الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي
جامعة سعيدة د. مولاي الطاهر
كلية الرياضيات و الإعلام الآلي و الاتصالات السلكية و اللاسلكية
قسم: الإعلام الآلي

Master's thesis in computer science

specialty : Computer Network and Distributed Systems

Theme

Eye disease detection using deep learning

- **Prosented by :**

KHALDOUN Douaa Amina

CHOUIH Amaur

- **Headed by :**

KOUIDRI Siham

academic year 2025-2026

Acknowledgements

In the name of Allah, the Most Gracious, the Most Merciful (Bismillah al-Rahman al-Rahim).

First and foremost, all praise and gratitude are due to Allah for His infinite wisdom, guidance, and strength throughout our dissertation journey. Without His blessings, this accomplishment would not have been possible.

We are particularly grateful to **Ms. KOUIDRI Siham** for her invaluable guidance and mentorship throughout this dissertation process. Her expertise, insightful feedback, and unwavering support were instrumental in shaping and refining our research.

Furthermore, our deepest appreciation goes to the members of the jury for their time and effort in evaluating this work. Your honorable presence and valuable insights are sincerely acknowledged.

We would like to express our deepest appreciation to our family for their unwavering love and support throughout our studies. Their constant encouragement and understanding during challenging times were invaluable.

We would also like to express our heartfelt thanks to everyone who has assisted us in any way, big or small. Your contributions, both tangible and intangible, have played a significant role in bringing this dissertation to fruition.

Dedication

In the name of Allah, the Most Gracious, the Most Merciful.

First and foremost, all praise and gratitude are due to Allah for His infinite wisdom, guidance, and strength throughout this journey. Without His blessings, this accomplishment would not have been possible.

To myself, for the perseverance, dedication, and countless hours of hard work invested in bringing this project to life. I thank myself for never giving up despite every challenge along the way.

To my beloved parents, whose unwavering love, sacrifices, and prayers have provided the foundation for my success. Any achievement I accomplish belongs to them before anyone else.

To my dear siblings, whose encouragement and support have been a continuous source of motivation and hope throughout this journey.

To my supervisor, for the invaluable guidance, advice, and mentorship provided during the realisation of this dissertation. Their expertise, insightful feedback, and unwavering support were instrumental in shaping and refining this work.

To my binome, for the teamwork, cooperation, and shared efforts that made this collaboration both successful and memorable.

To all my friends who helped, supported, and encouraged me along the way your presence made this journey easier and more meaningful.

To everyone who contributed, directly or indirectly, to the completion of this work, I extend my deepest gratitude.

To my future self, may you always remember the hard work and dedication that brought you here, and may you continue to strive for excellence in all your endeavors. Your achievements are a testament to your determination and ambition.

KHALDOUN Douaa Amina

Dedication

First of all, I would like to sincerely thank myself for all the effort, patience, and determination that helped me complete this work despite every challenge along the way.

I dedicate this success to my beloved parents, especially my father, Djilali, and my mother, whose love, sacrifices, support, and prayers have always been my greatest source of strength. Any success that comes from this work belongs to them before anyone else.

I would also like to thank all my family members and friends for their continuous encouragement, support, and motivation throughout this journey.

My sincere thanks as well to Binome for the teamwork, cooperation, and shared efforts during this project.

Finally, I would like to express my deep gratitude to my supervisor for the valuable guidance, advice, and support provided during the realisation of this dissertation.

Thank you to everyone who contributed, directly or indirectly, to the completion of this work.

CHOUIH Ameur

الملخص:

تقترح هذه الأطروحة نهجًا متعدد المهام للتعلم العميق للتشخيص الآلي لأمراض الشبكية من صور قاع العين. يعتمد النظام على خط أنابيب مشروط من مرحلتين يجمع بين EfficienNet-B3 -للتصنيف متعدد العلامات لأمراض العين و YOLO-V8 للكشف عن آفات اعتلال الشبكية السكري. يقوم نموذج التصنيف الذي تم تدريبه على مجموعة بيانات ODIR-5K ، بتنسيط خط أنابيب المعالجة المسبقة الذي يدمج CLAHE و Grad-CAM و SAHI بشكل مشروط عند اكتشاف اعتلال الشبكية السكري قبل إجراء اكتشاف الآفة على مجموعة بيانات DDR.

وتظهر النتائج التجريبية تحسنًا كبيرًا في الأداء مقارنة بالطرق الحالية. كما أن دمج التصورات القابلة للتفسير المستندة

إلى Grad-CAM يعزز أيضًا دعم القرار السريري. وأخيرا يسلط هذا البحث الضوء على قدرة تعميم البنية خفيفة الوزن والمحسنة مما يجعلها مناسبة للنشر في البيئات الطبية ذات الموارد المحدودة.

الكلمات الرئيسية : التعلم العميق، التعلم متعدد المهام، تشخيص أمراض الشبكية، اعتلال شبكية السكري، EfficienNet-B3

YOLO-V8 اكتشاف الآفات التعميم.

Abstract

This thesis proposes a multi-task deep learning approach for the automated diagnosis of retinal diseases from fundus images. The system is based on a two-stage conditional pipeline combining EfficientNet-B3 for the multi-label classification of ocular pathologies and YOLOv8s for the detection of diabetic retinopathy lesions. The classification model, trained on the ODIR-5K dataset, conditionally activates a preprocessing pipeline integrating CLAHE, Grad-CAM, and SAHI when diabetic retinopathy is detected, before performing lesion detection on the DDR dataset.

Experimental results demonstrate a significant improvement in performance compared to existing approaches. The integration of interpretable Grad-CAM-based visualizations also enhances clinical decision support. Finally, this research highlights the generalization capability of a lightweight and optimized architecture, making it suitable for deployment in resource-limited medical environments.

Keywords: *Deep Learning, Multi-Task Learning, Retinal Disease Diagnosis, Diabetic Retinopathy, EfficientNet-B3, YOLOv8s, Lesion Detection, Generalization.*

Résumé

Cette thèse propose une approche d'apprentissage profond multi-tâches pour le diagnostic automatisé des maladies rétiniennes à partir d'images du fond de l'œil. Le système est basé sur un pipeline conditionnel en deux étapes combinant EfficientNet-B3 pour la classification multi-label des pathologies oculaires et YOLOv8s pour la détection des lésions de rétinopathie diabétique. Le modèle de classification, entraîné sur l'ensemble de données ODIR-5K, active conditionnellement un pipeline de prétraitement intégrant CLAHE, Grad-CAM et SAHI lorsque une rétinopathie diabétique est détectée, avant d'effectuer la détection des lésions sur l'ensemble de données DDR.

Les résultats expérimentaux démontrent une amélioration significative des performances par rapport aux approches existantes. L'intégration de visualisations interprétables basées sur la Grad-CAM améliore également l'aide à la décision clinique. Enfin, cette recherche met en évidence la capacité de généralisation d'une architecture légère et optimisée, ce qui la rend adaptée au déploiement dans des environnements médicaux à ressources limitées.

Mots-clés : *Apprentissage profond, apprentissage multi-tâches, diagnostic des maladies rétiniennes, rétinopathie diabétique, EfficientNet-B3, YOLOv8s, détection de lésions, Généralisation.*

Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under the ROC Curve
AMP	Automatic Mixed Precision
AP	Average Precision
BCE	Binary Cross-Entropy
CAD	Computer-Aided Diagnosis
CAM	Class Activation Mapping
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
DDR	Deep Diabetic Retinopathy (Dataset)
DL	Deep Learning
DNN	Deep Neural Network
DR	Diabetic Retinopathy
EX	Hard Exudates
F1	F1-Score (Harmonic mean of Precision and Recall)
FPN	Feature Pyramid Network
Grad-CAM	Gradient-weighted Class Activation Mapping
GRU	Gated Recurrent Unit
HE	Hemorrhages
IEEE	Institute of Electrical and Electronics Engineers
IoU	Intersection over Union
KNN	K-Nearest Neighbors
LR	Learning Rate
LSTM	Long Short-Term Memory
MA	Microaneurysms
mAP	Mean Average Precision
MBConv	Mobile Inverted Bottleneck Convolution
ML	Machine Learning
NPDR	Non-Proliferative Diabetic Retinopathy
OCT	Optical Coherence Tomography
OCT-A	Optical Coherence Tomography Angiography
OD	Optic Disc
ODIR	Ocular Disease Intelligent Recognition
PDR	Proliferative Diabetic Retinopathy
RGB	Red, Green, Blue (Color Space)
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SAHI	Slicing Aided Hyper Inference
SE	Soft Exudates / Squeeze-and-Excitation (context-dependent)
SSD	Single Shot MultiBox Detector
SVM	Support Vector Machine
TML	Traditional Machine Learning
TTA	Test Time Augmentation
UV	Ultraviolet
ViT	Vision Transformer
YOLO	You Only Look Once (Object Detection Architecture)

General Introduction

Context and Motivation

Ophthalmic diseases are considered among the most significant public health challenges worldwide. Among these disorders, **Diabetic Retinopathy (DR)** is one of the primary causes of acquired blindness in the working-age population, mainly due to the increasing prevalence of diabetes mellitus. Studies from Algeria indicate that type 2 diabetes is the most common form among diabetic patients [15][16], and its prevalence varies by region and clinical setting, reflecting the growing burden of this disease in the country.

These findings are similar to those observed in many developing countries, where insufficient access to specialized ophthalmic services increases the risk of avoidable vision loss.

Besides diabetic retinopathy, other ocular diseases such as **Age-related Macular Degeneration (AMD), glaucoma, cataract, myopia, and hypertensive retinopathy** also play a major role in irreversible visual impairment worldwide. Although these diseases differ in their pathological origins, they all require early diagnosis to improve treatment outcomes and reduce the progression of vision deterioration. Unfortunately, many of these conditions remain asymptomatic in their early phases, making systematic screening essential for effective clinical management.

In this context, **Artificial Intelligence (AI)** and deep learning technologies have become increasingly important in ophthalmic medical imaging. Convolutional Neural Networks (CNNs), trained on large-scale retinal image datasets, have demonstrated strong capabilities in detecting subtle pathological features such as microaneurysms, hemorrhages, exudates, and optic disc abnormalities. Consequently, AI-based diagnostic systems offer promising solutions for automated retinal screening and clinical decision support, particularly in regions where access to experienced ophthalmologists and advanced diagnostic resources is limited.

Problem Statement

Despite the importance of early retinal screening, traditional diagnostic methods remain limited because they rely on the manual analysis of fundus images by ophthalmologists. This process is often time-consuming, expensive, and highly dependent on specialist expertise. Moreover, the growing prevalence of vision-threatening eye diseases continues to increase the demand for retinal screening beyond the available healthcare capacity [17, 18].

Beyond these practical limitations, automated retinal analysis faces important technical challenges. One major issue is multi-label disease classification under highly imbalanced datasets, where a single fundus image may contain multiple ocular pathologies at the same time. In such datasets, rare diseases like hypertensive retinopathy are significantly underrepresented, causing standard classification models to perform poorly on minority classes and reducing their diagnostic sensitivity [19].

A second major challenge is the precise localization of retinal lesions for disease grading. Diabetic retinopathy includes lesions with highly variable sizes and shapes, ranging from extremely small microaneurysms to large hemorrhages. Microaneurysms are particularly difficult to detect because they may become nearly invisible after image resizing, while hemorrhages exhibit strong morphological variability. These factors significantly reduce the effectiveness of conventional object detection models and make accurate lesion localization more challenging in retinal imaging tasks [20].

Existing automated systems generally treat disease classification and lesion localization as separate tasks, reducing both interpretability and diagnostic reliability. To address this limitation, this work proposes a two-stage retinal analysis pipeline that combines multi-label classification with lesion-level localization through a unified spatial guidance mechanism.

Objectives of the Thesis

The main objective of this work is to develop an automated deep learning–based ophthalmic diagnostic system capable of performing both disease classification and lesion localization. The proposed framework aims to classify multiple ocular diseases from fundus images using the ODIR-5K dataset, while also detecting diabetic retinopathy lesions such as microaneurysms, hemorrhages, and exudates using the DDR dataset.

To achieve this, a two-stage conditional pipeline is designed, where the classification stage activates the lesion detection stage when diabetic retinopathy is identified. In addition, Grad-CAM is integrated to improve model interpretability by highlighting disease-relevant regions, thereby enhancing clinical explainability and supporting ophthalmologists during diagnosis.

Thesis Structure

This thesis is organized into four chapters followed by a general conclusion. The first chapter presents the theoretical foundations of deep learning and convolutional neural networks. The second chapter introduces the medical context of the studied retinal diseases, ophthalmic imaging modalities, and a state-of-the-art analysis of existing approaches. The third chapter details the proposed methodology, including the datasets used, the architecture of the two-stage conditional pipeline, and the preprocessing strategies adopted. The fourth chapter presents the experimental results and their comparative analysis with existing works. Finally,

the general conclusion summarizes the main scientific contributions of this research and highlights future directions for clinical deployment and further system improvements.

Chapitre I

Deep Learning

1.1 Introduction

Artificial Intelligence has become a major force behind today's technological progress, especially in areas that require high accuracy, such as medical diagnosis. In the past, most research relied on traditional machine learning methods, where experts had to manually extract features from the data. With the development of deep learning, this process has changed significantly. Deep learning models can automatically learn complex patterns from large datasets, particularly from high-resolution images, making them very effective for medical image analysis.

The goal of this chapter is to present the theoretical background of our work. It begins with an overview of traditional machine learning techniques and then moves toward more advanced deep learning architectures, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). In addition, well-known models such as VGG, ResNet, and MobileNet are discussed to provide a clear understanding of the methods that can be used for the automatic detection of eye diseases in the practical part of this study.

1.2 The Evolution of Machine Learning to Deep Learning

1.2.1 What is machine learning?

Machine learning is a branch of artificial intelligence (AI) that allows computers to learn from data and make decisions without being explicitly programmed for every task. Instead of following fixed instructions, machine learning systems analyze data, identify patterns, and use this information to make predictions or decisions.

It includes a variety of techniques and algorithms that enable models to improve their performance over time as they are exposed to more data. Through this learning process, systems become more accurate and efficient in tasks such as classification, prediction, and pattern recognition. Machine learning is therefore considered a key component of AI, focusing specifically on data-driven learning and model improvement[21].

1.2.2 Traditional Machine Learning Algorithm

Traditional Machine Learning (TML) methods remain widely used in biomedical image analysis due to their interpretability, low computational cost, and effectiveness on relatively small datasets. They also provide a practical framework for combining information from multiple medical data sources, making them suitable for various healthcare applications.

Unlike deep learning approaches, which automatically learn relevant features directly from raw data, traditional machine learning techniques rely on handcrafted feature extraction. In this process, discriminative image features are first manually designed and extracted, after which classical classifiers such as Support Vector Machines (SVM) or k-Nearest Neighbors (k-NN) are employed to perform the final classification. This dependence on manually engineered features represents the main distinction between traditional machine learning and deep learning methods[22].

Support Vector Machines (SVM)

Support Vector Machines (SVMs) are supervised learning algorithms that determine the optimal decision hyperplane by maximizing the margin between classes in high-dimensional feature spaces. They are widely used for both classification and regression tasks due to their strong theoretical foundation and effectiveness in handling complex data distributions[23].

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a simple supervised learning algorithm used mainly for classification, but it can also handle regression tasks. It works by comparing a new data point to its *k* closest neighbors and assigning a class based on the most frequent label among them. The algorithm relies on the idea that similar data points are located near each other in the feature space[24].

Decision Tree

A Decision Tree is a non-parametric supervised learning algorithm used for both classification and regression tasks. It has a tree-like structure composed of a root node, internal nodes, branches, and leaf nodes.

The model works by recursively splitting the dataset into smaller subsets based on the most informative features. Using a greedy divide-and-conquer strategy, it selects the best split at each step to separate the data into more homogeneous groups. This process continues in a top-down manner until most data points are assigned to specific class labels[25].

Random Forest

Random Forest is a versatile machine learning algorithm that combines the predictions of multiple decision trees to improve accuracy and robustness. It can handle both classification and regression tasks. Each decision tree splits data using key features, and Random Forest aggregates the results of all trees, reducing overfitting and enhancing overall performance[26].

Naïve Bayes

Naïve Bayes is a supervised machine learning algorithm commonly used for classification tasks, especially in areas such as text classification and spam detection. It is based on probability theory and applies Bayes' theorem to predict the class of a given data point.

Naïve Bayes belongs to the family of generative learning algorithms. This means it models how the data is distributed within each class, rather than directly learning the boundary between classes. In contrast to discriminative models like logistic regression, which focus on distinguishing between categories, Naïve Bayes estimates the probability that a data point belongs to a particular class based on its features[27].

Logistic Regression

Logistic Regression is a widely used supervised machine learning algorithm for classification tasks. It predicts discrete or categorical outcomes based on input features. For example, it can determine whether a loan should be approved by analyzing factors like savings, income, and credit score.

As one of the most popular classification methods in machine learning and AI, logistic regression is versatile and interpretable. It can be applied to many problems, and there are different types of logistic regression depending on the number of classes and the nature of the data. Additionally, understanding logistic regression involves exploring its underlying regression analysis, use cases, and variations to choose the best approach for a given problem[28].

1.2.3 Types of Machine Learning

In the machine learning field, there are two main types of tasks: supervised and unsupervised.

What is Supervised Learning?

Reformulated version:

Supervised learning is a machine learning approach that relies on labeled data, where each input sample is associated with a known target output. During training, the model

learns the relationship between input features and their corresponding labels by using examples with predefined outcomes. The objective is to build a predictive function capable of accurately mapping inputs to outputs.

Once the learning process is completed, the trained model can analyze new, unseen data and predict the most appropriate labels based on the patterns it has learned. The performance of the model is then assessed by comparing its predictions with the true labels in a separate test dataset, allowing its accuracy and generalization capability to be evaluated[1].

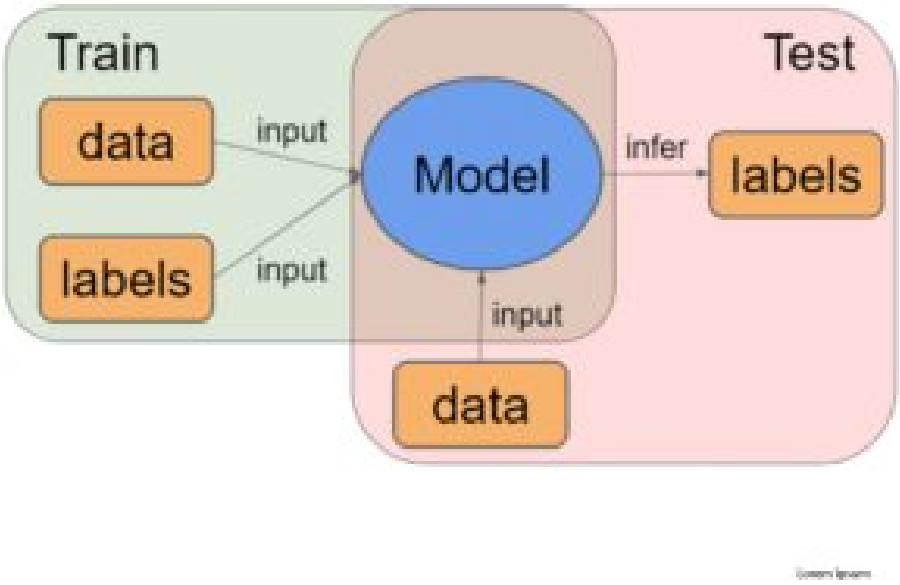


Figure 1.1: Supervised learning workflow[1].

What is Unsupervised Learning?

Unsupervised learning is a type of machine learning that operates without labeled data or ground truth. It relies on algorithms to analyze datasets and automatically identify patterns, similarities, and differences within the data. The main objective is to uncover the underlying structure of the dataset.

Unlike supervised learning, the model does not receive predefined outputs. Instead, it processes the input data and infers results on its own, which may appear as clusters or extracted features. One of the most common techniques is clustering, where data points are grouped based on their similarities, making it useful for exploratory data analysis and discovering hidden relationships.

In practice, unsupervised learning is especially valuable when dealing with large datasets that are difficult or costly to label. It allows models to learn from vast amounts of data and can be combined with supervised learning to enhance overall performance, accuracy, and robustness, particularly in applications such as medical imaging.

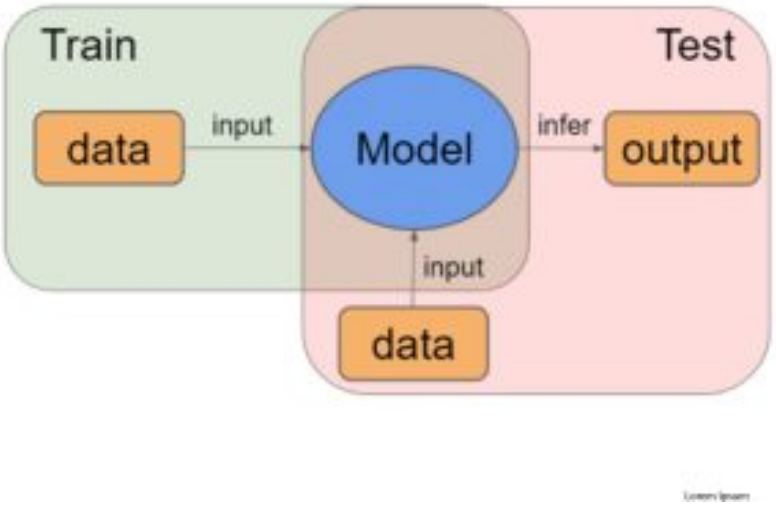


Figure 1.2: Unsupervised learning workflow[1].

1.2.4 How deep learning differs from traditional machine learning

Although machine learning has already had a major impact in many fields, deep learning goes even further by automating tasks that usually require human expertise. Deep learning is a specialized branch of machine learning that relies on neural networks with three or more layers.

These multi-layered neural networks are designed to mimic, in a simplified way, how the human brain processes information. By using large amounts of data, deep learning models are able to automatically learn complex patterns and representations without the need for manual feature extraction. While they do not truly replicate the intelligence of the human brain, they are highly effective at learning from vast datasets and solving complex problems such as image analysis, speech recognition, and medical diagnosis[21].

1.3 Deep learning

1.3.1 Introduction

Deep learning has transformed image recognition by removing the need for manual feature engineering. In traditional approaches, experts had to carefully design and extract features from images before classification. With deep learning, this process is automated.

