UNIVERSIDAD
NACIONAL
DE COLOMBIA

Deep learning analysis of eye fundus images to support medical diagnosis

Oscar Julián Perdomo Charry

Universidad Nacional de Colombia
Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2020

Deep learning analysis of eye fundus images to support medical diagnosis

Oscar Julián Perdomo Charry

Thesis submitted as requirement to obtain the title of:
PhD. in Systems and Computer Engineering

Advisor:
Fabio A. González, Ph.D.

Universidad Nacional de Colombia
Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2020

Deep learning analysis of eye fundus images to support medical diagnosis

Oscar Perdomo

Dedication

To my muse and beautiful wife Ana
To my parents Álvaro and Isabel
To my brothers Kike and Cesar

Abstract

Machine learning techniques have been successfully applied to support medical decision making of cancer, heart diseases and degenerative diseases of the brain. In particular, deep learning methods have been used for early detection of abnormalities in the eye that could improve the diagnosis of different ocular diseases, especially in developing countries, where there are major limitations to access to specialized medical treatment. However, the early detection of clinical signs such as blood vessel, optic disc alterations, exudates, hemorrhages, drusen, and microaneurysms presents three main challenges: the ocular images can be affected by noise artifact, the features of the clinical signs depend specifically on the acquisition source, and the combination of local signs and grading disease label is not an easy task.

This research approaches the problem of combining local signs and global labels of different acquisition sources of medical information as a valuable tool to support medical decision making in ocular diseases. Different models for different eye diseases were developed. Four models were developed using eye fundus images: for DME, it was designed a two-stages model that uses a shallow model to predict an exudate binary mask. Then, the binary mask is stacked with the raw fundus image into a 4-channel array as an input of a deep convolutional neural network for diabetic macular edema diagnosis; for glaucoma, it was developed three deep learning models. First, it was defined a deep learning model based on three-stages that contains an initial stage for automatically segment two binary masks containing optic disc and physiological cup segmentation, followed by an automatic morphometric features extraction stage from previous segmentations, and a final classification stage that supports the glaucoma diagnosis with intermediate medical information. Two late-data-fusion methods that fused morphometric features from cartesian and polar segmentation of the optic disc and physiological cup with features extracted from raw eye fundus images. On the other hand, two models were defined using optical coherence tomography. First, a customized convolutional neural network termed as OCT-NET to extract features from OCT volumes to classify DME, DR-DME and AMD conditions. In addition, this model generates images with highlighted local information about the clinical signs, and it estimates the number of slides inside a volume with local abnormalities. Finally, a 3D-Deep learning model that uses OCT volumes as an input to estimate the retinal thickness map useful to grade AMD.

The methods were systematically evaluated using ten free public datasets. The methods were compared and validated against other state-of-the-art algorithms and the results were also qualitatively evaluated by ophthalmology experts from Fundación Oftalmológica Nacional. In addition, the proposed methods were tested as a diagnosis support tool of diabetic macular edema, glaucoma, diabetic retinopathy and age-related macular degeneration using two different ocular imaging representations. Thus, we consider that this research could be potentially a big step in building telemedicine tools that could support medical personnel for detecting ocular diseases using eye fundus images and optical coherence tomography.

Esta tesis de doctorado se sustentó el 4 de Junio de 2020 a las 9AM,
y fue evaluada por los siguientes jurados:

Edgar Eduardo Romero Castro Ph.D.
Juan Carlos Caicedo Rueda Ph.D.
Pablo Andrés Arbeláez Escalante Ph.D.
Flavio Augusto Prieto Ortiz Ph.D.

Contents

List of Figures

List of Tables

1 Introduction

The main aim of this thesis is to develop automatic analysis of ocular images based on deep learning in order to support medical diagnosis in eye conditions related to diabetes. This kind of images are an invaluable resource for the diagnosis of different eye diseases. However, the manual analysis of these images by an ophthalmologist requires extensive training and, even for the experienced ones, it is a cumbersome and prone to error process. Diabetes Mellitus is considered by the World Health Organization (WHO)[1] and by the International Diabetes Federation (IDF) [2] as a global epidemic. According to the IDF diabetes atlas 2017, the worldwide associated costs to diabetes were $727 billion (USD), 415 million people worldwide suffer diabetes, but unfortunately, only half of these people were correctly diagnosed [1]. In Colombia, the data is not encouraging, it was estimated that around $3,8$ million (8% of the Colombian population) people between 20 and 79 years present diabetes, a disease to which are $18,640$ deaths attributed and 1 million people still have not been well diagnosed [1, 2]. Some patients in early stages of diabetes, have previous symptoms such as blurred vision, slow loss of vision, floaters, shadows or areas of diminished vision and difficulty for seeing at night [3, 4]. Unfortunately, most patients with diabetes do not show symptoms, pain, or loss of vision, but as the disease progresses serious conditions such as: diabetic retinopathy, glaucoma, macular edema and retinal detachment occurs leading to a partial or total loss of vision [5–7].

The research of new techniques based on deep learning to extract local information related with clinical signs as: microaneurysm, exudates, hemorrhages and drusen, to be combined with global grading disease label, is the start point that justifies to focus on this doctoral research. The focus thesis is the analysis of ocular images used to diagnose eye conditions associated with diabetes as: diabetic retinopathy (DR), diabetic macular edema (DME), glaucoma and age-related macular degeneration (AMD). These diseases are the leading cause of blindness worldwide in productive age (20-69 years), with the main problem that 25% of diabetics worldwide will have visual problems along diabetes, and without a preventive diagnosis and promptly treatment, these subjects will suffer an irreversible blindness [1–5].

1.1 Problem statement

Eye conditions related to diabetes at productive age is a worldwide health problem with a large amount of budget allocated by the World Health Organization [8]. In addition to this, the large amount of available information (images, signals and clinical data) to study eye condition, and the need of design new deep learning strategies as a support tool of different ocular diseases, make this proposal as a doctorate challenge from the point of engineering with social responsibility to

[1] https://www.who.int/diabetes/goal/en/
[2] https://www.idf.org/aboutdiabetes/complications.html

solve global issues in vulnerable populations. Hence, our hypothesis is that deep learning techniques applied to ocular images may improve the diagnosis of eye conditions related to diabetes. Moreover, the integration of local features such as blood vessel patterns, exudates, microaneurysm, drusen, hemorrhages, abnormalities of the macula and others could improve the classification of different eye conditions related to diabetes, but that has not been explored deeply yet.

The focus of this research is to develop deep learning models for analysis of ocular images, that combine the detection of local labels (features) and global labels (grade of diabetes-related eye conditions) to support medical personnel in a faster and more precise diagnosis.

The main research questions of this work are:

- ¿How to apply deep learning techniques to ocular images to improve the diagnosis of eye conditions related to diabetes?
- ¿Could the integration of domain knowledge improve the classification of different eye conditions related to diabetes?

Specifically, the guiding goals of this research are:

Main goal To develop interpretable deep learning methods for automatic analysis of ocular images to support the diagnosis of different diabetes-related conditions.

Specific goals

- To design and implement a deep learning method for automatically identifying different eye features (e.g. exudates, microaneurysm, blood vessel and optic disc segmentation) useful for diagnosis of diabetes-related conditions.
- To design and implement a deep learning model to automatically classify the grade of different diabetes-related conditions.
- To systematically evaluate the performance of the method on different dataset for diagnosis of diabetes-related conditions.

1.2 Main contributions

This research presents novel deep learning strategies to tackle the automatic analysis of ocular image using different sources of information. The following is the outline of the main contributions of this work.

- A CNN that combines a patches classifier used for detecting local abnormalities (exudates) stacked with the raw eye fundus image as a fourth-channel array for diabetic macular edema diagnosis.
- A three-stages strategy based on deep learning methods to support glaucoma diagnosis with relevant intermediate medical information such as: segmentations of part of the eyes and the morphometric features that describes these parts.

- A deep learning late fusion strategy that merges features extracted from images with polar or cartesian morphometric features to support glaucoma diagnosis.
- A method to classify three diabetes-related conditions using optical coherence tomography volumes based on a CNN (OCT-NET). The CNN outperformes other state-of-the-art models, generating clinically interpretable information to support the medical diagnosis.
- The design of a deep learning method to estimate the retinal thickness map from SD-OCT volumes.

This thesis contributed to the generation of new research products and the financial support of four master student, one fellowship/training ophthalmologist and one doctorate student. As result, it was listed the grants [G] and awards [A] based on this thesis:

G1 Google Research Awards for Latin America (Doctorate Student Award) by Google. *Computational Learning Model for the Eye Fundus Analysis to Support Medical Diagnosis.* Budget: USD$23.400 for one year, in collaboration with: Fundación Oftalmológica Nacional and Universidad del Rosario.[3]

G2 Projects of Science, Technology and Innovation in Health 2018 by COLCIENCIAS. *Detección temprana de daño ocular en diabéticos usando un sistema de inteligencia artificial en imágenes de fondo de ojo.* Budget: COP$448′094.113 for two years, in collaboration with: Fundación Oftalmológica Nacional and Haute école spécialisée de Suisse occidentale.

G3 National call for the support to projects of reserach and artistic creation by Universidad Nacional de Colombia. *Clasificación de retinopatía diabética y edema macular diabético en imágenes de fondo de ojo mediante redes neuronales convolucionales.* Budget: COP$40′000.000 for 14 months, in collaboration with Fundación Oftalmológica Nacional.

G4 Global Ophthalmology Awards Program (Fellowship Award 2018) by BAYER . *Neural networks for detection of diabetic retinopathy and diabetic macular edema.* Budget: US$50.000 for one year, in collaboration with: Universidad del Rosario and Fundación Oftalmológica Nacional.[4]

A1 **Second place**, XVI Premio Internacional de Investigación en Ciencias de la Salud, Sanitas Internacional, November 2016.[5]

A2 **Second place**, Premio a la investigación ACOREV-ALLERGAN 2017-2018, February 2018.

A3 **Second place**, IX Congreso Anual de Oftalmología, Fundación Oftalmológica Nacional - Sociedad de Cirugía Ocular, February 2018.

[3] https://ai.google/research/outreach/latin-america-research-awards/
[4] https://www.bayer-ophthalmology-awards.com/alumni
[5] http://www.unisanitas.edu.co/decimo-septima-edicion

The following is a list of papers that have been published during the development of this research:

JOURNAL PAPERS

1. **Perdomo, Oscar** and Carpio, Vanessa and Rosenstiehl, Shirley M. and González, Fabio A., "Deep data fusion of polar images and polar morphometric features for glaucoma detection", To be submitted to Plos One, (2020).

2. Pinilla, Carlos and **Perdomo, Oscar** and Rios, Hernán A. and Rosenstiehl, Shirley M. and Gómez, Flor E. and González, Fabio A. and Rodríguez, Francisco J., "Assessment of OCT-NET based deep learning model for the diagnosis of diabetic macular edema using 5-scans from optical coherence tomography volumes", To be submitted to Investigative Ophthalmology & Visual Science, (2020).

3. **Perdomo, Oscar** and González, Fabio A., "A systematic review of deep learning methods applied to ocular images", Ciencia e Ingeniería Neogranadina 30, (2020), pp. 1–21 [9].

4. **Perdomo, Oscar** and Rios, Hernán and Rodríguez, Francisco and Otálora, Sebastián and Meriaudeau, Fabrice and Müller, Henning and González, Fabio A., "Classification of diabetes-related retinal diseases using a deep learning approach in optical coherence tomography", Computer Methods and Programs in Biomedicine 178, (2019), pp. 181–189 [10].

CONFERENCE PAPERS

5. Pérez, Andrés and **Perdomo, Oscar** and González, Fabio A., "A lightweight deep learning model for mobile eye fundus image quality assessment", 15th International Symposium on Medical Information Processing and Analysis 10975, (2019), pp. 109750I-1 [11].

6. **Perdomo, Oscar** and Rios, Hernán A. and Rodríguez, Francisco J and González, Fabio A., "3D deep convolutional neural network for predicting neurosensory retinal thickness map from spectral domain optical coherence tomography volumes", 14th International Symposium on Medical Information Processing and Analysis 10975, (2018), pp. 109750I-1 [12].

7. **Perdomo, Oscar** and Andrearczyk, Vincent and Meriaudeau, Fabrice and Müller, Henning and González, Fabio A., "Glaucoma diagnosis from eye fundus images based on deep morphometric feature estimation", Computational pathology and ophthalmic medical image analysis, (2018), pp. 319–327 [13].

8. **Perdomo, Oscar** and Otálora, Sebastian and González, Fabio A. and Meriaudeau, Fabrice and Müller, Henning, "OCT-NET: A convolutional network for automatic classification of normal and diabetic macular edema using SD-OCT volumes", 2018 IEEE 15th International Symposium on Biomedical Imaging, (2018), pp. 1423–1426 [14].

9. **Perdomo, Oscar** and Arevalo, John and González, Fabio A., "Combining morphometric features and convolutional networks fusion for glaucoma diagnosis", 13th International Conference on Medical Information Processing and Analysis 10572, (2017), pp. 105721G [15].

10. **Perdomo, Oscar** and Arevalo, John and González, Fabio A., "Convolutional network to detect exudates in eye fundus images of diabetic subjects", 12th International Symposium on Medical Information Processing and Analysis 10160, (2017), pp. 101600T [16].

11. **Perdomo, Oscar** and Otalora, Sebastian and Rodríguez, Francisco and Arevalo, John and González, Fabio A., "A novel machine learning model based on exudate localization to detect diabetic macular edema", Computational pathology and ophthalmic medical image analysis, (2016), pp. 167–144 [17].

12. Kamble, Ravi M and Chan, Genevieve CY and **Perdomo, Oscar** and Kokare, Manesh and González, Fabio A and Müller, Henning and Mériaudeau, Fabrice, "Automated diabetic macular edema (DME) analysis using fine tuning with Inception-Resnet-v2 on OCT images", 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), (2018), pp. 442–446 [18].

13. Otálora, Sebastian and **Perdomo, Oscar** and González, Fabio and Müller, Henning, "Training deep convolutional neural networks with active learning for exudate classification in eye fundus images", Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, (2017), pp. 146–154 [19].

SHORT MEDICAL PAPERS

14. Rios, Hernán Andrés and **Perdomo, Oscar** and Carpio, Vanessa and Carvajal, Claudia Rosa and González, Fabio A. and Müller, Henning and Rodriguez, Francisco J., "Automatic prediction of capillarity patterns on optical coherence tomography angiography images", Investigative Ophthalmology & Visual Science, (2019) [20].

15. Rios, Hernán A. and **Perdomo, Oscar** and Rosenstiehl, Shirley M. and Gómez, Flor E. and González, Fabio A. and Rodríguez, Francisco J., "Deep learning method to identify diabetic retinopathy and diabetic macular edema characteristics", Investigative Ophthalmology & Visual Science, (2019) [21].

16. Belalcazar, Sandra and Carpio, Vanessa and **Perdomo, Oscar** and Rios, Hernán A. and Carvajal, Claudia and Rosenstiehl, Shirley M. and González, Fabio A. and Rodríguez, Francisco J., "Concordance between color photo interpretation of the optic nerve and an unsupervised learning algorithm to determine optic nerve damage", Investigative Ophthalmology & Visual Science (2019) [22].

17. Pinilla, Carlos M. and **Perdomo, Oscar** and Rios, Hernán A. and Carpio, Vanessa and Gómez, Flor E. and Rosenstiehl, Shirley M. and González, Fabio A. and Rodríguez, Francisco J., "Exactitud de una red neuronal artificial para el diagnóstico del edema macular diabético con imágenes de tomografía de coherencia óptica", Revista Científica de la Sociedad Colombia de Oftalmología, (2018), pp. 8–9 [23].

18. Belalcazar, Sandra and Rios, Hernán Andrés and Carpio, Vanessa and **Perdomo, Oscar** and Carvajal, Claudia Rosa and Gonzalez, Fabio A and Müller, Henning, "Convolutional neural

networks for identification and classification of optic nerve damage features", Investigative Ophthalmology & Visual Science, (2018), pp. 1719–1719 [24].

19. Rodriguez, Francisco J Jose and **Perdomo, Oscar** and Gomez, Flor Edith and Rosenstiehl, Shirley Margarita and Rios, Hernán Andrés and Gonzalez, Fabio A., "Neural network for detection of diabetic macular edema in fundus color images", Investigative Ophthalmology & Visual Science, (2017), pp. 688–688 [25].

Other non-related work:

20. Contreras, Victor H and Lara, Juan S and **Perdomo, Oscar** and González, Fabio A., "Supervised online matrix factorization for histopathological multimodal retrieval", 14th International Symposium on Medical Information Processing and Analysis 10975, (2018), pp. 109750Y [26].

21. Otálora, Sebastian and **Perdomo, Oscar** and Atzori, Manfredo and Andersson, Mats and Jacobsson, Ludwig and Hedlund, Martin and Müller, Henning, "Determining the scale of image patches using a deep learning approach", 2018 IEEE 15th International Symposium on Biomedical Imaging, (2018), pp. 843–846 [27].

1.3 Outline

The remainder of this document is organized as follows: Chapter 2 gives an overall background of automatic analysis of eye fundus images to support medical diagnosis. In Chapter 3 is defined a deep learning strategy applied to eye fundus images, to combine local information of clinical signs with global grading label to diagnose of DME. In Chapter 4, a deep learning model that uses a late fusion to merge features extracted from eye fundus images with polar or cartesian morphometric features for glaucoma diagnosis. In Chapter 5, a three-stages model to classify glaucoma disease is explained in detail. This model delivers eye's part segmentations, morphometric features about the segmentations, and the classification of control, suspicious and glaucoma conditions. In Chapter 6, OCT-NET, an end-to-end deep learning method is applied to SD-OCT to the diagnosis of three ocular diseases. This customized model learns to extract features about a volume to outperform previous state-of-the-art results in two different databases. In Chapter 7, a 3D deep learning model to estimate retinal thickness map is proposed. This proposed model delivers a 2D-retinal thickness map from a SD-OCT volume. Finally, Chapter 8 summarizes the main results of this research, discuses some relevant topics, and provides ideas for future works in deep learning analysis applied to medical images.

2 Background and related work

This chapter presents an overview of the state-of-the-art deep learning methods used in ophthalmic images, databases and potential challenges for ocular diagnosis. In ophthalmology, deep learning methods have primarily been applied to eye fundus images and optical coherence tomography. On the one hand, these methods have achieved an outstanding performance in the detection of ocular diseases such as: diabetic retinopathy, glaucoma, diabetic macular degeneration and age-related macular degeneration. On the other hand, several worldwide challenges have shared big eye imaging data sets with segmentation of part of the eyes, clinical signs and the ocular diagnostic performed by experts. In addition, these methods are breaking the stigma of black-box models, with the delivering of interpretable clinically information. This reviewed was published in the journal of *Ciencia E Ingeniería Neogranadina* [9].

2.1 Introduction

The diagnosis of ophthalmologic diseases is done with different kinds of clinical exams. Exams may be non-invasive such as: slit-lamp exam, visual acuity, eye fundus image (EFI), ultrasound, optical coherence tomography (OCT); or invasive exams as fluorescein angiography [6]. The non-invasive clinical exams are easier to take, have no contraindications and do not affect the eye's natural response to external factors in comparison to the invasive exams. Therefore, EFI and OCT exams are high patient compliance, quick and simple techniques, with the main advantages that images can be easily saved to be analyzed at a later time, and the prognosis, diagnosis and follow-up of diseases can be monitored over time.

The automatic analysis of EFIs and OCTs as a tool to support medical diagnosis has been an engineering challenge in terms of achieving the best performance, the lowest computational cost and lowest run time among the different algorithms [28–32]. Thus, the choice of the best method to represent, analyze and make a diagnosis using ocular images is a complex computational problem [33–37]. On the other hand, deep learning techniques have been applied with some success to several eye conditions using as evidence individual sources of information [38–40]. Some researchers have studied how to support the diagnosis with different methodologies. Vandarkuhali and Ravichandran [38] detected the retinal blood vessels with an extreme learning machine approach and probabilistic neural networks, Gurudath et al. [28] worked with machine learning identification from fundus images with a three layered artificial neural network and a support vector machine to classify retinal images, and Priyadarshini et al. studied clustering and classifications with data mining to give some useful prediction applied to diabetic retinopathy diagnosis [29]. Despite good results, the main problem with these works is: the data sets are small and the need for labels is expensive and cumbersome work. Deep learning (DL) offers some advantages such as the capacity of processing lots of data with the use of graphic processing units (GPUs) and tensor processing units

(TPUs); and the ability to automatically learn the data representation from raw data. Thanks to these features DL has been able to outperform the traditional methods in several computer vision and image analysis tasks. This success has motivated its application to medical image analysis including, of course, ophthalmology images.

This chapter focuses on the review and analysis of deep learning methods applied to ocular images for the diagnosis of: diabetic retinopathy (DR), glaucoma, diabetic macular edema (DME) and age-related macular degeneration (AMD). These diseases are related with diabetes as one of the four major types of chronic noncommunicable disease and they are the leading cause of blindness worldwide in productive age (20-69 years), with the main problem that 25% of diabetics worldwide will have visual problems along diabetes, and without a preventive diagnosis and treatment promptly, these subjects will suffer irreversible blindness [1–4, 41]. The research approach of the chapter is summarized on four main areas: an overview of the medical background about ocular diseases and medical information sources, the free public available ocular data sets, the most common performance metrics used by deep learning methods and an overview of the main deep learning methods for each source of medical information.

2.2 Medical background

2.2.1 Ocular diseases

Diabetic retinopathy

The diabetic retinopathy is caused by a side effect of diabetes which reduced blood supply to the retina, including include lesions appearing on the retinal surface [42]. DR-related lesions can be categorized into red lesions such as microaneurysms and hemorrhages and bright lesions such as exudates and cotton-wool spots [43], as shown in Figure 2.1.

Diabetic macular edema

The diabetes macular edema is a complication of DR that occurs when the vessels of the central part of the retina (macula) are affected by accumulation of fluid and exudate formation in different parts of the eye [45], as depicted in Figure 2.2.

Glaucoma

The glaucoma is related to the progressive degeneration of optic nerve fibers and structural changes of the optic nerve head [42]. Although glaucoma cannot be cured, its progression can be slowed down by treatment. Therefore, timely diagnosis of this disease is vital to avoid blindness [48, 49]. Glaucoma diagnosis detection is based on manual assessment of the Optic Disc (OD) through ophthalmoscopy, looking morphological parameters for the central bright zone called the optic cup and a peripheral region called the neuroretinal rim [50], as reported in Figure 2.3.

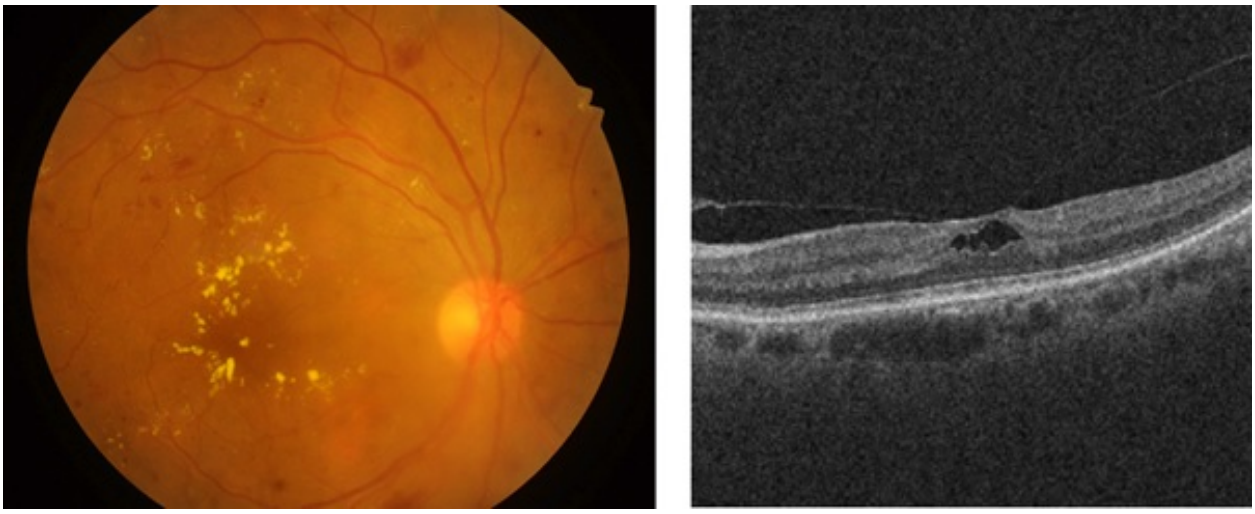

Figure 2.1: [Left] A color eye fundus image showing multiple microaneurysms, intraretinal hemorrhages, and exudation affecting the fovea in a patient with severe non-proliferative diabetic retinopathy with severe diabetic macular edema [44], and [Right] A b-scan OCT showing a vitreo-macular traction affecting the foveal depression [18]

.

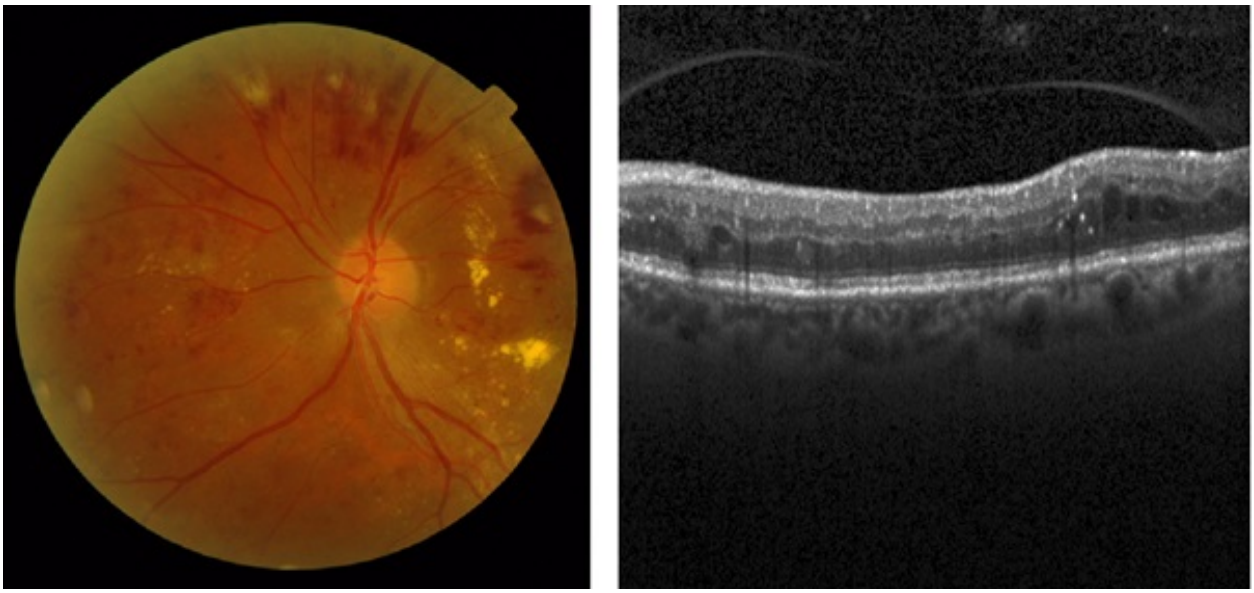

Figure 2.2: [Left] A color eye fundus image showing multiple dot and flame hemorrhages, cotton wool spots and macular exudation in a patient with severe nonproliferative diabetic retinopathy with diabetic macular edema [46], and [Right] A b-scan OCT showing multiple intraretinal hyper reflective dots and pseudo-cystic spaces in the middle retinal layers in a patient with diabetic macular edema [47]

.

