy of previous studies. It should be noted that, although the fovea location was based on the previous OD location, less false fovea detections than false OD detections were obtained. This is because some false OD detections

(a)   (b)   (c)
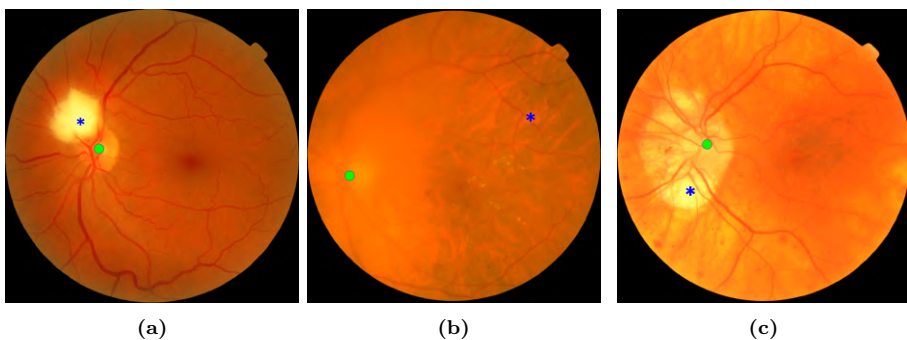
Figure 6.3: Examples where the optic disc is incorrectly detected. The green circle represents the ground truth and the blue mark represents the automatic detection. (a) Image with a bright structure located near the optic disc. (b) Very blurred image. (c) Image with choroidal and retinal thickness showing a large bright peripapillary region.

| (a) | (b) | (c) | (d) |

| (e) | (f) | (g) | (h) |

| (i) | (j) | (k) | (l) |
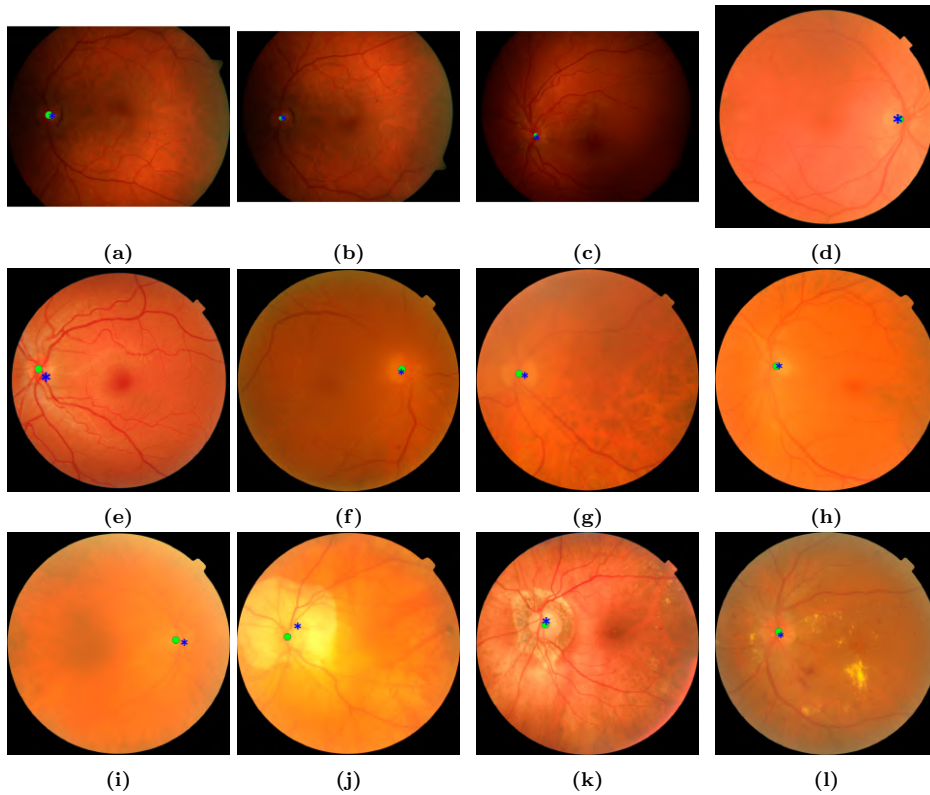
Figure 6.4: Examples where the optic disc is successfully detected. The green circle represents the ground truth and the blue mark represents the automatic detection. (a-e) Images where the optic disc lacks the brightness characteristic. (f-i) Poor quality images, where the edge of the optic disc is not well defined and some vessels are uncertain. (j-k) Images with choroidal and retinal thickness, showing a large bright peripapillary region. (l) Image having prominent exudates. The different aspect ratio of the example images is due to the fact that they belong to different databases.

were not far from the actual center annotated by the ophthalmologist, as shown in Figure 6.5.

However it should be noted that some false detections of the fovea were also obtained. The poor quality in Figure 6.6a and the irregular background in Figure 6.6b make the fovea unclear. In Figure 6.6c we have the same example as in Figure 6.4b. Since the OD detection was far from the ground truth, the subsequent fovea detection was incorrect. However, our method succeeded in detecting the fovea center in several complicated examples. In Figures 6.7a to 6.7b, correct detections over poor quality images are shown. In Figures 6.7c to 6.7d, large dark areas hide the fovea boundaries. Finally, in Figures 6.7e to 6.7h we can see that

(a)   (b)
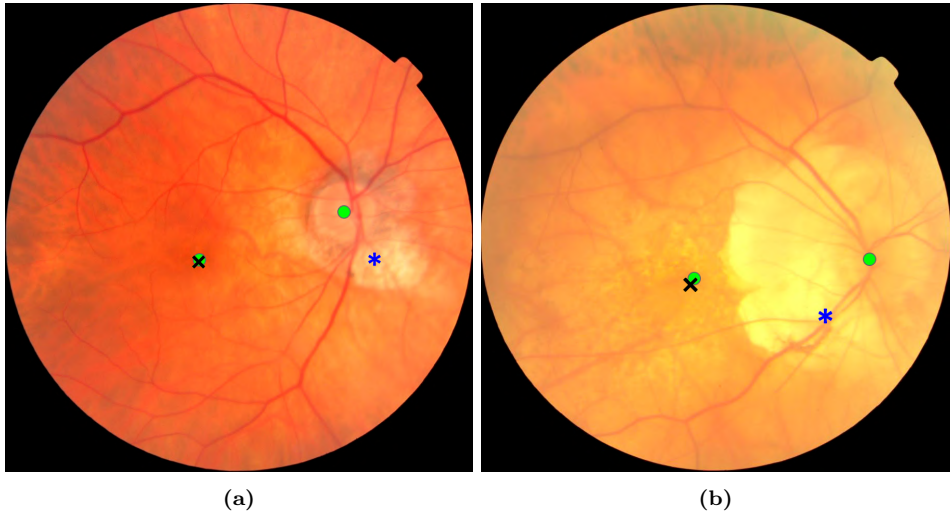
Figure 6.5: Images where the OD is incorrectly detected. However, since the OD detection is close to the real center, the automatic detection of the fovea is correct. The green circles represent the ground truths, the blue mark represents the automatic detection of the OD and the black mark represents the automatic detection of the fovea.

our method properly locates the fovea even in the presence of different types of lesions.

The proposed methods for the OD and fovea location have proved effective and robust. The novel algorithm for the retinal background extraction is valuable for fundus image processing and the proposed saliency maps showed a great generalization and interpretation ability. These reasons bring the clinical viability of the proposed solution closer. Since the OD and fovea locations are crucial for the detection of multiple diseases, such as glaucoma, age-related macular degeneration

Table 6.3: Results for the automatic location of the fovea.

Study	DRIVE	DIARETDB1	Messidor
Welfer et al. (2011)	100	92.13	-
Qureshi et al. (2012)	91.73	98.74	-
Gegundez-Arias et al. (2013)	-	-	96.92
Giachetti et al. (2013)	-	-	99.10
Aquino (2014)	-	94.38	98.24
GeethaRamani and Balasubramanian (2018)	100	97.75	99.33
Al-Bander et al. (2018)	-	-	96.60
Proposed method	**100**	**100**	**99.67**

(a)   **(b)**   **(c)**
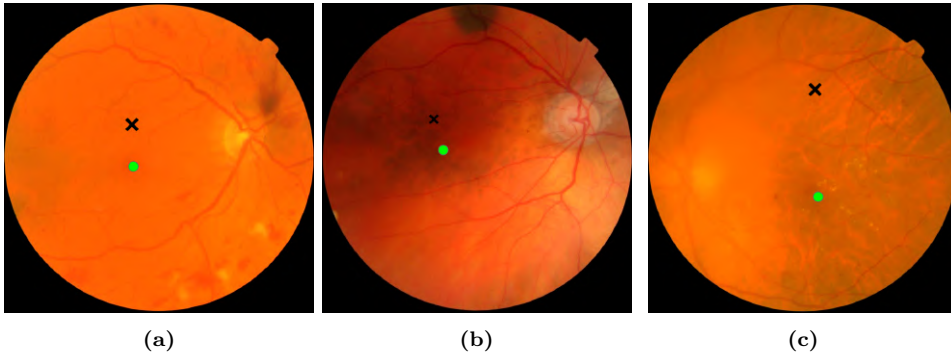
Figure 6.6: Examples where the fovea is incorrectly detected. The green circle repre-
sents the ground truth and the black mark represents the automatic detection. (a) Poor
quality image. (b) Image with irregular background. (c) Image where the fovea detection
is incorrect because of the previous incorrect detection of the OD (see Figure 6.3b).

(a)   **(b)**   **(c)**   **(d)**

(e)   **(f)**   **(g)**   **(h)**
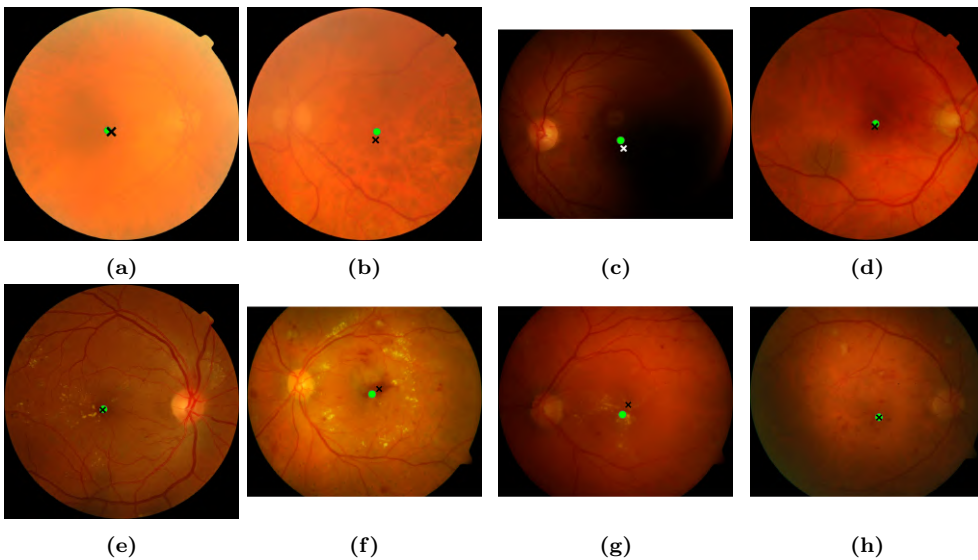
Figure 6.7: Examples where the fovea is successfully detected. The green circle rep-
resents the ground truth and the black mark represents the automatic detection. (a-b)
Poor quality images. (c-d) Images where the fovea boundaries are hidden by large dark
areas. (e-h) Images showing different types of lesions. The different aspect ratio of the
example images is due to the fact that they belong to different databases.

and DR, the proposed method could be useful in different CAD systems. This has
a great impact in the context of visual care.

6.3 Detection of red lesions

6.3.1 Method based on the classification of superpixels

In this study, we proposed an automatic method to detect RLs in fundus images based on superpixel classification. Firstly, the preprocessing stage detailed in section 4.2 was carried out to normalize the inter-image and intra-image appearance as well as to enhance the retinal structures. Secondly, the potential RL candidates were segmented. For this task, dark pixel detection was performed. Then, we separated the different elements of the image on the basis of their edges using the ERS method. RL candidates were later combined to reduce inaccurate fragmentation within structures. A set of 39 features related to shape, pixel intensity or structure distances was extracted from each superpixel. Then, a subset of 16 features was selected based on their relevancy and redundancy using the FCBF method. Finally, each candidate was classified by an MLP to yield the final segmentation of RLs.

The detection of RLs in fundus images has been addressed in previous studies. Different methods can be found in the literature, including mathematical morphology, region growing, wavelet, pixel classification and hybrid approaches (Mookiah et al., 2013). More recent studies used flooding (Seoud et al., 2016) and deep learning (Orlando et al., 2018) techniques. Commonly, these algorithms start by segmenting the normal anatomical structures. Particularly, the blood vessels are generally extracted since they can be easily detected as RLs (Abramoff et al., 2010). An important contribution of this study is that non-RL regions are eliminated during the classification stage, avoiding the need for an independent method for blood vessel segmentation. Additionally, almost all of the previously proposed methods consider pixels as the basic unit in the image. However, superpixels are more consistent with human visual cognition. Unlike pixels, the superpixels provide a novel and meaningful representation of the natural entities of the image (Zhou et al., 2017a). Superpixel segmentation for the detection of RLs in retinal images has been addressed in previous works since it can provide a significant representation of these lesions in the image (Romero Oraá et al., 2018; Zhou et al., 2017a). In these studies, the SLIC method was chosen to segment image superpixels. However, this algorithm has important drawbacks, producing some under-segmented superpixels and poorer accuracy for border segmentation (Xie et al., 2019). Therefore, we have followed a different approach to extract the superpixels in the image, using the ERS algorithm. To the best of our knowledge, this algorithm has never been applied in retinal image processing. We have verified

that the segmentation accuracy improves when comparing the results of the ERS with the SLIC method (Romero Oraá et al., 2018) on the images in our database. In order to improve the performance of the method, we have included a novel improvement to the ERS by combining the superpixels that belong to the same structure in order to reduce the number of candidates for the classification stage.

The proposed method was evaluated on a set of 283 fundus images, of which 120 presented RLs. Therefore, 163 images belonged to retinas lacking RLs. The images in our database showed a wide variability in terms of color, luminosity, contrast and varying levels of quality. Furthermore, different degrees of DR severity could be found and RLs were very variable in terms of appearance and size. The results obtained were measured with two different criteria. With the pixel-oriented criterion, a SE_p of 81.43% and a PPV_p of 86.59% were reached. With the image-oriented criterion, we obtained $SE_i = 84.04\%$, $SP_i = 85.00\%$ and $ACC_i = 84.45\%$. These results are similar to those of other studies for RL detection according to the image-oriented criterion, as shown in Table 6.4. However, comparisons should be made with caution since the databases and the performance measures vary among studies.

In the work of Seoud et al. (2016) three different databases were used to evaluate their method in a per-image basis. Regarding Messidor database and the private CARA1006 database, they reached high SE_i but an SP_i of 50.00%. In the Erlangen database, the method obtained both a SE_i and a SP_i of 93.30%. However, this database was composed of only 45 images. García et al. (2010) tested

Table 6.4: Performance comparison of some methods for the detection of RLs in fundus images according to the image-oriented criterion.

Authors	Database	Nb. im.	SE_i	SP_i
Seoud et al. (2016)	Messidor	1200	93.90%	50.00%
Seoud et al. (2016)	Erlangen	45	93.30%	93.03%
Seoud et al. (2016)	CARA1006	1006	96.10%	50.00%
García et al. (2010)	Private	115	100%	56.00%
Zhou et al. (2017a)	DIARETDB1	89	83.30%	97.30%
Orlando et al. (2018)	Messidor	1200	91.10%	50.00%
Roychowdhury et al. (2012)	DIARETDB1	89	75.50%	93.73%
Sánchez et al. (2011)	Messidor	1200	92.20%	50.00%
Niemeijer et al. (2005)	Private	100	100%	87.00%
Grisan and Ruggeri (2005)	Private	260	71.00%	99.00%
Proposed method	**Private**	**564**	**84.04%**	**85.00%**
Proposed method	**DIARETDB1**	**89**	**84.00%**	**88.89%**

their proposal in 65 fundus images from a private database. They obtained a SE_i of 100% but an unbalanced SP_i of 56.00%. In the study where SLIC superpixel segmentation was used, the proposed method reached $SE_i = 83.30\%$ and $SP_i = 97.30\%$ (Zhou et al., 2017a). However, it was evaluated over only 49 images from DIARETDB1. Orlando et al. (Orlando et al., 2018) used the Messidor data set to evaluate their method. Results showed a $SE_i = 91.10\%$ when $SP_i = 50.00\%$. In the work of Roychowdhury et al. (2012), the proposed method was evaluated using the DIARETDB1 database, obtaining $SE_i = 75.50\%$ and $SP_i = 93.73\%$. Sánchez et al. (2011) validated their method using the Messidor data set. They reached $SP_i = 92.20\%$ and $SP_i = 50.00\%$. Niemeijer et al. (2005) obtained $SP_i = 100\%$ and $SP_i = 93.73\%$ with pixel classification. However, they used a private database of only 100 images. The approach proposed by Grisan and Ruggeri (2005) showed $SP_i = 99.00\%$ yet $SE_i = 71.00\%$. Moreover, they used a private data set. Since our method has also been assessed on the test set of the public database DIARETDB1, a direct comparison with the methods proposed by Roychowdhury et al. (2012) and Zhou et al. (2017a) on the same database can be made. Roychowdhury et al. (2012) obtained $SE_i = 75.50\%$, lower than that obtained with our proposal ($SE_i = 84.00\%$). However, they obtained $SP_i = 93.73\%$, which is higher than the results in our study ($SP_i = 88.89\%$). Similar results were obtained in the study by Zhou et al. (2017a). Their SE_i reached $SE_i = 83.30\%$, which is slightly lower than in our approach. In addition, their SP_i also improves our results, reaching $SP_i = 97.30\%$. When comparing our results with previous approaches, it should be noted that the proposed method has been configured and trained using only images from our private database. These images differ from the images of the public database DIARETDB1 in multiple aspects. Firstly, the images in our database have a higher resolution. Secondly, they have been captured using a different protocol. Additionally, the FOV in the images of our private database is 45 degree, while the FOV in the images of DIARETDB1 is 50 degree. They also were selected with different quality criteria (Kauppi et al., 2007). In spite of these differences, our results on the test set of DIARETDB1 were reasonably good, which confirms the robustness of the proposed method. It is also important to point out, for global comparisons, that our private database consisted of 564 fundus images with a resolution of 1956x1934 pixels. The performance criteria are also heterogeneous among studies. This is especially relevant when a region-based or pixel-based criterion is used. When comparing to the ophthalmologist annotations, the detected regions may overlap completely, partially or not at all with any RL. Thus, there is ambiguity between studies in considering

whether a RL is correctly detected. In this sense, individual pixels belonging to an overlapping region could be sometimes considered as RLs or not depending on the criteria of the study.

The visual inspection of the results on the test images showed that the proposed approach adjusted considerably well to the edges of the true RLs (see Figure 5.4). However, in certain images, some regions were wrongly detected as RLs. In Figure 6.8a, we can see a fundus image corresponding to a healthy retina according to the expert annotations. Figure 6.8b shows the regions that the proposed method detects as RLs. In this image, the background shows a number of strongly visible choroidal vessels as well as a non-uniform texture. As a consequence, many superpixels were considered as RL candidates and our method fails to eliminate all of them as non-RLs.

The proposed method deals with superpixels instead of pixels to identify the entities of the image. From a clinical point of view, we believe that the concept of superpixel is suitable to separate the different structures of the image. Moreover, the extracted features over the superpixels are easily explainable. Although the algorithm does not work well in the presence of choroidal vessels, our results suggest that the proposed method could be useful for the detection of RLs in retinal images.