Convolutional Neural Networks (CNNs), inspired by the human visual cortex, introduced a structure that allows models to learn features directly from raw images. In the early layers, the network detects low-level features such as edges and textures. In the intermediate layers, it identifies mid-level features like shapes or object parts. Finally, the deeper layers capture high-level features that represent more complex concepts, such as complete objects or specific patterns.

This hierarchical learning process enables CNNs to generalize well across different datasets and environments. Unlike traditional machine learning methods, deep learning models perform better as the size of the dataset increases. The more labeled examples they are trained on, the more accurate and robust their predictions become [29].

A clear example of the application of deep learning is eye disease detection using retinal fundus images. By training Convolutional Neural Networks (CNNs) on large annotated datasets, such as ODIR-5K, which include images of both healthy and diseased eyes, the models learn to identify subtle visual patterns. These patterns may include microaneurysms, hemorrhages, and changes in the optic disc that are associated with diseases such as diabetic retinopathy or glaucoma.

Through this learning process, the network becomes capable of distinguishing between normal and pathological cases with high accuracy. This makes it possible to develop automated diagnostic systems that can support early screening in clinical environments. Such systems improve accessibility to eye care services, assist ophthalmologists in their decision-making, and help ensure timely intervention to prevent vision loss [30].

1.4 Artificial Neural Networks ANNs

Artificial Neural Networks (ANNs) form the foundation of Deep Learning. They are inspired by the structure of the human brain and consist of interconnected units called neurons, organized into layers. The input layer receives the raw data, while the output layer generates the final prediction or decision. Between these two layers, one or more hidden layers process the information, transform it, and extract meaningful features from the data.

To enable the network to learn, it is trained using algorithms such as backpropagation. This technique updates the weights and biases of the connections between neurons in order to minimize the difference between the predicted output and the true output. Once the training process is complete, ANNs can effectively perform a wide range of tasks, including regression, classification, and pattern recognition.[31].

1.5 Deep Neural Networks DNNs

Deep Neural Networks (DNNs) are a more advanced form of Artificial Neural Networks that contain multiple hidden layers rather than just one or two. These extra layers allow the model to learn in a hierarchical manner: the first layers capture simple features, while the deeper layers combine them to detect more complex patterns. The depth of the network, defined by the number of hidden layers, enables DNNs to model intricate relationships within data. For this reason, they are highly effective in tasks such as image recognition, speech recognition, and natural language processing.

In the context of our eye disease detection project, DNNs play an important role in analyzing retinal images. They can automatically extract meaningful features from medical

images and identify subtle abnormalities that may indicate different eye diseases. By learning complex visual patterns, DNNs help improve classification accuracy and support early and reliable diagnosis[31].

1.6 Convolutional neural network (CNN)

1.6.1 Introduction to CNN

Convolutional Neural Networks (CNNs) are a special type of Artificial Neural Network designed mainly for working with images. They were developed to handle image data more efficiently and to reduce the complexity involved in recognizing visual patterns. Today, CNNs are one of the most widely used techniques in machine learning, especially in image processing tasks such as object detection and image segmentation.

CNNs operate by applying convolutional layers that automatically extract important features from images, such as edges, shapes, and textures. These layers are usually followed by pooling layers, which reduce the size of the data while keeping the most relevant information. After several convolution and pooling steps, the extracted features are passed to fully connected layers that perform the final classification. Thanks to this structure, CNNs are able to learn hierarchical representations of images, making them highly effective for tasks like image recognition and object detection.

A Convolutional Neural Network (CNN) consists of three main parts: an input layer, several hidden layers, and an output layer. The input layer receives the image data, usually in the form of pixel values. Between the input and output layers, there are multiple hidden layers that perform different operations, such as convolution and pooling, to extract and refine important features from the image. Finally, the output layer produces the final result.

In the context of our eye disease detection project, these algorithms are mainly used to analyze retinal images and automatically identify different types of eye diseases. They help classify medical images, detect abnormalities, and distinguish between healthy and diseased eyes. This makes them valuable tools for supporting early diagnosis and assisting ophthalmologists in making accurate and efficient clinical decisions[32].

1.6.2 VGG

The VGG network was introduced in the influential paper ”Very Deep Convolutional Networks for Large-Scale Image Recognition” by Simonyan and Zisserman. The main motivation behind its development was to better understand how increasing the depth of a neural network affects its performance in image recognition tasks[33].

VGG is characterized by its simple and uniform architecture. Instead of using large convolutional filters, it relies on small 3×3 convolutional filters stacked on top of each other. By significantly increasing the number of layers while keeping the filter size small, VGG is able to capture complex and detailed features from images. This deeper structure allows the net-

work to learn more abstract representations at higher levels, which improves its performance in image classification and other computer vision applications[30].

1.6.3 ResNet

ResNet (Residual Network) addresses the vanishing gradient problem in deep neural networks. This architecture uses skip connections, which allow the model to learn residual mappings rather than trying to learn complex transformations directly. As a result, deep networks become much easier to train and more stable.

One of the main advantages of ResNet is that it reduces the vanishing gradient problem, which often affects very deep models. The skip connections enable information and gradients to flow directly across layers, improving learning efficiency. Thanks to this design, ResNet makes it possible to build extremely deep networks with hundreds or even thousands of layers. Today, it is widely applied in computer vision tasks such as image classification, object detection, and medical image analysis, including eye disease detection [2].

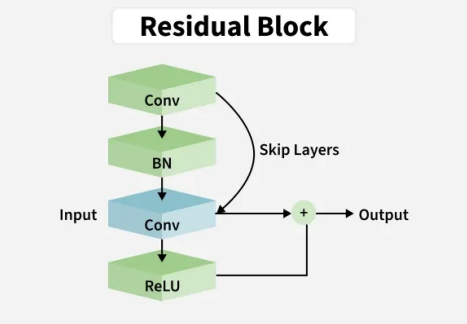


Figure 1.3: Residual Block ([2]).

A residual block allows the network to bypass one or more layers by directly adding the original input to the output of those layers. Instead of learning a completely new transformation, the network only learns the difference (or residual) between the input and the desired output. This simple idea makes very deep networks easier to train, improves gradient flow, and helps prevent performance degradation as the network depth increases[2].

1.6.4 EfficientNet

EfficientNet is a family of Convolutional Neural Networks (CNNs) designed around the principle of compound scaling, which aims to improve model performance while maintaining computational efficiency. Unlike traditional approaches that increase only the network depth, width, or input image resolution independently, EfficientNet scales these three dimensions in a balanced and coordinated manner. Width determines the number of channels in each layer, depth corresponds to the number of layers in the network, and resolution refers to the size of the input images. By jointly optimizing these factors, EfficientNet achieves high classification accuracy with fewer parameters and lower computational requirements.

Due to its efficiency and strong feature extraction capabilities, EfficientNet has become widely used in medical image analysis, particularly for retinal disease detection. Variants such as EfficientNet-B0, B1, B2, and B3 are capable of analyzing fundus images and identifying pathological patterns associated with different ocular diseases. Their ability to deliver high performance while maintaining a relatively low computational cost makes them well suited for automated ophthalmic screening systems t[34].

Features	Top-1 Acc	Top-5 Acc	Params	FLOPs
EfficientNet-B0	77.1%	93.3%	5.3M	0.39B
EfficientNet-B1	79.1%	94.4%	7.8M	0.70B
EfficientNet-B2	80.1%	94.9%	9.2M	1.0B
EfficientNet-B3	81.6%	95.7%	12M	1.8B
EfficientNet-B4	82.9%	96.4%	19M	4.2B
EfficientNet-B5	83.6%	96.7%	30M	9.9B
EfficientNet-B6	84.0%	96.8%	43M	19B
EfficientNet-B7	84.3%	97.0%	66M	37B

Figure 1.4: The EfficientNet Family([3]).

1.6.5 ImageNet

ImageNet is a large-scale dataset of annotated images created to support research in visual object recognition. It contains more than 14 million images, covering thousands of object categories. Each image is labeled according to WordNet synsets, which organize words into groups of related concepts. This structured annotation makes the dataset highly valuable for training and evaluating computer vision models.

ImageNet has played a major role in the development of deep learning, especially through the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Many state-of-the-art Convolutional Neural Networks were first trained and tested on this dataset. Because of its size and diversity, ImageNet remains one of the most important resources for building and benchmarking deep learning models in image classification and other computer vision tasks[35].

1.6.6 MobileNet

MobileNet is a family of efficient Convolutional Neural Network (CNN) architectures specifically designed for mobile and embedded vision applications. These models use a technique called depthwise separable convolutions, which significantly reduces the number of parameters and computational cost compared to standard convolution operations. As a result, Mo-

bileNet models are lightweight, faster, and well suited for devices with limited processing power and memory[36].

Because of their efficiency, MobileNet architectures are ideal for real-time applications in resource-constrained environments. In our eye disease detection project, such models can be deployed on portable medical devices or edge systems to perform fast and accurate retinal image classification. This makes automated screening more accessible and practical, especially in settings where high-end computational resources are not available[30].

1.6.7 DenseNet

Densely Connected Convolutional Networks (DenseNets) are built on a unique and efficient connectivity pattern. Unlike some neural network architectures that combine features using summation, DenseNets use concatenation to merge feature maps. This means that instead of adding outputs together, they stack them, preserving more information from previous layers.

A key characteristic of DenseNet is that each layer receives as input the feature maps from all the preceding layers within the same dense block. In turn, the output of that layer is passed to every subsequent layer in the block. This dense connectivity encourages feature reuse, improves information flow throughout the network, and helps reduce the vanishing gradient problem. As a result, DenseNets can achieve high performance while using fewer parameters compared to traditional deep networks, making them well suited for tasks such as medical image classification, including eye disease detection[37].

1.6.8 AlexNet

AlexNet is considered one of the first breakthrough Convolutional Neural Network (CNN) architectures in modern deep learning. It gained worldwide attention after winning the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 by a significant margin, demonstrating the power of deep learning for image classification.

One of the main contributions of AlexNet was the introduction and popularization of several important techniques. It used the ReLU (Rectified Linear Unit) activation function, which allowed faster and more effective training compared to traditional activation functions. It also applied dropout regularization to reduce overfitting and improve generalization. In addition, data augmentation techniques were used to artificially increase the size of the training dataset, enhancing the model's robustness [38].

1.6.9 CNN Object Detection Algorithms

CNN-based object detection algorithms are designed to both locate and identify objects within an image. In addition to classifying each object, they provide the coordinates of bounding boxes that define the object's position. These capabilities make them essential for

applications that require precise visual analysis, such as automated crop monitoring, weed detection, and counting fruits in agricultural settings.

1.6.9.1 YOLO

YOLO (You Only Look Once) is an object detection algorithm that identifies objects and their locations in an image using a single pass. Unlike traditional methods that treat detection as a classification problem, YOLO approaches it as a regression task, predicting both bounding boxes and class probabilities directly.

The input image is divided into a grid, and a Convolutional Neural Network (CNN) predicts bounding boxes and class probabilities for each grid cell. One key advantage of YOLO is that it performs detection in a single network pass, eliminating the need for separate region proposal networks. This design makes YOLO highly efficient, offering faster processing and strong overall performance for real-time object detection tasks[39].

1.6.9.2 RetinaNet

RetinaNet is a one-stage object detection model designed to improve the accuracy limitations often found in single-stage detectors. It incorporates important components such as Focal Loss and a Feature Pyramid Network (FPN), which help the model handle class imbalance and detect objects at different scales more effectively.

Object detection is a key task in computer vision that involves both locating objects in an image and classifying them. In general, object detection methods are divided into two categories: one-stage detectors (such as YOLO and SSD), which predict object classes and bounding boxes directly from the image, and two-stage detectors (such as R-CNN), which first generate region proposals and then classify them. RetinaNet combines the speed advantage of one-stage methods with improved accuracy, making it a powerful model for many object detection applications[40].

1.7 Recurrent Neural Networks (RNNs)

A Recurrent Neural Network (RNN) is a type of deep neural network designed to work with sequential or time-series data. It processes information step by step, using previous inputs to influence the current prediction. This ability allows RNNs to capture patterns and relationships in sequences of data.

RNNs are commonly used in tasks where the order of data is important. For example, they can predict future values in time-series data such as weather or flood levels based on past observations. They are also widely applied in fields like natural language processing, language translation, sentiment analysis, speech recognition, and image captioning[41].

1.7.1 Long short-term memory (LSTM)

Long Short-Term Memory (LSTM) is a type of RNN designed to solve the vanishing gradient problem and learn long-term dependencies. It introduces memory cells and gates (input, forget, output) that control the flow of information. LSTMs are widely used in tasks such as time series prediction, speech recognition, and natural language processing[41].

1.7.2 Gated recurrent units (GRUs)

A Gated Recurrent Unit (GRU) is a type of recurrent neural network designed to solve the short-term memory problem found in traditional RNNs. Like Long Short-Term Memory (LSTM) networks, GRUs help the model retain important information over longer sequences.

However, GRUs use a simpler structure. Instead of a separate cell state, they rely only on hidden states and use two gates: the reset gate and the update gate. These gates control how much past information should be kept or forgotten during the learning process.

Due to this simpler architecture, GRUs require fewer parameters and are more computationally efficient than LSTMs. As a result, they train faster and are often suitable for real-time or resource-constrained applications[41].

1.8 Conclusion

To conclude this chapter, we presented an overview of the main concepts related to Deep Learning. We first highlighted the differences between traditional machine learning approaches and deep learning architectures. Then, we explored the fundamentals of Convolutional Neural Networks (CNNs), which represent the core technology used in our project, and reviewed several well-known models that support transfer learning. In addition, we briefly discussed Recurrent Neural Networks (RNNs) and their advanced units such as LSTM, which are designed to handle sequential data.

Based on this theoretical foundation, we can now move toward the medical aspect of the study. In the next chapter (Chapter 2), we will focus on the medical context of eye diseases and the methods used for their diagnosis. After that, we will present a state-of-the-art review with a comparative analysis of existing approaches. This will help us evaluate previous work and clearly identify the added value of the model proposed in this research.

Chapitre II

State of art on retinal disease

2.1 Introduction

Age-related Macular Degeneration (AMD), glaucoma, and diabetic retinopathy are among the leading causes of irreversible vision loss and blindness worldwide. These ophthalmic diseases represent a major public health concern due to their high prevalence and their progressive impact on visual function.

In modern ophthalmology, advances in medical imaging technologies have significantly improved both diagnosis and patient management. In particular, imaging modalities enable clinicians to assess disease progression and evaluate the therapeutic effectiveness of treatments such as Anti-VEGF injections, which are widely used in the management of retinal disorders. As a result, ophthalmologists increasingly depend on imaging techniques like fluorescein angiography and Optical Coherence Tomography (OCT) for accurate diagnosis, continuous monitoring, and informed treatment decision-making.

This chapter first introduces the anatomy of the eye, with a particular emphasis on the retina and its key structural components. It then presents the main retinal pathologies relevant to this study. Finally, it defines and discusses the principles of fundus photography, angiography, and OCT imaging.

Although these imaging modalities provide highly detailed and clinically valuable information, their manual interpretation often involving large volumes of scans remains ,time consuming and subject to inter observer variability. This limitation highlights the growing importance of Deep Learning approaches, which offer promising solutions for automating disease detection and supporting clinical decision making. This aspect will be further explored in the subsequent chapters.

2.2 Anatomy of the eye

Vision is one of the five human senses and is considered the most complex. It represents the main source of information for the brain. The eye is the principal organ of the visual system.

It captures light information and converts it into electrical signals, which are then trans-

mitted through the optic nerve to the brain. A good understanding of the anatomy of the human eye is essential for better comprehending its functioning as well as its potential disorders[4].

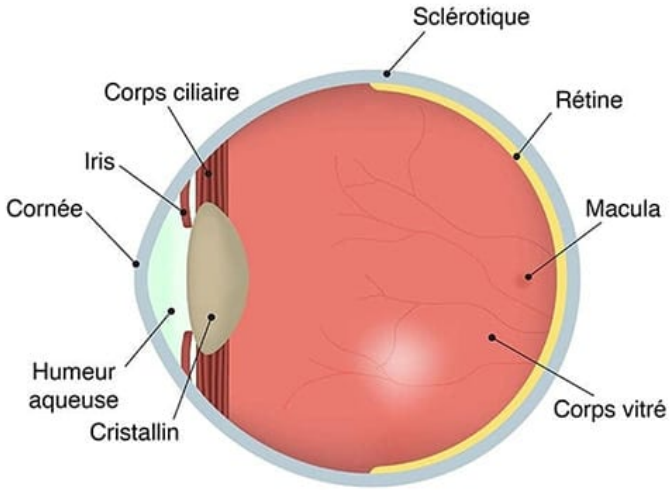


Figure 2.1: Simplified diagram of the human eye.[4]

The eyeball is a spherical organ with an average length of approximately 22 mm. Its structure is primarily composed of three distinct layers, or tunics:

The Outer Tunic (Fibrous Layer): This is the eye's protective shell. It consists of two main parts: The Cornea, The Sclera.

The Middle Tunic (The Uvea): This vascular layer is responsible for nourishment and light control. It includes: The Iris, The Choroid.

The Inner Tunic (Sensory Layer): The Retina.

2.2.1 The Outer Tunic

The Cornea

Anatomically, the cornea is a transparent, non-vascularized anterior tissue with a mean thickness of 530 microns, composed primarily of water and collagen. It is stratified into distinct layers principally the outermost epithelium and the thick underlying stroma, which is the primary target for laser refractive surgery, separated by Bowman's membrane. Physiologically, the cornea provides two-thirds of the eye's total refractive power, supplemented by

the internal crystalline lens. Consequently, structural variations in its spherical geometry directly induce specific refractive errors: excessive curvature causes myopia, a flattened profile results in hypermetropia, and an asymmetric toric shape generates astigmatism[4].

The Sclera

As an extension of the cornea, the sclera constitutes what we commonly call the ”white of the eye,” covering 4/5 of the eyeball’s outer surface. It is a vascularized, fibrous, and durable acellular membrane that serves as the primary attachment point for the extraocular muscles (the muscles that move the eye). Its physiological function is primarily structural. As the most resilient layer of the eye, it is responsible for maintaining the eye’s shape, internal pressure, and overall physical integrity[4].

The Conjunctiva

The conjunctiva is the transparent membrane covering the outer surface of the eyeball. It coats the sclera (known as the bulbar conjunctiva) as well as the inner lining of the eyelids (known as the tarsal conjunctiva). The physiology of the conjunctiva serves a protective and nourishing (trophic) function, as it produces a specialized mucus rich in antibodies to defend the eye[4].

2.2.2 The Middle Tunic

The Iris

The iris, or anterior uvea, is a contractile, circular, and pigmented membrane. It is the colored part of the eye, featuring a central perforated opening that defines the pupil. The physiology of the iris is characterized by variations in pupillary diameter, which dilates or contracts depending on the intensity of light[4].

The Choroid

The choroid is the intermediate layer of the eyeball, situated between the peripheral sclera and the internal retina. It is highly vascularized and rich in pigmented cells (melanin); its dark brown color provides opacity, ensuring that light rays can only enter through the pupil. The physiology of the choroid highlights its essential nourishing function, extending from the optic nerve head (papilla) to the ciliary bodies[4].

The Ciliary Body

The ciliary body is a structure composed of smooth muscles that allow for the modification of the lens curvature. This process, known as accommodation, is essential for achieving sharp, clear vision at various distances. The physiology of the ciliary body is centered on

the production of aqueous humor, the clear fluid that fills the anterior (front) chamber of the eye[4].

The Lens (Crystalline Lens)

The lens is a biconvex structure that provides 1/3 of the eye's refractive power. Initially transparent and avascular (lacking blood vessels), its clouding over time leads to a common condition known as cataracts. The physiology of the lens is characterized by its ability to contract under the influence of the ciliary muscles, playing a major role in accommodation. The gradual loss of this elasticity over the years results in presbyopia (age-related difficulty focusing on nearby objects)[4].

2.2.3 The Inner Tunic

The Vitreous (Vitreous Humor)

The vitreous is a transparent gel that fills 4/5 of the ocular volume behind the lens, maintaining the overall shape and volume of the eye. The physiology of the vitreous shows that it plays a dual role: it is involved in nutrient exchange and acts as a shock absorber for the posterior (back) part of the eyeball, thereby protecting the retina[4].

The Retina

The retina is a complex tissue lining the inner surface of the choroid. The anatomy of the retina distinguishes three different zones: the macula at the center, the fovea in the middle of the macula, and the optic disc (papilla), from which the optic nerve originates. The retina consists of highly specialized tissues, featuring both a neurosensory component (cones and rods) and a pigmented epithelium. The physiology of the retina can be summarized into two major functions:

1. A primary sensory function that enables vision via phototransduction: the cones (responsible for color and central vision) and rods (responsible for night and peripheral vision) convert light signals into electrical signals, which are then transmitted to the optic nerve to reach the brain.

2. A regulatory function of the pigmented epithelium, which acts as a protective screen.

2.3 Retinal pathologies

Retinal pathologies are primarily, though not exclusively, linked to aging. While current medical and surgical therapeutic advances allow for an improved prognosis, they are not always sufficient to halt secondary visual impairment[42].

2.3.1 Diabetic Retinopathy

Diabetic retinopathy is considered the leading cause of acquired blindness among individuals under the age of 65 in Algeria. It affects approximately 61.3% of diabetic patients at CHU Mustapha in Algiers (n = 362, mean HbA1c = 8.2%). In Khenchela, its prevalence has been reported at 21.2%, with 93% of cases occurring in patients with type 2 diabetes. These rates, which are comparable to those observed in other developing countries, highlight the urgent need for automated screening systems[43][44][45][46].

- **Causes and Risks:** Diabetic retinopathy is linked to hyperglycemia inherent in diabetes. The risk of development or worsening decreases as diabetes is better balanced.
- **Co-factors:** Arterial hypertension, common in diabetic patients, significantly contributes to the occurrence and aggravation of the condition.
- **Clinical Consultation:** Patients should provide recent blood tests, prescriptions, and correspondence from their diabetologist to assist in the diagnostic process [42].
- **Pathology:**
  - Characterized by the progressive rarefaction (occlusion) of retinal capillaries.
  - Leads to oxygenation defects in retinal cells and secondary vascular abnormalities.
  - Micro-aneurysms can rupture, causing hemorrhages and abnormal vessel growth.
  - Severe cases may result in intraocular bleeding or retinal detachment.

