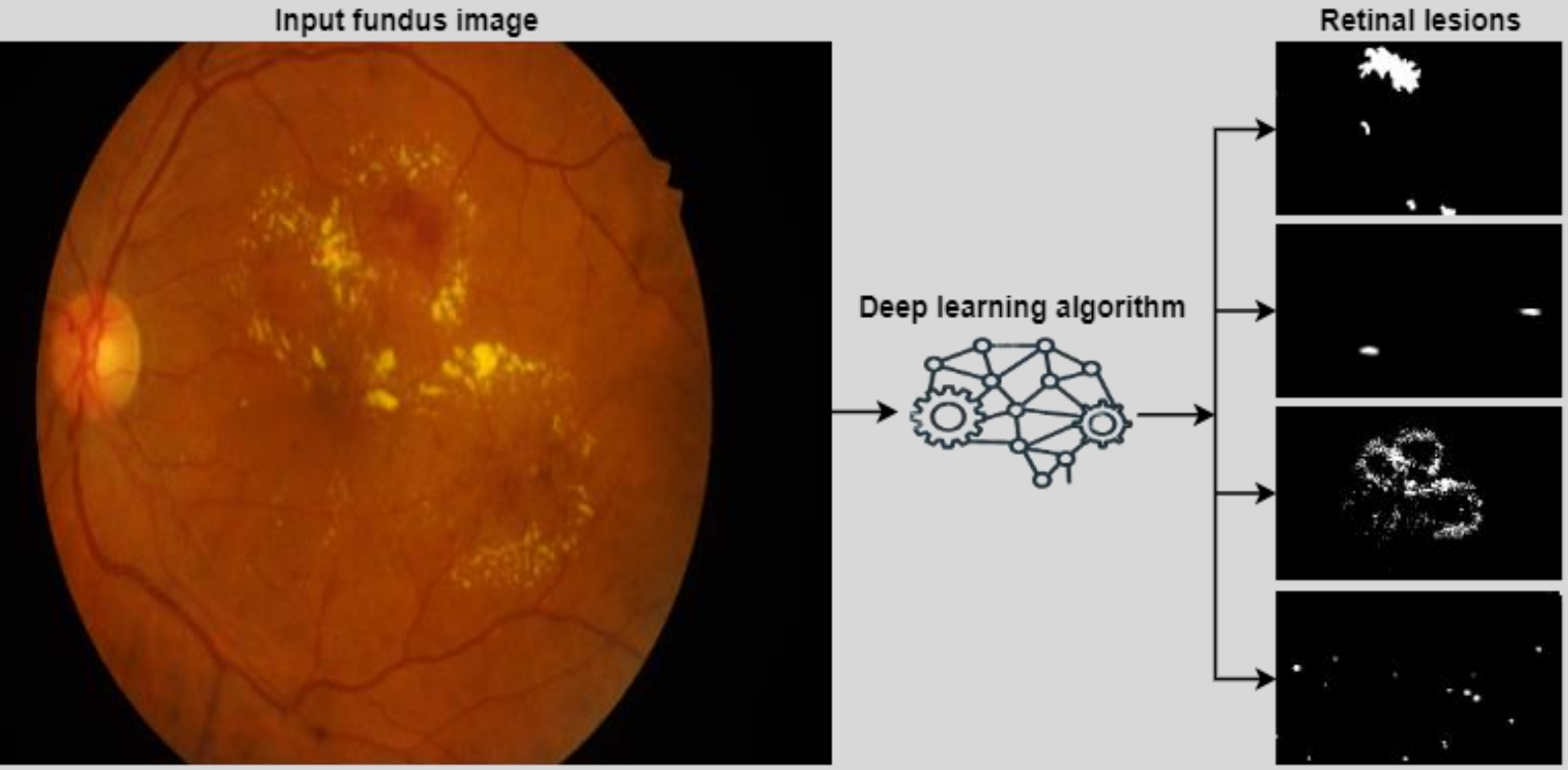
DEEP LEARNING-BASED METHODS FOR EXTRACTING FUNDUS IMAGE LANDMARKS AND SIGNS OF EYE DISEASES

Mohammed Yousef Salem Ali

UNIVERSITAT ROVIRA i VIRGILI

Deep Learning-based Methods for Extracting Fundus Image Landmarks and Signs of Eye Diseases

Mohammed Yousef Salem Ali



DOCTORAL THESIS
2022

Deep Learning-based Methods for Extracting Fundus Image Landmarks and Signs of Eye Diseases

DOCTORAL THESIS

Author:
Mohammed Yousef Salem Ali

Advisors:
Dr. Aïda Valls Mateu
Dr. Mohamed Abdelnasser Mohamed Mahmoud
Dr. Marc Baget Bernaldiz

Departament d'Enginyeria Informàtica i Matemàtiques

UNIVERSITAT ROVIRA I VIRGILI
Tarragona
2022

UNIVERSITAT ROVIRA I VIRGILI
DEEP LEARNING-BASED METHODS FOR EXTRACTING FUNDUS IMAGE LANDMARKS AND SIGNS OF EYE DISEASES
Mohammed Yousef Salem Ali

UNIVERSITAT
ROVIRA I VIRGILI

Departament d'Enginyeria Informàtica
i Matemàtiques

Av. Paisos Catalans, 27
43007 Tarragona
Tel. +34 977 55 95 95
Fax. +34 977 55 95 97

We STATE that the present study, entitled "Deep Learning-based Methods for Extracting Fundus Image Landmarks and Signs of Eye Diseases," presented by Mohammed Yousef Salem Ali, for the award of the degree of Doctor, has been carried out under our supervision at the Department of Computer Engineering and Mathematics of this university.

Tarragona, October 2022.

Doctoral Thesis Supervisors,

Dr. Aïda Valls Mateu

Dr. Mohamed Abdelnasser Mohamed Mahmoud

Dr. Marc Baget Bernaldiz

UNIVERSITAT ROVIRA I VIRGILI
DEEP LEARNING-BASED METHODS FOR EXTRACTING FUNDUS IMAGE LANDMARKS AND SIGNS OF EYE DISEASES
Mohammed Yousef Salem Ali

Dedication

To the soul of my eldest son Abood, who passed away while I was starting to study for a Ph.D.
To my uncle Dr. Abdullah, my wife, my parents, my brothers, and my sons (Yomna, Yousef, and Aiman).
To those whose financial support made this study a success.

I will never forget you; I promise that I will continue doing good things that bring happiness to humanity.

UNIVERSITAT ROVIRA I VIRGILI
DEEP LEARNING-BASED METHODS FOR EXTRACTING FUNDUS IMAGE LANDMARKS AND SIGNS OF EYE DISEASES
Mohammed Yousef Salem Ali

Acknowledgements

This work and the author were funded by the research projects PI21/00064 and PI18/00169 from Instituto de Salud Carlos III & FEDER funds, and the University Rovira i Virgili with projects 2022PFR-URV-41, 2021PFR-B2-103 and 2019PFR-B2-61.

I would like to express my gratitude to my supervisors, Dr. Aïda Valls, Dr. Mohamed AbdelNasser, and Dr. Marc Baget, for their helpful guidance, insightful comments, and considerable encouragement in completing this thesis. They have guided me to pursue fundamental problems that will have a practical impact and were always available to guide me whenever I approached them. Their extensive knowledge and exceptional ability to find new approaches to complex problems were pivotal in this work and my development as a researcher. This work would not have been completed without their encouragement and patience. In addition, because of the research environment sustained by Dr. Aïda Valls, I have crossed paths with many graduate students and postdocs who have influenced and enhanced my research.

My special thanks go to our ITAKA group member, Dr. Mohammed Jabreel, who has motivated me to work on deep learning and led me to discuss ideas and revise the manuscripts; thanks for all his help and support.

My heartfelt gratitude goes to Mr. Najeeb Jabreel for the stimulating discussions and sleepless nights we spent working together.

Also, I thank my fellow labmates, Mrs. Najla Alzyadi and Mr. Jordi Pascual, and all my friends, Dr. Fadi, Saddam, Saif, Nadeem, Ammar, Loay, and Hani Saeed, for their support and all the moments we spent together.

Finally, I would have never managed to accomplish this task without God's blessings and the loving support of my family. I cordially thank my wife, parents, brothers, sisters, uncle Dr. Abdullah, and my sons.

List of Abbreviations

AI Artificial Intelligence

ANN Artificial Neural Network

DL Deep Learning

DR Diabetic Retinopathy

AMD Age-related Macular Degeneration

CAD Computer Aided Diagnosis

ROI Region of Interest

CNN Convolutional Neural Network

EX Hard Exudates

SE Soft Exudates

HE Hemorrhages

MA Microaneurysms

OWA Ordered Weighted Averaging

WPM Weighted Power Mean

EXM EXponential Mean

DA Average partial Disjunction

CA Average partial Conjunction

SVM Support Vector Machine

FCN Fully Convolutional Network

TP True Positive

FP False Positive

FN False Negative

TN True Negative

BCE Binary Cross Entropy

SGD Stochastic Gradient Descent

ACC Accuracy

IOU Intersection over Union

SEN Sensitivity

Re Recall

Pre Precision

Spe Specificity

AUC Area Under Curve

AUPR Area Under Precision-Recall

ROC Receiver Operating Curve

OD Optic Disc

OC Optic Cup

CDR Cup to Disc Ratio

VCDR Vertical Cup to Disc Ratio

WHO World Health Organization

GSCs Gated Skip Connections

SJR Sant Joan de Reus

IDRiD Indian Diabetic Retinopathy Image Dataset

ASPP Atrous Spatial Pyramid Pooling

GSAB Unet+GSC+SacaleAtBegining

SAT multi-scale attention

SOTA state-of-the-art

DDN Dual Decoder based Network for Exudates Segmentation

MSN Multi-scale based Network for Exudates Segmentation

LezioSeg Eye Lesions Segmenting Utilizing Hybrid Multi-Scale Attention Modules

Abstract

The early detection and treatment of eye illnesses such as Diabetic retinopathy (DR) and Glaucoma disease in the retina are crucial to avoid vision loss. However, manual detection of small lesions in the fundus image is a painstaking process that consumes ophthalmologists' time and effort. The complex structure of lesions, various sizes, differences in brightness, and the inter-class similarity with other fundus tissues make it more challenging, even for ophthalmology experts. Therefore, it would be more beneficial to have a computer-aided diagnosis (CAD) system that can automatically outline the possible disease regions to the doctor.

In this thesis, we consider four different tasks on the eye: optic disc segmentation, glaucoma detection, segmenting exudate lesions, and segmenting other kinds of retinal eye DR lesions from fundus images. These tasks are extremely challenging due to several sources of variability in the image-capturing processes, the complex structure, similarities of lesions, and a small amount of annotated images.

Optic disc (OD) carries essential information linked to DR and glaucoma. Therefore, The first contribution in this thesis is a deep learning-based system for OD segmentation based on an ensemble of efficient semantic segmentation models. The aggregation was performed with the ordered weighted averaging operators on the different deep learning models. We propose an andness-directed set of weights to give a different contribution to the models according to their performance results. Tests were done with in-house dataset from Hospital Sant Joan de Reus.

The second contribution is an efficient CAD system for diagnosing glaucoma utilizing transfer learning and fuzzy aggregation operators. Specifically, the proposed CAD system includes three stages: 1) Detection of the region of interest of the optic disc using an efficient deep learning network, 2) Classification based on different pre-trained deep convolutional neural networks and support vector machines, and 3)

Use of fuzzy aggregation operators to fuse the predictions of glaucoma classifiers: ordered weighted average, weighted power mean, and exponential mean. Three public datasets were used: DRISHTI-GS1, RIM-ONE, and REFUGE.

Next, the thesis deals with the automatic detection of DR lesions in the eye. The third contribution of the thesis is two novel exudate segmentation methods. The first method is based on a dual-decoder boosted network with a pre-trained ImageNet ResNet50 encoder, Atrous Spatial Pyramid Pooling (ASPP), and gated skip connections (GSCs). The second method is based on two multi-scale modules with ImageNet MobileNet encoder, ASPP, and GSCs. The effectiveness of the two methods was assessed on the IDRiD dataset. They obtained promising results outperforming numerous previous methods.

As a fourth contribution, the thesis proposes an integrated retinal lesion segmentation model called LezioSeg, for segmenting the four typesof lesions: hard exudates (EX), soft exudates (SE), hemorrhages (HE), and microaneurysms (MA). It comprised four main elements: two multi-scale mules, ASPP at the neck of the network and multi-scale attention (SAT) unit after the decoder of the network, MobileNet backbone encoder, and several GSCs in the decoder block. It is 10 Million parameters, much lighter than those models that use ResNets or VGGNets backbones or those models that use the dual networks and leads to high segmentation performance of retinal eye lesions with the IDRiD and E-ophtha datasets. Extensive experiments had shown that the method showed superiority over other models and a competitive performance when it was generalized on the DDR dataset, which took in different imaging conditions. It can work well in the cases of medical images with small objects from different fields.

Keywords: Medical image analysis, Optic disc, Retinal eye lesions, Glaucoma, Deep learning, Convolutional neural network, Segmentation, Classification, Aggregation operators.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation

This thesis is devoted to designing and developing automatic computer vision tools for helping in the diagnosis of some eye diseases that cause vision loss, specifically Glaucoma and Diabetic Retinopathy.

Diabetic Retinopathy (DR) is the most common reason for visual loss in working-age adults. In 2040, it is predicted that over 200 million people will have DR [2]. Glaucoma affects about 75 million individuals worldwide, and it is called the silent thief of sight [3]. Therefore, early diagnosis of glaucoma and DR requires an efficient screening procedure. Medical centers regularly perform eye checkups for patients, especially diabetic ones, to minimize the risk of blindness development [4]. Unfortunately, examining eye fundus images is labor-intensive, time-consuming, expensive, and error-prone. Specialized ophthalmologists need much practice to achieve a high experience of accurate diagnosis from the observation of eye fundus images.

To assist the medical personnel, it is possible to build computer systems that analyze eye fundus images. Computer vision is a branch of Artificial Intelligence (AI) that refers to the technologies that allow computers to interpret medical images. Computer vision empowered with other intelligent tools has led to many automated tools, computer-aided diagnosis (CAD), for the healthcare sector [5]. These tools have achieved performance similar to humans in image classification, object localization and recognition, and image segmentation tasks.

Emerging AI techniques, like deep learning and especially the convolutional neural networks (CNNs) [6] have significantly improved CAD systems' performance. Developing advanced deep learning-based methods for detecting fundus images' landmarks and segmenting eye lesions helps ophthalmologists to make a more reliable diagnosis for eye diseases and treatment. In this line, this thesis presents various

AI-based tools for optic disc segmentation, glaucoma diagnoses, and retinal lesions segmentation from eye fundus images.

This work is part of a Spanish research project conducted between Universitat Rovira i Virgili and Institut d'Investigació Sanitària Pere Virgili, in Catalonia. It is funded by Instituto de Investigación Carlos III and Fondos Feder. The ophthalmology group of Hospital Universitari Sant Joan de Reus provides data, knowledge, and supervision to this work.

The next section introduces some details about the diseases studied in this thesis. After, the goals and the main contributions of this dissertation are presented.

1.2 Human Eye Lesions

The human eye is a natural optical device used to see objects by human beings. It is like a camera that has a lens and screen system. Figure 1.1 illustrates the anatomical structure of the human eye. Below, we introduce an overview of the most human eye parts related to this thesis.

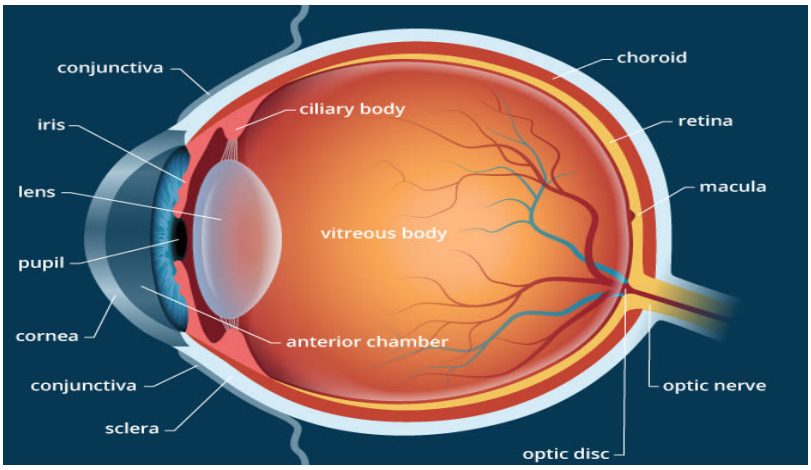


Figure 1.1: Anatomical structure of the human eye.
Source: https://www.allaboutvision.com/resources/anatomy.htm

- Optic Nerve Head (ONH) is commonly referred to as the optic disc and

sometimes as the optic papilla. It takes the image to the brain in the form of electrical signals. It is a circular area where the optic nerve enters the retina coming from the brain, which does not contain receptors itself, and is thereby the blind spot of the eye [7].

- Retina is a light-sensitive screen inside the eye on which image is formed by the lens and converted into signals that reach the brain through the optic nerve. It contains light-sensitive cells called cones and rods, responsible for daytime and night vision, respectively [8].
- Macula or Fovea is a yellow oval spot close to the retina's center. It is a part of the retina responsible for sharp, detailed central vision. Light rays from an object are concentrated on the macula when an eye looks at it directly. In addition, the macula, also called the fovea, contains a very high concentration of cones, which provides accurate, focused eyesight.

Many eye diseases may lead to irreversible vision loss, such as glaucoma, DR, and age-related macular degeneration (AMD). Hence, regular eye screening is essential for detecting these eye diseases early and providing appropriate treatment. Glaucoma

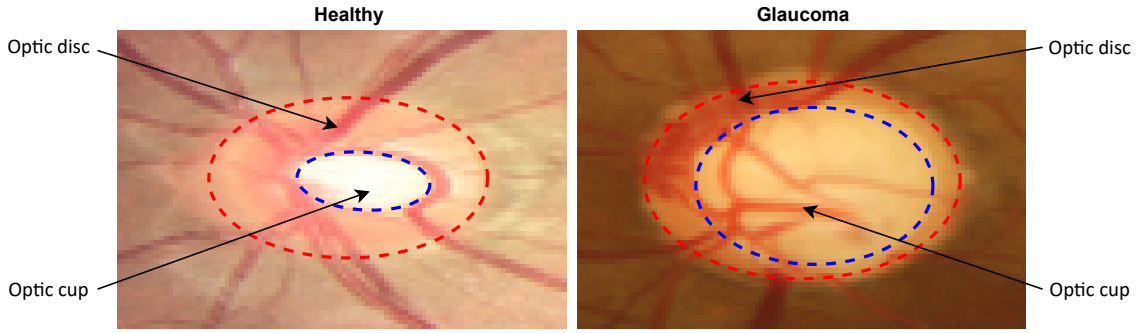


Figure 1.2: Healthy (left) and glaucoma (right) fundus image.

affects about 75 million individuals worldwide. Glaucoma is more common in elderly people, and because it typically takes a long time for eyesight to be lost, it has earned the name "silent thief of sight." Glaucoma, after cataracts, is the second-leading cause of blindness globally [3]. Figure 1.2 shows samples of healthy (left) and glaucoma

(right) fundus images. It is noticeable from the figure that the size of the optic cup is closer to the size of the optic disc in glaucoma. Diabetes mellitus causes

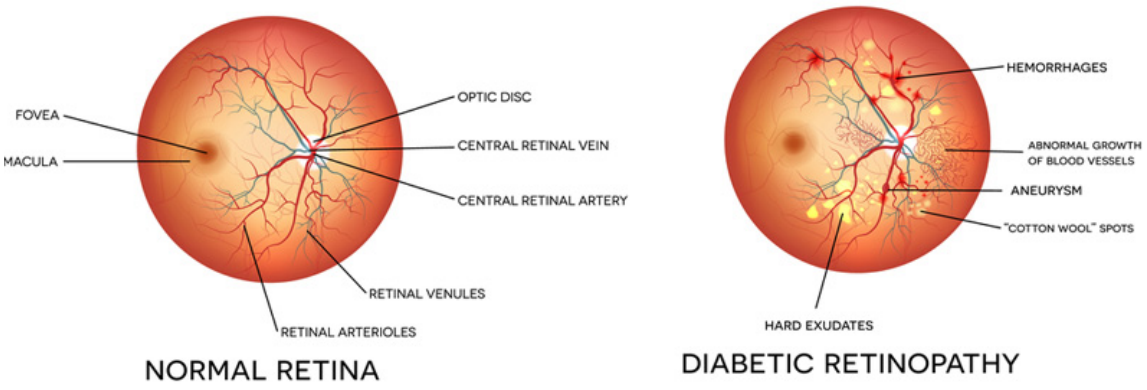


Figure 1.3: Fundus images of healthy (left) and DR (right) cases.
Source: https://neoretina.com/blog/diabetic-retinopathy-can-it-be-reversed/

damage to the retina, a condition known as diabetic retinopathy, commonly referred to as diabetic eye disease. By 2040, it is predicted that over 600 million people will have diabetes [2], with one-third of those obtaining DR, the world's most common cause of visual loss in working-age adults [4]. Figure 1.3 shows a fundus image with (right) and without DR (left). Many lesions appear on the DR fundus image, such as HE, MA, EX, and SE; SE is also known as cotton wool. AMD is another

Figure 1.4: Healthy (left) and AMD (right) fundus image.

type of eye infection that is the leading cause of blindness in people 50 years or

older. It is caused by macula degeneration in the retina. The macula is in charge of maintaining sharp vision. The central vision is affected when the macula degenerates. In addition, small yellowish sediments (drusens) grow beneath the retina in AMD. These sediments cause blood leakage and retinal damage, as shown in Figure 1.4. Globally, 288 million people are expected to be affected by the AMD by 2040 [9]. Human vision loss can be prevented using fundus image analysis to identify and treat eye illnesses, including glaucoma, DR, and AMD.

1.3 Thesis Objectives

After presenting the motivation and problems of eye diseases, The goals of this thesis are related to the automatic analysis of eye fundus images captured by opthalmologists. The main objectives faced in this doctoral thesis can be summarized as:

- To develop an efficient method for optic disc segmentation. An accurate identification of the optic disc is crucial to the posterior diagnosis of DR and Glaucoma.
- To present a fully automatic CAD system for glaucoma detection from the eye fundus images.
- To propose robust deep learning-based methods for segmenting different eye lesions due to Diabetic Retinopathy, like microaneurysm, hard exudates, soft exudates, and hemorrhages.

1.4 Contributions and Scientific Dissemination

This thesis is focused on extracting fundus image landmarks and signs of eye disease, including optic disc and eye lesions segmentation in fundus images. In the first part

of this research, we propose an efficient strategy for segmenting the OD in fundus images of the human eye based on accurate deep learning-based segmentation models and an ensemble method of deep learning models. After that, we have designed an accurate glaucoma prediction CAD system based on three main components: the segmented OD, glaucoma classifiers based on transfer learning of pre-trained convolutional neural networks (CNNs) feature extractors, and dynamic aggregation operators to merge the outputs of the different CNNs. The results of this study have been published in the following papers:

1. Mohammed Yousef Salem Ali, Mohammed Jabreel, Aida Valls, Marc Baget Bernaldiz, and Mohamed Abdel-Nasser. ***"Glaucoma Detection in Retinal Fundus Images Based on Deep Transfer Learning and Fuzzy Aggregation Operators"***, International Journal on Artificial Intelligence Tools. ISI-JCR Impact Factor: 1.208 **(Q3)**. In press., 2022.
2. Mohammed Yousef Salem Ali, Mohammed Jabreel, Mohamed Abdel-Nasser, Aida Valls, Marc Baget Bernaldiz, ***"An ensemble of deep learning models to segment the optic disc in eye fundus images"***, published in CARS 2021 Computer Assisted Radiology and Surgery - 35th International Congress and Exhibition, pp. s21-s22, 2021.
3. Mohammed Yousef Salem Ali, Mohammed Jabreel, Mohamed Abdel-Nasser, Aida Valls, Marc Baget Bernaldiz, ***"Segmenting the Optic Disc using a Deep Learning Ensemble Model based on OWA Operators"***, 23^{rd} International Conference of the Catalan Association for Artificial Intelligence (CCIA 2021), pp. 305-314, IOS Press Book, 2021.

In the second part, we propose efficient some DR lesions segmentation methods, in order to help detect Diabetic Retinopahy.

Firstly, we present two novel exudates lesions segmentation methods 1) Uses Dual-Decoder Boosted Network with pre-trained ImageNet ResNet50 encoder,

ASPP, and GSCs. 2) Uses Two multi-scale modules with ImageNet MobileNet encoder, ASPP, and GSCs. The effectiveness of the two methods was assessed on the IDRiD publicly dataset. They led to high segmentation performance of exudates outperforming numerous state-of-the-art approaches.

Second, we propose an integrated deep-learning model for segmenting all kinds of lesions: soft and hard exudates, hemorrhages and microanurysms. It comprised four main elements: two multi-scale modules, ASPP at the neck of the network and SAT unit after the decoder of the network, MobileNet backbone encoder, and modified UNet decoder block using several GSCs. It showed a competitive performance in different fundus image datasets. The results of this study have been published in the following papers:

4. Mohammed Yousef Salem Ali, Mohammed Jabreel, Aida Valls, Marc Baget Bernaldiz, and Mohamed Abdel-Nasser. ***"LezioSeg: Segmenting Eye Lesions in Fundus Images Utilizing Deep CNN with Hybrid Multi-Scale Attention Modules"***, Biomedical Signal Processing and Control. Impact Factor: 5.07, **(Q2)**. (Submitted).
5. Mohammed Yousef Salem Ali, Mohamed Abdel-Nasser, Mohammed Jabreel, Aida Valls, Marc Baget Bernaldiz, ***"Exu-Eye: Retinal Exudates Segmentation based on Multi-Scale Modules and Gated Skip Connection"***, 5^{th} IEEE International Conference on Multimedia, signal processing and communication and Technologies (IMPACT-2022), To appear, November 2022.
6. Mohammed Yousef Salem Ali, Mohamed Abdel-Nasser, Aida Valls, Marc Baget Bernaldiz, Mohammed Jabreel ***"EDBNet: Efficient Dual-Decoder Boosted Network for Eye Retinal Exudates Segmentation"***, 24^{th} International Conference of the Catalan Association for Artificial Intelligence (CCIA 2022), pp. 308-317, IOS Press Book, 2022.

1.5 Thesis Organization

The thesis contains 7 chapters. Below, we briefly describe the work done in each chapter:

- Chapter 2 presents concepts related to various aspects of fundus image analysis, particularly in the field of ophthalmology. The chapter reviews the methods and algorithms used to detect retinal abnormalities and segment fundus images. It also proposes the fundus datasets and evaluation metrics used in the thesis.
- Chapter 3 introduces a CAD system to segment the OD from the eye fundus image. This chapter presents the deep learning models and policies of the OWA operator's ensemble method of optic disc segmentation, the results of different ways of using ensemble methods for segmenting OD on the SJR dataset, and conclusions.
- Chapter 4 proposes a fully automated yet efficient glaucoma prediction CAD system in retinal fundus images. It proposes OD ROI detection using deep GSCs, glaucoma classifiers based on transfer learning, aggregation with different policies of OWA, WPM, and EXM, and glaucoma detection results of three publicly datasets DRISHTI-GS1, RIM-ONE, and REFUGE, and conclusions.
- Chapter 5 presents two novel methods of exudate lesions segmentation. First, presents the study of the dual-decoder boosted network method; second, the study of the two multi-scale modules method, the results of the baseline models and the two methods on the IDRiD publicly dataset, and the conclusions.
- Chapter 6 proposes an integrated retinal lesion segmentation model to segment the four kinds of retinal lesions (SE, EX, HE, and MA) from fundus images. It shows the detail of model contents, results of experiments on the IDRiD, E-ophtha, and DDR datasets, and conclusions.

- Chapter 7 presents the thesis concluding remarks and some future research lines.

CHAPTER 2

Background

2.1 Deep Learning

Deep learning is a machine learning (ML) technique to learn complex models from large-scale data based on Neural Networks. Machine learning comprises algorithms that enable a computer to determine significant patterns from data without human intervention. With the rapid advancement of computational power and the availability of large amounts of data, deep learning has become a good machine learning approach. It can determine significantly higher complex patterns, which the traditional machine learning methods cannot handle [10]. As a result, deep learning can perform a better generalization, improving decision-making.

Deep learning has been a great and robust way to foster Artificial Intelligence in recent years. It has performed remarkable or superior human-level performance on computer vision tasks [11] like classification, localization, and semantic segmentation. Moreover, it is efficiently applied to image data analysis in many disciplines, especially Health Care. Convolutional Neural Networks (CNNs) are the most common deep learning technique for image processing, classification, and segmentation. CNN is a neural network that depends on one or more convolutional layers. A convolution essentially slides a filter over the input to look at smaller portions of the image to find certain features [10]. It is a supervised method that requires a large set of annotated data to learn from. CNNs are commonly implemented as a Unet shape with an encoder-decoder network style. Unets are used in situations where the output must be of a similar size to the input, such as fundus image segmentation. Figure 2.1 shows Unet architecture [12].