(a)   (b)
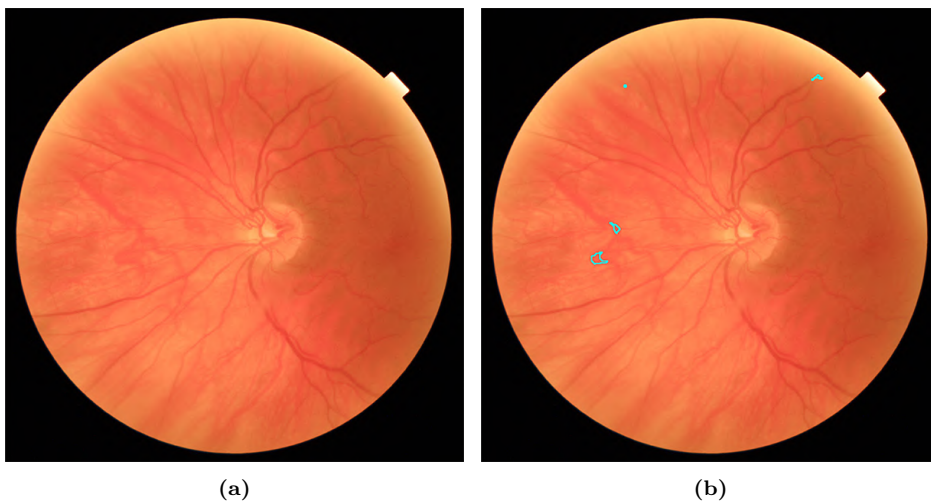
Figure 6.8: Fundus image example corresponding to a healthy retina. (**a**) original image; (**b**) wrongly detected RLs over the original image.

6.3.2 Method based on the decomposition of the fundus image into layers

This method was proposed for the joint detection of RLs and EX. In this section, we focus on RLs. The fundus image was decomposed into various layers representing different structures of the retina, which is the main contribution of this paper. Among these layers, the lesion candidates, the choroidal vasculature visible in tigroid retinas, and the reflective features were included, having proved useful for the classification of RLs. To the best of our knowledge, there are no previous studies that have decomposed the image separating the relevant retinal structures such as the choroidal vessels and the reflective features to detect RLs.

The proposed method for retinal lesion detection was evaluated on a set of 283 fundus images. Among them, 196 showed RLs. The private database was very heterogeneous, showing variations in color, luminosity, contrast, and quality among images. In the same way, variable lesions in terms of appearance and size could be found. The method was also evaluated on the test set of the public database DIARETDB1, composed of 61 images. The results for the detection of RLs were measured using a pixel-based criterion and an image-based criterion (see section 5.3.1). Results can be seen in Table 5.3.

All of these results can be compared to those obtained in previous studies according to the image-based criterion, as shown in Tables 6.5. However, comparisons should be made with caution, since the databases and the performance measures vary among studies. We have found four methods that have been evaluated using the DIARETDB1 database in order to establish a direct comparison with the proposed method for RL detection. Jaafar et al. (2011) obtained a high $SE_i = 98.80\%$, yet a $SP_i = 86.20\%$ lower than ours ($SP_i = 91.67\%$). In addition, they tested their method using the database DIARETDB0 together with DIARETDB1. In the work of Roychowdhury et al. (2012), they obtained $SP_i = 93.73\%$, but the $SE_i = 75.50\%$ was lower. In (Zhou et al., 2017a), $SP_i = 91.67\%$ was obtained. However, our value of SE_i (88.00%) improves their SE_i (83.30%). Table 6.5 also shows that the proposed method achieves better results than our previous work based on superpixels (Romero-Oraá et al., 2019). For this reason, the subsequent detection of EXs was carried out using this approach.

The method has been developed based on the indications provided by the ophthalmologists to identify the different structures of the retina. For this reason, the interpretability is straightforward and clear. This allowed us to study the existing relation between the RLs and other structures, such as the choroidal

Table 6.5: Comparison of some methods for red lesion detection.

Authors	Database	Nb. im.	SE_i	SP_i
Jaafar et al. (2011)	89	DIARETDB1	98.80%	86.20%
Roychowdhury et al. (2012)	DIARETDB1	89	75.50%	93.73%
Zhou et al. (2017a)	DIARETDB1	89	83.30%	97.30%
Romero-Oraá et al. (2019)	**DIARETDB1**	**89**	**84.00%**	**88.89%**
García et al. (2010)	Private	115	100%	56.00%
Niemeijer et al. (2005)	Private	100	100%	87.00%
Grisan and Ruggeri (2005)	Private	260	71.00%	99.00%
Seoud et al. (2016)	Messidor	1200	93.90%	50.00%
Orlando et al. (2018)	Messidor	1200	91.10%	50.00%
Sánchez et al. (2011)	Messidor	1200	92.20%	50.00%
Proposed method	**Private**	**564**	**85.00%**	**90.80%**
Proposed method	**DIARETDB1**	**89**	**88.00%**	**91.67%**

vessels and the reflective features. The extracted layers using our decomposition method have proven to be useful to detect RLs. With further validation, the proposed method could be used as part of an automatic DR screening system. Thus, it could be a diagnostic aid for the early detection of DR, reducing the workload of specialists and improving the management of diabetic patients.

6.4 Detection of hard exudates

As mentioned in the previous section, the detection of EXs was conducted using the same methodology as for RLs. However, in this section we focus on the discussion about EX detection. The fundus image was decomposed into various layers including the EX candidates, the choroidal vasculature visible in tigroid retinas, and the reflective features, which have proved useful for the classification of EXs. To the best of our knowledge, no previous studies have followed a similar approach.

The proposed method was evaluated on a set of 283 fundus images, of which 168 showed EXs. The method was also evaluated on the test set of the public database DIARETDB1, composed of 61 images. The results for the detection of EXs were measured using a pixel-based criterion and an image-based criterion (see section 5.3.1). Results can be seen in Table 5.4.

We have also found several methods aimed at EX detection that have been assessed using the DIARETDB1 database (Table 6.6). Walter et al. (2002) obtained $SE_i = 86.00\%$ and $SP_i = 69.00\%$. In (Harangi and Hajdu, 2014), values of $SE_i = 92.00\%$ and $SP_i = 68.00\%$ were obtained. Liu et al. (2016) achieved

Table 6.6: Comparison of some methods for exudate detection.

Authors	Database	Nb. im.	SE_i	SP_i
Walter et al. (2002)	DIARETDB1	89	86.00	69.00
Harangi and Hajdu (2015)	DIARETDB1	89	92.00	68.00
Liu et al. (2016)	DIARETDB1	89	83.00	75.00
Zhou et al. (2017b)	DIARETDB1	89	88.00	95.00
Kaur and Mittal (2018)	DIARETDB1	89	91.00	94.00
Adem (2018)	DIARETDB1	89	99.20	97.97
Proposed method	**Private**	**564**	**88.04%**	**98.95%**
Proposed method	**DIARETDB1**	**89**	**95.00%**	**90.24%**

$SE_i = 83.00\%$ and $SP_i = 75.00\%$. The method proposed in (Zhou et al., 2017b) showed a $SE_i = 88.00\%$ and $SP_i = 95.00\%$. Kaur and Mittal (2018) obtained $SE_i = 91.00\%$ and $SP_i = 94.00\%$. Finally, the work of Adem (2018) presented high values of $SE_i = 99.20\%$ and $SP_i = 97.97\%$. The value of SE_i achieved with our method (95.00%) is higher than those obtained in previous studies, with only one exception (Adem, 2018). It should be noted that the test set in (Adem, 2018) was composed of images from DIARETDB0 and DRIMDB databases in addition to DIARETDB1. Moreover, the training set of the DIARETDB1 database was used in the training phase and the appearance of these images is similar to the ones in the test set of the same database. Therefore, this could facilitate EX detection, leading to better results but poor robustness. When comparing our results with previous approaches, it should be noted that the proposed method has been developed using only images from the private database. This dataset is different from the public database DIARETDB1 in several aspects. Firstly, the images in our database have a higher resolution. Secondly, they have been captured using a different protocol. Thirdly, the FOV in the images of the proprietary database is $45°$, while the FOV in the images of DIARETDB1 is $50°$. In addition, they were selected considering different quality criteria (Kauppi et al., 2007). In spite of these differences, the results on the test set of DIARETDB1 are comparable to those obtained in previous studies, which proves the robustness of the proposed method.

The method has been developed based on the indications provided by the ophthalmologists to identify the different structures of the retina. For this reason, the interpretability is straightforward and clear. This allowed us to study the existing relation between the EXs and other structures, such as the choroidal vessels and the reflective features. The extracted layers using our decomposition

method have proven to be useful to detect EXs. With further validation, the proposed method could be used as part of an automatic DR screening system. Thus, it could be a diagnostic aid for the early detection of DR, reducing the workload of specialists and improving the management of diabetic patients.

6.5   Diabetic retinopathy severity grading

This study presents a new deep learning framework for DR grading where the proposed attention mechanism stands out, which is separately applied to the bright and dark pixels of the fundus image. The used dataset contained multiple poor-quality images, which makes them unsuitable for medical analysis. Consequently, an image quality assessment stage was required. For this task, we applied the automatic algorithm described in section 4.1.2, discarding 35,729 out of 88,702 images. The number of images discarded by the validated algorithm was very high. On one hand, this means that the overall quality of the dataset is far from optimal. On the other hand, we can assume that the quality distribution of the dataset is close to real clinical scenarios, which makes it a reliable research material. Other authors estimated that 75% of the dataset was gradable (Rakhlin, 2018).

Before using the CNN architecture, an image decomposition process was required to separate the bright and the dark structures of the fundus image. For this task, we applied the multiscale algorithm defined in sections 4.4.2 and 4.5. Despite the great success of deep learning in computer vision, this study shows that traditional image processing methods could still be very useful. They can provide context for the problem at hand and, when combined with the powerful optimization capacity of deep networks, they can improve results and allow more complex problems to be approached. Therefore, the fundamentals of classical methods can be successfully combined with state-of-the-art deep networks. This idea has already been proved in the context of automated fundus image classification by enhancing the performance of CNNs through the use of hand-crafted features (Bogacsovics et al., 2022).

We achieved a QWK of 0.78 on the test set of the Kaggle DR detection dataset. Thus, the proposed method achieves similar QWK to other state-of-the-art methods for DR grading, as shown in Table 6.7. In order to establish a direct comparison, the results shown from the other studies were obtained using the same database and the same metric (QWK). However, the train, validation and test sets could differ among studies and, therefore, the comparisons should be carefully considered. With our approach, we achieved a higher QWK than that obtained by

Table 6.7: Comparison of some methods for DR grading in therms of QWK using the Kaggle DR detection dataset.

Study	QWK
Wang et al. (2017)	0.85
González-Gonzalo et al. (2018)	0.72
de la Torre et al. (2018)	0.72
Krause et al. (2018)	0.84
de la Torre et al. (2020)	0.80
Araújo et al. (2020)	0.74
Proposed method	**0.78**

González-Gonzalo et al. (2018), de la Torre et al. (2018) and the recent study of Araújo et al. (2020). The later work of de la Torre et al. (2020) achieved a higher QWK=0.80 than ours. However, they accomplished the training phase including part of the images meant for testing and, therefore, they tested their method on a reduced subset of 10,000 images. Krause et al. (2018) also outperformed our method in terms of QWK, with a value of 0.84. However, their model was trained on a large, private dataset with more than 1.6 million fundus images, which is unavailable in most studies. Moreover, their test set was composed of 1,818 images, much smaller than ours (32,017). Finally, Wang et al. (2017) reported the highest QWK=0.85. However, their method required the annotated labels of the lesions of some images and the fundus images of both eyes, which are not always available. On the contrary, our model feeds exclusively on image labels.

Analyzing the obtained confusion matrix allows us to extract several conclusions on the results for the proposed method. First, the class 0 was detected with high accuracy: only 2.9% (706 out of 23,962) of those images were over-diagnosed. More importantly, only 0.0005% (12 out of 23,962) of them were rated class 3 or 4. Conversely, the class 1 was easily misguided with the classes 0 and 2, becoming the most misclassified severity degree. On one hand, this is because some images in this category only present tiny MAs that are hard to detect, as shown in Figure 6.9a. It should be noted that this issue could have important diagnostic consequences depending on the evolution of the lesions (Abramoff et al., 2010). Regarding class 2, an acceptable detection accuracy of 77.2% (2,636 out of 3,414 images) was achieved. However, most of the misclassified images were classified as class 0, which should be taken into account for patient management. Under-diagnosing the moderate NPDR degree could lead to a risky situation for the vision if not detected in upcoming screenings (Mookiah et al., 2013). Images of

(a)	(b)	(c)
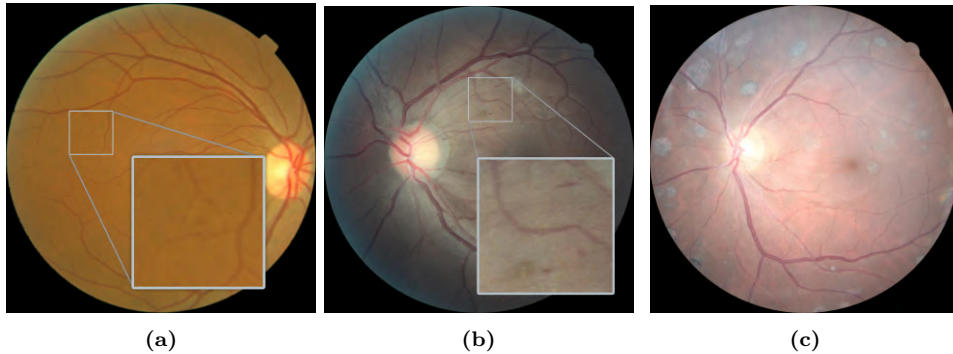
Figure 6.9: Misclassified examples. (a) Class 1 image with a tiny MA on the left part (see zoomed-in area) and misclassified as class 0. (b) Class 3 image misclassified as class 2 due to the similarity of the lesions (see zoomed-in area). (c) Class 4 image with laser marks and misclassified as R3.

class 3 were often misclassified as class 2, as the example in Figure 6.9b. However, only 7.2% (50 out of 694) of the severe NPDR images were classified as any of the other degrees. The incorrect detection of the class 3 as class 2 is not crucial since both degrees would involve a manual examination by a specialist. Finally, poor detection accuracy for class 4 was obtained. As shown in the confusion matrix, 40.2% (193 out of 480) of the class 4 images were diagnosed as class 3. The main reason for this result is that multiple eye fundus images contain photo-coagulation treatment and laser marks, hindering the detection of the characteristic signs of the proliferative DR, such as neovessels and pre-retinal hemorrhages. Figure 6.9c exhibits one of these examples. Nevertheless, laser-treated retinas should never reach screening scenarios since they are already under supervision. It is important to mention that the RIQA stage does not classify this type of images as inadequate quality. Just because they have laser marks does not mean they lack image quality.

As exposed in section 4.6.8, we performed 2 additional ablation studies in order to evaluate the influence of the proposed attention mechanism on the results:

1. In the first experiment, based on a standard architecture without attention mechanism, the network did not converge adequately due to class imbalance. The model learnt to classify every sample as the most dominant class (no DR). This experiment proved that the attention mechanism helps to deal with class imbalance.

2. In the second experiment, the attention mechanism was applied but the

separation of the bright and the dark pixels was not considered. We obtained QWK=0.76. Results were good, yet inferior to the ones obtained with the proposed method. This result shows that the separate optimization of bright and dark regions can improve the results of the classification task.

We propose a novel deep learning framework for DR grading based on an attention mechanism. Unlike previous methods, our approach performs separate attention for the bright and the dark pixels in the retinal image. On the one hand, this separation allows to improve model optimization. On the other hand, dividing the problem makes the model easier to manage. For these reasons, the proposed solution makes a great impact. However, the model performance should be taken into account. As previously mentioned, the proposed approach tends to fail in the classification of classes 1 and 4. Additionally, class 2 tends to be under-diagnosed. This is especially critical when the automatic estimation misclassifies the images as class 0 (no DR), since it would prevent patients with threatened vision from being seen by a doctor, putting their vision at risk. For clinical viability, the detection of referable DR needs to be optimized, assuring that no referable case remains undetected. Finally, the use of XAI techniques is required to provide explanatory capability. Despite the room for improvement, our results suggest that the proposed method could be a diagnostic aid for the early detection of DR. In this way, diabetic patients could receive better attention for their ocular health avoiding vision loss.

6.6 Limitations of the study

Despite the relevance of the contributions presented in this Doctoral Thesis for the automatic diagnosis of DR, some limitations should be pointed out.

One of the main limitations is related to the databases used in our studies. First of all, it would be desirable to increase the number of available images with which to develop the methods and evaluate the results. This is particularly relevant for the studies based on deep learning techniques. Although we have applied data augmentation and transfer learning to alleviate the issue, deep learning techniques usually require large amounts of data. Secondly, it is important to have a great variability of images, captured with different protocols, resolutions, and using diverse fundus cameras. We have used different databases, especially in the study of the location of the OD and the fovea, where four databases were used. However, it would be desirable to have more images with different visual charac-

teristics. Increasing the number and variability of the images would improve the generalization ability of the methods. Another important aspect is that the ground truths provided by human graders are not free from subjectivity. In this regard, it would be interesting to have the opinion of different evaluators to correctly analyze the results obtained with the automatic methods. Nevertheless, we are aware of the difficult challenge of building a large database annotated by multiple experts. Efforts to create an annotated database of these characteristics would be highly beneficial. However, at present, the only way to validate the robustness of the methods is to use public databases, which offer considerable variability but are not exempt from limitations. The lack of appropriate clinical validation is common to all CAD systems developed to date, which conditions their practical utility (Bellemo et al., 2019; Li et al., 2021a).

Regarding the methods developed to assess the quality of fundus images, it should be noted that the two classes of images (adequate and inadequate quality) were quite unbalanced, showing a small number of poor quality images. In this scenario, classifiers tend to get biased towards the prediction of adequate quality images. Therefore, it would be desirable to have a similar number of images for both classes. This way, we can focus our efforts on optimizing the classification without having to deal with the imbalance problem.

The proposed methods for the location of the OD and the fovea have also some specific limitations that should be mentioned. We have assumed the presence of both the OD and the fovea within the FOV, giving an estimation of the centers even if any of the landmarks were not present. In addition, the OD and the fovea must be approximately located at the same vertical level in the fundus image. If a different capture protocol is considered, the proposed method would have to be adapted. Moreover, the method for the automatic fovea location relies on the previous OD detection. Nonetheless, a close estimation of the center of the OD is sufficient.

When it comes to the detection of RLs based on superpixel classification, it should be noted that the ERS algorithm does not always properly manage to isolate the MAs or RLs of small size. The size of the superpixels is a commitment parameter. Small superpixels approximate small lesions better. However, they also become closer to the concept of individual pixels, losing the representation of the entities of the image. A separate detection of MAs and HEs would allow adjusting the parameters of the method more properly.

The study associated with the image decomposition for the joint detection of RLs and EXs has additional limitations. First, segmenting the blood vessels as

an independent stage has some drawbacks. When an RL is detected as part of the vascular network, it is discarded as a possible lesion, regardless of the later stages. Second, various parameters of the method were empirically set using the training images of our private database, so they may not adapt properly to other images with different characteristics. However, the value of these parameters is not critical, and the performance is not significantly affected as long as they are around the selected values. We noticed that small deviations of these values hardly produced changes in the output. Third, the classification stage is based on a set of handcrafted features. Exploring deep learning-based approaches would be desirable to find useful features in an optimized way.

Finally, the proposed algorithm for DR severity grading shows also some limitations that should be mentioned. The image decomposition algorithm previous to the CNN is time-consuming, requiring approximately 3 seconds per image. This time would not be a very important problem in a clinical setting, since the time required for image capture is considerably longer. However, it would be desirable to find a faster algorithm for this task. Another limitation is directly related to model performance. As previously mentioned, the proposed approach tends to fail in the classification of classes 1 and 4. Additionally, class 2 tends to be under-diagnosed. This is especially critical when the automatic estimation misclassifies the images as class 0 (no DR), since it would prevent patients with threatened vision from being seen by a doctor, putting their vision at risk. Therefore, it would be desirable to improve the results by focusing on ensuring that no patient with DR remains undetected.