2.3.2 Screening and diagnosis of diabetic retinopathy

Screening for diabetic retinopathy is essential to prevent vision loss in people with diabetes. The diagnosis of the disease is based on a complete eye examination after dilation of the pupil. It can include additional tests such as fluorescein angiography, optical coherence tomography. In addition, blood sugar and blood pressure can also be controlled. Some screening modes may rely on the cooperation between an orthoptist trained in performing retinography and a reader doctor (ophthalmologist), allowing delayed reading of the obtained images. The use of non-invasive imaging techniques and the development of artificial intelligence are also promising avenues for more accurate screening for diabetic retinopathy. Annual ophthalmologic follow-up is recommended for people affected by diabetes. [47]

2.4 Diagnosis of Diabetic Retinopathy

The diagnosis of diabetic retinopathy (DR) is primarily based on a biomicroscopic fundus examination performed after pupillary dilation, often supplemented by retinal fundus pho-

tography. This examination allows the identification of the main clinical signs of diabetic retinopathy.

Retinal microaneurysms are considered the earliest ophthalmoscopic signs of diabetic retinopathy. They appear as small red punctiform lesions, mainly located in the posterior pole of the retina. Although some microaneurysms may thrombose and disappear spontaneously, an increase in their number is generally considered an indicator of disease progression[48]. -Punctate retinal hemorrhages may also be observed and are often associated with microaneurysms. Other retinal findings may indicate severe retinal ischemia, particularly in advanced stages of diabetic retinopathy:

- **Pre-retinal and pre-papillary neovascularization** are characteristic features of proliferative diabetic retinopathy. These abnormal new vessels usually develop at the border of ischemic retinal areas or on the optic disc when the non-perfused retinal surface is extensive.

Additional signs may reflect increased capillary permeability at the macular level:

- **Cystoid macular edema (CME)** is characterized by microcystic thickening of the macular retina.
- **Hard exudates** are lipid-rich deposits that accumulate within the edematous retina. They often appear as circinate deposits surrounding areas of microvascular abnormalities from which they originate[48].

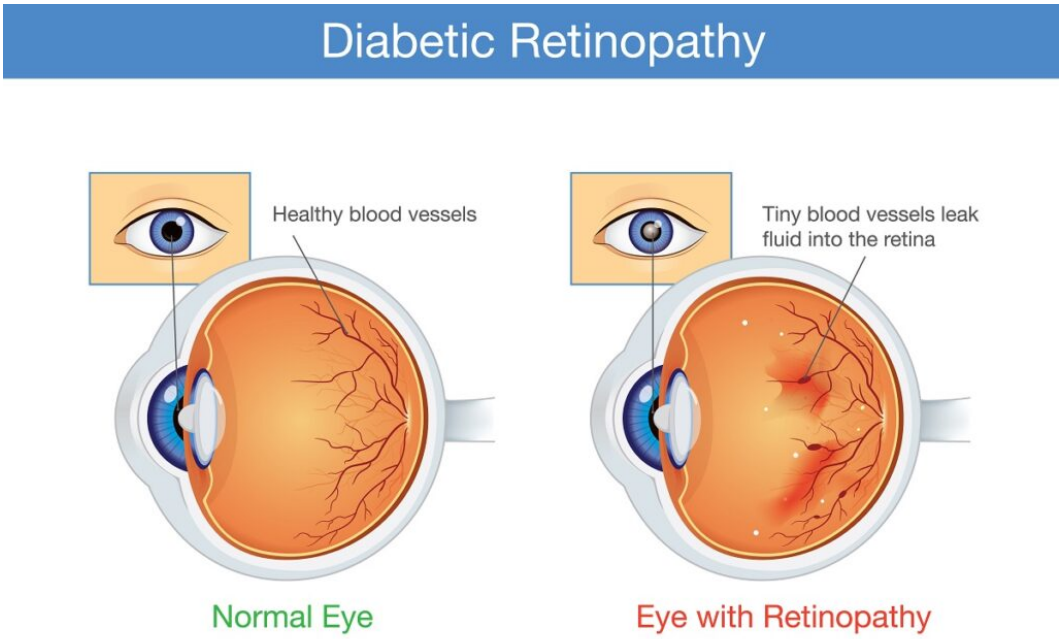


Figure 2.2: Diabetic retinopathy diagram.[5]

2.5 Age-related macular degeneration (AMD)

2.5.1 Definition

Age-related macular degeneration (AMD) is an eye disease whose development is mainly associated with aging, as well as genetic predisposition. It affects the eye, and more specifically the macula, which is the central part of the retina located at the back of the eye and responsible for central vision [49]. When age-related macular degeneration (AMD) becomes degenerative, it occurs in two main forms: atrophic (dry) and exudative (wet). The atrophic or dry form is the most common, accounting for approximately **80%** of cases. It is characterized by a shrinking of the macula, a reduction in retinal pigment epithelium cells, and a loss of photoreceptor cells. Currently, there is no treatment available to slow the progression of dry AMD[50]. Exudative, neovascular, or wet AMD is characterized by the formation of new blood vessels beneath the retina. These abnormal vessels can leak fluids such as serum or blood, leading to retinal hemorrhages[50]. The development of AMD is also influenced by genetic factors, smoking, and exposure to ultraviolet (UV) radiation. Smoking and UV exposure can further accelerate disease progression. It is important to note that psychological factors may also be involved, as individuals with AMD often experience increased levels of stress and anxiety[50].

2.5.2 AMD Diagnosis

AMD is suspected in the presence of a progressive decrease in visual acuity for both near and distant vision, corresponding to the slow worsening of lesions (drusen, atrophic form). A sudden drop in visual acuity, associated with metamorphopsia (a sensation of object deformation, often described by patients as wavy lines instead of straight ones), is most often linked to the development of choroidal neovascularization, which causes exudation and fluid accumulation in the macula. A central scotoma may also appear, corresponding to advanced stages of both atrophic and exudative forms[51].

2.6 Cataract

Algeria is considered one of the emerging countries where visual impairment is caused by cataracts in nearly **50%** of cases. It is estimated that between 100,000 and 200,000 patients suffering from cataracts require surgery each year.

Cataract is defined as the partial or complete opacification of the crystalline lens. It is a very common eye disease, and cataract surgery is considered the most frequently performed surgical procedure worldwide, across all medical specialties.

The increasing prevalence of this condition is mainly associated with aging, which explains its higher occurrence as the population continues to grow older [52].

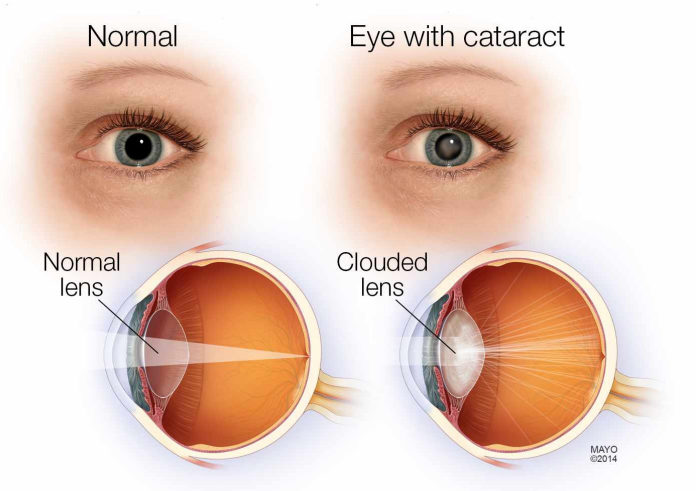


Figure 2.3: Normal eye vs eye with cataract.[6]

2.6.1 Types of cataract

There are several types of cataracts that affect the lens of the eye:
Nuclear cataracts: develop in the center of the lens and are more common with age.
Cortical cataracts: start at the edge of the lens and gradually move inward.
Posterior subcapsular cataracts: form at the back of the lens and are more frequent in younger adults.
Congenital cataracts: present at birth and require early treatment to prevent vision problems[53].

Symptom	Cataract Type	Mechanism
Progressive blurred vision	Nuclear	Visual axis opacity
Glare / halos	Posterior subcapsular (PSC)	Backlight scattering
Index myopia (”second sight”)	Nuclear	Decreased refractive index
Faded colors	All types	Spectral filtering
Monocular double vision	Cortical	Irregular astigmatism

Table 2.1: Symptoms, cataract types, and associated mechanisms[10].

2.7 Myopia

Myopia (nearsightedness) is a refractive error in which parallel light rays focus in front of the retina when the eye is relaxed. It is most often due to an elongated eyeball, but can also result from an overly curved cornea or increased lens power. It affects both children and adults and is usually corrected with glasses or contact lenses.

Myopia can be classified in several ways. By cause, it includes:

- **Axial** (most common, due to longer eyeball),
- **Curvatural** (increased curvature of cornea/lens),

- **Index** (change in lens refractive index),
- **Positional** (forward displacement of the lens).

Clinically, it may appear as simple, degenerative (high and progressive), pseudo (due to excessive accommodation), nocturnal (in low light), or induced (temporary, caused by external factors). By severity, it is classified as low (< -3D), moderate (-3D to -6D), and high (> -6D).

Myopia can also be categorized by age of onset, including congenital, youth-onset, early adult, and late adult forms. Additionally, intraocular pressure (IOP), regulated by the balance between production and drainage of aqueous humor, plays an important role in overall eye physiology, although it is not directly influenced by the vitreous body[54].

2.7.1 Diagnosis and Tests

Myopia is diagnosed through routine eye examinations performed by an eye care specialist. Although it is most commonly detected during childhood, it can also appear in adults, especially due to visual strain or conditions such as diabetes.

2.7.1.1 Adults

For adults, the examination focuses on how the eyes focus light and the level of correction needed.

- Visual acuity is first assessed using an eye chart to measure clarity of vision.
- A retinoscope is used to observe how light reflects from the retina.
- A phoropter may be used to determine the exact refractive error by testing different lenses and identifying the most suitable correction.

2.7.1.2 Children

For children, eye evaluations begin early, ideally before the age of one, and continue regularly as they grow. Pediatricians often perform initial screenings, especially if there is a family history of vision problems. If needed, children are referred to specialists such as optometrists or pediatric ophthalmologists. During the exam, doctors assess eye health and visual response, and for children aged 3 to 5, vision tests may include charts, symbols, or interactive methods like the ”tumbling E” test.

Regular follow-up is essential, as children’s vision can change over time. Screenings are recommended before starting school and periodically afterward. Myopia is often classified as mild, moderate, or high, depending on the degree of refractive error, and it may occur alongside other vision conditions such as astigmatism [55].

2.8 Hypertensive retinopathy

Hypertensive retinopathy is an eye condition caused by long-term high blood pressure, which damages the blood vessels in the retina. Over time, these vessels become narrow, thick, and less flexible, reducing blood flow and affecting vision. In most cases, the disease has no early symptoms, but in advanced stages it can lead to blurred vision or vision loss. The main risk factor is chronic hypertension, along with conditions such as smoking, high cholesterol, diabetes, and cardiovascular disease.

Doctors diagnose it through eye examinations and retinal imaging, where signs like vessel narrowing, hemorrhages, or swelling of the optic nerve may appear. The disease is classified into different stages, ranging from mild vascular changes to severe retinal damage.

Treatment mainly focuses on controlling blood pressure through lifestyle changes (diet, exercise) and medication. Early treatment can reverse or reduce retinal damage, while delayed management may lead to serious complications such as blindness, retinal detachment, or stroke risk.

Overall, hypertensive retinopathy is both a vision-threatening condition and an important warning sign of systemic cardiovascular disease, making early detection and proper management essential[56].

2.8.1 High blood pressure (diagnosis & treatment)

High blood pressure (hypertension) is usually diagnosed by measuring blood pressure levels using a cuff device. A diagnosis is confirmed when readings are consistently high over time. Doctors may also perform additional tests (blood tests, urine tests, ECG) to check for complications or underlying causes.

Treatment mainly focuses on reducing blood pressure to prevent complications such as heart disease, stroke, and eye damage. This is achieved through a combination of:

- **Lifestyle changes**, including a healthy diet (low salt), regular physical activity, weight control, limiting alcohol, and quitting smoking.
- **Medications**, such as diuretics, ACE inhibitors, calcium channel blockers, or beta-blockers, depending on the patient's condition.

Early detection and proper management are essential, as controlling blood pressure can prevent serious complications and protect overall health, including vision[57].

2.9 Glaucoma

Glaucoma is a degenerative disease of the optic nerve that can progressively lead to irreversible vision loss if not diagnosed and treated in time. It is considered one of the leading causes of blindness worldwide and is currently the second most common cause of blindness

globally. Glaucoma includes a wide range of clinical forms. In general, its classification is based on its origin and the configuration of the iridocorneal angle, distinguishing between primary and secondary glaucoma, as well as open-angle and angle-closure glaucoma[58].

2.9.1 Types of glaucoma

There are several forms of glaucoma that can affect both adults and children. However, most cases are generally classified into two main categories: open-angle glaucoma and angle-closure glaucoma[59].

Open-angle glaucoma

Open-angle glaucoma is the most common form of glaucoma. In this type, the drainage channels of the eye gradually become blocked by tiny microscopic deposits over a period of months or years. It is called open-angle because the angle between the iris and the cornea remains open, even though the fluid drainage is reduced. As a result, the aqueous humor is produced normally but drains too slowly, causing a gradual increase in intraocular pressure[59].

Angle-closure glaucoma

Angle-closure glaucoma is less common than open-angle glaucoma. In this type, the drainage channels become blocked because the angle between the iris and the cornea is too narrow or completely closed. This obstruction may occur suddenly, resulting in acute angle-closure glaucoma, or it may develop slowly over time, leading to chronic angle-closure glaucoma. When the blockage is sudden, intraocular pressure rises rapidly and may cause severe symptoms. When it develops gradually, the increase in eye pressure is slower, similar to open-angle glaucoma[59].

2.9.2 What is the cause of glaucoma?

Glaucoma is most commonly caused by an increase in intraocular pressure resulting from impaired drainage of the aqueous humor, which leads to fluid accumulation inside the eye. This elevated pressure can damage the optic nerve by compressing nearby blood vessels and visual nerve fibers. The risk of glaucoma also increases with age and is higher in individuals with myopia, cardiovascular diseases, diabetes, hypothyroidism, or certain genetic and ethnic predispositions[58].

2.9.3 Glaucoma Diagnosis

Glaucoma is often diagnosed late, sometimes several years after the abnormal increase in intraocular pressure begins. In many cases, it is discovered during an eye examination for another condition, such as presbyopia, early cataract, or glasses renewal[60].

Measurement of Intraocular Pressure

Intraocular pressure is measured using tonometry, either with a tonometer placed on the cornea after anesthetic eye drops or with a puff of air. The result may vary depending on the thickness of the cornea and the time of day[60].

Measurement of Corneal Thickness

Corneal thickness differs from one person to another and is important for interpreting intraocular pressure correctly. A thin cornea may lead to underestimation of eye pressure, while a thick cornea may make normal pressure appear higher[60].

Fundus Examination

A fundus examination allows the ophthalmologist to observe the retina and the optic nerve head in order to detect signs of optic nerve damage. It can also assess the retinal blood vessels and monitor disease progression over time[60].

Visual Field Test

The visual field test is used to detect early loss of peripheral vision caused by optic nerve damage. It is repeated regularly to evaluate treatment effectiveness and to check whether glaucoma is progressing[60].

Gonioscopy

Gonioscopy is performed to examine the angle between the iris and the cornea, where the drainage system of the eye is located. This test helps identify whether the angle is open or narrow and assesses the risk of angle-closure glaucoma[60].

2.10 Medical imaging analysis techniques

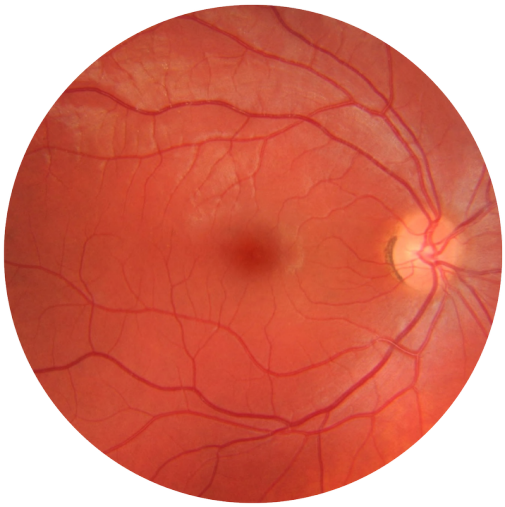
2.10.1 Eye fundus examination

Fundus is a simple examination that allows the ophthalmologist to observe the retina with its blood vessels, the head of the optic nerve (papilla) and the macula. There are several ways to observe the patient's fundus. Among all the existing devices, we will mention two important ones that allow doctors to observe the retina and at the same time offer photographs of the fundus.[48].

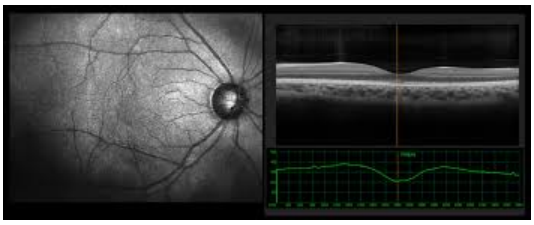
2.10.2 Retinal imaging techniques

Retinography

Retinography is an examination using a medical device (fundus camera) to photograph the retina, or back of the eye, through the pupil. The images allow the ophthalmologist to study the different layers of the retina and identify possible retinal diseases. This examination is non-invasive, meaning that no object or medical device comes into direct contact with the eye.[7].



Fundus photography



OCT scan

Figure 2.4: Medical imaging analysis techniques.[7]

Fluorescence angiography

Fluorescein angiography is an ophthalmic imaging technique used to visualize the blood vessels of the retina and choroid. It involves the injection of a fluorescent dye, called fluorescein, into the bloodstream, usually through a vein in the arm. As the dye circulates through the ocular vessels, a specialized camera with blue light captures sequential images, allowing the ophthalmologist to clearly observe the retinal circulation and identify vascular abnormalities. This examination provides important information about blood flow in the eye and is useful for detecting abnormalities such as leakage, vessel obstruction, or abnormal neovascularization. These findings are particularly relevant in the diagnosis and monitoring of several retinal diseases[61].

Fundus Photography Techniques

Definition

The fundus of the eye refers to the inner posterior surface visible through the pupil. It includes important structures such as the retina, macula, optic disc, and retinal blood vessels. In ophthalmology, examination of the fundus is essential for evaluating the health of the retina and optic nerve, as well as for detecting various ocular diseases [62][8].

Figure 2.5: fundus machine[8].

Principle of fundus examination

Fundus examination, also known as ophthalmoscopy, is a procedure used to examine the back part of the eye. During the examination, light is directed through the pupil using an ophthalmoscope or a fundus camera to visualize structures such as the retina, macula, optic nerve, and retinal blood vessels. In many cases, the pupil is dilated beforehand to provide a clearer and wider view of these ocular structures[8][63].

2.10.3 Fundus angiography

Fundus fluorescein angiography (FFA) is an invasive ophthalmic imaging technique used to examine the retinal and choroidal blood circulation in detail. The procedure involves

the intravenous injection of sodium fluorescein dye, followed by rapid sequential imaging of the fundus as the dye flows through the ocular vessels. Since fluorescein produces a yellow-green fluorescence under blue light, the fundus camera can clearly visualize both normal vascular perfusion and pathological abnormalities, including leakage, ischemia, and neovascularization[64].

This examination is particularly important in ophthalmology because it provides dynamic visualization of retinal and choroidal blood circulation, which cannot be fully assessed using standard fundus photography alone. Fundus fluorescein angiography (FFA) enables clinicians to identify abnormalities such as vascular leakage, capillary non-perfusion, delayed vascular filling, ischemia, macular edema, and neovascularization. It is widely used in the diagnosis and monitoring of diseases including diabetic retinopathy, retinal vein and artery occlusions, age-related macular degeneration, and choroidal neovascular membranes. In addition, FFA plays a significant role in guiding therapeutic decisions such as laser photocoagulation and intravitreal injections[65].

Although fundus fluorescein angiography is generally considered safe, it remains an invasive procedure because it requires the intravenous injection of fluorescein dye. Some patients may experience mild temporary side effects, including nausea, blurred vision caused by pupil dilation, or yellow discoloration of the skin and urine. Severe allergic reactions are uncommon but may occur in rare cases. Despite these limitations, FFA remains an essential examination in retinal practice, as it provides precise information about retinal circulation and vascular leakage, supporting accurate diagnosis, treatment planning, and patient follow-up[66].

2.11 Comparative Analysis of State-of-the-Art Approaches

To systematically evaluate existing deep learning approaches for retinal disease detection, we analyzed studies published between 2020 and 2026 based on nine key criteria: (1) targeted disease(s), (2) dataset, (3) model architecture, (4) reported performance, (5) generalization through external validation, (6) multitask capability (classification and localization), (7) multistage pipeline, (8) multimodality (e.g., Fundus + OCT), and (9) publication year. Table 2.2summarizes the main findings. For direct comparison, the last row presents the characteristics of our proposed approach.