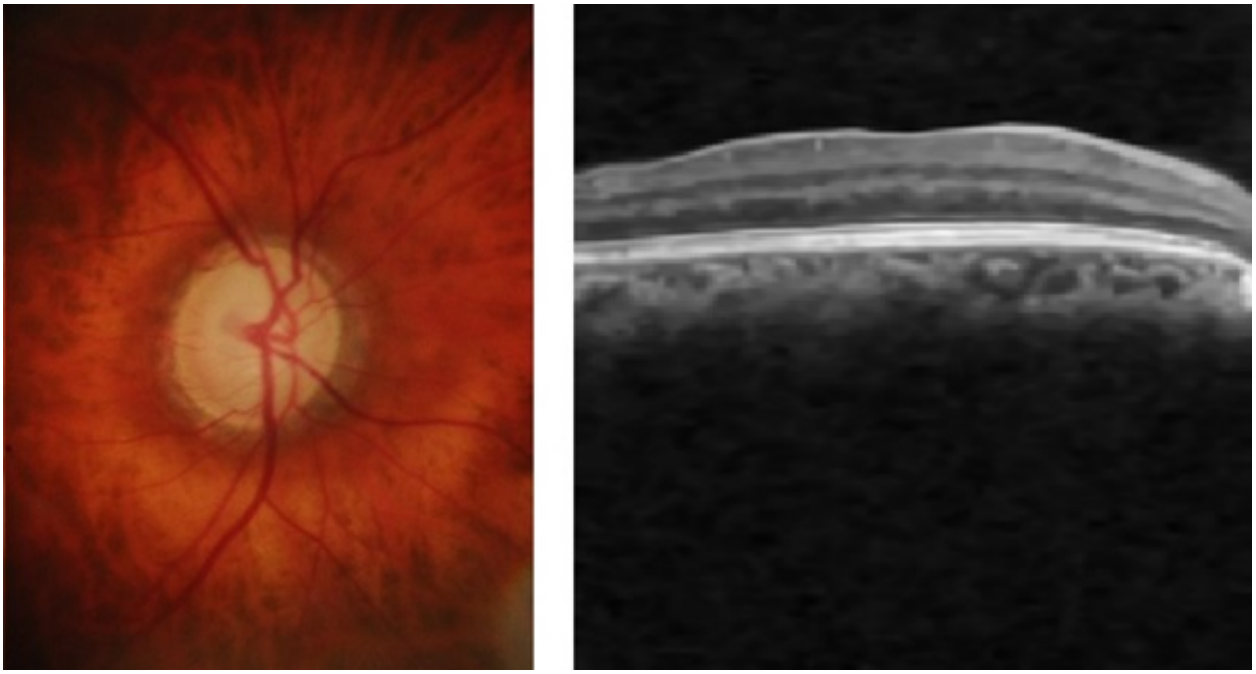

Figure 2.3: [Left] An optic disc color image showing an absence of the neural ring with a total excavation in a patient with advanced glaucoma [51], and [Right] A b-scan OCT showing a thinning in the nerve fiber layer in a patient with Glaucoma [52]

.

Age-related macular degeneration

The age-related macular degeneration (AMD) causes vision loss at the central region and distortion at the peripheral region [42]. The main symptom and clinical indicator of dry AMD is drusen. The major symptom of wet AMD is the presence of exudates [53], as presented in Figure 2.4.

2.2.2 Medical information sources

There are different types of clinical exams for the diagnosis of ocular disease. Some researchers documented techniques of digital signal and image processing of the eye, such as: electrooculogram (EOC) [7], electroretinogram (ERG) [56, 57], visual evoked potentials [58–61], dynamic pupillometry [62, 63], among other methods [64]. The two non-invasive techniques widely used by ophthalmologist to diagnose ocular condition are EFIs and OCT. On the one hand, the eye fundus is represented as a 2D image of the eye that allows to check faster and easily parts of the eyes (i.e. optic disc, blood vessels, and others), but also some retinal abnormalities (i.e. micro aneurysms, exudates, among others). On the other hand, the OCT uses near-infrared light based on low coherence interferometry principles to record the set of retinal layers. The OCT depicts the information in a 3D volume with a resolution of a cross-sectional area with a defined number of scans as shown in Figure 2.5. In the two cases, the diagnosis performed by experts depends crucially on the clinical findings located during the exam.

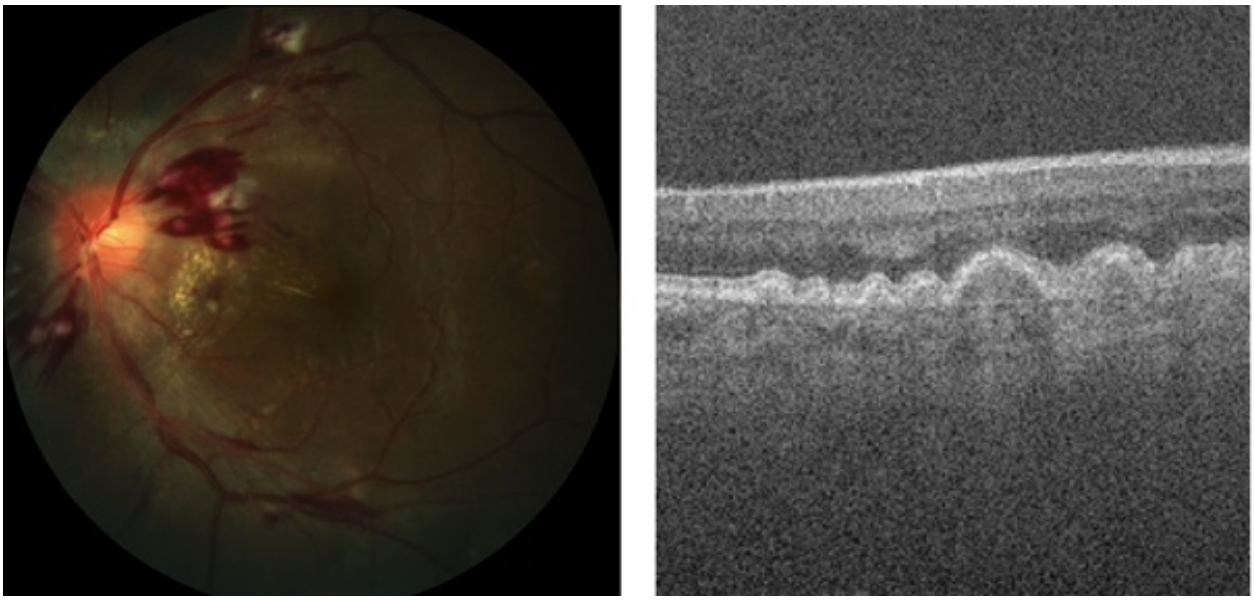

Figure 2.4: [Left] A color eye fundus image showing multiple flame hemorrhages, cotton wool spots and macular exudation [54], and [Right] A b-scan OCT showing the presence of soft drusen in the EPR-coriocapilar complex in a patient with Age-related Macular Degeneration [55]

.

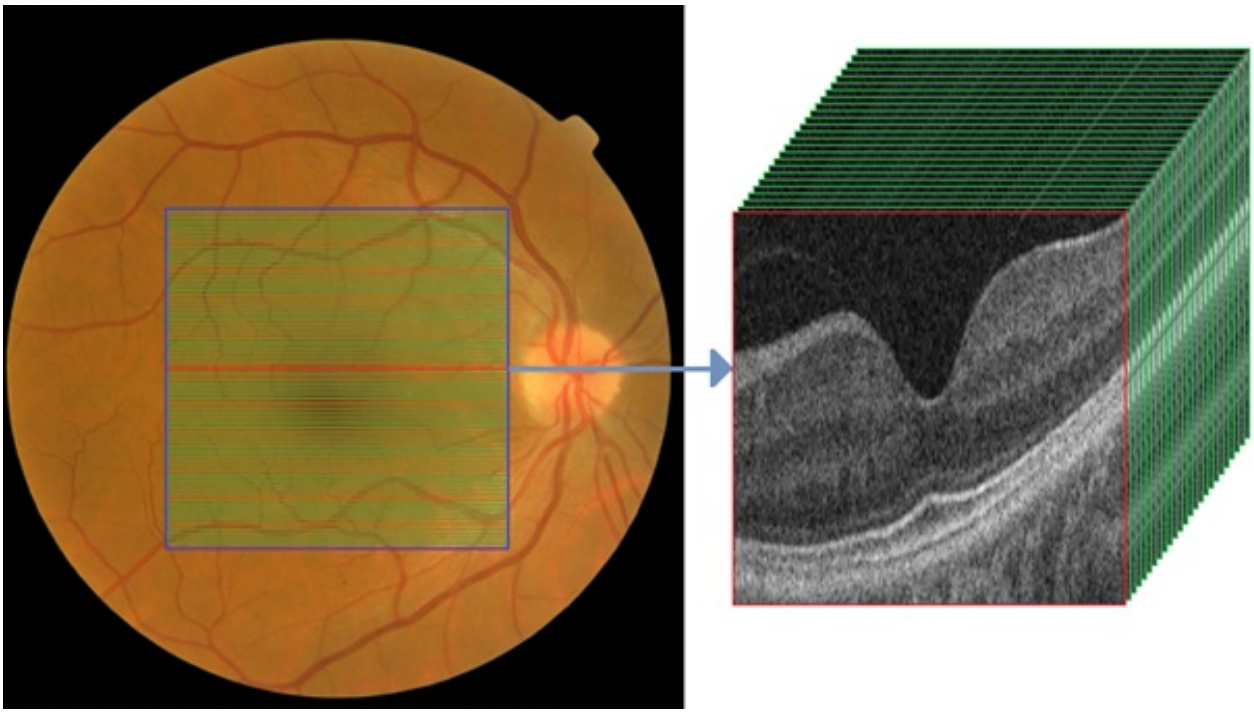

Figure 2.5: EFI and OCT volume containing cross-sectional b-scans from a healthy subject [65]

.

2.3 Ocular image data sets

In recent years, the detection of clinical signs and the grading of ocular diseases have been considering as engineering challenging tasks. In addition, worldwide researchers have published their methods and a set of EFIs and OCTs databases with different ocular conditions, population, acquisition devices and image resolution. The available ocular data sets for each ocular disease, the type of ocular image and the study population are presented in Table **2-1**.

2.4 Methods performance

Deep learning approaches have shown astonishing results in problem domains like recognition system, natural language processing, medical sciences, and in many other fields. Google, Facebook, Twitter, Instagram, and other big companies use deep learning in order to provide better applications and services to their customers [94]. Deep learning approaches have active applications using Deep Convolutional Neural Networks (DCNN) in object recognition [95–98], speech recognition [99, 100], natural language processing [101], theoretical science [102], medical science [103, 104], etc. In medical field, some researchers apply deep learning to solve different medical problems like diabetic retinopathy [105], detection of cancer cells in human body [106], spine imaging [107] and many others [12, 19]. Although unsupervised learning is applicable in the field of medical science where sufficient labeled data sets for a particular type of disease are not available. In particular, the state-of-the-art methods in ocular images are based on supervised learning techniques.

2.4.1 Performance metrics in deep learning models

The performance comparison of deep learning methods in classification tasks is performed by the calculation of statistical metrics. These metrics assess the agreement and disagreement between the expert and the proposed method to grade an ocular disease [13–17, 55, 81, 93, 108–118]. The performance metrics used in state-of-the-art works are presented in equations 2-1 to 2-7 as follows:

$$\textit{Area under the curve (AUC)} = \frac{\sum Rank(+) - |+| * \frac{|+|+1}{2}}{|+|+|-|} \tag{2-1}$$

$$\textit{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2-2}$$

$$\textit{Sensitivity} = \frac{TP}{TP + FN} \tag{2-3}$$

$$\textit{Specificity} = \frac{TN}{TN + FP} \tag{2-4}$$

$$\textit{Precision} = \frac{TP}{TP + FP} \tag{2-5}$$

Ocular disease	Dataset	Dataset description
DR	DRIVE [66]	40 eye fundus images with resolution of images is 768 × 584 pixels. The dataset contains 7 images graded by experts as mild DR and 33 images as normal.
	DIARETDB0 [67]	130 eye fundus images with 110 DR and 20 normal images. The images labelled as DR contains the segmentation of clinical signs: hard exudates, soft exudates, microaneurysms, hemorrhages and neovascularization.
	DIARETDB1 [68]	89 eye fundus images where 84 images has mil DR and 5 images labelled as normal.
	ROC [69]	100 digital color fundus images with microaneurysms in all the images. This dataset was randomly split into training and test data sets with 50 images.
	CHASE-DB [70]	28 eye fundus images with blood vessel segmentations.
	E-OPHTHA [71]	Two subsets: a set of 47 eye fundus images with the segmentation of exudates and 35 images without lesions labelled as normal. The second set has 148 images with microaneurysms and 233 images labelled as normal.
	EYE PACS [72]	Two subsets: the training set has 35126 and the test set has 53576. The images were labelled as normal, mild, moderate, severe and proliferative DR.
	APTOS [73]	13000 images with normal, mild, moderate, severe and proliferative DR.
DR,GLAUCOMA	ONHSD [74]	49 eye fundus images with the optic head segmentation and the grading of DR and glaucoma.
	HRF [75]	45 eye fundus images with 15 healthy, 15 DR and 15 glaucomatous subjects. The images have the detection and segmentation of clinical signs provided by experts.
DR, DME	MESSIDOR [46]	1200 eye fundus images with DR and DME labels.
	iDRID [44]	516 images with resolution of 4288x2848 pixels with the grading of DME and DR performed by experts.
DR, AMD	STARE [76, 77]	400 eye fundus images and 400 black and white mask with blood vessel annotations.
	ARIA [78, 79]	143 color fundus images with resolution of 768 × 576 pixels. The images were grading as: 23 AMD, 59 DR and 61 normal.
	OCTID [80]	500 OCTs with normal, macula hole, AMD, central serous retinopathy and DR.
GLAUCOMA	DRIONS-DB [81]	110 color fundus images with optic nerve head segmentation. The images were labelled as: 26 glaucomatous and 84 with eye hypertension.
	ORIGA650 [82]	650 eye fundus images with the classification of glaucoma.

Ocular disease	Dataset	Dataset description
GLAUCOMA	INSPIRE-AVR [83]	40 color images with the blood vessels, optic disc and arterio-venous reference.
	RIM-ONE [51]	783 images with glaucomatous, suspicious of glaucoma and normal conditions.
	ACHIKO-K [84]	258 eye fundus images with 144 normal and 114 glaucomatous subjects.
	DRISHTI-GS [85, 86]	101 images with optic disc and optic cup segmentations and glaucoma condition.
	RIGA [87]	760 retinal fundus images with glaucoma labels.
	REFUGE [88]	1200 eye fundus images with optic disc and cup segmentations with normal and glaucoma conditions.
	POAG [52]	1110 scans where 263 were diagnosed as healthy and 847 with primary open angle glaucoma (POAG).
AMD	AREDS [89]	206500 eye fundus images with AMD and non-AMD conditions.
	iCHALLENGE [54]	1200 eye fundus images with early AMD and non-AMD conditions.
	A2A SD-OCT [55]	385 OCTs with 269 AMD and 115 normal subjects. Each OCT volume has 100 B-scan with resolution of 512×1000 pixels.
	HEIDELBERG [90]	15 OCT volumes with the retinal layer segmentation performed by expert. The database was labelled with AMD condition.
DME	HEI-MED [91]	169 eye fundus images with mild, moderate and severe DME.
DME, AMD	DUKE-45 [47]	45 OCTs with 15 AMD, 15 DME and 15 normal subjects. Each OCT volume has 100 B-scan with resolution of 512×1000 pixels.
	NOOR HOSPITAL [92]	148 OCTs as follows: 50 DME, 50 normal and 48 AMD subjects.
	ZHANG LABDATA [93]	109309 scans of subjects with DME, drusen, choroidal neovascularization and normal conditions.
DME, AMD, DR	SERI-CUHK [18]	75 OCTs labelled as: 16 normal, 20 DME and 39 DR-DME. The OCT volume contains 128 B-scans with resolution of 512×1024 pixels.
DR, AMD, GLAUCOMA	UK BIOBANK [65]	231806 OCTs and eye fundus images with the labels of glaucoma, DR and AMD.

Table **2-1**: A summary of free public ocular data sets with the ocular diseases graded by experts, the dataset name and the dataset description.

$$f\text{-}score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{2-6}$$

$$Kappa\ Coefficient = \frac{p_o - p_e}{1 - p_e} \tag{2-7}$$

where,

- *TP=True positive (the ground-truth and predicted are non-control class).*
- *TN=True Negative (the ground-truth and predicted are control class).*
- *FP=False Positive (predicted as non-control class but the ground-truth is control class).*
- *FN=False Negative (predicted as control class but the ground-truth is non-control class).*
- *po=Probability of agreement or correct classification among raters.*
- *pe=Probability of chance agreement among the raters.*

2.5 Deep learning methods for diagnosis support

2.5.1 Deep learning methods using eye fundus images

The state-of-the-art DL methods to classify ocular diseases using EFIs are focused in conventional or vanilla CNN and multi-stage CNN. The most common vanilla CNN used with EFIs are the pretrained inception-V1 and V3 models on ImageNet database[1]. The inception-V1 is a CNN that contains different sizes of convolutions for the same input to be stacked as a unique output. Another difference with normal CNN is that the inclusion of convolutional layers with kernel size of 1×1 at the middle and global average pooling at the final of its architecture [98]. On the other hand, the inception-V3 is an improved version batch normalization and label smoothing strategies to prevent overfitting [119].

Aujih et al. [109] used the U-Net model proposed by Ronneberger et al. [120] to segment the retinal vessel from EFIs. Then, two new data sets were created with and without the vessels to be used as inputs in the inception-V1. This method obtained an AUC of 0.9772 in the detection of DR in DRIU dataset. Gulshan et al. [113] and Gao et al. [111] proposed a patch-based model composed by pretrained inception-V3 to detect DR in EYEPAC dataset. Gulshan et al. used private dataset with segmentations of clinical signs to classify an EFI into normal or referable DR with a sensitivity of 93.4% and specificity of 93.9% [113]. The ensembled of four inception-V3 CNN by Gao et al. reached an accuracy of 88.72%, a precision of 95.77% and a recall of 94.84% [111].

The multistage CNN are centered first in the detection of clinical signs to sequentially grade the ocular disease. Yang et al. [110] located different type of lesions to integrate an imbalanced weighting map to focus the model attention in the local signs to classify DR obtaining an AUC of

[1] http://www.image-net.org/

0.9590. Quellec et al. [112] used a similar approach to generate heat maps with the detected lesion as an attention model to grade in an image-level the DR with an AUC of 0.954. Perdomo et al. [16] uses a four-layers CNN as a patches-based model to segment exudates and the generated exudate mask was used to diagnose DME reporting an accuracy of 82.5% and a Kappa coefficient of 0.6 [17]. Then, Perdomo et al. [13, 15] proposes a three-stage DL model: optic and cup segmentations, morphometric features estimation and glaucoma grading, with an accuracy of 89.4%, a sensitivity of 89.5% and a specificity of 88.9%. Finally, Wang et al. [114] proposed a model to segment optic disc and cup and calculate a normalized cup-disc-ratio to discriminate healthy and glaucomatous optic nerve of EFIs. The table **2-2** presents a brief summary of DL methods in eye fundus images used to support the ocular diagnosis using.

Ocular disease	**Dataset used**	**Authors**	**Methods**	**Performance**
DR	DRIVE	[108]	Gaussian Mixture Model with an ensemble classifier	AUC of 0.94
		[109]	Pre-trained Inception V1	AUC of 0.9772
	EYEPACS	[110]	DCNN with two stages	AUC of 0.959
		[111]	An ensemble of 4 pre-trained Inception V3	Acc. of 88.72% Prec. of 95.77% Recall of 94.84%
	EYEPACS & E-OPHTHA	[112]	Two linked DCNN	AUC of 0.954 and 0.949.
	EYEPACS & MESSIDOR & Private dataset	[113]	A pre-trained Inception V3	Sens. of 93.4% Spec. of 93.9%
DME	MESSIDOR & E-OPHTHA	[16, 17]	DCNN with two stages	Acc. of 82.5% Kappa of 0.6
GLAUCOMA	DRISHTI-GS & REFUGE	[114]	DCNN with two stages	AUC of 0.8583
	DRISHTI-GS & RIM-ONE	[115]	Classical filters and an active disc formulation with a local energy function	Acc. of 0.8380 and 0.8456
		[15, 115]	DCNN with three stages	Acc. of 89.4% Sens. of 89.5% Spec. of 88.9% Kappa of 0.82
AMD	AREDS	[13]	DCNN	Acc of 75.7%

Table **2-2**: An overview of the main state-of-the-art DL methods to ocular diagnosis using EFIs. The dataset and the method used in the study, with the performance obtained by the method.

2.5.2 Deep learning methods using optical coherence tomography

The most representative DL methods to detect abnormalities in OCT obtained an outstanding performance using vanilla CNN models as reported with: ResNet [55, 80], VGG-16 [121] and Inception-V3 [122]. The VGG-16 CNN contains five-blocks of convolutional layers and max-pooling to perform the feature extraction [97]. The final block is composed of three fully connected layers to discriminate among a number of classes. The ResNet model contains a chain of interlaced layers that adds the information from previous layers to future layers to learn residuals errors [123]. Gholami et al. used a pretrained ResNet to differentiate healthy OCT volumes from DR with an accuracy of 97.55%, a precision of 94.49% and a recall of 94.33% [80]. Kamble et al. combined the Inception and the ResNet model into a model termed as inception-ResNet-V2 [18]. This model was able to classify DME scans with an accuracy of 100% using the SERI dataset. On the other hand, the best DCNN model using OCT volumes as input are customized models with two or three stages. In particular, these DL models used two or more data sets reported in Table **2-1** to perform the feature extraction of local signs, added to a classification stage for grading the ocular diseases as reported for OCTs in [14, 116, 117]. De Fauw et al. defined a two-stage DL method to segment abnormalities from the OCT volume into a 3D representation [122]. The generated segmentation was stacked with the 43 most representative cross-sectional scans from an OCT volume. This model obtained an AUC of 0.9921 to determine the grade of AMD in private data sets. Finally, Perdomo et al. proposed a customized DL method called OCTNET [10]. This CNN is based in four blocks of convolutional and max-pooling layers, and a final block with two dense layers and a dropout layer to avoid overfitting during training. In addition, the proposed model classifies in scan and volume levels, delivering highlighted images with the most relevant areas for the model. The model was assessed for DR and DME detection with a precision of 93%, an AUC of 0.86 and a Kappa coefficient of agreement of 0.71. The proposed model presented a sensitivity of 99% and an AUC of 0.99 for the classification task of OCT volumes as healthy and AMD. The table **2-3** reports an overview of the most prominent works used to support the diagnosis of ocular conditions using OCTs.

2.5.3 Discussion

The number of free public available data sets contributes to the design of new DL methods for classifying eye conditions as indicated in Table **2-1**. However, the use of a private data set limits the comparison between performance metrics achieved by DL methods [93, 113, 121, 122]. Replication of studies reported by Gulsan et at. [113] and De Fauw et al. [122] have criticized the lack of information related to method description and the hyperparameters used as reported by [124].

The main DL methods in the detection of eye diseases by EFIs are focused on the development of pre-trained CNNs such as Inception V1 [109] and Inception V3 [111]. In addition, pre-trained CNNs applied to OCT performed outstandingly performance as reported with pre-trained ResNet [55, 116], VGG-16 [121], and Inception V3 [93] as presented in Tables **2-2** and **2-3**. Despite the good results obtained, these studies did not report the same performance measures on a common image database, so performance figures are not directly comparable, making it difficult to replicate and evaluate the true performance of these methods.

Finally, the main issues were summarized in two areas: these methods are limited to conventional

or pretrained CNNs, the reproducibility problems of more advanced methods and the lack of domain knowledge; in the data sets, high variability of illumination, artifact noise, resolution and quality between images was found; different types of magnification (diopter lenses) in the same dataset and a variable number of scans depending on the acquisition device. In this doctoral thesis we verified our initial hypothesis that automatic analysis of eye images based on deep learning models could help support medical diagnosis in diabetes-related eye conditions. Furthermore, the integration of domain knowledge may improve the classification of different diabetes-related eye conditions, as will be assessed and reported in the following sections.

Ocular disease	**Dataset**	**Authors**	**Methods**	**Performance**
DR	OCTID	[116]	Pre-trained ResNet model	Acc. of 97.55; Prec. of 94.49; Recall of 94.33.
DME	SERI	[55]	Pretrained Inception-ResNet-V2	Acc. of 100%
	SERI + CUHK	[14, 117]	OCTNET with 16 layers, class activation maps and medical feedback	Precision of 93,0%; Kappa of 0.71; AUC of 0.86
GLAUCOMA	POAG	[81]	A 3D-DCNN with 6 layers	AUC of 0.89
AMD	A2A SD-OCT	[118]	HOG Feature Extraction and PCA, with SVM and Multi-Instance SVM classifiers	Acc. 94,4%, Sens. 96.8 % Spec. 92.1%
		[10, 118]	OCTNET with 16 layers, class activation maps and medical feedback	Sens. of 99%; AUC of 0.99
	PRIVATE DATASET	[122]	DCNN with two stages by Google	AUC of 0.9921
		[121]	Pretrained VGG-16 model	AUC of 0.9382
		[93]	Pretrained Inception-V3 model	AUC of 0.9745; Acc. of 93.45%.

Table **2-3**: An overview of the main state-of-the-art DL methods to ocular diagnosis using OCTs. The dataset and the method used in the study, with the performance obtained by the method.

3 Two-stage deep learning for DME classification

This chapter presents a deep learning-based method to localize exudates and to grade the diabetic macular edema using eye fundus images. The proposed method contains two stage: first, a deep learning stage to automatically localize exudates in eye fundus images without taking in consideration the image size. This stage delivers a binary-mask with the segmentation of the exudates. Then, another deep learning model uses a stacked binary mask with the raw RGB fundus image as 4-channel input image for DME diagnosis. The results of the patch-based method to localize exudates using eye fundus images were published in the *12th International Symposium on Medical Information Processing and Analysis* [16] and, the results about the two-stage method for DME detection were published in the *Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop, OMIA 2016, Held in Conjunction with MICCAI 2016* [17].

3.1 Introduction

Diabetes mellitus is one of the leading causes of death according to the World Health Organization (WHO) and it is treated as an epidemic due to the worldwide associated costs and the number of people that suffer it [1]. Diabetic retinopathy is a condition caused by prolonged DM, causing blindness worldwide in persons in their productive age (20-69 years) [125]. However, one of the main problems associated to DR is that most people have no symptoms, and suffer the disease without being diagnosed [1]. DR is defined as one of the microvascular complications of diabetes, which affects the small vessels of the retina (the innermost eye light sensitive layer), being a major cause of blindness [126]. Thus, an effective prognosis of DR or an early diagnosis of diabetes may help to avoid loss of vision or an irreversible blindness [4, 5].

Diabetes macular edema is a complication of DR that occurs when the vessels of the central part of the retina (macula) are affected by accumulation of fluid and exudate formation in different parts of the eye [127]. DME classification is performed by experts as follows:

- Class 0: No signs of retinopathy.
- Class 1: Mild. Presence of exudates in the macula, but without affecting the central macular area.
- Class 2: Moderate. Presence of exudates in central macular region without affecting the fovea.
- Class 3: Severe, Retinal thickening or presence of exudates that affect the foveal area or retinal thickening [127, 128].

[1] https://www.who.int/diabetes/goal/en/

The fundus imaging is a common clinical exams that presents some advantages as: high patient compliance, dilation not required in most cases, quick and simple technique, images can be saved and analyzed at a later time, and progression of diseases can be monitored over time [6, 129, 130]. A faster diagnosis of the disease grade may help to a proper clinical treatment, improving the patients quality of life. Macular exudates can be detected by examining color photographs of the back of the eye, which is interpreted by a specialist ophthalmologist in retina. This process has some drawbacks such as being very time consuming and repetitive for clinical personnel, depending on the ophthalmologist's experience, and its susceptibleness to inter-observer variability. Moreover, the analysis of lots of images without any pathologies increases the work time, but decreases the time of analysis to others images with pathologies. Computer-aided systems (CADx) are an interesting alternative to tackle with these problems, CADx systems perform an automatic assessment of the disease grade, they may increase the number of patients diagnosed and may reduce the time to detect ocular diseases [5, 128].

Besides, the huge work-time used to analyse lots of images from a healthy population, have repercussions in a decreasing available time to analyse pathologic images. Computer-aided systems (CADx) are an explored alternative to tackle these problems. These systems perform an automatic grading disease, they could increase the patients diagnosed, and may support experts and reduce the work-time to detect ocular diseases [28]. Machine learning algorithms have been widely used successfully to DR and DME diagnosis using eye fundus images [28–30, 38–40].

The development of automatic retinal image algorithms is the basis for the building of a screening tool for the detection of ocular diseases. Vandarkuhali and Ravichandran [38] explored the use of extreme learning machine and probabilistic neural network applied to the detection of blood vessels in eye fundus images. In addition, Priyadarshini et al. [29] applied modified extreme learning machine to cluster the patients with diabetic and non-diabetic features using clinical data. Gurudath et al. [28] worked in machine learning using a three-layers artificial neural network with a support vector machines to classify retinal images with healthy, non-proliferative diabetic retinopathy and proliferative diabetic retinopathy.

Other researchers have reported algorithms to detect the presence of exudates in retinal images as a way to diagnose DME. Chanwimaluang et al, and Narasimha-Iyer et al. [131, 132] studied bayesian models that combines area-based and feature-based methods to localize exudates and non-exudate patches. Giancardo et al. [91] defined a set of features representing the colour, decomposition, wavelet and automatic lesion segmentation of exudates. These features were used for training a classifier for the automatic detection of exudates and classification of DME. Vasanthi and Banu [133] studied an adaptive neuro-fuzzy inference system (ANFIS) combined with an extreme learning machine to detect hard exudates for classifying DME. Akram et al. proposed method that combines a Gaussian mixture model and support vector machine to detect exudates and classify of different stages of diabetic macular edema [134]. Finally, Kunwar et al [135], proposed a method to combine texture feature extraction from regions with exudates next to the macula of retinal fundus image and it uses a SVM to grade DME.