Chapter 7

Conclusions

The present Doctoral Thesis aims to contribute to the early diagnosis and grading of DR. The studies conducted during this research share a common thread covering all the stages of a complete, automatic CAD system. These include the image quality assessment, the preprocessing stage, the detection of the OD, the fovea and the vasculature, the segmentation of RLs and EXs, and the DR severity grading. Given the clear sequential relationship, the set of proposed methods can be seen as a single framework for DR screening. However, it is important to mention that each of the methods is very useful separately. In fact, several of the algorithms could also be helpful in diagnosing other diseases such as age-related macular degeneration or glaucoma.

This research work has been carried out to overcome some of the limitations of the previous studies found in the literature. Two different approaches have been proposed to automatically assess the quality of fundus images. The first approach allowed us to find a new set of global image features that complement each other satisfactorily. However, the second method showed a superior performance using a deep CNN. This fact justifies the current trend towards deep learning-based approaches in almost any automatic image classification and analysis task. Another study developed during the Thesis was aimed at the automatic location of the OD and the fovea. The proposed method demonstrated great robustness with different databases. The superpixel classification-based approach to detect RLs also showed interesting results. The ERS technique made it possible to separate the natural entities from the image and thus facilitate the segmentation of the lesions. The study devoted to the joint detection of RLs and EXs proved the usefulness of novel image indicators, such as the reflective features of the retina

and the choroidal vasculature visible in tigroid retinas. Finally, the study focused on the severity grading provided a novel attention mechanism which contribute to increase the diagnostic ability DR.

This chapter gathers the original contributions of the present Doctoral Thesis in section 7.1. Section 7.2 states the main conclusions that can be extracted from all the studies carried out. Finally, future research lines are outlined in section 7.3.

7.1 Contributions

The main contributions obtained from the present research work are described below:

1) Creation of a private database of fundus images with different quality levels and types of DR-related lesions. A total number of 2773 images were captured. Two experienced ophthalmologists manually annotated the quality level, the centers of the OD and the fovea, and the presence of RLs and EXs.

2) New generic features extracted from fundus images for automatic quality assessment (Jiménez-García et al., 2019). These are related to SSEQ, which has never been applied to this task, NIQE, sharpness and luminosity. The combination of the mentioned features proved to provide relevant and complementary information.

3) Novel deep learning model based on a InceptionResNetV2 CNN and transfer learning to automatically assess the quality in fundus images (Romero Oraá et al., 2020). To the best of our knowledge, the mentioned architecture has never been used with this purpose.

4) Novel preprocessing stage to normalize the inter-image and intra-image appearance in fundus images as well as to enhance the retinal structures (Romero-Oraá et al., 2019). It was based on five sequential operations: bright border artifact removal, background extension, illumination and color equalization, denoising, and contrast enhancement.

5) Retinal background extraction based on region-growing (Romero-Oraá et al., 2020). The estimation of the retinal background is valuable to isolate the structures of interest, such as lesions or anatomical landmarks. In addition to this estimation, we provided a background version which exclusively preserves the bright structures and another version which exclusively preserves the dark structures.

6) Development and combination of novel saliency maps with direct interpretability to create a robust method for the automatic location of the OD and the fovea in fundus images (Romero-Oraá et al., 2020).

7) The use of the ERS algorithm for RL detection (Romero-Oraá et al., 2019). Although superpixel segmentation has been slightly explored in fundus image analysis, the mentioned algorithm and its advantages have never been taken into account.

8) Novel algorithm for the segmentation of RLs and EXs based on the decomposition of the fundus image into various layers (Romero-Oraá et al., 2020). For the first time, the choroidal vasculature visible in tigroid retinas and the reflective features of the retina were considered to aid in the classification of DR-related lesions. The proposed method was founded on the clues observed by the ophthalmologists to identify the different structures of the retina, which allowed us to study the existing relation between the lesions and other retinal structures.

9) End-to-end deep learning framework for automatic DR grading. Our approach is based on a novel attention mechanism which performs a separate attention of the dark and the bright structures of the retina. The proposed method integrates a multiscale algorithm, based on traditional image processing techniques, into a deep neural network architecture.

7.2 Main conclusions

The comprehensive analysis and discussion of the results obtained in the studies conducted during the course of this Doctoral Thesis lead to extract the following conclusions:

1) The feature selection analysis performed using the FCBF technique allowed us to conclude that the SSEQ and NIQE methods are useful to assess retinal image quality, and can be complementary with sharpness and luminosity features.

2) Deep learning approaches are outperforming the traditional techniques for many image processing problems and, in particular, for fundus image quality assessment. In this context, the architecture InceptionResNetV2 together with transfer learning has allowed to achieve great results and a high generalization ability without the need for an excessively large database.

3) The saliency maps proposed for the automatic location of the OD and fovea have shown a high accuracy even in the images where those structures did not have the standard appearance. The evaluation over various databases has proven that the proposed method is very robust and effective.

4) The decomposition of the fundus image in various layers has allowed to extract a set of features relevant for the detection of RLs and EXs, as analyzed using the FCBF technique. In this decomposition, the choroidal vasculature visible in tigroid retinas and the reflective features of the retina have demonstrated to be useful.

5) The concept of the superpixel, as opposed to the pixel, is suitable to represent the natural entities and separate the different structures of the retina. In particular, the ERS algorithm has proven fast and effective for RL segmentation, outperforming the SLIC method.

6) The proposed deep learning framework for DR grading achieved results comparable to previously published methods. The separate attention for the bright and the dark pixels in the retinal image allowed to improve model optimization and made it easier to manage.

7) The results for any of the proposed methods in this Doctoral Thesis suggest that they could be used, separately or together, as part of an automatic DR screening system, being a diagnostic aid for the early detection of DR. In this way, diabetic patients could receive better attention for their ocular health avoiding vision loss.

7.3   Future research lines

Despite the numerous contributions that have emerged from this Doctoral Thesis, much research work remains to be done in the future to improve diagnostic aid in DR. Further efforts must be done to care for the eye health of diabetic patients. The most interesting future research lines are discussed below.

In this research work, a relatively large private database of fundus images has been built. However, the classes are unbalanced and more images are needed to provide more generalizable results. In addition, it would be desirable to have images captured with different protocols and cameras to evaluate the robustness of the methods. In this regard, several public databases are available in the literature. However, they are usually small databases and obtained with old cameras.

Larger databases have some important limitations. For these reasons, it would be important in the future to increase the number of images in the database, covering a greater variety of patients, cameras and capture protocols.

Performing *t*-Distributed Stochastic Neighbor Embedding (*t*-SNE) analysis in our studies is another future line to consider. This technique allows to visualize high-dimensional data by giving each datapoint a location in a two or three-dimensional map (Van Der Maaten and Hinton, 2008). *t*-SNE is capable of capturing much of the local structure of the high-dimensional data very well, while also revealing global structure such as the presence of clusters at several scales (Van Der Maaten and Hinton, 2008). In some of our studies, where feature selection is involved, this type of analysis would be beneficial to show the discriminative power of the selected features.

Deep learning is revolutionizing artificial vision in recent years. Automatic analysis of fundus images is already benefiting from the enormous potential of this field. However, it is experiencing rapid growth and much remains to be explored. In this context, the development of new, more advanced deep learning architectures is definitely the most important future line in the coming years. Among these architectures, vision transformers are a good example of innovative solutions that can bring great results.

Another interesting future research line would be to combine the most successful traditional methods with deep learning architectures. Traditional methods are more easily interpretable and have been developed based on clinical indicators, so they are more similar to human diagnosis. In this way, retinal image analysis problems could be solved from a clinical perspective yet taking advantage of the great optimization capacity of deep networks.

To deal with the lack of interpretability characterizing the deep architectures (the well-known "black box" problem), another interesting future research line could be the application of XAI techniques. They would make it possible to explain automatic decisions and discover unknown patterns that influence diagnosis. In the context of fundus images, we could detect the relevant pixels and check if the segmentation/classification task is misguided by artifacts or undetected lesions, for instance. Understanding the decisions of the automatic model would also allow to solve potential issues and improve the performance of the method.

Along with fundus images, OCT images are the diagnostic method for DR used in clinical practice. In fact, some signs of this disease can be detected earlier with this imaging modality. Therefore, a future line of research that combines fundus image analysis with OCT would be very interesting.

Appendix A

Papers included in this Doctoral Thesis

Combination of Global Features for the Automatic Quality Assessment of Retinal Images
Entropy Rate Superpixel Classification for Automatic Red Lesion Detection in Fundus Images
A Robust Method for the Automatic Location of the Optic disc and the Fovea in Fundus Images
Effective Fundus Image Decomposition for the Detection of Red Lesions and Hard Exudates to Aid in the Diagnosis of Diabetic Retinopathy

Appendix B

Scientific achievements

B.1 Publications

B.1.1 Papers indexed in the JCR

1. Jorge Jiménez-García, **Roberto Romero-Oraá**, María García, María I. López, Roberto Hornero, "Combination of Global Features for the Automatic Quality Assessment of Retinal Images", *Entropy*, vol. 21 (3), pp. 311, March, 2019, DOI: .

2. **Roberto Romero-Oraá**, Jorge Jiménez-García, María García, María I. López, Javier Oraá-Pérez, Roberto Hornero, "Entropy Rate Superpixel Classification for Automatic Red Lesion Detection in Fundus Images", *Entropy*, vol. 21 (4), pp. 417, April, 2019, DOI: .

3. **Roberto Romero-Oraá**, María García, Javier Oraá-Pérez, María I. López, Roberto Hornero, "A robust method for the automatic location of the optic disc and the fovea in fundus images", *Computer Methods and Programs in Biomedicine*, vol. 196, pp. 105599, June, 2020, DOI: .

4. **Roberto Romero-Oraá**, María García, Javier Oraá-Pérez, María I. López, Roberto Hornero, "Effective Fundus Image Decomposition for the Detection of Red Lesions and Hard Exudates to Aid in the Diagnosis of Diabetic Retinopathy", *Sensors*, vol. 20 (22), pp. 6549, November, 2020, DOI: .

5. **Roberto Romero-Oraá**, María García, María I. López, Roberto Hornero, "Attention-based Deep Learning Framework for Automatic Fundus Image

Processing to Aid in Diabetic Retinopathy Grading", *Computer Methods and Programs in Biomedicine, Under review.*

B.1.2 International conferences

1. **Roberto Romero-Oraá**, María García, Javier Oraá-Pérez, María I. López, Roberto Hornero, "Automatic fundus image quality assessment: diagnostic accuracy in clinical practice", *The Association for Research in Vision and Ophthalmology Annual Meeting (ARVO 2020)*, pp. 2033, Baltimore (United States), May 3 - May 7, 2020.

B.1.3 National conferences

1. **Roberto Romero-Oraá**, María García, María I. López, Félix Manco Lavado, Roberto Hornero, "Localización automática de la papila y la fóvea en retinografías", *XXXIV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2016)*, ISBN: 978-84-9048-531-6, pp. 173-176, Valencia (Spain), November 23 - November 25, 2016.

2. **Roberto Romero-Oraá**, María García, Jorge Jiménez-García, María I. López, Roberto Hornero, "Detección de lesiones rojizas en imágenes de fondo de ojo aplicando diferencias de color en el espacio CIELAB", *XXXV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2017)*, ISBN: 978-84-9082-797-0, pp. 525-528, Bilbao (Spain), November 29 - December 1, 2017.

3. Verónica Barroso-García, Gonzalo C. Gutiérrez-Tobal, Leila Kheirandish-Gozal, Daniel Álvarez, Fernando Vaquerizo-Villar, **Roberto Romero-Oraá**, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, "Análisis de diferencias de segundo orden aplicado a la señal de flujo aéreo monocanal para la ayuda al diagnóstico del síndrome de la apnea-hipopnea del sueño en niños", *XXXV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2017)*, ISBN: 978-84-9082-797-0, pp. 481-484, Bilbao (Spain), November 29 - December 1, 2017.

4. Fernando Vaquerizo-Villar, Daniel Álvarez, Leila Kheirandish-Gozal, Gonzalo C. Gutiérrez-Tobal, Verónica Barroso-García, **Roberto Romero-Oraá**, Andrea Crespo, Félix del Campo, David Gozal, Roberto Hornero, "Análisis de fluctuaciones sin tendencias (DFA) en los registros de oximetría

para la ayuda en el diagnóstico del síndrome de la apnea-hipopnea del sueño infantil", *XXXV Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2017)*, ISBN: 978-84-9082-797-0, pp. 209-212, Bilbao (Spain), November 29 - December 1, 2017.

5. Jorge Jiménez-García, **Roberto Romero-Oraá**, María García, María I. López, Roberto Hornero, "Evaluación automática de la calidad en retinografías mediante clasificación de características globales de imágenes", *XXXVI Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2018)*, ISBN: 978-84-09-06253-9, pp. 29-32, Ciudad Real (Spain), November 21 - November 23, 2018.

6. **Roberto Romero-Oraá**, María García, Jorge Jiménez-García, María I. López, Roberto Hornero, "Clasificación de superpíxeles para la detección automática de lesiones rojizas en imágenes de fondo de ojo", *XXXVI Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2018)*, ISBN: 978-84-09-06253-9, pp. 25-28, Ciudad Real (Spain), November 21 - November 23, 2018.

7. **Roberto Romero-Oraá**, María García, Javier Oraá-Pérez, María I. López, Roberto Hornero, "Transfer learning para evaluar de forma automática la calidad en imágenes de fondo de ojo", *XXXVII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2019)*, ISBN: 978-84-09-16707-4, pp. 175-178, Santander (Spain), November 27 - November 29, 2019.

8. **Roberto Romero-Oraá**, Cristina Pinar Muñoz-Zamarro, María García, Javier Oraá-Pérez, María I. López, Roberto Hornero, "Detección automática de patología en imágenes de fondo de ojo utilizando técnicas de deep learning", *XXXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica (CASEIB 2020)*, ISBN: 978-84-09-25491-0, pp. 484-487, Valladolid (Spain), November 25 - November 27, 2020.

B.2 International internship

Three-month research internship at Biomedical Imaging Lab at the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC) of the University of Porto, Portugal.

i. **Purpose of the internship**

The main objective of the research stay was to deepen into advanced methods of fundus image processing to aid in the diagnosis of DR. In order to achieve this general objective, the following specific objectives were proposed: (i) Investigate the latest deep learning techniques for the analysis of medical images and, in particular, fundus images. (ii) Propose a study to help diagnose diabetic retinopathy. (iii) Develop and evaluate an automatic fundus image analysis method. Another important purpose of the stay was to begin a significant collaboration between the Biomedical Engineering Group of the University of Valladolid and prestigious international institutions such as the Institute for Systems and Computer Engineering, Technology and Science of the University of Porto. This cooperation will not only improve the quality of the future research, but will also favor potential collaborations in international research projects.

ii. **Quality indicators of the institution**

The University of Porto was founded in 1911, although its origins date back to the 18th century. It is the largest university in Portugal and is among the top 300 universities in the world according to the Academic Ranking of World Universities. It currently has fourteen faculties, a Business School and more than 70 scientific research units that serve nearly 28,000 students, 2,300 teachers and researchers, and 1,700 non-teaching staff. This university is recognized internationally for the scientific research it produces. Its laboratories subscribe to more than a fifth of the Portuguese scientific articles. In addition, the University of Porto has cooperation protocols with more than 500 foreign universities, so it has extensive experience in hosting students through mobility programs.

C-BER is one of the research centers of the Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), which currently has more than 700 researchers and is directly associated with the University of Porto. The Biomedical Imaging Lab is one of three C-BER laboratories. The main focus of the laboratory is the development of advanced image processing and analysis methodologies, particularly medical and biological images, with the aim of creating computer-aided diagnostic tools to support medical decision-making. Among his lines of research, it is worth highlighting those related to the processing of ophthalmological images. In this sense, it has three important active projects (Screen-DR, RetinaCAD and Choroid-

CAD) and numerous publications of great impact.

The supervisor during the stay was Dr. Ana María Mendonça, a researcher at the Biomedical Imaging Lab of INESC TEC who obtained a PhD in Electrical Engineering from the Faculty of Engineering of the University of Porto (FEUP) in 1994. She is currently an associate professor in the Department of Engineering Electrical and Computer Science (DEEC) of FEUP. From 2006 to 2014, she was a member of the Executive Board of DEEC, where she was Deputy Director from 2010 to 2014. She was also director of the Master in Biomedical Engineering and is now a member of the Scientific Committee of the Doctoral Program in Biomedical Engineering. Since October 2014, she has been Vice-Dean of FEUP. Ana Maria Mendonça was a researcher at the Institute of Biomedical Engineering (INEB) from 1989 to 2014, where she was also a member of the Board of Directors from 2002 to 2014 and President of the Board from 2012 to 2014. Her research work has focused mainly on the development of medical image analysis methods with the aim of extracting essential information to aid diagnosis. Thus, her research on retinal pathologies stands out, having published numerous articles in indexed journals and international conferences, currently reaching an h-index of 20.

B.3 Awards and honors

05/2017: **Prize to the best poster** at "III Jornada de Investigación en Tecnologías de la Información y las Telecomunicaciones" carried out by **Roberto Romero Oraá**.

12/2017: Prize to the best poster at "II Jornada de doctorandos en Investigación Biomédica' awarded by the 'Instituto de Biología y Genética Molecular (IBGM)" and carried out by **Roberto Romero Oraá**.

11/2019: **Winner of the Three Minute Thesis (3MT®) competition** for the work "Avoiding blindness: diabetic retinopathy under control", awarded by the Doctoral School of the University of Valladolid (EsDUVa). **Roberto Romero Oraá**.

Apéndice C

Resumen en castellano

C.1   Introducción

La diabetes mellitus (DM) es una enfermedad grave, sin cura y de gran impacto a nivel mundial que se ha convertido en una de las diez causas de muerte más frecuentes en el mundo (Saeedi et al., 2019). Se trata de un trastorno metabólico caracterizado por la hiperglucemia, es decir, la presencia de niveles elevados de glucosa en sangre, como resultado de la producción insuficiente o anormal de insulina. La Federación Internacional de Diabetes (FID) estima que la población mundial con DM es de 463 millones y calcula que será de 700 millones para 2045 (Saeedi et al., 2019). Además, el rápido envejecimiento de la población mundial, el aumento de la esperanza de vida de las personas y los cambios en el estilo de vida conducen a un mayor riesgo de DM (Teo et al., 2021). El progreso de la DM implica daños a largo plazo, disfunción e insuficiencia de diferentes órganos, especialmente los ojos, los riñones, los nervios, el corazón y los vasos sanguíneos (American Diabetes Association, 2014). La más común de estas complicaciones es la retinopatía diabética (RD), una de las principales causas de ceguera evitable en la población adulta con edad de trabajar (Teo et al., 2021). Se estima que la prevalencia de la RD es del 22.27 % a nivel mundial dentro de la población con DM (Teo et al., 2021).

Muchas enfermedades oculares y otras relacionadas con la circulación sanguínea se manifiestan inicialmente en la retina. Entre las enfermedades oculares se incluyen la degeneración macular, el glaucoma y la RD, que son las causas más importantes de ceguera en el mundo desarrollado. En el caso de la RD, se sabe que la hiperglucemia, fruto de la DM, daña las paredes de los vasos de la retina

(Abramoff et al., 2010). Esto provoca diversas anomalías en la retina, produciendo los siguientes signos clínicos visibles en el fondo de ojo: microaneurismas (MAs), exudados duros (EXs), exudados algodonosos, hemorragias (HEs), neovascularización y edema macular. Las lesiones rojizas (LRs), que abarcan MAs y HEs, junto con los EXs, son las lesiones que aparecen durante las primeras etapas de la enfermedad, antes de que la visión se haya visto afectada.