Year	Study	Disease(target)	Dataset	Model(architecture)	Results(Performance)	Generalized?	Multitask?	Multistage?
2020	Exudate Detection for Diabetic Retinopathy Using Pretrained CNNs [67]	DR(exudate detection)	e-Ophtha, DIARETDB1	Inception-v3, ResNet-50, VGG-19	Fused model: Acc 98.2%, Sens 97.8%, Spec 98.5% (e-Ophtha);Acc 97.1%, Sens 96.5%, Spec 97.6% (DIARETDB1)	No	No(lesion only)	Yes(ROI → feature → classify)
2021	Detection of Exudates and Microaneurysms in the Retina [68]	DR lesions (EX, MA)	MESSIDOR	LAB model / RGB model	specificity of 0.94 and a sensitivity of 0.97	No	No (lesion focus)	Not applicable(not available)
2022	Screening of Common Retinal Diseases Using Six-Category Models Based on EfficientNet [69]	Multi-disease (Normal, RVO, High Myopia, Glaucoma, DR, MD)	Custom (2,400 train,1,315 test fundus)	EfficientNet-B4	Overall: Acc 95.59%, Kappa 94.61%; Normal: F1 99.83%, AUC <0.95; RVO: F1 93.09%; High Myopia: F1 97.37% ;Glaucoma: F1 94.62%; DR: F1 93.3%; MD: F1 91.51%	Limited (internal test only)	No	No

2023	Multi-Fundus Diseases Classification Using Retinal OCT Images [70]	OCT diseases(CNV /DME/Drusen/ Normal)	OCT2017, OCT-C8	Swin Transformer V2 + PolyLoss	OCT2017: Avg Acc 99.9%; OCT_C8: Avg Acc 99.5% (all AUCs >0.99)	No	No	No
2024	A deep learning framework for the early detection of multi-retinal diseases [71]	Multi-disease(DR, ODC, MH, WNL)	RFMiD, RFMiD 2.0	12/20-layer CNN	20-layer w/ aug: Val Acc 90.34%,Test Acc 89.59%; F1 89.2%; notes overfitting	No	No	No
2024	MultiEYE: Dataset and Benchmark for OCT-Enhanced Retinal Disease Recognition [72]	Multi-disease recognition	MultiEYE (58k fundus + 45k OCT)	Benchmark (various; focuses OCT-distil for fundus cls)	Dataset paper; e.g., OCT-distil boosts fundus Acc by 1-3% on DR/AMD/glaucoma	Not applicable(not available)	No	No
2025	Automated Multi-label Classification of Eleven Retinal Diseases [73]	Multi-label retinal diseases (11 classes)	AIROGS (glaucoma),Unified DR, RFMiD	ConvNeXtV2 + ensemble	AUC 0.9126(AIROGS), macro AUC 0.8800(RFMiD)	yes	No	No
2025	Diabetic retinopathy detection via exudates and hemorrhages [74]	DR detection via EX/HE	Exudate Hemorrhage Healthy retinal images	INRG-HSV / INRG-WS_HSV (thresholding+ region growing)	HE det: Acc 90.27%; EX det: Acc 88.14%; DR det: Acc 89.89%	No	Yes(lesion + DR)	Yes

2026	Diabetic Retinopathy Lesion Segmentation through Attention Mechanisms [75]	DR lesions (MA/HE/EX/SE segmentation)	DDR dataset	DeepLabV3+ + CBAM attentio+	Mean AP: 0.3010 → 0.3326; MA AP: 0.0205 → 0.0763 (+272%)	Not applicable(not available)	No	No

Table 2.2: Comparative state of the art for retinal disease diagnosis.

Discussion

Table 2.2 provides a systematic overview of pivotal research contributions spanning from 2020 to 2026 within the domain of automated retinal pathology identification via deep learning framework architectures.

The synthesized literature demonstrates a definitive methodological evolution, transitioning from baseline, single pathology classification schemes predominantly targeting Diabetic Retinopathy utilizing canonical networks such as Inceptionv3 and ResNet50 toward highly sophisticated, multi-disease and multi-label diagnostic pipelines that leverage contemporary backbones, including EfficientNetB4, Swin TransformerV2, and ConvNeXtV2. However, despite documenting highly optimized performance metrics across benchmark datasets like MESSIDOR, ODIR-5K, and DDR, a critical analysis reveals that the majority of existing frameworks remain constrained by generalization bottlenecks and typically lack integrated multi-stage or multi-task architectural designs.

The most recent state of the art implementations (2025–2026) exhibit continuous refinements in fine-grained lesion segmentation and complex multi-label classification, thereby emphasizing the ongoing academic and clinical imperative to develop more robust, generalizable, and translationally viable computer aided diagnostic systems.

2.12 Comparative Analysis of Multi-label Classification (ODIR-5K)

The evaluation of multi-label classification on the ODIR-5K dataset shows a clear progression from early simple models to more advanced ensemble approaches, as summarized in Table 2.3. presents a comparative evaluation of multi-label classification methods on the ODIR-5K dataset. Early benchmarks, such as Islam et al. ([76]), utilized shallow CNN architectures that struggled with minority classes, resulting in a Macro F1 score of 0.49. While Wang et al. ([77]) achieved a high weighted F1 score using a dual EfficientNet approach, their AUC-ROC remained relatively low at 0.73.

Study	**Year**	**Method**	**Macro F1**	**AUC-ROC**	**Notes**
Islam et al.[76]	2020	Shallow CNN, single model	0.49	––	Early baseline; struggles with minority classes
Wang et al.[77]	2021	Dual EfficientNet (gray + color), feature context	0.88*	0.73	*reported as weighted F1; AUC lower than peers
Proposed (EDD MT)	2026	EfficientNet-B3, weighted BCE, per-class threshold tuning	0.5566	0.805	Single model; test set (639 patients); no metadata; Colab T4 constraint

Table 2.3: Multi-label classification methods on ODIR-5K

2.13 Comparative Analysis of DR Lesion Detection (DDR Dataset)

The advancement of Diabetic Retinopathy (DR) lesion detection is analyzed using the DDR dataset, with performance comparisons based on mean Average Precision (mAP) presented in Table 4.5.1.4. the performance of various Diabetic Retinopathy (DR) lesion detection methods on the DDR dataset at $IoU = 0.50$. The baseline established in the original DDR paper ([78]) reported a mean Average Precision (mAP) below 0.15. Subsequent efforts by Santos et al. ([79]) improved this to 0.263 through the use of YOLOv5 and tiling techniques. While the YOLOv9e architecture by Coello et al. ([80]) achieved the highest overall mAP of 0.359 without preprocessing.

Study	**Year**	**Method**	**mAP@0.5**	**AP_{MA}**	**AP_{HE}**	**AP_{EX}**	**Notes**
Li et al. (DDR)[78]	2019	RetinaNet, SSD, YOLOv3 baselines	<0.15	–	–	–	Original DDR paper; detection treated as baseline only
Santos et al.[79]	2022	YOLOv5 + image processing + tiling	0.263	0.105	–	–	Val set; F1 = 0.35; PMC open access
Santos et al.[79]	2022	YOLOv5 + Adam + tiling	0.263	0.111	–	–	Variant with Adam optimiser
Coello et al. (YOLOv8m)[80]	2024	YOLOv8-mono, tiling, RGB, no preprocessing	0.339	0.182	0.418	0.417	MDPI Vision; no coding preprocessing
Coello et al. (YOLOv9e)[80]	2024	YOLOv9-extended, tiling, RGB	0.359	0.205	0.484	0.200	Best published result on DDR without preprocessing
Proposed (EDD MT)	2026	YOLOv8s + CLAHE + CAM masking + SAHI	0.279	0.278	0.159	0.400	Val set; 3 classes; CAM-guided preprocessing

Table 2.4: DR lesion detection methods on DDR dataset.

2.14 Conclusion

In this chapter, we have explored the biological and clinical foundations necessary to understand eye diseases, specifically focusing on the retina, glaucoma, and diabetic retinopathy,cataract,myopia,hypertension. We have also reviewed the various medical imaging modalities, such as OCT and Fundus photography, which serve as the primary data sources for diagnostic procedures.

The state-of-the-art review has demonstrated that while deep learning models have achieved remarkable success in medical imaging, challenges remain—particularly regarding the need for multi-class detection and the optimization of models for diverse datasets.

While the existing literature provides a strong foundation, the increasing volume of medical data requires more robust, automated systems. This leads us to the following chapter, where we will detail our specific methodological approach and the datasets used to address these clinical challenges.

Chapitre III

Proposed Approach in eye disease diagnosis

3.1 Introduction

The development of an effective Computer-Aided Diagnosis (CAD) system for ophthalmology requires a structured framework capable of managing the complexity and variability of clinical data. While many existing studies focus on single-disease detection, real clinical practice requires a more comprehensive approach that can identify multiple eye conditions and accurately locate lesions to ensure reliable and transparent diagnosis.

This chapter presents the general architecture and specific technical details of our contribution. We propose a unified, multi-task and multi-stage diagnostic pipeline designed to connect global screening with detailed clinical analysis.

3.2 Dataset acquisition and Characteristics

To develop a system capable of both detecting diseases and accurately locating clinical lesions, this study relies on two dedicated ophthalmic datasets. Each dataset is assigned a specific function within the proposed multi-stage framework.

3.2.1 The ODIR-5K Dataset (Stage 1: Classification)

The Ocular Disease Intelligent Recognition (ODIR-5K) dataset was selected as the primary training corpus for Stage 1. It presents a multi-label, multi-class classification challenge that reflects real-world clinical conditions: extreme label noise, multi-disease co-occurrence, non-standardised image acquisition, and severe class imbalance [11]. The dataset was introduced and benchmarked by Li et al. [81].

Attribute	Value
Total patients	˜5,000
Number of diagnostic classes	8
Classification type	Multi-label (a patient may carry multiple diagnosis)
Image format	Raw fundus photos, varied lighting and black borders
Train split	5,113 images
Validation split	640 images
Test split	639 images
Input resolution (after preprocessing)	224×224 pixels

Table 3.1: ODIR-5K dataset summary [11]

3.2.1.1 Diagnostic Classes

Each patient record carries an 8-dimensional binary label vector. The classes and their clinical definitions are:

Code	Disease	Clinical Significance
N	Normal	No pathological findings
D	Diabetes	Diabetic retinopathy signs
G	Glaucoma	Optic disc structural changes
C	Cataract	Lens opacity
A	Age-related Macular Degeneration (AMD)	Drusen, geographic atrophy
H	Hypertension	Arteriovenous nicking, vessel changes
M	Myopia	Myopic degeneration, posterior staphyloma
O	Other	Residual pathologies not fitting above categories

Table 3.2: ODIR-5K diagnostic class definitions [11]

3.2.2 The DDR Dataset (Stage 2: Lesion Detection)

While global classification provides a diagnostic label, clinical validation requires the precise identification of pathological markers. For this purpose, we utilize the DDR (Diabetic Retinopathy: Segmentation and Grading) dataset, a large-scale, multi-center benchmark specifically designed to support the detection of fine-grained retinal lesions [12].

3.2.2.1 Dataset Structure

The raw DDR dataset used in this project contains fundus images split across three partitions. The following table summarises the image and label counts per split:

Split	Images	Label files	Notes
Train	427	383	44 images without labels (background or unannotated)
Validation	143	149	6 extra label files (orphaned annotations)
Test	187	225	38 extra label files

Table 3.3: DDR dataset split statistics [12]

3.3 Technical Analysis and Diagnostic Classes

While global classification provides a diagnostic category, clinical validation requires the precise identification of pathological markers. The DDR (Diabetic Retinopathy: Segmentation and Grading) dataset was selected as the foundation for our Stage 2 localization pipeline due to its high-quality, manually-verified bounding box annotations.

Clinical Origin: The DDR dataset is a large-scale, multi-center benchmark consisting of 13,673 fundus images collected from 147 hospitals. It reflects real-world clinical diversity, featuring images captured with different fundus cameras and varying levels of illumination and noise[12]. As described by Li et al [13].

Target Pathologies: For this research, we focus on the detection of three hallmark lesions of Diabetic Retinopathy:

Class	Total Boxes	Percentage
Microaneurysms (MA)	16,163	20.4%
Hemorrhages (HE)	45,936	57.9%
Hard Exudates (EX)	17,277	21.8%
Soft Exudates (SE)	0	0.0% — ABSENT
TOTAL	**79,376**	**100%**

Table 3.4: DDR class distribution and annotation statistics (adapted from [12] and [13]).

Inherent Technical Challenges: Our analysis of the DDR characteristics identified three primary obstacles that distinguish it from standard object detection datasets:

Challenge	Description
Sub-pixel Scale	Images feature high resolutions (up to 2,592 $\times$ 1,728 pixels). When resized to 640 $\times$ 640, small lesions like microaneurysms can occupy less than 0.005% of the total image area.
Severe Class Imbalance	Hemorrhages dominate at 57.9%, while microaneurysms represent only 20.4%, making it difficult for models to learn the scarcer, critical signs.
Annotation Noise	A systematic review identified a 41.3% duplicate annotation rate, necessitating a rigorous deduplication process to prevent overfitting on redundant signals.

Table 3.5: Challenges in DDR dataset for lesion detection [12]

3.4 Proposed Architecture

Our proposed system is a two-stage conditional pipeline for retinal disease diagnosis. The pipeline begins with a global classification stage (Stage 1) and conditionally activates a detailed lesion detection stage (Stage 2) only when Diabetic Retinopathy is identified. This design mirrors the clinical workflow of screening followed by detailed investigation, ensuring computational efficiency while maintaining diagnostic comprehensiveness.

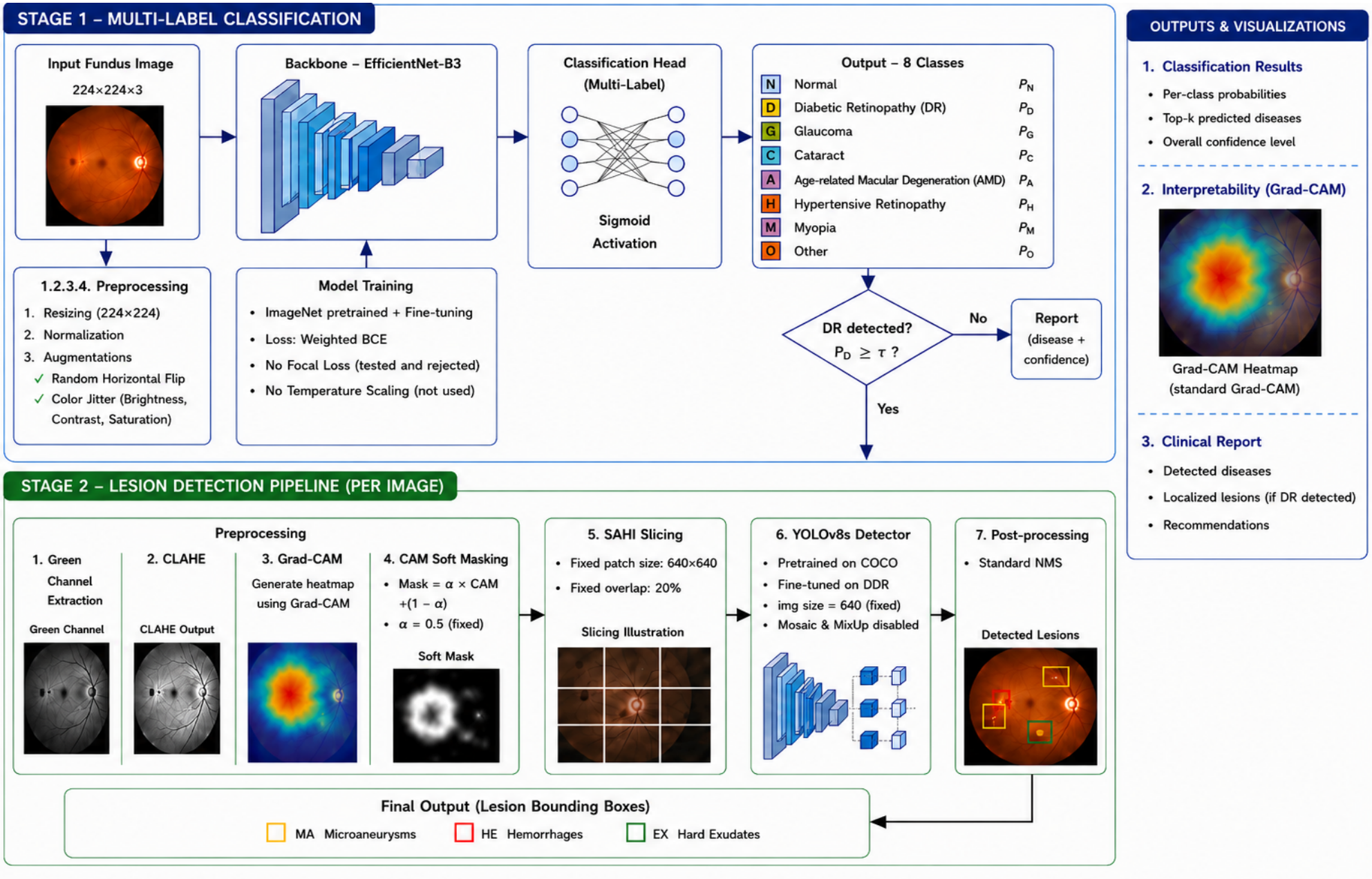


Figure 3.1: Proposed two-stage conditional pipeline for retinal disease diagnosis.

3.4.1 Stage 1: EfficientNet B3 for Image Classification

EfficientNet, introduced by Google Research in 2019, represents a significant advancement in convolutional neural networks by prioritizing computational efficiency alongside high accuracy. The architecture is built upon several core innovations that distinguish it from traditional CNN designs:

Core Innovations and Compound Scaling

The most significant innovation of EfficientNet is the compound scaling method, which scales depth, width, and resolution simultaneously using a fixed set of scaling coefficients. This approach ensures that as the network grows deeper to capture complex features, it also grows wider and uses higher-resolution inputs to maintain a balance between the receptive field and fine-grained detail.

The architecture is fundamentally built upon the Mobile Inverted Bottleneck Convolution (MBConv) block, which utilizes an inverted residual structure to reduce computational operations while maintaining representational capacity. Each MBConv block integrates a Squeeze-and-Excitation (SE) module that employs a channel-wise attention mechanism, allowing the network to prioritize the most informative features. This is a critical capability for medical imaging where pathological signs may be subtle. Furthermore, the model replaces standard ReLU with the Swish activation function, defined as $Swish(x) = x \cdot \sigma(x)$, which enhances gradient flow and leads to superior accuracy[9].

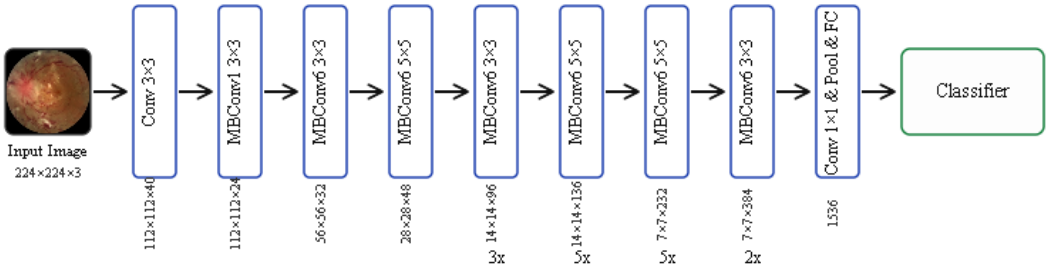


Figure 3.2: The EfficientNet architecture [9].

Model Selection: Why EfficientNet-B3?

The selection of **EfficientNet-B3** as the backbone for Stage 1 followed a systematic experimental comparison on the ODIR-5K dataset. Key factors in this selection include:

- **Comparative Evaluation:** The B3 variant was tested against ResNet-50, EfficientNet-B0, and EfficientNet-B4 under identical experimental conditions, including dataset splits and preprocessing. The results of this comparative evaluation are documented in Chapter 4.

3.4.2 Stage 2:Lesions Detection

To mitigate the challenge of identifying micro-scale pathologies, a specialized preprocessing pipeline was designed to precede the lesion detection stage. This workflow consists of three sequential operations:

3.4.3 Preprocessing Pipeline

The raw DDR dataset, in its unprocessed form, is unsuitable for direct training of a standard object detector. The preprocessing pipeline developed in this work addresses the identified challenges through three sequential stages,The raw DDR dataset, in its unprocessed form, is unsuitable for direct training of a standard object detector. The preprocessing pipeline developed in this work addresses the identified challenges through three sequential stages.

3.4.3.1 Step 1: Green Channel CLAHE Enhancement

In fundus photography, the red channel dominates due to the vascular nature of the retina, which can obscure fine lesion details. The blue channel typically contains elevated noise. The green channel, by contrast, provides the highest contrast for retinal vessels, microaneurysms, and hemorrhages. This is a well-established finding in retinal image analysis, supported by published literature.
CLAHE (Contrast Limited Adaptive Histogram Equalization) was applied exclusively to the green channel to enhance local contrast without amplifying noise. The enhanced green channel was subsequently merged back into the RGB image, producing a modified image with improved lesion visibility while retaining the full colour information[82].

Validation

Five fundus images were randomly selected from the DDR training set to validate the effect of green channel CLAHE. The following figures compare the original images, the isolated green channels, CLAHE-enhanced green channels, and the final merged RGB images

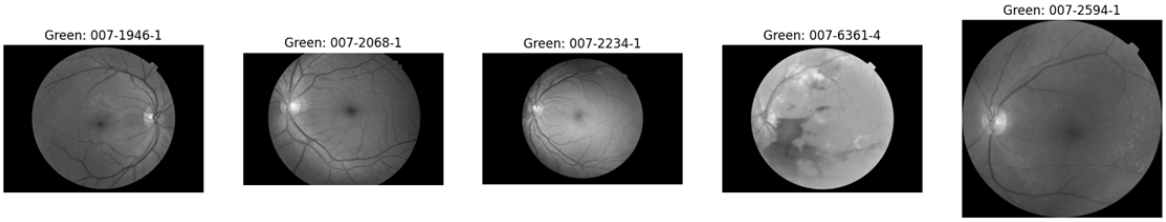


Figure 3.3: Isolated green channels from 5 randomly selected DDR images. Vessel and lesion structures are clearly visible.

Figure 3.4: CLAHE-enhanced green channels.

Discussion

This figures3.4 and 3.3 illustrates the effect of the first step of our preprocessing pipeline. By extracting the green channel and applying Contrast Limited Adaptive Histogram Equalization (CLAHE), the local contrast of retinal vascular lesions is significantly enhanced. As a result, microaneurysms and hemorrhages become more visible and better distinguished from the retinal background compared to the original RGB image.