This chapter reportes a two stage deep learning model: the first stage is a 48×48 patch classifier based on LeNet deep learning method [136]. Our patches classifier output generates a binary mask with the exudate segmentation in an eye fundus image. The second stage is a convolutional neural network that combines the raw input image with the binary mask segmentation to predict

DME class. The proposed method was assessed using two publicly available dataset and presents outstanding results when compared to state-of-the-art methods.

3.2 A deep learning model for exudate detection-DME classification

An overall view of our method is depicted in Figure 3.1. Our method is based on two convolutional neural networks. For successful training of deep neural networks a large number of training samples is required, in our case we need to have enough healthy and exudate samples, nevertheless, as is usual in many medical tasks, here the number of samples is limited and the class imbalance between the healthy and exudate patches is important and needs to be addressed. A standard way to obtain more samples is to rotate and flip the image, this was the data augmentation used in our experiments and is described in more detail in Section 3.2.1.

Our method (ED-DME classification) works in two stages. In the first one, the training of an exudate-detection model is done. This model is based on a LeNet [136] with 4 layers using the E-ophtha database with RGB patches of 48×48 pixels as input. The second stage is the processing of the Messidor database with 1200 images of size 512×512 pixels to generate a new dataset of 1200 grayscale mask images using the previous trained CNN model as predictor. The final stage is to train a DME-detection model based on the AlexNet architecture [95], which is composed of 8 layers using the RGB eye fundus images of 512×512 pixels plus the previous generated grayscale masks as the $4 - th$ input channel. One interesting aspect of our model is that it incorporates the mask of exudates detection trained in the first stage to serve as input to the DME-detection model, thus, giving more insights to the later model of what kind of patterns are we looking for.

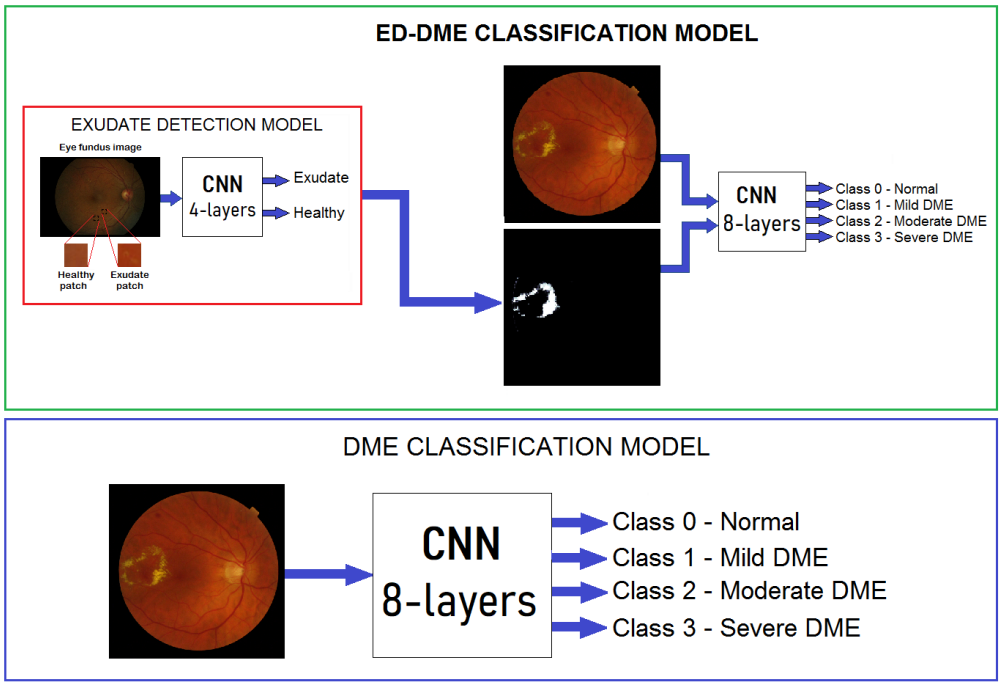

Figure 3.1: Block diagram with the proposed ED-DME Classification model and the DME classification model.

3.2.1 A deep learning exudate detection model

An overview of the patch-classifier is depicted in Figure 3.2. The first stage contain the preprocessing stages that detects and crops the input image with sizes ranging from 1440 $\times$960 to 2540 $\times$1690 pixels into non-exudate and exudate patches of 48 $\times$48 pixels. Then , the second stage is the LeNet CNN of four layers that uses the generated patches as an input, followed by a two sequential blocks of convolutional and max pooling layers to extract the features, and two fully connected layers with a softmax classier as the output layer.

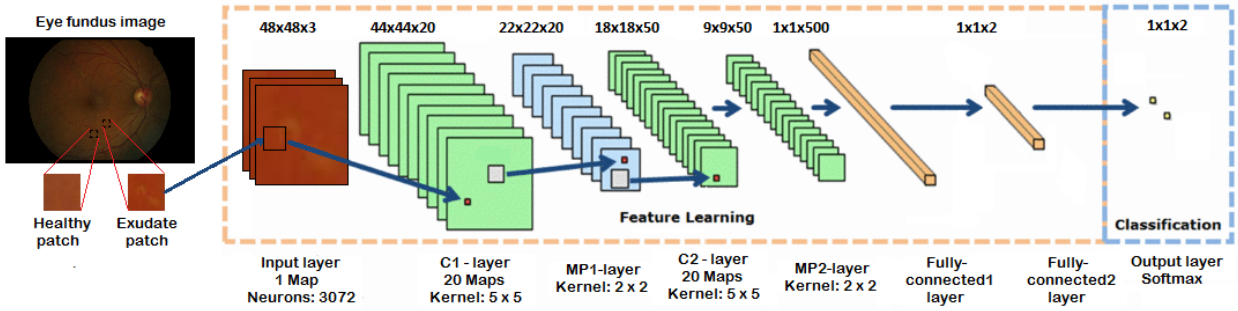

Figure 3.2: An overview of a CNN to classify patches from eye fundus images

3.2.2 Preprocessing stage

In the preprocessing stage the application of a set of transformations that helps to improve performance in the following stages is made. The first step is to envelope the exudate in a bounding box in order to extract the Region of Interest (RoI) from the eye fundus image of e-ophtha database in order to train the exudate-detection model as shown in Figure 3.3. Secondly, we define an oversampling strategy using to get more patches and help to prevent overfitting during training, this oversampling step showed significant improvements in our experimentation in comparison with no doing it. Finally, the resize of all the Messidor database images to a 512 $\times$ 512 sized images, keeping the aspect ratio.

Cropping

CADx systems use the previously identified RoI to classify in the whole film image. This RoI can be drawn by a time-consuming manual segmentation by an expert or automatically detected by an algorithm. The e-ophtha dataset contains binary masks with the manually segmentation of exudates and microaneurysm, where we fixed the average of exudates bounding-box to RoIs of 48 $\times$ 48 pixels according to the average lesion size. In addition, the RoIs were extracted without scaling and keeping the surrounding region. Then, these patches were cropped and labelled as an exudate patch if the intersection between the patch and the RoI was greater than the 60% of the RoI. Otherwise, the cropped patch was labelled as non-exudate or healthy patch. A set of 2136 patches with size of 48 $\times$ 48 that contain exudates information were cropped from e-ophtha dataset.

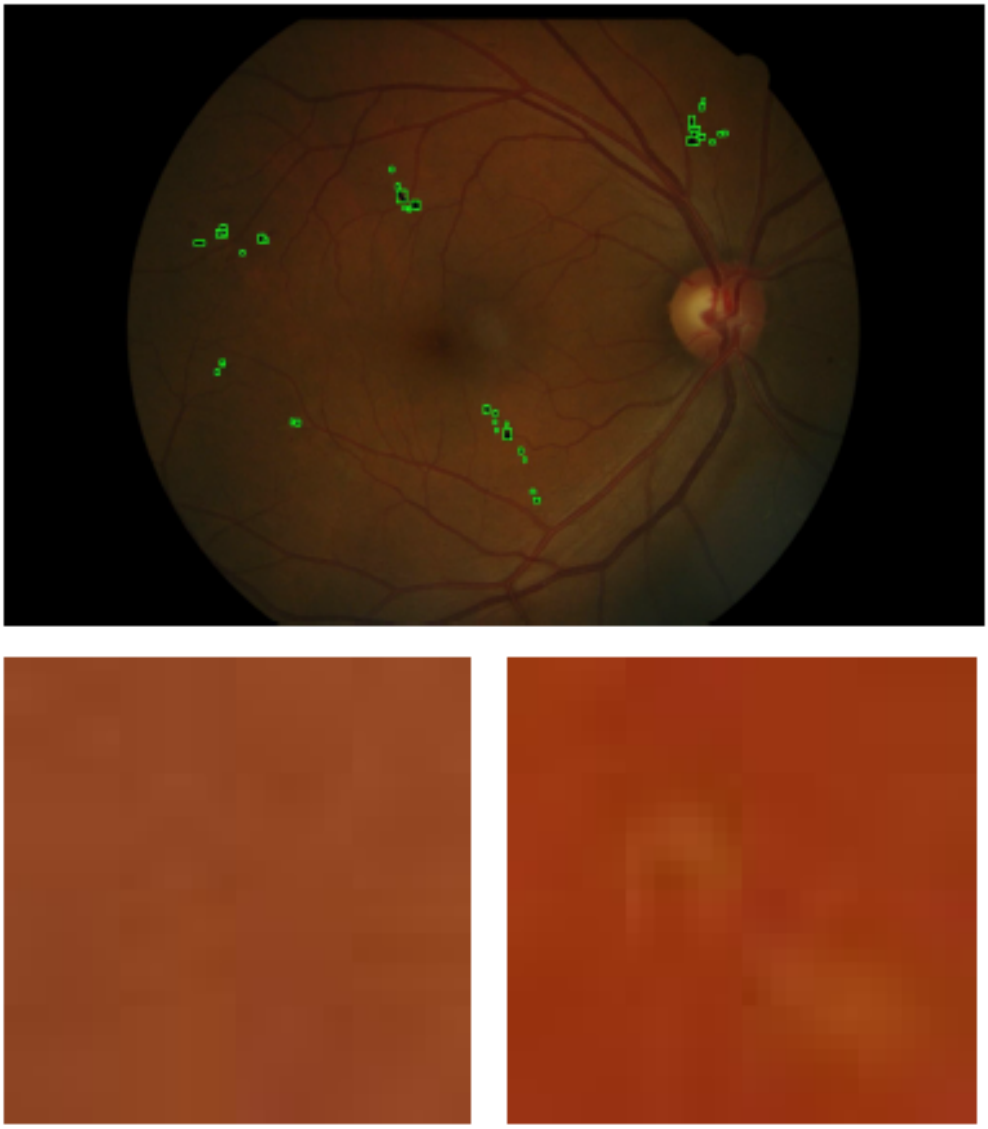

Figure 3.3: [Top] Eye fundus image with bounding boxes (green) and exudates (black). [Bottom] Eye fundus image with healthy and exudate patches respectively.

Data augmentation

The limitation of deep learning models comes mainly by the large number of parameters to learn. However, the design of complex models also increases the chance of overfitting the training data. Data augmentation strategy is a common way to tackle this issue. We apply the data augmentation strategy to artificially generate new samples using a set of flip and rotation transformations to the cropped patches' image. In classification problems, the data augmentation technique makes sense because exudates can be represented in any particular orientation. Thus, the patch-classifier model also should be able to learn from these representations. In total, We have artificially generated seven new label-preserving samples using a combination of the two transformations. On the other

hand, data augmentation was applied to the raw retinal image to increase the images as follows: 5 new label-preserving samples for the classes mild (Class 1 of DME) and moderate (Class 1 of DME) and 15 new label-preserving samples for the class severe (Class 2 of DME) to the Messidor dataset.

3.2.3 LeNet Convolutional Network

Convolutional Neural Networks (CNN) a type of bio-inspired artificial neural network based on the animal visual cortex. CNN are trained with the ability to respond to the variability and invariant to the exact position of the pattern in order to detect visual patterns with minimal preprocessing. The architecture of this CNN contains convolutional, pooling, and fully connected layers. Convolutional layer is a set of learnable windows (filters) moving through a stride with a kernel size that represent the receptive field. Each window is convolved computing the dot product between the filter and the input generating an activation map of that filter. Pooling layer is a non-linear function to reduce the size of the convolutional layer but extracting the most representative value in a windows defined by a kernel size with a stride. Max Pooling is the used due to this choose the maximum value of the filter. Fully-connected layer is a layer where all the neurons have full connections among all the neurons in the previous layer. The LeNet is kind of convolutional neural network [137] with the following structure as shown in Figure 3.2.

3.2.4 Exudate detection for DME classification model

The architecture of this model is summarized as follows. It contains five convolutional and three dense layers. Below, we describe convolutional, pooling and fully-connected layers, but this architecture has normalization layers proposed according bioinspired-based on inhibition schemes presented in the brain. The activations of normalization layer applies a transformation that maintains the values of mean activation close to 0 keeping the activation standard deviation close to 1 [95]. In the Table **3-1**, we summarized the layers, feature mas, layer size, kernel size and stride parameters as follows:

3.3 Experimental setup

3.3.1 e-Ophtha Dataset

The e-ophtha database contains 47 color fundus images with size ranging from 1440×960 to 2540×1690 pixels, which were segmented in order to find exudates by three ophthalmologists from the OPHDIAT Tele-medical network under the the French Research Agency (ANR) project, specially designed for scientific research in DR [71]. The labelled patches dataset created of 48×48 pixels with exudate and healthy classes after the preprocessing steps of cropping and data augmentation. These patches were randomly split by images where an image could only belong to a group with the following datasets distribution: training set with 8760 patches by each class, validation set has 328 by class and test set has 986 by class, representing 87%, 3.2% and 9.8% respectively.

Layer	Feature map	Layer size	Kernel size	Stride
Input	4	512×512	-	-
Convolutional	96	125×125	11×11	4 -
Batch Normalization	96	125×125	-	-
Max Pooling	96	62×62	3×3	2
Convolutional	256	58×58	5×5	1
Batch Normalization	256	58×58	-	-
Max Pooling	256	29×29	3×3	2
Convolutional	384	27×27	3×3	1
Batch Normalization	384	27×27	-	-
Convolutional	384	25×25	3×3	1
Batch Normalization	384	25×52	-	-
Max Pooling	384	12×12	3×3	2
Convolutional	256	10×10	3×3	1
Batch Normalization	256	10×10	-	-
Max Pooling	-	5×5	3×3	2
Dense	-	4096	-	-
Dense	-	2048	-	-
Dense	-	4	-	-
Output	-	4	-	-

Table **3-1**: Structure of CNN to classify DME with parameters used in each layer.

3.3.2 Messidor Dataset

The methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology or Messidor database contains images of the posterior pole of eyes from subjects diagnosed with DM, without data with respect to time of disease progression, sex or age of the subjects. Messidor is a research program funded by the Ministry of Research and Defense of France in a program TECHNO-VISION 2004. The database contains 1200 Messidor color photos Fundus provides information on the posterior pole. It was acquired by three ophthalmologic departments using a 3CCD color video camera in a retinograph TRC NW6 Topcon non-mydriatic a 45 degree field of vision. These images were captured using 8 bits per color with images ranged among 1440×960, 2240×1488 and 2304×1536 pixels. 800 images were acquired with pupil dilation (with a previous drop of tropicamide 0.5%) and 400 without dilation. The International Clinical Disease Severity Scale for DR concerning diabetic macular edema was performed by two ophthalmologists from Fundación Oftalmológica Nacional to classify as mild, moderate and severe according to on the distance of the exudates and thickening from the center of the fovea and normal without any exudates.The results after the classification of the four classes [128, 138] presented: 878, 140, 146 and 36 images respectively. After data augmentation stage for classes 1, 2 and, 3 applied to the training dataset, the distribution of the three sets was as follows: training dataset with 2215 images, validation dataset has 121 by class and test dataset has 239 by class, representing 70%, 10%, and 20% respectively.

3.3.3 Evaluation

The three classification models (ED, DME and ED-DME) were trained using stochastic gradient descent. The learning rate and batch size parameters were explored in a grid search manner and their values are listed in Table 1, using 30 as the number of epochs to train the model. The results of DME classification CNN was chosen as baseline. In this CNN, the performance was 92% in sensitivity and 40.6% in specificity applied in a image-based experimental setup using only the Messidor dataset. The proposed approach was implemented with Python 2.7 using GeForce GTX TITAN X from NVIDIA. During all the experiments, training loss and validation loss, as well as the accuracy over the validation set were monitored.

3.4 Results

3.4.1 Exudate detection

We choose the model proposed by Decencieriere et al as a baseline for the exudate detection[71]. In such work, they proposed a method based on machine learning and image processing techniques to detect exudates in eye fundus images reporting specificity and sensitivity in a patchwise experimental setup using the e-ophtha dataset.

We validated the proposed method applied to the test dataset. In Table **3-2** is reported the accuracy,sensitivity and specificity of the proposed method compared with the baseline method. The proposed method clearly outperforms Decencière et al. [71] in both sensitivity and specificity. This shows that the proposed method is able to better capture the visual features that characterize exudates.

Method	Accuracy	Sensitivity	Specificity
Decencière et al. [71]	-	90%	70%
Proposed model	99.6%	99.8%	99.6%

Table **3-2**: Performance measures in the baseline model and the proposed method in exudate detection task.

We assessed the ability of the ED model with the best parameters to create a mask from the test dataset that was randomly selected. Figure 3.4 presented an eye fundus image of 1440×960 pixels, ground truth generated by expert ophthalmologist and the mask generated using the exudate-detection CNN to classify patches as exudates if the output of the softmax layer is above 0.75; otherwise, it was classified as a healthy patch. The generation of the whole mask with the labeled exudates performed in 18000 patches of 48×48 pixels takes on average 1 minute per image.

The results of the systematic exploration are reported in Table **3-3**. The best performance of the ED model was obtained with a learning rate of 0.01 and a batch size of 64, and the best performance of DME-classification model was obtained with a learning rate of 0.01 and a batch size of 32.

Figure 3.4: [Right] Eye fundus image from test dataset;[Center] exudates manual label by ophthalmologist; [Left] exudates automatically labelled by LeNet CNN

CNN model	Learning rate	Batch size (train)	Loss (train)	Loss (val)	Accuracy (val)
ED model	0.01	64	0.0034	0.0002	99.80
ED model	0.001	64	0.4550	0.4763	80.00
ED model	0.00001	64	0.5850	0.5956	70.39
DME model	0.01	32	0.85	0.78	77.00
DME model	0.001	32	0.915	0.825	73.50
DME model	0.0001	32	1.388	1.301	64.50

Table **3-3**: Performance of ED and DME models with different batch size and learning rates during training.

3.4.2 DME classification

Then, we validated the proposed method comparing the results of the DME and ED-DME classification models applied to the test dataset. In Table **3-4** is reported the the performance metrics of sensitivity and specificity of the two methods. The proposed method clearly outperforms in detection of classes 0, 1 and 2. This shows that the proposed method is able to better capture the visual features that characterize exudates to classify diabetic macular edema.

3.5 Conclusion

We have presented two novel models for exudate detection and DME classification using eye fundus which are able to detect effectively exudates and the patterns in the 4 classes of diabetic macular edema.

The experimental results showed that the proposed method is highly effective to detect normal subjects in eye fundus images. The results clearly improved the sensitivity and specificity results produced by the single DME-method, thanks to the ability of the proposed method to first locate exudates, generating information as a mask in the input layer to improve the classification of DME. The proposed method is a good option to detect other diseases where the previous detection of abnormalities may improve the disease classification. Other kind of images of the eye, such as, angiography mask with the vessels segmentation may improve the performance to classify other

Method	Classes	Sensitivity (%)	Specificity (%)
DME model	Non-DME	87.1	56.6
	Mild DME	28.1	90.8
	Moderate DME	37.5	90.6
	Severe DME	15.0	98.1
	mean ± std	41.9 ± 31.5	84.0 ± 18.6
ED-DME model	Non-DME	94.9	87.3
	Mild DME	45.8	92.0
	Moderate DME	55.2	93.8
	Severe DME	30.0	98.2
	mean ± std	**56.5 ± 27.6**	**92.8 ± 4.5**

Table **3-4**: Comparison of performance metrics for DME and ED-DME models

ophthalmological disease, such as diabetic retinopathy.

Early detection of clinical signs in macular edema may improve diagnosis of the disease and the classification between the different stages of DME. This research is a first step in building telemedicine tools that can support medical personnel for detecting ophthalmic diseases using convolutional neural networks in eye fundus images.

4 Deep Late fusion for Glaucoma classification

This chapter presents two deep learning-based methods that fuse features extracted from eye fundus images with morphometric features from the optic disc and the cup segmentation for glaucoma classification. The two methods use cartesian and polar representations of optic nerve and cup, as novel methodologies to support glaucoma diagnosis. The first model presents a late-fusion strategy to merge 19 morphometric features extracted from the cartesian representation of optic disc and cup segmentation, with features obtained from the eye fundus image using a customized deep convolutional neural network. The second method includes the following sequential stages: segmentation polar transformation of the optic disc and physiological cup, estimation of 448 radii from previous segmentations, and the data fusion of polar morphometric features with features extracted from polar fundus image using a ResNet50. The results of the first method were published in the *13th International Symposium on Medical Information Processing and Analysis* [15] and the other results will be submitted to *Plos One*.

4.1 Introduction

Glaucoma is related to the progressive degeneration of optic nerve fibers and structural changes of the optic nerve head [42]. Although glaucoma cannot be cured, its progression can be slowed down by treatment. Therefore, timely diagnosis of this disease is vital to avoid blindness [50]. Glaucoma diagnosis detection is based on manual assessment of the Optic Disc (OD) through ophthalmoscopy and posterior eye fundus image analysis, looking morphological parameters for the central bright zone called the physiologic cup and a peripheral region called the neuroretinal rim [42, 50].

The automatic cup-to-disc ratio (CDR) in eye fundus images has been used as the main physiological characteristics in the diagnosis of glaucoma [139]. Some researchers have been focusing on global and regional features such as texture, gray-scale and wavelet energy of the Optic Nerve Head (ONH) to classify normal and glaucoma images [49]. Other study focused in texture property computed over the total image using Haralick features combined with neural networks [140]. Deep learning architecture has been explored for other eye conditions diagnosis [17] and automated glaucoma diagnosis using cartesian representations. Sevastopolsky used a modified U-Net CNN on publicly available eye fundus images DRIONS-DB, RIM-ONE v3, DRISHTI-GS to do optic disc and cup segmentation [141]. Chen et al. built a six CNN layers to get best performance of glaucoma diagnosis [142]. Orlando et al. studied pre-trained OverFeat and VGG-S CNN from non-medical data applied to eye fundus images in order to detect glaucoma [143]. On the other hand, the works related with polar domain in eye fundus images is focused to segmentation task of OD and PC. Fu et al. [144] proposed the a CNN named M-Net to segment the OD and PC jointly in a one-stage using polar fundus images. Liu et al. [145] combined two-branch to learn translation equivariant and rotation equivariant representations from Cartesian and polar domains respectively to segmen-

tation task of OD and PC. Liu et al. [146] formulated the joint segmentation task of OD and PC using a spatial-aware neural network that uses polar fundus images. Fu et al. [147]proposes the DEnet that combines four deep hierarchical context: global image, segmentation masks, cartesian optic disc and optic disc polar transformation. The output merges the probabilities of each output as the final glaucoma screening result. This chapter presents two deep learning late fusion models for automatic classification of eye fundus images to classify glaucoma condition. The first proposed model explores the strategy of using a deep convolutional neural network architecture merged with cup-to-disc morphological features to diagnose glaucoma. Also, the second proposed model uses the OD and PC segmentations to estimate radii from the polar binary mask of cup-to-disc to be merged with features obtained of polar fundus images using a DCNN for glaucoma diagnosis. The proposed methods were assessed using two publicly available datasets, and these methods report good performance metrics when compared to state-of-the-art methods.

4.2 Combining morphometric features and convolutional networks fusion for glaucoma classification

The proposed data fusion model is depicted in Figure 4.1. In the first stage, 19 morphological features are extracted using disc and cup segmentation. The second stage learns a set of features using a DCNN. The final stage combines both, morphological and convolutional features merging them to feed the loss function. Kappa loss function is preferred over the traditional softmax function since, there is a relation between grades of diagnosis [16]. The model is trained jointly by applying stochastic gradient descent.

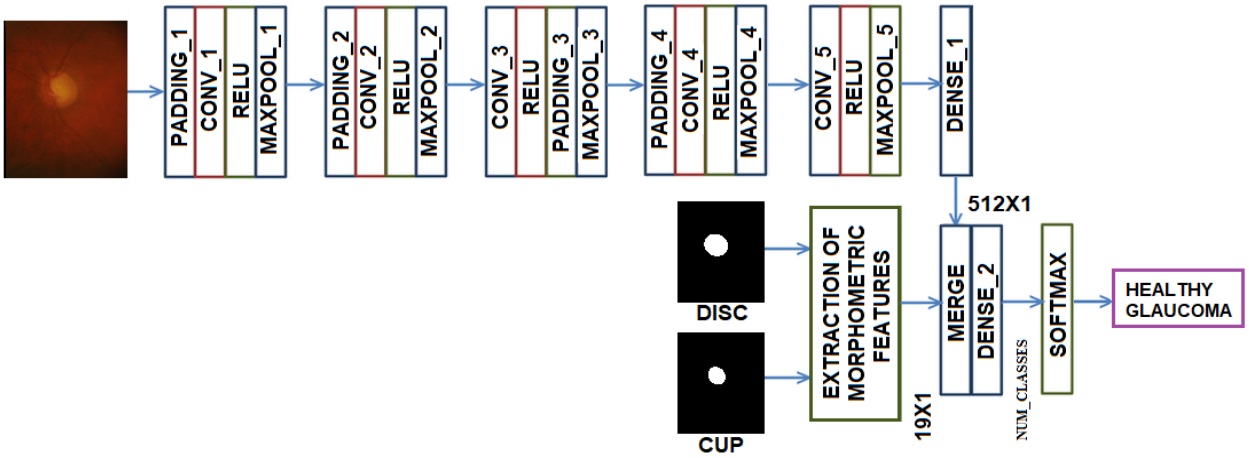

Figure 4.1: Late data fusion of features extracted and morphometric features to classify glaucoma condition in eye fundus images

4.2.1 Automatic extraction of cartesian morphometric features of eye fundus images

It has been shown that the eye morphometry in fundus images helps to glaucoma diagnosis [139]. This work proposed a set of 19 morphometric features based on the optic disc and physiologic cup

segmentations. The first step is to extract the four quadrants from the image as shown in Figure 4.2. Secondly, the cup-to-disc distance for each quadrant and the major radio and the minor radio for

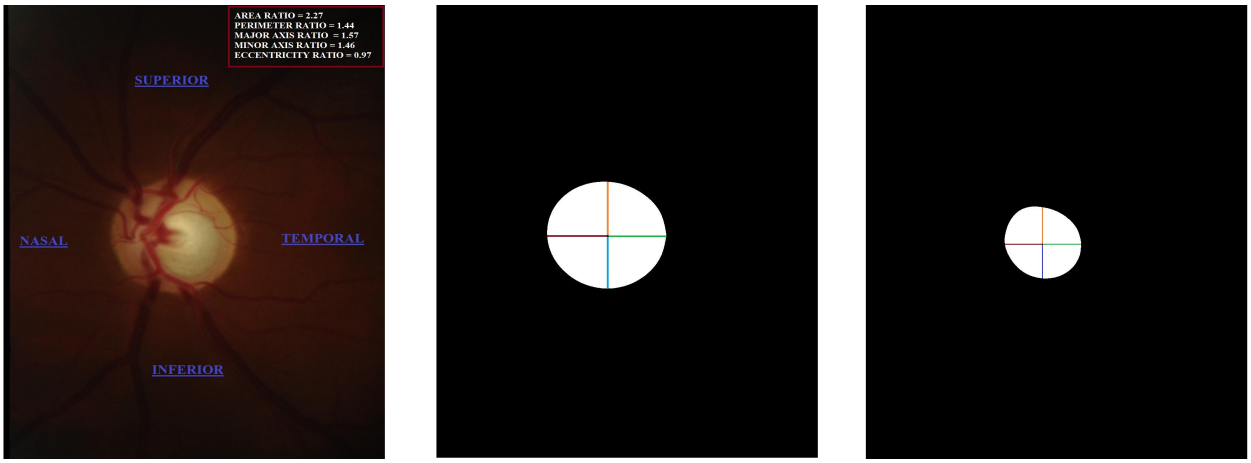

Figure 4.2: Morphometric features from cartesian representation of disc and cup segmentation.

optic disc and cup were calculated. Also, geometric features such as: perimeter, area, eccentricity were calculated for both, optic disc and cup. Additionally, 5 ratios were included in order to capture disproportions between optic disc and cup. Table **4-1** summarizes the 19 morphometric features proposed in this work.