La pérdida de visión asociada a la RD no se puede recuperar pero sí se puede prevenir desde las primera etapas de la enfermedad, cuando los tratamientos son efectivos y permiten retrasar su evolución. Por esta razón, es importante diagnosticar la RD de manera precoz. Sin embargo, esta enfermedad es asintomática hasta etapas avanzadas y el paciente no es consciente de padecerla hasta que la visión ya se ha visto afectada y el tratamiento puede resultar complicado (Mookiah et al., 2013). Por estos motivos, los pacientes con RD deben someterse a exámenes oculares periódicos a través de programas de cribado, con el objetivo de identificar los signos clínicos de la RD. Debido a su seguridad y rentabilidad, la imagen de fondo de ojo o retinografía es la modalidad de imagen retiniana más establecida en la clínica para realizar estos exámenes (Abramoff et al., 2010). Tradicionalmente, los programas de cribado se basan en la inspección visual de estas imágenes por parte de un especialista capacitado. Sin embargo, este análisis manual requiere mucho tiempo y es costoso. Con la creciente incidencia de la DM y el limitado número de médicos y recursos sanitarios existentes, la detección precoz de la RD está saturando las consultas sin llegar a tener bajo control la enfermedad (Stolte and Fang, 2020). Además, existe cierta subjetividad relacionada con el diagnóstico (alrededor del 11 % de discrepancia entre especialistas). Por todo ello, los sistemas de diagnóstico asistido por ordenador (CAD) son necesarios para ayudar a los especialistas a realizar un diagnóstico rápido y fiable, que permita reducir la carga de trabajo y los costes asociados (Abramoff et al., 2010; Stolte and Fang, 2020).

Se han desarrollado múltiples sistemas CAD como ayuda a la detección automática de la RD utilizando retinografías. La complejidad de este diagnóstico sugiere que estos sistemas se dividan en varias etapas. En primer lugar, se debe realizar una evaluación de la calidad de la imagen para evitar diagnósticos erróneos (Paulus et al., 2010). En segundo lugar, se requiere una etapa de preprocesado para preparar la imagen para las etapas posteriores. En tercer lugar, deben identificarse las principales estructuras de referencia del fondo de ojo normal, que son el disco óptico (DO), la fóvea y los vasos de la retina, como paso previo a la detección de los signos patológicos (Patton et al., 2006). A continuación, se debe realizar la segmentación de las anomalías. Las principales lesiones visibles de interés que

caracterizan a la RD son las LRs y EXs. Finalmente, en la práctica, el tratamiento potencialmente necesario depende del grado de severidad de la enfermedad. En consecuencia, la graduación de la severidad de la RD es la última etapa de los sistemas de detección automática. En este contexto, existen distintas escalas publicadas. En esta Tesis Doctoral se utilizó la Escala clínica internacional de la RD, puesto que es la más utilizada en la literatura (Wilkinson et al., 2003). Esta escala clasifica la enfermedad en 5 niveles: sin RD, RD No Proliferativa (NPDR) leve, NPDR moderada, NPDR severa y RD proliferativa (Wilkinson et al., 2003).

La presente Tesis Doctoral se presenta como un compendio de cuatro publicaciones indexadas en el *Journal Citation Reports* (JCR) entre los años 2019 y 2020. Además, se ha enviado un quinto artículo que se encuentra actualmente en fase de revisión. El objetivo de esta investigación fue desarrollar nuevos métodos automáticos de análisis de retinografías para la ayuda al diagnóstico de la RD. La primera publicación se centró en la evaluación automática de la calidad de las retinografías (Jiménez-García et al., 2019). Combinamos características derivadas de los métodos *spatial and spectral entropy-based quality* (SSEQ) y *natural images quality evaluator* (NIQE) junto con otras novedosas métricas de nitidez y luminosidad basadas en la transformada *wavelet* continua (CWT) y el modelo de color tono-saturación-valor (HSV). Además de este método, presentamos un nuevo enfoque basado en *deep learning* en una conferencia internacional (Romero Oraá et al., 2020). Utilizamos una red neuronal convolucional (CNN) con arquitectura InceptionResNetV2 y las técnicas *data augmentation*, *transfer learning* y *fine tuning*. El segundo artículo tenía como objetivo desarrollar un método para detectar automáticamente las LRs en retinografías, incluyendo HEs y MAs, en base al concepto de superpixels (Romero-Oraá et al., 2019). Se aplicó el algoritmo Entropy Rate Superpixel (ERS) en combinación con una red neuronal perceptrón multicapa (MLP). En este estudio, también propusimos una etapa de preprocesado novedosa para normalizar la apariencia de la imagen y realzar las estructuras de la retina. En el tercer artículo publicado, el objetivo principal fue desarrollar métodos robustos para localizar automáticamente los centros del DO y la fóvea (Romero-Oraá et al., 2020). El método propuesto se basó en la combinación de nuevos mapa de saliencia que representan las relaciones espaciales entre algunas estructuras de la retina y la apariencia visual del DO y la fóvea. En este artículo, se propuso además un método para extraer el fondo retiniano basado en *region growing*. La cuarta publicación del compendio se centró en la segmentación conjunta de LRs y EXs y se basó en la descomposición de la imagen en varias capas (Romero-Oraá et al., 2020). Entre estas capas, se incluyeron las características

reflectantes de la retina y la vasculatura coroidea visible en las retinas atigradas. El último artículo escrito durante la Tesis, aún en fase revisión, tenía por objetivo desarrollar un método automático para clasificar el grado de severidad de la RD. Propusimos un enfoque de *deep learning* basado en un novedoso mecanismo de atención que separa las estructuras brillantes de la retina de las oscuras.

C.2   Hipótesis y objetivos

El diagnóstico automático de la RD normalmente implica dividir el problema en varias etapas. Aunque se pueden encontrar en la literatura numerosos métodos para llevar a cabo cada una de estas etapas, estos no están exentos de limitaciones y todavía quedan margen de mejora y técnicas por explorar. En este contexto, planteamos la hipótesis de que *el desarrollo de nuevos métodos automáticos de análisis de retinografías podría contribuir al diagnóstico precoz de la RD*. Además, la combinación de dichos métodos podría proporcionar un sistema CAD completo como herramienta de diagnóstico en la práctica clínica. No obstante, aunque este sistema sería muy beneficioso, es importante destacar que cada uno de los métodos que cubren las distintas etapas de diagnóstico tiene utilidad en sí mismo. En esta Tesis Doctoral llevamos a cabo varios estudios para cubrir cada una de esas etapas. Las hipótesis particulares que motivaron estos estudios se exponen a continuación.

En el primero de los estudios del compendio de publicaciones de esta Tesis (Jiménez-García et al., 2019), se propuso un nuevo método para la evaluación de la calidad de las retinografías. Es importante asegurarse de que las imágenes de entrada tengan calidad suficiente para su análisis. De lo contrario, se podrían obtener diagnósticos erróneos. En los trabajos previos con el mismo objetivo, se estudiaron individualmente características genéricas de imagen que son insuficientes para representar la calidad de la imagen. Nosotros planteamos la hipótesis de que *combinar métodos no-reference image quality assessment (NR-IQA) basados en Natural Scenes Statistics (NSS) con características genéricas basadas en la nitidez y la luminosidad puede ser útil para evaluar la calidad de las retinografías*.

Durante el transcurso de la presente Tesis, se desarrolló un método adicional para la evaluación de la calidad de las retinografías, que se presentó en un congreso internacional (Romero Oraá et al., 2020). En los últimos años, las arquitecturas de *deep learning*, como las CNN, han cobrado importancia en el procesado de imágenes. En este contexto, planteamos la hipótesis de que *el uso de transfer learning y fine tuning con la arquitectura InceptionResNetV2 podría alcanzar resultados sólidos y superar los métodos convencionales*.

La etapa de preprocesado es primordial para normalizar la apariencia entre imágenes y dentro de cada imagen, así como para realzar las estructuras de la retina. Con el método de preprocesado propuesto en (Romero-Oraá et al., 2019), planteamos la hipótesis de que *la aplicación secuencial de ciertas operaciones nos permitiría resaltar las lesiones, evitar efectos de borde en etapas posteriores y normalizar la apariencia de todas las imágenes.*

El tercer artículo del compendio de publicaciones (Romero-Oraá et al., 2020) tenía por objetivo localizar los centros del DO y la fóvea. La mayoría de los métodos propuestos anteriormente se centran en criterios específicos que son insuficientes para representar el OD y la fóvea en todas las imágenes (Lalonde et al., 2001b; Sinthanayothin et al., 1999). Algunos estudios combinan varios indicadores, como la entropía de la imagen o la convergencia de los vasos principales, pero la forma en que se utilizan esos indicadores es muy restrictiva. Por lo tanto, estos algoritmos fallan cuando cualquiera de estos indicadores se desvía del patrón estándar (Giachetti et al., 2013; Qureshi et al., 2012). Partimos de la hipótesis de que *la combinación de nuevos indicadores complementarios permitiría representar con precisión las áreas donde se encuentran tanto el OD como la fóvea. De esta forma, el cálculo de ciertos mapas de prominencia sería suficiente para detectar las ubicaciones de los centros incluso para los casos específicos en los que el OD y la fóvea no muestran una apariencia estándar.*

La detección conjunta de EXs y LRs se estudió en (Romero-Oraá et al., 2020). Ninguno de los estudios previos ha considerado individualmente la presencia de otras estructuras de la retina más allá del DO, la fóvea y la vasculatura. Nuestra hipótesis es que *las características reflectantes de la retina y la vasculatura coroidea visible en las retinas atigradas también podrían ser útiles para la detección de lesiones retinianas. De esta forma, las retinografías podrían descomponerse en varias capas para facilitar la segmentación de EXs y LRs.*

La detección de LRs también se estudió en (Romero-Oraá et al., 2019). Casi todos los métodos anteriores consideraban los píxeles como unidad básica de la imagen. Sin embargo, el concepto de superpíxeles, entendido como un grupo de píxeles que representan entidades naturales, es más consistente con la cognición visual humana y contiene menos redundancia (Zhou et al., 2017a). Este tipo de enfoque ya se ha utilizado con Simple Linear Iterative Clustering (SLIC). Sin embargo, esta técnica tiene limitaciones importantes en términos de precisión y adherencia a los límites (Xie et al., 2019). En (Romero-Oraá et al., 2019), planteamos la hipótesis de que *el algoritmo de segmentación Entropy Rate Superpixel (ERS) podría ser útil para segmentar RL en retinografías, superando al algoritmo*

SLIC.

Finalmente, el último estudio realizado durante esta Tesis se centró en la graduación de la severidad. Todos los estudios previos con el mismo propósito estimaban la severidad detectando al mismo tiempo todos los signos de RD presentes en la imagen, independientemente de su tipo. Según nuestra hipótesis, *esta detección conjunta hace que la tarea de clasificación sea más difícil de optimizar que una detección separada de las lesiones claras y las oscuras.*

Definidas las hipótesis, el objetivo principal de la Tesis es *estudiar, diseñar y desarrollar nuevos métodos basados en el análisis automático de retinografías para ayudar en la detección, el diagnóstico y el tratamiento de la RD en cada una de las etapas.* Para llevar a cabo este objetivo, se plantearon los siguientes objetivos específicos:

 I. Revisar el estado del arte acerca del procesado de imágenes médicas y, en particular, del análisis de retinografías orientado al diagnóstico de la RD. Esta revisión involucra el estudio de métodos para las diferentes etapas de un sistema CAD para la RD.

 II. Construir una base de datos privada de retinografías con distintos niveles de calidad y diversos tipos de lesiones, incluidos pacientes con RD y controles sanos.

 III. Seleccionar las bases de datos públicas de retinografías útiles para el desarrollo y validación de los estudios involucrados en esta Tesis Doctoral.

 IV. Implementar y optimizar las técnicas de procesado de imagen existentes estudiadas para cada etapa del proceso de detección de RD. Este objetivo incluye la evaluación de la calidad de la imagen, la detección del DO y la fóvea, la segmentación de LRs y EXs y la graduación de la severidad de RD.

 V. Obtener los resultados que permitan evaluar el rendimiento de los métodos propuestos.

 VI. Discutir y comparar los resultados obtenidos con los de los estudios del estado del arte.

 VII. Extraer las conclusiones adecuadas en base a la discusión anterior.

VIII. Difundir los principales resultados y conclusiones de los estudios realizados en revistas indexadas al JCR y en foros científicos como congresos nacionales e internacionales. Asimismo, se pretende realizar otras actividades de

divulgación como presentarse al concurso universitario Three Minute Thesis
(3MT®) o asistir a distintas jornadas de investigación.

C.3 Materiales

Durante el desarrollo de la Tesis Doctoral, se creó una base de datos privada for-
mada por 2107 retinografías pertenecientes a 688 pacientes y proporcionadas por
el Instituto de Oftalmobiología Aplicada (IOBA) de la Universidad de Valladolid
y el Hospital Clínico Universitario de Valladolid. Todas las imágenes se captu-
raron utilizando la cámara Topcon TRC-NW400 (Topcon Medical Systems, Inc.,
Oakland, NJ, EE. UU.) con un campo de visión (FOV) de 45 grados y se alma-
cenaron utilizando el formato JPEG de 24 bits con un tamaño de 1956 × 1934
píxeles. Por cada ojo, se capturó una imagen centrada en la fóvea y una imagen
centrada en el DO.

Dos oftalmólogos especialistas determinaron, para cada imagen, si tenía calidad
suficiente para ser analizada o no. De esta manera, 1810 de las 2107 imágenes
se consideraron imágenes de calidad adecuada, mientras que las 297 imágenes
restantes tenían una calidad inadecuada. Esta clasificación manual de la calidad
de las imágenes sirvió como *gold standard* para nuestro estudio (Jiménez-García
et al., 2019), destinado a la evaluación automática de la calidad de las retinografías.

Entre las retinografías con calidad suficiente para ser analizadas, se seleccionó
un subconjunto reducido para anotar las lesiones. Los oftalmólogos dibujaron ma-
nualmente los contornos de las LRs y los EXs en 564 imágenes. De este subcon-
junto, 270 retinografías mostraban signos de RD mientras que las 294 imágenes
restantes carecían de cualquier tipo de lesión. Entre las 270 imágenes patológicas,
183 mostraban EXs, 239 mostraban LRs y 152 imágenes incluían tanto EXs como
LRs. La anotación manual de estas lesiones nos sirvió como *gold standard* para
el estudio (Romero-Oraá et al., 2019), orientado a la detección de LRs, y para el
estudio (Romero-Oraá et al., 2020), centrado en la detección conjunta de LRs y
EXs.

Finalmente, los oftalmólogos anotaron manualmente los centros del DO y la
fóvea en el mismo subconjunto de retinografías seleccionadas para la anotación de
lesiones. De esta manera, también contábamos con las coordenadas *gold standard*
para estas dos importantes estructuras de referencia. Esto nos permitió utilizar esta
base de datos para realizar nuestro estudio (Romero-Oraá et al., 2020), orientado
a la localización automática del DO y la fóvea.

Además de la base de datos privada, en esta Tesis Doctoral se utilizaron cinco

bases de datos públicas de retinografías:

- **DRIMDB**. Esta base de datos fue proporcionada por el Departamento de Oftalmología de Retina de la Facultad de Medicina de la Universidad Técnica de Karadeniz (Turquía) (Sevik et al., 2014). Todas las imágenes se obtuvieron con una cámara Canon CF-60UVi con un FOV de 60 grados y se almacenaron en archivos JPEG con una resolución de 570 × 760 píxeles. Esta base de datos consta de 216 retinografías. Un experto identificó tres clases de calidad: buena (125 imágenes), mala (69 imágenes) y *outlier* (22 imágenes). Esta base de datos se utilizó en nuestro estudio (Romero Oraá et al., 2020).

- **DIARETDB1**. Esta base de datos está compuesta por 89 imágenes capturadas en el Hospital Universitario de Kuopio (Finlandia) (Kauppi et al., 2007). Fueron capturadas con un FOV de 50 grados y tenían una resolución de 1500 × 1552 píxeles. Solo se capturó una imagen por ojo (centrada en la fóvea). En esta base de datos, cuatro expertos médicos anotaron cualquier signo de MAs, HEs y EXs. Según las anotaciones, 27 imágenes se asociaron con retinas sanas, 7 imágenes con RD leve, 28 imágenes con RD no proliferativa moderada y grave y 27 imágenes con RD proliferativa (Kauppi et al., 2007). Esta base de datos se utilizó en nuestros estudios (Romero-Oraá et al., 2019) y (Romero-Oraá et al., 2020).

- **DRIVE**. Esta base de datos consta de 40 imágenes obtenidas en un programa de cribado de la RD llevado a cabo en los Países Bajos (Staal et al., 2004). Las imágenes se capturaron con la cámara no midriática 3 Canon CR5 con un FOV de 45 grados y se comprimieron en formato JPEG de 24 bits. El tamaño de las imágenes es de 768 × 584 píxeles y todas ellas están centradas en la fóvea. En este conjunto de datos, 7 de las 40 imágenes presentan signos patológicos. Esta base de datos se utilizó para nuestro estudio (Romero-Oraá et al., 2020).

- **Messidor**. Esta base de datos contiene 1200 imágenes capturadas con la cámara Topcon TRC NW6 con un FOV de 45 grados (Decencière et al., 2014). Se almacenaron en formato TIFF de 24 bits a tres resoluciones diferentes: 1440 × 960, 2240 × 1488 y 2304 × 1536 píxeles. Todas las imágenes estaban centradas en la fóvea. La base de datos contiene un diagnóstico médico para cada imagen, pero no la anotación de las lesiones. Esta base de datos se empleó en nuestro estudio (Romero-Oraá et al., 2020).

- **Kaggle**. Este conjunto de retinografías fue proporcionado por EyePACS para el concurso de detección de la RD publicado en Kaggle (Kaggle, 2015), de ahí el nombre por el que se conoce. Es la base de datos pública relacionada con DR más grande disponible con 88,702 imágenes. En esta base de datos, un médico determinó el grado de gravedad de la RD según la Escala clínica internacional de RD (Wilkinson et al., 2003), que tiene 5 niveles: 65,343 imágenes sin RD, 6,205 imágenes con NPDR leve, 13,153 imágenes con NPDR moderada, 2,087 imágenes con NPDR severa y 1914 imágenes con RD proliferativa. Para cada paciente, se incluyó una retinografía de cada ojo. Esta base de datos se utilizó en el último estudio de esta Tesis, destinado a la graduación automática de la severidad de la RD.

C.4   Métodos

Los métodos propuestos en esta Tesis Doctoral pueden agruparse de acuerdo a las distintas etapas de diagnóstico de la RD: evaluación de la calidad de las imágenes, preprocesado, localización del DO y la fóvea, segmentación de LRs, segmentación de EXs y graduación de la severidad.

C.4.1   Evaluación de la calidad de las imágenes

La evaluación de la calidad de las retinografías se llevó con dos enfoques distintos. El primero se basó en características globales de las imágenes. El segundo enfoque consistió en una arquitectura de *deep learning*.

Método basado en la combinación de características globales

Este método se propuso en (Jiménez-García et al., 2019) como parte del compendio de publicaciones y se compone de cuatro etapas:

1. **Preprocesado**

   En esta etapa se aplicó un algoritmo iterativo para rellenar el borde negro que rodea el FOV con una estimación del fondo retiniano basada en los píxeles vecinos. Esto permitió reducir los efectos de borde en etapas posteriores (Soares et al., 2006).

2. **Extracción de características**

   En primer lugar se utilizó el método SSEQ sobre la imagen preprocesada en tres escalas diferentes (1, 1/2 y 1/3). Las imágenes se dividieron en bloques

no superpuestos y se calculó la entropía espacial y la entropía espectral en cada uno de ellos (Liu et al., 2014). A continuación, se calculó la media y la oblicuidad de ambas medidas sobre los bloques para cada escala de la imagen, extrayendo así 12 características.

En segundo lugar, se utilizó el método NIQE, que se basa en comparar cada imagen con un modelo de referencia (Mittal et al., 2013). Para ello construimos, por un lado, un modelo NSS de referencia para representar el conjunto de imágenes de calidad adecuada. Por otro lado, construimos otro modelo NSS para representar la imagen a evaluar. Para construir estos modelos, calculamos algunos parámetros de una distribución Gaussiana generalizada para alimentar un modelo Gaussiano multivariante. Este proceso se aplicó distintas versiones escaladas de la imagen preprocesada normalizada (Mittal et al., 2013). Finalmente, la comparación de estos modelos permitió extraer 1 característica, correspondiente al índice de calidad NIQE, que puede considerarse una medida de naturalidad de la imagen.