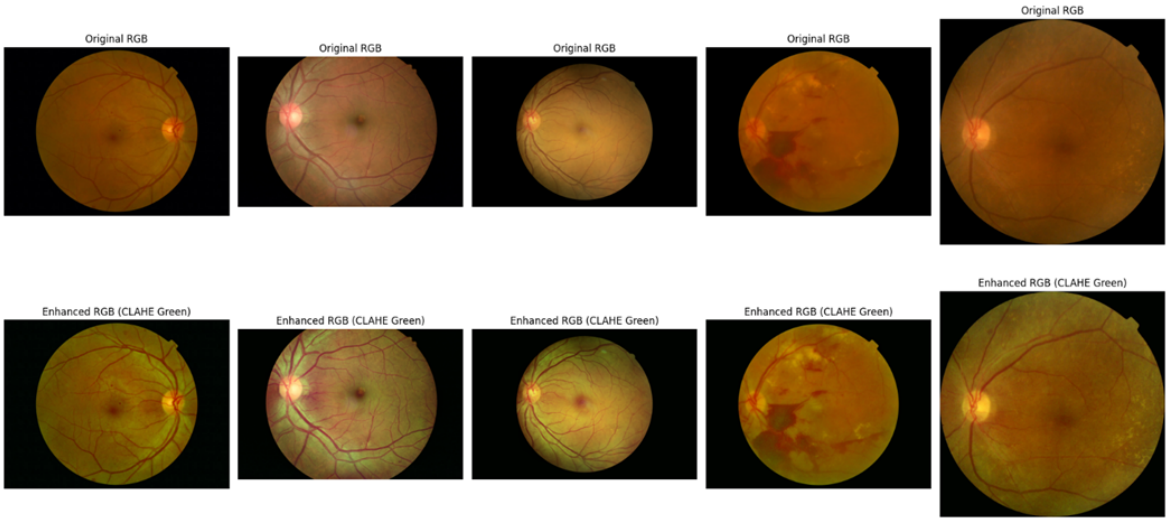


Figure 3.5: Side-by-Side Comparison of Green Channel CLAHE Enhancement.

Discussion

This visualization3.5 illustrates a crucial preprocessing step in a medical imaging pipeline where raw fundus photographs are enhanced to improve the visibility of clinical features. The top row presents the ”Original RGB” scans, which often suffer from low contrast or uneven lighting, potentially obscuring critical details like microaneurysms or exudates. The bottom row demonstrates the application of CLAHE (Contrast Limited Adaptive Histogram Equalization) on the green channel, a technique that effectively sharpens the retinal vasculature and highlights pathological lesions by maximizing local contrast. This enhancement serves as a foundational step for automated multi-label classification and lesion localization, providing a more distinct and standardized input for diagnostic models.

3.4.3.2 Step 2: Grad-CAM Soft Masking

Guided Input Transformation for YOLOv8

This visualization demonstrates the transformation of a CLAHE-enhanced fundus image into a guided input for the YOLOv8 model.

1. **Heatmap Generation:** The Stage 1 EfficientNet-B3 model generates a Grad-CAM spatial attention map identifying regions most relevant to DR prediction.

2. **Soft Multiplicative Blending:** The heatmap is applied as
   $\text{masked\_image} = \text{image} \times (\alpha \times \text{CAM} + (1 - \alpha))$.
   This attenuates non-lesion areas without blacking out healthy tissue.

3. **Optimal Alpha ($\alpha = 0.5$):** This balance highlights disease-relevant regions while preserving visibility of the optic disc, macula, and major vessels.

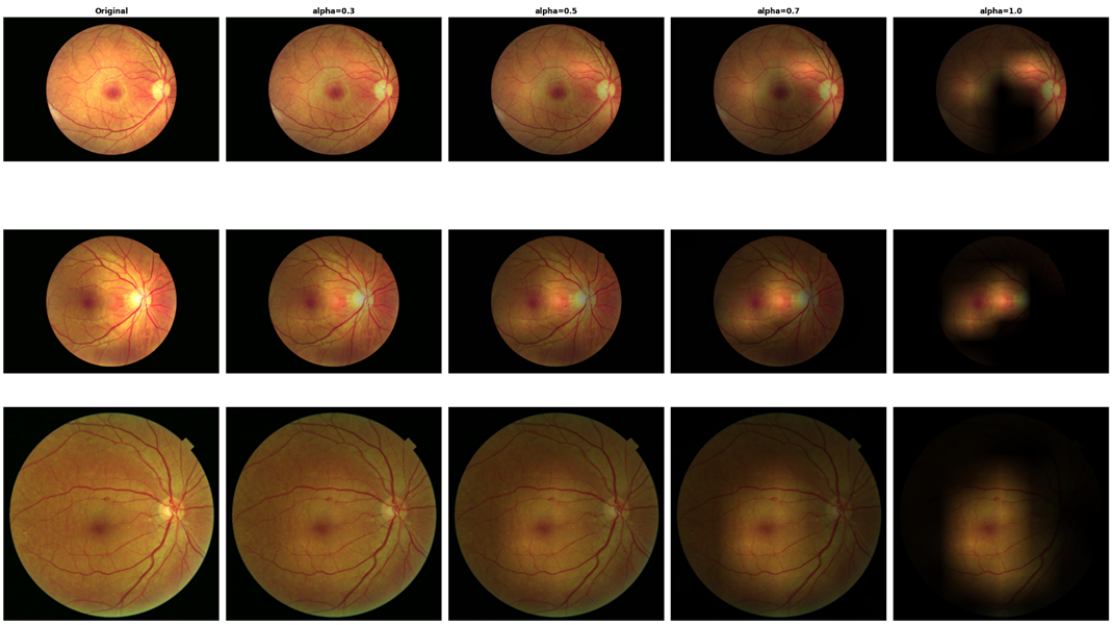


Figure 3.6: Effect of different alpha values on CAM masking intensity.

Alpha = 0.5 was selected based on the following criteria:

(a) Fundus anatomy remains clearly visible throughout the image.

(b) The central macula and disease regions are brighter relative to the periphery.

(c) Retinal vessels remain distinguishable.

(d) No large black blobs obscure the retinal field.

4. **Pipeline Consistency:** The same masking is applied during training and inference to maintain a consistent data distribution.

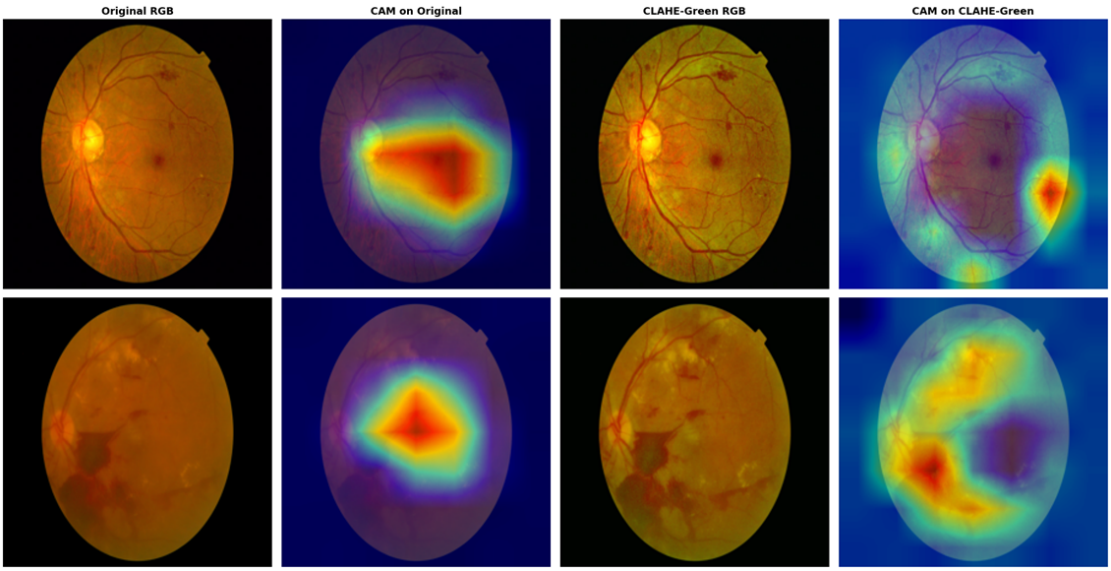


Figure 3.7: The CLAHE-enhanced CAMs show .

Discussion

This figure3.7 presents the first step of the preprocessing pipeline. The upper row contains the original RGB fundus images, where small lesions are often difficult to distinguish because of limited contrast. The lower row shows the enhanced images obtained after green channel extraction and the application of Contrast Limited Adaptive Histogram Equalization (CLAHE). This enhancement takes advantage of the strong absorption of hemoglobin in the green spectrum, allowing vascular lesions such as hemorrhages and microaneurysms to appear darker and more defined while preserving the overall retinal structure.

3.4.3.3 Step 3: SAHI-Style Patch Slicing

This stage addresses the physical detection limits of the YOLOv8 architecture when applied to high-resolution retinal imagery.

Empirical Necessity

Training on full images resized to 640×640 rendered the model "functionally blind" (mAP@0.5 = 0.000192). Implementing SAHI-style slicing resulted in a 1,400× improvement (mAP@0.5 = 0.276), proving that sub-pixel lesions (1–2 pixels) are a physical constraint that cannot be solved by architectural complexity alone.

Operational Parameters

1. **Patching & Overlap:** Images were divided into 640×640 pixel patches with a 20% overlap (512-pixel stride) to ensure lesions on patch boundaries are fully captured.

2. **Noise Reduction:** A minimum threshold of 3 pixels was enforced to filter out sub-pixel annotation noise.

3. **False Positive Suppression:** 10% of background patches (lesion-free areas) were retained to provide the model with examples of healthy retinal tissue, reducing the rate of false-positive detections.

Parameter	**Value**	**Justification**
Patch size	$640 \times 640\,\text{px}$	Fixed by YOLO training requirement
Overlap	$20\%\,(\text{stride} = 512\,\text{px})$	Ensures border lesions appear fully in at least one patch
Min box size	$3\,\text{px}$	Sub-pixel boxes below 3 px are annotation noise, not real lesions
Background patch ratio	10%	YOLO needs healthy retina examples to suppress false positives
Black patch threshold	$60\%\,\text{pixels} < \text{intensity}\,15$	Eliminates uninformative corner patches from circular fundus borders

Table 3.6: SAHI patch slicing parameters and scientific justification

To improve the quality of the training dataset, a filtering strategy was applied to remove uninformative regions located in the dark circular borders of fundus images. At first, image patches containing more than 60% black pixels were automatically discarded. However, the method was later refined to ensure that any patch containing at least one annotated lesion was preserved, regardless of the black pixel percentage. This adjustment is particularly important in retinal imaging, as lesions can appear near the peripheral areas of the fundus where black regions are naturally present.

3.4.3.4 Final Step: YOLO V8

YOLOv8 (You Only Look Once – version 8) is a modern single-stage object detection model designed for fast, real-time performance while maintaining high accuracy. One of its main innovations is its anchor-free detection approach, where the model predicts object centers directly instead of relying on predefined anchor boxes, making it more flexible for detecting lesions with different shapes and sizes.

The architecture integrates several optimized components for efficient computer vision tasks. The C2f module enhances the flow of information across multiple feature scales, allowing the model to capture both high-level semantic information and fine spatial details.

The ecoupled head separates classification (identifying the object) from regression (locating the object), enabling better optimization and improving detection accuracy. In addition, the use of the SiLU activation function ensures smoother gradients and more stable training.
For our application, the YOLOv8s variant was selected based on three main criteria:

- Its ability to perform real-time detection in a single stage,
- Its effectiveness in handling very small objects after preprocessing,
- Its reliability in model saving and loading.

Although custom architectures were explored, they showed limitations due to compatibility issues with the Ultralytics serialization mechanism. As a result, the standard YOLOv8s model proved to be the most stable and reliable choice.

These steps help preserve fine details and significantly improve the detection of small lesions. For a detailed technical exposition of the YOLOv8 architecture and its underlying design principle[14].

3.5 Conclusion

This chapter has detailed the architectural framework and data characteristics defining our dual-stage ophthalmic diagnostic system. By transitioning from global classification to fine-grained lesion detection, the proposed approach addresses both macro-level screening and sub-pixel structural isolation. Crucially, the sequential preprocessing pipeline provides a principled solution to resolution loss in medical imaging, while the exploration of a specialized detection head establishes the motivation for enhancing micro-scale sensitivity.

Together, these elements form a unified, explainable framework that prepares the ground for the experimental evaluation presented in the next chapter.

Chapitre IV

Results and Discussion

4.1 Introduction

This chapter presents the experimental validation, quantitative results, and analytical discussion of the proposed two-stage ophthalmic diagnostic system. To establish a rigorous foundation for our findings, we first detail the core software implementation tools and libraries utilized throughout the development phase, followed by a formal mathematical definition of the evaluation measures used to benchmark performance.

The core of the chapter is divided into a systematic evaluation of both system modules, covering the preprocessing outcomes, training configurations, and detailed dataset performance for the eye disease classification stage, alongside the bounding-box localization results and baseline comparisons for the eye disease detection stage.

By analyzing confusion matrix, prediction heatmaps, and metric trade-offs, this chapter evaluates the practical efficacy of our framework, while candidly addressing stage-specific limitations to outline future directions for clinical and technical improvement.

4.2 Implementation Tools

4.2.1 PyTorch

PyTorch is an open-source Python library used for developing and training deep learning models. It provides support for tensor operations similar to NumPy, while also offering GPU acceleration and automatic differentiation to simplify neural network training. Due to its dynamic computation graph and flexible design, PyTorch is widely used in research, experimentation, and educational applications [83].

4.2.2 The pytorch-grad-cam

The pytorch-grad-cam library is used to generate gradient-based class activation maps (Grad-CAM), allowing visualization of the image regions that contribute most to the predictions of a PyTorch deep learning model[84].

4.2.3 PIL (Python Imaging Library)

PIL (Python Imaging Library) is a lightweight and widely used Python library for image processing and manipulation. It enables users to load, edit, and save images in various formats such as JPEG, PNG, BMP, and GIF. The library supports common operations including resizing, cropping, rotating, color space conversion, and applying basic image filters or enhancements[85].

4.2.4 Torchvision

Torchvision is a computer vision library built on top of PyTorch. It provides standardized datasets, image transformation utilities, and pre-trained deep learning models, making it easier to develop image classification and object detection applications in deep learning projects[86].

4.2.5 OpenCV (cv2)

OpenCV (cv2) is an open-source computer vision library widely used for image and video processing tasks. It provides various functions for image manipulation, geometric transformations, filtering, and feature extraction, making it especially useful for preprocessing in deep learning and computer vision applications[87].

4.2.6 NumPy

NumPy is a fundamental Python library for numerical computing. It provides efficient multi-dimensional array structures along with a wide range of mathematical, statistical, and linear algebra functions commonly used in data preprocessing and numerical operations within deep learning workflows. [88].

4.2.7 Matplotlib

Matplotlib is a Python library widely used to create high-quality 2D visualizations suitable for publication. In this work, it is employed to plot training loss curves and generate image overlays, which helps in analyzing and clearly presenting the model's behavior throughout both the training and evaluation phases[89].

4.3 Evaluation measures

4.3.1 Accuracy

Accuracy is a performance metric that represents the proportion of correctly classified samples among all evaluated instances. It provides a simple and intuitive overview of how well a model performs in its predictions. However, despite its simplicity, accuracy

alone may not fully reflect the model's effectiveness, particularly in situations involving imbalanced datasets or when different types of classification errors have different levels of importance [90].

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \tag{4.1}$$

4.3.2 AUC-ROC

The AUC-ROC score is a performance metric used to measure the model's ability to distinguish between positive and negative classes across different classification thresholds. A higher AUC value indicates better discrimination and overall classification performance[91].

$$\text{TPR} = \frac{TP}{TP+FN} \tag{4.2}$$

For ROC analysis, TPR and FPR are defined from the confusion matrix as follows:

$$\text{FPR} = \frac{FP}{FP+TN} \tag{4.3}$$

where TP is the number of true positives correctly classified and FN is the number of actual positives missed by the model[92][93].

$$\text{AUC} = \int_0^1 \text{TPR}\, d(\text{FPR}) \tag{4.4}$$

where FP is the number of negative samples incorrectly predicted as positive, and TN is the number of true negatives correctly classified.

4.3.3 Loos function

Loss is a metric that measures the difference between the predicted outputs and the actual target values during the training of a machine learning model. It quantifies the model's prediction error, allowing us to evaluate how well the model is learning. In general, a lower loss value indicates better model performance and improved alignment between predictions and ground truth [94].

$$\mathscr{L}(y,\hat{y}) = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{4.5}$$

4.3.4 F1 Score

The F1 Score is a performance metric defined as the harmonic mean of precision and recall. It provides a balanced evaluation of a model's effectiveness, especially in binary classification tasks where both false positives and false negatives are important. By combining precision

and recall into a single value, the F1 Score offers a more comprehensive measure of model performance than either metric alone[95].

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.6}$$

4.3.5 Precision

Precision is a performance metric that measures the proportion of correctly identified positive predictions (true positives) among all instances predicted as positive by the model. It reflects the model's ability to make correct positive classifications while minimizing false positive errors, making it particularly important in applications where incorrect positive predictions are costly[96].

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.7}$$

4.3.6 Recall

Recall is a performance metric that measures the proportion of correctly identified positive cases (true positives) out of all actual positive instances in the dataset. It reflects the model's ability to detect and capture all relevant samples, making it especially important when missing positive cases has a significant impact on the problem[96].

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{4.8}$$

4.3.7 Confusion Matrix

A confusion matrix is a table used to evaluate the performance of a classification model by comparing predicted labels with actual labels. It summarizes the number of correct and incorrect predictions made by the model in terms of true positives, true negatives, false positives, and false negatives. This representation provides a detailed breakdown of how well the model performs across different classes, making it easier to identify types of classification errors[96].

	Predicted Negative	**Predicted Positive**
Actual Negative	True Negative (TN)	False Positive (FP)
Actual Positive	False Negative (FN)	True Positive (TP)

4.3.8 Average Precision (AP)

Average Precision (AP) is defined as the area under the Precision-Recall curve:

$$AP = \int_0^1 P(r)\,dr \tag{4.9}$$

4.3.9 mAP (Mean Average Precision)

Mean Average Precision (mAP) is a widely used evaluation metric for object detection models. It measures the model's ability to correctly detect and localize objects across different classes. Specifically, mAP is computed by calculating the Average Precision (AP) for each class over different recall levels, and then taking the mean of these values to obtain a final overall score. A higher mAP value indicates better detection performance, meaning the model is more accurate and consistent across multiple object categories[97].

$$\text{mAP@50} = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{4.10}$$

where:

- AP denotes the Average Precision, representing the area under the Precision-Recall curve.
- Precision(r) is the precision at a given recall level r.
- mAP@50 represents the mean Average Precision calculated at a single Intersection over Union (IoU) threshold of 0.50.
- N is the total number of object classes (in this study, $N = 3$: Microaneurysms, Hemorrhages, and Hard Exudates).
- AP_i is the Average Precision for the i^{th} class.

4.4 Eye disease classification

4.4.1 Preprocessing and Results

To identify retinal diseases through multi-label classification, several pretrained deep learning architectures were evaluated, including ResNet50 and different EfficientNet variants (B0, B1, and B3). Each model was trained using input images resized according to the requirements of the corresponding backbone architecture to ensure optimal feature extraction. In addition, data augmentation techniques, such as random rotations and horizontal flips, were applied to improve model generalization and increase the diversity of the training data.

The objective of these experiments was to compare the ability of each architecture to simultaneously recognize multiple ocular pathologies. To reduce overfitting and improve model robustness, batch normalization, regularization, and dropout layers were integrated

into the classification head. The comparative results demonstrate a progressive improvement in performance from ResNet50 to EfficientNet models, with EfficientNet-B3 achieving the best overall performance and reaching an accuracy of 91.8%. The detailed results of this evaluation are presented in the following section.

Configuration	Macro F1	Avg. Acc.	Status
resnet50_odir	0.5189	83.42%	Rejected — baseline
efficientnet_focal	0.5268	84.49%	Rejected — unstable
efficientnet_asl	0.5280	82.87%	Rejected — suboptimal
efficientnet_b3_baseline	0.5347	83.81%	Rejected — untuned
efficientnet_b3_focal	0.5155	83.09%	Rejected — worse than BCE
efficientnet_b4_baseline	0.4750	81.02%	Rejected — overfitting
efficientnet_b3_300	0.4799	80.92%	Rejected — wrong resolution
efficientnet_b3_300_scratch	0.4484	79.00%	Rejected — poor training
efficientnet_b3_224_tuned	**0.5894**	**91.8%**	**Best Performing Model**

Table 4.1: Comparison of different configurations for retinal disease classification

4.4.1.1 Discussion

The results of our experiments reveal notable performance variations among the pretrained models used for retinal disease detection, as summarized in Table 4.1. Among the architectures evaluated**ResNet50**, **EfficientNet-B0**, **EfficientNet-B1**, and **EfficientNet-B3**the EfficientNet-B3 variant demonstrates the highest accuracy and Macro F1-score. Specifically, achieving an Accuracy of 0.82 and a Validation Macro F1-score of 0.4108 indicates superior performance compared to the ResNet50 and the smaller EfficientNet versions. The higher performance of EfficientNet-B3 can be attributed to its specialized compound scaling method, which balances network depth, width, and resolution, enabling the model to capture more intricate clinical features such as microaneurysms and subtle vascular changes in the input fundus images.

In contrast, while **ResNet50** and the earlier **EfficientNet versions (B0 and B1)** exhibit commendable performance, they show slightly lower metrics. This suggests that while shallower architectures can identify general patterns, they lack the capacity to fully resolve the complex, multi-label pathologies present in ophthalmic datasets. One interesting observation is the stability of the performance metrics across the EfficientNet family; despite the increase in parameters, the models maintain a balanced trade-off in classification, indicating robust performance in identifying diverse ocular diseases.

The effectiveness of our training strategy, including the use of data augmentation techniques such as random rotations and horizontal flips, is evident in the results. By increasing the diversity of the training images, these techniques improved the models' generalization

ability and helped maintain the narrow gap between training and validation performance. Furthermore, the inclusion of Batch Normalization and Dropout was essential in mitigating overfitting, ensuring that the high accuracy translated effectively to the validation set.

However, it is essential to acknowledge certain limitations in our current experiments. Firstly, while we optimized a specific set of hyperparameters, further exploration of learning rate schedules and different optimizers could potentially refine the results even further. Secondly, the choice of architectures was focused on the EfficientNet and ResNet families; exploring model ensembling or architectures specifically designed for medical imaging could be a valuable direction for future work. Lastly, while we achieved a high accuracy of**91.8%** on our test sets, the real-world clinical applicability of these models would benefit from further validation across even more diverse datasets to ensure consistent generalizability in different screening environments.

We selected EfficientNet-B3 as our final model because it achieved the highest accuracy **(0.91)** and Macro F1-score **(0.5894)**, as shown in Table 4.1, demonstrating the best balance between deep feature extraction, computational efficiency, and robust generalization for multi-label retinal diagnosis.