Table **4-1**: List of morphometric measures extracted from disc and cup segmentation.

Distance features	Geometric features	Ratio features
Superior distance cup-to-disc		
Inferior distance cup-to-disc	Area optic disc	Cup-to-disc area ratio
Temporal distance cup-to-disc	Area physiological disc	Cup-to-disc major axis ratio
Nasal distance cup-to-disc	Perimeter optic disc	Cup-to-disc minor axis ratio
Major axis optic disc	Perimeter physiological cup	Cup-to-disc perimeter ratio
Minor axis optic disc	Eccentricity optic disc	Cup-to-disc eccentricity ratio
Major axis physiological cup	Eccentricity physiological cup	
Minor axis physiological cup		

4.2.2 DCNN for features extraction of eye fundus images

The DCNN is a model designed with a big number of layers to learn a representation of data containing spatial relations. This is the case of eye fundus images, where spatial patterns are determinant to diagnose different eye diseases, e.g. glaucoma, making the DCNN a suitable approach for image classification. DCNN learns a set of features using a minimal preprocessing while, with the properly supervised training, may respond to distortion, variability and invariant patterns. The DCNN is composed of 5 convolutional layers with kernel size of 3×3 and strides of 1×1, 5

max-pooling layers with pool size of 2×2 and stride of 2×2, 4 zero-padding layers with padding of 1×1, and 2 fully-connected layers with 512 and x-class number units as presented in Figure 4.1. A convolutional layer is composed of a set of learnable filters that convolved with the input generating an activation map for each filter. The convolutional layer output is the input of a max pooling layer that is a non-linear size reducer that is applied to the activation choosing the maximum value of a set of contiguous pixels. A zero-padding layer adds a set of pixels of value 0 to increase the image size but without affecting the image information, these layers were applied in order to ensure an even dimension at max-pooling layer's output. Finally, the fully-connected layer connects all the neurons in the previous layer to the next layer. The DCCN architecture used in this work is described in Table **4-2**.

Table **4-2**: Architecture of the DCNN with values used in each layer.

N	Name	Channels	Width	Height	Filter size	Stride
0	Input	3	224	224	-	-
1	Padding2D	3	226	226	-	-
2	Conv2D	32	224	224	3×3	1×1
3	MaxPool2D	32	112	112	2×2	2×2
4	Padding2D	32	114	224	-	-
5	Conv2D	64	112	112	3×3	1×1
6	MaxPool2D	64	56	56	2×2	2×2
7	Conv2D	64	54	54	3×3	1×1
8	Padding2D	64	56	56	-	-
9	MaxPool2D	64	28	28	2×2	2×2
10	Padding2D	64	30	30	-	-
11	Conv2D	64	28	28	3×3	1×1
12	MaxPool2D	64	14	14	2×2	2×2
13	Conv2D	64	12	12	3×3	1×1
14	MaxPool2D	64	6	6	2×2	2×2
15	Dense	512	-	-	-	-
16	Dense	`Num_classes`	-	-	-	-
17	`Kappa_Loss`	`Num_classes`	-	-	-	-

4.2.3 Experimental setup

RIM-ONE v3 dataset

The RIM-ONE v3 database with eye color fundus images was used in this study. The database contains 159 images with size of 1072×1424 pixels, 85 images from healthy subjects, 35 images with a suspected glaucoma diagnosis and 39 images with glaucoma diagnosis. The images were labeled and two binary masks with optic disc and cup were performed by two ophthalmologist experts from the Department of Ophthalmology at the Hospital Universitario de Canarias in Spain [51]. The first

proposed method was evaluated in 2 setups: an unbalanced 3-class classification setup (healthy, suspicious and glaucoma) and a binary classification setup (healthy vs. suspicious+glaucoma), this was created to balance the classes and to assess the detection ability of models to discriminate between healthy class vs non-healthy class. This dataset was randomly split in a patient basis training (60%), validation (10%) and test (30%) subsets with stratified sampling.

4.2.4 Evaluation

Two configurations of the proposed model were evaluated a configuration using only the convolutional network (DCNN) and a configuration that also includes the morphometric features (DCNN + MFs). The models were trained using stochastic gradient descent on both the 2-classes and the 3-classes problems. The learning rate (lr) parameter was explored for each model using a grid search strategy, the best performing values found in validation are listed in Table 3, using 200 as the number of epochs to train the model. The 8-layers DCNN was chosen as baseline. The proposed approach was implemented with Keras using GeForce GTX TITAN X from NVIDIA. The Kappa coefficient was implemented as a cost function, and loss, precision, recall, f-score and Kappa measures were reported for both training and test sets. Finally, SVM and RF were evaluated as baseline methods and trained using the 19 cartesian morphometric features normalized with (mean = 0) and (variance = 1). C parameter for the linear SVM was explored.

4.2.5 Results

Experimental results are reported in Table **4-3**. The best performance of the proposed model was obtained with a learning rate of 0.01 for two classes and a learning rate of 0.0001 for three classes both with a batch size of 32. We evaluated the proposed model with the best parameters applied to the test dataset. Table **4-4** presents the precision, recall, macro averaged f-score and Kappa coefficient results of the proposed methods compared with baseline methods. The proposed method clearly outperforms SVM and RF in Kappa coefficient that measures inter-rater agreement among the binary classification (balanced setup) and 3-class problems (unbalanced setup) [148]. This coefficient is widely used in medicine to compare classification performances regardless of balanced or unbalanced setups [148]. According to [148] the Kappa presented at the two classification problems presented good (0.60 to 0.80) and moderate (0.40 to 0.60) agreements respectively. This showed the proposed method is able to capture the visual features and morphological features that characterize glaucoma and combine them to improve the glaucoma diagnosis.

4.3 Deep data fusion of polar fundus images and polar morphometric features for glaucoma detection

The CDR measures the ratio between of optic disc and physiological cup in the thinnest distance of the rim. Althought, the CDR approach is the gold standard accepted for the most experts, clinical application for glaucoma diagnosis, we discuss about two main problems: the detection of the thinnest distance is not an easy and objective task in cartesian domain and, the optic disc and the physiological cup could have more than one point with a close distance [149]. The motivation

Method	Num_classes	lr	Precision	Recall	f-score	Kappa
DCNN	2 Classes	0.1	0.0	0.0	0.0	0.0
		0.01	0.59	0.77	0.67	0.23
		0.001	0.55	0.23	0.32	0.0
		0.0001	0.50	0.05	0.08	0.0
	3 Classes	0.1	0.17	0.33	0.23	0.0
		0.01	0.17	0.33	0.23	0.0
		0.001	0.17	0.33	0.23	0.0
		0.0001	0.17	0.33	0.23	0.0
DCNN + MFs	2 Classes	0.1	0.74	0.77	0.76	0.52
		0.01	**0.90**	**0.86**	**0.88**	**0.78**
		0.001	0.81	0.77	0.79	0.61
		0.0001	0.80	0.73	0.76	0.54
	3 Classes	0.1	0.47	0.47	0.47	0.27
		0.01	0.42	0.40	0.41	0.08
		0.001	0.49	0.54	0.51	0.39
		0.0001	**0.46**	**0.56**	**0.50**	**0.42**

Table **4-3**: Performance of the two models with different learning rates in validation dataset. [In boldface] the best performance achieved at 2 setups.

Method	Num_classes	Precision	Recall	f-score	Kappa
SVM	2 Classes	0.74	0.77	0.75	0.52
RF		0.88	0.68	0.76	0.61
Propose method		**0.90**	**0.86**	**0.88**	**0.78**
SVM	3 Classes	0.63	0.56	0.55	0.33
RF		**0.64**	**0.57**	**0.58**	0.35
Propose method		0.46	0.56	0.50	**0.42**

Table **4-4**: Performance measures in the baseline models and the proposed method in test dataset. [In boldface] the best performance metrics achieved at binary class and three-classes classification problems

to represent images of the fundus in another domain is based on the difficulty for humans to find the CDR parameter in the narrowing of the edge of the asymmetric optic nerve in the Cartesian domain [150]. Thus, the polar domain could easily and reliably represent the distance between the optical disc and the cup in a set of radii and angles.

An overview of the method of deep data fusion using extracted polar characteristics and polar morphometric characteristics is shown in Figure 4.3. The proposed model contains four main stages with three deep learning-based architectures that perform segmentation, feature extraction and regression tasks. In the first stage, it is performed a RoI detection to crop the optic disc, and this cropped image is resized and transformed to polar coordinates. The second stage is a customized DCNN to segment the optic disc and cup using the polar representation of the eye

fundus image. Concurrently, a DCNN model uses previous binary masks to automatically estimate 448 radii from disc and cup, and the other stage learns a set of features using a ResNet50 DCNN and merge both features extracted, polar morphometric and convolutional, to calculate the loss function to grade glaucoma condition.

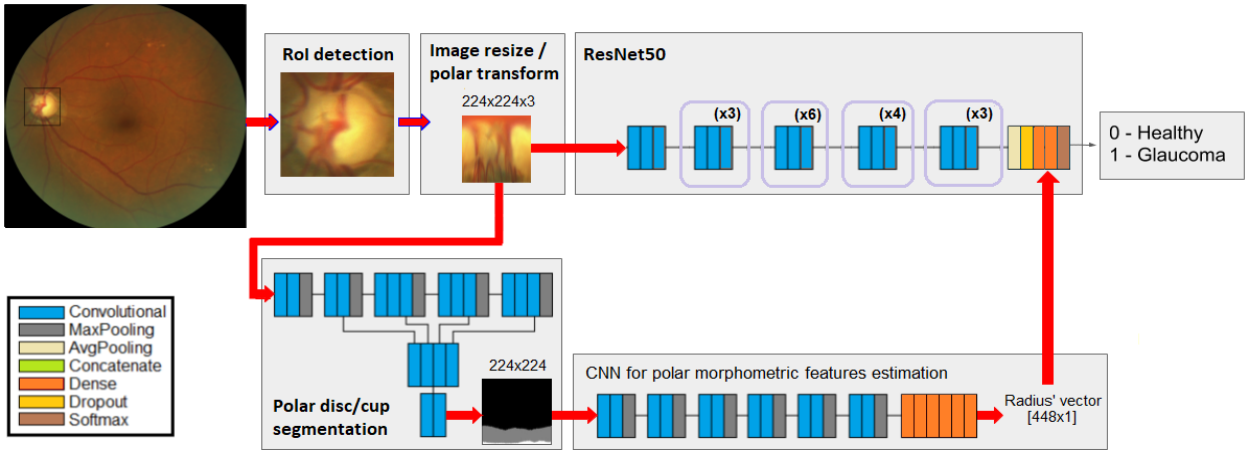

Figure 4.3: Late data fusion of features extracted and morphometric features to classify glaucoma condition in eye fundus images

4.3.1 Preprocessing stage

The preprocessing stage includes the automatic RoI detection and the resizing and polar transformation of cropped images. The automatic RoI detection in each fundus image is performed by searching the brightest and roundest part of the eye. This RoI is used to calculate a bounding-box of the optic disc and to crop the fundus image and the binary mask. The cropped images are resized to 224×224 keeping the aspect ratio. Then, the cartesian resized images are transformed to a polar coordinate system (r, φ), where $r = \sqrt{x^2 + y^2}$ and $\varphi = \text{atan2}(y, x)$. Based on the polar representation, the image is unfolded in such a way that the φ polar coordinate corresponds to the x Cartesian coordinate and the r polar coordinate corresponds to the y Cartesian coordinate. This is illustrated in Figure 4.4. These polar images are used to feed the segmentation and features extractions DCNNs.

4.3.2 Deep learning method for automatic segmentation of optic disc and physiological cup

The second stage of the model corresponds to a DCNN that receives as input an RGB eye fundus image and calculates a segmentation of the optic disc and the optic cup. The DCNN is based on a deep retinal image understanding (DRIU) model using the last four sub-blocks called coarse feature maps [151], but with two additional convolutional layers. The DRIU model contains 13 convolutional layers with different filters sizes and 4 max-pooling layers [151]. The DRIU model is initialized with VGG weights pretrained on ImageNet. It is fine-tuned for 1,000 epochs with a learning rate of $1e-5$, which is gradually decreased as the training process proceeds.

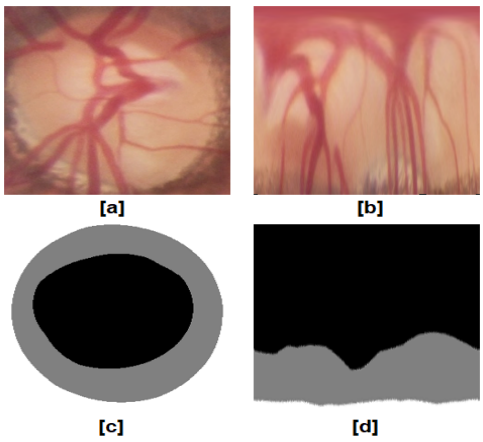

Figure 4.4: Images from a glaucomatous subjects: [a,c] Fundus and mask of optic disc and optic cup on cartesian coordinates; [b,d] Fundus and mask of optic disc and optic cup on polar coordinates

The coarse DRIU feature maps of the 4^{th}, 7^{th}, 10^{th} and 13^{th} convolutional layers are extracted, stacked and up-sampled to generate a binary mask with size 224×224. A modification of the original DRIU was done adding two convolutional layers in cascade with one-padding and kernels size of 3×3 and 1×1 to improve resolution details at the optic and cup edges, without affecting the final size. The model is trained using the combination of dice coefficient plus binary cross-entropy as loss function with manually ground-truth masks of the OD and the PC provided by experts.

4.3.3 Automatic extraction of polar morphometric features of eye fundus images

The automatic extraction of polar radii of OD and PC was done using a customized DCNN. This DCNN is based on VGG-16 and VGG-19 architectures with 6 sequential blocks of two convolutional layers and a max-pooling layer with an increasing number of filters to perform the feature extraction in the optic disc and optic cup masks [97]. The final block is composed by 6 dense layers with LeakyReLU activations with a negative slope coefficient alpha of 0.2 to make a regressor model which delivers the 448 radii given an image. The model is trained using mean absolute error as loss function, with 300 as the number of epochs and Adams optimizer.

4.3.4 DCNN for features extraction of eye fundus images

The DCNN is a model designed with a big number of layers to learn a representation of data containing spatial relations. This is the case of eye fundus images, where spatial patterns are determinant to diagnose different eye diseases, e.g. glaucoma, making the DCNN a suitable approach for image classification. DCNN learns a set of features using a minimal preprocessing while, with the properly supervised training, may respond to distortion, variability and invariant patterns. The DCNN is composed of 5 convolutional layers with kernel size of 3×3 and strides of 1×1, 5 max-pooling layers with pool size of 2×2 and stride of 2×2, 4 zero-padding layers with padding of 1×1, and 2 fully-connected layers with 512 and x-class number units. A convolutional layer is

composed of a set of learnable filters that convolve the input image generating an activation map for each filter. The convolutional layer output is the input of a max pooling layer that is a non-linear size reducer that is applied to the activation choosing the maximum value of a set of contiguous pixels. A zero-padding layer adds a set of pixels of value 0 to increase the image size but without affecting the image information, these layers were applied in order to ensure an even dimension at max-pooling layer's output. Finally, the fully-connected layer connects all the neurons in the previous layer to the next layer. The DCCN architecture used in this work is described in Table **4-5**.

Table **4-5**: Architecture of the ResNet50 with values used in each layer.

Name	Branch / Times	Channels	Width	Height	Filter size	Stride
Input	-	3	224	224	-	-
Padding2D	-	3	230	230	-	-
Conv2D	-	64	112	112	7×7	2×2
MaxPool2D	-	64	56	56	–	2×2
Conv2D	1 / (3x)	64	56	56	1×1	1×1
Conv2D	1 / (3x)	64	54	54	3×3	1×1
Conv2D	1 / (3x)	256	56	56	1×1	1×1
Conv2D	2 / (4x)	128	28	28	1×1	2×2
Conv2D	2 / (4x)	128	26	26	3×3	1×1
Conv2D	2 / (4x)	512	28	28	1×1	1×1
Conv2D	3 / (6x)	256	14	14	1×1	2×2
Conv2D	3 / (6x)	256	12	12	3×3	1×1
Conv2D	3 / (6x)	1024	14	14	1×1	1×1
Conv2D	4 / (3x)	512	7	7	1×1	2×2
Conv2D	4 / (3x)	512	5	5	3×3	1×1
Conv2D	4 / (3x)	2048	7	7	1×1	1×1
GlobalAveragePool2D	-	2048	-	-	-	-
Dense	-	1024	-	-	-	-
Dense	-	256	-	-	-	-
Dense	-	2	-	-	-	-

4.3.5 Experimental setup

REFUGE dataset

The REFUGE challenge database contains 1200 retinal fundus images for glaucoma diagnosis. The images are centered at fovea with high-quality, and the macula, the optic disc and the retinal nerve fiber layer visibles. The database was split into three equally-sized subsets with 400 images for training, validation, and test, with 40 glaucomatous and 360 non-glaucomatous images for each subset. The retinal images from the training set were acquired using a Zeiss Visucam 500 fundus

camera with a resolution of 2124×2056 pixels. On the other hand, the validation and test set images were acquired with a Canon CR-2 device with a resolution of 1634×1634 pixels. The optic disc and optic cup were manually segmented by experts from the Zhongshan Ophthalmic Center (China). The proposed method was trained using the training and validation sets as an one dataset with 800 images. The evaluation of the proposed method was performed using the test set.

Evaluation

The proposed approach (DCNN + polar MFs) was implemented with Keras and run in a computer with a GeForce GTX TITAN X from NVIDIA. The binary cross entropy was used as loss function. The accuracy, precision, recall, and f-score were reported in the test set. Three configurations of the proposed model were evaluated: two configurations using only the cartesian and polar fundus images applied as an input of the ResNet50 DCNN, and one configuration that fuses the cartesian and polar morphometric features with the cartesian and polar features extracted respectively from the ResNet50 (DCNN + Cartesian MFs) as explained in section 4.2. The previous models were trained using Adams in all the configurations. The number of epochs to train the model was set to 100, and the learning rate and batch size parameter were explored for each model using a grid search strategy, the best performance metrics obtained in validation set was evaluated in test set. In addition, five machine learning models were implemented: LR, KNN, RF, SVM and MLP were evaluated as baseline methods and trained using the 448 polar morphometric features normalized with (mean = 0) and (variance = 1). C parameter for the LR and the linear SVM, number of neighbors for the KNN, number of estimators for the RF, and number of hidden layers for MLP were explored. Moreover, the segmentation task of optic disc and optic cup was evaluated using dice coefficient. The estimation of polar morphometric features was assessed using mean absolute error.

4.3.6 Results

The best performance of the proposed model in the automatic segmentation of optic disc and physiologic cup (second stage) was obtained with a learning rate of $1e-5$, a batch size of 2, a number of samples per epoch of 400, and a number of epochs of 100. The Dice coefficient (DC) were monitored during training (not reported here), and the best model evaluated on the test set for optic disc segmentation with a mean of $DC = 0.972$, and physiologic cup segmentation with a mean of $DC = 0.886$. For the third stage, the estimated radii presented a MAE during training of 2.11%, and a MAE in the test set of 2.45% compared to the ground truth radii.

We evaluated the proposed method with baseline methods as reported in Table **4-6**. The table presents the comparison between the methods, source of information and performance metrics of sensitivity, specificity, accuracy, and AUC evaluated in the glaucoma classification task.

Finally, we compared the proposed method against the 12 best teams in the REFUGE Challenge. The REFUGE challenge was organized by the artificial intelligence innovation business, Baidu Inc and, it was a satellite event of the MICCAI 2018 conference in the 5th MICCAI Workshop on Ophthalmic Medical Image Analysis (OMIA) in Granada, Spain [152]. The segmentation and classification results are shown in Table **4-7** and Figure 4.5 respectively.

Table **4-6**: Performance metrics for baseline models and the proposed method on the test data set. The bold values show the two best score for each performance metric.

Method	Source	Sensitivity	Specificity	Accuracy	AUC
LR	Cartesian MFs	0.954	0.697	0.932	0.938
	Polar MFs	0.957	0.727	0.938	0.936
KNN	Cartesian MFs	0.959	**0.833**	0.950	0.892
	Polar MFs	0.959	0.806	0.948	0.929
RF	Cartesian MFs	**0.975**	0.705	**0.957**	0.910
	Polar MFs	0.959	0.714	0.938	**0.952**
SVM	Cartesian MFs	0.951	0.688	0.930	0.925
	Polar MFs	0.929	0.765	0.922	0.903
MLP	Cartesian MFs	0.962	**0.812**	0.950	0.941
	Polar MFs	0.959	0.694	0.935	0.938
DCNN	Cartesian images	0.934	0.682	0.920	0.908
	Polar images	0.949	0.700	0.930	0.939
	Cartesian images and MFs	0.937	0.762	0.928	0.934
	Polar images and MFs †	**0.983**	0.733	**0.953**	**0.971**

† The proposed model fusing polar images and polar MFs of OD and PC.

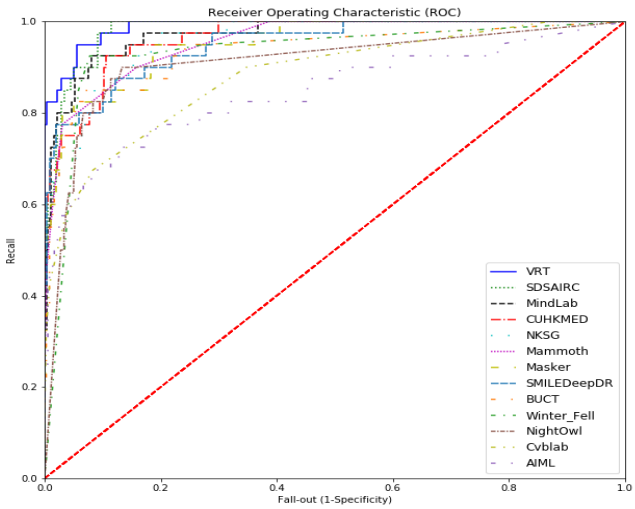

Figure 4.5: ROC curves of participating teams in REFUGE challenge. The legend is sorted according highest values of AUC

Table **4-7**: Methods and performance metrics for teams on REFUGE test set. The bold values show the three best score for each performance metric.

Team	Segmentation & classification methods	Mean OD Dice	Mean PC Dice	AUC
MindLab†	Customized DCNN & late-fusion of polar MFs and features extracted from ResNet-50 using polar fundus images	**0.9720**	**0.8867**	**0.971**
CUHKMED [153]	Combination of U-Net and DeepLabv3 & CDR calculated using segmentations	**0.9602**	**0.8826**	0.9644
Masker [154]	Combination of a Mask-RCNN, an U-Net and a M-Net & Combination of CDR and ResNet-50, 101 and 152	0.9464	**0.8837**	0.9524
BUCT [152]	Combination of two U-Net & a Xception using as an input grayscale images	0.9525	0.8728	0.9348
NKSG [155]	The DeepLabv3 using OD area & SENet using data augmentation	0.9488	0.8643	0.9587
VRT [152]	Combination of an U-Net and a vessel-based network CNNs & Combination of three CNNs using semi-supervised learning approach	**0.9532**	0.8600	**0.9885**
AIML [156]	Combination of dilated-ResNet & ensemble of ResNet-50, 101, 152 and 38 using the whole fundus images and local regions around the OD	0.9505	0.8519	0.8458
Mammoth [157]	A Dense U-Net & a combination of a ResNet-18 and a CatGAN using 4-folds	0.9361	0.8667	0.9555
SMILEDeep [158]	Three modified U-Net & a modified DeepLabv3 using cross-entropy loss	0.9386	0.8367	0.9508
NightOwl [159]	Two U-shaped networks 'CoarseNet' and 'FineNet' & 10-fold cross-validation using FineNet encoders	0.8257	0.9487	0.9101
SDSAIRC [160]	M-Net using the OD area & the fine tuning of a pretrained ResNet-50 combined with a LR for CDR estimation	0.9436	0.8315	**0.9817**
Cvblab [152]	A two-stage model using a modified U-Net & an ensemble of VGG19, InceptionV3, ResNet-50 and Xception	0.9077	0.7728	0.8806
Winter_Fell [161]	Combination of a Faster R-CNN and a ResU-Net & an ensemble of two ResNets (101 and 152), and two DenseNets (169 and 201)	0.8772	0.6861	0.9327

† The proposed model fusing polar images and polar MFs of OD and PC.

4.4 Discussion and conclusions

Experimental results for the first proposed model showed that the DCNN model combined with cartesian morphological features is highly correlated with the ground truth labels for glaucoma condition classification according to the Kappa coefficient as shown in table **4-4**. In addition, our proposed method is able to obtain a good performance for binary and three-classes classification problems using unbalanced sets. The main advantage of the proposed method is that it uses as input the raw image and the optic disc and physiologic cup segmentations to calculate morphologic features merging those two sources of information in a vector to improve the glaucoma classification.

The pipeline for the second proposed method based on the deep data fusion of polar features presented good results in the two segmentation tasks, overcoming the other methods with a mean optic disc dice of 0.9720 and a mean physiological cup dice of 0.8867 as reported in Table **4-7**. The use of RoI of optic disc in the polar domain to perform the segmentation of optic disc and physiologic cup was the differential compared to other methods that use cartesian representation. Additionally, the proposed method in classification task reported good results in sensitivity, accuracy, and AUC compared to classical machine learning approaches as reported in Table **4-6**. The comparison with the 12 best teams participating in the REFUGE challenge positions us in third place with an AUC of 0.971 as shown in presented in Table **4-7** and Figure 4.5.

The proposed methods compared to baseline methods deliver intermediate morphometric features in two domains that could offer additional information to expert medical decision making. In addition, the estimated radii of optic disc and physiological cup are useful for the detection of the thinnest area of the ring for the detection of clinical CDR parameter. The combination of DCNN with morphometric features showed good preliminary results in glaucoma detection in two different datasets. However, its application to other datasets will be the subject of our future work.

5 Three-stages Deep Learning for Glaucoma Classification

This chapter presents a deep learning-based method to automatically segment optic disc and physiological cup, extract morphometric features and use the features extraction and the morphometric features extracted to grade glaucoma using eye fundus images. This method is a multi-stage deep learning model for glaucoma diagnosis based on a three-learning-task strategy. The model is sequentially trained to solve incrementally difficult tasks. Our proposed model includes the following stages: segmentation of the optic disc and physiological cup, prediction of morphometric features from segmentations, and prediction of disease level (healthy, suspicious and glaucoma). The results of this method were published in the *Proceedings of the Ophthalmic Medical Image Analysis Fifth International Workshop, OMIA 2018, Held in Conjunction with MICCAI 2018* [13].

5.1 Introduction

Glaucoma is one of the leading causes of vision loss and blindness worldwide [162]. It is defined as an increment of intraocular pressure producing morphological changes in the optic disc (OD) and physiological cup (PC) affecting the ability of the optic nerve to transmit images to the brain [163]. The main problem with glaucoma is associated with a delayed diagnosis causing an irreversible damage to the eye [162, 163]. The examination of the optic disc, physiological cup and neuroretinal rim structures is important for an early detection and proper treatment [162].

The ocular tonometry or measurement of intraocular pressure does not quantify the damage or glaucoma progression [164]. Thus, a complete ophthalmoscopy examination through an eye fundus image is widely used to grade and monitor the disease [164, 165]. Additionally, an accurate and objective diagnosis is required to avoid the minimal damage to the eye structure [165]. Thus, the design of computer-aided diagnosis models for automatic disease assessment is important to improve the glaucoma detection and minimize the subjectivity in the diagnosis.

The cup-to-disc ratio (CDR) is the most typical morphometric feature used in the diagnosis of glaucoma. However, locating and segmenting the OD or optic nerve head (ONH) and the physiological cup are not easy tasks. Septiarini et al. proposed an automatic glaucoma detection method extracting statistical features from the intensity in ONH: the mean, smoothness and 3rd moment, and using a k-nearest-neighbor algorithm as a classifier [166]. Pardha et al. reported a region-based active contour model using multiple image channels and gray level properties for optic disc and cup segmentations [165].

Deep learning models, such as Deep Convolutional Neural Networks (DCNNs) have been applied with success to different medical image analysis tasks and, in particular, to automatically discriminate between glaucoma and non-glaucoma patterns in eye fundus images. Al-Bander et al.

presented an 8-layer CNN model to automatically extract features of the optic disc from the raw images and a linear Support Vector Machine (SVM) classifier to classify the images into normal or glaucoma subjects [167]. On the other hand, Chen et al. reported a deep learning model that contains four convolutional layers and two fully-connected layers, where a dropout layer and data augmentation strategies are used to improve the performance of glaucoma diagnosis [142]. In addition to this, Orlando et al. fine-tuned two deep learning approaches: OverFeat and VGG pre-trained from non-medical data for automated glaucoma detection on the DRISHTI-GS1 dataset [143]. Sevastopolsky reported a DCNN for automatic OD and PC segmentations, using a modification of the U-Net CNN tested on the DRIONS-DB, RIM-ONE v.3 and DRISHTI-GS1 database [141]. Finally, Abbas presented an unsupervised CNN architecture to extract the features and used a deep belief network model to select the most discriminative deep features with a softmax linear classifier to differentiate between glaucoma and non-glaucoma retinal fundus image [168]. Despite the results obtained, these studies have only been tested on a binary classification task, as many images are not clear cases but in a continuum between healthy and glaucoma.