En tercer lugar, se calculó la CWT del canal verde de la imagen preprocesada para 6 escalas distintas (2, 4, 8, 16, 32 y 64) y se utilizó el sombrero mexicano como *wavelet* madre (Antoine and Murenzi, 1996; Rangayyan, 2004). A continuación, se obtuvieron los mapas de varianza local de esas transformadas (Aja-Fernández et al., 2006; Gonzalez and Woods, 2009). Para cada escala, se calculó la entropía de la CWT y la media y la desviación típica de los mapas de varianza local, dando lugar a 18 nuevas características.

Por último, se utilizó el canal V del modelo de color HSV para cuantificar la iluminación (Zhou et al., 2018). Se aplicó un pequeño filtro de mediana para reducir el ruido y, a continuación, un filtro Gausiano de gran tamaño para estimar el fondo de las imágenes (Gonzalez and Woods, 2009). Sobre la imagen resultante, se calcularon los valores de luminosidad correspondientes a los percentiles 1 %, 5 %, 10 %, 15 % y 20 %. Adicionalmente, se calcularon las diferencias entre percentiles de luminosidad consecutivos (5-1, 10-5, 15-10 y 20-15) para representar las variaciones de intensidad en el fondo. Esta etapa dio lugar a 9 características.

3. **Selección de características: *Fast Correlation-Based Filter***

   Se extrajeron un total de 40 características globales para cada imagen. Con el objetivo de identificar cuáles son relevantes y no redundantes, se aplicó el algoritmo *fast correlation-based filter* (FCBF) (Yu and Liu, 2004). Este método es independiente del clasificador y se basa en la métrica *symmetrical*

uncertainty (SU). Con el objetivo de mejorar la robustez del proceso de selección de características, se utilizó la técnica *bootstrapping* (Witten et al., 2016).

4. **Clasificación:** ***Multilayer Perceptron Neural Network***

Para clasificar las imágenes según su nivel de calidad, se utilizó una red neuronal MLP de 3 capas (entrada, oculta y salida). La capa de entrada tenía un número de neuronas igual al número de características seleccionadas. La capa de salida tenía solo una neurona para realizar la clasificación binaria (Bishop, 1995; Witten et al., 2016). El número de neuronas ocultas (N_{HIDDEN}) se obtuvo experimentalmente (Bishop, 2006; Witten et al., 2016). En la capa oculta se utilizó la función de activación tangente hiperbólica sigmoide (Bishop, 1995). En la neurona de salida se empleó la función de activación logística sigmoidea (Bishop, 1995). La función de error utilizada fue la entropía cruzada (Bishop, 1995). Además, para evitar el sobreentrenamiento se utilizó la regularización *weight decay*, cuyo parámetro (η) se obtuvo experimentalmente. Por último, para lidiar con el desbalanceo de clases, se aplicó la técnica *synthetic minority oversampling technique* (SMOTE) (Chawla et al., 2002).

Método basado en *deep learning*

El segundo enfoque para evaluar la calidad de las retinografías se publicó en (Romero Oraá et al., 2020). Este método requería, en primer lugar, redimensionar las imágenes al mismo tamaño y normalizarlas. Para aumentar el número de imágenes de entrenamiento, se utilizó la técnica *data augmentation* (Perez and Wang, 2017). Las nuevas imágenes se obtuvieron aplicando rotaciones, desplazamientos, volteos y escalado aleatorios (Zago et al., 2018). Es importante destacar en este trabajo el uso de *transfer learning*. Esta técnica permite resolver un problema de aprendizaje automático en un dominio particular de interés con el conocimiento aprendido de los datos de entrenamiento de otro dominio de interés (Pan and Yang, 2010). Para ello, en este estudio se utilizó un modelo inicializado con un conjunto de pesos previamente entrenados con imágenes del proyecto ImageNet (Jia Deng et al., 2009).

La arquitectura de *deep learning* propuesta consistió en una CNN. Este tipo de redes están formadas por capas convolucionales, así como otras de tipo *pooling* y *fully connected* (Saha et al., 2018; Zago et al., 2018). Las CNNs se caracterizan por tener "interacciones dispersas", donde cada neurona solo está conectada a un

subconjunto vecino, y "pesos compartidos", donde todas las neuronas de capa usan los mismos pesos. Esto permite mejorar enormemente el rendimiento del modelo y es ideal para el análisis de imagen (Wang et al., 2019). En este trabajo se utilizó la arquitectura InceptionResNetV2, que es una combinación de Inception (Szegedy et al., 2015) y ResNet (He et al., 2016). Para adaptar esta arquitectura a nuestro problema de clasificación binaria, reemplazamos las últimas 3 capas por 3 capas *fully connected* de 1024, 512 y 1 neurona, respectivamente (Zago et al., 2018). Las dos primeras tenían una función de activación ReLU (Zago et al., 2018) mientras que para la última se utilizó una función de activación sigmoidea.

C.4.2   Preprocesado

El procesado de retinografías se abarcó en el estudio (Romero-Oraá et al., 2019) y estaba formado por cinco operaciones:

1. **Eliminación de artefactos brillantes en el borde de la FOV**. Se aplicó umbralización y operaciones morfológicas (Zhang et al., 2014).

2. **Extensión del fondo retiniano**. Se aplicó un algoritmo iterativo para sustituir el borde negro que rodea el FOV con el valor medio de los píxeles vecinos (Soares et al., 2006).

3. **Ecualización de la iluminación y el color**. Se utilizó el espacio de color Hue-Saturation-Intensity (HSI), normalizando cada canal por separado en base a operaciones matemáticas directas (Hoover and Goldbaum, 2003).

4. **Eliminación de ruido**. Se aplicó un filtro de media de tamaño pequeño (Seoud et al., 2016).

5. **Realce del contraste**. Se empleó el método contrast limited adaptive histogram equalization (CLAHE) (Rasta et al., 2015).

C.4.3   Localización del disco óptico y la fóvea

Esta sección describe el algoritmo propuesto en (Romero-Oraá et al., 2020). Después de aplicar el preprocesado anterior, se extrajo el fondo retiniano, obteniendo 3 nuevas imágenes: I_{bg}, que es el fondo estimado de la retinografía después de eliminar las estructuras brillantes y las oscuras; I_{bg-bri}, que es el fondo estimado después de eliminar las estructuras oscuras y conservar las brillantes; y $I_{bg-dark}$, que es el fondo estimado después de eliminar las estructuras brillantes y conservar las oscuras. Para obtener estas estimaciones, aplicamos un novedoso algoritmo

multiescala y una etapa de umbralización (Romero-Oraá et al., 2019). A continuación, obtuvimos una segmentación de la vasculatura (Mendonça et al., 2006). Como novedad, utilizamos la imagen $I_{bg-dark}$ como entrada para evitar los falsos positivos causados por las estructuras brillantes. Después, se calcularon varios mapas de saliencia para representar el DO y la fóvea, que se describen a continuación.

Para detectar el DO, se obtuvo un mapa de saliencia que representaba las arcadas principales en su paso vertical por el DO. El segundo mapa de saliencia se obtuvo aplicando *template matching* para representar el aspecto circular y brillante que caracteriza al DO. El tercer mapa se obtuvo utilizando un filtro de desviación estándar local sobre el canal rojo de la imagen preprocesada. La combinación de estos tres mapas permitió identificar el centro del DO.

Para detectar la fóvea, calculamos un primer mapa de saliencia que representaba la ausencia de vasos sanguíneos que la caracteriza, para lo que se utilizó la previa segmentación de la vasculatura. El segundo mapa de saliencia se obtuvo aplicando *template matching* para representar la apariencia oscura de la zona macular. A continuación, consideramos el hecho de que la fóvea siempre se encuentra a una distancia aproximadamente constante del DO (entre tres y siete veces el radio del DO). Por ello, construimos un anillo para delimitar esta región. Igualmente, consideramos que la fóvea siempre se encuentra aproximadamente al mismo nivel vertical que el DO en las retinografías. Por ello, construimos una banda vertical alineada con el DO. En el siguiente paso, consideramos que la densidad vascular es mayor en el lado temporal de la retina (donde se encuentra la fóvea). Por este motivo generamos un nuevo mapa de saliencia que representaba esta medida a cada lado del DO. La combinación de estos cinco mapas permitió identificar el centro de la fóvea.

C.4.4   Detección de lesiones rojizas

La detección de LRs en esta Tesis Doctoral se llevó a cabo desde dos enfoques distintos. En (Romero-Oraá et al., 2019) se aplicó la clasificación de superpíxeles. En (Romero-Oraá et al., 2020) se descompuso la imagen en capas.

Método basado en la clasificación de superpíxeles

En primer lugar, se detectaron los píxeles oscuros de la retinografía utilizando el algoritmo multiescala presentado anteriormente para extraer el fondo retiniano. A continuación, se aplicó la segmentación ERS, que dividió la imagen en superpíxeles. Un superpíxel es una región perceptualmente uniforme en la imagen. Agrupa píxe-

les con colores y texturas similares y se adapta a los bordes de la imagen (Zhou et al., 2017a). Dado que representa las entidades naturales de la imagen, el superpíxel es una región adecuada de la que extraer características (Zhou et al., 2017a). Aplicar el algoritmo ERS requiere ajustar correctamente los parámetros K, λ y σ. Estos parámetros determinan la formación de los superpíxeles en términos de compacidad, homogeneidad, tamaño y adherencia a los bordes. En este trabajo se eligió experimentalmente que $K = 2000$, $\lambda = 0{,}08$ y $\sigma = 2$. Tras la segmentación ERS, se trató de reducir el número de candidatos para simplificar la posterior tarea de clasificación. La mayoría de los superpíxeles detectados no eran regiones oscuras, por lo que fueron descartados como candidatos. Además, combinamos todos aquellos superpíxeles con un color medio similar. Para calcular este color medio se utilizó el espacio de color CIELAB.

A continuación, se extrajo un conjunto de 39 características para representar cada uno de los superpíxeles candidatos a ser LR. Entre estas características, se encuentran el área de la región, la excentricidad o la distancia al centro del DO, entre otras. En el siguiente paso se aplicó el algoritmo FCBF, anteriormente descrito, para seleccionar las características más relevantes y no redundantes. Finalmente, se utilizó una red neuronal MLP de 3 capas para detectar cuáles de los superpíxeles candidatos eran verdaderas LRs. La capa de entrada tenía tantas neuronas como características fueron seleccionadas por el método FCBF. La capa de salida estaba formada por una sola neurona, suficiente para un problema dicotómico. El número de neuronas de la capa oculta, n_{hid}, se obtuvo experimentalmente. En esta capa se utilizó la función de activación tangente hiperbólica sigmoide ($tanh$) (Bishop, 1995). En la neurona de salida se utilizó la función logística sigmoide. La función de activación fue la entropía cruzada. Además, se aplicó la regularización *weight decay*, cuyo parámetro (υ) se obtuvo experimentalmente.

Método basado en la descomposición de la imagen en capas

Este método tenía por objetivo la detección conjunta de LRs y EXs pero aquí nos centramos exclusivamente en la detección de LRs. Se partió de la imagen preprocesada y la detección del DO, la fóvea y la vasculatura, detallados anteriormente. A continuación, se llevó a cabo la segmentación de regiones candidatas a ser LR. Para ello, descompusimos la retinografía en distintas capas. Aquí partimos de la imagen I_{dark}, que se obtiene utilizando la estimación del fondo y representa las estructuras oscuras de la imagen, eliminando por completo el resto de estructuras. Después eliminamos también la vasculatura. La imagen resultante se descompuso haciendo uso del modelo de color HSV, umbralizando empíricamente cada uno de

los canales. De esta manera, conseguimos separar, en primer lugar, la vasculatura coroidea subyacente visible en las retinas atigradas. En segundo lugar, la atrofia coroidea. Y, por último, los píxeles asociados a LRs candidatas. A continuación, extrajimos un total de 100 características para cada región candidata. Muchas de estas características se calcularon a partir de las capas previamente separadas. Luego, seleccionamos un subconjunto de las 24 características más relevantes utilizando la técnica FCBF. Por último, utilizamos una red neuronal MLP para detectar las verdaderas LRs de entre las candidatas. Esta red tenía la misma arquitectura que la utilizada en el método basado en la detección de superpíxeles.

C.4.5 Detección de exudados duros

La detección de EXs en esta Tesis Doctoral se llevó a cabo junto a la detección de LRs con el método basado en la descomposición de la imagen en capas (Romero-Oraá et al., 2020). Al igual que para las LRs, se partió de la imagen preprocesada y la detección del DO, la fóvea y la vasculatura. A continuación, se llevó a cabo la segmentación de regiones candidatas a ser EX para lo que descompusimos la imagen en capas. En este caso partimos de la imagen I_{bri}, que también se obtiene utilizando la estimación del fondo retiniano y representa las estructuras brillantes de la imagen, eliminando por completo las estructuras oscuras y el fondo. Esta imagen se descompuso en capas empleando de nuevo el modelo de color HSV y umbralizando empíricamente cada uno de los canales. De esta forma, obtuvimos, en primer lugar, la capa correspondiente a los vasos coroideos visibles en las retinas atigradas. En segundo lugar, la capa que representa las caracterísicas reflectantes fruto de las fibras nerviosas. Estos reflejos son comunes en pacientes jóvenes. Finalmente, obtuvimos la capa de EXs candidatos. A continuación, extrajimos un total de 100 características para cada región candidata. Varias de estas características se calcularon a partir de las capas previamente separadas. Luego, seleccionamos un subconjunto de las 34 características más relevantes utilizando la técnica FCBF. Por último, utilizamos una red neuronal MLP para detectar los verdaderos EXs de entre las regiones candidatas. Esta red tenía la misma arquitectura que la utilizada en el método de detección de LRs.

C.4.6 Graduación de la severidad

Este método estaba basado en técnicas de *deep learning* y tenía por objetivo clasificar la severidad de la RD en 5 niveles, de acuerdo a la escala internacional (Wilkinson et al., 2003). Primero se evaluó la calidad de la imagen utilizando el

método basado en *deep learning* descrito anteriormente. A continuación, se llevó a cabo una pequeña etapa de preprocesado para adaptar la imagen a la red profunda. En esta etapa se recortó el borde negro externo al FOV, se redimensionó la imagen y se normalizó. A continuación, partimos de la imagen preprocesada y de las imágenes I_{bri} y I_{dark}, que ya se han presentado anteriormente. La imagen I_{bri} representa las estructuras brillantes de la imagen mientras que la imagen I_{dark} representa las estructuras oscuras. En el método propuesto se utilizó una red completamente convolucional (FCN) como extractor de características para las tres imágenes mencionadas. La arquitectura seleccionada para esta red fue Xception (versión eXtrema de Inception) (Chollet, 2017). Cabe destacar en este trabajo el uso de *data augmentation* y *transfer learning* (Costa et al., 2019; Perez and Wang, 2017). Las características extraídas por las redes FCN sirvieron de entrada a un novedoso mecanismo de atención. Este mecanismo permitió centrar el aprendizaje en las características más relevantes para la clasificación. Como resultado, obtuvimos dos mapas de atención independientes para la detección de LRs y EXs. A continuación, combinamos estos mapas con las características extraídas de la imagen preprocesada. Al final de la arquitectura, incluimos 3 capas *fully connected*. Las 2 primeras capas tenían 1024 y 512 neuronas, respectivamente, y una función de activación de ReLU. La última capa tenía 5 neuronas, una para cada grado de severidad de RD, y una función de activación *softmax*.

C.5   Resultados y discusión

Los resultados obtenidos para cada estudio de esta Tesis Doctoral se exponen a continuación acompañados de una discusión.

C.5.1   Evaluación de la calidad de las imágenes

Método basado en la combinación de características globales

El método basado en la combinación de características globales se desarrolló con un conjunto de entrenamiento de 1053 imágenes y se evaluó con un conjunto de test de 1054 imágenes de la base de datos privada. De las 40 características extraídas, el método FCBF seleccionó un subconjunto de 10 características relevantes y no redundantes. Cabe destacar que se incluyeron características de los cuatro tipos analizados en este estudio: SSEQ, QNIQE, nitidez (CWT) y luminosidad. Por lo tanto, se demuestra que las características estudiadas son complementarias y útiles para la evaluación de la calidad de las retinografías. Para la red MLP, las pruebas

experimentales determinaron que los valores óptimos para el número de neuronas en la capa oculta y el parámetro de regularización fueron 21 y 0.1, respectivamente. Una vez entrenada la red con estos parámetros, los resultados sobre el conjunto de test alcanzaron una sensibilidad (Se) del 92.04 %, una especificidad (Sp) del 87.92 %, una precisión (Acc) del 91.46 %, un valor predictivo positivo (PPV) del 97.88 % y un valor F_1 de 0.9487.

Estos resultados son similares a los obtenidos en estudios anteriores. Sin embargo, las comparaciones deben hacerse con precaución, ya que las bases de datos utilizadas son distintas. Los métodos estructurales y genéricos alcanzaron una Se y Sp en torno al 90 % con una Acc superior 90 % en la mayoría de los casos. Entre los métodos estructurales, Fleming et al. (2012) alcanzaron $Se = 92,60 \%$ y $Sp = 90,00 \%$ con 98 imágenes. Otros autores validaron su método utilizando 400 imágenes, alcanzando $Se = 95,33 \%$ y $Sp = 91,13 \%$ (Welikala et al., 2016). Wang et al. (2016) lograron $Se = 87,45 \%$ y $Sp = 91,66 \%$ usando 536 imágenes, mientras que Abdel-Hamid et al. (2017) alcanzaron $F_1 = 0,8780$ usando 190 imágenes. Otros autores también combinaron el índice NIQE con características estructurales y de iluminación, alcanzando $Acc = 93,60 \%$ con 194 imágenes. Los mejores resultados entre los métodos genéricos se lograron utilizando la red neuronal MLP, con $Se = 99,49 \%$ y $Sp = 99,76 \%$ en una base de datos formada por 848 imágenes (Pires Dias et al., 2014). Los enfoques híbridos también mostraron resultados notables, aunque son más complejos que los métodos genéricos (Paulus et al., 2010; Sevik et al., 2014). Paulus et al. (2010) emplearon 301 imágenes, alcanzando $Acc = 91,70 \%$. En otros estudios, F_1 alcanzó 0.9960 usando 194 imágenes (Sevik et al., 2014). También se han propuesto métodos basados en *deep learning*, consiguiendo un rendimiento excelente. Entre ellos, cabe destacar el estudio de Saha et al. (2018), que alcanzó $Acc = 100 \%$ sobre 3425 imágenes. Sin embargo, cabe destacar que sólo 123 (3.6 %) de estas imágenes correspondían a imágenes de calidad inadecuada, por lo que los datos estaban muy desbalanceados.

Método basado en *deep learning*

El modelo CNN se entrenó con 2348 imágenes de la base de datos privada. La evaluación del método se llevó a cabo con 425 imágenes de la misma base de datos y con las 194 imágenes de la base de datos DRIMDB. Sobre la base de datos privada se obtuvo $Se = 96,82 \%$, $Sp = 91,00 \%$ y $Acc = 95,29 \%$, mientras que sobre DRIMDB se alcanzó $Se = 99,20 \%$, $Sp = 100,00 \%$ y $Acc = 99,48 \%$. Los resultados para DRIMDB son superiores a los obtenidos para la base de datos privada. Esto puede deberse a que en esta última es menos evidente la diferencia

de calidad entre las imágenes de las dos clases.