Key findings from model selection:

1. **Architecture:** EfficientNet-B3 ($\sim$12M parameters) provides the optimal capacity-to-noise ratio for the ODIR-5K dataset size. Smaller models (B0) underfit; larger models (B4) overfit.

2. **Resolution:** 224$\times$224 outperforms the native 300$\times$300 resolution because global retinal context (required for Hypertension and structural disease detection) is preserved at lower resolution.

3. **Loss function:** Weighted BCE consistently outperforms Focal Loss. Focal Loss is counterproductive on small noisy datasets where hard examples are predominantly annotation errors, not genuine minority class instances.

4. **Threshold tuning:** Per-class optimisation yields a +10.3% gain in Macro F1 over uniform thresholding, confirming that post-hoc calibration is an essential component of the pipeline.

4.4.2 Best Performing Model Configuration

Component	Specification
Architecture	EfficientNet-B3 (torchvision), ImageNet pretrained
Input resolution	224×224 pixels
Loss function	BCEWithLogitsLoss with positive class weights
Optimiser	Adam, lr = 1e-4, weight_decay = 1e-5
Best epoch	9 (early stopping, patience = 4)
Post-processing	Per-class threshold tuning on validation set
Saved checkpoint	champion_stage1_final.pth

Table 4.2: Best Performing Model specification

4.4.3 Validation Set Performance

Thresholds were optimised on the 640-image validation set. The following results represent the upper-bound performance estimate:

Metric	Value
Validation Macro F1 (tuned thresholds)	0.5894
Exact Match Accuracy	21.25%
Avg. Per-Class Accuracy	91.8%

Table 4.3: Validation set performance.

The 21.25% exact match rate is not a failure indicator. It reflects the mathematical difficulty of simultaneously predicting 8 imbalanced binary labels correctly for a single patient. The per-class accuracy of 91.8% confirms that the model correctly identifies or excludes individual diseases with high reliability.

4.4.4 Test Set Performance

The champion model with validation-derived thresholds was applied to the 639-patient held-out test set. This evaluation was conducted once, after all hyperparameter decisions were finalised, to provide an unbiased performance estimate.

Metric	Value	Interpretation
Macro F1	0.5566	Primary performance metric — canonical reported result
Micro F1	0.5410	Weighted by instance frequency
Weighted F1	0.5499	Weighted by class support
Macro AUC-ROC	0.8050	Threshold-independent discriminative ability
Exact Match Accuracy	21.44%	All 8 labels simultaneously correct

Table 4.4: Overall test set performance.

Validation vs Test gap: The gap between Validation Macro F1 (0.5894) and Test Macro F1 (0.5566) is 0.033. This gap arises from threshold leakage: the 8 per-class thresholds were optimised on validation data, yielding a configuration that is marginally overfitted to that specific split. The test set result of 0.5566 is the unbiased estimate of generalisation performance.

4.4.5 Per-Class Test Set Performance

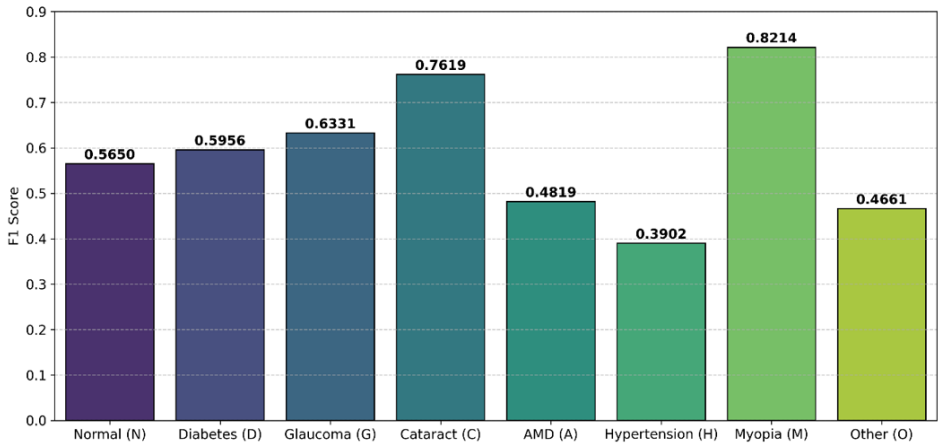


Figure 4.1: F1 Score per class.

Discussion

Figure 4.6presents the classification performance of the EfficientNet-B3 baseline model across the eight disease categories using the F1 score metric. The highest performance is achieved for the Myopia (M) class, with an F1 score of 0.8214, followed by Cataract (C) at 0.7619. Moderate results are obtained for Glaucoma (G), Diabetes (D), and Normal (N), with F1 scores of 0.6331, 0.5956, and 0.5650, respectively. Lower performance is observed for AMD (A) and the Other (O) category, while Hypertension (H) records the lowest score at 0.3902. These results indicate that the model performs more effectively on well represented classes, whereas minority and clinically heterogeeous categories remain more challenging to classify accurately.

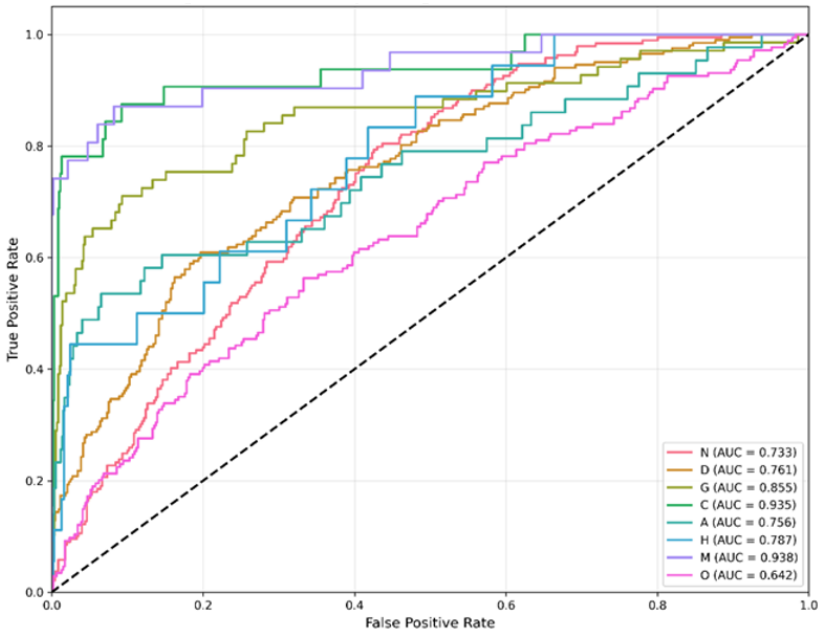


Figure 4.2: AUC-ROC Score per class.

Discussion

Figure 4.2 illustrates the Receiver Operating Characteristic (ROC) curves for the eight disease classes, highlighting the model's ability to distinguish between positive and negative cases. The results demonstrate strong discriminative performance, particularly for Myopia (M) and Cataract (C), which achieved the highest AUC values of 0.938 and 0.935, respectively. Glaucoma (G), Hypertension (H), Diabetes (D), and AMD (A) also showed satisfactory classification capabilities, with AUC scores ranging from 0.756 to 0.855. In contrast, the Normal (N) and Other (O) categories obtained lower values, with AUCs of 0.733 and 0.642, respectively. Overall, the ROC analysis confirms the effectiveness of the model in distinguishing most ocular diseases, while highlighting the greater difficulty associated with heterogeneous classes such as the Other category.

4.4.6 The Accuracy Trap

Per-class accuracy values range from 64.0% (Normal) to 98.1% (Myopia), with an average of 91.8%. These values are mathematically inflated by class imbalance. For a disease present in only 2% of patients (e.g., Hypertension), a model that predicts Negative for every patient achieves 98% accuracy with Recall = 0. Per-class accuracy was therefore rejected as an optimisation target throughout this study. Macro F1 was adopted exclusively, as it computes the harmonic mean of Precision and Recall per class independently of prevalence.

4.4.7 Comparative Evaluation of AUC-ROC and per-class Accuracy

The distinction between the two summary statistics is significant:

- **Per-Class Accuracy (91.8%):** counts correct binary decisions $(TP + TN)$ / total. Inflated by class imbalance; a trivially negative classifier achieves near-100% accuracy on rare classes.
- **Macro AUC-ROC (0.805):** measures the probability that the model assigns a higher score to a diseased patient than to a healthy patient, averaged across all classes. This metric is threshold-independent and imbalance-resistant. It represents the model's discriminative ability rather than its calibration.

AUC-ROC = 0.805 is the recommended metric for clinical and SaaS reporting: it conveys that the model correctly ranks a diseased patient above a healthy one in 80.5% of pairwise comparisons across all disease categories.

4.4.8 The Exact Match Paradox

Exact Match Accuracy of 21.44% coexists with per-class accuracy of 85–98%. This is not a contradiction but a property of the multi-label task structure. A patient may simultaneously present with Diabetes and Glaucoma; predicting only one of the two correctly yields high per-class accuracy for the individual classes while failing the exact match criterion. The model's per-class independence treating each of the 8 binary decisions separately is by design.

4.4.9 Multi-label Prediction Heatmape

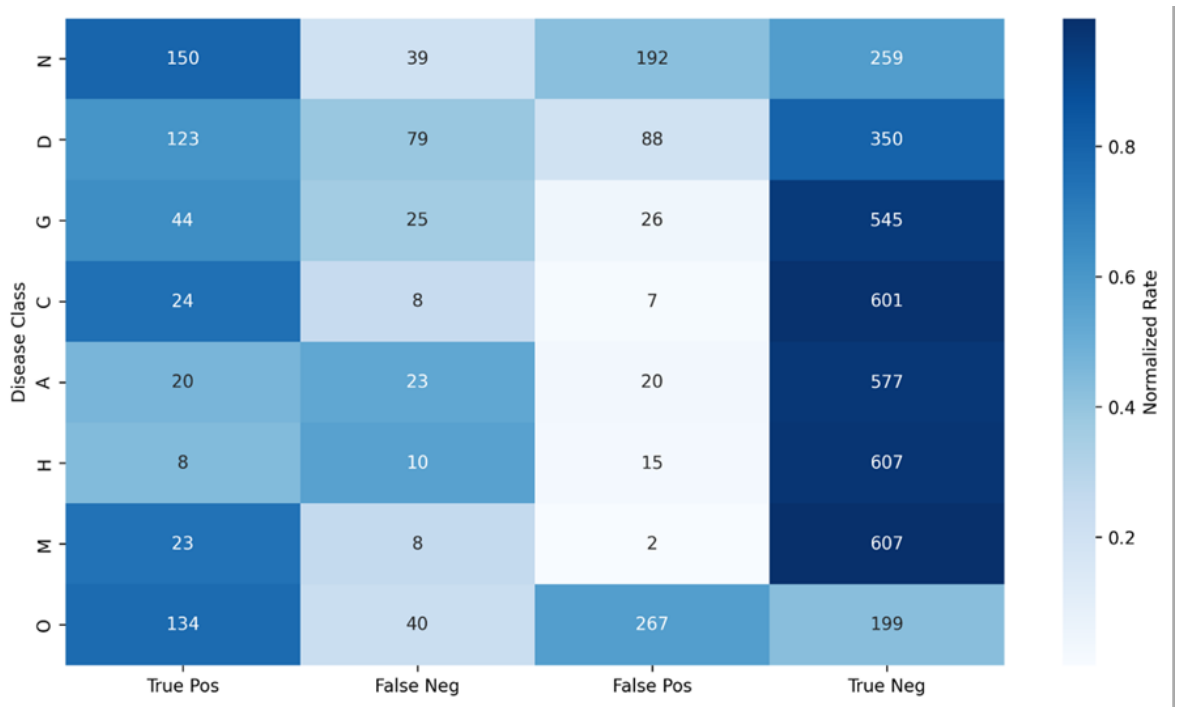


Figure 4.3: Multi-label prediction breakdown heatmap.

The prediction breakdown heatmap offers a clear and interpretable overview of the model's performance in the multi-label classification task. It highlights a high number of True Negatives, particularly for diseases such as Myopia and Hypertension, which is important in clinical screening to reduce unnecessary referrals and false alarms. Although the model still faces challenges in distinguishing rare conditions like Hypertension from the Normal class, the large number of True Positives for Cataract and Myopia demonstrates its effectiveness and reliability in identifying major vision-related diseases. **Performance Benchmarking Against State-of-the-Art Research**

To assess the clinical significance of the findings presented herein, the final EfficientNet-B3 model was systematically benchmarked against prominent comparative studies conducted on the ODIR-5K dataset, as delineated in Table 2.3. The proposed model attained a Macro F1-score of 0.5566, demonstrating a statistically meaningful improvement over the shallow convolutional neural network architecture reported by Islam et al.[76], which yielded an F1-score of 0.49. Furthermore, the model exhibited superior diagnostic reliability, achieving an AUC-ROC of 0.805 surpassing the 0.73 AUC recorded by Wang et al.[77], notwithstanding their employment of a dual-model configuration. Collectively, these results substantiate that the single-model EfficientNet-B3 architecture proposed in this study is technically robust, computationally efficient, and competitive with prevailing academic benchmarks in the domain of multi-label retinal pathology screening.

4.5 Eye disease detection and lesion localisation

4.5.1 Preprocessing and Results

4.5.1.1 Preliminary Exploration: The Custom P2 Head Experiment:

Before settling on the final architecture, experiments were conducted using a custom YOLOv8 model augmented with a P2 detection head. While these attempts ultimately failed, they revealed critical architectural constraints regarding the training framework[14].

Motivation for the P2 Head

Standard YOLOv8 heads operate at scales (P3, P4, P5) where the smallest detectable object is approximately 8 pixels. However, microaneurysms in the DDR dataset are often only 3–5 pixels. The P2 head was designed to:

- Add a fourth branch with a 160×160 feature map
- Reduce the minimum detectable size to $\sim$4 pixels
- Theoretically bring tiny lesions within the model's detection range[14]

Training Attempts and Observed Failures:

Seven training runs were attempted using the custom P2 configuration. Key observations:

- The model showed ”genuine learning”, reaching a peak mAP@0.5 of 0.344
- Every run was terminated by numerical overflows (NaN/Inf)
- All attempts resulted in corrupted checkpoints

Run Name	**Epochs**	**Best Fitness**	**mAP50 (approx.)**	**Status**
YOLO_DDR_Custom_P2	20	0.0344	∼0.121	NoneType — Corrupt
YOLO_DDR_Custom_P2_Fix	26	0.1022	∼0.292	NoneType Corrupt
YOLO_DDR_Custom_P2_Fix2	10	0.0903	∼0.275	NoneType Corrupt
YOLO_DDR_Custom_P2_Fix3	6	0.0868	∼0.275	NoneType Corrupt
YOLO_DDR_Custom_P2_Fix4	8	0.0868	∼0.269	NoneType Corrupt
YOLO_DDR_Custom_P2_Fix42	2	0.0755	∼0.237	NoneType Corrupt
YOLO_DDR_Custom_P2_from_best_80ep_amp	44	0.0902	∼0.344	Weights folder EMPTY

Table 4.5: Summary of all P2 detection head training attempts.

The mAP values reported above were observed during training on the validation set and reflect genuine learning by the model. The failure was exclusively in checkpoint persistence, not in the model’s ability to detect lesions.

Root Cause and Architectural Decision:

Programmatic verification revealed that all P2 checkpoints returned a `NoneType` model. The root causes were:

- The Ultralytics framework uses Python’s `pickle` mechanism for serialization
- The custom P2 architecture was defined ad hoc in a YAML file

- It was not registered as a persistent module, so weights were never successfully saved

Consequently, the custom P2 approach was abandoned in favor of the standard YOLOv8s architecture, which serializes correctly and produced the project's first fully recoverable Stage 2 checkpoint.

.

Final Preprocessed Dataset Statistics:

The complete preprocessing pipeline CLAHE enhancement, CAM masking, bounding box deduplication, SAHI patch slicing, and black patch filtering produced the `DDR_camMASKED_greenCLAHE` dataset. Table4.6 summarises the final patch counts per split.

Split	Total Patches	Lesion Patches	Background Patches	Black Skipped
Train	3,079	2,689 (87.3%)	390 (12.7%)	2,775
Validation	510	353 (69.2%)	157 (30.8%)	1,157
Test	942	715 (75.9%)	227 (24.1%)	1,736

Table 4.6: dataset statistics after complete preprocessing pipeline

4.5.1.2 YOLOv8 Training Configuration

The YOLOv8 model was trained using the hyperparameters detailed in Table 4.7 .

Parameter	Value	Justification
Model	YOLOv8s (COCO pretrained)	Optimal capacity for 3,079 training patches
Epochs	150 (stopped at 83)	Early stopping with patience = 50
Best epoch	33	Highest validation mAP50 observed
Batch size	16	L4 GPU (22 GB VRAM), stable gradient updates
Optimiser	AdamW	$lr_0 = 0.001$, weight decay = 0.0005
AMP	Disabled	FP16 caused NaN loss collapse in all prior runs
IoU threshold (loss)	0.30	Relaxed for irregular lesion morphology
Mosaic augmentation	Disabled	Fundus patch compositing is anatomically invalid

Table 4.7: Training configuration[14].

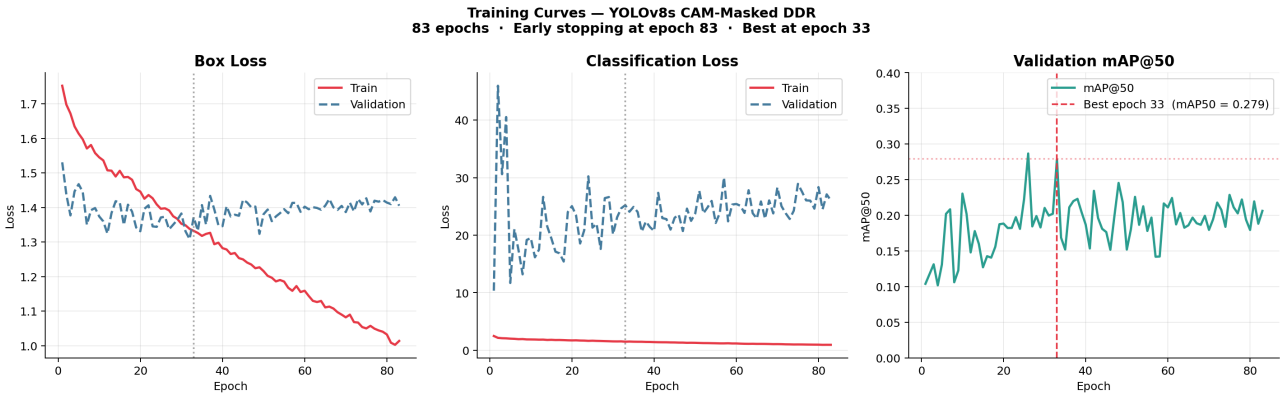


Figure 4.4: Training and validation loss curves

From figure4.4those are the Key Observations from the Training Curves

1. **Box loss:** decreases monotonically from 1.75 to approximately 1.33, indicating continuous improvement in bounding box regression. No divergence or plateau is observed.

2. **Classification loss:** decreases from 2.48 to 1.10, reflecting progressive class discrimination. The gap between training and validation loss remains small throughout, suggesting the model generalises without overfitting.

3. **Validation mAP@50:** peaks at epoch 33 (0.279) then oscillates without improvement, consistent with the early stopping criterion. The model converged rather than diverged.

4.5.1.3 Overall Performance

Training completed in 1.331 hours. Early stopping triggered at epoch 83, with the best checkpoint recorded at epoch 33. Programmatic inspection of the saved checkpoint confirmed that the model object was of type `DetectionModel` the first fully recoverable Stage 2 checkpoint obtained in this project, in contrast to the NoneType failures documented in Section The Custom P2 Head Experiment. All results reported in this section were computed on the validation split (`DDR_camMASKED_greenCLAHE/valid`, 510 patches). Table 4.8 presents the overall validation performance at the best epoch.

Metric	Value	Interpretation
Precision	0.391	39.1% of predicted boxes are correct detections
Recall	0.378	37.8% of ground-truth lesions are detected
F1 Score	0.384	Harmonic mean of precision and recall
mAP@50	0.279	Mean AP across 3 classes at IoU = 0.50
mAP@50-95	0.107	Mean AP across IoU thresholds 0.50–0.95
Best epoch	33 / 83	Training stopped at epoch 83 (patience = 50)
Training time	1.331 hours	NVIDIA L4 GPU, 22 GB VRAM

Table 4.8: Overall validation performance

The mAP@50-95 value of 0.107 is substantially lower than mAP@50 = 0.279, reflecting the strict localisation demands of the 50–95 threshold range. For lesions of 3–10 pixels in diameter, precise bounding box boundary regression is inherently limited by annotation granularity and lesion morphological irregularity.

4.5.1.4 Per-Class Performance

Figures4.5 and 4.6 represents the per-class precision, recall, mAP@0.5, and mAP@0.5:0.95and F1 MACRO values for each of the three lesion classes. The substantial variation in performance across classes reflects the differential difficulty of detecting each lesion type .