2.1.1 Encoder-Decoder Networks

In this architecture, we have two blocks. The CNN deep learning encoder block acts as the feature extractor and learns an abstract representation of the input

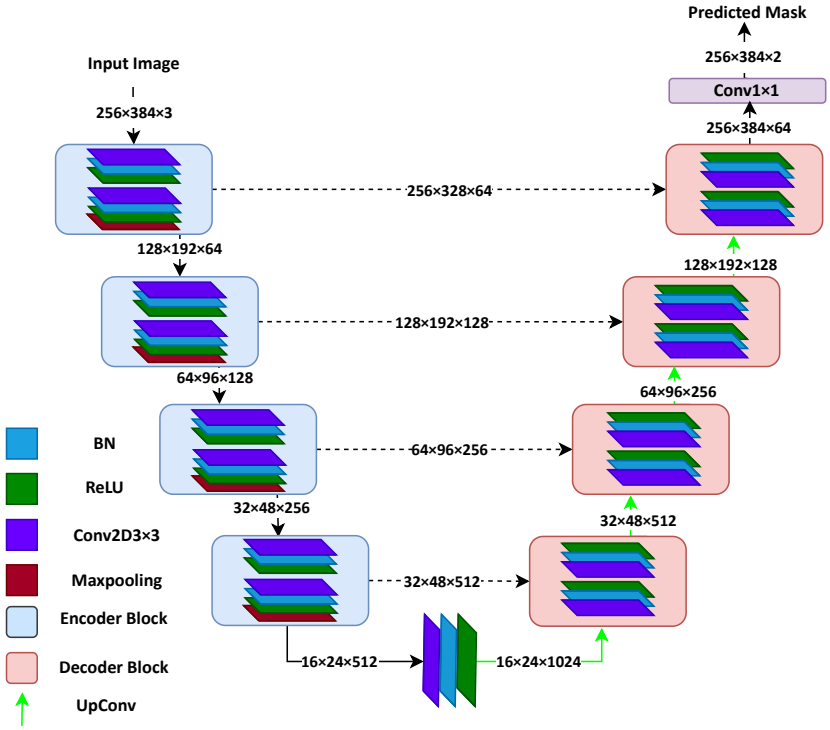


Figure 2.1: Unet architecture.

eye fundus image through a sequence of the encoder layers. Each encoder layers consists of one or more convolutions depending on the backbone architecture, such as VGGNets [13], ResNets [14], [15], and MobileNets [16], where each convolution layer is followed by rectified linear unit (ReLU) activation function. The ReLU activation function introduces non-linearity into the network, which helps better generalize the training data. In addition, the output of the ReLU acts as a skip connection for the corresponding decoder layer. Usually, next is followed by a 2x2 max-pooling or another technique, where the feature maps' spatial dimensions (height and width) are reduced by half. Again, this reduces the number of trainable parameters and computational costs.

The decoder block uses the encoder layers' abstract representation to generate a semantic segmentation mask of the input image. The decoder block begins with a transpose convolution of 2x2 in each layer. The skip connection feature map is then concatenated with the encoder layers' corresponding. Skip connections provide

features from previous layers that are sometimes lost due to network depth. Then, one or more convolutions are used, each followed by an activation function, such as a ReLU activation function. The last decoder layer output is passed through a 1x1 convolution with a sigmoid activation function for binary classification or softmax for multiple classifications. This activation function returns the pixel-wise classification mask.

2.1.2 Transfer Learning

Deep learning has evidenced remarkable success in a wide range of computer vision tasks, particularly with a large amount of labeled training data [17]. Nevertheless, sometimes we lack sufficient training data in the target task [18], such as annotated images of eye lesions to detect DR.

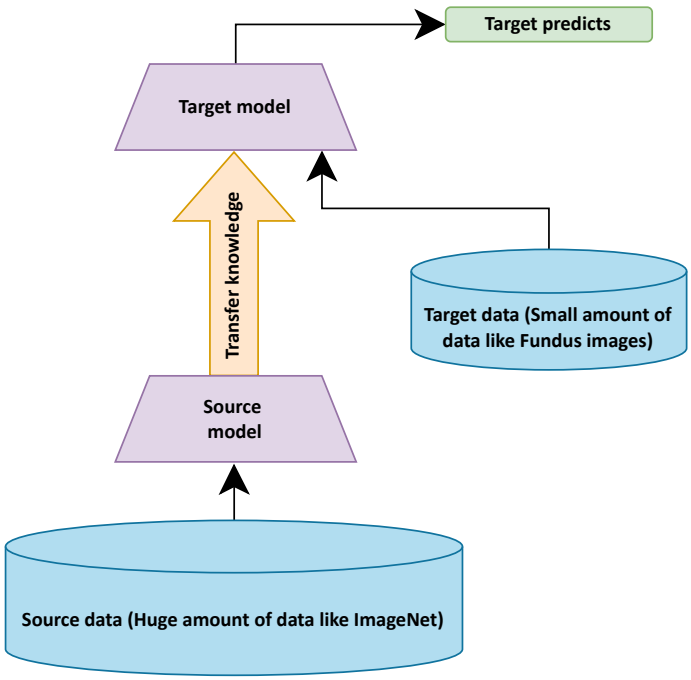


Figure 2.2: Transfer learning.

In this case, transfer learning assists us in transferring knowledge from a related source task. Furthermore, transfer learning has accelerated computer vision

development by allowing us to utilize previously trained models on large datasets for our tasks. Figure 2.2 illustrates the steps of transfer learning. As a result, the cost of training pre-trained models is reduced than training from scratch, and we can be confident in the datasets' quality because they have been verified [19]. The most familiar strategy for knowledge transfer in the context of deep learning is fine-tuning a pre-trained network model such as ImageNet on a new dataset. It starts with a model that has already been trained on the source task, such as VGGNets, ResNets, and MobileNets, and then trains them on the target task, such as optic disc and retinal lesions segmentation, which can notably improve performance [20]. The models used in this thesis will be explained in the corresponding chapters.

There are many types of deep transfer learning, such as domain adaptation, domain confusion, multitask learning, few-shot and zero-shot learning:

- **Domain adaptation learning** is commonly used when the marginal probabilities of the source and target domains differ. The data distribution of the origin and target domains is shifted or drifted, requiring adjustments to transfer the learning [21].
- **Domain confusion learning** is different layers in a deep learning network that capture different sets of features due to domain confusion. It can be used to learn domain-invariant features and improve their cross-domain transferability. Instead of letting the model learn any representation, we encourage both domains' representations to be as similar as possible [22].
- **Multitask learning** aims to leverage helpful information in multiple learning tasks to help learn more accurately for each task. Based on the assumption that all the tasks, or at least a subset, are related, jointly learning multiple tasks is empirically and theoretically found to lead to better performance than learning them independently [23].
- **Few-shot learning** is a type of transfer learning that aims to learn information

concerning object classes from one, or only a few, training images [24].

- **Zero-shot learning** aims to train a model that can classify objects from unseen classes (target domain) by transferring knowledge from other seen classes (source domain) using semantic information. Sometimes relying on no labeled examples to learn a task [25].

2.2 Aggregation Operators

This thesis will study the use of ensembles that use multiple models to make the image analysis. The fusion of the information will be done utilizing aggregation operators. From the different available methods, we have focused on fuzzy operators that permit establishing a certain degree of andness/orness when performing the aggregation of the input values. In particular, we have studied three well-known operators:

- OWA: Ordered Weighted Average
- WPM: Weighted Power Mean
- EXM: Exponential Mean

In [26; 27], we can find a formalization in terms of andness-directed aggregation, which permits finding the appropriate set of weights dynamically from a value that indicates the degree of simultaneity desired (i.e. tradeoff between conjunction and disjunction). Below, we introduce OWA, WPM, and EXM in detail.

2.2.1 OWA-based Aggregation Function

An OWA operator performs a weighted average of a vector of input values $X = (x_1, x_2, ...x_m)$, but before averaging, the values are sorted from the best to the worst (denoted with σ) [28]. In that way, the weighting vector W is associated with the importance of each position in the ordered input vector, regardless of the source of

the value in that position. This is expressed in Equation 2.1. Notice that the sum of weights must be equal to 1.

$$F_W\left(x_1, x_2, \ldots, x_m\right) = \sum_{k=1}^{m} w_k x_{\sigma_k} \tag{2.1}$$

The key property of this operator is that the weights can model different aggregation policies, ranging from situations of full andness (when $w_m = 1$ and rest are 0) to full orness (when $w_1 = 1$ and rest are 0). Any possibility in between can be represented with an appropriate combination of the weights.

Manually setting the weights for a specific policy may be difficult. One usual way of generating the weighting vector automatically is through Regular Increasing Monotone fuzzy linguistic quantifiers [28]. Each weight is then obtained using the expression in Equation 2.2, where $Q(z)$ is a function that corresponds to a fuzzy quantifier. This quantifier can generate weights representing different "quantities" of agreement, which is a factor in determining the degree of andness/orness of the aggregation.

$$w_k = Q\left(\frac{k}{m}\right) - Q\left(\frac{k-1}{m}\right) \tag{2.2}$$

Three common aggregation strategies that can be defined using a fuzzy quantifier are the following:

1. **At least half**—This policy is more tolerant and focuses only on a small subset of the best values aggregated. It corresponds to a situation of hard disjunction. The *at least half* policy, Q^1, can be expressed as follows:

$$Q^1(z) = \begin{cases} 2z & \text{if } 0 \leq z \leq 0.5 \\ 1 & \text{if } 0.5 < z \leq 1 \end{cases} \tag{2.3}$$

2. **As many as possible**—It gives the most important to the worst values

aggregated (quite strict in the aggregation). It represents a situation of hard conjunction. The *as many as possible* policies, Q^2, can be formulated as follows:

$$Q^2(z) = \begin{cases} 0 & \text{if } 0 \leq z \leq 0.5 \\ 2z - 1.0 & \text{if } 0.5 < z \leq 1 \end{cases} \tag{2.4}$$

3. **Most**—This policy represents a case of finding the majority value. It corresponds to a situation of soft conjunction (close to neutrality). The *most* policy, Q^3, can be expressed as follows:

$$Q^3(z) = \begin{cases} 0 & \text{if } 0 \leq z \leq 0.3 \\ 2(z - 0.3) & \text{if } 0.3 < z \leq 0.8 \\ 1 & \text{if } 0.8 < z \leq 1 \end{cases} \tag{2.5}$$

These three quantifiers are the ones that have been used for aggregating the probability values obtained with the different deep-learning models for each class. The aggregation is applied at each pixel of the image studied. A threshold of 0.5 is applied to the result, and the class with the maximum activation is taken as a label.

2.2.2 Weighted Power and Exponential Means

Given m arguments $p = (p_1, p_2, ...p_m)$, WPM and EXM are two aggregation operators that calculate a mean value for p. These operators have a parameter that allows tuning the aggregation polarity (conjunctive/disjunctive) by fixing the value of andness in $0 \leq \alpha \leq 1$. These operators are defined as follows [26; 29]:

$$A_{WPM}(p_1, \ldots, p_m; W, \alpha) = \left(w_1 * p_1^{r_m(\alpha)} + \ldots + w_m * p_m^{r_m(\alpha)} \right)^{1/r_m(\alpha)} \tag{2.6}$$

$$A_{EXM}\left(p_1,\ldots,p_m;W,\alpha\right)=\frac{1}{t_m(\alpha)}\ln\left(\sum_{i=1}^{m}w_i*\exp(t_m(\alpha)p_i)\right) \tag{2.7}$$

where $p_i \in [0,1]$, $i=(1,\ldots,m)$, $m\geq 2$, W is a vector of weights where $w_i=(1/m)$, and $r_m(\alpha)$ and $t_m(\alpha)$ are obtained from tables in [1].

It is worth noting that an $\alpha \geq 0.5$ determines the threshold between hard and soft conjunction, whereas $\alpha < 0.5$ is the threshold for hard/soft disjunction. An α value of 0.75 represents medium partial conjunction. Hard and soft conjunction properties are defined as follows:

$$\begin{aligned} &A\left(p_1,\ldots,p_m;W,\alpha\right)=0, \quad 0.75\leq\alpha\leq 1, \quad \exists p_i=0, \quad \text{(hard conjunction)}\\ &A\left(p_1,\ldots,p_m;W,\alpha\right)>0, \quad 0.5<\alpha<0.75, \quad \exists p_i>0, \quad \text{(soft conjunction)} \end{aligned} \tag{2.8}$$

In turn, the case of $\alpha = 0.25$ is the medium partial disjunction point. Hard and soft disjunction properties are defined as follows:

$$\begin{aligned} &A\left(p_1,\ldots,p_m;W,\alpha\right)>0, \quad 0.25<\alpha<0.5, \quad \exists p_i>0, \quad \text{(hard disconjunction)}\\ &A\left(p_1,\ldots,p_m;W,\alpha\right)=1, \quad 0\leq\alpha\leq 0.25, \quad \exists p_i=1, \quad \text{(soft disconjunction)} \end{aligned} \tag{2.9}$$

2.3 Fundus Image Datasets

This thesis has done experiments with public datasets of labeled images for glaucoma detection and optic disc and lesions segmentation. The used datasets are the standard ones used in related work so that we can compare the performance of the new methods with previous ones. In addition, for the case of optic disc and segmentation, we have been able to work with a private dataset collected from Hospital Sant Joan de Reus (SJR). They are now described in more detail.

For the first task, optic disc segmentation and glaucoma prediction, 4 fundus image datasets were used (see Table 2.1).

Table 2.1: Optic disc and glaucoma datasets description.

Dataset	Count of Images			Image Size
	Glaucoma	Healthy	Total	
REFUGE	120	1080	1200	1634 × 1634
DRISHTI-GS1	70	31	101	2048 × 1760
RIM-ONE-r2	200	255	455	Not fixed
SJR	-	-	105	3008 × 2000

- The DRISHTI-GS1 database [30] was collected at Aravind eye hospital, Madurai, from visitors to the hospital. In 2014, it was publicly released. The selected glaucoma patients were 40-80 years of age, and clinical investigators selected them based on clinical findings during examination. DRISHTI-GS1 images were taken with the eyes dilated and centered on the OD with a Field of View of 30 degrees. In addition, manual segmentations for each image were collected for OD and cup region from four different human experts with varying clinical experience (experts with experience of 3, 5, 9, and 20 years). The DRISHTI-GS1 dataset consists of 101 images with an original image size

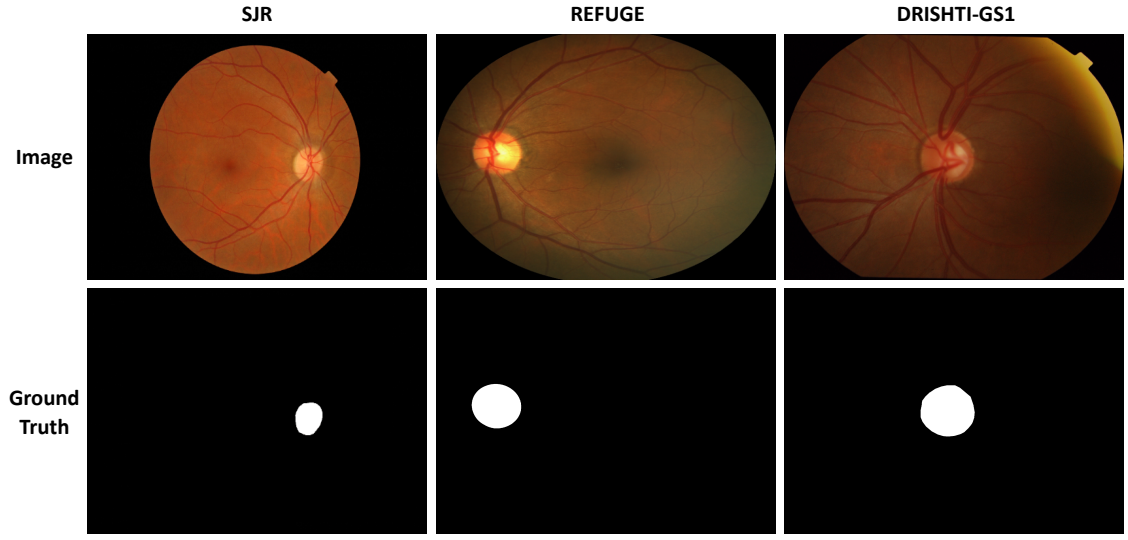


Figure 2.3: Examples of images and masks for optic disc segmentation in 3 datasets.

of 2048 × 1760 pixels and PNG uncompressed image format. A total of 50 images were used for training, and the rest 51 were used for testing for optic

disc segmentation and glaucoma prediction. The partitions of the dataset are standard as it is provided by the source providers, which facilitates the comparison with other works.

- The REFUGE database [31] was released in the REFUGE challenge 2018. The database consists of 1200 images stored in JPEG format, acquired by ophthalmologists or technicians from patients using one of two devices: a Zeiss Visucam 500 fundus camera with a resolution of 2124 × 2056 pixels (400 images) and a Canon CR-2 device with a resolution of 1634 × 1634 pixels (800 images). The images are centered at the posterior pole, with both the macula and the optic disc visible, to allow the assessment of the optic nerve head and potential retinal nerve fiber layer defects. These pictures correspond to Chinese patients visiting eye clinics from multiple sources, including hospitals and clinical studies. Only high-quality images were selected to ensure proper labeling. Manual annotations of the OD and the OC were provided by seven independent glaucoma specialists from the Zhongshan Ophthalmic Center (Sun Yat-sen University, China) with experience in the field (ranging from 5 to 10 years). In addition, a senior specialist with more than 10 years of experience in glaucoma obtained the single segmentation per image by taking the majority voting of the annotations of the seven experts, stored in JPEG format. Each image in the REFUGE data set includes a reference, trustworthy glaucomatous / non-glaucomatous label. The partitioning is standard and given by the source providers: 800 images are used for training and 400 for testing. Some examples are given in Figure 2.3.
- The RIM-ONE-r2 database [32] was publicly released in 2014 for glaucoma as a part of a research project developed in collaboration between three Spanish hospitals: Hospital Universitario de Canarias, Hospital Clínico San Carlos, and Hospital Universitario Miguel Servet. The aim of the project is the design an

automated software system for supporting glaucoma diagnosis. RIMONE-r2 consists of 455 cropped images of OD ROI of different sizes, of which 200 are glaucoma and 255 are normal. These images are obtained from full fundus images and have different resolutions. This dataset comes from the Medical Image Analysis Group and is available online. This thesis used the RIM-ONE-r2 dataset for glaucoma prediction, and It has been randomly split into 80% for training and 20% for testing.

- The private SJR database is used for optic disc segmentation. It includes 105 images that ophthalmologists manually labeled in the hospital Sant Joan de Reus (Catalonia, Spain). The original size of the images is 3008×2000 pixels. Images were captured using Topcon TRC-NW8F non-mydriatic fundus camera at 45°. All images were manually annotated of the OD by a senior ophthalmologist from the Hospital Sant Joan of Reus (Catalonia, Spain) and stored as binary images. Images of the dataset were randomly divided into 3 subsets: training set (56 images), validation set (14 images), and testing set (35 images). Figure 2.4 shows examples from the datasets of glaucoma.

Figure 2.4: Examples from the datasets of glaucoma prediction.

For the second task of the thesis, lesions segmentation, we have taken three popular datasets that enable comparison with other works. Table 2.2 shows the details of lesions segmentation datasets. Figure 2.5 shows examples of images and masks from the three lesion segmentation databases.

Table 2.2: Eye lesions datasets description.

Dataset	EX	HE	MA	SE	Total	Image Size
IDRiD	81	80	81	40	81	4288×2848
E-ophtha	47	-	148	-	195	2544×1696
DDR	171	194	124	42	225	Not fixed

- The Indian Diabetic Retinopathy Image Dataset (IDRiD) database [33] was publicly released in the challenges of the IEEE International Symposium on Biomedical Imaging (ISBI-2018). IDRiD dataset created from real clinical exams acquired at an eye clinic in Nanded (M.S.), India. The retinal photographs of people affected by diabetes were captured with a focus on the macula using a Kowa $VX-10\alpha$ fundus camera. Before image acquisition, pupils of all subjects were dilated with one drop of tropicamide at 0.5% concentration. The captured images have a 50° field of view. Annotated at a pixel level of the retinal lesions. IDRiD comprises 81 high-resolution retinal

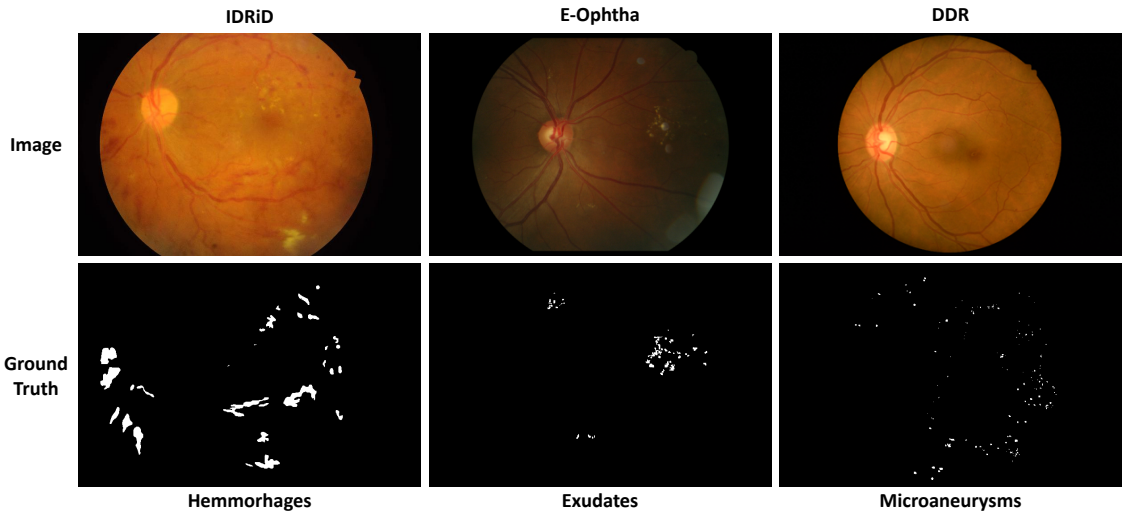


Figure 2.5: Examples from the datasets of lesions segmentation.

fundus images of eye lesions, jpg format of 4288×2848 pixels. Each image contains at least one binary mask labeled in the tif file format as one of four types of DR lesions: EX, SE, MA, and HE. The dataset providers have split it into 54 images as a training set and the rest of 27 as a testing set.

- The E-ophtha database [34] was gathered between 2008 and 2009 via the OPHDIAT network and released online in 2013. It is made of retinal images with different types of lesions (EX and MA) manually annotated by ophthalmology experts. The OPHDIAT network, located in the Ile-de-France region, comprises 16 screening centers: 11 centers in the diabetes departments of 11 hospitals, one diabetic retinopathy screening center in northern Paris, three in healthcare centers, and one in prison. The e-ophtha dataset introduced 47 annotated images of EX and 148 of MA. All images are stored in JPEG format of 2544×1696 pixels. It randomly split the dataset into 80% of images as a training set and 20% as a testing set and used the dataset for lesions segmentation.
- The DDR database [35] was provided a 757 color fundus images annotated for four DR lesions, including MA, EX, SE, and HE, released online in 2019 and collected from 147 hospitals between 2016 and 2018, covering 23 provinces in China. All images were provided by 9598 patients whose ages ranged from 1 to 100, with an average age of 54.13. Among these images, 48.23% are from male patients, and 51.77% are female patients. It was captured using 42 types of fundus cameras with a 45° field of view, mainly Topcon D7000, Topcon TRC NW48, Nikon D5200, and Canon CR 2 cameras. Six annotators performed the pixel-level annotations using special software. The annotators focus on whether lesions and boundaries are correctly identified. The annotators manually annotate lesions not identified by the annotation tool and remove false lesions. After an image is annotated, it is saved as a MAT file and a color image; 4 colors can distinguish four lesions. Finally, the ground truth for each lesion is extracted from the MAT files. This thesis used the 225 testing set images and ground truth for lesions segmentation, including MA, EX, SE, and HE.

2.4 Evaluation Metrics

In automatic medical image analysis, for detecting a region of interest (ROI) (e.g. landmark or abnormality), each pixel is classified by the CAD system into two classes: presence or absence of ROI. Then, the class obtained is compared to the knowledge provided by a human expert who has marked each pixel with a mask as positive or negative. Hence, to make a quantitative and objective evaluation of the performance of the methods proposed in this thesis, the results achieved in the experiments are evaluated with many metrics commonly used in the detection and segmentation tasks. These metrics can be formally defined in terms of four values: TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives):

1. **Accuracy (ACC)**: It is a metric that generally describes how the model performs across all classes. It is computed as the ratio between the number of correct predictions to the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.10}$$

2. **Sensitivity (Sen)/Recall (Re)**: It denotes the proportion of real landmark or abnormalities pixels (ROI) that are correctly detected, that is, that are classified as landmark or abnormalities pixels. Formally it is defined as the following:

$$Recall = \frac{TP}{TP + FN} \tag{2.11}$$

3. **Precision (Pre)**: It refers to the number of correctly classified pixels for landmarks or abnormalities divided by the total number of positive predictions. Formally it is defined as the following:

$$Precision = \frac{TP}{TP + FP} \tag{2.12}$$

4. **Specificity (Spe)**: It refers to the probability of a negative classification, conditioned on truly being negative.

$$Specificity = \frac{TN}{TN + FP} \tag{2.13}$$

5. **F1-score**: It stands for the harmonic mean of precision and recall. It can be expressed as follows:

$$F1 = 2 \cdot \frac{Precision * Recall}{Precision + Recall} \tag{2.14}$$

6. **Area Under Precision-Recall curve (AUPR)**: This curve shows the tradeoff between precision and recalls for different threshold values. The high area under the curve represents high recall and precision, indicating good performance. It is known to be a realistic measure for lesion segmentation performance like lesions [36].

In these expressions, TP refers to the true positive (the pixels were labeled as foreground, i.e., retinal lesion pixels, and correctly classified). Then, FP means false positive (the pixels were labeled as background and misclassified as foreground). Also, TN is a true negative, referring to the network's healthy pixels correctly classified. At the same time, FN is a false negative representing the lesion pixels misclassified as healthy pixels. Some other standard evaluation measures in image segmentation are based on comparing two regions, A and B. Where A corresponds to the experts' ground truth pixels of ROI, and B indicates the pixels that belong to the predicted mask (i.e. the region obtained as the output of segmentation).

7. **Intersection-Over-Union (IOU)/Jaccard**: It uses to check if there each landmark or abnormalities detector is over or under-segmentation of landmark or abnormalities regions [37]. The IOU is the intersection ratio between the

two masks concerning their union.

$$IOU = \frac{|A \cap B|}{|A \cup B|} \tag{2.15}$$

8. **Dice Coefficient**: It is 2 times the overlap between the ground truth and the obtained mask, divided by the total number of pixels in both images. It can be expressed as follows [37]:

$$Dice = \frac{2 * |A \cap B|}{|A| + |B|} \tag{2.16}$$

2.5 Chapter summary

This chapter introduced background concepts related to the thesis, such as deep learning and its role in health care, encoder-decoder networks, transfer learning types, and different aggregation operators. It also provides an overview of image datasets used in this thesis for optic disc segmentation, glaucoma prediction, and retinal lesion segmentation. Finally, the evaluation metrics commonly used for prediction and segmentation tasks have been defined.

CHAPTER 3

Optic Disc Segmentation Utilizing Deep Learning Networks and OWA Operators

3.1 Introduction

Retinal diseases like diabetic retinopathy and glaucoma highly affect the area of the optic disc of the human eye, also called the optic nerve head. To make the diagnosis, ophthalmologists inspect eye fundus images to detect the signs of such diseases in or near the optic disc (OD). For this reason, the automatic detection of the OD is relevant to improve the subsequent analysis of the image. Figure 3.1 shows the sample of fundus image and the ground truth of optic disc [33]. Several deep learning-based automated systems have recently been proposed

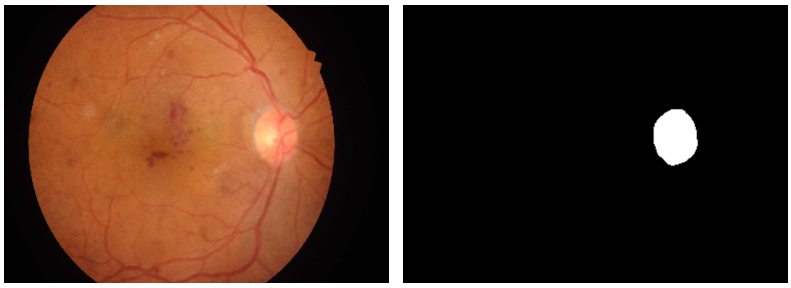


Figure 3.1: Example of fundus image (left) and optic disc ground truth (right).

to segment the optic disc automatically. Most use CNNs to automatically learn representative and high-level features from the input fundus images to achieve accurate segmentation [38]. For instance, the authors of [39] modified the U-Net model by reducing the size of the filters in all CNN layers to segment the optic disc directly. They stated that this modification produced a lighter model and improved the segmentation performance. With the DRIONS-DB dataset, they achieved IOU and Dice scores of 89 and 94%, respectively. With RIM-ONE v.3, they obtained IOU and Dice scores of 89 and 95%, respectively. The authors of [40] modified the U-Net and DeepLabv3+ models by inserting an attention module between the encoder and decoder networks and adding a conditional random field layer in the output layer. The method was evaluated using DRIONS-DB, RIM-ONE v.3, and DRISHTI-GS fundus image datasets. They achieved a 95% Dice and 91% Jaccard

with DRIONS-DB, a 97% Dice and a 94% Jaccard with RIM-ONE, and a 96% Dice and a 92% Jaccard with DRISHTI-GS.

In [41], a multi-label deep neural network called M-Net was proposed jointly for segmenting the optic disc and optic cup. M-Net includes a multi-scale input layer, U-shape CNN, a side-output layer, and a multi-label loss function based on the dice loss. The side-output layer works as an initial classifier to provide a local prediction map for different scale layers. M-Net achieved an accuracy of 98.30% with the ORIGA dataset. The study confirmed that the joint accurate segmentation process of OD and optic cup is essential to glaucoma detection and diagnosis. The authors of [42] leveraged multi-task learning and dense connections to jointly segment the optic disc and cup. They evaluated their work on three datasets: DRISHTI-GS, RIM-ONE, and REFUGE. With the OD segmentation task, they achieved a 91.83% Jaccard and a 95.97% Dice with DRISHTI-GS, a 91.01% Jaccard and a 95.82% Dice with RIM-ONE, and an 88.37% Jaccard with REFUGE.

Although the above methods achieved good OD segmentation accuracy, no individual model performs best under all conditions. For this reason, some works perform a fusion of the predictions of individual models in order to improve the segmentation accuracy [43]. In [44], a CNN model was trained using probability masks instead of binary masks to segment the optic disc and optic cup. Each probability mask was obtained by fusing segmentation masks made by multiple experts. In [45], an ensemble of different CNN architectures for medical image classification is used. The study demonstrated that different CNNs learn different levels of semantic image representation; therefore, an ensemble of CNNs produces richer features. In paper [46], authors merge the predicted labels of five CNNs models (Resnet50, Inceptionv3, Xception, Dense121, and Dense169) to improve the results of DR classification in fundus images.

However, most existing methods use fixed weights associated with each model

to construct the ensemble. This strategy may not be appropriate for all situations. In this chapter, we employ the OWA [28] to aggregate the information provided by several OD segmentation models. More specifically, this chapter proposes an ensemble-based OD segmentation system based on a set of deep learning models and an aggregation stage based on the OWA operator (see Section 2.2).