Nuestros resultados son similares a los obtenidos en estudios previos. La comparación más directa se puede establecer con los trabajos de Chalakkal et al. (2019) y Zago et al. (2018), donde se utilizó la misma base de datos pública. Chalakkal et al. (2019) obtuvieron $Acc = 95{,}29\,\%$ con la arquitectura Inception v1 mientras que Zago et al. (2018) utilizaron una arquitectura Inception v3 para obtener $Acc = 98{,}55\,\%$. En (Yu et al., 2017) se empleó la arquitectura Alexnet y se alcanzó $Acc = 95{,}42\,\%$. En el estudio realizado por Saha et al. (2018) también se utilizó la arquitectura Alexnet y obtuvieron $Acc = 100\,\%$. Sin embargo, en este estudio se descartaron imágenes de calidad ambigua, que son las más propensas a error.

C.5.2   Localización del disco óptico y la fóvea

Los métodos para la localización del DO y la fovea se evaluaron sobre el conjunto de test de la base de datos privada (283 imágenes) y las bases de datos públicas DiaretDB1 (89 imágenes), DRIVE (40 imágenes) y Messidor (1200 imágenes). Para el DO, alcanzamos una precisión del $100\,\%$ para todas las bases de datos excepto Messidor (99.50 %). En cuanto a la localización de la fóvea, también alcanzamos un $100\,\%$ de precisión para todas las bases de datos excepto Messidor (99.67 %). Estos resultados demuestran la gran robustez de los mapas de saliencia propuestos.

En comparación con otros estudios, nuestros resultados para la detección del DO son similares. Los resultados de la base de datos DRIVE alcanzaron la máxima precisión posible (100 %), al igual que en otros estudios Basit and Fraz (2015); Chalakkal et al. (2018); Hsiao et al. (2012); Pereira et al. (2013); Qureshi et al. (2012); Welfer et al. (2011); Yu et al. (2015). En cuanto a DiaretDB1, solo dos estudios anteriores reportaron una precisión del $100\,\%$ (Abdullah et al., 2016; Abed et al., 2016), superando así a muchos otros estudios (Abdullah et al., 2016; Alshayeji et al., 2017; Basit and Fraz, 2015; Chalakkal et al., 2018; Díaz-Pernil et al., 2016; Harangi and Hajdu, 2015,?; Lu and Lim, 2011; Pereira et al., 2013; Qureshi et al., 2012; Rahebi and Hardalaç, 2016,?; Welfer et al., 2011; Yu et al., 2015). Nuestro método también supera la precisión de localización obtenida en todos los métodos anteriores para la base de datos Messidor, excepto (Giachetti et al., 2013) y (Yu et al., 2015). Sin embargo, en (Giachetti et al., 2013) no se proporcionaron resultados para las bases de datos DRIVE y DiaretDB1 y el método propuesto por Yu et al. (2015) obtuvo una menor precisión que nuestro método para la base de datos DiaretDB1 (99.88 %).

Igualmente, para la detección de la fóvea se alcanzó una precisión similar a la

de otros estudios previos. En la base de datos DRIVE, GeethaRamani and Balasu-bramanian (2018) obtuvieron la misma precisión que nosotros, pero usaron solo 35 de las 40 imágenes disponibles. La precisión obtenida por Qureshi et al. (2012) fue menor utilizando la base de datos DRIVE (91.73 %). En cuanto a la base de datos DiaretDB1, nuestro método obtuvo la mayor precisión en la literatura, localizando correctamente las fóveas en todas las imágenes de DiaretDB1. En relación a la base de datos de Messidor, nuestro método también superó la precisión de todos los estudios anteriores Al-Bander et al. (2018); Aquino (2014); GeethaRamani and Balasubramanian (2018); Gegundez-Arias et al. (2013); Giachetti et al. (2013).

Cabe destacar que el método para la extracción de fondo retiniano es otra contribución importante en este estudio, que podría ser útil en estudios adicionales relacionados con el procesado de retinografías. Con este método, mejoramos la segmentación de la red vascular, evitando que los bordes de los EX y otras regiones brillantes se detecten erróneamente como segmentos de vasos sanguíneos.

C.5.3 Detección de lesiones rojizas

A continuación se muestran los resultados para los dos métodos propuestos en esta Tesis orientados la segmentación de LRs.

Método basado en la clasificación de superpíxeles

La elección de parámetros y el entrenamiento de la red se llevaron a cabo con un conjunto de entrenamiento de la base de datos privada formado por 281 retinografías. El método se evaluó con un conjunto de test de la misma base de datos (283 imágenes) y con la base de datos DIARETDB1 (61 imágenes). Los resultados se obtuvieron en base a dos criterios distintos. En primer lugar, con un criterio orientado a píxel, se consideraron como detecciones correctas todos los píxeles pertenecientes a una lesión correctamente detectada y se calcularon el valor predictivo positivo (PPV_p) y la sensibilidad (SE_p). Por otro lado, con un criterio orientado a imagen, se consideró que una imagen correspondía a una retina sana si se detectaban menos de 30 píxeles de lesión (0.000008 % de la imagen) y patológica en caso contrario. Según este criterio se calculó la precisión (ACC_i), la sensibilidad (SE_i) y la especificidad (SP_i).

De las 39 características extraídas de cada superpíxel, el algoritmo FCBF seleccionó un subconjunto de 16. Para la red MLP, las pruebas experimentales determinaron que los valores óptimos para el número de neuronas en la capa oculta y el parámetro de regularización fueron 30 y 0.6, respectivamente. Una vez entrenada

la red con estos parámetros, los resultados sobre el conjunto de test de la base de datos privada alcanzaron $SE_p = 81{,}43\%$, $PPV_p = 86{,}59\%$, $SE_i = 84{,}04\%$, $SP_i = 85{,}00\%$ y $ACC_i = 84{,}45\%$. Para la base de datos DIARETDB1 se obtuvo $SE_p = 88{,}10\%$, $PPV_p = 93{,}10\%$, $SE_i = 84{,}00\%$, $SP_i = 88{,}89\%$ y $ACC_i = 86{,}89\%$.

Estos resultados son similares a los de otros estudios para la detección de LRs según el criterio orientado a imagen. Sin embargo, las comparaciones deben hacerse con precaución ya que las bases de datos y la forma de medir los resultados varían entre estudios. En algunos trabajos se obtuvo una SP_i por debajo del 60 % (García et al., 2010; Seoud et al., 2016). En otros estudios se emplearon conjunto de test muy pequeños como para extraer conclusiones generalizables (García et al., 2010; Niemeijer et al., 2005; Zhou et al., 2017a). Orlando et al. (2018) utilizaron la base de datos Messidor para evaluar su método. Los resultados mostraron $SE_i = 91{,}10\%$ y $SP_i = 50{,}00\%$. (Sánchez et al., 2011) alcanzaron $SP_i = 92{,}20\%$ y $SP_i = 50{,}00\%$ con la misma base de datos. Dado que nuestro método también se evaluó con la base de datos DIARETDB1, se puede establecer una comparación directa con los métodos propuestos por Roychowdhury et al. (2012) y Zhou et al. (2017a). Roychowdhury et al. (2012) alcanzaron $SE_i = 75{,}50\%$ y $SP_i = 93{,}73\%$. Zhou et al. (2017a) obtuvieron $SE_i = 83{,}30\%$, que es ligeramente inferior a la de nuestro enfoque pero su SP_i supera nuestros resultados, alcanzando $SP_i = 97{,}30\%$. Al comparar nuestros resultados con enfoques anteriores, debe tenerse en cuenta que el método propuesto se ha configurado y entrenado utilizando solo imágenes de nuestra base de datos privada.

Una contribución importante de este estudio es que no requiere una etapa de segmentación de la vasculatura, ya que se elimina en la etapa de clasificación. Otra contribución importante es el algoritmo de preprocesado descrito anteriormente, que fue publicado en este estudio. Este algoritmo se utilizó como etapa inicial de los métodos de localización del DO y la fóvea y de detección de LRs y EXs.

Método basado en la descomposición de la imagen en capas

Como en el método anterior, el desarrollo de este método se llevó a cabo con 281 retinografías de la base de datos privada y la evaluación se hizo con 283 imágenes de la misma base de datos y 61 imágenes de la base de datos DIARETDB1. Igualmente, los resultados se obtuvieron con los mismos criterios.

En cuanto a la red MLP, los valores óptimos para el número de neuronas en la capa oculta y el parámetro de regularización fueron 51 y 0.5, respectivamente. Sobre la base de datos privada se obtuvo $SE_p = 82{,}25\%$, $PPV_p = 91{,}07\%$, $SE_i = $

85,00 %, $SP_i = 90,80$ % y $ACC_i = 88,34$ %. Sobre la base de datos pública los resultados mostraron $SE_p = 84,79$ %, $PPV_p = 96,25$ %, $SE_i = 88,00$ % $SP_i = 91,67$ % y $ACC_i = 90,16$ %.

Todos estos resultados son comparables con los obtenidos en estudios previos según el criterio basado en imagen. Hemos encontrado cuatro métodos que han sido evaluados utilizando la base de datos DIARETDB1 con el fin de establecer una comparación directa con el método propuesto. Jaafar et al. (2011) obtuvieron una $SE_i = 98,80$ % alta, pero una $SP_i = 86,20$ % más baja que la nuestra. Además, evaluaron su método sobre un conjunto de imágenes que combina las bases de datos DiaretDB0 y DiaretDB1. En el trabajo de Roychowdhury et al. (2012), se alcanzó una $SP_i = 93,73$ %, pero la $SE_i = 75,50$ % fue menor. En (Zhou et al., 2017a), se obtuvo $SP_i = 91,67$ %. Sin embargo, nuestro valor de SE_i mejora su SE_i del 83.30 %. Nuestro método basado en la descomposición de la imagen en capas también logra mejores resultados que los obtenidos con el método basado en la clasificación de superpíxeles.

C.5.4 Detección de exudados duros

Al igual que para la detección de LRs, se utilizaron 281 retinografías de la base de datos privada como conjunto de entrenamiento y 283 imágenes de la misma base de datos junto con 61 imágenes de la base de datos DIARETDB1 como conjuntos de test. De nuevo, los resultados se obtuvieron aplicando los mismos criterios.

En cuanto a la red MLP, los valores óptimos para el número de neuronas en la capa oculta y el parámetro de regularización fueron 55 y 0.4, respectivamente. Sobre la base de datos privada se obtuvo $SE_p = 89,42$ %, $PPV_p = 96,01$ %, $SE_i = 88,04$ %, $SP_i = 89,42$ % y $ACC_i = 95,41$ %. Sobre la base de datos pública los resultados mostraron $SE_p = 91,65$ %, $PPV_p = 98,59$ %, $SE_i = 95,00$ % $SP_i = 90,24$ % y $ACC_i = 91,80$ %.

Los resultados son comparables con los obtenidos en estudios previos según el criterio basado en imagen. Hemos encontrado varios métodos que se han evaluado utilizando la base de datos DIARETDB1. Walter et al. (2002) obtuvieron $SE_i = 86,00$ % y $SP_i = 69,00$ %. En (Harangi and Hajdu, 2014) se obtuvieron valores de $SE_i = 92,00$ % y $SP_i = 68,00$ %. Liu et al. (2016) alcanzaron $SE_i = 83,00$ % y $SP_i = 75,00$ %. El método propuesto en (Zhou et al., 2017b) mostró $SE_i = 88,00$ % y $SP_i = 95,00$ %. Finalmente, el trabajo de Adem (2018) presentó valores altos de $SE_i = 99,20$ % y $SP_i = 97,97$ %. El valor de SE_i alcanzado con nuestro método es superior a los obtenidos en estudios anteriores, con una sola excepción (Adem,

2018). Cabe señalar que el conjunto de test en (Adem, 2018) estaba compuesto por imágenes de las bases de datos DIARETDB0 y DRIMDB además de DIARETDB1. Además, en la fase de entrenamiento se utilizó el conjunto de entrenamiento de la base de datos DiaretDB1. Al comparar nuestros resultados con enfoques anteriores, debe tenerse en cuenta que el método propuesto se ha desarrollado utilizando solo imágenes de la base de datos privada. A pesar de las diferencias entre bases de datos, nuestros resultados sobre DIARETDB1 demuestran la robustez del método propuesto.

C.5.5   Graduación de la severidad

El método propuesto se evaluó sobre 32,017 retinografías de calidad adecuada de la base de datos Kaggle, alcanzando un valor de Kappa ponderado cuadrático (QWK) de 0.78. Este resultado es similar al de otros métodos previos. Para establecer una comparación directa, se incluyen los estudios en los que se ha utilizado la misma base de datos, aunque los conjuntos de entrenamiento, validación y test pueden variar y, por tanto, la comparación debe hacerse con cuidado. Con nuestro método, obtuvimos un QWK superior al obtenido por González-Gonzalo et al. (2018), de la Torre et al. (2018) y el reciente estudio de Araújo et al. (2020). El trabajo posterior de de la Torre et al. (2020) alcanzó un $QWK = 0{,}80$ superior al nuestro. Sin embargo, utilizaron para la fase de entrenamiento parte de las imágenes destinadas a test y, por lo tanto, evaluaron su método sobre un subconjunto reducido de 10.000 imágenes. Krause et al. (2018) también superaron nuestro método en términos de QWK, con un valor de 0.84. Sin embargo, su modelo se entrenó con un gran conjunto de datos privado compuesto por más de 1.6 millones de retinografías. Además, su conjunto de test estaba formado por 1,818 imágenes, mucho menor que el nuestro (32,017). Finalmente, Wang et al. (2017) alcanzaron el $QWK = 0{,}85$ más alto. Sin embargo, su método requería la anotación manual de las lesiones de algunas imágenes así como las retinografías de los dos ojos de cada paciente, que no siempre están disponibles.

C.6   Conclusiones

En base a los resultados obtenidos en la presente Tesis Doctoral se pueden extraer las siguientes conclusiones:

   1) El análisis de selección de características realizado con la técnica FCBF nos
      permitió concluir que los métodos SSEQ y NIQE son útiles para evaluar la

calidad de la imagen retiniana y pueden ser complementarios con las características de nitidez y luminosidad.

2) Los enfoques de *deep learning* están superando a las técnicas tradicionales para muchos problemas de procesado de imagen y, en particular, para la evaluación de la calidad de las retinografías. En este contexto, la arquitectura InceptionResNetV2 junto con *transfer learning* ha permitido lograr buenos resultados y una alta capacidad de generalización sin necesidad de una base de datos excesivamente grande.

3) Los mapas de saliencia propuestos para la localización automática del DO y la fóvea han demostrado una alta precisión incluso en las imágenes donde dichas estructuras no tenían el aspecto estándar. La evaluación sobre varias bases de datos ha demostrado que el método propuesto es muy robusto y efectivo.

4) La descomposición de la retinografía en varias capas ha permitido extraer un conjunto de características relevantes para la detección de LRs y EXs, analizadas mediante la técnica FCBF. En esta descomposición, la vasculatura coroidea visible en las retinas atigradas y las características reflectantes de la retina han demostrado ser útiles.

5) El concepto de superpíxel, a diferencia del píxel, es adecuado para representar las entidades naturales y separar las diferentes estructuras de la retina. En particular, el algoritmo ERS ha demostrado ser rápido y efectivo para la segmentación de LRs, superando al método SLIC.

6) El método de *deep learning* propuesto para la calificación de DR alcanzó resultados comparables a los métodos publicados anteriormente. Separar la atención de los píxeles brillantes y oscuros en la retinografía permitió mejorar la optimización del modelo y facilitar su gestión.

7) Los resultados de cualquiera de los métodos propuestos en esta Tesis Doctoral sugieren que podrían ser utilizados, por separado o en conjunto, como parte de un sistema de cribado automático, suponiendo una ayuda diagnóstica para la detección precoz de RD. De esta forma, los pacientes diabéticos podrían recibir una mejor atención para su salud ocular evitando la pérdida de visión.

Bibliography

Abdel-Hamid, L., El-Rafei, A., El-Ramly, S., Michelson, G., Hornegger, J., sep 2016. Retinal image quality assessment based on image clarity and content. Journal of biomedical optics 21 (9), 096007.

Abdel-Hamid, L., El-Rafei, A., Michelson, G., nov 2017. No-reference quality index for color retinal images. Computers in biology and medicine 90, 68–75.

Abdullah, M., Fraz, M. M., Barman, S. A., 2016. Localization and segmentation of optic disc in retinal images using circular Hough transform and grow-cut algorithm. PeerJ 2016 (5), e2003.

Abed, S., Al-Roomi, S. A., Al-Shayeji, M., dec 2016. Effective optic disc detection method based on swarm intelligence techniques and novel pre-processing steps. Applied Soft Computing Journal 49, 146–163.

Abramoff, M. D., Garvin, M. K., Sonka, M., 2010. Retinal imaging and image analysis. IEEE Reviews in Biomedical Engineering 3, 169–208.

Abràmoff, M. D., Folk, J. C., Han, D. P., Walker, J. D., Williams, D. F., Russell, S. R., Massin, P., Cochener, B., Gain, P., Tang, L., Lamard, M., Moga, D. C., Quellec, G., Niemeijer, M., 2013. Automated analysis of retinal images for detection of referable diabetic retinopathy. JAMA Ophthalmology 131 (3), 351–357.

Abràmoff, M. D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J. C., Niemeijer, M., oct 2016. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. Investigative Ophthalmology and Visual Science 57 (13), 5200–5206.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (11), 2274–2281.

Adem, K., 2018. Exudate detection for diabetic retinopathy with circular Hough transformation and convolutional neural networks. Expert Systems with Applications 114, 289–295.

Aja-Fernández, S., Estépar, R. S. J., Alberola-López, C., Westin, C. F., 2006. Image quality assessment based on local variance. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference 2006, 4815–4818.

Al-Bander, B., Al-Nuaimy, W., Williams, B. M., Zheng, Y., feb 2018. Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc. Biomedical Signal Processing and Control 40, 91–101.

Alshayeji, M., Al-Roomi, S. A., Abed, S., jun 2017. Optic disc detection in retinal fundus images using gravitational law-based edge detection. Medical and Biological Engineering and Computing 55 (6), 935–948.

American Diabetes Association, jan 2014. Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 37 (1), S81–S90.

American Diabetes Association Professional Practice Committee, jan 2022. 9. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes—2022. Diabetes Care 45 (Supplement 1), S125–S143.

Angenent, S., Pichon, E., Tannenbaum, A., 2006. Mathematical methods in medical image processing. Bulletin (new series) of the American Mathematical Society 43 (3), 365.

Antoine, J. P., Carrette, P., Murenzi, R., Piette, B., apr 1993. Image analysis with two-dimensional continuous wavelet transform. Signal Processing 31 (3), 241–272.

Antoine, J. P., Murenzi, R., aug 1996. Two-dimensional directional wavelets and the scale-angle representation. Signal Processing 52 (3), 259–281.

Aquino, A., dec 2014. Establishing the macular grading grid by means of fovea centre detection using anatomical-based and visual-based features. Computers in Biology and Medicine 55, 61–73.

Araújo, T., Aresta, G., Mendonça, L., Penas, S., Maia, C., Carneiro, Â., Mendonça, A. M., Campilho, A., jul 2020. DR—GRADUATE: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images. Medical Image Analysis 63, 101715.

Bachiller, A., Lubeiro, A., Díez, Á., Suazo, V., Domínguez, C., Blanco, J. A., Ayuso, M., Hornero, R., Poza, J., Molina, V., sep 2015. Decreased entropy modulation of EEG response to novelty and relevance in schizophrenia during a P300 task. European archives of psychiatry and clinical neuroscience 265 (6), 525–535.

Bankman, I. N. I. N., 2000. Handbook of medical imaging: processing and analysis. Academic Press.

Bartling, H., Wanger, P., Martin, L., sep 2009. Automated quality evaluation of digital fundus photographs. Acta ophthalmologica 87 (6), 643–647.

Basit, A., Fraz, M. M., apr 2015. Optic disc detection and boundary extraction in retinal images. Applied optics 54 (11), 3440–7.

Bellemo, V., Lim, Z. W., Lim, G., Nguyen, Q. D., Xie, Y., Yip, M. Y., Hamzah, H., Ho, J., Lee, X. Q., Hsu, W., Lee, M. L., Musonda, L., Chandran, M., Chipalo-Mutati, G., Muma, M., Tan, G. S., Sivaprasad, S., Menon, G., Wong, T. Y., Ting, D. S., may 2019. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. The Lancet. Digital health 1 (1), e35–e44.