Deep convolutional neural networks have been highly successful in solving several image analysis tasks. However, they require a large number of labeled samples for training, which is not necessarily the case when dealing with medical images. To mitigate this problem, we devised a strategy that trains the network in stages that solve incrementally more complex tasks [169]. The division of the problem into subtasks allows to better train the different network modules even with a small number of samples. The sequence of tasks is motivated by the current practice for glaucoma diagnosis from eye fundus images by specialists who use morphometric measures estimated from the optical disc and physiological cup segmentations. Thus, the proposed deep learning method is composed of three stages: OD and PC segmentations, morphometric feature estimation and glaucoma detection, which are sequentially trained using a three-stages strategy.

This chapter presents a novel deep learning model for automatic analysis of eye fundus images to support glaucoma diagnosis. The proposed model is based on a three-stage deep learning model: the first stage is a DCNN used for automatic segmentation of optic disc and cup, the second stage is a DCNN for extraction of Morphometric Features (MFs), and the last stage is a multilayer perceptron neural network used for glaucoma diagnosis. The proposed method was assessed using three publicly available dataset and presents outstanding results when compared to state-of-the-art methods.

5.2 Three-stages model for Glaucoma Classification

Figure 5.1 shows the architecture of the deep neural network model for automatic analysis of eye fundus images to support glaucoma diagnosis. The model is organized in three consecutive stages that are sequentially trained using a three-stages approach, i.e. at each stage the training process focuses on different learning goals. This learning strategy regularizes the optimization process to converge faster, guiding the search towards better local minima [169]. The network stages were designed following a process analogous to the one followed by experts. The first stage performs the segmentation of the OD and PC using a 15-layer DCCN. The second stage uses as input the two segmentations generated by the first stage, stacking a third image mask corresponding to the

union of the OD and PC segmentations to create a 3D-binary mask, which are fed to a 12-layers DCNN. The goal of this stage is to calculate different morphometric features, which are generally used by experts to diagnose glaucoma. Finally, the third stage applies a multilayer neural network to produce the final prediction that classifies the input image into three possible classes: normal, suspicious or glaucoma. The following subsections discuss the details of the three stages.

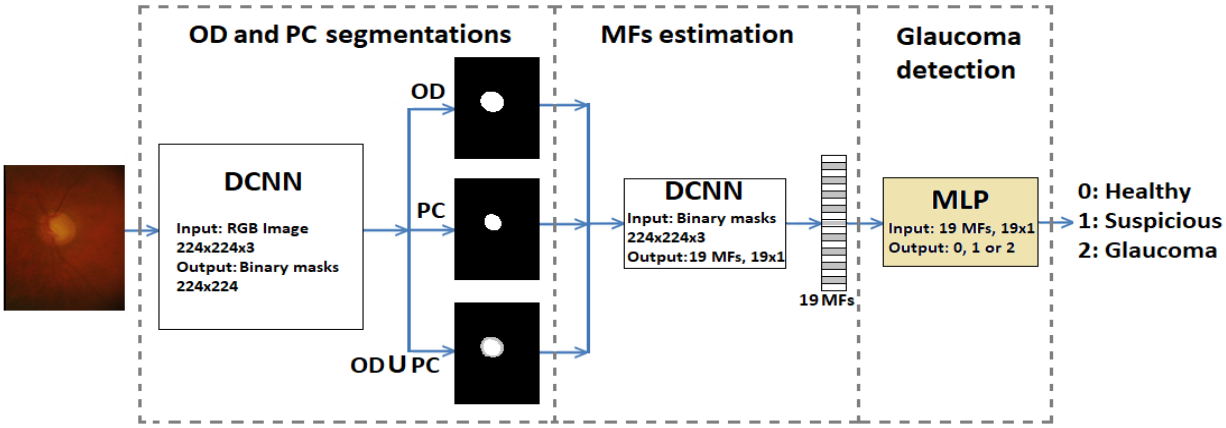

Figure 5.1: Block diagram used to segment binary masks (first block), to extract morphometric features (second block), and to classify into healthy, suspicious and glaucoma classes (third block).

DCNN for automatic segmentation of optic disc and physiological cup

The first stage of the model corresponds to a DCNN that uses as input an RGB eye fundus image to automatically segment the OD and the PC. The DCNN is based on the DRIU model with two additional convolutional layers as explained in detail in subsection 4.3.2.The DRIU model is initialized with VGG weights pretrained on ImageNet. It is fine-tuned for 10,000 epochs with a learning rate of $1e-6$, which is gradually decreased as the training process proceeds. A real time data augmentation strategy is implemented to grow the training data. Class weights of 0.1 and 0.9 for background and foreground respectively are used to handle the imbalance of the number of background pixels (nor disc neither cup pixels) compared to foreground pixels (disc and cup pixels). The model is trained using binary cross-entropy as loss function with ground-truth segmentations of the OD and the PC provided by experts.

DCNN for automatic morphometric estimation

The second stage takes as input the segmentations of the OD and the PC, along with a third image corresponding to the union, to calculate 19 morphometric features used by Perdomo et al. [15]. The 19 morphometric features can be divided into four subsets: geometric, distances, axis and ratio. The geometric subset contains areas and perimeters of OD and PC; the distance subset is composed of the superior, inferior, nasal and temporal distances between the OD and PC; the axis subset is defined by major and minor axis of the OD and the PC, finally, the ratio subset includes

the eccentricity of the OD and PC and the five ratios between upper OD and PC parameters to seek disproportions and relationships between optic disc and the physiologic cup [15].

The DCNN designed for automatic morphometric feature estimation is composed of five convolutional, five pooling layers and two fully connected layers. The four max-pooling and the global average pooling layers are non-linear size reducers that are applied to reduce the spatial dimensions, to minimize overfitting and the number of parameters in the model. Finally, the first fully-connected layer connects all the neurons obtained from the global average-pooling layer with 512 neurons to the next fully-connected layer with the number of morphometric features to predict. The model is trained using mean absolute error as the loss and, morphometric features directly predicted from the OD, PC and union segmentations as ground-truth.

Multilayer perceptron for glaucoma classification from morphometric features

The final stage corresponds to a multilayer perceptron (MLP) that receives as input the 19 morphometric features and classifies them into three possible classes: normal, suspicious or glaucoma. The MLP is composed of two fully-connected layers with 64 hidden and 3 output units. The batch size, number of epochs and initial learning rate used were explored using a grid search strategy, the best performing parameters found experimentally were 16, 500 and 0.01 respectively.

5.3 Experimental setup

5.3.1 Eye fundus image databases for glaucoma detection

The DRISHTI-GS1, RIM-ONE_v1 and RIM-ONE_v3 databases are used in this study [51, 85]. The DRISHTI-GS1 dataset has been acquired and labeled as healthy and glaucomatous by Aravind Eye Hospital (India), and it contains 101 color fundus images distributed in two subsets: 50 images for training, and the 51 remaining for the testing subset [170]. RIM-ONE_v1 and RIM-ONE_v3 focus on optic nerve head segmentation for glaucoma detection with manual reference segmentations as gold standard with 455 and 159 images respectively, created by ophthalmologists from the Department of Ophthalmology at the Hospital Universitario de Canarias in Spain [51]. The RIM-ONE_v1 dataset was labeled according to a binary classification (healthy vs. glaucomatous), and RIM-ONE_v3 was labeled as a 3-class classification problem (healthy, suspicious and glaucoma).

The proposed method was evaluated in 2 setups: binary-classification task (DRISHTI-GS and RIM-ONE_v1) and 3-class classification (RIM-ONE_v3). Additionally, the proposed method used a stratified sampling to randomly divide two RIM-ONE datasets into three subsets with 60%, 10% and 30% for each class that correspond to training, validation and test sets respectively.

5.3.2 Evaluation

The proposed method used several performance metrics in each stage. The OD and PC segmentations stage were assessed with the Jaccard index and the Dice coefficient. The MF estimation stage was evaluated using the Mean Average Percentage Error (MAPE) among all the predicted MFs. Moreover, the MFs calculated from the OD and PC segmentations by the experts called Real Morphometric Features (RMFs) were compared to the Estimated Morphometric Features (EMFs)

or the MFs calculated from the OD and PC segmentations by the first stage as reported in Table **5-1**. Finally, the complete proposed method for glaucoma detection was evaluated using accuracy, sensitivity, specificity, Area Under the Curve (AUC), precision, recall, f-score, Kappa coefficient, and Overall Accuracy (OA) performance metrics reported on the test sets.

For the binary task, a combination of an 8-layer CNN model and a linear SVM applied to RIM-ONE_v1 [167], and a fine-tuning of VGG pre-trained on ImageNet applied to DRISHTI-GS1 [143] were chosen as binary-classification baseline models, as reported for the test set in Table **5-2**. The deep learning data-fusion model that combines raw color fundus images and RMFs was chosen as the 3-class classification baseline [15] on the RIM-ONE_v3 dataset. Furthermore, a DCNN feeding with color fundus images and a 3D-binary mask described in Section 4.2.2 was compared to the proposed method. The proposed approach was implemented with Keras[1] using a GeForce GTX TITAN X from NVIDIA.

5.4 Results

The best performance of the proposed model for OD and PC segmentations (first stage) was obtained with a learning rate of $1e-6$, a batch size of 2, a number of samples per epoch of 300, and a number of epochs of 10.000. The Jaccard index (JI) and Dice coefficient (DC) were monitored during training in the first stage (not reported here), and these parameters were evaluated on the test set for OD segmentation with of $JI = 0.9975$ and $DC = 0.9987$ of, and PC segmentation with $JI = 0.9983$ and $DC = 0.9991$ respectively.

For the second stage, it was assessed the EMFs compared with RMFs, the EMFs presented a MAPE during training of 3.57%, and a MAPE in the test set of 6.30% compared to RMFs. Table **5-1** presents a comparison between Real Morphometric Features (RMFs) and Estimated Morphometric Features (EMFs) for 3-class classification on the RIM-ONE_v3 dataset using SVM and Random Forest (RF) classifiers as reported in [15].

Table **5-1**: Performance measures for the RMFs and EMFs on the test RIM-ONE_v3 dataset, bold values show the best score for each performance metric.

Method	Source	Precision	Recall	f-score	Kappa	OA
SVM [15]	RMFs	0.63	0.56	0.55	0.35	0.66
RF [15]	RMFs	**0.64**	0.57	0.58	**0.37**	**0.65**
SVM	EMFs	0.54	0.41	0.39	0.23	0.57
RF	EMFs	0.60	**0.64**	**0.61**	0.35	0.64

Finally, we evaluated the proposed method with baseline methods for binary and 3-class classifications (third stage) as reported in Tables **5-2** and **5-3** respectively. These tables present the comparison between the methods, information sources and performance metrics evaluated in classification tasks for the two experimental setups respectively.

[1] http://keras.io

Table **5-2**: Binary-classification performance metrics for baseline models and the proposed method on the test data set. The bold values show the best score for each performance metric.

Method	Source	Accuracy	Sensitivity	Specificity	AUC
Al-Bander et al. [167]	RGB	0.882	0.85	**0.908**	–
Orlando et al. [143]	RGB	–	–	–	0.763
Proposed method	RGB	**0.894**	**0.895**	0.889	**0.82**

Table **5-3**: Comparison of performance metrics for 3-class classification for baseline models and the proposed method on the test dataset. The bold values show the best score for each performance metric.

Method	Source	Precision	Recall	f-score	Kappa	OA
DCNN	RGB	0.48	0.55	0.51	0.20	0.55
DCNN	OD -PC-Union	0.60	0.60	0.59	0.29	0.60
Data fusion model [15]	RMF,RGB	0.46	0.56	0.50	0.42	0.68
Three-stages model	RGB	**0.76**	**0.72**	**0.69**	**0.48**	**0.70**

5.5 Discussion and conclusions

We present a novel method for automatic glaucoma assessment from eye fundus images based on DCNNs. The results show that the three-stages model for glaucoma classification is competitive with the best results reported for each dataset: RIM-ONE-v1 accuracy of 89.4% vs. 88.2% reported by [167], and DRISHTI-GS AUC of 0.82 vs 0.76 reported by [143], as shown in Table **5-2**.

The most remarkable characteristic of this model is its architecture and training strategy. The model is organized in stages that follow a conventional process for glaucoma diagnosis based on the calculus of morphometric features. The multistage architecture allows us to train the model using a three-stages approach, which gradually trains the model to accomplish subtasks with increasing complexity. This approach allows training a complex deep learning model with a reduced set of training samples, resulting in an improved performance of the model as corroborated in section 5.4.

In particular, the experimental results showed that the multistage architecture along with the sequential training, produces better results than conventional DCNNs, as reported in Table **5-3**. As the resulting model is a three-stage model, it is capable of directly producing a prediction from the input image without the need for the manual intermediate segmentation required by the conventional diagnostic protocol from images of the fundus.

The work shows that it is possible to involve domain knowledge in deep learning models. Additionally, intermediate results produced by the model (segmentations and morphometric features) can help the interpretability of the model predictions, making them more useful in support of the diagnosis process. We hypothesize that this approach can be extended to other medical image analysis applications and exploring this hypothesis will be the focus of our future work.

6 Deep learning for ocular diseases detection using OCT volumes

This chapter presents a new deep learning model, OCT-NET, which is a customized convolutional neural network for processing scans extracted from optical coherence tomography volumes. OCT-NET is applied to the classification of three conditions seen in SD-OCT volumes. Additionally, the proposed model includes a feedback stage that highlights the areas of the scans to support the interpretation of the results. This information is potentially useful for a medical specialist while assessing the prediction produced by the model. The architecture of OCT-NET was reported and published in the *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* [14]. In addition, the OCT-NET model was tested on two public data sets containing different ocular conditions and producing interpretable clinical information in the form of highlighting of the regions of the image that most contribute to the classifier decision. This results was published in the Journal of *Computer Methods and Programs in Biomedicine Volume 178, September 2019* [10].

6.1 Introduction

Ophthalmic diseases related to Diabetes Mellitus are characterized by a vascular permeability of retinal vessels with fluid accumulating in retinal layers [171]. Diabetic retinopathy and diabetic macular edema are two non-exclusive complications that affect the visual field [172]. The diagnosis of these complications is not an easy task, since edema can occur in subjects with and without DM at any stage of DR, with similar symptoms but with different treatment strategies and associated costs [173]. Age-related Macular Degeneration is linked to macular changes derived from non-modifiable and modifiable risk factors. The diagnosis is based on typical changes related to aging and visual loss and prognosis is related to the severity of the either geographic atrophy or choroidal neovascular membrane [174].

The Spectral Domain Optical Coherence Tomography (SD-OCT) is a widely accepted noninvasive imaging approach that contains images of the depth of the retina through a set of B-scan (2D images) used to detect abnormalities among the ten retinal layers with an accurate diagnosis of retinal disorders [175]. A typical ophthalmological examination of the retina may include an analysis of eye fundus images and in some cases SD-OCT to locate retinal vascular damage and changes in choroidal thickness [42]. The DR and DME diagnoses are performed by looking for the presence of microaneurysms, intraretinal hemorrhages, exudates and edema [176–178]. The evaluation of the thickness of the neurosensory retina, retinal pigment epithelium, and choroid are analyzed independently for the AMD diagnosis [179, 180].

Automatic image analysis methods based on machine learning have shown to be a valuable tool to support medical decision making [181, 182]. In particular, deep neural network methods have

been explored in several medical domains exhibiting promising results. The results of deep neural networks include: the detection of red lesions in fundus images [183], prediction of breast, lung and stomach cancers using RNA-sequence data [184], early diagnosis of Alzheimer's disease using CT brain images [185] and the recognition of emotions using multimodal physiological signals [186].

Deep learning methods applied to SD-OCT presented outstanding results in automatic segmentation and disease classification tasks. For a segmentation task, the state of the art presents an overall Dice coefficient (mean of all tissues) ranging between 0.90 and 0.95 using known architectures such as VGG [121], U-Net [187–190] or DenseNet [191, 192]. The classification of SD-OCT volumes has mainly focused on two approaches: (1) the manual or automatic feature extraction combined with ensemble classifiers, and (2) the use of end-to-end deep learning models.

This paper presents a deep learning-based method with a feedback stage for automatic classification of B-scans inside a volume for three retinal diseases. The method is able to automatically identify visual patterns associated with several pathologies and use them to make accurate predictions. The model has the ability to highlight the patterns in the input image, allowing the expert to better understand the model prediction. The remainder of this article is organized as follows: in Section 6.2 the main work for retinal disease classification using SD-OCT volumes is summarized. Then, the volume preprocessing and the convolutional neural network-based model architecture are presented in Section 6.3. The data sets and baseline models used are described in Section 6.4. The experimental results are reported in Section 6.5. Finally, Section 6.6 discusses the outcomes and finishes with conclusions.

6.2 Related work

The end-to-end OCT-NET model was tested on a data set that contains 32 SD-OCT volumes with healthy and DME patients commonly known as the SERI (Singapore Eye Research Institute) data set as explained in details in subsection 6.4.1. In this previous work, the OCT-NET model obtained an outstanding performance using a leave-one-patient-out evaluation methodology with an accuracy of $93.75 \pm 3.125\%$ and a sensitivity and a specificity of 93.75% [14]. This paper presents an extended version that addresses three main challenges: (1) the qualitative evaluation of B-scans to highlight medical findings using a visualization stage; (2) the quantitative evaluation of SD-OCT volumes with three retinal diseases from two OCT scanners, and (3) the medical feedback of quantitative and qualitative evaluations to validate the usefulness of the methodology.

The main work reported on the SERI data set is characterized by using deep learning architectures pre-trained on ImageNet[1] combined with ensemble classifiers. First, Awais et al. [193] presented a method that used block-matching and 3D filtering (BM3D) for removing the speckle noise in SD-OCT. The new filtered volumes fed a pre-trained VGG-16 with a k-Nearest Neighbors (kNN) algorithm to classify features from the three dense layers with an accuracy, sensitivity, and specificity of 93%, 87% and 100% respectively. In a similar way, Chan et al. [194] designed a method that applies a BM3D filter and saturation removal. The processed volumes are then used as input of three pre-trained architectures known as Alexnet, VGG and GoogleNet. The last convolutional layers of each model are fused and a feature space reduction is performed using Principal

[1] http://www.image-net.org/

Component Analysis. The volumes are classified using a Support Vector Machine with a precision, sensitivity and specificity of 93.75%. Finally, Kamble et al. [18] proposed the fusion of residual connections with an inception architecture termed as inception-ResNet-v2. This model used as an input filtered volumes with a BM3D filter stage and presented a performance of 100% in accuracy, specificity and sensitivity.

The most representative work classifying SD-OCT volumes on the A2A SD-OCT data set is mainly reported in three papers [118, 195, 196]. Sun et al. [118] manually cropped patches based on the annotation of interest points to calculate a Histograms of Oriented Gradient (HOG) and merged them as the training set. Then, Principal Component Analysis (PCA) was performed for reducing the length of the HOG features. Finally, a multiple instance SVM classifier was trained with the PCA-transformed patch representation and tested to classify volumes on the test data set obtaining an accuracy, sensitivity and specificity of 94.4%, 96.8% and 92.1% respectively.

Venhuizen et al. [195] developed an unsupervised clustering stage to extract interest points in 31 B-scans per volume centered at the fovea of 284 SD-OCT volumes as the training data set. The number of 9×9 patches is reduced by the application of a uniform subsampling by a factor of 8. Then, the patches are normalized to zero mean and unit variance before the extraction of 9 principal components through PCA. A bag of words is created using k-means clustering with an experimental value of $k = 100$ on the total set of PCA-transformed patches. Finally, the unsupervised clustering is combined with a supervised training stage that uses a random forest classifier with a number of trees set to 100 trained to differentiate healthy subjects from AMD subjects. The performance on the test data set after classifying 50 AMD and 50 healthy subjects was an AUC of 0.984.

Chakravarty et al. [196] designed a two-stage retinal atlas for macular SD-OCT volumes that comprises a pre-processing and a classification stages. First, the pre-processing step resized the images to a pixel dimensions of 3.6 μm by 8.6 μm. A denoising and intensity standardization is applied to reduce the speckle noise in the volumes. Moreover, a retinal curvature flattening of the SD-OCT volumes is performed, where each B-scan is flattened and aligned across the volume. Finally, the Region of Interest (RoI) by SD-OCT volume is defined to a set of 31 B-scans with the 108 axial scans centered at the macula, where the histogram is calculated and concatenated across all the RoIs. The binary classification stage is done using a linear SVM presenting an accuracy and an AUC of 98% and 0.996 respectively for AMD classification.

The state of the art for classifying real-world scans is mainly focused in three deep learning approaches. De Fauw et al. [122] trained with two private data sets a two-stage deep learning-based pipeline: a deep segmentation network with a three-dimensional U-Net architecture and a deep classification network to predict the diagnosis probability and the referral suggestions using the segmentation. The deep segmentation network was trained using 877 SD-OCT volumes acquired by Moorfields Eye Hospital with a Topcon 3D OCT device, where the three most representative slices were manually segmented in a detailed tissue-segmentation map with 15 classes including anatomy, pathology and image artifacts. The classification network was trained using the 43 most representative slices of $14,884$ SD-OCT volumes obtained from $7,621$ patients referred by the experts as subjects with symptoms suggestive of macular pathology. The model was tested with 997 patients obtaining an area under the ROC curve of 99.21%. Lee et al. [121] used a VGG-16 convolutional neural network applied on a private data set for the classification of normal and AMD. The deep learning model receives as an input a scan with a size of 192×124 and

performed 13 convolutional layers with an incremental number of filters and 3 dense layer to classify the two classes. The image database was acquired using a Heidelberg Spectralis OCT device with 80,839 images in the training set and the test set contains 20,163 images. The results at image level presented an area under the ROC curve of 92.78% with an accuracy of 87.63%, a sensitivity of 84.63% and a specificity of 91.54%. Additionally, an occlusion test identified the RoIs with the areas contributing most to the deep neural network's probability. Finally, Kermany et al. [93] used a pretrained Inception V3 from ImageNet to predict four classes: Normal, Choroidal Neovascularization (CNV), DME and Drusen. The method was trained using a public data set acquired by Heidelberg Spectralis OCT device with 108,312 images for training and tested in 1,000 with 250 per class. The best results on the test set presented an accuracy of 96.6%, a sensitivity of 97.8%, and a specificity of 97.4%. In addition, a sliding window of 20×20 was systematically moved across 491 images to record the probabilities of the disease.

Although previous work reported very good results in the classification task, the performance of these methods is crucially dependent of the manual extraction of RoIs and in some case limiting the number of scans from the SD-OCT volumes. The proposed model provides the highlighted areas in all scans into volumes with a validation performed by two ophthalmology experts. Distinctively from previous work, our approach automatically classifies AMD but also produces useful medical information at qualitative and quantitative levels inside an SD-OCT volume to support medical decision making in the diagnosis of AMD.

6.3 Methods

This section presents the details of the SD-OCT classification model based on deep neural networks and more specifically OCT-NET. The method comprises six stages as shown in Figure 6.1 and it is available in a repository of Github[2]. The first stage (1) receives a raw SD-OCT volume with speckle noise that hinders layers and abnormalities among the layers as an input. Then, the volume preprocessing stage (2) makes the detection of the Internal Limiting Membrane (ILM) and the Retinal Pigment Epithelium (RPE) layers, to resize the volumes in order to crop the relevant raw pixels into the volumes as presented in subsection 6.3.1. Furthermore, the OCT-NET CNN (3) performs the feature extraction and classification of each B-scan from a SD-OCT volume as healthy or non-healthy. Simultaneously, the disease classification stage (4) based on the number of images calculates with a majority rule the prediction for the SD-OCT volume as explained in detail in subsection 6.3.2. Then, the Class Activation Map (CAM) visualization stage (5) allows to highlight the relevant zones of the scans used by the OCT-NET model to classify a specific retinal disease as reported in subsection 6.3.3. Finally, the expert feedback stage (6) evaluates the provided information in the disease classification and the CAM visualization stages to qualitatively validate the obtained results.

6.3.1 SD-OCT volume preprocessing

A spectral domain optical coherence tomography is a volumetric array $V(n, a, b)$ that can be defined as a set n of 2D-images called B-scans or cross-sectional scans $I \in IR^{a \times b}$, with a corresponding

[2]https://github.com/Ojperdomoc/OCT-NET.git

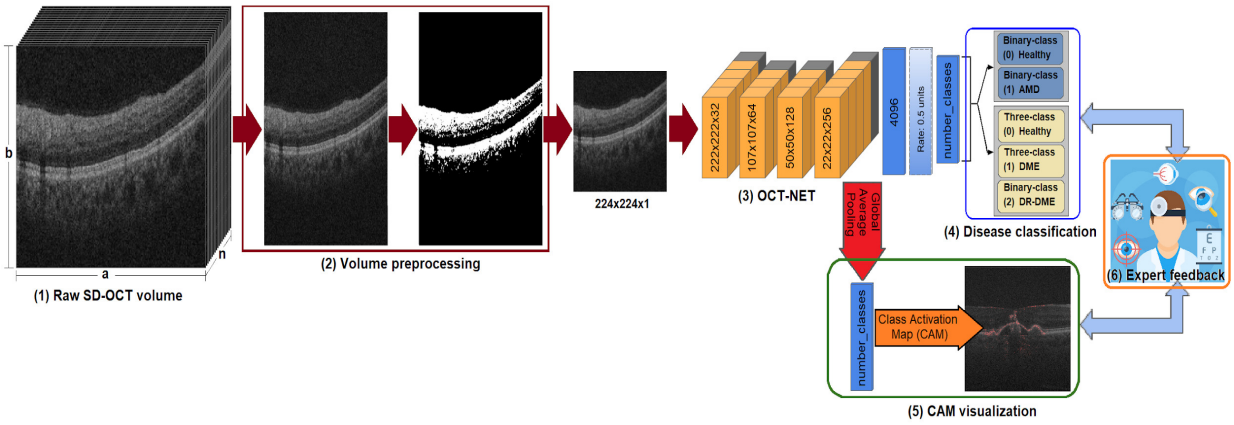

Figure 6.1: Overview of the six-stage proposed deep learning model for retinal disease classification. (1) The raw SD-OCT volume for a database; (2) The volume preprocessing stage to resize and crop to an input size; (3) the OCT-NET model to extract features from all the scans in a volume; and the interpretability three-stage (4,5,6) to classify and evaluate qualitative and quantitative the scans and the SD-OCT volumes.

label $l \subset$ `Healthy, DME, DR-DME and AMD`. The input for the customized OCT-NET was set for scans with size of $224 \times 224 \times 1$ as described in subsection 6.3.2. Therefore, a set of transformations are needed for automatically extracting a RoI per scan in the SD-OCT volume.

A median filter was applied to SD-OCT scans with a threshold to differentiate speckle noise from retinal layers as reported in [14, 197]. First, the RoI was automatically detected using a median filter with a kernel size of 3×3 and a threshold of 0.5 to highlight the top layer (ILM) and bottom layer (RPE) in the volume as shown in stage (2) of Figure 6.1.

Each image cropped without scaling in such a way that the resulting image fully contains the RoI. This is independently done for each B-scan hence that the process is not affected by a pronounced tilt. Finally, these cropped images are resized keeping aspect ratio to ensure the relevant information in a volume dimension of $V_{input} \in IR^{\,n \times 224 \times 224}$.

6.3.2 OCT-NET model

OCT-NET is a customized Convolutional Neural Network (CNN) inspired by the VGG model reported by Simonyan and Zisserman [97]. The proposed model is based on the combination of convolutional and max-pooling layers in four sub-blocks that are responsible for the feature extraction in the CNN and the remaining layers conforming the classification sub-block to classify a scan from an OCT-volume. In summary, the OCT-NET model contains 10 convolutional layers, 4 max-pooling layer, 2 fully connected layers, and 1 dropout layer as shown in detail in Table **6-1**.

The input layer receives an image with a size of $224 \times 224 \times 1$ as reported in subsection 6.3.1. The number of filters of the convolutional layers in the four sub-blocks is inspired by the VGG model [97], with the difference that OCT-NET has a number of filters f_n defined by an arithmetic

Table **6-1**: Structure of OCT-NET with the parameter layer, output shape and trainable parameters of each layer.