OWA is an aggregation operator that uses a set of weights to define the input's aggregation policy, which can vary from full conjunctiveness to full disjunctiveness. These weights do not indicate the importance of each source of information but the importance of the merged input values. The definition of the proper weights is a critical issue in this aggregation operator. A dynamic way of generating weights is through quantifier functions [28], which can be directed by an andness value. In that way, this OWA-based ensemble of deep learning models will allow the construction of a dynamic OD segmentation system.

3.2 Optic Disc Segmentation with a Dynamic Ensemble

The overall idea is to train several deep-learning segmentation models to develop a reliable and efficient OD segmentation system. For each new eye fundus image, it is fed into N individual OD segmentation models. Next, the OWA operator aggregates the different segmentation masks given by the models to produce the final segmentation mask (final prediction). Figure 3.2 shows the architecture of the procedure of this composed segmentation method for OD. The next sections will clarify the details of each stage.

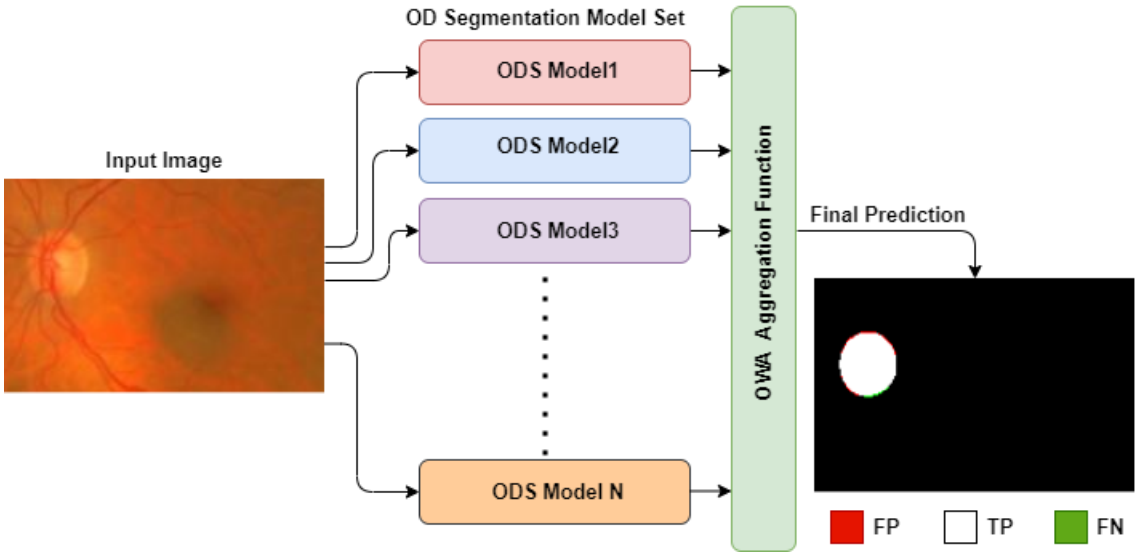


Figure 3.2: Structure of the proposed ensembling method for OD segmentation.

3.2.1 Constructing Individual OD Segmentation Models

In this study, we developed ten OD segmentation models based on different state-of-the-art deep learning-based semantic segmentation models. Specifically, we used the Unet, GSCs, DoubleU-Net, DeepLabV3+, CGNet, ERFNet, SegNet, ESNet, LinkNet, and SQNet models. Each of them was trained to maximize the segmentation accuracy. The architecture of all models is based on the encoding and decoding method for work except the CGNet. The binary cross-entropy loss function and ADAM optimizer were used to train all models. We briefly introduce each model below.

- **Unet** [12]—The developed Unet model has five blocks in the contracting path (downsampling path) and five blocks in expanding path (upsampling path). The Unet model contains a total of twenty-three convolutional layers. The implementation of Unet is available at https://github.com/jakeret/unet/blob/master/src/unet/unet.py.
- **Gated Skip Connections (GSCs)** [47]—Is a modified version of Unet. It had five encoder blocks with two convolutional layers, five

decoder blocks with four convolutional layers, a total of thirty-three convolutional layers, and a gated skip connection mechanism in each decoder block with a convolutional layer. The implementation of GSCs is available at https://github.com/Mohammed-Horbi/GSCs/blob/main/UnetGatedSkipConnection-UnetGSCs.py

- **DoubleU-Net** [48]—The employed DoubleU-Net model includes two cascaded Unets. The first Unet is based on a pre-trained VGG19, meaning there are four pre-trained VGG19 blocks in the encoder and decoder networks. The second Unet is a standard Unet[12] that includes four blocks in the encoder network, four blocks in the decoder network, and an atrous spatial pyramid pooling between the encoder and decoder networks. The implementation of DoubleU-Net is available at https://github.com/DebeshJha/2020-CBMS-DoubleU-Net.
- **DeepLabV3+** [49]—The DeepLabV3+ based ROI detection model employs the ResNet50 backbone with deep separation and Atrous separable convolution blocks. The implementation of DeepLabV3+ is available at https://github.com/srihari-humbarwadi/DeepLabV3_Plus-Tensorflow2.0/blob/master/deeplab.py.
- **CGNet** [50]—It has a total of fifty-one convolutional layers with no decoder modules. CGNet involves three down-sampling stages with context-guided blocks in each stage. The implementation of CGNet is available at https://github.com/xiaoyufenfei/Efficient-Segmentation-Networks/blob/master/model.
- **ERFNet** [51]—The developed ERFNet model has 23 convolutional layers. It includes two types of residual layer design, namely bottleneck, and non-bottleneck. It uses skip connections between the encoder and decoder blocks. The implementation of ERFNet is available at https://github.com/

xiaoyufenfei/Efficient-Segmentation-Networks/blob/master/model.

- **SegNet** [52]—The encoder network of the SegNet-based OD ROI detection model uses a VGG16 backbone with 13 convolutional layers in four blocks. Each encoder layer has a corresponding decoder layer; hence, the decoder network has 13 deconvolutional layers. The encoder and decoder of SegNet (26 layers). The implementation of SegNet is available at https://github.com/xiaoyufenfei/Efficient-Segmentation-Networks/blob/master/model.
- **ESNet** [53]—It consists of four key components, namely the down-sampling unit, upsampling unit, factorized convolution unit and its parallel counterparts (a total of 18 convolutional layers). The factorized convolution unit uses the 1D factorized convolution in the residual layers, whereas the parallel version utilizes a transform-split transform-merge strategy in the residual modules. The implementation of ESNet is available at https://github.com/xiaoyufenfei/Efficient-Segmentation-Networks/blob/master/model.
- **LinkNet** [54]—There are four blocks in the encoder and decoder of LinkNet (26 layers). The encoder consists of residual blocks. The model also uses skip connections to bypass spatial information from the encoder blocks to the corresponding decoder blocks. The implementation of LinkNet is available at https://github.com/xiaoyufenfei/Efficient-Segmentation-Networks/blob/master/model.
- **SQNet** [55]—It has a SqueezeNet-like encoder followed by four parallel dilated convolutions, while the decoder includes SharpMask-like refinement modules. SQNet also employs Exponential-Linear-Units to avoid the bias shift. The implementation of SQNet is available at https://github.com/xiaoyufenfei/Efficient-Segmentation-Networks/blob/master/model.

3.2.2 OWA-based Aggregation of OD masks

For the aggregation step, we propose using the OWA operator defined by [28]. This operator performs a weighted average of a vector of input values $X = (x_1, x_2, ...x_m)$, but before averaging, the values are sorted from the best to the worst (denoted with σ). OWA is explained in Section 2.2. In this study, we will consider the three different quantifiers policy (Q^1, Q^2, and Q^3) presented in Section 2.2. They correspond to Hard Disjunction (Q^1), Hard Conjunction (Q^2), and Soft Conjunction (Q^3). For each of the pixels of the image, the output values (i.e. probability of belonging to OD) obtained from the 10 individual classifiers are averaged using OWA for each one of these 3 policies. A threshold of 0.5 is applied to the result, and the class with the maximum activation is taken as a label.

3.3 Experimental Results

This section summarizes the experiments carried out to evaluate the performance of the proposed model, including a description of the experimental setup and the analysis of the outcomes.

3.3.1 Experimental Setup

Here, we used an in-house fundus image dataset collected from Hospital Universitari Sant Joan de Reus. It includes 105 images that ophthalmologists manually labeled in SJR. The original size of the images is 3008×2000. We randomly divided the dataset into three subsets: the training set (56 images), the validation set (14 images), and the testing set (35 images). We used flip horizontal, flip vertical, and rotations data augmentation techniques to increase the training data set size by 20 times, generating a training set with 1120 images. Then, we resized the images and the ground truths to 384×256 to make the training process faster. The validation set was used to save

Table 3.1: Performance comparison on the SJR dataset.

Models	Evaluation Metrics (%)			
	IOU	**Dice**	**Pre.**	**Rec.**
ODS1: GSCs	**95.1**	**94.9**	97.0	92.9
ODS2: DoubleU-Net	94.4	94.2	93.3	**95.0**
ODS3: DeepLabV3+	94.1	93.8	97.7	90.1
ODS4: U-Net	92.4	91.8	96.6	87.5
ODS5: LinkNet	92.4	91.8	99.7	85.3
ODS6: SQNet	91.9	91.2	97.6	84.8
ODS7: ESNet	91.9	91.2	99.5	84.3
ODS8: CGNet	91.6	90.8	99.6	83.7
ODS9: ERFNet	91.5	90.7	**99.8**	83.4
ODS10: SegNet	90.5	89.4	99.3	81.8

the best checkpoint of the trained models. We employed the binary cross-entropy loss function in all models. We trained the models with 50 epochs using an Adam optimizer with a learning rate of 0.001 and a batch size of 4 images.

3.3.2 Results of Individual Models

We trained and tested 10 state-of-the-art deep CNN independently for OD image segmentation on the SJR dataset. The results of segmentation are shown in Table 3.1. We noted that the GSCs model has the best result of IOU and Dice. Then ERFNet gives the best result of the Precision metric. Also, the DoubleU-Net has the best results in the Recall metric. So, we can see no unique winner method in all quality indices. This fact motivates the idea of using an ensemble of a subset of these models. To illustrate the results, we displayed sample masks obtained for the three best IOUs OD segmentation models and the worst ones in Figure 3.3. Pixels in red are the FP, in green the FN, while white ones TP are the correctly classified.

3.3.3 Results of the OWA-based Ensemble Method

In this section, we compared the three OWA aggregation policies, defined in Section 2.2: Q^1 uses a disjunctive approach, considering only the models with the best

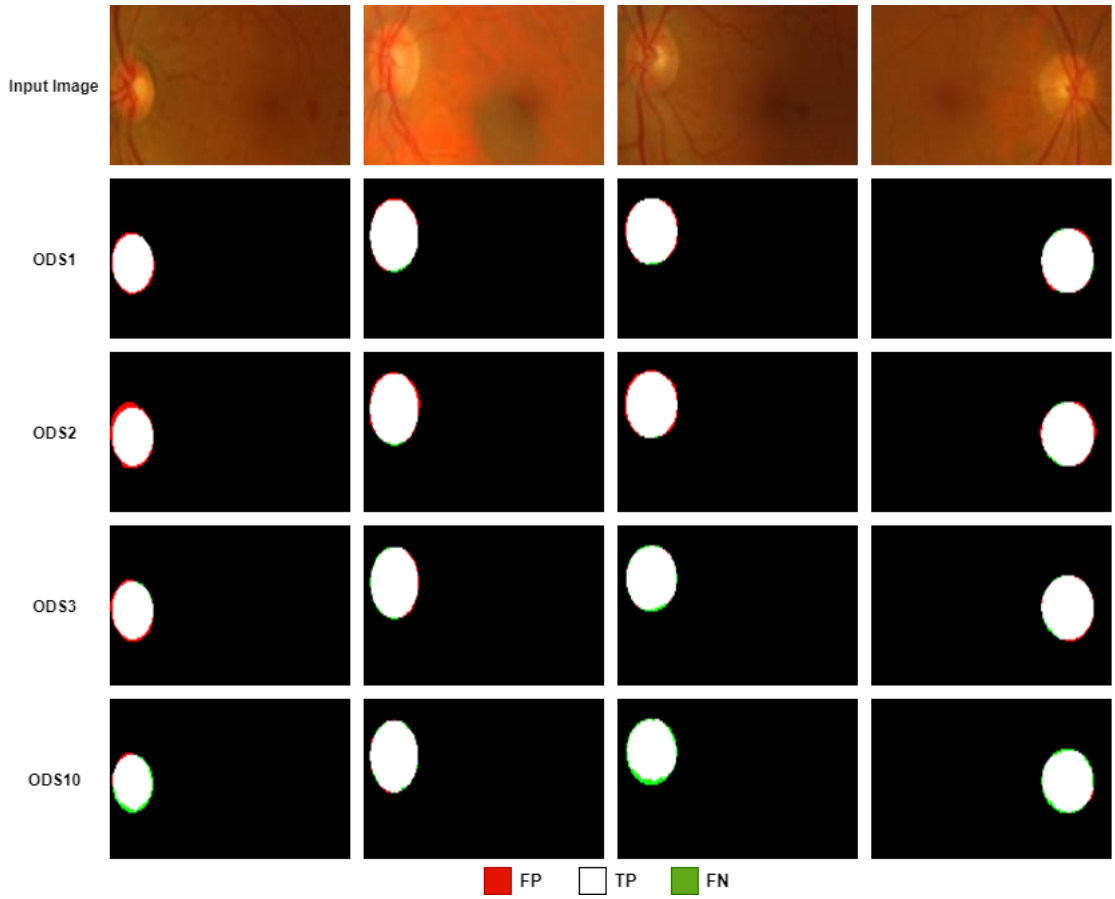


Figure 3.3: Results of the best three OD segmentation models and the worst one. Here ODS1, ODS2, ODS3, and ODS10 indicate to GSCs, DoubleU-Net, DeepLabV3+, and SegNet models.

performance, Q^2 applies a robust conjunctive approach, requiring that most models agree on the same output. In contrast, Q^3 takes an intermediate point between the previous ones.

We used different subsets of models to build the ensemble. Reducing the number of models is interesting to decrease the execution time. After ranking the deep learning models with the IOU index, we compared a set with the Top-7 models: GSCs, DoubleU-Net, DeepLabV3+, U-Net, LinkNet, SQNet, and ESNet; a set with the Top-5 models: GSCs, DoubleU-Net, DeepLabV3+, U-Net, and LinkNet; and with the Top-3 models: GSCs, DoubleU-Net, and DeepLabV3+. Results are shown in Tables 3.2, 3.3, 3.4. With the three aggregation policies, we can see that performance indicators are better when we reduce the number of models in the

Table 3.2: Results of evaluation metrics of At least half, Q^1 OWA policy with a different number of aggregated models.

Aggregated Models	IOU	Dice	Pre	Rec
10 Models	93.4	92.9	**99.6**	87.3
Top-7	94.1	93.8	99.2	89.2
Top-5	94.9	94.6	98.3	91.4
Top-3	**95.2**	**95.0**	97.4	**93.0**

ensemble. The best results are obtained with the Top-3 ensemble, which achieved an IOU of 95.4% and Dice of 95.1%. We observe that the IOU and Dice scores of the aggregated models are better than the ones of the individual models. However, precision is higher when using more models. An improvement in recall means that

Table 3.3: Results of evaluation metrics of As many as possible, Q^2 OWA policy with a different number of aggregated models.

Aggregated Models	IOU	Dice	Pre	Rec
10 Models	93.7	93.2	**99.5**	87.9
Top-7	94.4	94.1	99.0	89.8
Top-5	95.1	94.8	98.3	91.9
Top-3	**95.4**	**95.1**	96.7	**93.9**

we reduce the FNs. The non-improvement of precision means that the number of FPs is the same or increases slightly. However, as IOU and Dice are better, it means that the improvement in FNs is considerably larger than the increase in FPs.

Regarding OWA, we can also observe that the most conjunctive policy (i.e. as many

Table 3.4: Results of evaluation metrics of Most, Q^3 OWA policy with a different number of aggregated models.

Aggregated Models	IOU	Dice	Pre	Rec
10 Models	93.5	93.0	**99.6**	87.5
Top-7	94.3	94.0	99.1	89.6
Top-5	95.0	94.7	98.2	91.7
Top-3	**95.3**	**95.1**	97.2	**93.3**

as possible, Q^2) outperforms the others. Comparing the results with the Q^2 ensemble

system and the individual models, we observe an improvement in the performance of Table 3.1. In this experiment, results improve when using a strict conjunctive aggregation policy, which requires a high agreement of the models to choose the final class.

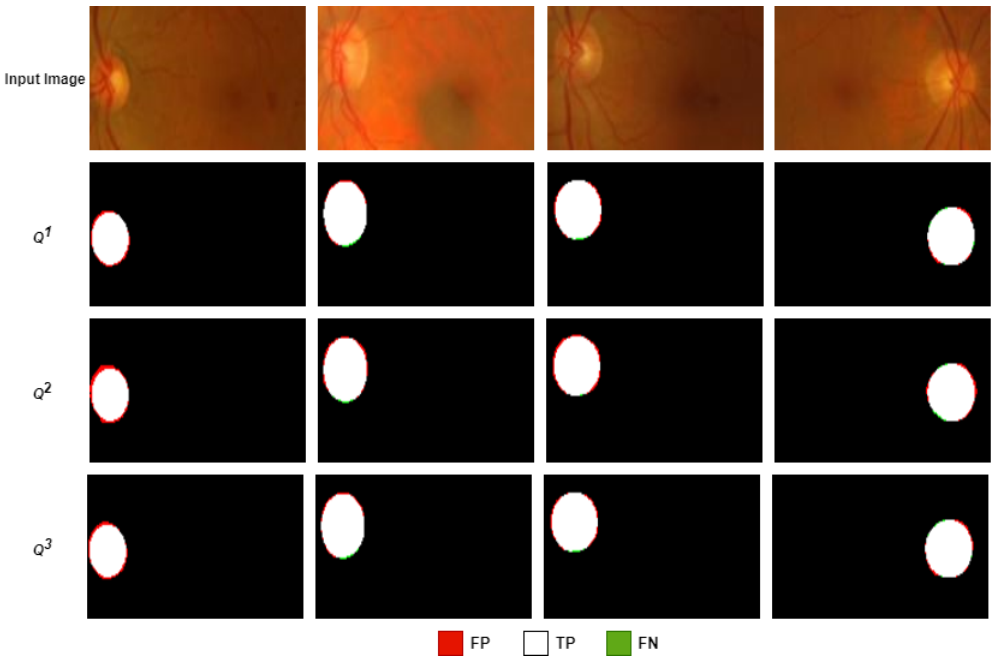


Figure 3.4: Results of the three policies of the OWA proposed method. Here Q1, Q2, and Q3 indicate the policies

As OWA is using the best-performing subset of n models, this aggregation considers the consensus of the n models regarding the class probability of each pixel (i.e. probability of each class given by each model). As this consensus may be difficult to find, it is reasonable that smaller subsets perform better than adding more models with fewer performance indicators. This observation is interesting because it is unnecessary to include many deep-learning models in the ensemble. From the improvement concerning individual models, we can conclude that the ensemble approach can improve the segmentation result, correcting some of the mistakes of individual models. Figure 3.4 displays the segmentation made with the three different OWA policies when using the Top-3 methods with the same images in the previous figure. It is interesting to determine the statistical significance of the differences in

performance between the proposed ensemble and GSCs (the best individual model) in terms of the IOU and Dice. To do so, we used Student's t-test (significance level < 0.05) to determine the difference in IOU and Dice values. The p-values obtained are higher than 0.05, indicating no statistical significance. However, with the OWA-based ensembling of the Top3 models and a Q^2 policy, we are able to increase the Recall when compared with a single GSCs model. This is a conjunctive policy, denoted with "as many as possible".

3.4 Conclusions

This chapter presented a deep learning-based system for segmenting the optic disc based on an ensemble of efficient semantic segmentation models for medical images. The aggregation of the different deep learning models was performed with OWA operators. We proposed using a dynamically generated set of weights that can give a distinct contribution to the models according to their performance during the optic disc segmentation in the eye fundus images. Reducing the number of models decreases the execution time and obtains the best result. The effectiveness of the proposed system was assessed on a fundus image dataset collected from the Hospital Sant Joan de Reus. We obtained Jaccard, Dice, Precision, and Recall scores of 95.40, 95.10, 96.70, and 93.90%, respectively.

CHAPTER 4

Glaucoma Detection Utilizing Transfer Learning and Fuzzy Aggregation Operators

4.1 Introduction

In this chapter, we will focus on the automatic detection of glaucoma disease. Early glaucoma detection and treatment are essential to avoid total vision loss for people over 60 years old [56]. Ophthalmologists inspect eye fundus images to detect the symptoms of glaucoma, especially to find a large cup-to-disc ratio. (CDR). Figure 4.1 shows optic disc and optic cup in fundus images of healthy (Left) and glaucoma cases (Right). In the two shown examples, the optic disc is marked by the blue dashed ellipse, while the green marks the optic cup dashed ellipse (the optic cup is a bright region inside the optic disc region). The measurement of the optic cup to optic disc ratio is used to assess glaucoma progression. In glaucoma severe cases, the optic cup region enlarges until it occupies most of the optic disc area. In other words, the size of the optic cup is much close to the size of the optic disc in severe glaucoma patients [57]. It should be noted that the standard optic cup to optic disc ratio value of normal cases is ≤ 0.4 [58]. Glaucoma patients require a regular examination to

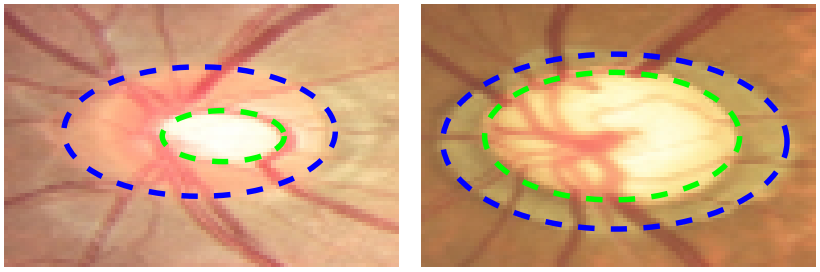


Figure 4.1: Examples of the optic disc and optic cup in fundus images of healthy (Left) and glaucoma cases (Right).

assess glaucoma progression, yielding a large number of fundus images. However, the manual analysis of many images is expensive in both time and effort. Besides, the precision of the diagnosis varies with the experience of ophthalmologists. Modern CAD systems can automatically analyze fundus images and provide diagnoses as accurately as ophthalmologists with many years of experience [59; 60].

Deep-learning technologies have been used for glaucoma detection, and two

categories can be distinguished: 1) systems based on optic cup-to-optic disc ratio calculation and 2) systems based on texture feature extraction. Some examples of the first case are Deep learning-based glaucoma detection systems such as [61; 41; 42] detect optic disc and optic cup and then calculate the optic cup to optic disc ratio. In turn, [62] focused on extracting the features of the optic disc and optic cup individually from the fundus images and then compared them with the optic cup to optic disc ratio results. Regarding approaches of the second type, deep feature extraction glaucoma detection systems include the study of [60], in which they used the whole fundus image as a region of interest (ROI) and employed five pre-trained deep CNNs for glaucoma. However, the most crucial ROI in the fundus image for glaucoma detection is the optic disc region, which also includes the optic cup region [57]. Therefore, some systems such as [63; 64] include two main steps: optic disc ROI detection and feature extraction. In the optic disc ROI detection step, a deep learning-based method automatically localizes and crops ROIs that include optic disc from the fundus images. Deep CNNs extract features from the extracted ROIs in the feature extraction step.

Some works use ensemble methods for improving performance. The typical approach is to use static ensemble techniques to aggregate the predictions of individual classifiers. For instance, [63] used the mean operator to aggregate the predictions of InceptionV3, XceptionNet, VGG16, VGG19, and ResNet50 CNN models. Indeed, the fusion of the predictions of individual models can improve the segmentation and classification accuracy [43; 65].

However, such ensemble techniques that use fixed weights associated with each model to construct the ensemble may not be suitable for practical situations because such fixed weights do not imply the importance of the predictions of glaucoma classifiers. It should be noted that the definition of proper weights is essential to construct a robust ensemble [28].

This chapter continues the work of the previous one and proposes an accurate glaucoma detection system, which includes three stages: 1) ROI detection (i.e., optic disc region) using a deep learning network, 2) construction of accurate individual glaucoma classifiers based on different CNNs and support vector machine classifiers, and 3) the use of fuzzy aggregation operators to dynamically fuse the predictions of the top-performing glaucoma classifiers.

The most significant novelty of this chapter is the use of parametrized fuzzy aggregation operators to build a dynamic ensemble, unlike existing glaucoma detection methods that employ a static ensemble of deep feature extractors. Specifically, three different aggregation operators (OWA, WPM, and EXM) are used to fuse the predictions of individual SVMs trained on the top of deep feature extractors.

It should be noted that these 3 operators permit the parametrization of the degree of compensation between high and low scores in the aggregation. In this case, the aggregated values are the activation levels of the individual glaucoma classifiers for a particular decision class. This study tested three different polarities (conjunctive/simultaneity/no compensation, disjunctive/replaceability/full compensation, and average=medium compensation). In this way, these aggregation procedures can adapt the weight given to each input classifier at each point, depending on the consistency among the conflicting glaucoma predictions of each SVM classifier. The chapter shows that this approach gives more accurate glaucoma detection results.

4.2 Related Work

Several deep learning-based CAD systems have recently been proposed to diagnose glaucoma automatically. The related work of glaucoma detection systems can

be categorized into the calculation of CDR, CNN-based glaucoma detection, and ensemble-based glaucoma detection. Below, we briefly discuss these methods' current state of the art.

4.2.1 CDR-based Glaucoma Detection

Table 4.1 presents and summarizes different CDR-based glaucoma detection related studies. [66] used trainable computer vision algorithms to estimate the vertical CDR (VCDR) value in retinal fundus images. They set the following rule for glaucoma detection: if the VCDR value is higher than or equal to 0.7, the fundus image is classified as glaucoma. Based on this rule, they achieved a mean VCDR difference of 0.11 and an AUC of 93.0% on a total of 850 test images from eight datasets. [31] summarized the REFUGE dataset challenge of glaucoma assessment from fundus image. The top-ranked team, CUHKMED, used an ensemble model for segmenting

Table 4.1: Summary of CDR-based glaucoma detection methods

Study	Result	Used Datasets	Year
Guo et al. [66]	AUC = 93.0%	850 images of Eight datasets	2019
Orlando et al. [31]	AUC = 98.9% ,Sen = 97.5%	REFUGE	2020
Hemelings et al. [67]	AUC = 94.0, 87.0%	Private UZL and REFUGE	2021
Harvella et al. [68]	AUC = 97.6%, 94.7%	REFUGE and DRISHTI-GS	2021

the OD and OC, and then they computed the VCDR of the two segmented ellipses of OD and OC. They achieved an AUC (i.e., area under the receiver operating characteristic curve) and sensitivity of 98.9 and 97.5%, respectively. [67] proposed a deep learning-based method for glaucoma detection by estimating VCDR. With a test dataset of 4765 images, they obtained 94.0 and 87.0% of AUC with UZL and REFUGEE datasets for glaucoma detection. [68] proposed a multi-task deep learning-based network for glaucoma classification directly and OD/OC segmentation to calculate the absolute difference between the predicted and ground truth VCDR. REFUGE and DRISHTI-GS datasets achieved AUC values of 97.6 and 94.7%,

respectively.

4.2.2 CNN-based Glaucoma Detection

Table 4.2 presents and summarizes different CNN-based glaucoma detection related studies. [69] employed a fine-tuned VGGNet CNN with an SVM classifier to predict glaucoma. Fundus images collected from Taipei Veterans General Hospital and DRISHTI-GS test dataset (187 images) were used for assessing the proposed method, achieving accuracy rates of 95.0% and 80.3%, respectively. [70] proposed a 13-layer deep CNNs for glaucoma diagnosis. With a total of 1113 fundus images (660 normal and 453 glaucomatous images), they obtained an accuracy of 93.9%. [60] fine-tuned five pre-trained deep CNNs, namely, standard CNN, VGG19, GoogleNet, ResNet50, and DENet, for glaucoma prediction and used the whole fundus image as an ROI. They obtained an AUC value of 94.0% with a test set containing 124 glaucoma and 455 normal images.

[62] aggregated the predictions of two networks. The first network estimates the CDR value from the localized OC and OD in the fundus images. The second network

Table 4.2: Summary of CNN-based glaucoma detection methods

Study	Result	Used Datasets	Year
Ko et al. [69]	ACC = 95.0%, 80.3%	Private dataset, Drishti-GS	2020
Ajitha et al. [70]	ACC = 93.9%	Mixed of four datasets	2021
Gomez Valverde et al. [60]	AUC = 94.0%	RIM-ONE, DRISHTI-GS, and ESPERANZA Private dataset	2019
Civit-Masot et al. [62]	ACC = 88%, AUC = 96%	Mixed of DRISHTI-GS1 and RIM-ONE	2020

extracts texture features from fundus images to predict glaucoma. With a mixed dataset from DRISHTI-GS1 and RIM-ONE, they obtained an accuracy of 88% and AUC of 96%.

4.2.3 Ensemble-based Glaucoma Detection

Table 4.3 presents and summarizes different Ensemble-based glaucoma detection related studies. [64] proposed a two-step glaucoma detection system. They employed a modified DeepLabv3+ architecture to detect OD in the first step. In the second step, they evaluated three glaucoma classification approaches: fine-tuning pre-trained CNNs, CNN features with SVM, and the ensemble of the best models of the previous approaches. They found that the ensemble model yields the best

Table 4.3: Summary of Ensemble-based glaucoma detection methods

Study	Result	Used Datasets	Year
Sreng et al. [64]	ACC = 97.37, 90.00, 86.84, 95.8 and 99.53%	RIM-ONE, ORIGA, DRISHTI-GS1, REFUGE, and ACRIMA	2020
Deepa et al. [65]	AUC = 98.9% ,Sen = 97.5%	Mixed of DRISHTI-GS, DRIONS-DB, and HRF datasets	2021
Diaz-Pinto [63]	AUC = 76.78, 83.54, 80.41, 85.75, and 77.39%	ACRIMA, HRF, DRISHTI-GS1, RIM-ONE, and sjchoi86-HRF	2019

detection results: accuracy rates of 97.37, 90.00, 86.84, 95.8, and 99.53% with RIM-ONE, ORIGA, DRISHTI-GS1, REFUGE, and ACRIMA datasets, respectively. [65] employed the maximum voting strategy to fuse the predictions of three fine-tuned models (GoogLeNet, VGGNet-16, and ResNet50). They tested their method on a mixture of DRISHTI-GS, DRIONS-DB, and HRF datasets, achieving accuracy rates of 88.96%. [63] presented an ensemble method by averaging five different CNN models (InceptionV3, XceptionNet, VGG16, VGG19, and ResNet50) to distinguish between glaucomatous and healthy eye fundus images. They evaluated their method on ACRIMA, HRF, DRISHTI-GS1, RIM-ONE, and sjchoi86-HRF datasets, obtaining AUC values of 76.78, 83.54, 80.41, 85.75, and 77.39%, respectively.