Besenczi, R., Tóth, J., Hajdu, A., jan 2016. A review on automatic analysis techniques for color fundus photographs. Computational and Structural Biotechnology Journal 14, 371–384.

Bhatkalkar, B. J., Vighnesh Nayak, S., Shenoy, S. V., Vijaya Arjunan, R., 2021. FundusPosNet: A Deep Learning Driven Heatmap Regression Model for the Joint Localization of Optic Disc and Fovea Centers in Color Fundus Images. IEEE Access 9, 159071–159080.

Bishop, C., 1995. Neural Networks for Pattern Recognition, 1st Edition. Vol. 1995. Oxford University Press, New York.

Bishop, C. M., 2006. Pattern Recognition and Machine Learning. Springer.

Bogacsovics, G., Toth, J., Hajdu, A., Harangi, B., jul 2022. Enhancing CNNs through the use of hand-crafted features in automated fundus image classification. Biomedical Signal Processing and Control 76, 103685.

Bronzino, J. D., Peterson, D. R., 2017. Biomedical Engineering Fundamentals, fourth edi Edition. CRC Press.

Chalakkal, R. J., Abdulla, W. H., Thulaseedharan, S. S., nov 2018. Automatic detection and segmentation of optic disc and fovea in retinal images. IET Image Processing 12 (11), 2100–2110.

Chalakkal, R. J., Abdulla, W. H., Thulaseedharan, S. S., 2019. Quality and content analysis of fundus images using deep learning. Computers in Biology and Medicine.

Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal Of Artificial Intelligence Research 16, 321–357.

Chen, L. C., Yang, Y., Wang, J., Xu, W., Yuille, A. L., dec 2016. Attention to Scale: Scale-Aware Semantic Image Segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol. 2016-Decem. IEEE Computer Society, pp. 3640–3649.

Chollet, F., nov 2017. Xception: Deep learning with depthwise separable convolutions. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. Vol. 2017-Janua. Institute of Electrical and Electronics Engineers Inc., pp. 1800–1807.

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20 (1), 37–46.

Cohen, J., oct 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological Bulletin 70 (4), 213–220.

Costa, P., Araujo, T., Aresta, G., Galdran, A., Mendonca, A. M., Smailagic, A., Campilho, A., may 2019. EyeWeS: Weakly supervised pre-trained convolutional neural networks for diabetic retinopathy detection. In: Proceedings of the 16th International Conference on Machine Vision Applications, MVA 2019. Institute of Electrical and Electronics Engineers Inc., pp. 1–6.

Cover, T. M., Thomas, J. A., 1991. Elements of Information Theory, 2nd Edition. Jon Wiley.

Coyner, A. S., Swan, R., Brown, J. M., Kalpathy-Cramer, J., Kim, S. J., Campbell, J. P., Jonas, K. E., Ostmo, S., Chan, R. V. P., Chiang, M. F., 2018. Deep Learning for Image Quality Assessment of Fundus Images in Retinopathy of Prematurity. AMIA ... Annual Symposium proceedings. AMIA Symposium 2018, 1224–1232.

Davis, H., Russell, S., Barriga, E., Abramoff, M., Soliz, P., 2009. Vision-based, real-time retinal image quality assessment. Proceedings - IEEE Symposium on Computer-Based Medical Systems.

de la Torre, J., Puig, D., Valls, A., apr 2018. Weighted kappa loss function for multi-class classification of ordinal data in deep learning. Pattern Recognition Letters 105, 144–154.

de la Torre, J., Valls, A., Puig, D., jul 2020. A deep learning interpretable classifier for diabetic retinopathy disease grading. Neurocomputing 396, 465–476.

Decencière, E., Zhang, X., Cazuguel, G., Laÿ, B., Cochener, B., Trone, C., Gain, P., Ordóñez-Varela, J. R., Massin, P., Erginay, A., Charton, B., Klein, J. C., 2014. Feedback on a publicly distributed image database: The Messidor database. Image Analysis and Stereology 33 (3), 231–234.

Department of Health and Social Care (Government of the United Kingdom), 2002. National Service Framework for Diabetes: Standards. Tech. rep., Department of Health, London.

Diabetic Retinopathy Study Research Group, 1981. A modification of the Airlie House classification of diabetic retinopathy. Invest Ophthalmol Vis Sci 21, 210–226.

Díaz-Pernil, D., Fondón, I., Peña-Cantillana, F., Gutiérrez-Naranjo, M. A., nov 2016. Fully automatized parallel segmentation of the optic disc in retinal fundus images. Pattern Recognition Letters 83, 99–107.

Engerman, R. L., oct 1989. Pathogenesis of Diabetic Retinopathy. Diabetes 38 (10), 1203–1206.

Fasih, M., Langlois, J. P., Ben Tahar, H., Cheriet, F., mar 2014. Retinal image quality assessment using generic features. In: Medical Imaging 2014: Computer-Aided Diagnosis. Vol. 9035. SPIE, p. 90352Z.

Fathi, A., Naghsh-Nilchi, A. R., jan 2013. Automatic wavelet-based retinal blood vessels segmentation and vessel diameter estimation. Biomedical Signal Processing and Control 8 (1), 71–80.

Fawcett, T., jun 2006. An introduction to ROC analysis. Pattern Recognition Letters 27 (8), 861–874.

Fleming, A., Philip, S., Goatman, K., Olson, J., Sharp, P., sep 2006. Automated microaneurysm detection using local contrast normalization and local vessel detection. IEEE Transactions on Medical Imaging 25 (9), 1223–1232.

Fleming, A. D., Goatmana, K. A., Williams, G. J., Philip, S., Sharpa, P. F., Olson, J. A., 2008. Automated detection of blot haemorrhages as a sign of referable diabetic retinopathy. In: Proceedings of the Medical Image Understanding and Analysis (MIUA2008). pp. 235–239.

Fleming, A. D., Philip, S., Goatman, K. A., Sharp, P. F., Olson, J. A., sep 2012. Automated clarity assessment of retinal images using regionally based structural and statistical measures. Medical Engineering and Physics 34 (7), 849–859.

Foracchia, M., Grisan, E., Ruggeri, A., oct 2004. Detection of optic disc in retinal images by means of a geometrical model of vessel structure. IEEE Transactions on Medical Imaging 23 (10), 1189–1195.

Foracchia, M., Grisan, E., Ruggeri, A., jun 2005. Luminosity and contrast normalization in retinal images. Medical Image Analysis 9 (3), 179–190.

García, M., Member, S., Sánchez, C. I., Member, S., López, M. I., Hornero, R., 2008. Automatic Detection of Red Lesions in Retinal Images Using a Multilayer Perceptron Neural Network. pp. 5425–5428.

García, M., Sánchez, C. I., López, M. I., Abásolo, D., Hornero, R., 2009. Neural network based detection of hard exudates in retinal images. Computer Methods and Programs in Biomedicine 93 (1), 9–19.

García, M., López, M. I., Álvarez, D., Hornero, R., 2010. Assessment of four neural network based classifiers to automatically detect red lesions in retinal images. Medical Engineering and Physics 32 (10), 1085–1093.

GeethaRamani, R., Balasubramanian, L., jul 2018. Macula segmentation and fovea localization employing image processing and heuristic based clustering for automated retinal screening. Computer Methods and Programs in Biomedicine 160, 153–163.

Gegundez-Arias, M. E., Marin, D., Bravo, J. M., Suero, A., jul 2013. Locating the fovea center position in digital fundus images using thresholding and feature extraction techniques. Computerized Medical Imaging and Graphics 37 (5-6), 386–393.

Giachetti, A., Ballerini, L., Trucco, E., Wilson, P. J., jul 2013. The use of radial symmetry to localize retinal landmarks. Computerized Medical Imaging and Graphics 37 (5-6), 369–376.

Giancardo, L., Chaum, E., Karnowski, T. P., Meriaudeau, F., Tobin, K. W., Li, Y., 2009. Bright retinal lesions detection using color fundus images containing reflective features. IFMBE Proceedings 25 (11), 292–295.

Giancardo, L., Meriaudeau, F., P, T., Chaum, E., Tobi, K., 2010. Quality Assessment of Retinal Fundus Images using Elliptical Local Vessel Density. New Developments in Biomedical Engineering Ed. Domeni, 201–224.

Gonzalez, R. C., Woods, R. E., 2009. Digital image processing. Dorling Kindersley.

González-Gonzalo, C., Liefers, B., Van Ginneken, B., Sánchez, C. I., 2018. Improving weakly-supervised lesion localization with iterative saliency map refinement. In: International Conference on Medical Imaging with Deep Learning.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning.

Grisan, E., Ruggeri, A., 2005. A hierarchical bayesian classification for non-vascular lesions detection in fundus images. Proceedings of the 3rd EuropeanMedical and Biological Enginerring Conference 11.

Grzybowski, A., Brona, P., Lim, G., Ruamviboonsuk, P., Tan, G. S., Abramoff, M., Ting, D. S., 2020. Artificial intelligence for diabetic retinopathy screening: a review. Eye (Basingstoke) 34 (3), 451–460.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., Webster, D. R., dec 2016. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 316 (22), 2402.

Guo, S., Wang, K., Kang, H., Liu, T., Gao, Y., Li, T., 2019. Bin loss for hard exudates segmentation in fundus images. Neurocomputing.

Guyon, I., Elisseeff, A., 2003. An Introduction to Variable and Feature Selection. Journal of Machine Learning Research 3, 1157–1182.

Harangi, B., Hajdu, A., 2014. Automatic exudate detection by fusing multiple active contours and regionwise classification. Computers in Biology and Medicine 54, 156–171.

Harangi, B., Hajdu, A., oct 2015. Detection of the optic disc in fundus images by combining probability models. Computers in Biology and Medicine 65, 10–24.

Hasan, M. K., Alam, M. A., Elahi, M. T. E., Roy, S., Martí, R., jan 2021. DRNet: Segmentation and localization of optic disc and Fovea from diabetic retinopathy image. Artificial Intelligence in Medicine 111, 102001.

Hatanaka, Y., Nakagawa, T., Hayashi, Y., Kakogawa, M., Sawada, A., Kawase, K., Hara, T., Fujita, H., mar 2008. Improvement of automatic hemorrhage detection methods using brightness correction on fundus images. In: Giger, M. L., Karssemeijer, N. (Eds.), Medical Imaging 2008: Computer-Aided Diagnosis. Vol. 6915.

Haykin, S. S., 1999. Neural networks: a comprehensive foundation. Prentice Hall.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol. 2016-Decem. pp. 770–778.

Hoover, A., Goldbaum, M., aug 2003. Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. IEEE Transactions on Medical Imaging 22 (8), 951–958.

Hsiao, H. K., Liu, C. C., Yu, C. Y., Kuo, S. W., Yu, S. S., 2012. A novel optic disc detection scheme on retinal images. Expert Systems with Applications 39 (12), 10600–10606.

Hsu, W., Pallawala, P. M., Lee, M. L., Eong, K. G. A., 2001. The role of domain knowledge in the detection of retinal hard exudates. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol. 2. IEEE Comput. Soc, pp. II–246–II–251.

Huang, G.-B., Chen, Y.-Q., H.A., B., may 2000. Classification ability of single hidden layer feedforward neural networks. IEEE Transactions on Neural Networks 11 (3), 799–801.

Huang, S., Li, J., Xiao, Y., Shen, N., Xu, T., 2022. RTNet: Relation Transformer Network for Diabetic Retinopathy Multi-lesion Segmentation. IEEE Transactions on Medical Imaging.

Imani, E., Pourreza, H. R., 2016. A novel method for retinal exudate segmentation using signal separation algorithm. Computer Methods and Programs in Biomedicine 133, 195–205.

Islam, M. M., Yang, H. C., Poly, T. N., Jian, W. S., (Jack) Li, Y. C., jul 2020. Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis. Computer Methods and Programs in Biomedicine 191, 105320.

Jaafar, H. F., Nandi, A. K., Al-Nuaimy, W., 2011. Automated detection of red lesions from digital colour fundus photographs. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS. pp. 6232–6235.

Jain, A. K., 1989. Digital Processing of Speckle Images. In: Fundamentals of Digital Image Processing. Prentice Hall, New Jersey, United States of America, pp. 68,71,73.

Jia Deng, Wei Dong, Socher, R., Li-Jia Li, Kai Li, Li Fei-Fei, 2009. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255.

Jiménez-García, J., Romero-Oraá, R., García, M., López-Gálvez, M., Hornero, R., Jiménez-García, J., Romero-Oraá, R., García, M., López-Gálvez, M. I., Hornero, R., mar 2019. Combination of Global Features for the Automatic Quality Assessment of Retinal Images. Entropy 21 (3), 311.

Kaggle, 2015. Diabetic Retinopathy Detection competition.

Kar, S. S., Maity, S. P., mar 2018. Automatic Detection of Retinal Lesions for Screening of Diabetic Retinopathy. IEEE Transactions on Biomedical Engineering 65 (3), 608–618.

Kauppi, T., Kalesnykiene, V., Kamarainen, J. K., Lensu, L., Sorri, I., Raninen, A., Voutilainen, R., Pietilä, J., Kälviäinen, H., Uusitalo, H., 2007. The DIARETDB1 diabetic retinopathy database and evaluation protocol. In: BMVC 2007 - Proceedings of the British Machine Vision Conference 2007.

Kaur, J., Mittal, D., 2018. A generalized method for the segmentation of exudates from pathological retinal fundus images. Biocybernetics and Biomedical Engineering 38 (1), 27–53.

Kels, B. D., Grzybowski, A., Grant-Kels, J. M., mar 2015. Human ocular anatomy. Clinics in Dermatology 33 (2), 140–146.

Khojasteh, P., Passos Júnior, L. A., Carvalho, T., Rezende, E., Aliahmad, B., Papa, J. P., Kumar, D. K., 2019. Exudate detection in fundus images using deeply-learnable features. Computers in Biology and Medicine 104, 62–69.

Koh, J. E., Acharya, U. R., Hagiwara, Y., Raghavendra, U., Tan, J. H., Sree, S. V., Bhandary, S. V., Rao, A. K., Sivaprasad, S., Chua, K. C., Laude, A., Tong, L., may 2017. Diagnosis of retinal health in digital fundus images using continuous wavelet transform (CWT) and entropies. Computers in biology and medicine 84, 89–97.

Krause, J., Gulshan, V., Rahimy, E., Karth, P., Widner, K., Corrado, G. S., Peng, L., Webster, D. R., aug 2018. Grader Variability and the Importance of Reference Standards for Evaluating Machine Learning Models for Diabetic Retinopathy. Ophthalmology 125 (8), 1264–1272.

Lalonde, M., Gagnon, L., Boucher, M. C., 2001a. Automatic image quality assessment in optical fundus images. CRIM Report CRIM-00/12-11, 259–271.

Lalonde, M., Beaulieu, M., Gagnon, L., 2001b. Fast and robust optic disc detection using pyramidal decomposition and hausdorff-based template matching. IEEE Transactions on Medical Imaging 20 (11), 1193–1200.

Lam, C., Yu, C., Huang, L., Rubin, D., jan 2018. Retinal lesion detection with deep learning using image patches. Investigative Ophthalmology and Visual Science 59 (1), 590–596.

Lazar, I., Hajdu, A., feb 2013. Retinal Microaneurysm Detection Through Local Rotating Cross-Section Profile Analysis. IEEE Transactions on Medical Imaging 32 (2), 400–407.

Lee, S. C., Wang, Y., may 1999. Automatic retinal image quality assessment and enhancement. https://doi.org/10.1117/12.348562 3661, 1581–1590.

Li, H., Chutatape, O., feb 2004. Automated Feature Extraction in Color Retinal Images by a Model Based Approach. IEEE Transactions on Biomedical Engineering 51 (2), 246–254.

Li, L., Xia, W., Lin, W., Fang, Y., Wang, S., may 2017. No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features. IEEE Transactions on Multimedia 19 (5), 1030–1040.

Li, S., Zhao, R., Zou, H., Guo, L., feb 2021a. Artificial intelligence for diabetic retinopathy. Chinese medical journal 135 (3), 253–260.

Li, T., Bo, W., Hu, C., Kang, H., Liu, H., Wang, K., Fu, H., apr 2021b. Applications of deep learning in fundus images: A review.

Lin, J. W., Weng, Q., Xue, L. Y., Cao, X. R., Yu, L., jun 2017a. A retinal image sharpness metric based on histogram of edge width:. http://dx.doi.org/10.1177/1748301817713184 11 (3), 292–300.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., aug 2017b. Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 42 (2), 318–327.

Liu, L., Liu, B., Huang, H., Bovik, A. C., 2014. No-reference image quality assessment based on spatial and spectral entropies. Signal Processing: Image Communication.

Liu, M. Y., Tuzel, O., Ramalingam, S., Chellappa, R., jun 2011. Entropy rate superpixel segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, pp. 2097–2104.

Liu, Q., Zou, B., Chen, J., Ke, W., Yue, K., Chen, Z., Zhao, G., 2016. A location-to-segmentation strategy for automatic exudate segmentation in colour retinal fundus images. Computerized Medical Imaging and Graphics 55, 78–86.

Lu, S., dec 2011. Accurate and efficient optic disc detection and segmentation by a circular transformation. IEEE Transactions on Medical Imaging 30 (12), 2126–2133.

Lu, S., Lim, J. H., jan 2011. Automatic optic disc detection from retinal images by a line operator. IEEE Transactions on Biomedical Engineering 58 (1), 88–94.

Lupaşcu, C. A., Di Rosa, L., Tegolo, D., 2008. Automated detection of optic disc location in retinal images. In: Proceedings - IEEE Symposium on Computer-Based Medical Systems. pp. 17–22.

Lyu, X., Li, H., Zhen, Y., Ji, X., Zhang, S., 2017. Deep tessellated retinal image detection using Convolutional Neural Networks. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 676–680.

Marín, D., Aquino, A., Gegúndez-Arias, M. E., Bravo, J. M., jan 2011. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. IEEE Transactions on Medical Imaging 30 (1), 146–158.

Marrugo, A. G., Millán, M. S., Baez, H. C. A., Cristóbal, G., Gabarda, S., jul 2012. Anisotropy-based robust focus measure for non-mydriatic retinal imaging. https://doi.org/10.1117/1.JBO.17.7.076021 17 (7), 076021.

Marrugo Hernández, A. G., Millán Garcia-Varela, M. S., Gabriel, C., Gabarda, S., Abril Báez, H., 2011. No-reference quality metrics for eye fundus imaging. Computer analysis of images and patterns : 14th International Conference, CAIP 2011, Seville, Spain, August 29-31, 2011, Proceedings 6854, 486–493.

Medhi, J. P., Dandapat, S., 2015. Automatic detection of fovea using property of vessel free region. 2015 21st National Conference on Communications, NCC 2015.

Mendonça, A. M., Member, S., Campilho, A., 2006. Segmentation of Retinal Blood Vessels by Combining the Detection of Centerlines and Morphological Reconstruction 25 (9), 1200–1213.

Mittal, A., Moorthy, A. K., Bovik, A. C., 2012. No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing 21 (12), 4695–4708.

Mittal, A., Soundararajan, R., Bovik, A. C., 2013. Making a 'completely blind' image quality analyzer. IEEE Signal Processing Letters 20 (3), 209–212.

Mohamed, Q., Gillies, M. C., Wong, T. Y., aug 2007. Management of Diabetic Retinopathy: A Systematic Review. JAMA 298 (8), 902–916.

Moller, M. F., 1993. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning. Tech. rep.

Mookiah, M. R. K., Acharya, U. R., Chua, C. K., Lim, C. M., Ng, E. Y. K., Laude, A., 2013. Computer-aided diagnosis of diabetic retinopathy: A review. Computers in Biology and Medicine 43 (12), 2136–2155.