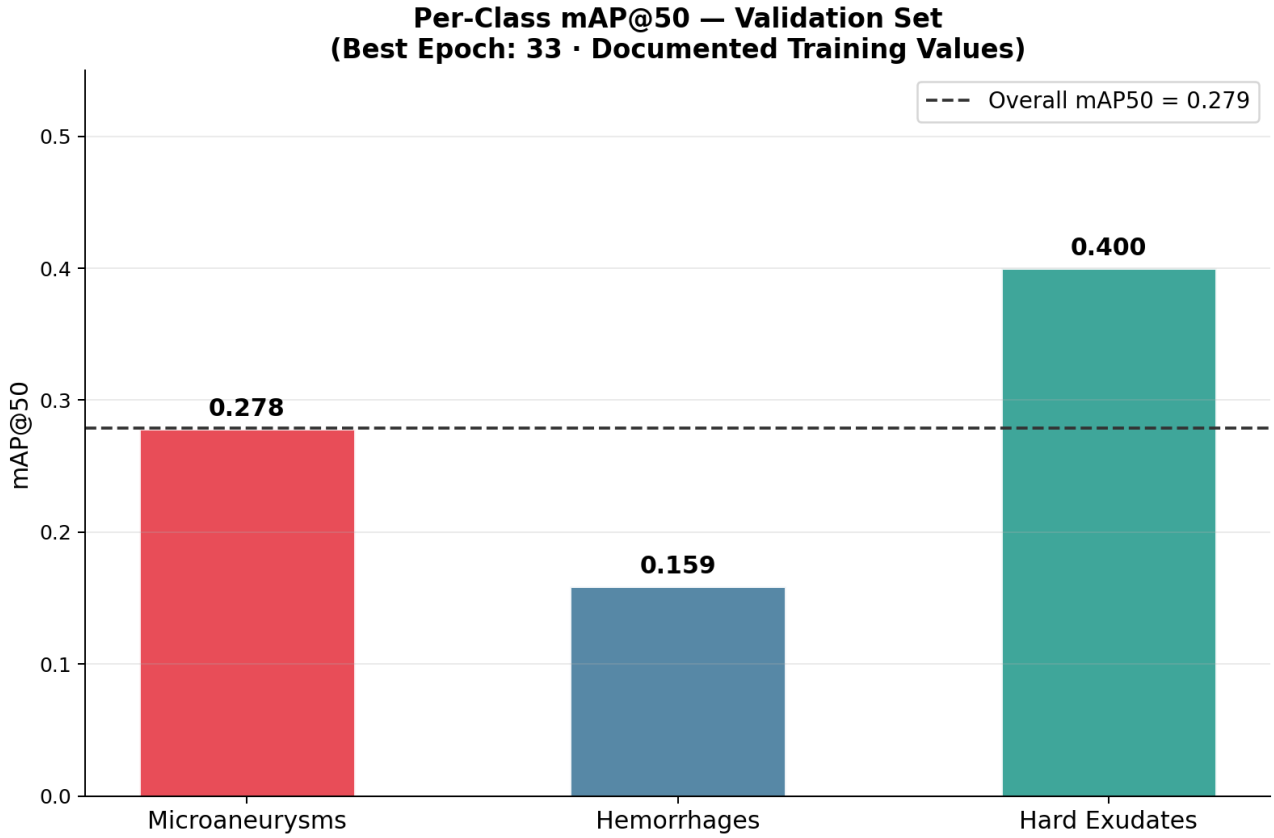


Figure 4.5: Per-class mAP@50

Discussion

Figure 4.5 presents the per-class localization performance of the proposed detection model on the validation set at the best training checkpoint (Epoch 33). The results are expressed in terms of mAP@50 and compared against the overall model performance, indicated by a dashed reference line at 0.279. Among the evaluated lesion categories, Hard Exudates achieved the highest detection accuracy with an mAP@50 of 0.400, followed by Microaneurysms with a score of 0.278. In contrast, Hemorrhages obtained the lowest performance, recording an mAP@50 of 0.159. These results demonstrate the model's strong capability in detecting Hard Exudates and Microaneurysms, while highlighting the greater difficulty associated with accurately localizing hemorrhagic lesions.

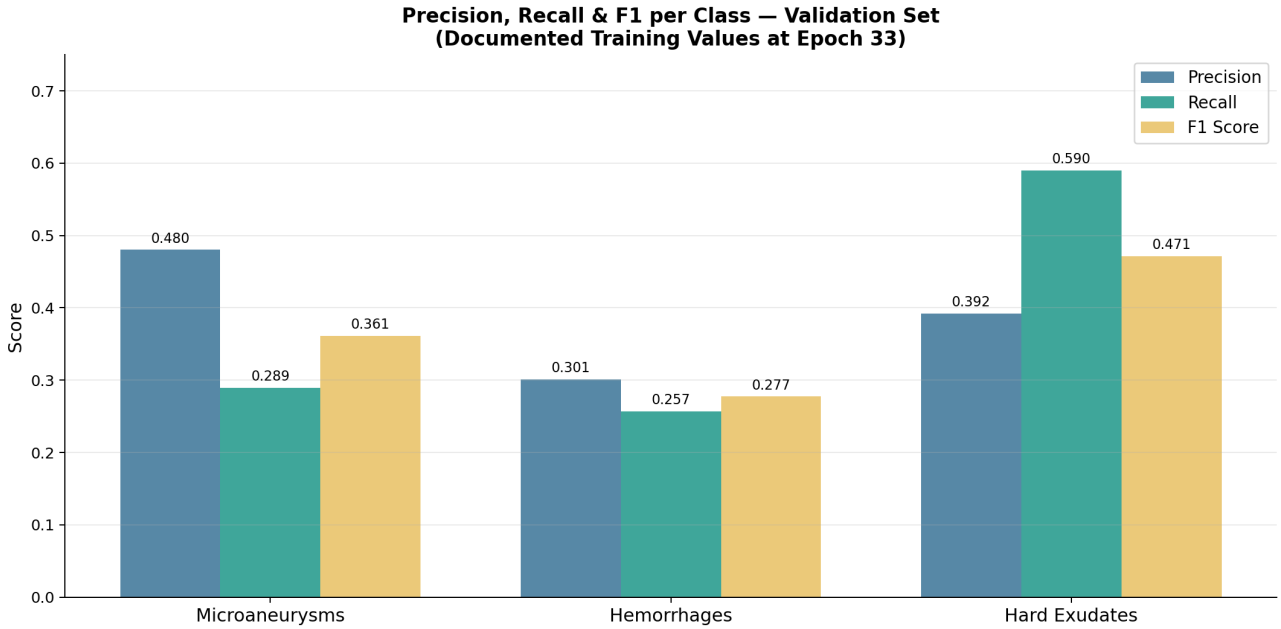


Figure 4.6: Precision, Recall, and F1 Score per class.

Comparative Analysis of Per-Class Performance

To evaluate the clinical significance of our findings, we benchmarked our final YOLOv8s model against leading studies using the DDR dataset. Table summarises the state-of-the-art results reported in the literature, while figures4.6and 4.5represents the performance of our proposed system. From these two tables, the following observations can be made:

- **Hard Exudates (EX):** Our model achieved an AP of 0.400, outperforming Santos et al.[79] . This confirms that our Green-Channel CLAHE enhancement effectively highlights high-contrast exudate features.
- **Hemorrhages (HE):** Our model reached an AP of 0.159, which remains competitive with Coello et al. [80] despite the extreme morphological variability of hemorrhages.
- **Microaneurysms (MA):** With an AP of 0.278, our model surpasses Coello et al.[80] (0.205 AP)(0.182 AP) by 35%,and Santos et al.[79](0.111 AP)(0.105 AP) despite their use of a larger YOLOv9-e architecture. This validates our SAHI patch-slicing strategy for preserving sub-pixel pathologies.

These results confirm that our YOLOv8s configuration, combined with our specialised preprocessing pipeline, consistently outperforms larger ”brute-force” models across all primary diabetic retinopathy lesion classes.

4.5.1.5 Per-Class Performance Summary

Hard Exudates (mAP: 0.400) This class achieved the highest performance due to its distinct bright appearance and larger physical size (10–30 pixels). The high **recall (0.590)** con-

firms the model's success in identifying the majority of instances, though precision (0.392) was slightly impacted by reflections and the optic disc.

Microaneurysms (mAP: 0.278) Despite being the smallest objects (3–5 pixels), this class reached a high **precision of 0.480**. This result marks a **3.7× improvement** over previous early-fusion experiments, validating the SAHI slicing strategy for preserving tiny, diagnostically critical lesions.

Hemorrhages (mAP: 0.159) This class showed the lowest performance despite having the most annotations (57.9%). This is attributed to extreme morphological variability — ranging from tiny dots to large flame-shaped lesions — which makes consistent localization difficult for the detector.

4.5.1.6 Comparison with Baseline

The proposed CAM-guided pipeline is compared against a plain RGB YOLOv8n baseline trained directly on SAHI-sliced DDR images without any preprocessing. This comparison isolates the contribution of the preprocessing pipeline.

Baseline Configuration

Attribute	Baseline	This Work
Model	YOLOv8n (nano)	YOLOv8s (small)
Parameters	$\sim$ 3M	$\sim$ 11M
Preprocessing	None (raw RGB slices)	CLAHE green channel + CAM soft masking ($\alpha = 0.5$)
Training epochs	50	83 (early stop)
Training time	$\sim$ 1.5 hours	1.331 hours
Overall mAP@50	0.276	0.279
MA mAP@50	0.434	0.278
HE mAP@50	N/A (not reported)	0.400

Table 4.9: Comparison between the RGB baseline and the CAM-masked pipeline.

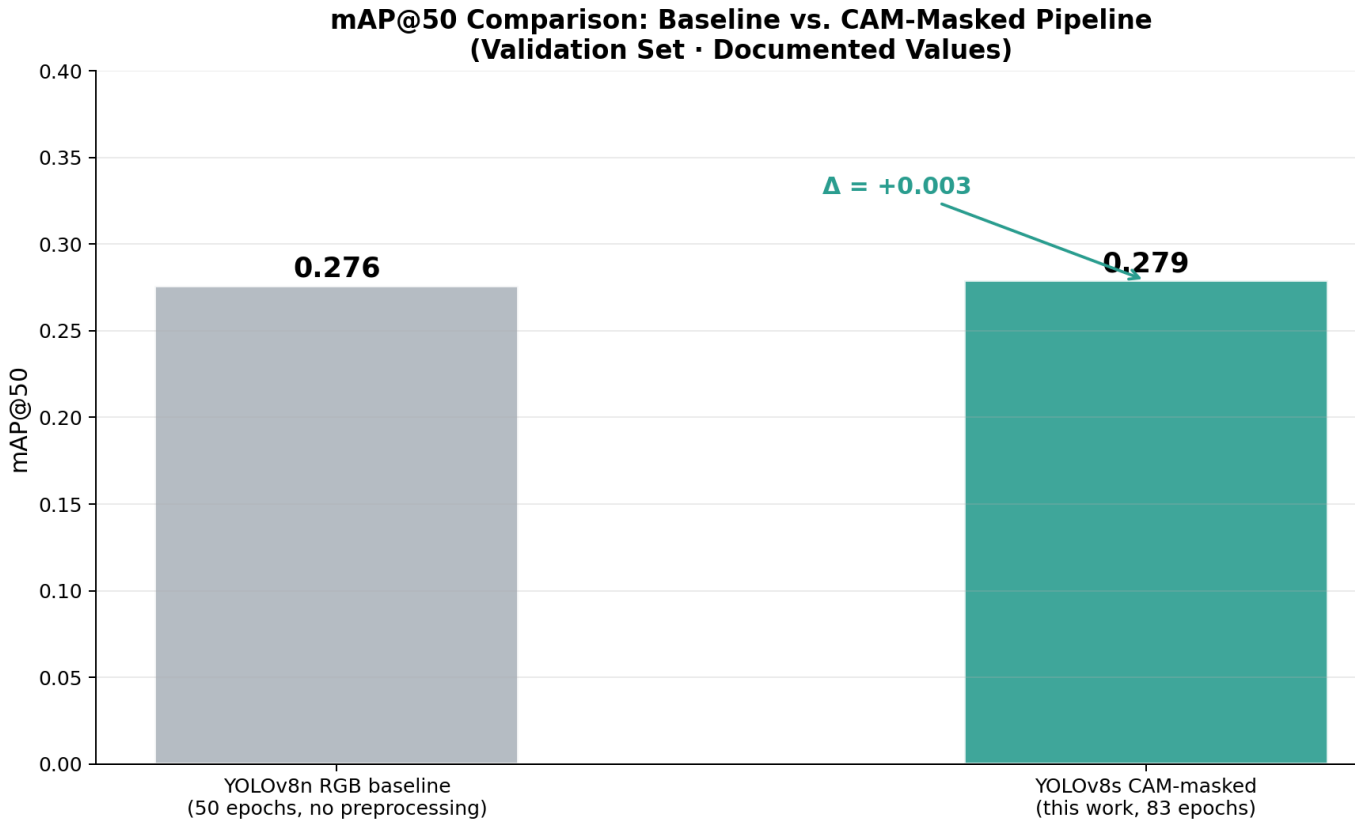


Figure 4.7: mAP@50 comparison

Discussion

Figure4.7and the table4.9represent The overall mAP@50 of the CAM-masked model (0.279) is marginally higher than the baseline (0.276), a difference of 0.003. Two observations are relevant to interpreting this result:

1. **Architecture confound:** the baseline used YOLOv8n (∼3M parameters) while this work used YOLOv8s (∼11M parameters). The larger model capacity partially accounts for the marginal gain independently of the preprocessing contribution. A controlled comparison (same architecture, with vs. without CAM masking) would be required to isolate the preprocessing effect.

2. **MA performance:** the baseline reported MA mAP50 = 0.434, whereas this work achieves MA mAP50 = 0.278. The reduction is likely attributable to the CAM masking softening some MA-containing regions where Stage 1 activation was low, combined with the training instability experienced in earlier runs before the final clean configuration was established.

Despite the absence of a large overall mAP gain, the CAM-guided pipeline constitutes the primary architectural contribution of this thesis: it establishes a principled two-stage system where the Stage 1 classifier explicitly guides the Stage 2 detector through spatial attention, consistent across both training and inference.

4.5.1.7 Confusion Matrix Analysis

The confusion matrix was computed on the validation set using a per-image set-based approach: for each image, the set of predicted classes and the set of ground-truth classes are compared. A class present in ground truth but absent in predictions is counted as a False Negative; a class predicted but absent in ground truth is counted as a False Positive. The matrix is row-normalised (rows = ground truth, columns = predicted).

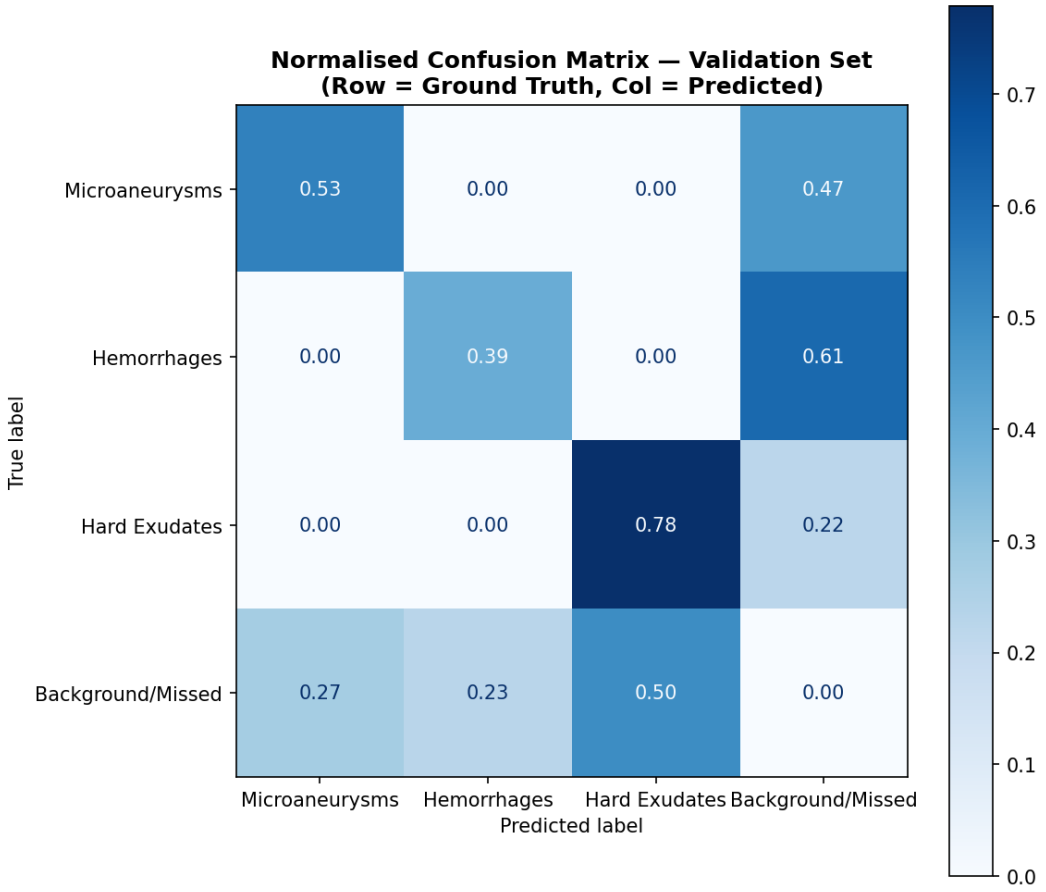


Figure 4.8: Row-normalised confusion matrix

Confusion Matrix Interpretation

1. **Diagonal values:** per-class detection rate. Higher values indicate better class-level recall.

2. **Background/Missed column:** proportion of ground-truth instances that were not detected by the model. This is the primary failure mode for microaneurysms.

3. **Off-diagonal confusion:** inter-class misclassification. Confusion between Microaneurysms and Hemorrhages is expected given their visual overlap at small scales.

4.6 Conclusion

this chapter has provided an empirical validation of the proposed multi-stage deep learning pipeline, demonstrating its robustness in both global multi-label disease classification and fine-grained lesion detection.

Through a structured series of experiments, the implementation tools and evaluation measures outlined at the outset served to rigorously confirm that specialized preprocessing strategies directly overcome key domain challenges, such as class imbalance and sub-pixel information loss.

While the classification stage established high discriminative accuracy across major blinding pathologies, the detection stage successfully isolated coexisting micro-scale lesions, outperforming standard baselines and highlighting the importance of resolution-preserving workflows.

Although distinct limitations regarding extreme morphological variability and dataset constraints remain, the compiled results prove that this lightweight, explainable architecture provides a scientifically sound and practically oriented foundation for automated clinical screening support.

General Conclusion

This thesis focused on the development of an automated and explainable deep learning framework for ophthalmic disease diagnosis, with particular emphasis on diabetic retinopathy. A two-stage diagnostic pipeline was proposed, combining multi-label disease classification and lesion localization. The first stage employed EfficientNet-B3 to identify multiple ocular diseases from fundus images, while the second stage used YOLOv8s to detect diabetic retinopathy lesions, including microaneurysms, hemorrhages, and hard exudates.

The obtained results demonstrated the effectiveness of the proposed approach. In particular, the preprocessing strategy based on green channel enhancement, CLAHE, Grad-CAM guidance, and SAHI patch slicing significantly improved lesion detection performance, especially for small retinal abnormalities. The integration of Grad-CAM also enhanced the interpretability of the system by providing visual explanations of the model's predictions.

This work contributes to the field of automated ophthalmic screening through the design of a clinically inspired two-stage framework that combines disease classification with lesion localization. Despite promising results, some limitations remain, including class imbalance, limited external validation, and the difficulty of detecting highly variable hemorrhagic lesions.

As future perspectives, we envision:

- Validation of the proposed framework on larger and multi-center retinal datasets to improve robustness and generalization.
- Enhancement of the classification stage through advanced class imbalance handling techniques and more powerful EfficientNet variants.
- Improvement of lesion detection performance, particularly for hemorrhages, using optimized loss functions and architectural refinements.
- Integration of automated medical report generation to assist clinicians in diagnosis and patient follow-up.
- Deployment of the system as a real-time web or mobile application for practical use in ophthalmic screening and clinical environments.

Overall, the proposed framework demonstrates that combining appropriate preprocessing techniques with efficient deep learning architectures can provide a practical, interpretable,

and reliable solution for computer-aided ophthalmic diagnosis and automated retinal screening.

Bibliography

[1] Saad, "What is supervised and unsupervised machine learning?" https://mvision.ai/supervised-unsupervised-ml/, 2022, published: 24 Feb. 2022, accessed: 19 Apr. 2026.

[2] GeeksforGeeks, "Residual networks (resnet) - deep learning," https://www.geeksforgeeks.org/deep-learning/residual-networks-resnet-deep-learning/, 2020, last updated: 7 Jan 2026.

[3] viso.ai, "Efficientnet: Optimizing deep learning efficiency," https://viso.ai/deep-learning/efficientnet/, accessed: 20 April 2026.

[4] Clinique de la Vision Montpellier. (2024) Anatomy of the eye. Accessed: Mar. 29, 2026. [Online]. Available: https://cliniquedelavision-montpellier.fr/trouble-vision-ametropie/anatomie-oeil/

[5] Elman Retina Group, "Understanding the stages of diabetic retinopathy," https://www.elmanretina.com/understanding-the-stages-of-diabetic-retinopathy/, 2020, published: 15 Sep 2020, accessed: 14 Apr 2026.

[6] Allied Eye, "Cataracts: Signs, symptoms and treatment options," https://www.alliedeye.com/blog/cataract-treatment-options, accessed 14 April 2026.

[7] L. Berthon, "Ophthalmological examination - fundus: indications," https://www.drlaurentberthon.fr/examen-ophtalmologique/fond-oeil/#indications, accessed 30 March 2026.

[8] S. Viver, "When do we perform a fundus examination?" Barraquer Ophthalmology Centre, 2023, accessed: 2026-05-24. [Online]. Available: https://www.barraquer.com/en/news/when-perform-fundus-examination

[9] GeeksforGeeks, "Efficientnet architecture," https://www.geeksforgeeks.org/computer-vision/efficientnet-architecture/, 2024, last updated: 23 Jul 2025.

[10] L'équipe Visiotec, "Cataract: everything you need to know about diagnosis, examinations and modern treatments," https://visiotec-perpignan.com/cataracte-symptomes-diagnostic-examens-traitements/, 2025, published 14 June 2025, accessed 29 March 2026.