Most exiting ensemble-based glaucoma detection methods employ static weights to construct the ensemble model [64; 65; 63]. The use of aggregation strategies that employ static weights may not be appropriate for practical situations because they do not embed the importance of the predictions of individual glaucoma classifiers in

different cases.

4.3 Glaucoma Detection System

The architecture proposed in this chapter for glaucoma detection is illustrated in Figure 4.2. It involves three main stages: 1) detection of ROI, 2) building accurate individual glaucoma classifiers based on different deep CNNs and transfer learning approaches, and 3) constructing an efficient aggregation method based on fuzzy aggregation operators to merge the predictions of individual glaucoma classifiers dynamically. Below, we explain each stage in the proposed system in detail.

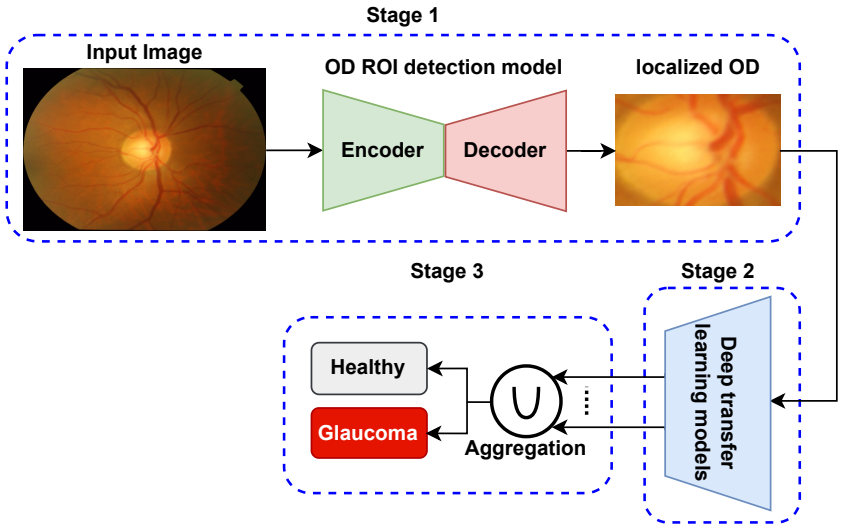


Figure 4.2: Workflow of the proposed glaucoma detection system.

4.3.1 Stage 1—ROI Detection

The OD region contains the essential characteristics for emphasizing glaucoma, so there is no need to use the whole eye fundus image in the glaucoma classification process. For this reason, the OD ROI detection stage is critical for glaucoma prediction and helps fasten the classification process. Therefore, an efficient OD ROI detection method is required to localize the OD region in the fundus images precisely.

This chapter proposes an OD ROI detection model based on deep GSCs [47]. As shown in Figure 4.3, the OD ROI detection model is Unet with a GSC mechanism in each decoder block. It has five encoder blocks with two convolutional layers and five decoder blocks with four convolutional layers. Figure 4.4 presents the GSCs

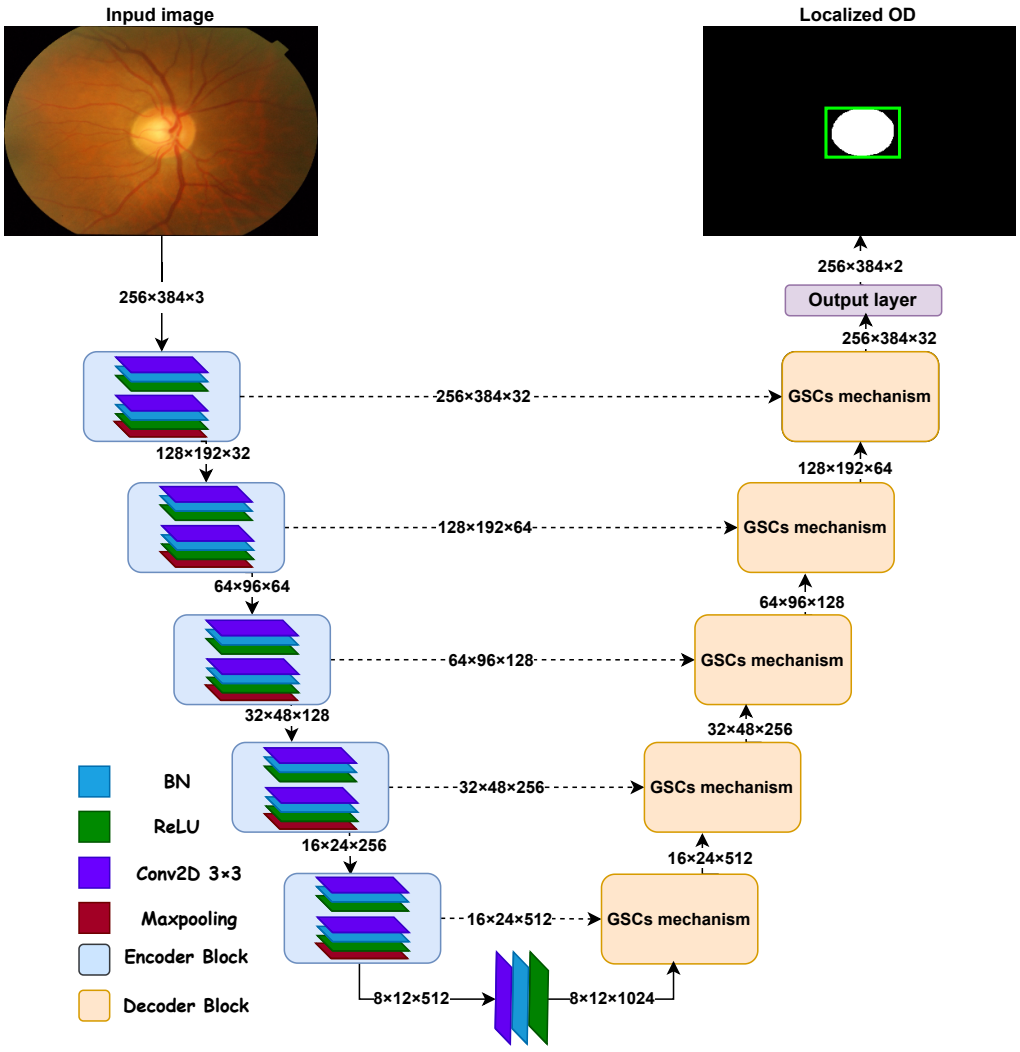


Figure 4.3: OD ROI detection model based on deep GSCs. BN stands for batch normalization.

mechanism of each decoder block. The GSCs mechanism receives feature maps from the corresponding encoder layers. It concatenates them with the feature maps produced by the previous block (either the bottleneck block or a decoder block). We express these feature maps as T_1, T_2. $T_1 \in \mathrm{R}^{h \times w \times f}$, $T_2 \in \mathrm{R}^{h/2 \times w/2 \times 2f}$. T_2 feeds to ConvTrans2D transposed convolution layer with a kernel size of 3×3 to produce

feature maps $\hat{T}_2$. $\hat{T}_2$ and T_1 should have equal sizes to perform the concatenation process as follows:

$$C = \varphi_{1\times1}([T_1||\hat{T}_2]) \tag{4.1}$$

In this expression, $\varphi_{1\times1}$ stands for the convolution operation with a kernel size of 1×1 and $||$ refers to the concatenation operation. C feature maps pass to a *sigmoid* activation function to generate the weights ϑ, which helps to improve the discrimination between the OD ROI pixels and background pixels. These weights multiply by T_1 and add the result to $\hat{T}_2$ as follows:

$$D = \vartheta(C) \cdot T_1 + \hat{T}_2 \tag{4.2}$$

After that, the improved feature maps, D, are fed into four convolution layers with the same architecture as the convolution layers of the encoder network. Final decoder

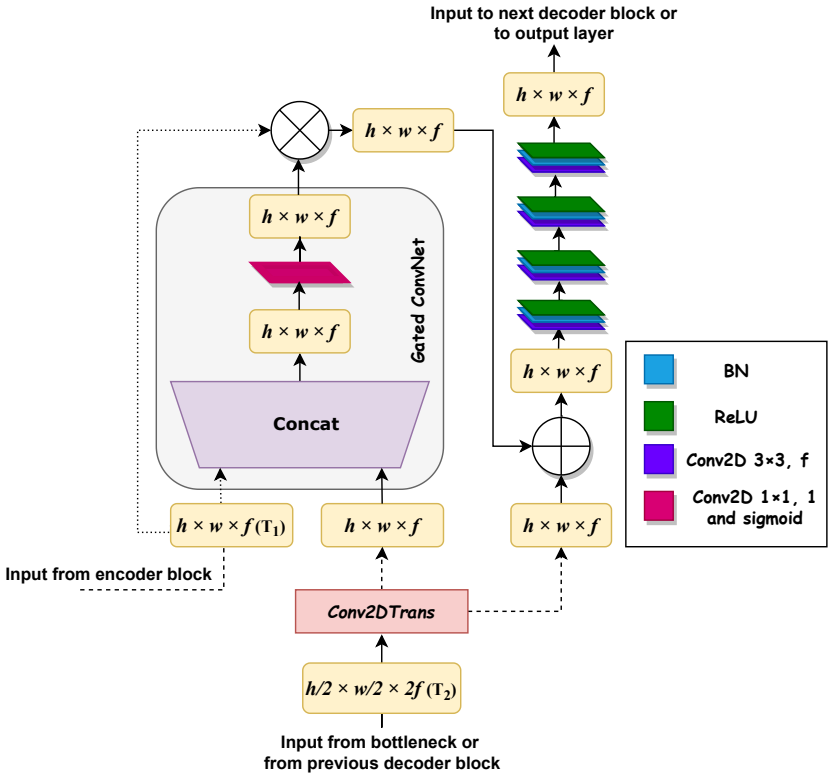


Figure 4.4: Illustration of gated skip connections mechanism (Decoder block).

block fed to a Conv2D as output layer with a kernel size of 1×1 followed by a sigmoid activation function to predict the OD ROI. The output is a binary image that includes the OD ROI and has a size identical to the size of the input image size (384×256). The implementation of GSCs is available at https://github.com/Mohammed-Horbi/GSCs/blob/main/UnetGatedSkipConnection-UnetGSCs.py

4.3.2 Stage 2—Constructing Glaucoma Classifiers Using Deep Transfer Learning

Figure 4.5 shows the second stage of the proposed glaucoma detection system, i.e., the construction of accurate individual glaucoma classifiers based on different deep CNNs and transfer learning approaches. The input to this stage is the OD ROI detected in the first stage. We employ several pre-trained deep CNNs and transfer learning techniques to extract glaucoma-relevant features.

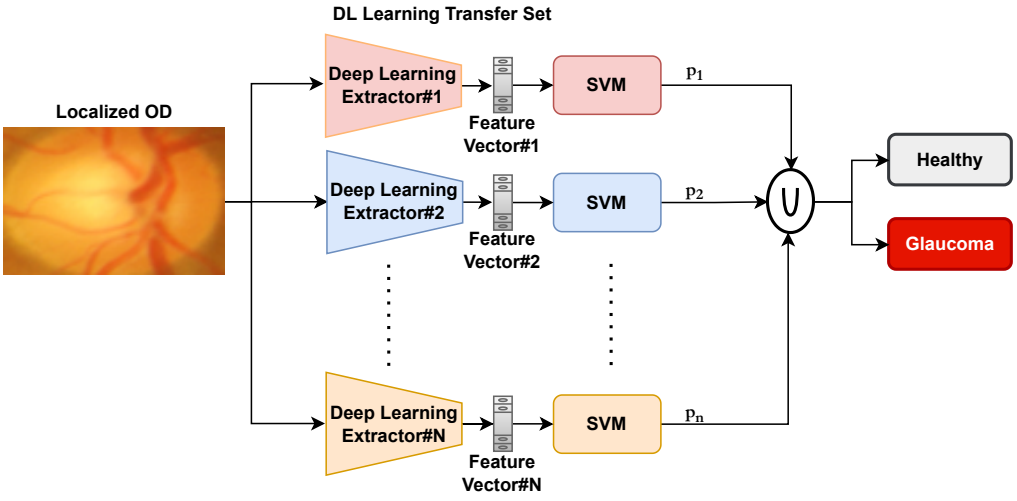


Figure 4.5: Constructing individual glaucoma classifiers based on different deep CNNs and SVM.

Specifically, we assessed the efficacy of 11 efficient deep feature extractors: MobileNet, InceptionV3, DenseNet201, InceptionResNetV2, VGG16, VGG19, Xception, NASNetMobile, ResNet50, ResNet50V2, and ResNet101V2. All of these networks are ImageNet pre-trained CNNs. Table 4.4 summarized the input image

size and the number of parameters of each deep feature extractor. Below, we briefly describe each deep feature extractor. All pre-trained models are available at https://keras.io/api/applications.

- **VGG architectures** [13]—VGG16 and VGG19 are two networks that use a set of 3×3 convolutional layers stacked on top of each other in an increasing depth. Reducing volume size is handled by max pooling. A softmax classifier then follows two fully-connected layers, each with 4,096 nodes. In the case of VGG16, and VGG19, 16 and 19 stand for the number of network layers.
- **ResNet architectures** [14], [15]—ResNet models are much deeper than VGG16 and VGG19. ResNet architectures depend on a collection of micro-architecture building blocks (along with standard Conv and pooling layers). Using the residual module in these architectures eases propagating the feature maps from one block to another. Many ResNet architectures are available, such as ResNet50, ResNet50V2, and ResNet101V2. Each ResNet architecture has a small difference in the internal construction and number of layers than other architectures.
- **DenseNet201** [17]—The architecture of DenseNet is fairly similar to ResNet, with some fundamental differences. ResNet uses an additive method that merges the previous layer with upcoming layers. In contrast, DenseNet concatenates each layer and obtains additional inputs from all preceding layers, and passes its feature maps to all subsequent layers. Here, 201 stands for the number of network layers.
- **InceptionV3** [71]—Like the ResNet architecture, InceptionV3 is based on a collection of micro-architecture building blocks. In this architecture, the *Inception* module produces multi-level features by applying different convolutions like 1×1, 3×3, and 5×5 within the same module on the input feature maps. Then, the outputs of filters are stacked along the channel

dimension before feeding them into the next layer of the network.

- **Xception** [72]—This network is very similar to the Inception network, which relies on ResNet architecture. However, In *Xception* architecture, depthwise separable convolutions are employed rather than the standard Inception modules to reduce the model weight and be much more efficient in computation time.
- **MobileNet** [16]—This convolutional neural network is designed for mobile and embedded vision applications. It uses Depth-wise convolution and Point-wise convolution to build light weight deep neural networks with low latency for mobile and embedded devices. In addition, it has two different global hyperparameters to reduce the computational cost further effectively. One is the width multiplayer, and another is the resolution-wise multiplayer.
- **InceptionResNetV2** [73]— This network architecture includes 164 layers deep CNNs based on a merging Inception structure and the Residual connection. In the Inception-Resnet block, multiple-sized convolutional filters are combined with residual connections. Using residual connections avoids the degradation problem caused by deep structures and reduces the training time.
- **NASNetMobile** [74]—It is based on a scalable CNN network. It consists of basic building blocks optimized by reinforcement learning. Each block has several separable convolutions and pooling. The blocks of NASNetMobile can be repeated multiple times according to the required capacity of the network. NASNetMobile has 12 blocks to reduce the training time significantly without sacrificing performance.

It should be noted that the optic disc ROI is resized to the exact size of the input image of each deep feature extractor before extracting the features. Transfer learning is performed by training the SVM classifier on the top of each deep feature extractor to differentiate between glaucoma and healthy cases. The top-performing SVM

Table 4.4: Summary of the input image size and the number of parameters of the deep feature extractors.

Deep Features Extractor	Input Size	Parameters
MobileNet [16]	224 × 224	3,228,864
InceptionV3 [71]	299 × 299	21,802,784
DenseNet201 [17]	224 × 224	18,321,984
InceptionResNetV2 [73]	299 × 299	54,336,736
VGG16 [13]	224 × 224	14,714,688
VGG19 [13]	224 × 224	20,024,384
Xception [72]	299 × 299	20,861,480
NASNetMobile [74]	224 × 224	4,269,716
ResNet50 [14]	224 × 224	23,587,712
ResNet50V2 [15]	224 × 224	23,564,800
ResNet101V2 [14]	224 × 224	42,626,560

classifiers are selected to be aggregated in the third stage. The SVM algorithm has achieved outstanding results with various CAD systems [75; 76]. It finds a decision surface, i.e., a hyperplane, to differentiate between glaucoma and healthy samples. Each glaucoma detection dataset includes a labeled training set having the form $[(Im_1,\ L_1),\ \ldots,\ (Im_i,\ L_i),\ \ldots,\ (Im_T,\ L_T)]$, where Im_i is the fundus image, $L_i \epsilon [-1, 1]$ is the label (i.e., healthy or glaucoma), T is the number of training images. Each deep feature extractor receives the training images $[Im_1, Im_2, \ldots, Im_T]$ and generates feature vectors $[F_1, F_2, \ldots, F_i, \ldots, F_T]$, where $F_i \in R^{N_g}$, N_g is the number of features generated by the deep feature extractor g, and $i = 1, 2, \ldots, T$

The SVM algorithm solves the optimization problem expressed in Eq. (4.3) to find the decision boundary that discriminates the glaucoma features from the healthy ones in the feature space.

$$\|\omega\|_{\omega}^{2} + C \sum_{i=1}^{k} \xi_i \quad \text{where} \quad L_i(\omega^T F_i + B_0) \geq (1 - \xi_i), \xi_i \geq 0 \tag{4.3}$$

Here, the soft margin parameter C is employed to decide how much is required to mitigate the misclassification of each image in the training set, ω is a weight vector,

ξ is a degree of flexibility to the algorithm when fitting the training data, and B_0 is the bias. In the non-linear SVM, the extracted features are mapped into a higher dimensional space using a kernel function:

$$K(F_i, F_j) = (\Phi^T(F_i) \cdot \Phi(F_j)) \tag{4.4}$$

In this study, we evaluated the performance of the linear kernel $K(F_i, F_j) = F_i \cdot F_j$, the polynomial kernel, and the radial basis function (RBF) kernel, finding that the linear kernel yields the best results.

4.3.3 Stage 3– Aggregation based on Fuzzy Operators

To fuse the probabilities of n SVM classifiers $(p_1, p_2, \ldots, p_n)$ into one consensus value P, we use an aggregation function A:

$$P = A_w\,(p_1, p_2, \ldots, p_n) \tag{4.5}$$

where w includes the weights of the n SVM classifiers. It should be noted that most exciting ensemble-based glaucoma detection methods use static weights to construct a glaucoma ensemble classifer [64; 65; 63]. However, aggregation mechanisms that depend on static weights may not be appropriate for practical situations because they ignore the importance of the predictions of individual glaucoma classifiers $(p_1, p_2, \ldots, p_N)$ when constructing the ensemble. Unlike existing ensemble-based glaucoma detection methods, we propose the use of fuzzy operators in the aggregation function A, noting that the use of a gradable operator is more appropriate, effective, and it suitably reflects the natural decision-making process.

Specifically, we use the three popular yet robust andness-directed aggregators: OWA, WPM, and EXM. Such aggregators can generate dynamic weights to determine the

degree of simultaneity (i.e. trade-off between conjunction and disjunction) in the aggregation stage. OWA depends on fuzzy linguistic quantifiers [28], while WPM and EXM use an internal parameter to control the trade-off between hard and soft conjunctiveness and disjunctiveness [26]. OWA, WPM, and EXM are explained in details in Section 2.2.

We have used the three quantifiers of OWA (Q^1, Q^2, and Q^3) for aggregating the probability values obtained with the different SVM classifiers for each class. In addition, for the WPM and EXM, It is worth noting that an $\alpha \geq 0.5$ determines the threshold between hard and soft conjunction, whereas $\alpha < 0.5$ is the threshold for hard/soft disjunction. An α value of 0.75 represents medium partial conjunction. The aggregations are applied to each image classified by the SVMs.

We must note that other more general aggregation operators, like the Choquet integral, could be used. Choquet integral adds the possibility of giving different importance to coalitions of input sources (using a fuzzy capacity measure). However, we did not consider Choquet integral in this study for the following reasons. 1) In the aggregation approach of this chapter, we are not interested in giving different importance to subsets of the individual glaucoma classifiers. 2) when using Choquet, the number of parameters increases exponentially with the number of inputs, and it usually requires a different learning stage to find the appropriate fuzzy capacity measure. This additional complexity is not needed for the aggregation in this work.

4.4 Experimental Results

This section outlines the experiments carried out to assess the performance of the proposed model, including a description of the datasets, the experimental setup, and an analysis of the outcomes.

4.4.1 Datasets

In this chapter, the efficacy of the proposed system for OD detection and glaucoma prediction was evaluated by three public datasets: DRISHTI-GS1, RIM-ONE-r2, and REFUGE. Besides, for the OD detection task, the in-house fundus image dataset, which was collected from Hospital Universitari Sant Joan de Reus (SJR), was also used. The four datasets have been described in Section 2.3.

4.4.2 Experimental Setup

We employed data augmentation techniques to increase the training dataset size of all datasets. This step helps enhance the generalization process of OD ROI detection and glaucoma detection models. For OD ROI detection models, the training data of each dataset is increased by 16 times with randomly chosen augmentation techniques like flipping, different rotation angles, and brightness. Then, both images and the corresponding ground truths are resized to 256×384 to make the training process faster. For glaucoma detection models, each dataset is increased by 4 times for training, using horizontal and vertical flipping and rotation angles of $90°$, $270°$ and resized the input images to 64×96. The validation set is employed to determine the best checkpoint of the trained models. As these datasets are unbalanced, we consider balancing the number of augmented images of healthy and glaucoma classes.

4.4.3 OD ROI Detection Results

Figure 4.6 illustrates some results of the proposed OD ROI detection model based on GSCs architecture. As shown, GSCs can correctly localize OD in fundus images of healthy and glaucoma cases. To demonstrate the efficacy of the proposed OD ROI detection model, we compare it with nine other promising deep learning networks: Unet, DoubleU-Net, DeepLabV3+, CGNet, ERFNet, SegNet, ESNet, LinkNet, and

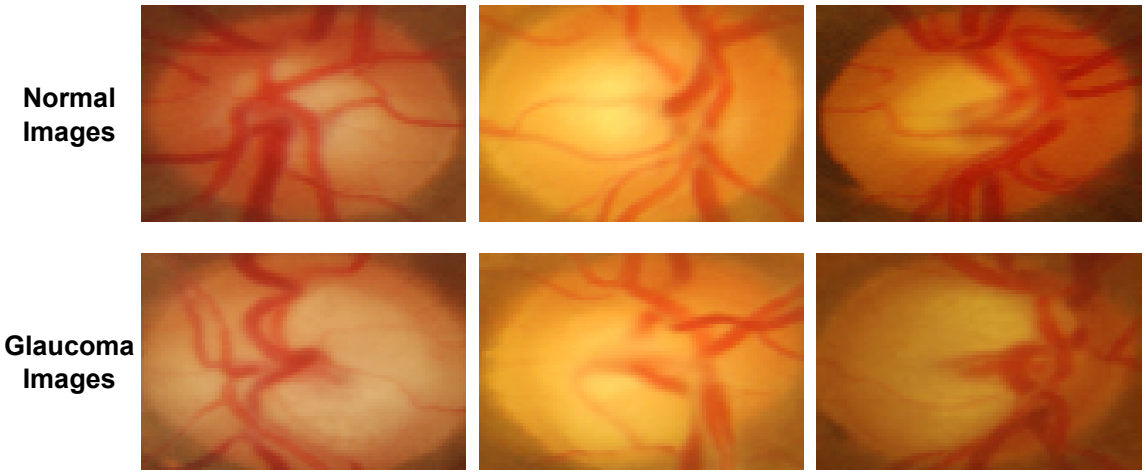


Figure 4.6: Examples of optic disc detected by GSCs.

SQNet models. All models follow the encoder-decoder architecture except CGNet. Each optic disc ROI detection model is briefly described in 3.2.1. Each OD ROI

Table 4.5: Performance comparison on SJR dataset.

Models	Evaluation Metrics	
	IOU(%)	Dice(%)
GSCs	**95.1**	**94.9**
DoubleU-Net	94.4	94.2
DeepLabV3+	94.1	93.8
U-Net	92.4	91.8
LinkNet	92.4	91.8
SQNet	91.9	91.2
ESNet	91.9	91.2
CGNet	91.6	90.8
ERFNet	91.5	90.7
SegNet	90.5	89.4

detection model is trained with the binary cross-entropy loss function, 50 epochs, Adam optimizer with a learning rate of 0.001, and a batch size of 4 images. Figure 4.7 (top) shows the OD ROIs detected by different models: the proposed optic model, Double-UNet, and SegNet.

The green color stands for under-segmentation for the OD region, meaning a part of the OD region is cropped and will not be included in the resulting OD ROIs. The red color stands for over-segmentation for the OD region. As shown in Figure 4.7 (bottom), over-segmented OD leads to a larger ROI, whereas

under-segmented OD leads to a smaller ROI. The under-segmentation of the OD

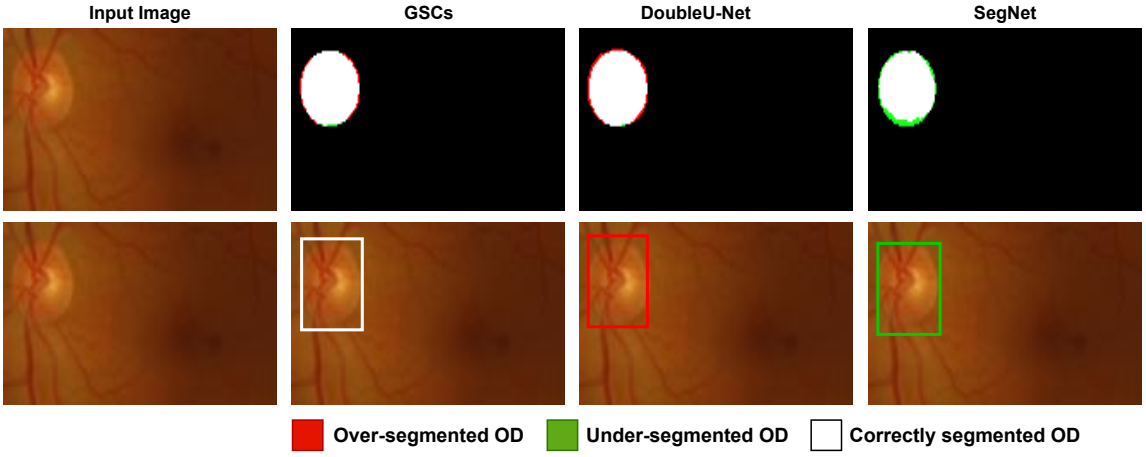


Figure 4.7: Examples of under and over-segmented optic disc detected by GSCs.

region can degrade the performance of the glaucoma detection system. Here, we compute the IOU and Dice scores of each OD ROI detector to check if there is over or under-segmentation of the OD region. IOU and Dice are two standard measures in image segmentation. Table 4.5 demonstrates that the proposed GSCs model achieves the best segmentation result compared to other models. It obtains IOU and Dice scores close to 95%. The Double-UNet obtains an IOU score of 94.4%, while SegNet obtains the lowest Dice and IOU scores. Such a model produces over-segmentation and under-segmentation of OD in many fundus images. The excellent IOU and dice scores obtained by GSCs mean that it has very few over-segmentation and under-segmentation cases. This will yield more accurate OD ROIs, as shown in Figure 4.7 (bottom).

4.4.4 Evaluating the Individual Glaucoma Classifiers

Table 4.6 presents the glaucoma detection results of the 11 state-of-the-art deep feature extractors: MobileNet, InceptionV3, ResNet50, ResNet50V2, DenseNet201, InceptionResNetV2, VGG16, VGG19, Xception, NASNetMobile, and ResNet101V2. With the DRISHTI-GS1 dataset, ResNet50V2 and DenseNet201 models obtain the

highest accuracy (90.2%). ResNet50V2 also achieves an AUC and specificity of 89.9 and 84.6%, respectively. In the case of the REFUGE dataset, ResNet101V2

Table 4.6: Performance of individual glaucoma classifier.

Deep Features	DRISHTI-GS1				REFUGE				RIM-ONE			
Extractor	ACC	AUC	Sen	Spe	ACC	AUC	Sen	Spe	ACC	AUC	Sen	Spe
MobileNet	80.4	81.2	89.5	53.9	87.0	89.2	82.5	87.5	96.4	98.1	95.0	97.4
InceptionV3	80.4	80.8	89.5	53.9	85.5	**92.6**	**85.0**	85.6	93.4	98.1	96.7	90.9
ResNet50	84.3	83.8	92.1	61.5	82.5	87.2	70.0	83.9	88.3	92.3	88.3	88.3
ResNet50V2	90.2	**89.9**	92.1	**84.6**	87.3	92.5	82.5	87.8	95.6	**99.9**	**100**	92.2
DenseNet201	**90.2**	83.6	94.7	76.9	86.8	88.8	75.0	88.1	92.0	98.4	96.7	88.3
InceptionResNetV2	82.4	82.0	84.2	77.0	91.8	91.9	75.0	93.6	94.9	99.5	98.3	92.2
VGG16	84.3	84.0	**97.4**	46.2	76.8	86.1	77.5	76.7	90.5	96.2	91.7	89.6
VGG19	82.4	83.4	84.2	76.9	71.5	81.8	82.5	70.3	90.5	95.6	91.7	89.6
Xception	78.4	75.9	89.5	46.2	87.8	90.9	70.0	89.7	93.4	97.8	95.0	92.2
NASNetMobile	84.3	87.5	84.2	84.6	86.0	91.3	77.5	86.9	92.0	93.9	85.0	97.4
ResNet101V2	82.4	88.1	86.8	69.2	**92.3**	92.2	80.0	**93.6**	**97.1**	99.5	96.7	**97.4**

obtains accuracy and specificity rates of 92.3 and 93.6%, respectively, which are much better than the scores of all other models except InceptionResNetV2. InceptionV3 achieves AUC and sensitivity rates of 92.6 and 85%, respectively, which are 0.4 and 2.5 points higher than the second-best models. With the RIM-ONE dataset, ResNet101V2 obtains an accuracy of 97.1% and specificity of 97.4%, while ResNet50V2 obtains AUC and sensitivity close to 100%. Figure 4.8 presents the ROC curves of the top-five individual glaucoma classifiers with each dataset. This analysis demonstrates that there is no superior deep feature extractor or individual glaucoma classifier. Based on our experiments and as shown in Figure 4.8, we notice that the top-performing individual glaucoma classifiers with the DRISHTI-GS1 dataset achieve sustainable performance with all datasets. Hence, we employ ResNet50V2, ResNet101V2, NASNetMobile, DenseNet201, and VGG16 to build the ensemble based on fuzzy aggregation operators.