Moorthy, A. K., Bovik, A. C., 2010. A two-step framework for constructing blind image quality indices. IEEE Signal Processing Letters 17 (5), 513–516.

Morales, S., Naranjo, V., Angulo, U., Alcaniz, M., 2013. Automatic detection of optic disc based on PCA and mathematical morphology. IEEE Transactions on Medical Imaging 32 (4), 786–796.

Nguyen, U. T., Bhuiyan, A., Park, L. A., Ramamohanarao, K., 2013. An effective retinal blood vessel segmentation method using multi-scale line detection. Pattern Recognition 46 (3), 703–715.

Niemeijer, M., Van Ginneken, B., Staal, J., Suttorp-Schulten, M. S. A., Abràmoff, M. D., 2005. Automatic detection of red lesions in digital color fundus photographs. IEEE Transactions on Medical Imaging 24 (5), 584–592.

Niemeijer, M., Abràmoff, M. D., van Ginneken, B., 2006. Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. Medical Image Analysis 10 (6), 888–898.

Niemeijer, M., Van Ginneken, B., Russell, S. R., Suttorp-Schulten, M. S. A., Abràmoff, M. D., 2007a. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. Investigative Ophthalmology and Visual Science 48 (5), 2260–2267.

Niemeijer, M., Abràmoff, M. D., Van Ginneken, B., jan 2007b. Segmentation of the optic disc, macula and vascular arch in fundus photographs. IEEE Transactions on Medical Imaging 26 (1), 116–127.

Niemeijer, M., Abràmoff, M. D., van Ginneken, B., 2009. Fast detection of the optic disc and fovea in color fundus photographs. Medical Image Analysis 13 (6), 859–870.

Orlando, J. I., Prokofyeva, E., del Fresno, M., Blaschko, M. B., jan 2018. An ensemble deep learning based approach for red lesion detection in fundus images. Computer Methods and Programs in Biomedicine 153, 115–127.

Osareh, A., 2004. Automated Identification of Diabetic Retinal Exudates and the Optic Disc. Ph.D. thesis.

Osareh, A., Shadgar, B., Markham, R., 2009. A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images. In: IEEE Transactions on Information Technology in Biomedicine. Vol. 13. pp. 535–545.

Pan, S. J., Yang, Q., oct 2010. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering 22 (10), 1345–1359.

Patton, N., Aslam, T. M., MacGillivray, T., Deary, I. J., Dhillon, B., Eikelboom, R. H., Yogesan, K., Constable, I. J., 2006. Retinal image analysis: Concepts, applications and potential. Progress in Retinal and Eye Research 25 (1), 99–127.

Paulus, J., Meier, J., Bock, R., Hornegger, J., Michelson, G., 2010. Automated quality assessment of retinal fundus photos. International Journal of Computer Assisted Radiology and Surgery 5 (6), 557–564.

Pereira, C., Gonçalves, L., Ferreira, M., mar 2013. Optic disc detection in color fundus images using ant colony optimization. Medical and Biological Engineering and Computing 51 (3), 295–303.

Pereira, C., Gonçalves, L., Ferreira, M., 2015. Exudate segmentation in fundus images using an ant colony optimization approach. Information Sciences 296, 14–24.

Perez, L., Wang, J., dec 2017. The Effectiveness of Data Augmentation in Image Classification using Deep Learning.

Pires Dias, J. M., Oliveira, C. M., Da Silva Cruz, L. A., 2014. Retinal image quality assessment using generic image quality indicators. Information Fusion 19 (1), 73–90.

Poza, J., Hornero, R., Abásolo, D., Fernández, A., García, M., dec 2007. Extraction of spectral based measures from MEG background oscillations in Alzheimer's disease. Medical Engineering and Physics 29 (10), 1073–1083.

Prentašić, P., Lončarić, S., 2016. Detection of exudates in fundus photographs using deep neural networks and anatomical landmark detection fusion. Computer Methods and Programs in Biomedicine 137, 281–292.

Quellec, G., Lamard, M., Josselin, P., Cazuguel, G., Cochener, B., Roux, C., sep 2008. Optimal Wavelet Transform for the Detection of Microaneurysms in Retina Photographs. IEEE Transactions on Medical Imaging 27 (9), 1230–1241.

Quellec, G., Charrière, K., Boudi, Y., Cochener, B., Lamard, M., jul 2017. Deep image mining for diabetic retinopathy screening. Medical Image Analysis 39, 178–193.

Qureshi, R. J., Kovacs, L., Harangi, B., Nagy, B., Peto, T., Hajdu, A., 2012. Combining algorithms for automatic detection of optic disc and macula in fundus images. Computer Vision and Image Understanding 116 (1), 138–145.

Rahebi, J., Hardalaç, F., mar 2016. A new approach to optic disc detection in human retinal images using the firefly algorithm. Medical and Biological Engineering and Computing 54 (2-3), 453–461.

Raj, A., Tiwari, A. K., Martini, M. G., jun 2019. Fundus image quality assessment: survey, challenges, and future scope. IET Image Processing 13 (8), 1211–1224.

Rakhlin, A., jun 2018. Diabetic Retinopathy detection through integration of Deep Learning classification framework. bioRxiv, 225508.

Rangayyan, R. M., 2004. Biomedical image analysis.

Rasta, S. H., Partovi, M. E., Seyedarabi, H., Javadzadeh, A., 2015. A comparative study on preprocessing techniques in diabetic retinopathy retinal images: illumination correction and contrast enhancement. Journal of medical signals and sensors 5 (1), 40–8.

Remeseiro, B., Mendonca, A. M., Campilho, A., jun 2017. Objective quality assessment of retinal images based on texture features. Proceedings of the International Joint Conference on Neural Networks 2017-May, 4520–4527.

Ritter, F., Boskamp, T., Homeyer, A., Laue, H., Schwier, M., Link, F., Peitgen, H. O., nov 2011. Medical image analysis. IEEE Pulse 2 (6), 60–70.

Rodger, J. A., feb 2019. QuantumIS: A Qualia Consciousness Awareness and Information Theory Quale Approach to Reducing Strategic Decision-Making Entropy. Entropy (Basel, Switzerland) 21 (2).

Romero Oraá, R., García Gadañón, M., López Gálvez, M. I., Manco Lavado, F., Hornero Sánchez, R., 2016. Localización automática de la papila y la fóvea en retinografías. Congreso Anual de la Sociedad Española de Ingeniería Biomédica, 173–176.

Romero Oraá, R., Jiménez García, J., García Gadañón, M., López Gálvez, M. I., Hornero Sánchez, R., 2018. Clasificación de superpíxeles para la detección automática de lesiones rojizas en imágenes de fondo de ojo. In: Congreso Anual de la Sociedad Española de Ingeniería Biomédica. pp. 25–28.

Romero-Oraá, R., Jiménez-García, J., García, M., López-Gálvez, M. I., Oraá-Pérez, J., Hornero, R., apr 2019. Entropy rate superpixel classification for automatic red lesion detection in fundus images. Entropy 21 (4).

Romero-Oraá, R., García, M., Oraá-Pérez, J., López, M. I., Hornero, R., jun 2020. A robust method for the automatic location of the optic disc and the fovea in fundus images. Computer Methods and Programs in Biomedicine 196, 105599.

Romero Oraá, R., García, M., Oraá-Pérez, J., María, ., López, I., Hornero, R., 2020. Automatic fundus image quality assessment: diagnostic accuracy in clinical practice. Investigative Ophthalmology and Visual Science 61, 2033.

Romero-Oraá, R., García, M., Oraá-Pérez, J., López-Gálvez, M. I., Hornero, R., nov 2020. Effective fundus image decomposition for the detection of red lesions and hard exudates to aid in the diagnosis of diabetic retinopathy. Sensors (Switzerland) 20 (22), 1–17.

Roychowdhury, S., Koozekanani, D. D., Parhi, K. K., nov 2012. Screening fundus images for diabetic retinopathy. In: Conference Record - Asilomar Conference on Signals, Systems and Computers. IEEE, pp. 1641–1645.

Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala, A. A., Ogurtsova, K., Shaw, J. E., Bright, D., Williams, R., nov 2019. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. Diabetes Research and Clinical Practice 157, 107843.

Saeys, Y., Inza, I., Larrañaga, P., oct 2007. A review of feature selection techniques in bioinformatics. Bioinformatics 23 (19), 2507–2517.

Saha, S. K., Fernando, B., Cuadros, J., Xiao, D., Kanagasingam, Y., dec 2018. Automated Quality Assessment of Colour Fundus Images for Diabetic Retinopathy Screening in Telemedicine. Journal of Digital Imaging 31 (6), 869–878.

Salazar-Gonzalez, A. G., Li, Y., Liu, X., 2011. Optic disc segmentation by incorporating blood vessel compensation. In: IEEE SSCI 2011 - Symposium Series on Computational Intelligence - CIMI 2011: 2011 IEEE 3rd International Workshop on Computational Intelligence in Medical Imaging. pp. 1–8.

Sánchez, C. I., Hornero, R., López, M. I., Aboy, M., Poza, J., Abásolo, D., 2008. A novel automatic image processing algorithm for detection of hard exudates based on retinal image analysis. Medical Engineering and Physics 30 (3), 350–357.

Sánchez, C. I., García, M., Mayo, A., López, M. I., Hornero, R., 2009. Retinal image analysis based on mixture models to detect hard exudates. Medical Image Analysis 13 (4), 650–658.

Sánchez, C. I., Niemeijer, M., Dumitrescu, A. V., Suttorp-Schulten, M. S. A., Abràmoff, M. D., van Ginneken, B., jun 2011. Evaluation of a Computer-Aided Diagnosis System for Diabetic Retinopathy Screening on Public Data. Investigative Opthalmology and Visual Science 52 (7), 4866.

Seoud, L., Hurtut, T., Chelbi, J., Cheriet, F., Langlois, J. M. P., 2016. Red Lesion Detection Using Dynamic Shape Features for Diabetic Retinopathy Screening. IEEE Transactions on Medical Imaging 35 (4), 1116–1126.

Sevik, U., Köse, C., Berber, T., Erdöl, H., apr 2014. Identification of suitable fundus images using automated quality assessment methods. Journal of Biomedical Optics 19 (4), 046006.

Shao, F., Yang, Y., Jiang, Q., Jiang, G., Ho, Y. S., nov 2017. Automated Quality Assessment of Fundus Images via Analysis of Illumination, Naturalness and Structure. IEEE Access 6, 806–817.

Sharafeldin, N., Kawaguchi, A., Sundaram, A., Campbell, S., Rudnisky, C., Weis, E., Tennant, M. T., Damji, K. F., nov 2018. Review of economic evaluations of teleophthalmology as a screening strategy for chronic eye disease in adults. The British journal of ophthalmology 102 (11), 1485–1491.

Sinthanayothin, C., Boyce, J. F., Cook, H. L., Williamson, T. H., 1999. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. British Journal of Ophthalmology 83 (8), 902–910.

Soares, J. V., Leandro, J. J., Cesar, R. M., Jelinek, H. F., Cree, M. J., 2006. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. IEEE Transactions on Medical Imaging 25 (9), 1214–1222.

Sopharak, A., Uyyanonvara, B., Barman, S., mar 2009. Automatic exudate detection for diabetic retinopathy screening. ScienceAsia 35 (1), 80–88.

Srivastava, R., Duan, L., Wong, D. W., Liu, J., Wong, T. Y., jan 2017. Detecting retinal microaneurysms and hemorrhages with robustness to the presence of blood vessels. Computer Methods and Programs in Biomedicine 138, 83–91.

Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., Van Ginneken, B., apr 2004. Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging 23 (4), 501–509.

Stolte, S., Fang, R., aug 2020. A survey on medical image analysis in diabetic retinopathy. Medical Image Analysis 64, 101742.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol. 07-12-June. pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the Inception Architecture for Computer Vision. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Vol. 2016-Decem. pp. 2818–2826.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A., 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning. In: 31st AAAI Conference on Artificial Intelligence, AAAI 2017. pp. 4278–4284.

Tennakoon, R., Mahapatra, D., Roy, P., Sedai, S., Garnavi, R., oct 2016. Image Quality Classification for DR Screening Using Convolutional Neural Networks. In: Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop. University of Iowa, Iowa City, IA, pp. 113–120.

Teo, Z. L., Tham, Y. C., Yu, M., Chee, M. L., Rim, T. H., Cheung, N., Bikbov, M. M., Wang, Y. X., Tang, Y., Lu, Y., Wong, I. Y., Ting, D. S. W., Tan, G. S. W., Jonas, J. B., Sabanayagam, C., Wong, T. Y., Cheng, C. Y., 2021. Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045: Systematic Review and Meta-analysis. Ophthalmology 128 (11), 1580–1591.

Theera-Umpon, N., Poonkasem, I., Auephanwiriyakul, S., Patikulsila, D., 2019. Hard exudate detection in retinal fundus images using supervised learning. Neural Computing and Applications.

Usher, D., Dumskyj, M., Himaga, M., Williamson, T. H., Nussey, S., Boyce, J., jan 2004. Automated detection of diabetic retinopathy in digital retinal images: A tool for diabetic retinopathy screening. Diabetic Medicine 21 (1), 84–90.

Usher, D. B., Himaga, M., Dumskyj, M. J., 2003. Automated assessment of digital fundus image quality using detected vessel area. Proceedings of Medical Image Understanding and Analysis, 81–84.

Van Der Maaten, L., Hinton, G., 2008. Visualizing Data using t-SNE. Journal of Machine Learning Research 9, 2579–2605.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I., jun 2017. Attention is all you need. In: Advances in Neural Information Processing Systems. Vol. 2017-Decem. Neural information processing systems foundation, pp. 5999–6009.

Veiga, D., Pereira, C., Ferreira, M., Gonçalves, L., Monteiro, J., apr 2014. Quality evaluation of digital fundus images through combined measures. Journal of Medical Imaging 1 (1), 014001.

Walter, T., Klein, J.-C., Massin, P., Erginay, A., 2002. A contribution of image processing to the diagnosis of diabetic retinopathy–detection of exudates in color fundus images of the human retina. IEEE transactions on medical imaging 21 (10), 1236–1243.

Wang, S., Jin, K., Lu, H., Cheng, C., Ye, J., Qian, D., 2016. Human Visual System-Based Fundus Image Quality Assessment of Portable Fundus Camera Photographs. IEEE Transactions on Medical Imaging 35 (4), 1046–1055.

Wang, W., Yang, Y., Wang, X., Wang, W., Li, J., apr 2019. Development of convolutional neural network and its application in image classification: a survey. https://doi.org/10.1117/1.OE.58.4.040901 58 (4), 040901.

Wang, Z., Yin, Y., Shi, J., Fang, W., Li, H., Wang, X., 2017. Zoom-in-net: Deep mining lesions for diabetic retinopathy detection. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Vol. 10435 LNCS. pp. 267–275.

Welfer, D., Scharcanski, J., Marinho, D. R., dec 2011. Fovea center detection based on the retina anatomy and mathematical morphology. Computer Methods and Programs in Biomedicine 104 (3), 397–409.

Welikala, R. A., Fraz, M. M., Foster, P. J., Whincup, P. H., Rudnicka, A. R., Owen, C. G., Strachan, D. P., Barman, S. A., Aslam, T., Barman, S., Bishop, P., Blows, P., Bunce, C., Carare, R., Chakravarthy, U., Chan, M., Chianca, A., Cipriani, V., Crabb, D., Cumberland, P., Day, A., Desai, P., Dhillon, B., Dick, A., Foster, P., Gallacher, J., Garway-Heath, D., rini Goverdhan, Guggenheim, J., Gupta, P., Hammond, C., Hogg, R., Hughes, A., Keane, P., Khaw, S. P. T., Khawaja, A., Lascaratos, G., Lotery, A., Luthert, P., Mac-Gillivray, T., Mackie, S., Martin, K., McGaughey, M., McGuinness, B., McKay, G., McKibbin, M., Mitry, D., Moore, T., Morgan, J., Muthy, Z., O'Sullivan, E., Owen, C., Patel, P., Peto, T., Rahi, J., Rudnicka, A., Grossi Sampedro, C., Steel, D., Stratton, I., Strouthidis, N., Sudlow, C., Thaung, C., Thomas, D., Trucco, E., Tufail, A., Vernon, S., Viswanathan, A., Williams, C., Williams, K., Yates, J., Yates, M., Yip, J., Zhu, H., 2016. Automated retinal image quality assessment on the UK Biobank dataset for epidemiological studies. Computers in Biology and Medicine 71, 67–76.

Wilkinson, C. P., Ferris, F. L., Klein, R. E., Lee, P. P., Agardh, C. D., Davis, M., Dills, D., Kampik, A., Pararajasegaram, R., Verdaguer, J. T., Lum, F., sep 2003. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. Ophthalmology 110 (9), 1677–1682.

Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., nov 2016. Data Mining: Practical Machine Learning Tools and Techniques. Data Mining: Practical Machine Learning Tools and Techniques, 1–621.

Wu, B., Zhu, W., Shi, F., Zhu, S., Chen, X., 2017. Automatic detection of microaneurysms in retinal fundus images. Computerized Medical Imaging and Graphics 55, 106–112.

Xiao, D., Bhuiyan, A., Frost, S., Vignarajan, J., Tay-Kearney, M. L., Kanagasingam, Y., apr 2019. Major automatic diabetic retinopathy screening systems and related core algorithms: a review. Machine Vision and Applications 30 (3), 423–446.

Xie, X., Xie, G., Xu, X., Cui, L., Ren, J., 2019. Automatic image segmentation with superpixels and image-level labels. IEEE Access 7, 10999–11009.

Yu, F., Sun, J., Li, A., Cheng, J., Wan, C., Liu, J., jul 2017. Image quality classification for DR screening using deep learning. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 664–667.

Yu, H., Barriga, E. S., Agurto, C., Echegaray, S., Pattichis, M. S., Bauman, W., Soliz, P., 2012. Fast localization and segmentation of optic disk in retinal images using directional matched filtering and level sets. IEEE Transactions on Information Technology in Biomedicine 16 (4), 644–657.

Yu, L., Liu, H., 2004. Efficient Feature Selection via Analysis of Relevance and Redundancy. Journal of Machine Learning Research 5, 1205–1224.

Yu, T., Ma, Y., Li, W., may 2015. Automatic localization and segmentation of optic disc in fundus image using morphology and level set. In: International Symposium on Medical Information and Communication Technology, ISMICT. Vol. 2015-May. IEEE Computer Society, pp. 195–199.

Zago, G. T., Andreão, R. V., Dorizzi, B., Teatini Salles, E. O., 2018. Retinal image quality assessment using deep learning. Computers in Biology and Medicine 103, 64–70.

Zhang, L., Lim, C. P., jul 2020. Intelligent optic disc segmentation using improved particle swarm optimization and evolving ensemble models. Applied Soft Computing 92, 106328.

Zhang, X., Chutatape, O., 2005. A SVM approach for detection of hemorrhages in background diabetic retinopathy. In: 2005 IEEE International Joint Conference on Neural Networks. pp. 2435–2440.

Zhang, X., Thibault, G., Decencière, E., Marcotegui, B., Laÿ, B., Danno, R., Cazuguel, G., Quellec, G., Lamard, M., Massin, P., Chabouis, A., Victor, Z., Erginay, A., 2014. Exudate detection in color retinal images for mass screening of diabetic retinopathy. Medical Image Analysis 18 (7), 1026–1043.

Zhou, M., Jin, K., Wang, S., Ye, J., Qian, D., mar 2018. Color Retinal Image Enhancement Based on Luminosity and Contrast Adjustment. IEEE Transactions on Biomedical Engineering 65 (3), 521–527.

Zhou, W., Wu, C., Chen, D., Wang, Z., Yi, Y., Du, W., 2017a. Automated Detection of Red Lesions Using Superpixel Multichannel Multifeature. Computational and Mathematical Methods in Medicine 2017 (4), 1–13.

Zhou, W., Wu, C., Yi, Y., Du, W., 2017b. Automatic Detection of Exudates in Digital Color Fundus Images Using Superpixel Multi-Feature Classification. IEEE Access 5, 17077 – 17088.