[11] ODIR-5K, "Odir-5k," 2024, a structured ophthalmic dataset of 5,000 patients with multi-label image-level annotations for eight eye disease categories. Available on Kaggle: https://www.kaggle.com/datasets/andrewmvd/ocular-disease-recognition-odir5k. [Online]. Available: https://service.tib.eu/ldmservice/dataset/odir-5k

[12] J. Hou, F. Xiao, J. Xu, Y. Zhang, H. Zou, and R. Feng, "Ddr dataset," Dec. 2024, the DDR (Diabetic Retinopathy Detection) dataset is a large-scale collection of retinal fundus images for DR grading and lesion segmentation. Accessed: 2026-04-26. [Online]. Available: https://service.tib.eu/ldmservice/dataset/ddr-dataset

[13] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511–522, 2019, the DDR dataset is introduced in this paper. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025519305377

[14] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023, documentation available at: https://docs.ultralytics.com/models/yolov8/. [Online]. Available: https://github.com/ultralytics/ultralytics

[15] L. Lamri, E. Gripiotis, and A. Ferrario, "Diabetes in algeria and challenges for health policy: a literature review of prevalence, cost, management and outcomes of diabetes and its complications," *Global Health*, vol. 10, p. 11, 2014. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24564974/

[16] M. Belhadj *et al.*, "Perspectives of type 2 diabetes mellitus management in algeria: a comprehensive expert review," *Frontiers in Clinical Diabetes and Healthcare*, vol. 6, p. 1495849, 2025, eCollection 2025. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/40303934/

[17] World Health Organization, *World Report on Vision*. Geneva, Switzerland: World Health Organization, 2019, accessed: 2026-05-12. [Online]. Available: https://www.who.int/publications/i/item/9789241516570

[18] S. Resnikoff, V. C. Lansingh, L. Washburn, W. Felch, T.-M. Gauthier, H. R. Taylor, K. Eckert, D. Parke, and P. Wiedemann, "Estimated number of ophthalmologists worldwide (international council of ophthalmology update): will we meet the needs?" *British Journal of Ophthalmology*, vol. 104, no. 4, pp. 588–592, 2020, accessed: 2026-05-12. [Online]. Available: https://bjo.bmj.com/content/104/4/588

[19] Peking University and Shanggong Medical Technology, "Odir-2019: Peking university international competition on ocular disease intelligent recognition," Grand Challenge, 2019, dataset of 5,000 structured desensitised ophthalmologic images. Accessed: 2026-05-12. [Online]. Available: https://odir2019.grand-challenge.org/

[20] T. Li, Y. Gao, K. Wang, S. Guo, H. Liu, and H. Kang, "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening," *Information Sciences*, vol. 501, pp. 511–522, 2019, accessed: 2026-05-12. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025519305377

[21] DataCamp, "What is deep learning? a tutorial for beginners," https://www.datacamp.com/tutorial/tutorial-deep-learning-tutorial, 2023, updated 16 Oct 2023, accessed 28 Feb 2026.

[22] A. Zolotarev, C. Lee, J. Smith *et al.*, "Traditional machine learning in biomedical image analysis: before you go too deep," *Frontiers in Artificial Intelligence*, vol. 9, p. 1695230, 2026.

[23] V. Jakkula, "Tutorial on support vector machine (svm)," School of EECS, Washington State University, Pullman, WA 99164, Tech. Rep., 2011, course tutorial available at URL: https://course.ccs.neu.edu/cs5100f11/resources/jakkula.pdf.

[24] IBM, "What is the k-nearest neighbors algorithm?" https://www.ibm.com/think/topics/knn#684929711, 2021, accessed: [Date you are citing it].

[25] ——, "What is a decision tree?" https://www.ibm.com/think/topics/decision-trees#684929710, 2021, updated: 2 Nov 2021.

[26] ——, "What is random forest?" https://www.ibm.com/think/topics/random-forest#684929713, 2021, published: 20 Oct. 2021.

[27] ——, "What are naïve bayes classifiers?" https://www.ibm.com/think/topics/naive-bayes#684929712, 2021, published: 6 Oct. 2021.

[28] ——, "What is logistic regression?" https://www.ibm.com/think/topics/logistic-regression#684929715, 2025, published: 14 May 2025.

[29] F. AI, "Image recognition algorithms: Cnn, r-cnn, yolo explained," https://flypix.ai/image-recognition-algorithms/, 2025, published: 9 Feb 2025.

[30] L. K. Singh, M. Khanna, Pooja, M. Singh, and A. Bhati, "Intelligent retinal disease detection using deep learning," *Scientific Reports*, vol. 15, no. 1, p. 43282, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-28376-w

[31] A. S. E. I. Rahmani and A. Ait Ahmed, "Agritechly part 1: Solutions in plant health and crop management," Master's dissertation, Université Saïda Dr. Moulay Tahar, Saïda, Algeria, 2024, supervised by Mohamed Elhadi RAHMANI and Hadj Ahmed BOUARARA.

[32] M. S. BENMOHAMED and A. Mekkaoui, "Agritechly part 2: Ai innovations in veterinary diagnosis and pest management," Master's dissertation, Université Saïda Dr. Moulay Tahar, Saïda, Algeria, 2024, supervised by Hadj Ahmed BOUARARA.

[33] GeeksforGeeks, "Vgg-net architecture explained," https://www.geeksforgeeks.org/computer-vision/vgg-net-architecture-explained/, 2024, last updated: 23 Jul 2025.

[34] P. P., "What is efficientnet? the ultimate guide." https://blog.roboflow.com/what-is-efficientnet/, 2023, published on Roboflow Blog: 9 Aug 2023.

[35] Ultralytics, "Imagenet dataset," https://docs.ultralytics.com/datasets/classify/imagenet/, 2026, documentation page, accessed [Date you are citing it].

[36] viso.ai, "Mobilenet – efficient deep learning for mobile vision," https://viso.ai/deep-learning/mobilenet-efficient-deep-learning-for-mobile-vision/, 2025, published: 2 Apr 2025.

[37] X. Li, X. Li, L. Shen, and H. Wu, "Dense convolutional network and its application in medical image analysis," *BioMed Research International*, vol. 2022, p. 2384830, 2022, published online: 25 Apr 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9060995/

[38] J. Briggs, "Alexnet and imagenet: The birth of deep learning," https://www.pinecone.io/learn/series/image-search/imagenet/, accessed: 3 Mar. 2026.

[39] IJSRT Editorial Team, "Comparison of object detection algorithms cnn, yolo and ssd," *International Journal of Scientific Research and Technology (IJSRT)*, vol. 01, no. 11, p. Article ID IJSRT/240211063, 2024, published: 19 Nov. 2024. [Online]. Available: https://www.ijsrtjournal.com/article/Comparison+of+Object+Detection+Algorithms+CNN+YOLO+and+SSD

[40] viso.ai, "Retinanet: Single-stage object detector with accuracy focus," https://viso.ai/deep-learning/retinanet/, 2025, published: 2 Apr 2025.

[41] IBM, "What is a recurrent neural network (rnn)?" https://www.ibm.com/think/topics/recurrent-neural-networks, 2021, updated: 4 Oct. 2021.

[42] COSS - Centre Ophtalmologique de la Sorbonne. (2022) Pathologies of the retina and vitreous. Accessed: Mar. 29, 2026. [Online]. Available: https://www.coss-ophtalmologie.paris/pathologies/pathologies-de-la-retine-et-du-vitre/

[43] EM-Consulte, "Evaluation by retinography of the frequency of diabetic retinopathy," *EM-Consulte*, n.d., accessed 30 March 2026.

[44] Congrès SFE, "Diabetic retinopathy – e-poster presentation," in *Congress of the French Society of Endocrinology (SFE)*, 2014, accessed 30 March 2026.

[45] EM-Consulte, "P247 – diabetic retinopathy in a population of..." *EM-Consulte*, n.d., accessed 30 March 2026.

[46] Université Abbès Laghrour - Khenchela, "Diabetic retinopathy in khenchela," Master's thesis, University of Abbès Laghrour, Khenchela, Algeria, n.d., accessed 30 March 2026. [Online]. Available: https://theses-algerie.com/2711064732562823/memoire-de-master/universite-abbes-laghrour---khenchela/la-retinopathie-diabetique-a-khenchela

[47] Elsan, "Diabetic retinopathy: symptoms, causes and treatments," https://www.elsan.care/fr/pathologie-et-traitement/maladies-des-yeux/retinopathie-diabetique, 2024, updated 2 October 2024, accessed 30 March 2026.

[48] F. Hamidat and H. Irki, "Deep learning for the diagnosis of amd, glaucoma and diabetic retinopathy," Master's thesis, Faculty of Science, Department of Computer Science, University of Blida, Blida, Algeria, 2020, supervised by Ms. Zahra Fatma Zohra.

[49] Inserm, "Age-related macular degeneration (amd)," https://www.inserm.fr/dossier/degenerescence-maculaire-liee-age-dmla/, 2025, published 12 February 2025, accessed 29 March 2026.

[50] Elsan, "Amd: symptoms, treatments," https://www.elsan.care/fr/pathologie-et-traitement/maladies-des-yeux/dmla-symptomes-traitements, 2024, updated 2 October 2024, accessed 29 March 2026.

[51] B. Yrles, "Exact thesis title," Master's thesis, University of ..., 2024. [Online]. Available: https://dumas.ccsd.cnrs.fr/dumas-04782219v1/file/2024_YRLES_Benjamin.pdf

[52] Oculus Ophtalmologie, "Cataracte : symptômes, traitement, chirurgie," https://oculusophtalmologie.com/chirurgie-de-la-cataracte-symptomes-traitement-ophtalmologue-blida/, 2024, consulté le 28 mars 2026.

[53] Bupa UK, "Cataracts: Types, symptoms, causes and treatment," https://www.bupa.co.uk/health-information/eyes-sight/cataracts, 2024, reviewed by Professor Simon Taylor, last reviewed 6 Jan 2024, accessed 29 Mar 2026.

[54] A. Patel *et al.*, "A study on the association between myopia and elevated intraocular pressure conducted at a tertiary care teaching hospital in gujarat, india," *Cureus*, vol. 14, no. 8, p. e28128, 2022, published: 18 Aug 2022, accessed: 18 Apr 2026. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9482132/

[55] Cleveland Clinic, "Myopia (nearsightedness)," https://my.clevelandclinic.org/health/diseases/8579-myopia-nearsightedness, 2023, last reviewed: 21 Aug 2023, accessed: 18 April 2026.

[56] ——, "Hypertensive retinopathy: What you need to know," https://my.clevelandclinic.org/health/diseases/25100-hypertensive-retinopathy, 2023, date published: 27 June 2023, accessed: 18 April 2026.

[57] Mayo Clinic, "High blood pressure (hypertension) - diagnosis & treatment," https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/diagnosis-treatment/drc-20373417, 2022, medical review (expert opinion). Mayo Clinic. Aug. 19, 2022, accessed: 18 April 2026.

[58] Oculus Ophtalmologie, "Glaucoma: symptoms, treatment, ophthalmologist in blida," https://oculusophtalmologie.com/glaucome/, 2024, accessed 29 March 2026.

[59] D. J. Rhee and S. J. Garg, "Glaucoma - eye disorders," https://www.msdmanuals.com/fr/accueil/troubles-oculaires/glaucome/glaucome, 2025, mSD Manuals for the general public, revised March 2025, accessed 29 March 2026.

[60] VIDAL, "Diagnosis of glaucoma," https://www.vidal.fr/maladies/yeux/glaucome/diagnostic.html, 2021, updated 5 February 2021, accessed 29 March 2026.

[61] Apollo Hospitals, "Fluorescein angiography," https://www.apollohospitals.com/fr/diagnostics-investigations/fluorescein-angiography, accessed 30 March 2026.

[62] RxList, "Medical definition of fundus, retinal," RxList, 2021, accessed: 2026-05-24. [Online]. Available: https://www.rxlist.com/fundus_retinal/definition.htm

[63] H. Schneiderman and J. D. Sapira, *The Funduscopic Examination*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds. Butterworths, 1990, accessed: 2026-05-24. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK221/

[64] Mater Private Network, "Fundus fluorescein angiogram," Mater Private Network Eye Care Services, accessed: 2026-05-24. [Online]. Available: https://www.materprivate.ie/our-services/eye-care/eye-care-dublin/fundus-fluorescein-angiogram

[65] ScienceDirect, "Fundus fluorescein angiography," ScienceDirect Topics, compounds information from multiple peer-reviewed sources. Accessed: 2026-05-24. [Online]. Available: https://www.sciencedirect.com/topics/medicine-and-dentistry/fundus-fluorescein-angiography

[66] S. Ruia and K. Tripathy, "Fluorescein angiography," *StatPearls [Internet]*, 2023, last update: August 25, 2023. Accessed: 2026-05-24. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK576378/

[67] A. Farkhooy, P. J. Rosenfeld, G. Gregori, J. P. Campbell, and M. F. Chiang, "Anatomy of the human eye," *Journal of Ophthalmology*, vol. 2020, p. 5801870, 2020, accessed: Mar. 31, 2026. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2020/5801870

[68] M. López-Venegas, J. García-López *et al.*, "Breve historia de la anatomía ocular," *Revista Mexicana de Oftalmología*, vol. 95, no. 2, pp. 105–112, 2021, accessed: Mar. 31, 2026. [Online]. Available: https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-95322021000200105

[69] X. Zhang, G. Ji *et al.*, "Application of artificial intelligence in diabetic retinopathy screenings," *Frontiers in Medicine*, vol. 9, p. 808402, 2022, accessed: Mar. 31, 2026. [Online]. Available: https://www.frontiersin.org/journals/medicine/articles/10.3389/fmed.2022.808402/full

[70] Z. Kondraki *et al.*, "Advanced imaging and anatomical considerations of the human eye," *PubMed Central (PMC)*, 2023, accessed: Mar. 31, 2026. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC10607340/

[71] S. Mishra *et al.*, "Deep learning for retinal image analysis," *PLOS ONE*, vol. 19, no. 7, p. e0307317, 2024, accessed: Mar. 31, 2026. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0307317

[72] X. Wang, Y. Zhang *et al.*, "Recent developments in deep learning for retinal vessel segmentation," *arXiv preprint arXiv:2412.09402v2*, 2024, accessed: Mar. 31, 2026. [Online]. Available: https://arxiv.org/html/2412.09402v2

[73] J. Cao-Xue, T. Comlekoglu, K. Xue, G. Wang, J. Li, and G. Laurie, "Automated multi-label classification of eleven retinal diseases: A benchmark of modern architectures and a meta-ensemble on a large synthetic dataset," *arXiv preprint arXiv:2508.15986*, 2025, accessed: Mar. 31, 2026. [Online]. Available: https://arxiv.org/abs/2508.15986

[74] Z. Al-Qudah *et al.*, "Advanced deep learning architectures for retinal vessel segmentation," *Scientific Reports (via PubMed Central)*, vol. 15, no. 1, p. 4892, 2025, accessed: Mar. 31, 2026. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11829032/

[75] A. Jithesh, C. Karumuri, V. K. R. Kotha, M. Doddapuneni, and T. Jeong, "Diabetic retinopathy lesion segmentation through attention mechanisms," *arXiv preprint arXiv:2602.07301*, 2026, accessed: Mar. 31, 2026. [Online]. Available: https://arxiv.org/abs/2602.07301

[76] M. Islam *et al.*, "Ocular disease recognition using cnn," *Some Journal*, 2020, early baseline; struggles with minority classes.

[77] L. Wang *et al.*, "Dual efficientnet for odir-5k," *Some Journal*, 2021, reported as weighted F1; AUC lower than peers.

[78] T. Li *et al.*, "Ddr dataset," *Information Sciences*, vol. 501, pp. 511–522, 2019, original DDR paper; detection treated as baseline only.

[79] C. Santos *et al.*, "Yolov5 for fundus lesion detection," *Computers*, 2022, pMC9460625; Val set; F1 = 0.35; PMC open access.

[80] P. Coelho *et al.*, "Yolov8 and yolov9 for dr lesion segmentation," *MDPI Vision*, 2024, pMC11417923; YOLOv8-mono (0.339 mAP) and YOLOv9-extended (0.359 mAP) results.

[81] N. Li, T. Li, C. Hu, K. Wang, and H. Kang, "A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection," in *Benchmarking, Measuring, and Optimizing (Bench 2020)*, ser. Lecture Notes in Computer Science, vol. 12614. Springer, 2020, pp. 177–193. [Online]. Available: https://arxiv.org/abs/2102.07978

[82] P. C and J. K. R, "Retinal image enhancement based on color dominance of image," *Scientific Reports*, vol. 13, p. 7172, 2023, open access. Accessed: 2026-05-14. [Online]. Available: https://www.nature.com/articles/s41598-023-34212-w

[83] NVIDIA Corporation, "What is pytorch?" NVIDIA Glossary, 2026, accessed: 2026-05-07. [Online]. Available: https://www.nvidia.com/en-us/glossary/pytorch/

[84] A. Sharma *et al.*, "Gradcam – enhancing neural network interpretability in the realm of explainable ai," LearnOpenCV, 2023, accessed: 2026-05-07. [Online]. Available: https://learnopencv.com/intro-to-gradcam/

[85] Pillow Contributors, "Pillow (pil fork) documentation," Pillow Read the Docs, 2025, accessed: 2026-05-07. [Online]. Available: https://pillow.readthedocs.io/en/stable/

[86] K. Hynkova, "Ultimate guide to torchvision library in python," Deepnote Blog, 2024, updated: 2025-08-22. Accessed: 2026-05-07. [Online]. Available: https://deepnote.com/blog/ultimate-guide-to-torchvision-library-in-python

[87] OpenCV team, "Opencv: Open source computer vision library," OpenCV Documentation, accessed: 2026-05-07. [Online]. Available: https://opencv.org/

[88] NumPy Developers, "Numpy: The fundamental package for scientific computing with python," NumPy Official Website, 2025, accessed: 2026-05-07. [Online]. Available: https://numpy.org/

[89] GeeksforGeeks, “Introduction to matplotlib,” GeeksforGeeks, 2026, last updated: 9 Apr 2026. Accessed: 2026-05-08. [Online]. Available: https://www.geeksforgeeks.org/python/python-introduction-matplotlib/

[90] M. S. BENMOHAMED and A. MEKKAOUI, “Agritechly part 2: Ai innovations in veterinary diagnosis and pest management,” Master’s thesis, University of Saida Dr. Moulay Tahar, Saida, Algeria, June 2024, promotion 2023-2024. Supervised by Dr. Hadj Ahmed BOUARARA.

[91] Innovatiana, “Roc & auc curve,” Innovatiana Glossary, 2025, accessed: 2026-05-10. [Online]. Available: https://www.innovatiana.com/es/glossary/roc-auc-curve

[92] H. Volk-Jesussek, *ROC Curve: A Beginner’s Guide*, Numiqo.com, Mar. 2026, statistics made easy (8th revised edition). [Online]. Available: https://numiqo.com/tutorial/roc-curve

[93] Iguazio, “What is the true positive rate in machine learning?” Iguazio Glossary, accessed: 2026-05-29. [Online]. Available: https://www.iguazio.com/glossary/true-positive-rate/

[94] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, chapter 8: Optimization for Training Deep Models. [Online]. Available: http://www.deeplearningbook.org/

[95] A. Vina, “What is f1 score? a computer vision guide,” Roboflow Blog, 2024, accessed: 2026-05-08. [Online]. Available: https://blog.roboflow.com/f1-score/

[96] J. I. Barrios, “La matriz de confusión y sus métricas,” Inteligencia Artificial en Salud, 2019, updated: December 2025. Accessed: 2026-05-10. [Online]. Available: https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/

[97] Nicolas, “Precisión media (map o map) para optimizar y evaluar sus modelos de ia,” Innovatiana Blog, 2024, accessed: 2026-05-10. [Online]. Available: https://www.innovatiana.com/es/post/mean-average-precision

ملخص

تقترح هذه الأطروحة نهجًا متعدد المهام للتعلم العميق للتشخيص الآلي لأمراض الشبكية من صور قاع العين. يعتمد النظام على خط أنابيب مشروط من مرحلتين يجمع بين EfficienNet-B3 -للتصنيف متعدد العلامات لأمراض العين، و YOLO-V8 للكشف عن آفات اعتلال الشبكية السكري. يقوم نموذج التصنيف الذي تم تدريبه على مجموعة بيانات ODIR-5K ، بتنشيط خط أنابيب المعالجة المسبقة الذي يدمج CLAHE و Grad-CAM و SAHI بشكل مشروط عند اكتشاف اعتلال الشبكية السكري قبل إجراء اكتشاف الآفة على مجموعة بيانات DDR.

وتظهر النتائج التجريبية تحسنًا كبيرًا في الأداء مقارنة بالطرق الحالية. كما أن دمج التصورات القابلة للتفسير المستندة إلى Grad-CAM يعزز أيضًا دعم القرار السريري. وأخيرا يسلط هذا البحث الضوء على قدرة تعميم البنية خفيفة الوزن والمحسنة مما يجعلها مناسبة للنشر في البيئات الطبية ذات الموارد المحدودة.

الكلمات الرئيسية : التعلم العميق، التعلم متعدد المهام، تشخيص أمراض الشبكية، اعتلال شبكية السكري، EfficienNet-B3 YOLO-V8 اكتشاف الآفات التعميم.

Abstract

This thesis proposes a multi-task deep learning approach for the automated diagnosis of retinal diseases from fundus images. The system is based on a two-stage conditional pipeline combining efficientNet-B3 for the multi-label classification of ocular pathologies and YOLOv8s for the detection of diabetic retinopathy lesions. The classification model, trained on the ODIR-5K dataset, conditionally activates a preprocessing pipeline integrating CLAHE, Grad-CAM, and SAHI when diabetic retinopathy is detected, before performing lesion detection on the DDR dataset. Experimental results demonstrate a significant improvement in performance compared to existing approaches. The integration of interpretable Grad-CAM-based visualizations also enhances clinical decision support. Finally, this research highlights the generalization capability of a lightweight and optimized architecture, making it suitable for deployment in resource-limited medical environments.

Keywords: Deep Learning, Multi-Task Learning, Retinal Disease Diagnosis, Diabetic Retinopathy, EfficientNet-B3, YOLOv8s, Lesion Detection, Generalization.

Résumé

Cette thèse propose une approche d'apprentissage profond multi-tâches pour le diagnostic automatisé des maladies rétiniennes à partir d'images du fond de l'œil. Le système est basé sur un pipeline conditionnel en deux étapes combinant EfficientNet-B3 pour la classification multi-label des pathologies oculaires et YOLOv8s pour la détection des lésions de rétinopathie diabétique. Le modèle de classification, entraîné sur l'ensemble de données ODIR-5K, active conditionnellement un pipeline de prétraitement intégrant CLAHE, Grad-CAM et SAHI lorsque une rétinopathie diabétique est détectée, avant d'effectuer la détection des lésions sur l'ensemble de données DDR.Les résultats expérimentaux démontrent une amélioration significative des performances par rapport aux approches existantes. L'intégration de visualisations interprétables basées sur la Grad-CAM améliore également l'aide à la décision clinique. Enfin, cette recherche met en évidence la capacité de généralisation d'une architecture légère et optimisée, ce qui la rend adaptée au déploiement dans des environnements médicaux à ressources limitées.

Mots-clés : Apprentissage profond, apprentissage multi-tâches, diagnostic des maladies rétiniennes, rétinopathie diabétique, EfficientNet-B3, YOLOv8s, détection de lésions, Généralisation.