4.4.5 Evaluating the Aggregation Methods

As mentioned in Section 4.4.4, five individual glaucoma classifiers are aggregated using the fuzzy operators; thus, $m = 5$ in Equations 2.6 and 2.7. Table 4.7 presents

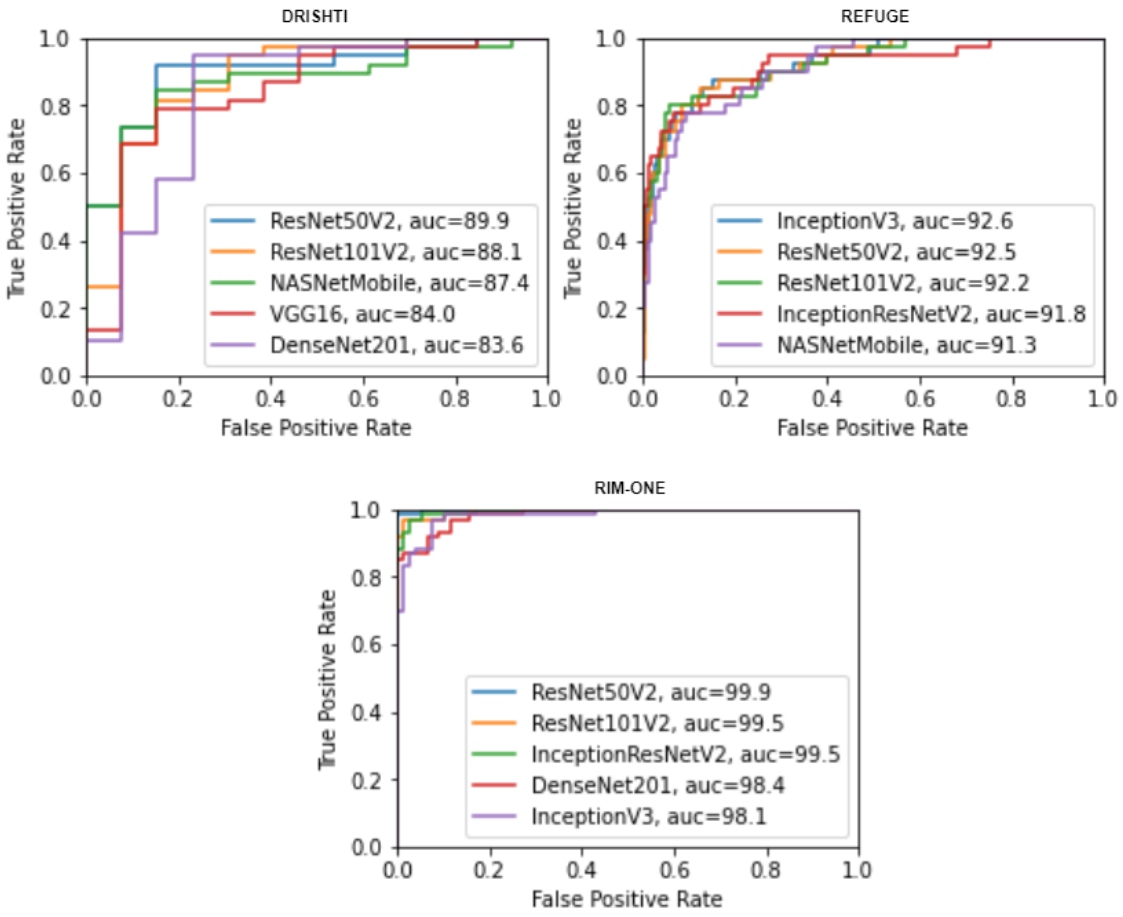


Figure 4.8: ROC curves of the top-five individual glaucoma classifiers with DRISHTI-GS1, RIM-ONE, and REFUGE datasets.

the values used in WPM (Equation 2.6) and EXM (Equation 2.7). Two cases are tested: average partial disjunction (DA) with $\alpha = 0.25$ and average partial conjunction (CA) with $\alpha = 0.75$. For OWA, three cases are tested: Q^1 (with andness $\alpha = 0.2$), Q^2 (with $\alpha = 0.8$) and Q^3 (with $\alpha = 0.55$). In this study, we

Table 4.7: Values of $r5(\alpha)$ and $t5(\alpha)$ used in WPM and EXM [1].

Method	**Global**		**Aggregators**	Params	$\mathbf{r_5/t_5}$ **values**
	Andness	**Orness**			
WPM	0.25	0.75	**DA**	r_5	5.1113
	0.75	0.25	**CA**		-0.7054
EXM	0.25	0.75	**DA**	t_5	6.5074
	0.75	0.25	**CA**		- 6.5074

consider that an operator is conjunctive when andness is greater than 0.5, which

means that orness is lower than 0.5. As we explained in Section 4.3.3, andness and orness are measured with the Yager definition [28]. For WPM and EXP, we use the definitions of Dujmovic [26]. Tables 4.8, 4.9, and 4.10 present the performance of three glaucoma detection systems based on three different aggregation methods (OWA, WPM, and EXM) with DRISHTI-GS1, RIM-ONE, and REFUGE datasets. Values in bold highlight the best values in accuracy and AUC for each dataset. The

Table 4.8: Results of different OWA's policies of best five aggregated models of glaucoma classification.

OWA Policy	Dataset	ACC	AUC	Sen	Spe
At least half, Q^1	DRISHTI-GS1	84.3	92.3	100	38.5
	REFUGE	85.3	96.6	95.0	84.2
	RIM-ONE	94.9	99.8	100	90.9
As many as possible, Q^2	DRISHTI-GS1	88.2	94.7	84.2	100
	REFUGE	**94.3**	**96.2**	65.0	97.5
	RIM-ONE	**97.8**	**99.8**	96.7	98.7
Most, Q^3	DRISHTI-GS1	**90.2**	**95.3**	92.1	84.6
	REFUGE	94.0	96.3	77.5	95.8
	RIM-ONE	95.6	99.8	96.7	94.8

results of each dataset will now be commented focusing on sensitivity and specificity indices, as they are two essential indicators in healthcare applications. With the DRISHTI-GS1 dataset, the Q^1 disjunctive policy obtains a sensitivity rate of 100% but quite a low specificity. The Q^2 conjunctive policy gives the opposite result, with a specificity rate of 100% and a lower sensitivity. In this case, the Q^3 policy (soft conjunction) obtains the highest accuracy of 90.2 and AUC of 95.3, which is closer to an arithmetic average.

Table 4.9: Results of different WPM's alpha values of best five aggregated models of glaucoma classification.

Dataset	Metrics of WPM, $\alpha = 0.25$				Metrics of WPM, $\alpha = 0.75$			
	ACC	AUC	Sen	Spe	ACC	AUC	Sen	Spe
DRISHTI-GS1	84.3	92.1	100	38.5	**86.3**	**93.9**	84.2	92.3
REFUGE	81.8	96.5	97.5	80.0	**94.0**	**96.2**	67.5	96.9
RIM-ONE	91.4	99.7	100	84.4	**97.1**	**99.8**	96.7	97.4

Table 4.10: Results of different EXM's alpha values of best five aggregated models of glaucoma classification.

Dataset	Metrics of EXM, $\alpha = 0.25$				Metrics of EXM, $\alpha = 0.75$			
	ACC	**AUC**	**Sen**	**Spe**	**ACC**	**AUC**	**Sen**	**Spe**
DRISHTI-GS1	84.3	91.9	100	38.5	**84.3**	**93.5**	81.6	92.3
REFUGE	83.8	96.6	97.5	82.2	**94.0**	**96.2**	62.5	97.5
RIM-ONE	92.0	99.8	100	85.7	**97.1**	**99.8**	96.7	97.4

We observe a similar behavior with WPM and EXP, where the best performance indicators are given in the conjunctive models. At the same time, very low and unacceptable specificity is obtained in the disjunctive case (38.5). It indicates that it is better an aggregation that gives the final classification mask from the conjunctive aggregation of most of the 5 deep learning models.

In the REFUGE dataset, the Q^2 conjunctive policy yields the best glaucoma segmentation results (accuracy of 94.3% and AUC of 96.2%). WPM and EXM with conjunctive model give similar quality results (accuracy of 94.0% and AUC of 96.2%). Analogous results are obtained with the RIM-ONE dataset. The OWA with conjunctive Q^2 leads to an accuracy of 97.8% and AUC of 99.8%, which are much better than other aggregation methods. Very close results are obtained with a conjunctive WPM and EXM.

In conclusion, we can say that the conjunctive aggregation models give better results than the disjunctive ones. A compensatory approach that focuses only on a small subset of the deep learning-based glaucoma classification models is insufficient to have good accuracy. The disjunctive policy gives low specificity values because it overestimates the positives (including false positives). The conjunctive approach, which considers the majority output of the deep learning-based glaucoma classification models, produces better results in the three datasets.

It should be noted that the accuracy and AUC for the two case models (CA and DA) are almost the same in the three aggregation operators. So, they are equivalent when the same *andness* level is fixed. Any of the 3 operators could be

used in the ensembling stage. This analysis concludes that OWA with Q^2 is the best configuration for REFUGE and RIM-ONE and a low conjunctive policy using Q^3 for DRISHTI-Gs.

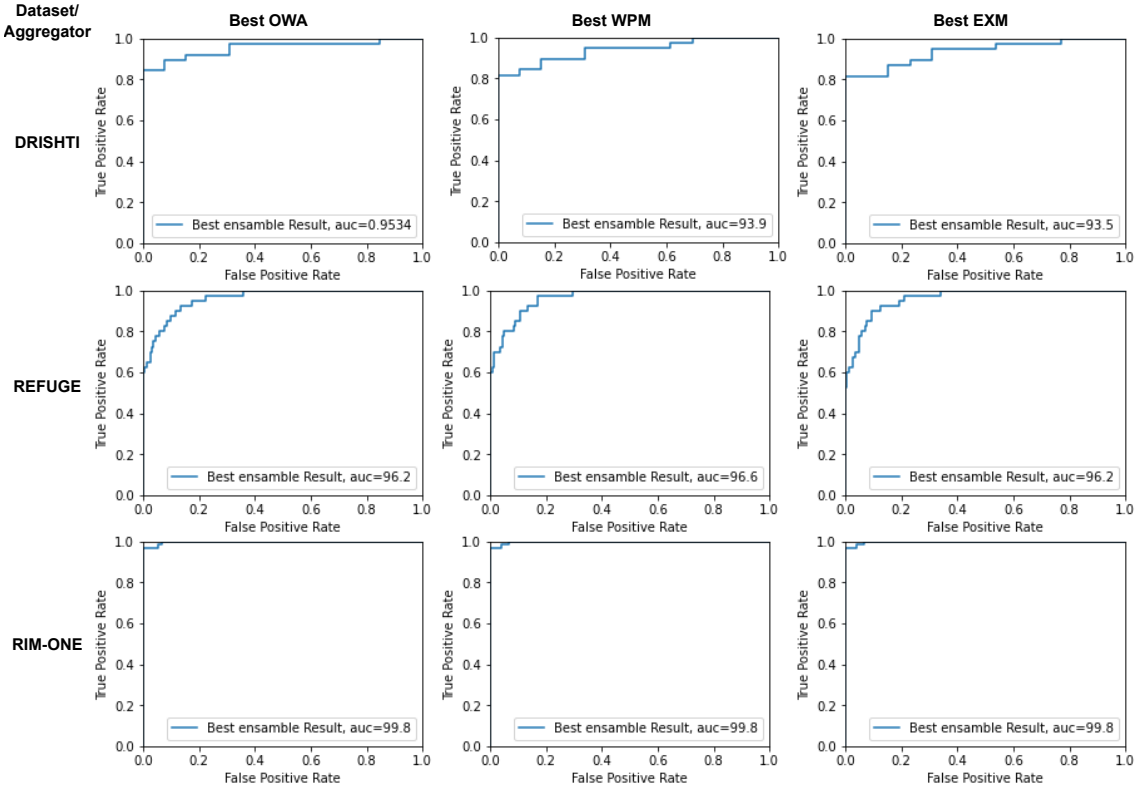


Figure 4.9: ROC curves for (left) OWA, (middle) WPM, and (right) EXM for the three datasets.

Figure 4.9 presents the ROC curves of the glaucoma CAD systems based on (left) OWA, (middle) WPM, and (right) EXM. These plots help to visually evaluate the performance of the OWA, WPM, and EXM ensemble methods. As we can see in the ROC curves, the glaucoma CAD systems based on the OWA aggregation policies maintain high detection results with the three datasets. Figure 4.10 shows the Bland-Altman plots for glaucoma prediction results of each aggregation method with DRISHTI-GS1, REFUGE, and RIM-ONE datasets. Bland-Altman plots show the difference between predictions of the proposed CAD system and ground truth. The solid line stands for the mean difference, and the dashed line stands for the standard deviation. The best aggregation policies are compared for each operator.

With DRISHTI-GS1, the 95% confidence interval of the mean difference of OWA, WPM and EXM are −0.59 to +0.63, −0.60 to +0.80, −0.62 to +0.86, respectively. In the case of REFUGE, the widest interval of the mean difference is achieved with EXM (−0.46 to +0.49). RIM-ONE has intervals of the mean difference narrower than DRISHTI-GS1, REFUGE, where the narrowest interval of the mean difference is achieved with OWA (−0.28 to +0.30). This analysis demonstrates the reliability and agreement of the results of the three aggregation methods.

This analysis concludes that OWA with a conjunctive policy (from now on, denoted *Conj-OWA*) yields the best results. In turn, WPM and EXP give almost the same results.

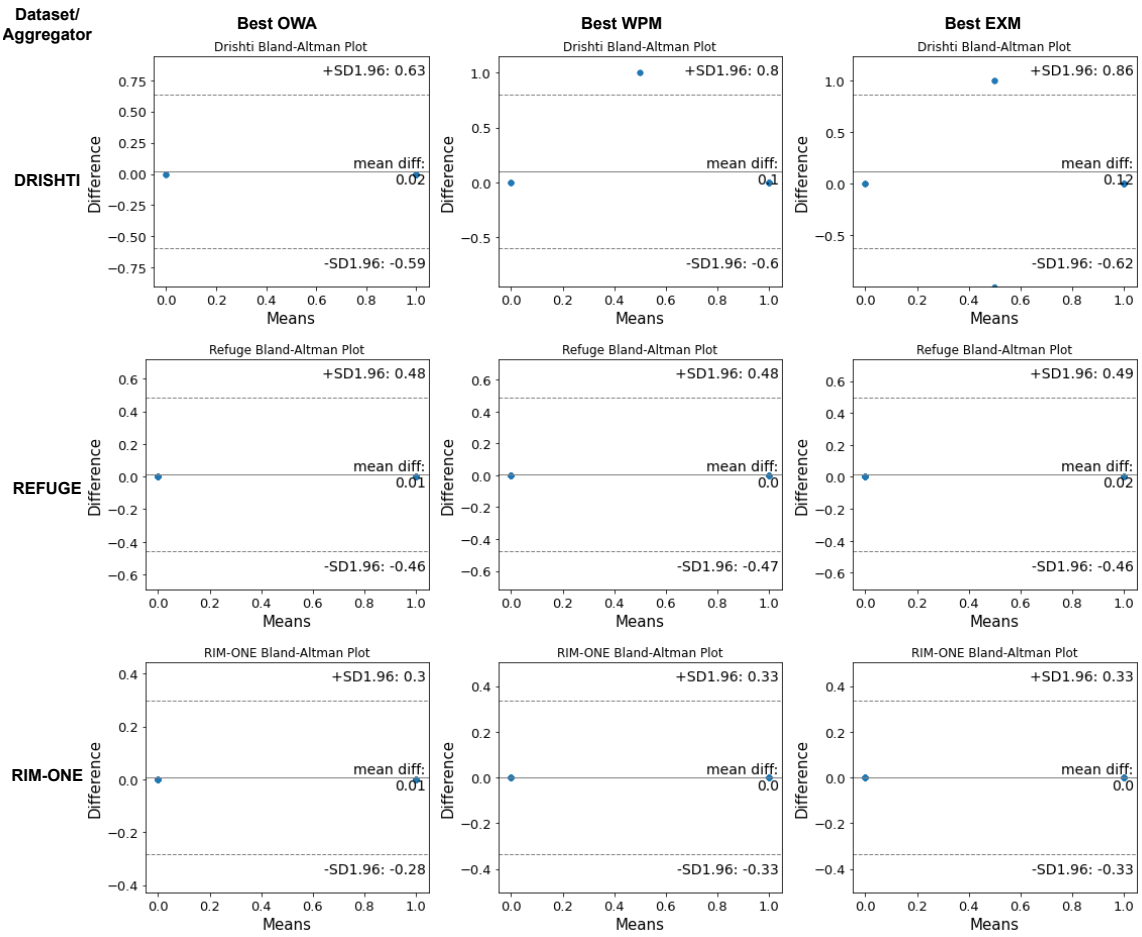


Figure 4.10: Bland-Altman plots for glaucoma prediction results of each aggregation method with DRISHTI-GS1, REFUGE, and RIM-ONE datasets.

4.4.6 Comparison with Other Aggregation Methods

Table 4.11 presents a comparison between the proposed aggregation method, *Conj-OWA*, and other aggregation approaches: arithmetic mean and median. In

Table 4.11: ACC for different aggregation methods

Dataset	Conj-OWA	Mean	Median
DRISHTI-GS1	**90.20**	84.30	**90.20**
REFUGE	**94.30**	93.30	93.80
RIM-ONE	**97.80**	95.60	96.40

the case of the DRISHTI-GS1 dataset, our proposed method obtained an accuracy 5.9 points higher than the mean aggregation method and produced the same result as the median aggregation method. Moreover, in the case of the REFUGE database, our proposed method achieves accuracy rates 1 and 0.5 higher than the mean and median aggregation methods. With the RIM-ONE database, it obtains accuracy rates of 1.20 and 1.40, higher than the other aggregation methods.

4.4.7 Comparison with Existing Methods

Table 4.12 compares the proposed glaucoma detection system with existing methods on three different datasets. With the DRISHTI-GS1 dataset, the proposed method with the conjunctive OWA aggregation (i.e., *Conj-OWA ensemble*) achieved an accuracy higher than 90%, which is 0.2 points better than the accuracy of the method proposed by [64]. Besides, the proposed method outperforms all methods in terms of AUC, sensitivity, and specificity.

In the case of the REFUGE dataset, the proposed method obtained glaucoma detection results on par with the existing methods, and it achieved a specificity higher than 97%. However, [31] achieved the best AUC value with the REFUGE dataset. [68] obtained 97.6%, which is better than our proposed method. Nevertheless, they used a multi-task and complex deep learning network with a multi-adaptive optimization which requires additional memory. In the case of RIM-ONE, we

Table 4.12: Performance of the proposed system for glaucoma detection. Here, (-) stands for 'not reported', and **'Conj-OWA ensemble'** stands for the conjunctive OWA aggregation.

Datasets	Method	Performance Measures (%)			
		ACC	**AUC**	**Sen**	**Spe**
DRISHTI-GS1	Chakravarty et al. [77]	76.8	78.0	-	-
	Diaz-Pinto et al. [63]	75.3	80.4	74.2	71.4
	Sreng et al. [64]	90.0	92.1	-	-
	Hervella et al. [68]	-	94.7	-	-
	Conj-OWA ensemble	**90.2**	**95.3**	**92.1**	**84.6**
REFUGE	Orlando et al. [31]	-	98.9	97.5	-
	Sreng et al. [64]	95.8	94.3	-	-
	Hervella et al. [68]	-	97.6	-	-
	Conj-OWA ensemble	94.3	96.2	65.0	**97.5**
RIM-ONE	Gomez-Valverde et al. [60]	-	94.0	87.0	89.0
	Diaz-Pinto et al. [63]	71.2	85.8	79.3	79.9
	Sreng et al. [64]	92.1	99.0	-	-
	Conj-OWA ensemble	**97.8**	**99.8**	**96.7**	**98.7**

achieved an AUC close to 100%. Also, the proposed method beats all other methods by achieving accuracy, sensitivity, and specificity rates higher than 96%.

4.5 Conclusions

This chapter proposed an efficient CAD system for diagnosing glaucoma based on fundus images, deep transfer learning, and fuzzy aggregation operators. Specifically, the proposed CAD system includes three stages: 1) Detection of the region of interest of the optic disc using an efficient deep learning network, 2) Classification of images based on different pre-trained deep convolutional neural networks and support vector machines, and 3) Use of fuzzy aggregation operators to fuse the predictions of glaucoma classifiers. We used three popular yet robust aggregators: OWA, WPM, and EXM operators. We assessed the efficacy of the proposed glaucoma CAD system on three public datasets: DRISHTI-GS1, RIM-ONE, and REFUGE. The proposed conjunctive OWA aggregation method (Conj-OWA) achieves the best glaucoma classification results. Specifically, it achieves accuracy values of 90.2, 97.8, and 94.3% and AUC values of 95.3, 99.8, and 96.2%, respectively, on DRISHTI-GS1, RIM-ONE, and REFUGE databases.

CHAPTER 5

Deep Learning-based Models for Exudates Segmentation in Fundus Images

5.1 Introduction

In this chapter, we move to Diabetic Retinopathy disease in the eyes. People with diabetes for a long time and not good health control are prone to suffer from Retinopathy. DR produces different retinal lesions in the retina due to blood leakage, such as hard exudates (EX) and soft exudates (SE) [56]. Figure 5.1 shows a fundus image with EX (box colored in green) and SE (box colored in red). Regular screening tests are done on diabetic people, taking photos of the eye fundus. Then, ophthalmologists inspect these images to detect the signs of such lesions. However, this manual analysis is quite difficult due to the complex structure of lesions, various sizes, differences in brightness, and the inter-class similarity with other fundus tissues. In addition, manually detecting tiny lesions consumes a lot of time and effort.

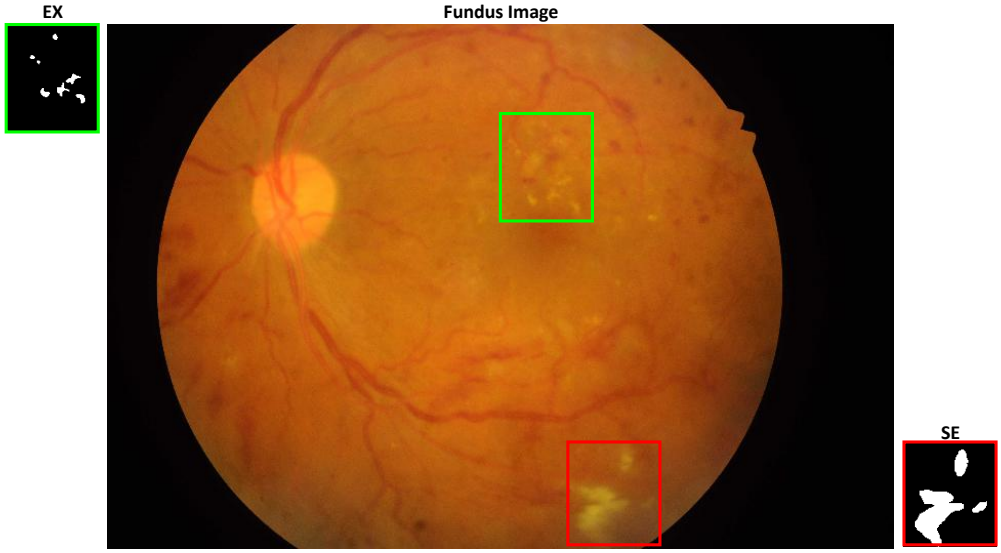


Figure 5.1: Example of hard exudates (green box) and soft exudates (red box).

Several deep learning-based automated systems have recently been proposed for segmenting retinal eye lesions. Most of them use CNNs like UNet [12] to automatically learn representative and high-level features from the input fundus images to achieve accurate segmentation. For instance, The authors of [78] proposed CARNet for multi-lesion segmentation. CARNet feeds the whole image

and patch image into ResNet50 and ResNet101 networks using a single attention refinement decoder. They used IDRiD, E-ophtha, and DDR datasets for evaluation. EAD-Net [79] presented a CNN-based system divided into an encoder module, dual attention module, and decoder module. They evaluated their work on two datasets: the E_ophtha_EX dataset for exudates and the IDRiD dataset for four kinds of lesions. In [80], authors developed a weakly-supervised framework for fundus lesion segmentation using grayscale and morphological features of lesions and a deep neural network with an attention mechanism and residual module. They evaluated their system by 1485 images extracted from the Messidor dataset and labeled by them. Finally, in [81], authors introduced scale-aware attention with different backbones to re-weight multi-scale features of decoders dynamically, and they evaluated it on IDRiD, E-ophtha, and DDR datasets.

Although there are many advantages offered by deep learning techniques, especially those based on the UNet models, there is still a problem with dealing with tiny lesions like retinal exudates, so we need to develop a boosted mechanism for dealing with them. This chapter proposes accurate eye retinal exudates segmentation models called dual decoder-based network (DDN) and multi-scale based network (MSN) that take as an input a fundus image and produces its corresponding lesion mask. The novelties of the proposed deep learning models are the following:

1. DDN uses dual decoders to boost the performance and produce an accurate mask of the input image. The key component in each decoder is the GSCs network [47]. The reason behind using the GSCs is the ability to focus on the decoder's most valuable feature based on the previous level's features.
2. The first decoder receives the skip connections from the encoder, where the second decoder takes the output of the first one as skip connections in a cascading manner. Hence, although we only use two decoders in this work, DDN can generalize to any level of cascaded boosted decoders, such

cascading and composing of multiple layers give the network the ability to grasp representations of data with multiple levels of abstraction [10].

3. The final output of DDN is obtained by fusing the output of the two decoders, which can be seen as a kind of online ensembling technique. We mean by online that the fusion operation is done in both the training and the inference phases.
4. Proposing the use of two effective multi-scale modules to encourage the MSN network to extract exudate-relevant features from the fundus images. The first multi-scale module is used at the beginning of the MSN network. It consists of a combination of 3×3 and 1×1 convolution layers. The second multi-scale module contains an atrous spatial pyramid pooling block used at the neck of MSN.
5. Integrating a gated skip connection mechanism at the decoder of MSN to help the network focus on exudates features.

We conducted extensive experiments on the well-known publicly available IDRiD dataset. DDN and MSN achieved competitive performance in one kind of lesion compared with the state-of-the-art systems in the IDRiD challenge and outperformed them in another.

5.2 Exudates Segmentation Methods

In this chapter, we propose two methods for exudate segmentation. The first method is based on a dual-decoder-based network for exudates segmentation (DDN). The second method uses a multi-scale network for exudates segmentation (MSN). Below we introduce each method.

5.2.1 Dual-decoder based network for exudates segmentation

This section describes the first proposed method, DDN. Figure 5.2 shows its architecture. It comprises the following parts: Backbone (colored in blue), also known as the encoder layer, which aims to encode the input image and produce feature maps at multiple levels of scales. The neck (colored in purple) is an Atrous Spatial Pyramid Pooling (ASPP) layer that helps to extract high-resolution features. Head, also known as decoder layer, a DDN uses a dual-decoder including GSCs network [47] followed by an output layer. We explain each part of the method in the following:

- **Encoder Layer**: we use a ResNet50 [14] encoder that was pre-trained on the ImageNet dataset as a backbone for our model. We selected this model because ResNet is the state-of-the-art backbone for many computer vision tasks [78; 82]. The main goal of the backbone in our proposed model is to encode the input eye fundus image and extract abstracted and meaningful features at different levels of scales. The key advantage of ResNet50 is the residual connection which is used to add the output from a previous layer to the next layer, which helps to avoid the gradient vanishing problem.
- **ASPP**: The ASPP uses to assist in extracting high-resolution feature maps and maximize capturing the contextual data of the micro lesions lost across the encoder multiple-scale. We provide the ASPP [83] as a neck to the link between the encoder and the decoder blocks.
- **Decoder Layer**: The main new component of our methodology is the design of the head part. The proposed method employs a shared ResNet50 encoder to improve the performance without computation and memory overload and decoder block for lesions segmentation. The decoder block has the same

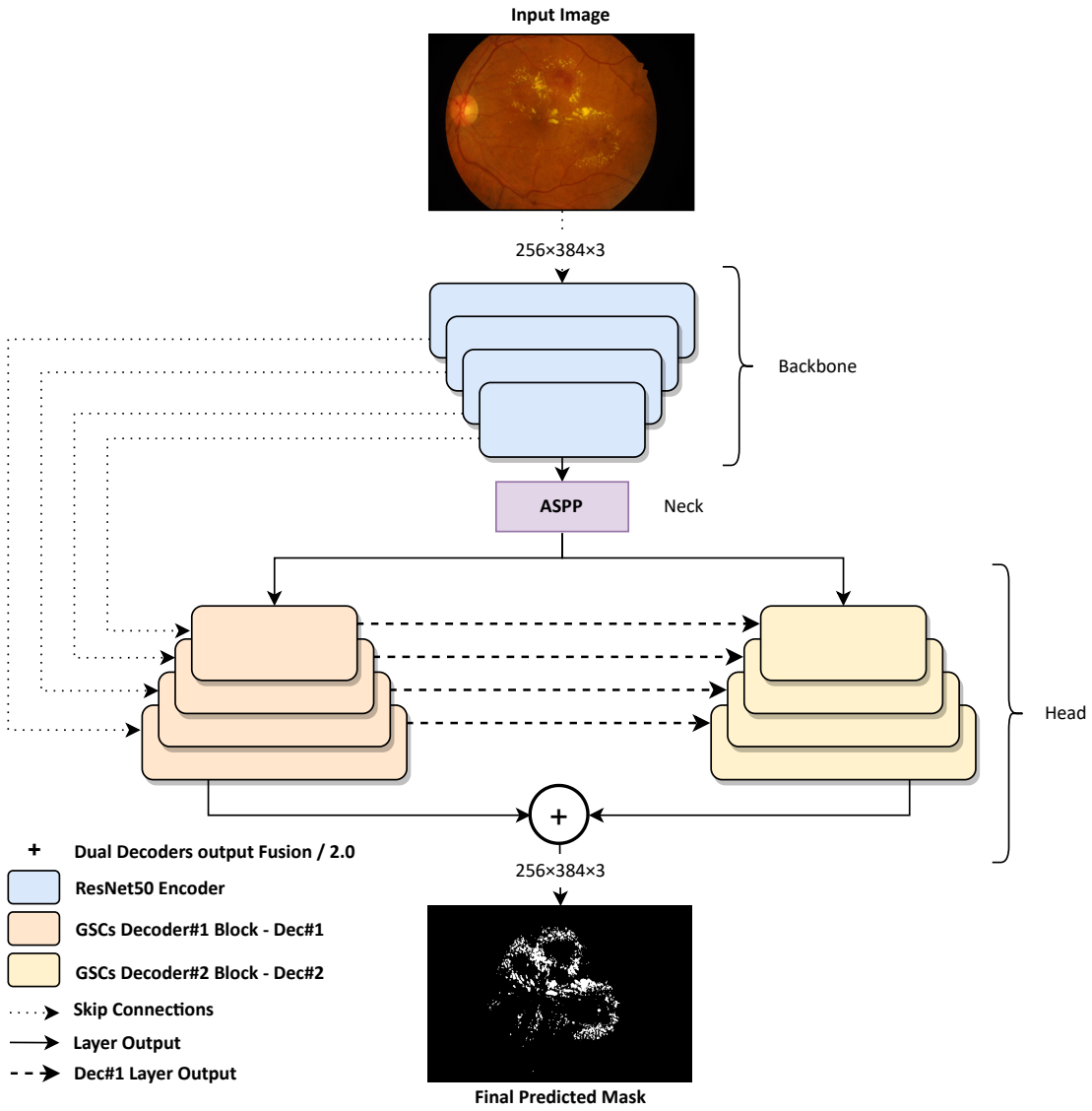


Figure 5.2: Structure of the DDN framework for Exudates segmentation.

internal GSCs architecture as each decoder side block. Figure 5.3 presents the GSCs mechanism of each decoder side block. GSCs modified the standard UNet decoder with a boosted feature maps production to enhance the discrimination between the lesion and background pixels for exudates segmentation.