N	Layer	Output shape	Number of parameters
0	Input	$224 \times 224 \times 1$	0
1	Conv2D (kernel size = 3×3)	$222 \times 222 \times 32$	320
2	Conv2D (kernel size = 3×3)	$220 \times 220 \times 32$	9248
3	Conv2D (kernel size = 3×3)	$218 \times 218 \times 32$	9248
4	MaxPooling2D (pool size = 2×2)	$109 \times 109 \times 32$	0
5	Conv2D (kernel size = 3×3)	$107 \times 107 \times 64$	18496
6	Conv2D (kernel size = 3×3)	$105 \times 105 \times 64$	36928
7	MaxPooling2D (pool size = 2×2)	$52 \times 52 \times 64$	0
8	Conv2D (kernel size = 3×3)	$50 \times 50 \times 128$	73856
9	Conv2D (kernel size = 3×3)	$48 \times 48 \times 128$	147584
10	MaxPooling2D (pool size = 2×2)	$24 \times 24 \times 128$	0
11	Conv2D (kernel size = 3×3)	$22 \times 22 \times 256$	295168
12	Conv2D (kernel size = 3×3)	$20 \times 20 \times 256$	590080
13	Conv2D (kernel size = 3×3)	$18 \times 18 \times 256$	590080
14	MaxPooling2D (pool size = 2×2)	$9 \times 9 \times 256$	0
15	Dense	4096	84938752
16	Dropout (rate = 0.5)	4096	0
17	Dense	number of classes	8194

series, as described in Eq. (6-1) as follows:

$$f_n = f_0 + 32 * (n - 1) \tag{6-1}$$

where the parameter $f_0 = 32$ and n is the number of sub-blocks with $1 \leq n \leq 4$. The cascading of four blocks of convolutional and max-pooling layers provides a translation invariance and a reduction of dimensionality: by applying a set of f_n learned filters with kernel size of 3×3 and stride of 1×1 and eliminating non-maximal values with pool size of 2×2 and stride of 2×2.

The classification sub-block is composed of three layers: one fully-connected layer with 4096 neurons, one dropout layer with a fraction of deactivation of units during training of 0.5, and a final fully-connected layer with *number of classes* as the number of neurons. The dropout layer allows to learn with different neurons the same information improving the generalization of the model and the number of neurons for the final fully-connected layer is set to 2 or 3 for binary and three-class data sets respectively. Additionally, the disease classification of one SD-OCT volume was determined using a majority rule, as such the volume was affected by the class that was the most preponderant among the B-scans as reported in [193].

6.3.3 Class activation map

The Class Activation Map (CAM) is defined as the sum of the weighted activation maps generated for each image at different spatial locations. The main use of a CAM focuses on the validation of a CNN model that indicates the discriminative image regions for a particular category. Thus, the CAM block adds a Global Average Pooling (GAP) after the last convolutional layer in the CNN model for obtaining an accurate discriminative localization as reported by Bolei et al. [198] and Selvaraju et al. [199]. The CAM highlights the magnitude of the activation at the spatial grid (x, y) to classify an image to class c. The CAM for class c is defined by w_k^c as the weight corresponding to class c for unit k applied to an input image $f_k(x, y)$ described in Equation 6-2 as follows [198]:

$$M_c(x, y) = \sum_k w_k^c f_k(x, y) \tag{6-2}$$

6.4 Experimental evaluation

6.4.1 SERI+CUHK dataset

The Singapore Eye Research Institute (SERI) database contains 32 SD-OCT volumes with 16 control and 16 DME SD-OCT volumes. Similarly, the Chinese University of Hong Kong (CUHK) database contains 43 SD-OCT volumes with 4 DME and 39 DR-DME SD-OCT volumes. The two data sets were combined into one three-class data set termed in this paper as SERI+CUHK data set. The SERI-CUHK data set was acquired with a CIRRUS SD-OCT device [3] and labeled by certified expert graders as control, DME and DR-DME volumes, according to findings among the retinal layers as shown in Fig 6.2.

The inclusion criterion was the presence of abnormal retinal thickening, hard exudates, intraretinal cystoid space formation and subretinal fluid among the retinal layers of working-age adult subjects. Finally, each SD-OCT volume contains 128 cross-sectional scans with a resolution of $512 \times 1,024$ pixels. The data set was cropped and resized (keeping the aspect ratio) to a dimension of $128 \times 224 \times 224$ as discussed in subsection 6.3.1.

6.4.2 A2A SD-OCT data set

The A2A SD-OCT is a binary data set from the Age-Related Eye Disease Study 2 (AREDS2) also known as Duke data set [55]. The images from the A2A SD-OCT study obtained the informed consent from all subjects and it was approved by the institutional review boards of the 4 A2A SD-OCT clinics: Devers Eye Institute, Duke Eye Center, Emory Eye Center, and National Eye Institute.

The Duke data set was acquired using imaging systems from Bioptigen, Inc (Research Triangle Park, NC) as shown in Fig. 6.2. The classification of each volume was done by certified SD-OCT readers. The inclusion criteria were defined as subjects between 50 and 85 years of age, exhibiting intermediate AMD with large drusen ($> 125mm$) in both eyes or large drusen in one eligible eye and advanced AMD in the fellow eye, with no history of vitreoretinal surgery or ophthalmic surgery. The Duke data set contains 384 SD-OCT volumes: 269 AMD and 115 control or normal eyes, with

[3] https://www.zeiss.com/meditec/int/products/oct-optical-coherence-tomography.html

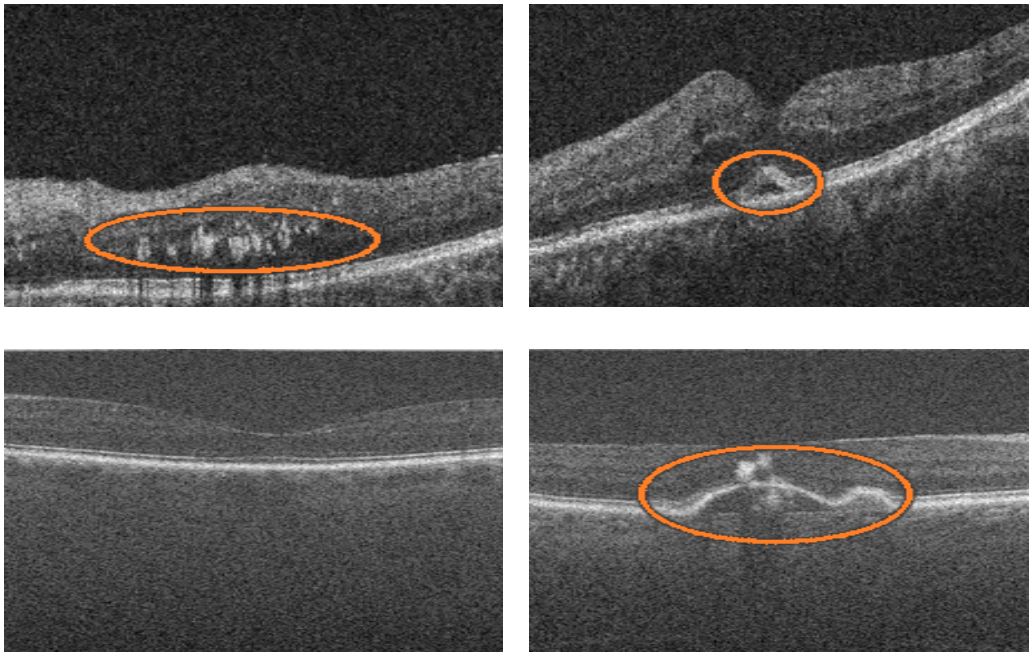

Figure 6.2: [Top-left] DME SD-OCT scan with hyper-reflective material in the middle layers of the retina, likely by exudates, and [Top-right] DR-DME SD-OCT scan with a retinal pigment epithelium detachment.[Bottom-left] Normal SD-OCT scan and [Bottom-right] AMD SD-OCT scan with a drusenoid detachment with migration of pigment to the inner layers of the retina.

100 B-scans per volume and a resolution of $1,000 \times 512$. The data set was resized to a volume dimension of $100 \times 224 \times 224$ as reported in subsection 6.3.1.

6.4.3 Experimental setup

The OCT-NET model was trained with random initialization weights using the Adam optimizer. The batch size, learning rate and number of epochs were experimentally set to 16, $1e-5$ and 10 respectively for all experiments as reported in a previous article [14]. Moreover, the classification of one SD-OCT volume was determined using a majority rule as explained in subsection 6.3.2.

The SERI+CUHK data set was randomly split in a stratified way into three independent data sets with 60%, 10% and 30% for training, validation and testing respectively as presented in Table **6-2**. On the other hand, the Duke data set was randomly split into 3 independent data sets with 67% and 10% for training and validation respectively. The remaining 23% corresponds to the test data set, with 50 AMD and 50 Healthy volumes as reported in Table **6-2**.

6.4.4 Baseline models and performance metrics

The work proposed by Venhuizen et al. [195], Chakravarty et al. [196] and Kermany et al. [93] were chosen as baseline models applied on the Duke data set as explained in Section 6.2. In

Table **6-2**: Retinal disease data sets used for training, validation and testing in the experimental evaluation.

Data set	Training	Validation	Test
SERI+CUHK	45 SD-OCT	8 SD-OCT	22 SD-OCT
Duke	246 SD-OCT	38 SD-OCT	100 SD-OCT

addition, the methods reported by Awais et al. [193] and Kermany et al. [93] were chosen as baseline models for the SERI-CUHK data set. Additionally, we performed a qualitative evaluation for the interpretability stage according to the ability of the proposed model to highlight medical findings in the scans. In this test, 40 SD-OCT volumes from the Duke test data set were randomly split with 20 healthy and 20 AMD samples. Two retina specialists from Fundación Oftalmológica Nacional manually labeled each B-scan of this subset without taking into account the given volume label. Finally, the two experts assessed the generated CAM visualizations plus the individual prediction from each scan individually from a SD-OCT volume.

The proposed model was implemented with Keras using a theano backend on a GeForce GTX TITAN X from NVIDIA. The loss and accuracy metrics were monitored on the training and validation data, and the best performance in the validation set was assessed on the test data set presented in Tables **6-3** and **6-4**. OCT-NET was evaluated on the test set of the Duke data using accuracy, sensitivity, specificity as performance metrics. In addition, the AUC was calculated according to the probability that our classifier will rank a randomly selected positive case higher than a randomly chosen negative case.

OCT-NET was evaluated on the SERI+CUHK test data set using precision, f-score (macro) and Kappa coefficient as multi-class performance metrics. Recall was another performance metric evaluated on the SERI+CUHK test data set. These performance measures were chosen so that the results can be compared with those reported by the state of the art.

6.5 Results

6.5.1 Volume classification performance

The performance classification was reported for the Duke and the SERI+CUHK databases explained in detail in Section 6.4. Moreover, we tested the performance metrics of the OCT-NET model using the hyper-parameters and monitoring the loss and the accuracy during training as shown in Figure 6.3. The training was set to 10 epochs as it obtained high accuracy while limiting the training time that is on average to 60 minutes per epoch. The computational time to evaluate a single B-scan of the test set from Duke was 0.33 ms, and 0.28 ms in SERI+CUHK.

For the Duke database, the performance of OCT-NET applied on the test data set is reported in Table **6-3**. Additionally, the performance of the proposed method was compared with the main related works using this database as presented in subsection 6.4.3. The OCT-NET was modified into three architectures with different values in Dropout (DO) layer and the last dense (D2) layer.

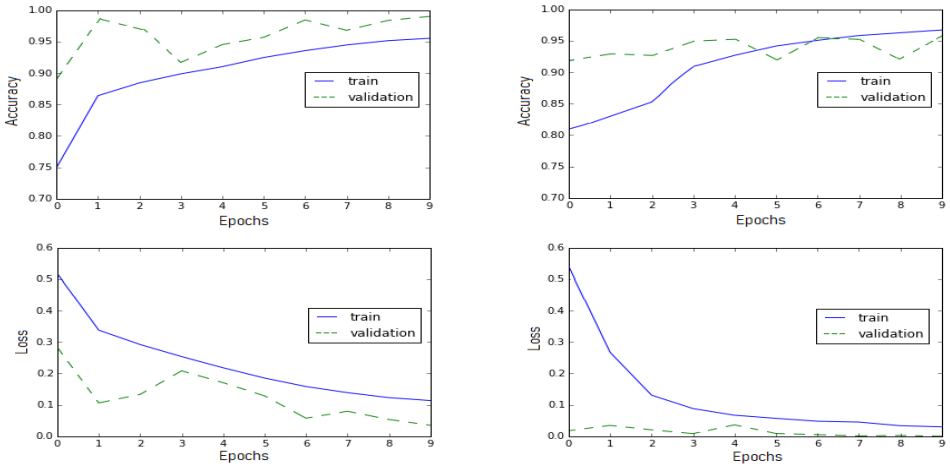

Figure 6.3: The monitoring of the accuracy and loss during training for the two experiments: [Left] SERI+CUHK, and [Right] Duke data sets. The blue solid and the green dashed lines represent the training and validation sets respectively.

The OCT-NET presented a similar AUC metric that method proposed by Chakravarty et al. [196], but outperforms baseline methods in sensitivity, specificity and accuracy as reported in Table **6-3**.

Table **6-3**: Performance measures of the baseline methods and the proposed method on the test data (Duke), bold values show the best score for each architecture.

Model	Sensitivity	Specificity	Accuracy	AUC
Venhuizen et al. [195]	0.96	0.92	0.94	0.984
Chakravarty et al.[196]	0.97	0.98	0.98	**0.99**
Kermany et al.[93]	0.98	0.89	0.94	0.94
OCT-NET	**0.99**	**0.99**	**0.99**	**0.99**
OCT-NET with DO=0.25	0.89	0.89	0.89	0.89
OCT-NET without DO	0.90	0.88	0.88	0.88
OCT-NET with D2=2048	0.95	0.95	0.95	0.95

The precision, recall, f-score Kappa coefficient and AUC were calculated to assess the performance of the proposed model applied to the SERI+CUHK data set, as reported in Table **6-4**. For the SERI+CUHK data set, the best performance of the proposed model on the test data is reported in Table **6-4**. Furthermore, we compared the performance of the baseline model reported by [93] and [193] using the three output dense layers (D1, D2 and D3) with three different ensemble classifiers: Decision Trees (DT), and KNNs with $K = 1$ and $K = 3$ applied on the test set.

The OCT-NET architecture presented the best performance and it outperforms baseline methods in precision, recall, f-score and Kappa coefficient as shown in Table **6-4**.

Table **6-4**: Performance metrics of OCT-NET on the test data (SERI+CUHK).

Model	Precision	Recall	f-score (macro)	Kappa coefficient	AUC
D1 with DT (depth=100) [193]	0.69	0.70	0.69	0.42	0.5
D2 with KNN (K=1)[193]	0.62	0.65	0.63	0.27	0.5
D3 with KNN (K=3)[193]	0.62	0.65	0.63	0.27	0.57
Kermany et al.[93]	0.91	0.78	0.74	0.59	**0.86**
OCT-NET	**0.93**	**0.83**	**0.85**	**0.71**	**0.86**

6.5.2 Qualitative analysis of CAM

The CAM visualization stage for the proposed model was validated according to the ability of locating medical findings that allow to highlight different retinal disorders as reported in subsection 6.3.3. The CAM output of the proposed model for the AMD class was highlighted in red with the corresponding medical findings of the ophthalmologist outlined in green as shown in Figure 6.4. Besides this, the ability of the proposed model to predict the condition of individual scans belonging to an AMD SD-OCT volume, compared with the diagnosis performed by an ophthalmology expert is presented in Figure 6.5.

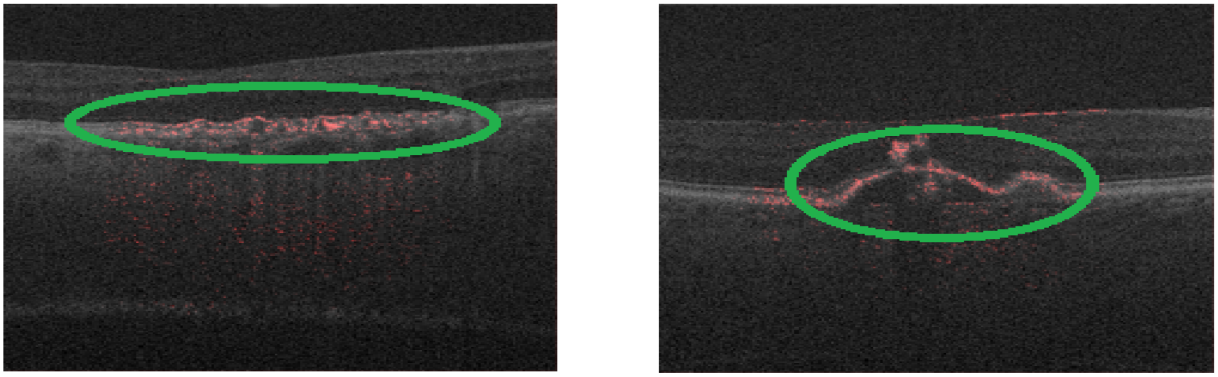

Figure 6.4: SD-OCT scans for subjects with AMD. [left] three large lesions on the outer layers and, [right] a drusenoid detachment with migration of pigment to the inner layers of the retina.

6.5.3 Individual B-scan classification

The SD-OCT data sets are commonly labeled to a volume level despite the retinal disease is present in a range of B-scans. This challenge motivated the evaluation of the proposed model to detect healthy and non-healthy scans regardless of the global labels of the volumes. Thus, the proposed method was validated with a subset of 4000 labeled B-scans annotated by the experts from Fundación Oftalmológica Nacional as presented in subsection 6.4.3. Table **6-5** presents the confusion

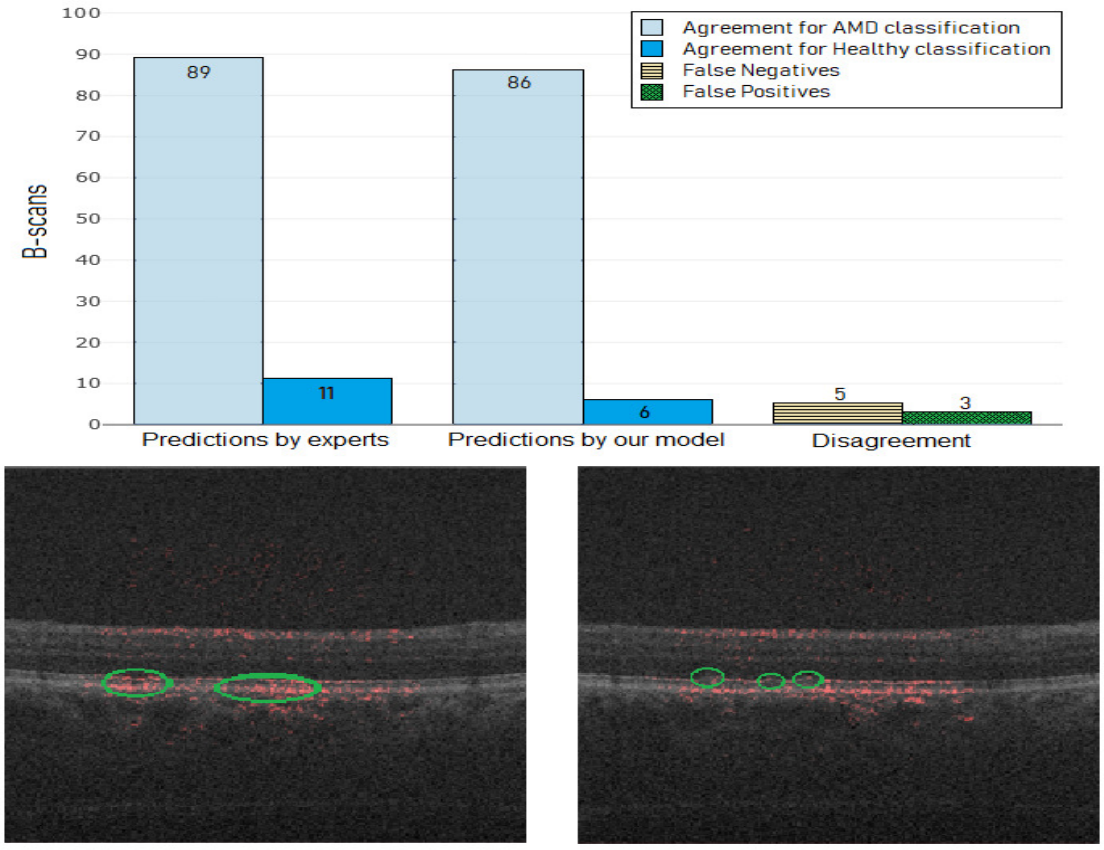

Figure 6.5: [Top] Classification of B-scans from an SD-OCT volume of an AMD subject by an ophthalmologist and predictions by the proposed model. [Left] False positive or a misclassified B-scan as AMD class due to an RPE layer presenting hyperreflectivity in these areas (green circles). This results in the proposed method incorrectly highlighting the areas like drusen or an RPE elevation. [Right] False negative or a misclassified B-scan as healthy class due to the RPE elevation not having enough hyperreflectivity (green circles); probably the proposed method is not able to detect these tiny drusen.

matrix for the 40 SD-OCT volumes from the Duke test data set as explained in Section 6.4. The major diagonal is equivalent to the agreements in the classification of the two classes. Otherwise, the sub-diagonal represents the erroneous classification of the proposed model. The overall accuracy in the prediction of the two classes was of 89% with a precision of 93% in the detection of AMD scans.

6.6 Discussion and conclusion

OCT-NET outperforms the state of the art methods for AMD diagnosis reported in [195, 196] in sensitivity, specificity, accuracy but it presents a similar AUC compared to the model proposed by Chakravarty et al. [196] as shown in Table **6-3**. The main two advantages of the proposed model

Table **6-5**: Confusion matrix describing the agreement in the predictions for healthy and AMD classes with true negatives (TN) and true positives (TP) respectively. The disagreement between the two classes is measured with false positives (FP) and false negatives (FN).

GroundTruth	Prediction by model: **Healthy**	**AMD**
Healthy	TN=2001	FP=332
AMD	FN=125	TP=1542

compared to the two-stage method [195] and the retinal atlas [196] are the automatic classification of raw scans without manual annotation of interest points or regions and the generation of qualitative and quantitative information to support medical decision making in a diagnosis of AMD as presented in Figure 6.4.

The experimental results of OCT-NET on the SERI+CUHK data set overcome the performance of the approach presented by Kermany et al. [93] in precision, recall, f-score and Kappa coefficient as reported in Table **6-4**. The proposed method shows an outstanding performance compared to the Inception-v3 pretrained with weights from ImageNet [93] without requiring a large database for training or selecting a limited numbers of scans with the condition by a patient. Finally, the Kappa coefficient or inter-rater agreement presented a substantial level of agreement of 0.71 between the model and the expert for the classification of healthy, DME and DR-DME SD-OCT volumes as reported in Section 6.5.

The global label of an SD-OCT volume is used without questioning local labels for each scan or the specific range of scans that contain the retinal disorder. We evaluated the prediction of B-scans belonging to an SD-OCT volume inspired in the manual classification performed by an ophthalmology expert as presented in subsection 6.5.3. The experimental results shown an agreement of 92.5% for AMD and 85.8% for healthy scans of a total of 4000 scans compared to the manually labeled scans, as reported in Table **6-5**. In addition, the range of scans with the retinal disorder in the volume, and the highlighted areas by CAM stage present a strong agreement with the delineations of the ophthalmologists as shown in subsection 6.5.2. This suggests that the information provided by the model could potentially be useful to deal with the lack of interpretability in deep learning models applied to medical images.

Despite the very good results for the SD-OCT volume AMD classification as reported in subsection 6.5.1, the evaluation of the qualitative analysis of CAM and the individual B-scan classification provided useful feedback about the medical findings in scans classified as false positives and false negatives as presented in subsection 6.5.2 and 6.5.3 respectively. The false positives were misclassified mainly in non-centered scans or poor resolution among the scans, which means that the layers are not defined in some scans inside the volume as shown in the scan [B] of Figure 6.5. On the other hand, false negatives could be due to the presence of subtle findings in some images. We hypothesized that the tiny drusen may be misleading the proposed method to classify these images as healthy as presented in scan [C] of Figure 6.5.

The speckle noise in images from medical devices was different among the SD-OCT volumes.

The OCT-NET model was trained with random weights presenting a better model generalization in classification tasks without being affected by the speckle noise. Finally, we want to emphasize that our approach was assessed with SD-OCT volumes acquired from different devices, with a different populations and several retinal diseases. However, these datasets are relatively small and our study lacks an evaluation over larger datasets, this will be part of our future work.

7 Deep learning for predicting neurosensory retinal thickness map using OCT volumes

Age-related macular degeneration is a common cause of vision loss in people aging 55 and older. The condition affects the light-sensing cells in the macula limiting the sharp and central vision. On the other hand, Spectral Domain Optical Coherence Tomography allow highlighting abnormalities and thickness in the retinal layers which are useful for age-related macular degeneration diagnosis and follow up. The Neurosensory Retina (NSR) map is defined as the thickness between the inner limiting membrane layer and the inner aspect of the retinal pigment epithelium complex. Additionally, the NSR map has been used to differentiate between healthy and subjects with macular problems, but the plotting of the retinal thickness map depends critically on additional manufacturer interpretation software to automatically drawing. Therefore, this paper presents a deep learning model based on a 3D convolutional neural network to automatically extract nine thickness mean values to draw the NSR map from an SD-OCT. The results of these method were published in the *14th International Symposium on Medical Information Processing and Analysis* [12].

7.1 Introduction

The eye normal condition can be evaluated with different non-invasive clinical exams. Spectral Domain Optical Coherence Tomography has become one of the most popular non-invasive exams for allowing qualitative and quantitative measurement of the macula [200]. The main advantage of SD-OCT technique lies in the accuracy to visualize the structure among all retinal layers. Therefore, ophthalmologists use SD-OCT to record and analyze the follow-up of diseases over time [201].

Age-related macular degeneration is the first cause of vision loss and irreversible blindness in developed countries [202]. Although AMD cannot be cured, the progression can be slowed down by an early detection and a proper treatment. The SD-OCT volumes have been widely recorded to locate retinal and subretinal changes in the first stage of AMD. [202, 203] Moreover, an SD-OCT volume can be represented into a 2D-image NSR map that presents thickness values in the fovea region and distances of $1mm$, $3mm$, and $6mm$ from the macular region [203].

The development of algorithms involving SD-OCT have been focused on two main tasks: ocular disease classification and nine-retinal layers segmentation. For disease classification, Lee et al. presented a model based on the VGG16 convolutional neural network that classifies OCT macula scans as normal or AMD scan [121]. Additionally, Rasti et al. devised a method based on on multiple-scale sub-images ensemble model to identify normal retina, dry age-related macular degeneration, and diabetic macular edema [92]. In contrast, Arabi at al. classified OCT images by automatically extracting and counting the white pixels in retinal pigmented epithelium layer of the images using a decision rule with a threshold value to classify the number of pixels [204]. Finally,

Perdomo et al. designed a deep learning model composed of a 12 layer CNN termed OCT-NET to classify an OCT volume using 2D-scans with a majority rule [14]. In retinal semgmentation task, Fang et al. combined CNN to extract features of all retinal layers with graph search methods, using probabilities maps to estimate the boundaries of the nine layers on SD-OCT images [205]. On the other hand, Roy et al. proposed an end-to-end CNN architecture based on encoders and decoders for semantic segmentation called RelayNet to highlight the retinal layers and fluid masses in OCT scans [206]. Despite the outstanding results obtained for each task, we have no knowledge of models for automatically plotting of the retinal thickness map from an SD-OCT volume.

The aim of the present work is to devise a deep learning model based on a 3D Deep Convolutional Neural Network (3D-DCNN) architecture to draw the NSR thickness map from an SD-OCT volume. The remainder of this paper is organized as follows: First, in Section 2, we give an overview of the proposed method including the parameters and layers of the 3D-DCNN. Then, in Section 3 we define the experimental setup describing the dataset and the evaluation performed. In Section 4, the results are presented and discussed. Finally, Section 5 presents conclusions and future work.

7.2 Methods

The proposed method is depicted in Figure 7.1. The model is composed by two branches, where each branch decomposes one dimension of the SD-OCT volume to create a set of 2D-images: the first branch contains the features from temporal-nasal plane, and the other branch represents for the superior-inferior plane. Next, these two three-dimensional tensors are reduced with a global-average-pooling layer in order to extract the nine parameters as explained in detail as follows.