DDB has four encoder blocks and four GSCs mechanism blocks for each decoder side. In dec#1, the GSCs mechanism receives feature maps from the corresponding ResNet encoder layers. It concatenates them with the feature maps produced by the previous block (either the ASPP neck block or a previous

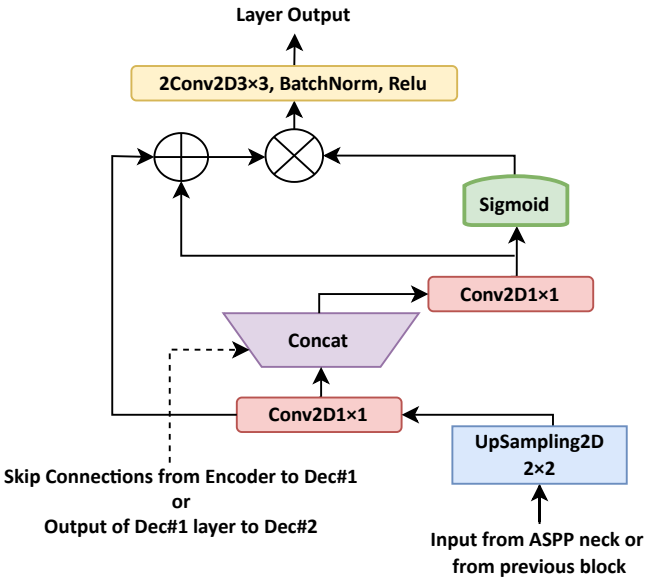


Figure 5.3: Gated skip connections network.

dec#1 block). For dec#2, it receives feature maps from the corresponding dec#1 layers. Then concatenate them with the previous block's feature maps (either the ASPP neck block or a previous dec#2 block). We can express these feature maps as $S_1 \in \mathrm{R}^{h \times w \times f}$, and $S_2 \in \mathrm{R}^{h/2 \times w/2 \times 2f}$. Then, S_2 feeds to UpSampled2D transposed convolution layer with a kernel size of 2×2 to produce feature maps $\hat{S}_2$. $\hat{S}_2$ and S_1 should have the same width and height to perform the concatenation process as follows:

$$C = \varphi_{1\times1}([S_1||\hat{S}_2]) \tag{5.1}$$

In this expression, $\varphi_{1\times1}$ stands for the convolution operation with a kernel size of 1×1 and $||$ refers to the concatenation operation. C feature maps are fed to a *sigmoid* activation function to generate the weights ϑ, which helps to improve the discrimination between the lesion and background pixels for EX segmentation tasks. These weights are multiplied by the sum of S_1 and $\hat{S}_2$:

$$D = \vartheta \cdot (S_1 + \hat{S}_2) \tag{5.2}$$

After that, the improved feature maps of D are fed into two convolution layers, batch normalization, and Relu activation function. The final output blocks of both dec#1 and dec#2 are followed by a sigmoid activation function and fed into a fusion process by the average weighted aggregation to take the benefits of multiple information sources and generate an optimal joint lesion segmentation [84]. In addition, the fusion aims to produce one final output with fewer output channels as follows:

$$M = (M_1 + M_2)/2.0 \tag{5.3}$$

where M stands for the final output of the proposed framework, whereas M_1 and M_2 indicate the output masks of dec#1 and dec#2, respectively. The final output mask is a binary image that includes the EX lesions and has a size identical to the size of the input image size (384×256).

5.2.2 Multi-scale based Network for Exudates Segmentation

This section describes the second proposal, a multi-scale network for exudates segmentation (MSN). As shown in Figure. 5.4, it is composed of the following parts: a multi-scale module at the beginning, which helps enhance image fundus feature extraction before it comes to the MobileNet backbone (the encoder), and another is the neck which appears after the encoder. The backbone, the encoder layer, aims to encode the input image and produce feature maps at multiple scale levels. The neck is an ASPP layer that helps extract high-resolution features and works as a second multi-scale. Finally, Head, also known as the decoder layer, GSCs mechanism in each layer [47] followed by an output layer. Below we introduce each part.

- **Multi-scale module at the beginning:** The main component of the MSN method is the design of the multi-scale part of the image before feeding it

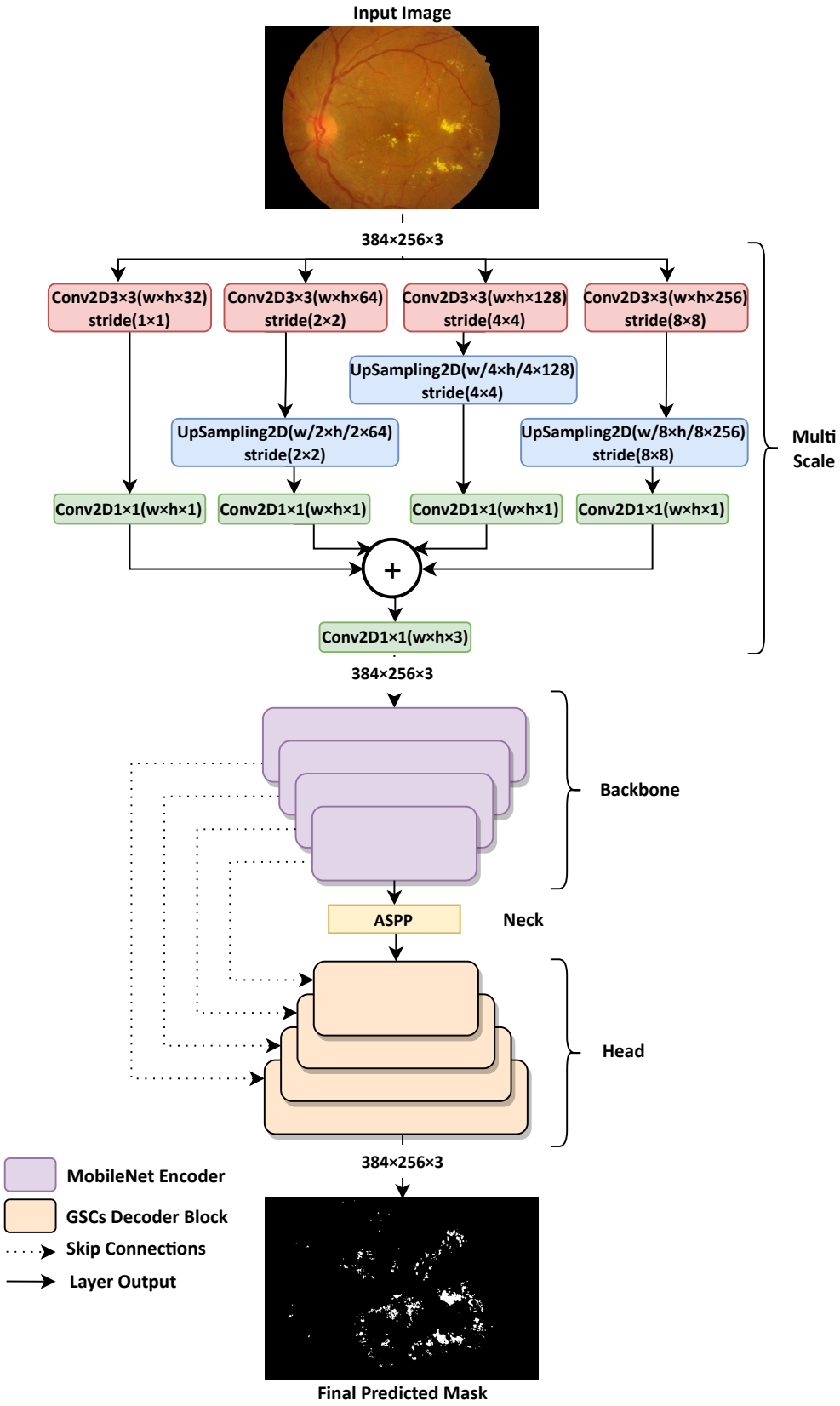


Figure 5.4: Schematic diagram of the MSN network for exudates segmentation in fundus images.

into the model. MSN maintains the multi-scale of the input image to consider features from different stages of the image. To extract features, we feed four copies of the image to a 3×3 convolution with different strides to down scales the image into the same, 1/2, 1/4, and 1/8 of the original image size. Next, we upsample each scale size to the original size of the input image by UpSampled2D transposed convolution to make four upscale feature copies of the downscale images. Then, we use a 1×1 convolution to reduce the dimension of features from coarser stages to the same as the finest scale. After that, we use the attention feature fusion to generate a 3D feature and feed it to the backbone of the model. Moreover, the ASPP neck is another multi-scale module; we introduced it in section 5.2.1.

- **Encoder Layer**: a MobileNet [85] encoder pre-trained on the ImageNet dataset uses as a backbone for our network. We selected this model because MobileNet is a lightweight deep neural network with higher classification accuracy and a state-of-the-art backbone for many computer vision tasks [86; 87]. The main goal of the backbone in our proposed model is to encode the input eye fundus image and extract abstracted and meaningful features at different levels of scales. The key advantage of MobileNet is the two different global hyperparameters to reduce the computational cost-effectively. One is the width multiplayer, and another is the resolution-wise multiplayer. ASPP follows the encoder as a bridge between the encoder and decoder.
- **Decoder Layer**: MSN decoder has four GSCs, modified the standard Unet decoder with a boosted feature maps production to enhance the discrimination between the exudate and background pixels. The four decoder blocks have the same internal GSC architecture in each decoder block. GSC is explained in Section 4.3.1.

5.3 Experimental Results

This section describes the conducted experiments to evaluate the effectiveness of the proposed model, including the experimental setup and the ablation study of the obtained results.

5.3.1 Experimental Setup

We used the popular IDRiD dataset in our experiments [33], which has been described in Section 2.3. It composed of 81 high-resolution retinal fundus images of 4288×2848. Each image contains at least one mask labeled as one of four types of DR lesions, EX, SE, MA, and HE. The dataset was split into 54 and 26 images as the training set for the EX and SE, respectively, and the rest of 27 and 14 as the testing set of EX and SE.

We used the following training pipeline (including some data augmentation techniques to enrich the data and improve the regularity of the model) to process the images in the training set. First, each image is divided into four non-overlapped sub-images, and the corresponding sub-masks are constructed. We ignored the negative sub-images, i.e., the sub-images only with the background mask. Hence, each example in the training process is a sub-image with the size of 2144×1424 pixels with its corresponding sub-mask. Next, we resized the sub-images and the sub-masks to 384×256. The interpolation mechanism is cubic for the images and the nearest neighbor for the masks.

After that, we applied flipping, rotation, and Gaussian noise augmentation techniques for 12 times. The total of training data calculated as: (e.g.,$((4 \times 54) - 20\%$ (for validation)) $\times 12$). We trained each model for 50 epochs using an Adam optimizer and a batch size of 4. The learning rate is set to 0.001. We sampled a subset (20%) from the training set and used it as a validation set to save the best checkpoint of

the trained models. We used the binary cross-entropy as a loss function to train the models.

During the inference phase, we only resize the input image to 768×512 and perform a full image segmentation process (i.e., no image splitting or image augmentation is used during the inference).

5.3.2 Ablation study

In this section, we evaluate the performance of the used methods on the test set images of the IDRiD dataset of EX and SE. Tables 5.1 and 5.2 present the performance of EX and SE retinal lesion segmentation models with the IDRiD dataset. We conduct four different experiments for each EX and SE retinal lesions

Table 5.1: Performance comparison on the IDRiD dataset of EX. Baseline refers to Unet with MobileNet encoder.

Method	Re	Pre	F1	AUPR
Baseline	65.30	81.33	72.44	80.93
Baseline + GSCs	74.0	81.24	77.45	85.25
DDN	68.30	**85.20**	75.82	85.30
MSN	**75.53**	83.54	**79.33**	**87.50**

segmentation—Baseline, Baseline + GSCs, DDN, and MSN. The DDN achieves the best results of Pre of 85.20% for EX and the best Re and F1 of 69.70 and 71.83% for SE lesion segmentation. In contrast, MSN achieves the best results of Re, F1, and AUPR of 75.53, 79.33, and 87.50% for EX and the Pre and AUPR of 74.32 and 74.65% for SE. In general, DDN and MSN proposed models to achieve a higher performance than the Baseline and Baseline + GSCs models, Specifically when we look at the AUPR metric, which is commonly applied in the IDRiD dataset challenges. At the same time, the MSN outperforms the Baseline, Baseline + GSCs, and DDN models in both EX and SE lesions segmentation by 87.50 and 74.64%,

Table 5.2: Performance comparison on the IDRiD dataset of SE. Baseline refers to Unet with MobileNet encoder.

Method	Re	Pre	F1	AUPR
Baseline	64.41	65.88	65.13	67.48
Baseline + GSCs	61.66	**75.94**	68.06	69.79
DDN	**69.70**	74.10	**71.83**	74.40
MSN	64.14	74.32	68.85	**74.65**

respectively. In addition, we present some sample masks obtained from the Baseline,

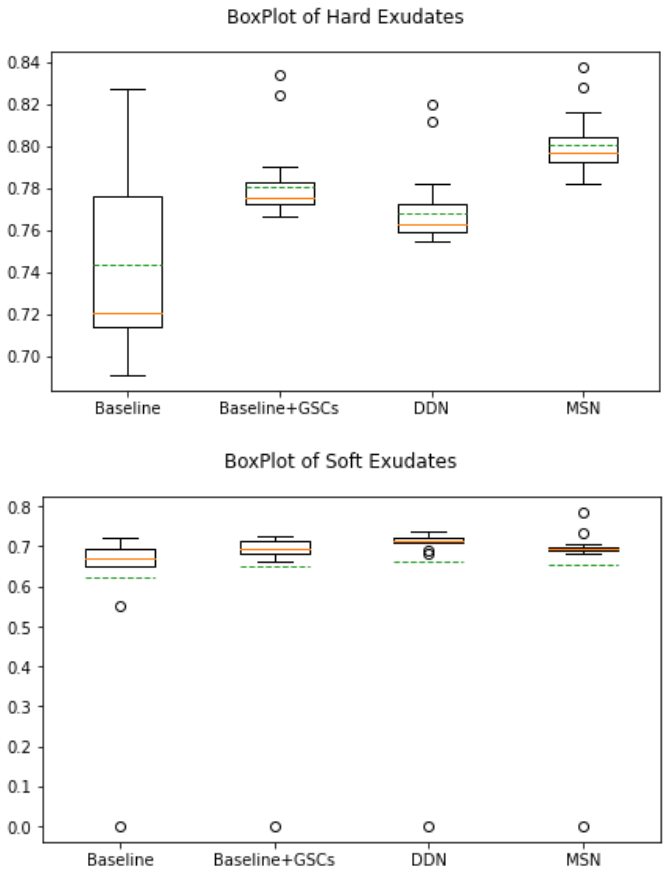


Figure 5.5: BoxPlot of Dice for Hard (top) and Soft (bottom) Exudates segmentation results (green dashed lines indicate the mean and the oranges indicate the median). All values outside the whiskers are considered outliers, which are marked with the (∘) symbol.

DDN, and MSN models for SE and EX lesion segmentation models to demonstrate the efficacy as shown in Figure 5.7. Figure 5.7 shows that the final output masks of DDN and MSN are better segmented than those final output masks of the baseline

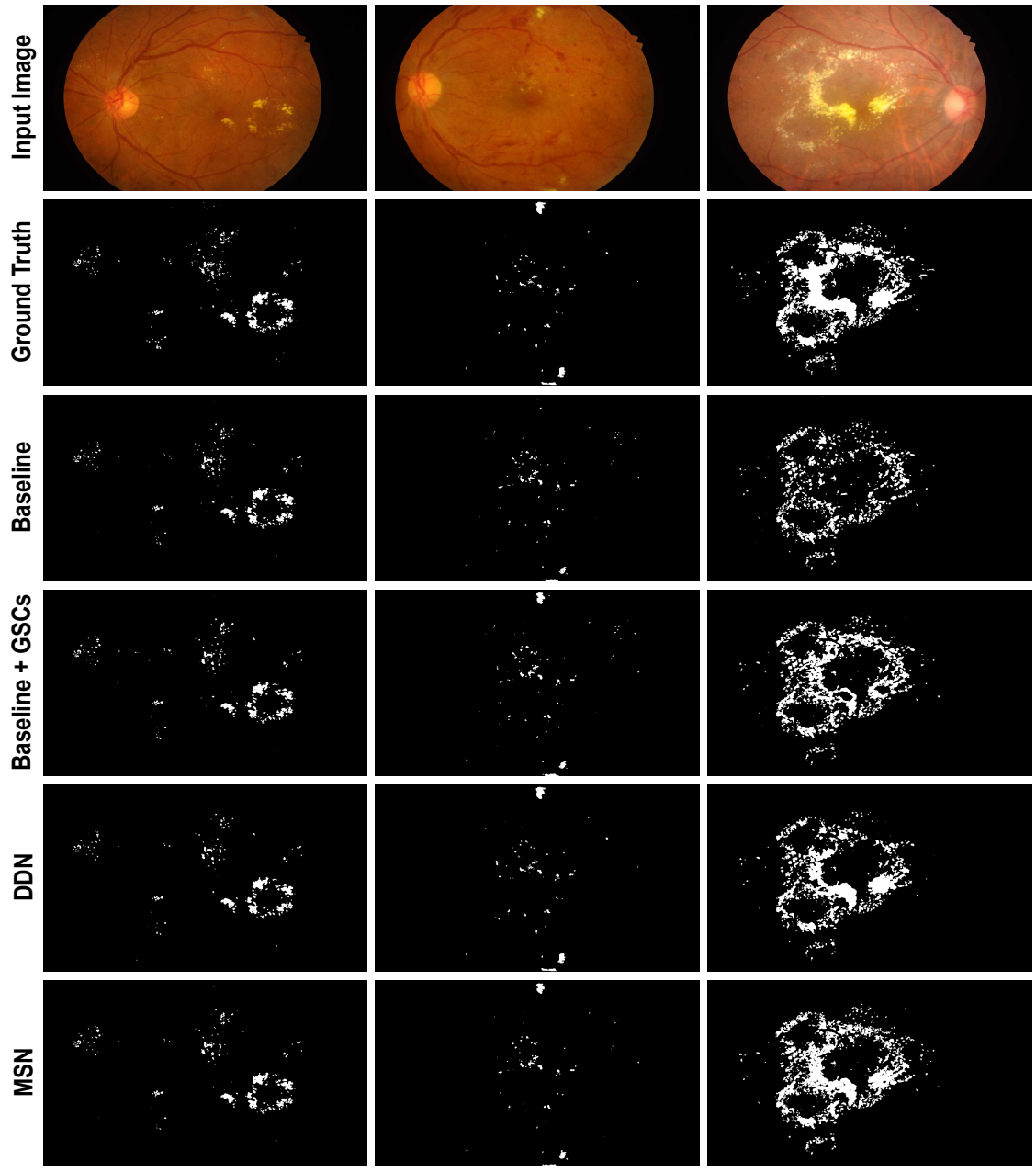


Figure 5.6: Hard Exudates Segmentation results.

model.

Finally, we show the boxplots of the F1 of the proposed DDN and MSN models, baseline, and baseline+GSCs. As shown in Figure 5.5 among the tested models, we can see that the MSN method achieves the highest mean and median for both EX and SE. Hence, the outliers in the positive whisker and data scattered are more consistent. At the same time, the DDN achieves the SE's highest mean and median.

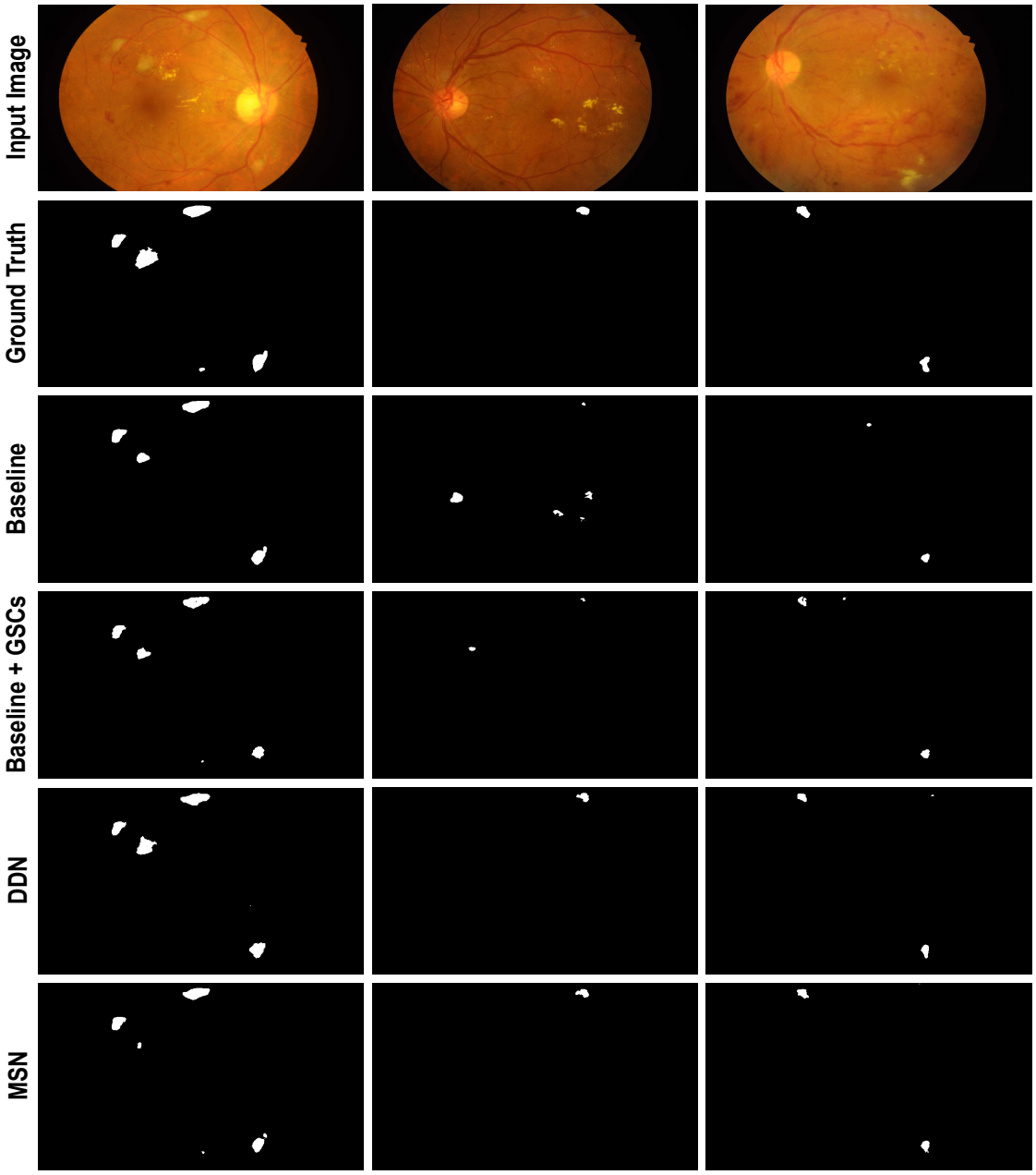


Figure 5.7: Soft Exudates Segmentation results.

It is good to refer to the statistical significance of the differences in performance between the proposed MSN and DDN compared with the Baseline model in terms of F1 for both EX and SE lesions. To do so, we used Student's t-test (significance level < 0.05) to specify the difference between F1 values. The p-values obtained are less than 0.05, indicating a statistical significance for EX; and higher than 0.05 for SE with the IDRiD dataset of both proposed models MSN and DDN.

5.3.3 Comparison with Existing Methods

To confirm the proposed method's effectiveness, we compare MSN and state-of-the-art on AUPR metric (same as the one used in the IDRiD competition) as shown in Table 5.3. The comparison includes the top 5 teams on the IDRiD competition [88] (the first five rows in the table), as well as CARNet [78], EAD-Net [79], L-Seg [89], and SAA [81]. DDN surpassed all the state-of-the-art

Table 5.3: Comparing the proposed models with the state-of-the-art methods on the IDRiD dataset.

Method	AUPR of SE	AUPR of EX
VRT (1st) [88]	69.95	71.27
PATech (2nd) [88]	-	**88.50**
IFLYTEK-MIG (3rd) [88]	65.88	87.41
SOONER (4th) [88]	53.95	73.90
SAIHST (5th) [88]	-	85.82
CARNet [78]	71.25	86.75
EAD-Net [79]	60.83	78.18
L-Seg [89]	71.13	79.45
SAA [81]	72.80	87.92
DDB - Proposed	74.40	85.30
MSM - Proposed	**74.65**	87.50

segmenting of the soft exudates segmentation by 1.55% of the AUPR metric. MSN surpassed all the state-of-the-art segmenting of the soft exudates segmentation by 1.80% of the AUPR metric. Also, achieved a comparable result for Hard exudates segmentation. The results of MSN of EX are comparable with SAA, while it had a good result of EX lesion segmentation but not with SE. PATech has the best results for EX retinal lesion segmentation but did not introduce SE retinal lesion segmentation results to compare. Therefore, as shown in Table 5.3, there is no method with the best results for the two retinal lesions segmentation at the same time.

5.4 Conclusions

This chapter presented two new deep-learning methods for the segmentation of hard and soft exudates, called DDN and MSN. Their effectiveness has been assessed on the IDRiD publicly dataset. The DDN obtains an F1 value of 71.83% and an AUPR value of 74.40% for soft exudates, while it performs a bit better for hard exudates, with a F1 score of 75.82% and an AUPR value of 85.30%. The second method proposed, MSM, improves the DDN method and achieves the best results in terms of F1 and AUPR metrics for both the soft and hard exudates. For SE, we have F1=68.85 and AUPR=74.65; for EX, we gave F1=79.33 and AUPR=87.5. The experimental results demonstrate that these methods outperform many state-of-the-art methods, especially in the case of soft exudates. In the next chapter, we will expand this study to other types of lesions caused by Diabetic Retinopathy.

CHAPTER 6

Eye Lesions Segmentation with Hybrid Multi-Scale Attention Modules

6.1 Introduction

The previous chapter has presented a first approach to segmenting exudates (EX and SE). Still, Diabetic Retinopathy produces other kinds of retinal eye lesions, such as microaneurysms (MA) and hemorrhages (HE). All of them are important to establish the degree of DR of a person, from mild to proliferative. Figure 6.1 shows some examples of those lesions. MA and HE are abnormal red lesions due to blood vessels breaking and indicate early stages of DR, whereas EX and SE appear as light lesions and indicate advanced stages of DR [79]. Several deep learning-based CADs

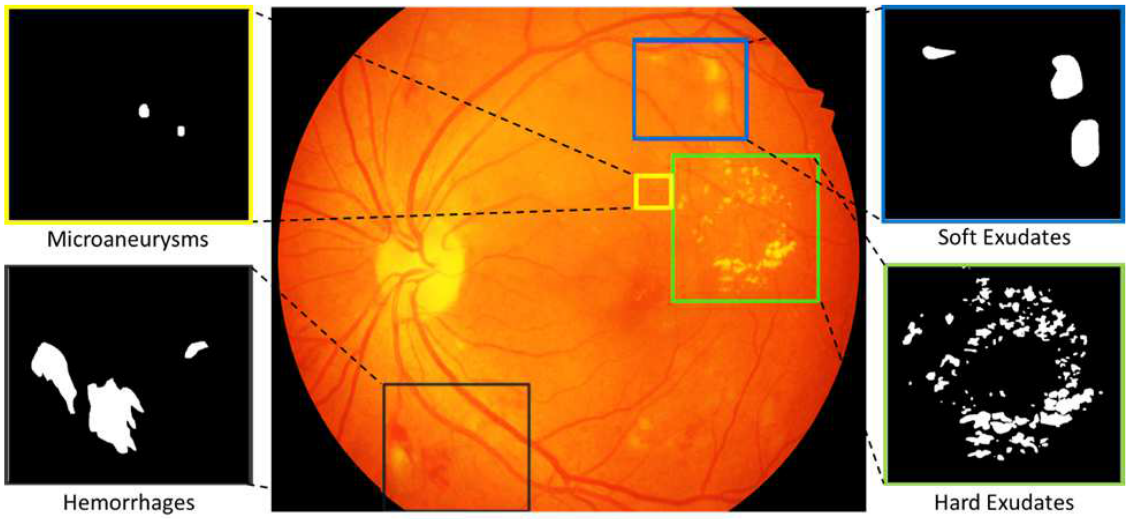


Figure 6.1: An example of a fundus image from the IDRiD dataset shows the retinal lesion types.

have been proposed for segmenting retinal eye lesions based on an encoder-decoder networks style, widely used in medical image segmentation [79; 90]. However, most of the deep learning models proposed in the literature have a large network size, such as [91; 92; 89; 78], which makes them computationally expensive during training or testing. Hence, we lose the advantage of reducing analysis time. Additionally, they may perform well with one type of lesion and fail with others [79; 81]. Therefore, a lightweight and accurate model is still needed to deal with retinal eye lesion segmentation.

In this chapter, we propose a precise retinal lesions segmentation model called LezioSeg that takes a fundus image as an input and produces its corresponding lesion mask for each lesion type individually.

The novelties of the proposed deep learning model are as follows:

1. Using a lightweight MobileNet deep neural network and higher classification accuracy as a backbone in our network.
2. Proposing the use of two effective multi-scale modules to encourage the LezioSeg network to extract lesion-relevant features from the fundus images. The first multi-scale module is used at the bottleneck of the LezioSeg network. It contains atrous spatial pyramid pooling blocks used at the neck of LezioSeg. The second multi-scale attention (SAT) module is used with the decoder to upsample the output of each decoder layer. It consists of a combination of 1×1, upsampling convolution layers, and an attention unit.
3. Integrating a GSCs mechanism at the decoder of LezioSeg to help the network focus on retinal lesion features.
4. Presenting extensive experiments on the well-known publicly available IDRiD and E-ophtha datasets and LezioSeg achieved competitive performance of all kinds of the lesion compared with the state-of-the-art systems with the IDRiD and outperformed them with the E-ophtha datasets.
5. Presenting a generalization of the IDRiD dataset experiments on the DDR dataset and LezioSeg achieved competitive performance.

6.2 LezioSeg

This section describes the proposed model, LezioSeg. As shown in Figure 6.2, it is composed of the following parts: the encoder network, also known as the Backbone, aims to encode the input image and produce feature maps. In addition, Multi-scale

modules help enhance image fundus feature extraction; one is after the encoder block at the LezioSeg neck, which is the ASPP layer that helps extract high-resolution features. Another SAT appears after the decoder network and works as a second multi-scale with an attention unit. Finally, the decoder network, also known as the Head, contains four layers with a GSCs mechanism in each layer [47] followed by SAT and an output layer. Each part can be explained in the following subsections.

6.2.1 Encoder Network

We use an ImageNet pre-trained MobileNet [85] encoder as a backbone in our network. We selected this model because MobileNet is a lightweight deep neural network with higher classification accuracy and a state-of-the-art backbone for many computer vision tasks [86; 87].

MobileNet uses depth-wise separable convolution, comprised of two layers, the depth-wise convolution, and the point-wise convolution. The depth-wise convolution layer applies a single filter into each input channel, and the point-wise convolution layer combines the output of depthwise using a 1×1 convolution to create a new feature. Also, MobileNet has two different global hyperparameters to reduce the computational cost effectively. One is the width multiplayer, and another is the resolution-wise multiplayer.

The main goal of the backbone in our proposed model is to encode the input eye fundus image and extract abstracted and meaningful features at different levels of scales.

6.2.2 Neck of LezioSeg

We employ ASPP to assist in extracting high-resolution feature maps and maximize capturing the contextual data of the micro lesions lost across the encoder

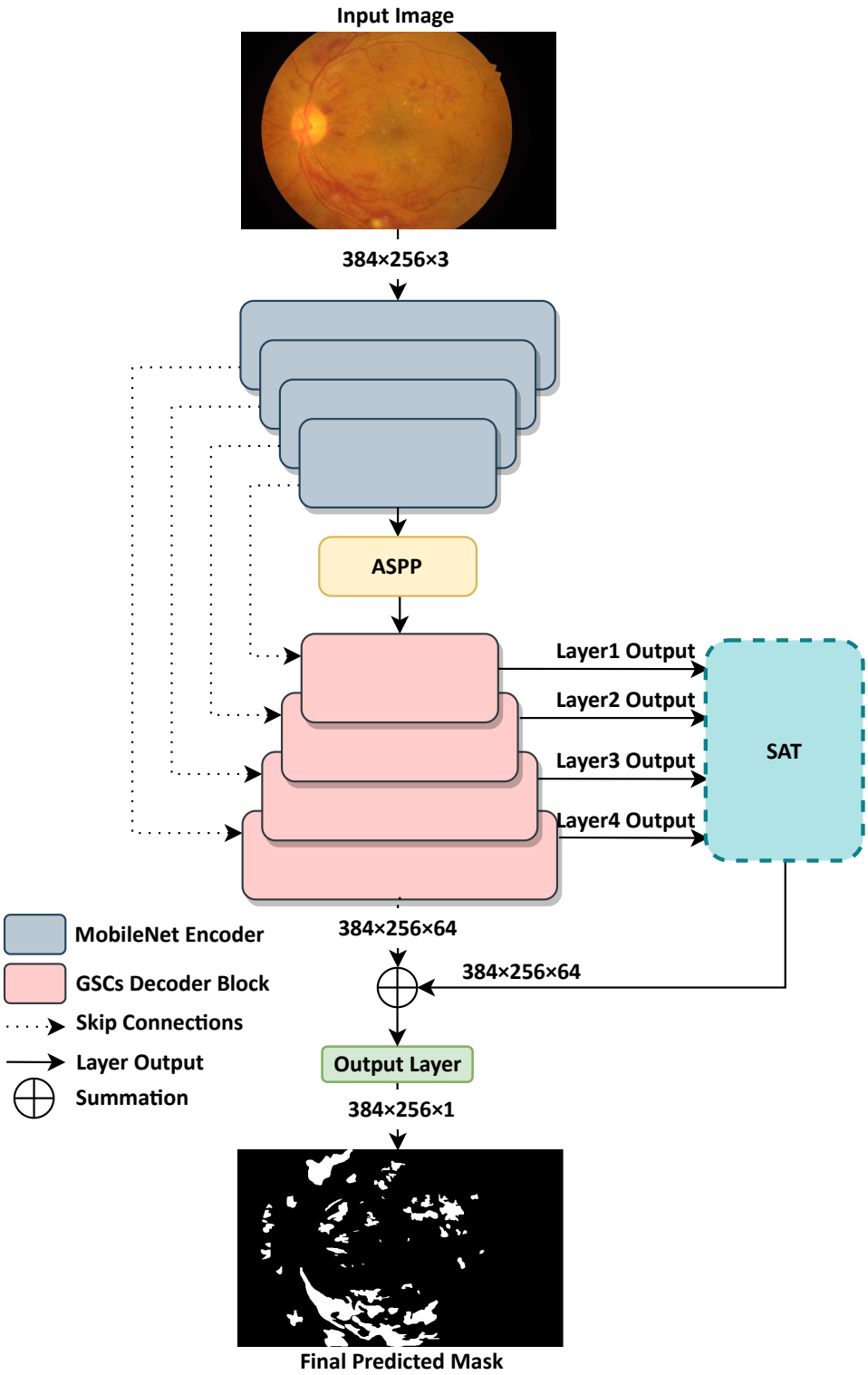


Figure 6.2: Schematic diagram of the LezioSeg network for lesions segmentation in fundus images.

multiple-scale. We provide the ASPP [83] as a link between the MobileNet encoder and the decoder to improve the proposed LezioSeg network. ASPP includes four

parallel atrous convolutions with different atrous rates. It is a mixture of atrous convolution and spatial pyramid pooling. It can capture contextual information at multiple scales for more accurate classification of the micro-objects of the retinal eye. ASPP can be expressed as follows:

$$y[p] = \sum_{k=1}^{K} x[p + r \cdot k] f[k] \tag{6.1}$$

For each pixel p on the output y and filter f, Atrous convolution is applied to the input x, where rate r determines the stride of sampling of the input image. Atrous convolution involves the input x with the filters produced by inserting $r - 1$ zeros between two consecutive filter values. By adjusting the rate r, we can modify the filter's receptive field. In this chapter, the ASPP module consists of one 1×1 convolution and three parallel 3×3 convolutions with rates of 6, 12, and 18, respectively, and an image-level feature produced by global average pooling. The resulting branches' resulting features are bilinearly upsampled to the input size and then concatenated and passed through another 1×1 convolution.

6.2.3 Decoder Network

The decoder block comprises four layers, SAT block, and the output layer. Each layer has the same GSCs mechanism internally. The output of each layer is fed into the SAT. The final output layer concatenates the decoder block output and the SAT output.

6.2.3.1 GSCs

LezioSeg has four GSCs blocks that boost feature map production to enhance the discrimination between the lesion and background pixels for retinal eye lesions segmentation. The four decoder blocks have the same internal GSCs architecture

in each decoder layer. GSCs are explained in Section 4.3.1. The GSCs mechanism in the decoder block receives feature maps from the corresponding Mobilenet encoder layers. Then, it concatenates them with the feature maps produced by the previous block (either the ASPP neck block or a previous decoder block).

Finally, the output of each decoder layer is fed into the SAT block. SAT will be explained in section 6.2.3.2. The final output mask is a binary image that includes a kind of retinal eye lesions and has a size identical to the size of the input image size.

6.2.3.2 Multi Scale Module Attention

The multi-scale with attention helps LezioSeg to maintain the multi-scale of each decoder layer output to consider features from the four decoder layers. Figure 6.3 shows the SAT block. In SAT, we first collect the four different copies of the features from the different stages of the decoder to extract features and to reduce the dimension of features from coarser stages to the same as the finest scale. Then, we unify the different scales using a 1×1 convolutional with a Kernal of 64. Next,

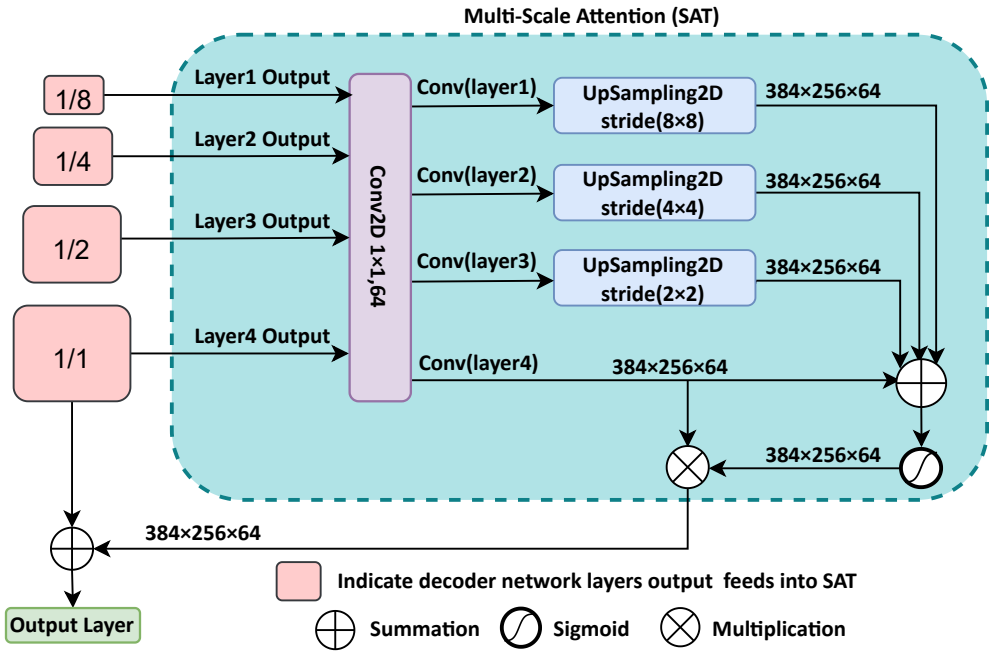


Figure 6.3: Structure of SAT block.

we upsample each scale size to the original size of the input image by UpSampled2D transposed convolution with different strides to make four upscaled features copies of the output features of the decoder layers. SAT can be expressed as follows:

$$SAT = \vartheta(\sum_{k=1}^{3} \uparrow (\varphi_{1\times1}(L_k))^{2*k} + \varphi_{1\times1}(L_4)) * L_4 \tag{6.2}$$

In this expression, L indicates the decoder network layers output, ϑ stands for the sigmoid activation function, $\uparrow$ indicates to UpSampled2D transposed convolution layer with a kernel size of 3×3, $\varphi_{1\times1}$ stands for the convolution operation with a kernel size of 1×1. where the $\varphi_{1\times1}(L_k)$, feature maps passes to a $\uparrow$ by $2 \times k$ strides, $k = (1, 2, 3)$, added with $\varphi_{1\times1}(L_4)$ are multiplied by sigmoid activation function weights, which helps to improve the discrimination between the lesion pixels and background pixels for lesion segmentation tasks, and multiplied of L_4. After that, we use another fusion to improve the discrimination process of lesion pixels by adding the SAT result to the final decoder network output, which is fed into 1×1 convolutional with Kernal 64 to be balanced with SAT output as follows:

$$Z = SAT + \varphi_{1\times1}(D_{out}) \tag{6.3}$$

where D_{out} stands for the final output of the decoder network.

Finally, Z feeds into the Lezioseg output layer to generate the predicted mask of lesion segmentation.

6.3 Experimental Results

This section outlines the experiments carried out to assess the performance of the proposed model, including a description of the experimental setup and an analysis of the outcomes.

6.3.1 Experimental Setup

To demonstrate the efficacy of LezioSeg, we employ three publicly datasets, namely, IDRiD [33], E-ophtha [34], and DDR [35]. Table 6.1 shows the general information of them. Each image of IDRiD has at least one mask labeled as one of four types

Table 6.1: Overview of experimental datasets

Dataset	EX	HE	MA	SE	Total	Image Size	Country
IDRiD	81	80	81	40	81	4288×2848	India
E-ophtha	47	-	148	-	195	2544×1696	France
DDR	171	194	124	42	225	Not fixed	China

of DR lesions EX, SE, MA, and HE. The dataset was split into 54 images as a training set (distributed of $54, 54, 54$, and 26 of EX, HE, MA, and SE) and the rest of 27 as a testing set (distributed of $26, 27, 27$, and 14 of EX, HE, MA, and SE). The e-Ophtha dataset has 47 images with masks labeled as exudate lesions and 148 labeled as microaneurism. We randomly split the dataset into 80% of images as a training set and 20% as a testing set. From the DDR dataset, we use the test set just for inference of our model to generalize the LezioSeg model.

The DDR test set is distributed as follows: $171, 194, 124$, and 42 for EX, HE, MA, and SE, respectively. We use the following training pipeline (including some data augmentation techniques to enrich the data and improve the regularity of the model) to process the images in the training set. First, each image is divided into four non-overlapped sub-images, and the corresponding sub-masks are constructed. We ignored the negative sub-images, i.e., the sub-images only with the background mask.

Next, we resize the sub-images and the sub-masks to 384×256. Then, we use the cubic interpolation mechanism for the images and the nearest neighbor for the masks.

After that, we apply horizontal flipping, affine, Gaussian noise, and rotation augmentation techniques for 12 repetition. We train each model for 50 epochs using an Adam optimizer and a batch size of 4. The 0.001 learning rate is used. We sampled a subset (20%) from the training set and used it as a validation set to save the best checkpoint of the trained models, and we used the binary cross-entropy as a loss function to train the models. During the inference phase, we only resize the input image to 768×512 and perform a full image segmentation process (i.e., no image splitting or image augmentation is used during the inference).

6.3.2 Ablation Study

In this section, we assess the performance of the different architectures on the test set images of the IDRiD and E-ophtha datasets. We conduct four experiments—Baseline, Baseline+GSCs, Baseline+GSCs+SacaleAtBegining (GSAB), and LezioSeg methods on the IDRiD dataset of each lesion EX, SE, HE, and MA separately. Tables 6.2, 6.3 present the performance of EX, SE, HE, and MA retinal lesion segmentation models. At the same time, we conduct two experiments of the same methods on the E-ophtha dataset of EX and MA. Table 6.4 presents the performance of EX and MA retinal lesion segmentation models with the E-ophtha dataset. We infer our methods on a single GTX 1070 GPU With 768×512 input image.

6.3.2.1 Experiments on the IDRiD Dataset

Tables 6.2, 6.3, shows that our proposed LezioSeg model achieves the best results on all metrics of SE, HE, and MA retinal lesions. Specifically, of the AUPR metric, the widespread usage of the IDRiD dataset challenge achieves a considerable improvement of SE, HE, and MA of 79.75, 38.07, and 67.44%, respectively. For EX evaluation, LezioSeg obtains the comparable with or slightly surpasses the baselines

except for the GSAB, which achieves a better of 87.50%. We show the boxplots of

Table 6.2: Experimental results of EX and SE on the IDRiD dataset

	EX				SE			
	ACC	**IOU**	**F1**	**AUPR**	**ACC**	**IOU**	**F1**	**AUPR**
Baseline	99.48	78.13	72.44	80.93	99.75	74.02	65.13	67.48
Baseline+GSCs	99.55	81.37	77.45	85.25	99.79	75.68	68.06	69.97
GSAB	**99.59**	**82.67**	**79.33**	**87.50**	99.79	76.14	68.85	74.65
LezioSeg	99.57	81.62	77.81	86.03	**99.82**	**79.66**	**74.60**	**79.75**

Table 6.3: Experimental results of MA and HE on the IDRiD dataset

	MA				HE			
	ACC	**IOU**	**F1**	**AUPR**	**ACC**	**IOU**	**F1**	**AUPR**
Baseline	99.9	57.03	24.81	32.56	99.24	69.15	56.20	62.53
Baseline+GSCs	99.9	60.02	**33.96**	33.69	99.22	67.63	53.01	58.56
GSAB	99.9	58.72	29.85	32.95	99.18	66.58	50.74	54.60
LezioSeg	**99.9**	**60.06**	33.64	**38.07**	**99.30**	**70.45**	**58.76**	**67.44**

the F1 of the proposed model, Baseline, Baseline+GSCs, and GSAB models on the IDRiD dataset. As shown in Figure 6.4 among the tested models, the proposed model has the highest mean, median, smallest standard deviation, and fewer outliers of the F1 metric for the SE and HE. Also, has the highest mean and median compared with the GSAB and Baseline, and the outliers are equal to the top whisker of the Baseline and the bottom whisker of the GSAB for the MA. Finally, For the EX, LezioSeg obtains the highest mean and median compared with Baseline and Baseline+GSCs, and the outliers in the positive whisker equal that of Baseline+GSCs and GSAB. The Figure shows that our method achieves the best performance for both SE and HE while it achieves the second-best performance for both EX and MA. It is good to refer to the statistical significance of the differences in performance between the proposed LezioSeg and the best model of Baseline, Baseline+GSCs, and GSAB in terms of F1 for each lesion. To do so, we used Student's t-test (significance level < 0.05) to specify the difference between F1 values. The p-values obtained are less than 0.05, indicating a statistical significance for EX, HE, and MA and higher than 0.05 for SE with the

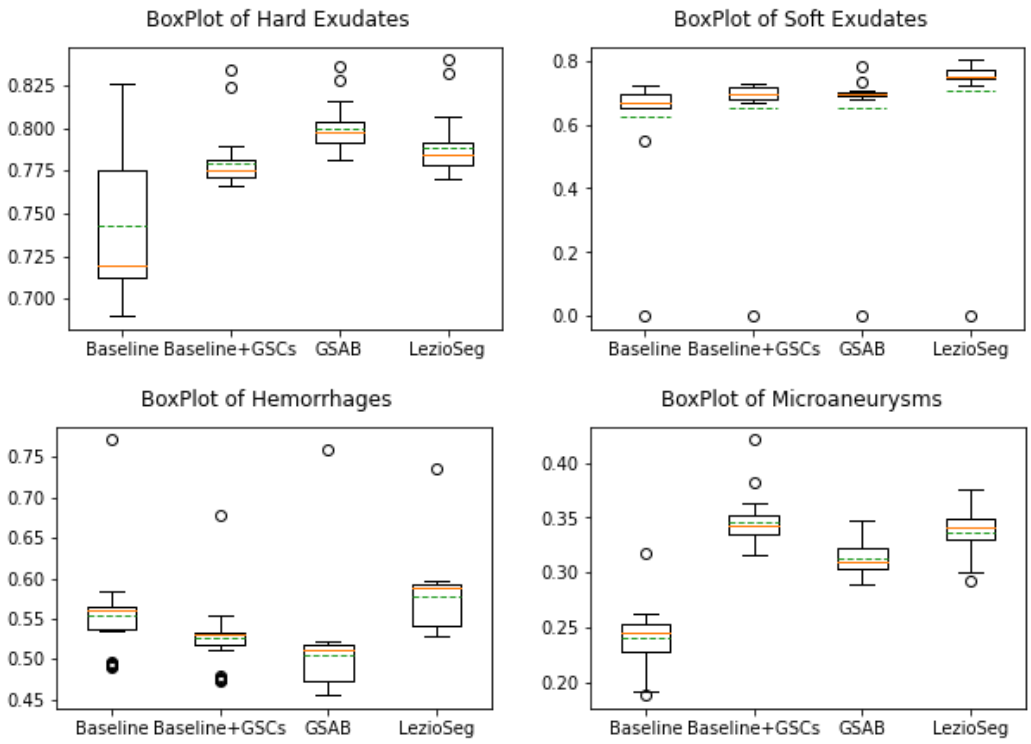


Figure 6.4: BoxPlot of F1 for EX, SE, HE, and MA Segmentation results of IDRiD dataset(green dashed lines indicate the mean, and the oranges indicate the median). All values outside the whiskers are demonstrated outliers, which are marked with the (∘) symbol.

IDRiD dataset. Finally, we present samples of predicted masks obtained for the LezioSeg proposed model, Baseline, Baseline+GSCs, and GSAB lesion segmentation methods to demonstrate the efficacy of the IDRiD dataset. Figure 6.5 shows samples of EX and SE, and Figure 6.6 shows the samples of HE and MA.

6.3.2.2 Experiments on the E-ophtha Dataset

We also conducted experiments on the E-ophtha dataset to show the impact of introducing LezioSeg to the same baseline model used in experiments of the IDRiD dataset. Table 6.4 presents the performance of EX and MA retinal lesion segmentation models of the E-ophtha dataset. As shown in Table 6.1 E-ophtha introduces labeled data of EX and MA. The LezioSeg model achieves the best segmentation results of all metrics of MA. For EX, LezioSeg achieves the best results of ACC, IOU, and F1 metrics of 99.85, 71.74, and 60.8%, respectively. In contrast,

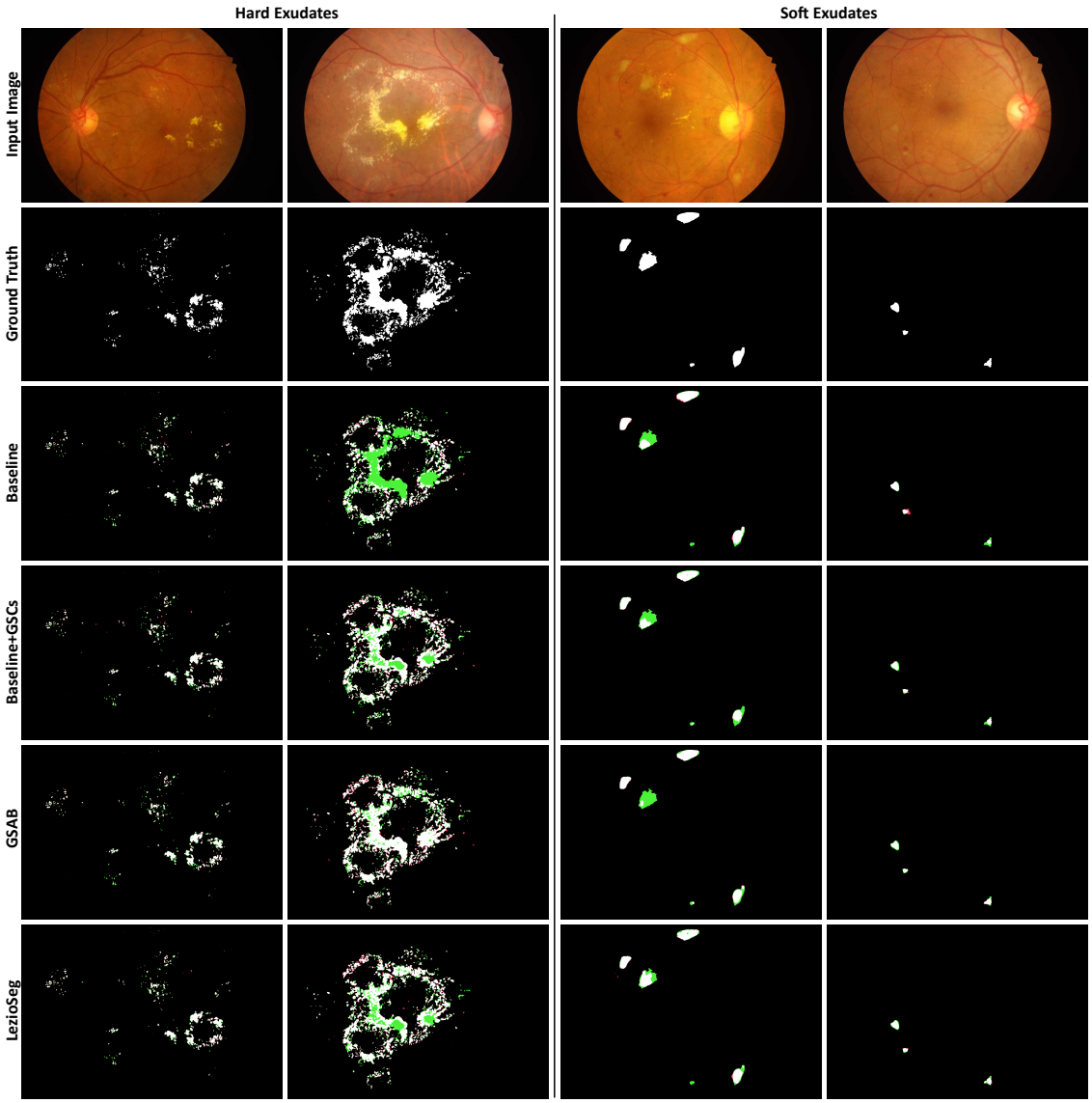


Figure 6.5: Hard and Soft exudates segmentation results on the IDRiD dataset.

Table 6.4: Experimental results of EX and MA on the E-ophtha dataset

	EX				MA			
	ACC	**IOU**	**F1**	**AUPR**	**ACC**	**IOU**	**F1**	**AUPR**
Baseline	99.85	69.43	56.13	62.84	99.98	60.09	33.6	30.01
Baseline+GSCs	99.84	70.68	58.67	**67.15**	99.98	60.43	34.57	29.6
GSAB	99.85	71.55	60.4	64.53	99.97	57.87	27.23	22.52
LezioSeg	**99.85**	**71.74**	**60.8**	63.04	**99.98**	**61.18**	**36.57**	**32.11**

the Baseline+GSCs model obtains the best of the AUPR metric of 67.15%. We show the boxplots of the F1 of the proposed model, Baseline, Baseline+GSCs, and GSAB. As shown in Figure 6.7 among the tested models, the proposed model has the highest

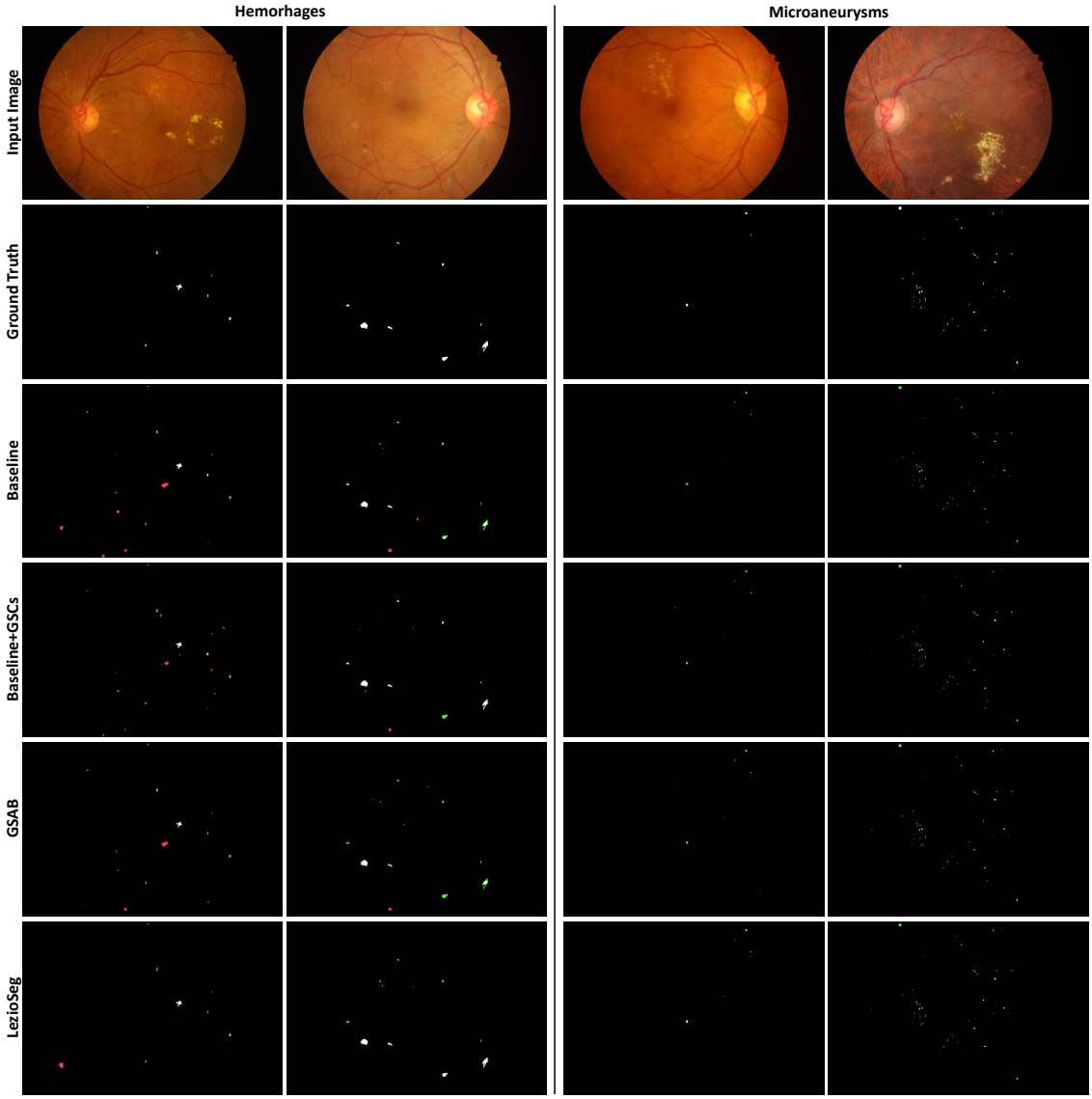


Figure 6.6: Hemmorages and Microaneurysms segmentation results on the IDRiD dataset.

mean and median of the F1 metric for the EX. For MA, the mean of LezioSeg is the best, has well-scattered data, has no outliers, and the bottom whisker is the best than the top whisker of GSAB. The Student's t-test of the statistical differences in performance between the proposed LezioSeg and the best model of Baseline, Baseline+GSCs, and GSAB in terms of F1 shows that the p-values obtained less than 0.05 indicate a statistical significance for MA with E-ophtha dataset. Finally, we present samples of predicted masks obtained for the LezioSeg proposed model, Baseline, Baseline+GSCs, and GSAB lesion segmentation methods of EX and MA

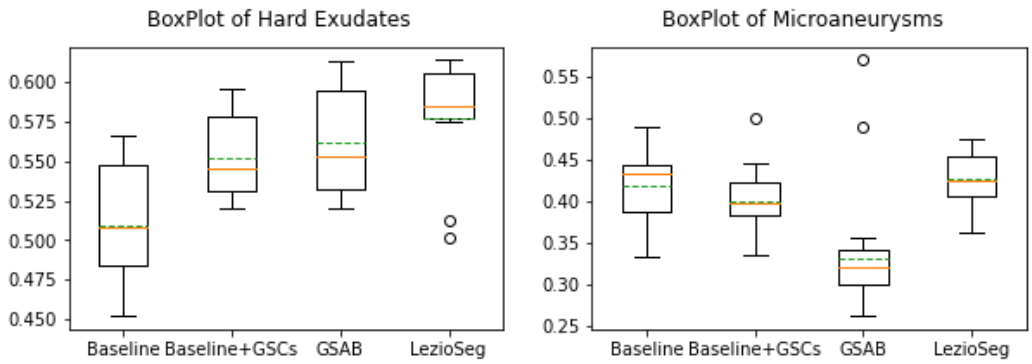


Figure 6.7: BoxPlot of F1 for EX and MA Segmentation results of the E-ophtha dataset(green dashed lines indicate the mean, and the oranges indicate the median). All values outside the whiskers are regarded outliers, which are marked with the (∘) symbol.

to demonstrate the efficacy of the E-ophtha dataset as shown in Figure 6.8.