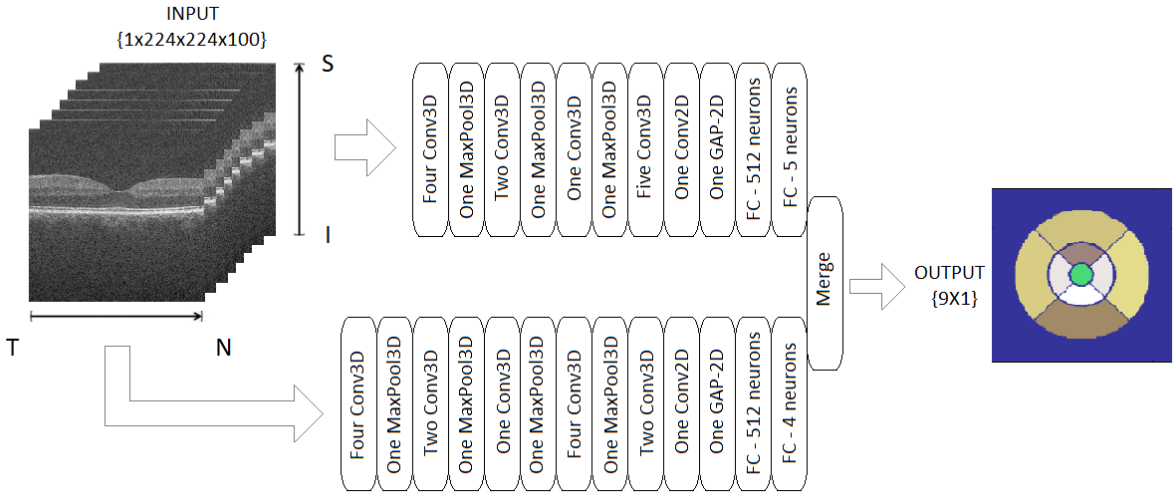

Figure 7.1: Block diagram used to estimate the nine-region values from an SD-OCT volume

7.2.1 Preprocessing stage to define the mean NSR thickness map

The interpretation of a NSR map is not an easy task, due to it contains a thousand values that represent the thickness as shown in Figure 2 (a). The interpretation is eased with a manufacturer visualization mechanism that draws a summary version to represent the mean of the nine regions. The regions are defined as: inner superior (I3), outer inferior (I6), fovea (F), inner nasal (N3), inner temporal (T3), outer nasal (N6) and outer temporal (T6) as presented in Figure 2(b). The nine regions of the NSR thickness map are obtained of drawing four equal parts with 90 degrees, and two concentric circles inside the NSR image. Finally, the mean in each region $\overline{X}$ is calculated using the equation 1, where x_i represents a pixel and, n the amount of pixels inside a region as shown in Figure 7.2. The mean NSR thickness map is plotted as presented in Figure 7.2 (c), where the lowest value is always located in fovea as reported by Farsiu et al. [55]

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{7-1}$$

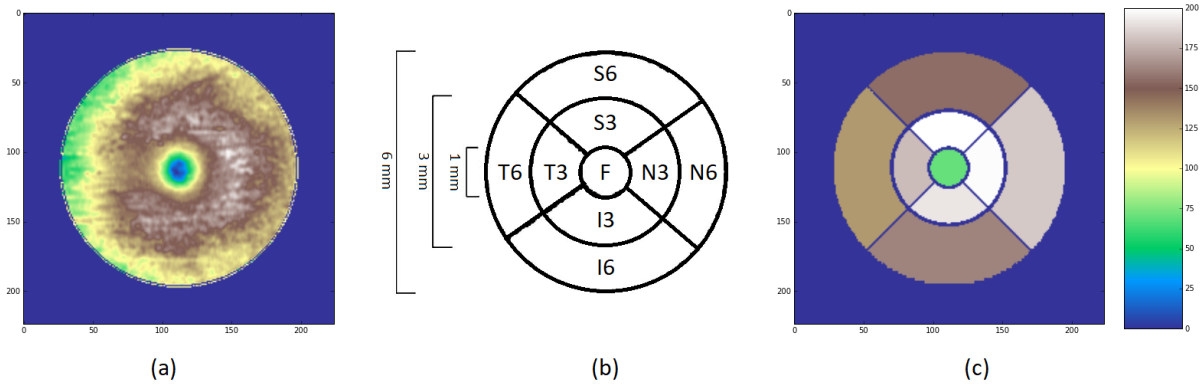

Figure 7.2: (a) Raw NSR thickness map; (b) Nine regions of the retinal thickness map; (c) Calculated mean NSR thickness map

7.2.2 3D Deep Convolutional Neural Network for drawing thickness map

The proposed model is based on a 3D-DCNN designed to learn a representation of data, and it predicts nine pixel-values between two planes of a SD-OCT volume. The first stage of the 3D-DCNN contains twenty-five 3D-Convolutional (Conv3D), two 2D-Convolutional (Conv2D), seven 3D-Max-Pooling (MaxPool3D) with kernel and pool sizes of 2 to perform the feature extraction among the two planes. Thus, two Global Average Pooling-2D (GAP-2D), and four Fully Connected (FC) layers with linear activations are used to predict the nine-regions mean values to plot the NSR thickness map.

The first branch to extract the four parameters from the temporal-nasal plane is described by 21 layers as follows: thirteen Conv3D, one Conv2D, four MaxPool3D, one GAP-2D and 2 FC layers with an output of 4 neurons that represents T6, T3, N3 and N6 as described in Figure 2 (b). On the other hand, the second branch can be described by 18 layers with: twelve Conv3D, one

Conv2D, three MaxPool3D and 2 FC layers with an the remaining 5 estimated parameters that corresponding to the superior-inferior plane named S6, S3, F, I3 and I6 as shown in Figure 7.2 (b). Finally, the two branches outputs are merged to create a calculated NSR thickness map containing the nine regions values as showed in Figure 7.2 (c). The 3D-DCNN architecture used in this work is described in details in Table **7-1**.

1st branch	**Output1 - Ch x W x H x D**	**2nd branch**	**Output2 - Ch x W x H x D**
Input	$1 \times 224 \times 224 \times 100$	Input	$1 \times 224 \times 224 \times 100$
4 Conv3D	$16 \times 224 \times 216 \times 100$	4 Conv3D	$16 \times 224 \times 224 \times 92$
1 MaxPool3D	$16 \times 224 \times 108 \times 100$	1 MaxPool3D	$16 \times 224 \times 224 \times 46$
2 Conv3D	$32 \times 226 \times 106 \times 102$	2 Conv3D	$32 \times 226 \times 226 \times 44$
1 MaxPool3D	$32 \times 226 \times 53 \times 102$	1 MaxPool3D	$32 \times 226 \times 226 \times 22$
1 Conv3D	$64 \times 226 \times 51 \times 102$	1 Conv3D	$64 \times 226 \times 226 \times 20$
1 MaxPool3D	$64 \times 226 \times 25 \times 102$	1 MaxPool3D	$64 \times 226 \times 226 \times 10$
4 Conv3D	$128 \times 226 \times 17 \times 102$	5 Conv3D	$128 \times 226 \times 226 \times 2$
1 MaxPool3D	$128 \times 226 \times 8 \times 102$	1 Conv2D	$256 \times 224 \times 224$
2 Conv3D	$128 \times 226 \times 2 \times 102$	1 GAP-2D	256
1 Conv2D	$256 \times 224 \times 100$	1 FC	512 units
1 GAP-2D	256	1 FC	5 units
1 FC	512 units	—	—
1 FC	4 units	—	—
Final Output		$[9 \times 1]$	

Table **7-1**: The architecture of the 3D-DCNN with output size in each sub-block, described as Channels (Ch) x Width (W) x Height (H) x Depth (D).

7.3 Experimental setup

7.3.1 A2A SD-OCT dataset

The dataset from the A2A SD-OCT study was approved by the institutional review boards of the 4 A2A SD-OCT clinics (Devers Eye Institute, Duke Eye Center, Emory Eye Center, and National Eye Institute). Moreover,the study follows the Declaration of Helsinki for ethical requirements, where the informed consents were obtained from all subjects [55].

The A2A SD-OCT study contains 384 SD-OCT of subjects aging 50 and 85 years of age, diagnosed with an intermediate AMD with large drusen ($> 125\ \mu m$) in both eyes or large drusen in one eligible eye, with no history of vitreoretinal surgery or ophthalmologic diseases that affect sharp vision in eyes. The SD-OCT volumes were acquired by imaging systems from Bioptigen, Inc (Research Triangle Park, NC), using a rectangular area of ($6.7mm \times 6.7mm$) by scan. The SD-OCT volumes were centered at the fovea with 1000 A-scans per B-scan and 100 B-scans per volume for both eyes. The expert SD-OCT readers assessed the scan quality for each volume and labeled 269 with intermediate AMD and 115 control (normal) eyes [55]. Finally, the dataset was

randomly split using a stratified sampling method as follows: training set (70%) with 268 SD-OCT volumes, validation set (10%) with 39 SD-OCT volumes, and the remaining (20%) for test subset. The data distribution of the nine regions for the NSR thickness map for the A2A SD-OCT dataset is presented in Figure 7.3 as follows:

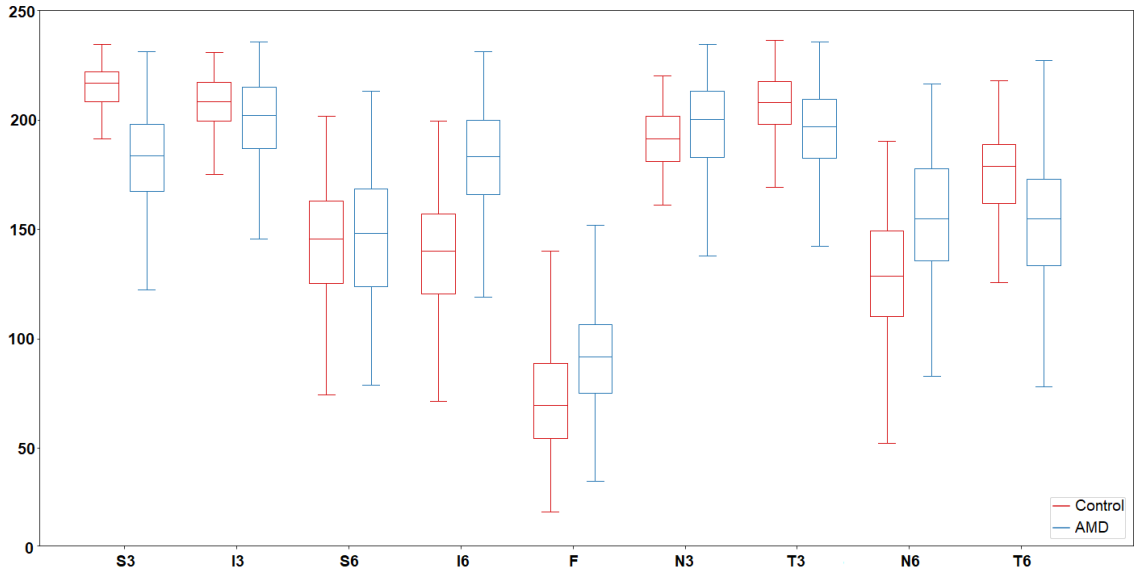

Figure 7.3: Data distributions for control (blue boxes) and AMD (red boxes) conditions in the nine regions for the NSR thickness map.

7.3.2 Evaluation

The proposed model was trained using stochastic gradient descent. The learning rate and the number of samples per epoch parameters were explored using a grid search strategy (results not reported). The best performance of the proposed model was obtained experimentally in validation with a learning rate of 0.001, a number of samples per epoch of 134 and a number of epochs of 10. The proposed approach was implemented with Keras framework [1] using GeForce GTX TITAN X from NVIDIA. The Mean Absolute Percentage Error (MAPE) rate between predicted and ground-truth thickness map was implemented as loss function. Additionally, MAPE loss and accuracy measures are monitored for both training and validation datasets.

7.4 Results

The preprocessing stage defined to obtain the mean NSR thickness maps was applied to all the three subsets of A2A SD-OCT dataset. The figure 3 depicts the box plot that summarizes the difference of values between control and AMD conditions for the nine regions.

The experimental results presented a MAPE for training and validation datasets of 11.30 and 15.58 respectively. We evaluated the proposed model with the best parameters applied to the test

[1] http://keras.io

dataset. The predicted NSR thickness maps from a Normal and AMD SD-OCT volumes of test dataset are presented in Figure 7.4.

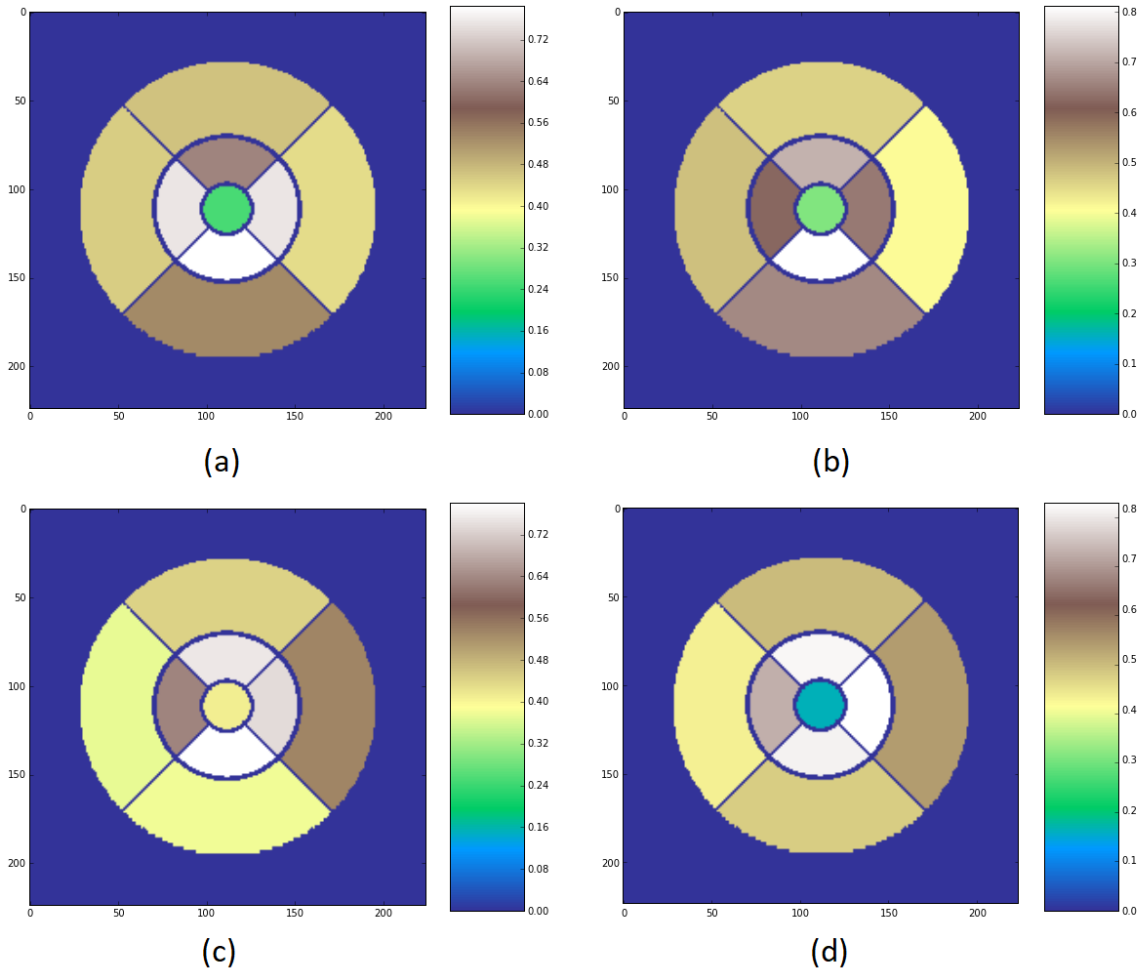

Figure 7.4: [Left] (a) Ground-truth of Normal NSR and (b) predicted NSR thickness maps; [Right] (c) ground-truth of AMD NSR, (d) predicted NSR thickness maps

7.5 Discussion and conclusion

Experimental results showed that the proposed model is highly effective to draw NSR thickness map from SD-OCT volume. The main advantage of the proposed method is the use of 3D convolutional neural network to extract features in the two planes of a volume (Superior-Inferior and Nasal-Temporal) useful to estimate the nine regions' mean of a 2D-image retinal map as shown in Figure 4.

The automatic extraction of thickness values from a SD-OCT volume could be useful for the AMD diagnosis, without the need of an additional manufacturer software to analyze the volumes. Its application to other datasets from another manufacturer, and other diseases are the subject of our future work.

8 Conclusions and future work

This thesis presented different new deep learning methods for automatically identifying different local eye features and predicting the grading of different diabetes-related conditions. The local information of these clinical findings as well as the characterization of parts of the eye deliver relevant information to improve the global classification task compared to other automatic analysis methods. Besides, the proposed methods were systematically evaluated with experts from the Fundación Oftalmológica Nacional on different datasets for diagnosis of diabetes-related conditions.

The main five contributions of the thesis to the state-of-the-art are summarized as follows: (1) the combination of an exudates-patch classifier with raw eye fundus image as a fourth-channel array for feeding a CNN for DME diagnosis. (2) A three-stages strategy inspired in the clinical routine performed by experts that deliver intermediate medical information to diagnose glaucoma. (3) A deep learning late fusion strategy that merges features extracted from CNN with polar or cartesian morphometric features to support glaucoma diagnosis. (4) A DL method termed OCT-NET that uses OCT volumes to classify DME, DR-DME and AMD delivering clinically interpretable information to support the ocular medical diagnosis. (5) The design of a 3D deep learning method to estimate the NSR thickness map from SD-OCT volumes. The proposed methods obtained outstanding results in the grading task of different diabetes-related conditions such as glaucoma, diabetic macular edema, age-related macular degeneration, and diabetic retinopathy.

When the research work of this thesis started, there were only a handful of works that applied deep learning methods to eye image analysis. This thesis contributed to show that deep learning is an effective approach for automatically finding patterns that support the diagnosis of eye diseases. One important conclusion of this work is that a straightforward application of deep learning methods to eye images does not produce the best results. It is important to take into account the particularities of eye images and to adapt the methods consequently to get the best of them.

The thesis generated different impacts during the last four years. The thesis research work was the basis for formulating projects with social impact such as the support of eye screening sessions in rural zones, academic impact with articles (overall, the thesis papers have got more than 80 citations), and interdisciplinary collaborative work with experts from Fundación Oftalmológica Nacional in the validation of the proposed models. The funding of these projects provided support for four master students, one ophthalmology fellow and one doctorate student.

The main issues found in this thesis and that constitute opportunities for future works are:

- In the medical context, new devices such as Optical Coherence Tomography-Angiography (OCTA) require new models to represent and extract features that supports the prognosis, diagnosis and follow-up of ocular diseases. Hence, the design of deep learning methods that use multimodal information such as: clinical reports, physiological data and other medical images is still an important issue.

- The validation of DL methods in clinical environment with real-word datasets and images acquired using low-cost devices could improve the social impact of the methods developed. Despite the outstanding results, there are some open challenges with these methods related with the interpretability and the feedback of medical personnel to the models. In addition, the application of DL models in medical centers could potentially increase the number of subjects diagnosed with the consequent improvement on the quality of life of the population. Realizing the potential of these techniques requires a coordinate, interdisciplinary effort of engineers and ophthalmologists focused on the patient to optimize the medical diagnosis time and costs.

- The number of free public available datasets contributes to the design of new DL methodologies to classify ocular conditions as reported in Table **2-1**. However, the use of private dataset limits the comparison among performance metrics reached by DL methods [93, 113, 121, 122]. The replication of studies reported by [113] and [122] have been criticized for the lack of information related with the description of the method and the hyperparameters used by them [124]. The use of public repositories as GitHub[1] to share datasets and codes is still a need.

- Nowadays, the growing interest of big technologies companies and medical centers to create open challenges has increased the number of ocular dataset such as: The DR detection by Kaggle [72, 105], the blindness detection by the Asia Pacific Tele-Ophthalmology Society (APTOS) [73] and iChallenge for AMD detection by Baidu [54]. These new datasets contain diverse information related to acquisition devices, image resolution and worldwide population. Moreover, DL techniques are leveraging the new data to the design of new robust approaches with outstanding performances as reported in Tables **2-2** and **2-3**.

- The lack of validation of DCNN models with real-world scans or fundus images is still a problem. We found a couple of methods validated with ocular images from medical centers [10, 111, 118, 121, 122]. However, the number of free public real-world ocular images is limited to five set of images [51, 65, 72, 73, 93]. The clinical acceptation of the proposed DCCN models depends critically of the validation in clinical studies and real-world datasets.

[1] https://github.com/

Bibliography

[1] NH1 Cho et al. "IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045". In: *Diabetes research and clinical practice* 138 (2018), pp. 271–281.

[2] International Diabetes Federation. *Diabetes Atlas 8th Edition*. 2017. URL: `https://www.idf.org/e-library/epidemiology-research/diabetes-atlas.html` (visited on 11/04/2019).

[3] American Diabetes Association et al. "2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2018". In: *Diabetes care* 41.Supplement 1 (2018), S13–S27.

[4] Carl W Baker, Yi Jiang, and Thomas Stone. "Recent advancements in diabetic retinopathy treatment from the Diabetic Retinopathy Clinical Research Network". In: *Current opinion in ophthalmology* 27.3 (2016), p. 210.

[5] Joanne WY Yau et al. "Global prevalence and major risk factors of diabetic retinopathy". In: *Diabetes care* 35.3 (2012), pp. 556–564.

[6] Alan W Stitt et al. "Advances in our understanding of diabetic retinopathy". In: *Clinical science* 125.1 (2013), pp. 1–17.

[7] Yuh-Fang Chen et al. "Macular Thickness and Aging in Retinitis Pigmentosa". In: *Optometry and Vision Science* 89.4 (2012), pp. 471–482.

[8] Gojka Roglic et al. "WHO Global report on diabetes: A summary". In: *International Journal of Noncommunicable Diseases* 1.1 (2016), p. 3.

[9] Oscar Perdomo and Fabio A. Gonz´lez. "A Systematic Review of Deep Learning Methods Applied to Ocular Images". In: *Ciencia e Ingenierıa Neogranadina* 30.1 (2020).

[10] Oscar Perdomo et al. "Classification of diabetes-related retinal diseases using a deep learning approach in optical coherence tomography". In: *Computer Methods and Programs in Biomedicine* (2019).

[11] Andrés Pérez, Oscar J Perdomo, and Fabio A González. "A lightweight deep learning model for mobile eye fundus image quality assessment"". In: *15th International Symposium on Medical Information Processing and Analysis*. Vol. 10975. International Society for Optics and Photonics. 2019, p. 109750I.

[12] Oscar J Perdomo et al. "3D deep convolutional neural network for predicting neurosensory retinal thickness map from spectral domain optical coherence tomography volumes". In: *14th International Symposium on Medical Information Processing and Analysis*. Vol. 10975. International Society for Optics and Photonics. 2018, p. 109750I.

[13] Oscar Perdomo et al. "Glaucoma diagnosis from eye fundus images based on deep morphometric feature estimation". In: *Computational pathology and ophthalmic medical image analysis*. Springer, 2018, pp. 319–327.

[14] Oscar Perdomo et al. "OCT-NET: A convolutional network for automatic classification of normal and diabetic macular edema using sd-oct volumes". In: *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE. 2018, pp. 1423–1426.

[15] Oscar Perdomo, John Arevalo, and Fabio A González. "Combining morphometric features and convolutional networks fusion for glaucoma diagnosis". In: *13th International Conference on Medical Information Processing and Analysis*. Vol. 10572. International Society for Optics and Photonics. 2017, 105721G.

[16] Oscar Perdomo, John Arevalo, and Fabio A González. "Convolutional network to detect exudates in eye fundus images of diabetic subjects". In: *12th International Symposium on Medical Information Processing and Analysis*. Vol. 10160. International Society for Optics and Photonics. 2017, 101600T.

[17] Oscar Perdomo et al. "A novel machine learning model based on exudate localization to detect diabetic macular edema". In: *Computational pathology and ophthalmic medical image analysis* (2016), pp. 137–144.

[18] Ravi M Kamble et al. "Automated diabetic macular edema (DME) analysis using fine tuning with Inception-Resnet-v2 on OCT images". In: *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)*. IEEE. 2018, pp. 442–446.

[19] Sebastián Otálora et al. "Training deep convolutional neural networks with active learning for exudate classification in eye fundus images". In: *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*. Springer, 2017, pp. 146–154.

[20] Hernán Andrés Ríos et al. "Automatic prediction of capillarity patterns on Optical Coherence Tomography Angiography images". In: *Investigative Ophthalmology & Visual Science* 59.9 (2018), pp. 1741–1741.

[21] Hernán Rios et al. "Deep Learning Method to Identify Diabetic Retinopathy and Diabetic Macular Edema Characteristics". In: *Investigative Ophthalmology & Visual Science* 60.9 (2019).

[22] Sandra Belalcazar et al. "Concordance between color photo interpretation of the optic nerve and an Unsupervised Learning Algorithm to determine optic nerve damage". In: *Investigative Ophthalmology & Visual Science* 60.9 (2019).

[23] Carlos M. Pinilla et al. "Exactitud de una red neuronal artificial para el diagnóstico del edema macular diabético con imágenes de tomografía de coherencia óptica". In: *Revista Científica de la Sociedad Colombia de Oftalmología* 1.1 (2018), pp. 8–9.

[24] Sandra Belalcazar et al. "Convolutional Neural Networks for identification and classification of optic nerve damage features". In: *Investigative Ophthalmology & Visual Science* 59.9 (2018), pp. 1719–1719.

[25] Francisco J Rodríguez et al. "Neural network for detection of diabetic macular edema in fundu color images". In: *Investigative Ophthalmology & Visual Science* 58.8 (2017), pp. 688–688.

[26] Victor H Contreras et al. "Supervised online matrix factorization for histopathological multimodal retrieval". In: *14th International Symposium on Medical Information Processing and Analysis*. Vol. 10975. International Society for Optics and Photonics. 2018, 109750Y.

[27] Sebastián Otálora et al. "Determining the scale of image patches using a deep learning approach". In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE. 2018, pp. 843–846.

[28] Nikita Gurudath, Mehmet Celenk, and H Bryan Riley. "Machine learning identification of diabetic retinopathy from fundus images". In: *2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE. 2014, pp. 1–7.

[29] Rojalina Priyadarshini, Nilamadhab Dash, and Rachita Mishra. "A Novel approach to predict diabetes mellitus using modified Extreme learning machine". In: *2014 International Conference on Electronics and Communication Systems (ICECS)*. IEEE. 2014, pp. 1–5.

[30] Gwénolé Quellec et al. "Automated assessment of diabetic retinopathy severity using content-based image retrieval in multimodal fundus photographs". In: *Investigative ophthalmology & visual science* 52.11 (2011), pp. 8342–8348.

[31] RA Welikala et al. "Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification". In: *Computer methods and programs in biomedicine* 114.3 (2014), pp. 247–261.

[32] Sohini Roychowdhury, Dara D Koozekanani, and Keshab K Parhi. "DREAM: diabetic retinopathy analysis using machine learning". In: *IEEE journal of biomedical and health informatics* 18.5 (2013), pp. 1717–1728.

[33] Dumskyj Usher et al. "Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening". In: *Diabetic Medicine* 21.1 (2004), pp. 84–90.

[34] Sam Philip et al. "The efficacy of automated "disease/no disease" grading for diabetic retinopathy in a systematic screening programme". In: *British Journal of Ophthalmology* 91.11 (2007), pp. 1512–1517.

[35] Shu-Chen Cheng and Yueh-Min Huang. "A novel approach to diagnose diabetes based on the fractal characteristics of retinal images". In: *IEEE Transactions on Information Technology in Biomedicine* 7.3 (2003), pp. 163–170.

[36] Maria Garcia et al. "Neural network based detection of hard exudates in retinal images". In: *Computer Methods and programs in biomedicine* 93.1 (2009), pp. 9–19.

[37] Wei Lu et al. "Applications of artificial intelligence in ophthalmology: general overview". In: *Journal of ophthalmology* 2018 (2018).

[38] DCS Vandarkuzhali and T Ravichandran. "Elm based detection of abnormality in retinal image of eye due to diabetic retinopathy". In: *Journal of theoretical and applied information technology* 6 (2005), pp. 423–428.

[39] Bálint Antal and András Hajdu. "An ensemble-based system for automatic screening of diabetic retinopathy". In: *Knowledge-based systems* 60 (2014), pp. 20–27.

[40] Tae Keun Yoo and Eun-Cheol Park. "Diabetic retinopathy risk prediction for fundus examination using sparse learning: a cross-sectional study". In: *BMC medical informatics and decision making* 13.1 (2013), p. 106.

[41] Leonor Guariguata et al. "An updated systematic review and meta-analysis on the social determinants of diabetes and related risk factors in the Caribbean". In: *Revista Panamericana de Salud Pública* 42 (2018).

[42] Zhuo Zhang et al. "A survey on computer aided diagnosis for ocular diseases". In: *BMC medical informatics and decision making* 14.1 (2014), p. 80.