6.3.3 Comparison with Existing Lesions Segmentation Methods

To ensure the proposed method's efficacy, we compare LezioSeg and state-of-the-art using the AUPR metric of the IDRiD and E-ophtha dataset. The comparison with IDRiD includes the top 3 teams in the IDRiD challenge [88], L-Seg [89], CASENet [93], DeepLabV3+ [49], HEDNet+cGAN [91], EAD-Net [79], and SAA [81]. Table 6.5 shows the proposed model LezioSeg has achieved the top performance of all the state-of-the-art results of segmenting the SE retinal lesion by achieving the AUPR of 79.75% and obtained the highest mAUPR value of 67.82% surpasses the SAA. Also, we achieved a second-best value of HE compared with the top 1 team VRT on the IDRiD dataset of 67.44% and LezioSeg ranked No.4 on EX segmentation.

For the E-ophtha dataset, the comparison includes CASENet [93], DeepLabV3+ [49], and L-Seg [89]. Table 6.6 shows the LezioSeg surpasses DeepLabV3+ by 7.92% in segmenting EX, 15.24% is over than the L-Seg in segmenting MA% of the AUPR metric. At the same time, the LezioSeg achieved the best mAUPR of 47.58%.

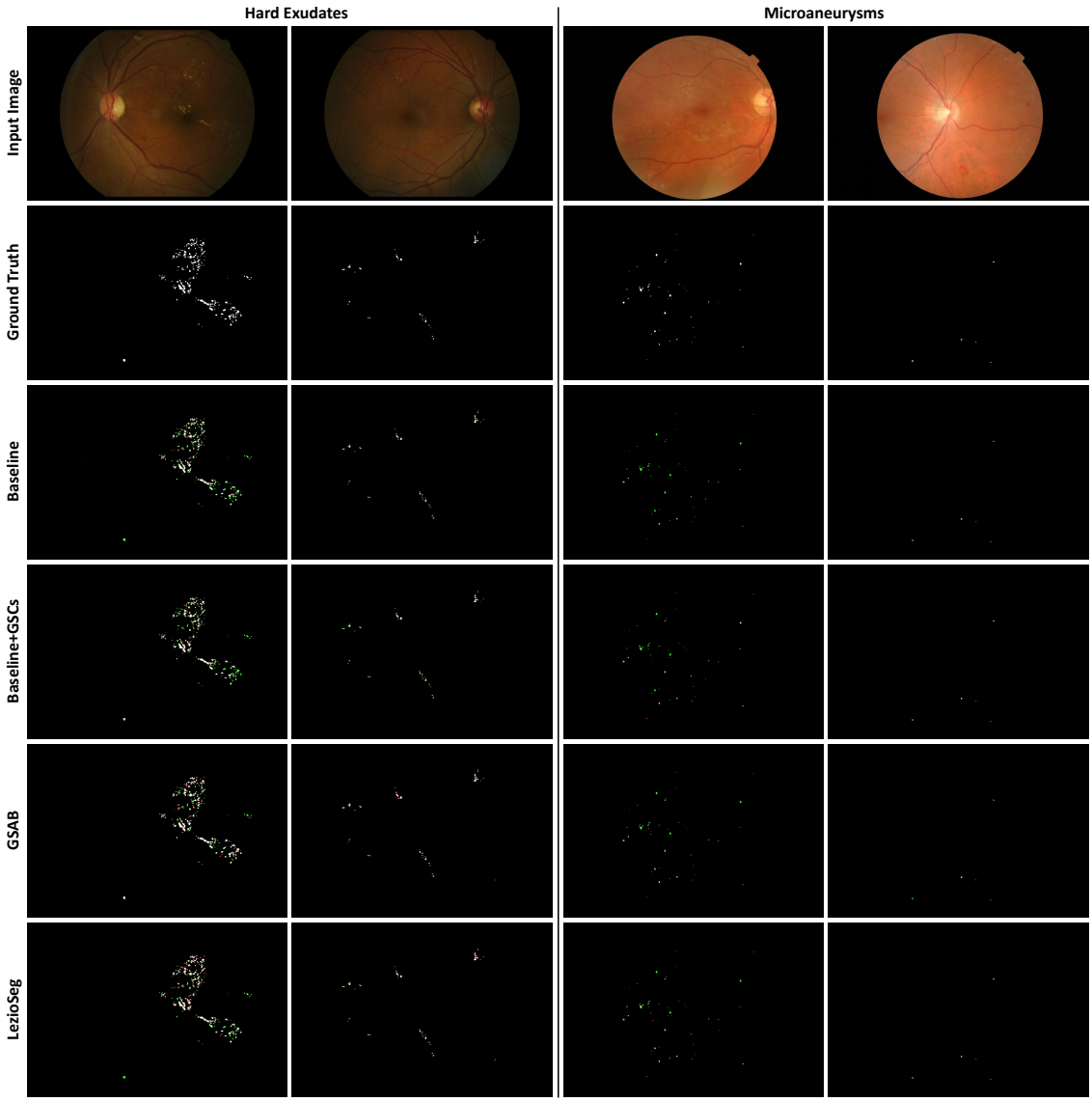


Figure 6.8: Hard exudates and Microaneurysms segmentation results on the E-ophtha dataset.

6.3.4 Evaluating the generalization capability of LezioSeg

It is also meaningful to study the generalization capability of the methods over different domains under different imaging conditions. In this section, we validate the generalization capability of proposed models by training with the images from the train set of the IDRiD dataset and testing on the test set of the DDR dataset, which is captured from another source. Table 6.7 shows that LezioSeg achieves the best performance for EX, MA, and SE of AUPR of 57.0, 19.90, and 30.35% and

Table 6.5: Comparison with existing methods for lesions segmentation on the IDRiD dataset.

Method	EX	HE	SE	MA	mAUPR
VRT (1st) [88]	71.27	**68.04**	69.95	49.51	64.69
PATech (2nd) [88]	88.5	64.9	-	47.4	-
iFLYTEK-MIG (3rd) [88]	87.41	55.88	65.88	**50.17**	64.84
L-Seg [89]	79.45	63.74	71.13	46.27	65.15
CASENet [93]	75.64	44.62	39.92	32.75	48.23
DeepLabV3+ [49]	71.18	47.72	59.12	15.14	48.29
HEDNet+cGAN [91]	84.05	48.12	48.39	43.92	56.12
EAD-Net [79]	78.18	56.49	60.83	24.08	54.90
SAA [81]	88.12	67.04	72.81	41.52	67.37
LezioSeg	86.03	67.44	**79.75**	38.07	**67.82**

(-) Denotes that the data is not reported in the paper.

Table 6.6: Comparison with existing methods for lesions segmentation on the E-ophtha dataset

Method	EX	MA	mAUPR
CASENet [93]	17.15	15.65	16.40
DeepLabV3+ [49]	55.12	0.45	27.78
L-Seg [89]	41.71	16.87	29.29
LezioSeg	**63.04**	**32.11**	**47.58**

Table 6.7: Performance comparison of generalization of the different methods.

Method	EX	HE	MA	SE	mAUPR
Unet++ [94]	50.87	24.39	13.70	9.73	24.67
DeepLabV3+ [49]	53.49	**38.94**	2.23	21.20	28.97
CASENet [93]	27.77	26.25	10.05	13.04	19.28
L-Seg [89]	55.46	35.86	10.52	26.48	32.08
LezioSeg (Ours)	**57.0**	32.55	**17.90**	**30.35**	**34.45**

mAUPR of 34.45%. On the other hand, LezioSeg surpassed all the state-of-the-art in segmenting of EX, MA, SE, and mAUPR by 1.56, 4.20, 3.87, and 2.37%, respectively.

6.4 Conclusions

This chapter proposed LezioSeg, an efficient method for lesion segmentation using multi-scale modules and gated skip connections. Different experiments were done to demonstrate the efficacy of LezioSeg on publicly available IDRiD and E-ophtha datasets. The IDRiD dataset obtained 79.75% of AUPR metrics for SE, 86.3% for EX, 67.44% for HE, and 38.07% for MA, respectively. On the other hand, the E-ophtha dataset achieved 63.04% of AUPR for EX and 32.11% for MA, respectively. In addition, when the trained model with IDRiD is tested with a different dataset, DDR, we could see that LezioSeg achieved the best performance for EX, MA, and SE compared to 4 other state-of-the-art methods. The experimental result demonstrates that LezioSeg achieves a good generalization model that can be used with images from different sources.

CHAPTER 7

Concluding Remarks and Future works

7.1 Summary and Conclusions

Diabetic Retinopathy and Glaucoma are diseases that affect the human eye and may cause blindness. Early detection can reduce eye illness progression. Such eye diseases can be detected by studying fundus images. However, analyzing thousands of fundus images is costly and requires the availability of solid, experienced ophthalmologists. Despite there are several methods that have proposed solutions for this problem, many weak points have been identified in the related work.

Therefore, I have done this work in the frame of the research projects funded by Instituto de Investigación Carlos III, done with collaboration of researchers from ITAKA group at Universitat Rovira i Virgili and ophthalmology group at Institut d'Investigació Pere Virgili. The goal approached in this doctoral thesis is to prove that computer vision technologies like deep learning techniques may be used for detecting different eye lesions and diseases with high performance indicators.

This thesis has proposed new efficient computer vision-based methods based on deep learning for segmenting the optic disc and eye lesions and detecting glaucoma in fundus images. Various publicly available and private fundus image datasets were used to develop and validate the proposed methods and to prove their generalization abilities. Different evaluation metrics and statistical analysis (e.g., Student's t-test and Bland–Altman plots) techniques were used to assess the performance of the proposed methods.

From the work done in this thesis, the following conclusions can be formulated:

- Current deep learning methods segmentation and classification for eye fundus images can be improved using ensembles of models. Chapter 3 has dealt the case of optic disc segmentation and Chapter 4 the classifiction of glaucoma.
- Using andness-directed aggregation operators (like, OWA, WPM and EXP) gives better results than using arithmetic averages. We found that the

conjunctive policy (simultaneity,high andness) with a reduced subset of the models gives the best results. We also proved that the three operators give equivalent results.

- The detection of the small and different types of DR lesions requires some sophisticated DL architectures. Multi-scale models have been studied, to encourage the deep learning network to extract relevant features about the lesions. We found that inserting the multi-scale techniques with the mechanism of gated skip connections at the head of the network improved spatial opacity, which enhanced the segmentation process results for small retinal lesions.
- The final architecture (LezioSeg), based on multi-scale and gated skip connections is also interesting as it uses less resources than other methods and give quite robust results when used with different datasets.

7.2 Future Research Lines

The work presented in this doctoral thesis contributes to the interpretation analysis of eye fundus images. Due to the impact of eye diseases in the quality of life of people, we believe this is an interesting and important field of research that still has some open problems. The support and encouragement of the ophthalmologists from Hospital Sant Joan de Reus also confirms that this topic is worth to being studied.

Several directions for future work have been identified during this work.

- For improving the fusion of the ensemble of models, it would be good to optimize the fuzzy OWA aggregation function presented in chapter 4 that dynamically fuses the top-performing glaucoma prediction classifiers, using metaheuristics optimization methods to find the best values for the parameters in OWA fuzzy aggregation.
- This thesis proposed a CAD system for lesion segmentation, which is used

to detect Diabetic Retinopahty, and a CAD system for Glaucoma prediction. However, having one single integrated CAD system to diagnose the most common and risky eye diseases (like DR grading, glaucoma, and age-related macular degeneration) will help ophthalmologists to make better decisions.

- The work in this thesis focused on a binary classification problem. In future work, the plan is to test the proposed approaches on problems with multiple classes and to study the effect of each of the contributions in this thesis on the multi-class situation. Different challenges may be considered, for example, the presence of imbalanced classes in the segmentation tasks or the order between classes if they represent different degrees of the same disease.
- Using deep learning methods as a black box for segmentation and prediction tasks introduced solutions with high performance to many problems, specifically medical image analysis. However, the decision-maker, such as an ophthalmologist, still needed ways to explain the causality of how the CAD gives a specific prediction for input fundus images. The explainability of deep learning models is nowadays in the focus of many researchers.

References

[1] Dujmovic J. Soft Computing Evaluation Logic: The LSP Decision Method and Its Applications. John Wiley & Sons; 2018.

[2] Ogurtsova K, da Rocha Fernandes J, Huang Y, Linnenkamp U, Guariguata L, Cho NH, et al. IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040. Diabetes research and clinical practice. 2017;128:40-50.

[3] Bicket AK, Le JT, Azuara-Blanco A, Gazzard G, Wormald R, Bunce C, et al. Minimally invasive glaucoma surgical techniques for open-angle glaucoma: An overview of cochrane systematic reviews and network meta-analysis. JAMA ophthalmology. 2021;139(9):983-9.

[4] Romero-Aroca P, Valls A, Moreno A, Sagarra-Alamo R, Basora-Gallisa J, Saleh E, et al. A clinical decision support system for diabetic retinopathy

screening: creating a clinical support application. Telemedicine and e-Health. 2019;25(1):31-40.

[5] Esteva A, Chou K, Yeung S, Naik N, Madani A, Mottaghi A, et al. Deep learning-enabled medical computer vision. NPJ digital medicine. 2021;4(1):1-9.

[6] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review. Computational intelligence and neuroscience. 2018;2018.

[7] Chakraborty S, Mukherjee A, Chatterjee D, Maji P, Acharjee S, Dey N. A semi-automated system for optic nerve head segmentation in digital retinal images. In: 2014 International conference on information technology. IEEE; 2014. p. 112-7.

[8] Ruiz-Alonso S, Villate-Beitia I, Gallego I, Lafuente-Merchan M, Puras G, Saenz-del Burgo L, et al. Current insights into 3D bioprinting: an advanced approach for eye tissue regeneration. Pharmaceutics. 2021;13(3):308.

[9] Wong WL, Su X, Li X, Cheung CMG, Klein R, Cheng CY, et al. Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. The Lancet Global Health. 2014;2(2):e106-16.

[10] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.

[11] Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. Annals of translational medicine. 2020;8(11).

[12] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for

biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer; 2015. p. 234-41.

[13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014.

[14] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770-8.

[15] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: European conference on computer vision. Springer; 2016. p. 630-45.

[16] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 4510-20.

[17] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700-8.

[18] Kornblith S, Shlens J, Le QV. Do better imagenet models transfer better? In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019. p. 2661-71.

[19] Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging. 2016;35(5):1299-312.

[20] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? Advances in neural information processing systems. 2014;27.

[21] Zhao S, Li B, Xu P, Keutzer K. Multi-source domain adaptation in the deep learning era: A systematic survey. arXiv preprint arXiv:200212169. 2020.

[22] Zhang Y, Tang H, Jia K, Tan M. Domain-symmetric networks for adversarial domain adaptation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019. p. 5031-40.

[23] Zhang Y, Yang Q. An overview of multi-task learning. National Science Review. 2018;5(1):30-43.

[24] Jadon S. An overview of deep learning architectures in few-shot learning domain. arXiv preprint arXiv:200806365. 2020.

[25] Pourpanah F, Abdar M, Luo Y, Zhou X, Wang R, Lim CP, et al. A review of generalized zero-shot learning methods. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022.

[26] Dujmović J. Numerical Comparison of Idempotent Andness-Directed Aggregators. In: International Conference on Modeling Decisions for Artificial Intelligence. Springer; 2021. p. 67-77.

[27] Dujmović J, Torra V. Properties and comparison of andness-characterized aggregators. International Journal of Intelligent Systems. 2021;36(3):1366-85.

[28] Yager RR. Quantifier guided aggregation using OWA operators. International Journal of Intelligent Systems. 1996;11(1):49-73.

[29] Dujmović J, Allen III WL. Soft computing logic decision making in strategic conservation planning for water quality protection. Ecological Informatics. 2021;61:101167.

[30] Sivaswamy J, Krishnadas S, Joshi GD, Jain M, Tabish AUS. Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation. In: 2014 IEEE 11th international symposium on biomedical imaging (ISBI). IEEE; 2014. p. 53-6.

[31] Orlando JI, Fu H, Breda JB, van Keer K, Bathula DR, Diaz-Pinto A, et al. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical image analysis. 2020;59:101570.

[32] Fumero F, Alayón S, Sanchez JL, Sigut J, Gonzalez-Hernandez M. RIM-ONE: An open retinal image database for optic nerve evaluation. In: 2011 24th international symposium on computer-based medical systems (CBMS). IEEE; 2011. p. 1-6.

[33] Porwal P, Pachade S, Kamble R, Kokare M, Deshmukh G, Sahasrabuddhe V, et al. Indian diabetic retinopathy image dataset (IDRiD): a database for diabetic retinopathy screening research. Data. 2018;3(3):25.

[34] Decenciere E, Cazuguel G, Zhang X, Thibault G, Klein JC, Meyer F, et al. TeleOphta: Machine learning and image processing methods for teleophthalmology. Irbm. 2013;34(2):196-203.

[35] Li T, Gao Y, Wang K, Guo S, Liu H, Kang H. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. Information Sciences. 2019;501:511-22.

[36] Boyd K, Eng KH, Page CD. Area under the precision-recall curve: point estimates and confidence intervals. In: Joint European conference on machine learning and knowledge discovery in databases. Springer; 2013. p. 451-66.

[37] Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC medical imaging. 2015;15(1):1-28.

[38] Fourcade A, Khonsari R. Deep learning in medical image analysis: A third eye for doctors. Journal of stomatology, oral and maxillofacial surgery. 2019;120(4):279-88.

[39] Sevastopolsky A. Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. Pattern Recognition and Image Analysis. 2017;27(3):618-24.

[40] Bhatkalkar BJ, Reddy DR, Prabhu S, Bhandary SV. Improving the performance of convolutional neural network for the segmentation of optic disc in fundus images using attention gates and conditional random fields. IEEE Access. 2020;8:29299-310.

[41] Fu H, Cheng J, Xu Y, Wong DWK, Liu J, Cao X. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE transactions on medical imaging. 2018;37(7):1597-605.

[42] Tabassum M, Khan TM, Arsalan M, Naqvi SS, Ahmed M, Madni HA, et al. CDED-Net: Joint segmentation of optic disc and optic cup for glaucoma screening. IEEE Access. 2020;8:102733-47.

[43] Dietterich TG. Ensemble Learning, The Handbook of Brain Theory and Neural Networks, MA Arbib. Cambridge, MA: MIT Press; 2002.

[44] Mangipudi PS, Pandey HM, Choudhary A. Improved optic disc and cup segmentation in Glaucomatic images using deep learning architecture. Multimedia Tools and Applications. 2021:1-21.

[45] Kumar A, Kim J, Lyndon D, Fulham M, Feng D. An ensemble of fine-tuned convolutional neural networks for medical image classification. IEEE journal of biomedical and health informatics. 2016;21(1):31-40.

[46] Qummar S, Khan FG, Shah S, Khan A, Shamshirband S, Rehman ZU, et al. A deep learning ensemble approach for diabetic retinopathy detection. IEEE Access. 2019;7:150530-9.

[47] Jabreel M, Abdel-Nasser M. Promising crack segmentation method based on gated skip connection. Electronics Letters. 2020;56(10):493-5.

[48] Jha D, Riegler MA, Johansen D, Halvorsen P, Johansen HD. Doubleu-net: A deep convolutional neural network for medical image segmentation. In: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS). IEEE; 2020. p. 558-64.

[49] Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV); 2018. p. 801-18.

[50] Wu T, Tang S, Zhang R, Cao J, Zhang Y. Cgnet: A light-weight context guided network for semantic segmentation. IEEE Transactions on Image Processing. 2020;30:1169-79.

[51] Romera E, Alvarez JM, Bergasa LM, Arroyo R. Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. IEEE Transactions on Intelligent Transportation Systems. 2017;19(1):263-72.

[52] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence. 2017;39(12):2481-95.

[53] Wang Y, Zhou Q, Xiong J, Wu X, Jin X. Esnet: An efficient symmetric network for real-time semantic segmentation. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). Springer; 2019. p. 41-52.

[54] Chaurasia A, Culurciello E. Linknet: Exploiting encoder representations for efficient semantic segmentation. In: 2017 IEEE Visual Communications and Image Processing (VCIP). IEEE; 2017. p. 1-4.

[55] Treml M, Arjona-Medina J, Unterthiner T, Durgesh R, Friedmann F, Schuberth P, et al. Speeding up semantic segmentation for autonomous driving. In: MLITS, NIPS Workshop. vol. 2; 2016. .

[56] Mary VS, Rajsingh EB, Naik GR. Retinal fundus image analysis for diagnosis of glaucoma: a comprehensive survey. IEEE Access. 2016.

[57] Xu Y, Liu J, Lin S, Xu D, Cheung CY, Aung T, et al. Efficient optic cup detection from intra-image learning with retinal structure priors. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2012. p. 58-65.

[58] Turgut B. Pearls for correct assessment of optic disc at glaucoma diagnosis. Journal-Pearls for Correct Assessment of Optic Disc at Glaucoma Diagnosis. 2017.

[59] Jani K, Srivastava R, Srivastava S, Anand A. Computer aided medical image analysis for capsule endoscopy using conventional machine learning and deep learning. In: 2019 7th International Conference on Smart Computing & Communications (ICSCC). IEEE; 2019. p. 1-5.

[60] Gómez-Valverde JJ, Antón A, Fatti G, Liefers B, Herranz A, Santos A, et al. Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning. Biomedical optics express. 2019;10(2):892-913.

[61] Kim J, Tran L, Chew EY, Antani S. Optic disc and cup segmentation for glaucoma characterization using deep learning. In: 2019 IEEE 32nd

International Symposium on Computer-Based Medical Systems (CBMS). IEEE; 2019. p. 489-94.

[62] Civit-Masot J, Domínguez-Morales MJ, Vicente-Díaz S, Civit A. Dual machine-learning system to aid glaucoma diagnosis using disc and cup feature extraction. IEEE Access. 2020;8:127519-29.

[63] Diaz-Pinto A, Morales S, Naranjo V, Köhler T, Mossi JM, Navea A. CNNs for automatic glaucoma assessment using fundus images: an extensive validation. Biomedical engineering online. 2019;18(1):1-19.

[64] Sreng S, Maneerat N, Hamamoto K, Win KY. Deep Learning for Optic Disc Segmentation and Glaucoma Diagnosis on Retinal Images. Applied Sciences. 2020;10(14):4916.

[65] Deepa N, Esakkirajan S, Keerthiveena B, Dhanalakshmi SB. Automatic Diagnosis of Glaucoma using Ensemble based Deep Learning Model. In: 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS). vol. 1. IEEE; 2021. p. 536-41.

[66] Guo J, Azzopardi G, Shi C, Jansonius NM, Petkov N. Automatic determination of vertical cup-to-disc ratio in retinal fundus images for glaucoma screening. IEEE Access. 2019;7:8527-41.

[67] Hemelings R, Elen B, Barbosa-Breda J, Blaschko MB, De Boever P, Stalmans I. Deep learning on fundus images detects glaucoma beyond the optic disc. Scientific Reports. 2021;11(1):1-12.

[68] Hervella ÁS, Rouco J, Novo J, Ortega M. End-to-end multi-task learning for simultaneous optic disc and cup segmentation and glaucoma classification in eye fundus images. Applied Soft Computing. 2021:108347.

[69] Ko YC, Wey SY, Chen WT, Chang YF, Chen MJ, Chiou SH, et al. Deep learning assisted detection of glaucomatous optic neuropathy and potential designs for a generalizable model. PloS one. 2020;15(5):e0233079.

[70] Ajitha S, Akkara JD, Judy M. Identification of glaucoma from fundus images using deep learning techniques. Indian Journal of Ophthalmology. 2021;69(10):2702.

[71] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 2818-26.

[72] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1251-8.

[73] Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence; 2017. .

[74] Zoph B, Vasudevan V, Shlens J, Le QV. Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 8697-710.

[75] Abdel-Nasser M, Melendez J, Moreno A, Puig D. The impact of pixel resolution, integration scale, preprocessing, and feature normalization on texture analysis for mass classification in mammograms. International Journal of Optics;2016.

[76] Mahmoud H, Abdel-Nasser M, Omer OA. Computer aided diagnosis system for skin lesions detection using texture analysis methods. In: 2018 International Conference on Innovative Trends in Computer Engineering (ITCE). IEEE; 2018. p. 140-4.

[77] Chakravarty A, Sivaswamy J. Glaucoma classification with a fusion of segmentation and image-based features. In: 2016 IEEE 13th international symposium on biomedical imaging (ISBI). IEEE; 2016. p. 689-92.

[78] Guo Y, Peng Y. CARNet: Cascade attentive RefineNet for multi-lesion segmentation of diabetic retinopathy images. Complex & Intelligent Systems. 2022:1-21.

[79] Wan C, Chen Y, Li H, Zheng B, Chen N, Yang W, et al. EAD-net: a novel lesion segmentation method in diabetic retinopathy using neural networks. Disease Markers. 2021;2021.

[80] Li Y, Zhu M, Sun G, Chen J, Zhu X, Yang J. Weakly supervised training for eye fundus lesion segmentation in patients with diabetic retinopathy. Mathematical Biosciences and Engineering. 2022;19(5):5293-311.

[81] Bo W, Li T, Liu X, Wang K. SAA: Scale-Aware Attention Block For Multi-Lesion Segmentation Of Fundus Images. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). IEEE; 2022. p. 1-5.

[82] Pan X, Jin K, Cao J, Liu Z, Wu J, You K, et al. Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus fluorescein angiography based on deep learning. Graefe's Archive for Clinical and Experimental Ophthalmology. 2020;258(4):779-85.

[83] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence. 2017;40(4):834-48.

[84] Zhang J. Multi-source remote sensing data fusion: status and trends. International Journal of Image and Data Fusion. 2010;1(1):5-24.

[85] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:170404861. 2017.

[86] Widiansyah M, Rasyid S, Wisnu P, Wibowo A. Image segmentation of skin cancer using MobileNet as an encoder and linknet as a decoder. In: Journal of Physics: Conference Series. vol. 1943. IOP Publishing; 2021. p. 012113.

[87] Mohamed NA, Zulkifley MA, Abdani SR. Spatial pyramid pooling with atrous convolutional for mobilenet. In: 2020 IEEE student conference on research and development (SCOReD). IEEE; 2020. p. 333-6.

[88] Porwal P, Pachade S, Kokare M, Deshmukh G, Son J, Bae W, et al. Idrid: Diabetic retinopathy–segmentation and grading challenge. Medical image analysis. 2020;59:101561.

[89] Guo S, Li T, Kang H, Li N, Zhang Y, Wang K. L-Seg: An end-to-end unified framework for multi-lesion segmentation of fundus images. Neurocomputing. 2019;349:52-63.

[90] Villaret M, et al. Segmenting the Optic Disc Using a Deep Learning Ensemble Model Based on OWA Operators. In: Artificial Intelligence Research and Development: Proceedings of the 23rd International Conference of the Catalan Association for Artificial Intelligence. vol. 339. IOS Press; 2021. p. 305.

[91] Xiao Q, Zou J, Yang M, Gaudio A, Kitani K, Smailagic A, et al. Improving lesion segmentation for diabetic retinopathy using adversarial learning. In: International Conference on Image Analysis and Recognition. Springer; 2019. p. 333-44.

[92] Yan Z, Han X, Wang C, Qiu Y, Xiong Z, Cui S. Learning mutually local-global u-nets for high-resolution retinal lesion segmentation in fundus images. In: 2019

IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE; 2019. p. 597-600.

[93] Yu Z, Feng C, Liu MY, Ramalingam S. Casenet: Deep category-aware semantic edge detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 5964-73.

[94] Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer; 2018. p. 3-11.

UNIVERSITAT
ROVIRA i VIRGILI