[43] Alan D Fleming et al. "Automated microaneurysm detection using local contrast normalization and local vessel detection". In: *IEEE transactions on medical imaging* 25.9 (2006), pp. 1223–1232.

[44] Prasanna Porwal et al. "Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research". In: *Data* 3.3 (2018), p. 25.

[45] Rui Bernardes and José Cunha-Vaz. *Optical coherence tomography: a clinical and technical update*. Springer Science & Business Media, 2012.

[46] Meindert Niemeijer et al. "Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs". In: *IEEE transactions on medical imaging* 29.1 (2009), pp. 185–195.

[47] Pratul P Srinivasan et al. "Fully automated detection of diabetic macular edema and dry age-related macular degeneration from optical coherence tomography images". In: *Biomedical optics express* 5.10 (2014), pp. 3568–3577.

[48] Di Zhao et al. "Improving follow-up and reducing barriers for eye screenings in communities: the stop glaucoma study". In: *American journal of ophthalmology* 188 (2018), pp. 19–28.

[49] Muthu Rama Krishnan Mookiah et al. "Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features". In: *Knowledge-Based Systems* 33 (2012), pp. 73–82.

[50] Rüdiger Bock et al. "Glaucoma risk index: automated glaucoma detection from color fundus images". In: *Medical image analysis* 14.3 (2010), pp. 471–481.

[51] Francisco Fumero et al. "RIM-ONE: An open retinal image database for optic nerve evaluation". In: *Computer-Based Medical Systems (CBMS), 24th International Symposium on*. IEEE. 2011, pp. 1–6.

[52] Stefan Maetschke et al. "A feature agnostic approach for glaucoma detection in OCT volumes". In: *PloS one* 14.7 (2019), e0219126.

[53] Eiko K de Jong, Maartje J Geerlings, and Anneke I den Hollander. "Age-related macular degeneration". In: *Genetics and Genomics of Eye Disease*. Elsevier, 2020, pp. 155–180.

[54] Baidu. *iChallenge-AMD*. 2019. URL: `http://ai.baidu.com` (visited on 11/04/2019).

[55] Sina Farsiu et al. “Quantitative classification of eyes with and without intermediate age-related macular degeneration using optical coherence tomography”. In: *Ophthalmology* 121.1 (2014), pp. 162–172.

[56] Helen Mactier, Michael S Bradnam, and Ruth Hamilton. “Dark-adapted oscillatory potentials in preterm infants with and without retinopathy of prematurity”. In: *Documenta Ophthalmologica* 127.1 (2013), pp. 33–40.

[57] Kavita P Dhamdhere et al. “Associations between local retinal thickness and function in early diabetes”. In: *Investigative ophthalmology & visual science* 53.10 (2012), pp. 6122–6128.

[58] Dobrila Karlica et al. “Visual evoked potential can be used to detect a prediabetic form of diabetic retinopathy in patients with diabetes mellitus type I”. In: *Collegium antropologicum* 34.2 (2010), pp. 525–529.

[59] Monica Lövestam-Adrian et al. “Multifocal visual evoked potentials (MFVEP) in diabetic patients with and without polyneuropathy”. In: *The open ophthalmology journal* 6 (2012), p. 98.

[60] Sangeeta Gupta et al. “Electrophysiological evaluation in patients with type 2 diabetes mellitus by pattern reversal visual evoked potentials”. In: *National Journal of Physiology, Pharmacy and Pharmacology* 7.5 (2017), p. 527.

[61] Javad Heravian et al. “Pattern visual evoked potentials in patients with type II diabetes mellitus”. In: *Journal of ophthalmic & vision research* 7.3 (2012), p. 225.

[62] Randy Kardon et al. “Chromatic pupillometry in patients with retinitis pigmentosa”. In: *Ophthalmology* 118.2 (2011), pp. 376–381.

[63] Maria Carolina Ortube et al. “Comparative regional pupillography as a noninvasive biosensor screening method for diabetic retinopathy”. In: *Investigative ophthalmology & visual science* 54.1 (2013), pp. 9–18.

[64] Jennifer Threatt et al. “Ocular disease, knowledge and technology applications in patients with diabetes”. In: *The American journal of the medical sciences* 345.4 (2013), pp. 266–270.

[65] Danny Mitry et al. “Crowdsourcing as a novel technique for retinal fundus photography classification: Analysis of Images in the EPIC Norfolk Cohort on behalf of the UKBiobank Eye and Vision Consortium”. In: *PloS one* 8.8 (2013), e71154.

[66] Joes Staal et al. “Ridge-based vessel segmentation in color images of the retina”. In: *IEEE transactions on medical imaging* 23.4 (2004), pp. 501–509.

[67] Tomi Kauppi et al. “DIARETDB0: Evaluation database and methodology for diabetic retinopathy algorithms”. In: *Machine Vision and Pattern Recognition Research Group, Lappeenranta University of Technology, Finland* 73 (2006), pp. 1–17.

[68] Tomi Kauppi et al. “The diaretdb1 diabetic retinopathy database and evaluation protocol.” In: *BMVC*. Vol. 1. 2007, pp. 1–10.

[69] Luca Giancardo et al. "Microaneurysm detection with radon transform-based classification on retina images". In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. 2011, pp. 5939–5942.

[70] Muhammad Moazam Fraz et al. "An ensemble classification-based approach applied to retinal blood vessel segmentation". In: *IEEE Transactions on Biomedical Engineering* 59.9 (2012), pp. 2538–2548.

[71] Etienne Decencière et al. "TeleOphta: Machine learning and image processing methods for teleophthalmology". In: *Irbm* 34.2 (2013), pp. 196–203.

[72] EyePACS. *Diabetic retinopathy detection of Kaggle*. 2015. URL: `https://www.kaggle.com/c/diabetic-retinopathy-detection/data` (visited on 11/04/2019).

[73] APTOS. *Aptos blindness detection*. 2019. URL: `https://www.kaggle.com/c/aptos2019-blindness-detection/data` (visited on 11/04/2019).

[74] James Lowell et al. "Optic nerve head segmentation". In: *IEEE Transactions on medical Imaging* 23.2 (2004), pp. 256–264.

[75] Attila Budai et al. "Robust vessel segmentation in fundus images". In: *International journal of biomedical imaging* 2013 (2013).

[76] AD Hoover, Valentina Kouznetsova, and Michael Goldbaum. "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response". In: *IEEE Transactions on Medical imaging* 19.3 (2000), pp. 203–210.

[77] Adam Hoover and Michael Goldbaum. "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels". In: *IEEE transactions on medical imaging* 22.8 (2003), pp. 951–958.

[78] Damian JJ Farnell et al. "Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators". In: *Journal of the Franklin institute* 345.7 (2008), pp. 748–765.

[79] Yalin Zheng, Mohd Hanafi Ahmad Hijazi, and Frans Coenen. "Automated "disease/no disease" grading of age-related macular degeneration by an image mining approach". In: *Investigative ophthalmology & visual science* 53.13 (2012), pp. 8310–8318.

[80] Peyman Gholami et al. "OCTID: Optical Coherence Tomography Image Database". In: *arXiv preprint arXiv:1812.07056* (2018).

[81] Enrique J Carmona et al. "Identification of the optic nerve head with genetic algorithms". In: *Artificial Intelligence in Medicine* 43.3 (2008), pp. 243–259.

[82] Zhuo Zhang et al. "Origa-light: An online retinal fundus image database for glaucoma analysis and research". In: *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE. 2010, pp. 3065–3068.

[83] Meindert Niemeijer et al. "Automated measurement of the arteriolar-to-venular width ratio in digital color fundus photographs". In: *IEEE Transactions on medical imaging* 30.11 (2011), pp. 1941–1950.

[84] Zhuo Zhang et al. "ACHIKO-K: Database of fundus images from glaucoma patients". In: *2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA)*. IEEE. 2013, pp. 228–231.

[85] Jayanthi Sivaswamy et al. "A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis". In: *JSM Biomedical Imaging Data Papers* 2.1 (2015), p. 1004.

[86] Jayanthi Sivaswamy et al. "Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation". In: *2014 IEEE 11th international symposium on biomedical imaging (ISBI)*. IEEE. 2014, pp. 53–56.

[87] Ahmed Almazroa et al. "Retinal fundus images for glaucoma analysis: the RIGA dataset". In: *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*. Vol. 10579. International Society for Optics and Photonics. 2018, 105790B.

[88] IEEE Dataport. *REFUGE: Retinal Fundus Glaucoma Challenge*. 2019. URL: `https://ieee-dataport.org/documents/refuge-retinal-fundus-glaucoma-challenge` (visited on 11/04/2019).

[89] Traci E Clemons et al. "National Eye Institute visual function questionnaire in the age-related eye disease study (AREDS): AREDS report no. 10". In: *Archives of Ophthalmology* 121.2 (2003), pp. 211–217.

[90] Mahdi Kazemian Jahromi et al. "An automatic algorithm for segmentation of the boundaries of corneal layers in optical coherence tomography images using gaussian mixture model". In: *Journal of medical signals and sensors* 4.3 (2014), p. 171.

[91] Luca Giancardo et al. "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets". In: *Medical image analysis* 16.1 (2012), pp. 216–226.

[92] Reza Rasti et al. "Macular OCT classification using a multi-scale convolutional neural network ensemble". In: *IEEE transactions on medical imaging* 37.4 (2017), pp. 1024–1034.

[93] Daniel S Kermany et al. "Identifying medical diagnoses and treatable diseases by image-based deep learning". In: *Cell* 172.5 (2018), pp. 1122–1131.

[94] Sandeep Paul, Lotika Singh, et al. "A review on advances in deep learning". In: *2015 IEEE Workshop on Computational Intelligence: Theories, Applications and Future Directions (WCI)*. IEEE. 2015, pp. 1–6.

[95] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[96] Matthew D Zeiler and Rob Fergus. "Visualizing and understanding convolutional networks". In: *European conference on computer vision*. Springer. 2014, pp. 818–833.

[97] Karen Simonyan and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[98] Christian Szegedy et al. "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 1–9.

[99] Geoffrey Hinton et al. "Deep neural networks for acoustic modeling in speech recognition". In: *IEEE Signal processing magazine* 29 (2012).

[100] Ossama Abdel-Hamid et al. "Convolutional neural networks for speech recognition". In: *IEEE/ACM Transactions on audio, speech, and language processing* 22.10 (2014), pp. 1533–1545.

[101] Tara N Sainath et al. "Deep convolutional neural networks for large-scale speech tasks". In: *Neural Networks* 64 (2015), pp. 39–48.

[102] Kaggle. *Kaggle: Higgs boson machine learning challenge*. 2014. URL: `http://www.kaggle.com/c/higgs-boson` (visited on 11/04/2019).

[103] Alexander de Brebisson and Giovanni Montana. "Deep neural networks for anatomical brain segmentation". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015, pp. 20–28.

[104] Hoo-Chang Shin et al. "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data". In: *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2012), pp. 1930–1943.

[105] Kaggle. *Kaggle: 1000 Fundus images with 39 categories*. 2019. URL: `https://www.kaggle.com/linchundan/fundusimage1000` (visited on 11/04/2019).

[106] TCIA Collections. *The cancer genome atlas*. 2019. URL: `http://www.cancerimagingarchive.net/` (visited on 11/04/2019).

[107] Spineweb. *Spineweb: Collaborative platform for research on spine imaging and image analysis*. 2019. URL: `http://spineweb.digitalimaginggroup.ca/` (visited on 11/04/2019).

[108] M Usman Akram et al. "Detection and classification of retinal lesions for grading of diabetic retinopathy". In: *Computers in biology and medicine* 45 (2014), pp. 161–171.

[109] AB Aujih et al. "Analysis of retinal vessel segmentation with deep learning and its effect on diabetic retinopathy classification". In: *2018 International conference on intelligent and advanced system (ICIAS)*. IEEE. 2018, pp. 1–6.

[110] Yehui Yang et al. "Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2017, pp. 533–540.

[111] Zhentao Gao et al. "Diagnosis of Diabetic Retinopathy Using Deep Neural Networks". In: *IEEE Access* 7 (2018), pp. 3360–3370.

[112] Gwenolé Quellec et al. "Deep image mining for diabetic retinopathy screening". In: *Medical image analysis* 39 (2017), pp. 178–193.

[113] Varun Gulshan et al. "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs". In: *Jama* 316.22 (2016), pp. 2402–2410.

[114] Shujun Wang et al. "Patch-based Output Space Adversarial Learning for Joint Optic Disc and Cup Segmentation". In: *arXiv preprint arXiv:1902.07519* (2019).

[115] JR Harish Kumar, Aditya Kumar Pediredla, and Chandra Sekhar Seelamantula. "Active discs for automated optic disc segmentation". In: *2015 IEEE global conference on signal and information processing (GlobalSIP)*. IEEE. 2015, pp. 225–229.

[116] Philippe M Burlina et al. "Use of deep learning for detailed severity characterization and estimation of 5-year risk among patients with age-related macular degeneration". In: *JAMA ophthalmology* 136.12 (2018), pp. 1359–1366.

[117] Peyman Gholami. "Developing algorithms for the analysis of retinal Optical Coherence Tomography images". MA thesis. University of Waterloo, 2018.

[118] Weiwei Sun, Xiaoming Liu, and Zhou Yang. "Automated detection of age-related macular degeneration in OCT images using multiple instance learning". In: *Ninth International Conference on Digital Image Processing (ICDIP 2017)*. Vol. 10420. International Society for Optics and Photonics. 2017, p. 104203V.

[119] Christian Szegedy et al. "Rethinking the inception architecture for computer vision". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.

[120] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[121] Cecilia S Lee, Doug M Baughman, and Aaron Y Lee. "Deep learning is effective for classifying normal versus age-related macular degeneration OCT images". In: *Ophthalmology Retina* 1.4 (2017), pp. 322–327.

[122] Jeffrey De Fauw et al. "Clinically applicable deep learning for diagnosis and referral in retinal disease". In: *Nature medicine* 24.9 (2018), p. 1342.

[123] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[124] Mike Voets, Kajsa Møllersen, and Lars Ailo Bongo. "Replication study: Development and validation of deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs". In: *arXiv preprint arXiv:1803.04337* (2018).

[125] Frank B Hu, Ambika Satija, and JoAnn E Manson. "Curbing the diabetes pandemic: the need for global policy solutions". In: *Jama* 313.23 (2015), pp. 2319–2320.

[126] Thomas C Gräsbeck et al. "Fundus photography as a screening method for diabetic retinopathy in children with type 1 diabetes: Outcome of the initial photography". In: *American journal of ophthalmology* 169 (2016), pp. 227–234.

[127] Lihteh Wu et al. "Classification of diabetic retinopathy and diabetic macular edema". In: *World journal of diabetes* 4.6 (2013), p. 290.

[128] Jie Ding and Tien Yin Wong. "Current epidemiology of diabetic retinopathy and diabetic macular edema". In: *Current diabetes reports* 12.4 (2012), pp. 346–354.

[129] Kenneth W Tobin et al. "Detection of anatomic structures in human retinal imagery". In: *IEEE transactions on medical imaging* 26.12 (2007), pp. 1729–1739.

[130] AP Shingade and AR Kasetwar. "A review on implementation of algorithms for detection of diabetic retinopathy". In: *International Journal of Research in Engineering and Technology* 3.3 (2014), pp. 87–94.

[131] Thitiporn Chanwimaluang, Guoliang Fan, and Stephen R Fransen. "Hybrid retinal image registration". In: *IEEE transactions on information technology in biomedicine* 10.1 (2006), pp. 129–142.

[132] Harihar Narasimha-Iyer et al. "Integrated analysis of vascular and nonvascular changes from color retinal fundus image sequences". In: *IEEE Transactions on Biomedical Engineering* 54.8 (2007), pp. 1436–1445.

[133] S Vasanthi and RSD Wahida Banu. "Automatic segmentation and classification of hard exudates to detect macular edema in fundus images." In: *Journal of Theoretical & Applied Information Technology* 66.3 (2014).

[134] M Usman Akram et al. "Automated detection of exudates and macula for grading of diabetic macular edema". In: *Computer methods and programs in biomedicine* 114.2 (2014), pp. 141–152.

[135] Aditya Kunwar, Shrey Magotra, and M Partha Sarathi. "Detection of high-risk macular edema using texture features and classification using SVM classifier". In: *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE. 2015, pp. 2285–2289.

[136] Yann LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[137] Yann LeCun. "Learning invariant feature hierarchies". In: *Computer Vision–ECCV 2012. Workshops and Demonstrations*. Springer. 2012, pp. 496–505.

[138] Pengcheng Wu et al. "Online multimodal deep similarity learning with application to image retrieval". In: *Proceedings of the 21st ACM international conference on Multimedia - MM '13*. MM '13. ACM Press, 2013, pp. 153–162. ISBN: 9781450324045. DOI: 10.1145/2502081.2502112.

[139] N Ranjith, C Saravanan, and MR Bibin. "Glaucoma Diagnosis by Optic Cup to Disc Ratio Estimation". In: *International Journal of Inventive Engineering and Sciences (IJIES)* 3.5 (2015), pp. 1–5.

[140] Sourav Samanta et al. "Haralick features based automated glaucoma classification using back propagation neural network". In: *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*. Springer. 2015, pp. 351–358.

[141] Artem Sevastopolsky. "Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network". In: *Pattern Recognition and Image Analysis* 27.3 (2017), pp. 618–624.

[142] Xiangyu Chen et al. "Glaucoma detection based on deep convolutional neural network". In: *EMBC, 37th Annual International Conference of the IEEE*. IEEE. 2015, pp. 715–718.

[143] José Ignacio Orlando et al. "Convolutional neural network transfer for automated glaucoma identification". In: *12th International Symposium on Medical Information Processing and Analysis*. Vol. 10160. International Society for Optics and Photonics. 2017, 101600U.

[144] Huazhu Fu et al. "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation". In: *IEEE transactions on medical imaging* 37.7 (2018), pp. 1597–1605.

[145] Qing Liu et al. "DDNet: Cartesian-polar Dual-domain Network for the Joint Optic Disc and Cup Segmentation". In: *arXiv preprint arXiv:1904.08773* (2019).

[146] Qing Liu et al. "A spatial-aware joint optic disc and cup segmentation method". In: *Neurocomputing* 359 (2019), pp. 285–297.

[147] Huazhu Fu et al. "Disc-aware ensemble network for glaucoma screening from fundus image". In: *IEEE transactions on medical imaging* 37.11 (2018), pp. 2493–2501.

[148] Douglas G Altman. *Practical statistics for medical research*. CRC press, 1990.

[149] George L Spaeth et al. "The disc damage likelihood scale: reproducibility of a new method of estimating the amount of optic nerve damage caused by glaucoma." In: *Transactions of the American Ophthalmological Society* 100 (2002), p. 181.

[150] Ian JC MacCormick et al. "Accurate, fast, data efficient and interpretable glaucoma diagnosis with automated spatial analysis of the whole cup to disc profile". In: *PloS one* 14.1 (2019).

[151] Kevis-Kokitsi Maninis et al. "Deep retinal image understanding". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016, pp. 140–148.

[152] José Ignacio Orlando et al. "REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs". In: *Medical image analysis* 59 (2020), p. 101570.

[153] Shujun Wang, Lequan Yu, and Pheng-Ann Heng. "Optic Disc and Cup Segmentation with Output Space Domain Adaptation". In: *Refuge reports* ().

[154] Pengshuai Yin et al. "Optic Disc and Cup Segmentation using Ensemble Deep Neural Networks". In: *Refuge reports* ().

[155] Hong Kang et al. "Pixel quantification for robust segmentation of optic cup". In: *Refuge reports* ().

[156] Zifeng Wu et al. "Optic Disc/Cup Segmentation and Glaucoma Classification from Fundus Images with Fully Convolutional Networks". In: *Refuge reports* ().

[157] Xuesheng Bian et al. "Automatic Optic Disc/Cup Segmentation and Glaucoma Classification and Fovea Localization?" In: *Refuge reports* ().

[158] Peng Liu and Ruogu Fang. "Regression and Learning with Pixel-wise Attention for Retinal Fundus Glaucoma Segmentation and Detection". In: *arXiv preprint arXiv:2001.01815* (2020).

[159] Apoorva Sikka, Sai Samarth R Phaye, and Deepti R Bathula. "REFUGE Challenge Submission: Using Dense U-Nets to Detect Glaucoma and Segment Optic Disc and Cup". In: *Refuge reports* ().

[160] Joonseok Lee et al. "Development of an End-to-End Deep Learning System for Glaucoma Screening Using Color Fundus Images". In: *Refuge reports* ().

[161] Sharath M Shankaranarayana et al. "Deep Learning based Retinal Image Analysis for evaluation of Glaucoma". In: *Refuge reports* ().

[162] Yih-Chung Tham et al. "Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis". In: *Ophthalmology* 121.11 (2014), pp. 2081–2090.

[163] Robert N Weinreb, Tin Aung, and Felipe A Medeiros. "The pathophysiology and treatment of glaucoma: a review". In: *Jama* 311.18 (2014), pp. 1901–1911.

[164] Daniel M Stein, Gadi Wollstein, and Joel S Schuman. "Imaging in glaucoma". In: *Ophthalmology clinics of North America* 17.1 (2004), p. 33.

[165] Pardha Saradhi Mittapalli and Giri Babu Kande. "Segmentation of optic disk and optic cup from digital fundus images for the assessment of glaucoma". In: *Biomedical Signal Processing and Control* 24 (2016), pp. 34–46.

[166] Anindita Septiarini et al. "Automatic Glaucoma Detection Method Applying a Statistical Approach to Fundus Images". In: *Healthcare informatics research* 24.1 (2018), pp. 53–60.

[167] Baidaa Al-Bander et al. "Automated glaucoma diagnosis using deep learning approach". In: *SSD, 14th International Multi-Conference on*. IEEE. 2017, pp. 207–210.

[168] Qaisar Abbas. "Glaucoma-Deep: Detection of Glaucoma Eye Disease on Retinal Fundus Images using Deep Learning". In: *International Journal of Advanced Computer Science and Applications* 8.6 (2017), pp. 41–45.

[169] Yoshua Bengio et al. "Curriculum learning". In: *Proceedings of the 26th annual international conference on machine learning*. ACM. 2009, pp. 41–48.

[170] Jayanthi Sivaswamy et al. "A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis". In: *JSM Biomedical Imaging Data Papers* 2.1 (2015), p. 1004.

[171] Undurti N Das. "Diabetic macular edema, retinopathy and age-related macular degeneration as inflammatory conditions". In: *Archives of medical science: AMS* 12.5 (2016), p. 1142.

[172] Mads Fonager Nørgaard and Jakob Grauslund. "Automated Screening for Diabetic Retinopathy–A Systematic Review". In: *Ophthalmic research* (2018).

[173] Ryan Lee, Tien Y Wong, and Charumathi Sabanayagam. "Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss". In: *Eye and vision* 2.1 (2015), p. 17.

[174] Yali Jia et al. "Quantitative optical coherence tomography angiography of choroidal neovascularization in age-related macular degeneration". In: *Ophthalmology* 121.7 (2014), pp. 1435–1444.

[175] Ryan L Shelton et al. “Optical coherence tomography for advanced screening in the primary care office”. In: *Journal of biophotonics* 7.7 (2014), pp. 525–533.

[176] Martin M Nentwich and Michael W Ulbig. “Diabetic retinopathy-ocular complications of diabetes mellitus”. In: *World journal of diabetes* 6.3 (2015), p. 489.

[177] Desire Sidibe et al. “An anomaly detection approach for the identification of DME patients using spectral domain optical coherence tomography images”. In: *Computer methods and programs in biomedicine* 139 (2017), pp. 109–117.

[178] Gabriela Samagaio et al. “Automatic Macular Edema Identification and Characterization Using OCT Images”. In: *Computer Methods and Programs in Biomedicine* (2018).

[179] E Talisa et al. “Spectral-domain optical coherence tomography angiography of choroidal neovascularization”. In: *Ophthalmology* 122.6 (2015), pp. 1228–1238.

[180] Adeel M Syed et al. “Automated diagnosis of macular edema and central serous retinopathy through robust reconstruction of 3D retinal surfaces”. In: *Computer methods and programs in biomedicine* 137 (2016), pp. 1–10.

[181] Oliver Faust et al. “Deep learning for healthcare applications based on physiological signals: a review”. In: *Computer methods and programs in biomedicine* (2018).

[182] Eli Gibson et al. “NiftyNet: a deep-learning platform for medical imaging”. In: *Computer methods and programs in biomedicine* 158 (2018), pp. 113–122.

[183] Jose Ignacio Orlando et al. “An ensemble deep learning based approach for red lesion detection in fundus images”. In: *Computer methods and programs in biomedicine* 153 (2018), pp. 115–127.

[184] Yawen Xiao et al. “A deep learning-based multi-model ensemble method for cancer prediction”. In: *Computer methods and programs in biomedicine* 153 (2018), pp. 1–9.

[185] Xiaohong W Gao, Rui Hui, and Zengmin Tian. “Classification of CT brain images based on deep learning networks”. In: *Computer methods and programs in biomedicine* 138 (2017), pp. 49–56.

[186] Zhong Yin et al. “Recognition of emotions using multimodal physiological signals and an ensemble deep learning model”. In: *Computer methods and programs in biomedicine* 140 (2017), pp. 93–110.

[187] Hyungwoo Lee et al. “Automated Segmentation of Lesions Including Subretinal Hyperreflective Material in Neovascular Age-related Macular Degeneration”. In: *American journal of ophthalmology* 191 (2018), pp. 64–75.

[188] Sripad Krishna Devalla et al. “DRUNET: a dilated-residual U-Net deep learning network to segment optic nerve head tissues in optical coherence tomography images”. In: *Biomedical optics express* 9.7 (2018), pp. 3244–3265.

[189] Freerk G Venhuizen et al. “Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks”. In: *Biomedical optics express* 8.7 (2017), pp. 3292–3316.

[190] Avi Ben-Cohen et al. "Retinal layers segmentation using Fully Convolutional Network in OCT images". In: *RSIP Vision* (2017).

[191] Mike Pekala et al. "Deep Learning based Retinal OCT Segmentation". In: *arXiv preprint arXiv:1801.09749* (2018).

[192] Yufan He et al. "Topology guaranteed segmentation of the human retina from OCT using convolutional neural networks". In: *arXiv preprint arXiv:1803.05120* (2018).

[193] Muhammad Awais et al. "Classification of SD-OCT images using a Deep learning approach". In: *Signal and Image Processing Applications (ICSIPA), 2017 IEEE International Conference on*. IEEE. 2017, pp. 489–492.

[194] Genevieve CY Chan et al. "Fusing Results of Several Deep Learning Architectures for Automatic Classification of Normal and Diabetic Macular Edema in Optical Coherence Tomography". In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2018, pp. 670–673.

[195] Freerk G Venhuizen et al. "Automated age-related macular degeneration classification in OCT using unsupervised feature learning". In: *Medical Imaging 2015: Computer-Aided Diagnosis*. Vol. 9414. International Society for Optics and Photonics. 2015, p. 94141I.

[196] Arunava Chakravarty, Divya Jyothi Gaddipati, and Jayanthi Sivaswamy. "Construction of a Retinal Atlas for Macular OCT Volumes". In: *International Conference Image Analysis and Recognition*. Springer. 2018, pp. 650–658.

[197] R Natarajan et al. "Comparative analysis of optical coherence tomography retinal images using multidimensional and cluster methods." In: *Biomedical Research* 26.2 (2015).

[198] Bolei Zhou et al. "Learning deep features for discriminative localization". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 2921–2929.

[199] Ramprasaath R Selvaraju et al. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization." In: *ICCV*. 2017, pp. 618–626.

[200] Carmelina Trimboli-Heidler, Kelly Vogt, and Avery Robert A. "Volume averaging of spectral-domain optical coherence tomography impacts retinal segmentation in children". In: *Translational vision science & technology* 5.4 (2016), pp. 1–9.

[201] Muna Bhende et al. "Optical coherence tomography: A guide to interpretation of common macular diseases". In: *Indian journal of ophthalmology* 66.1 (2017), pp. 20–35.

[202] SB Velaga et al. "Impact of optical coherence tomography scanning density on quantitative analyses in neovascular age-related macular degeneration". In: *Eye* 31.1 (2017), pp. 53–61.

[203] Sijie Niu et al. "Automated detection of foveal center in SD-OCT images using the saliency of retinal thickness maps". In: *Medical physics* 44.12 (2017), pp. 6390–6403.

[204] Punal M Arabi et al. "Identification of Age-Related Macular Degeneration Using OCT Images". In: *IOP Conference Series: Materials Science and Engineering*. Vol. 310. 1. IOP Publishing. 2018, p. 012096.

[205] Leyuan Fang et al. “Automatic segmentation of nine retinal layer boundaries in OCT images of non-exudative AMD patients using deep learning and graph search”. In: *Biomedical optics express* 8.5 (2017), pp. 2732–2744.

[206] Abhijit Guha Roy et al. “ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks”. In: *Biomedical optics express* 8.8 (2017), pp. 3627–3642.