Université Mohamed Khider de Biskra
Faculté des Sciences et de la Technologie
Département de génie électrique

# MÉMOIRE DE MASTER

Sciences et Technologies
Filière d'Automatique
Spécialité d'Automatique et Informatique Industrielle

Réf. : Entrez la référence du document

---

Présenté et soutenu par :

Bouzahar Hadil et Benchabane Imane

Le : Mercredi 04 juin 2025

## Classification automatisée de la rétinopathie diabétique basée sur l'apprentissage profond

---

**Jury :**

| | | | | |
|---|---|---|---|---|
| Dr | M. Rezig | MCA | Univ Biskra | Président |
| Dr | A. Messaoudi | MCA | Univ Biskra | Rapporteur |
| Dr | D-E. Boukhari | Dr | CRSTRA | Co-Encadrant |
| Pr. | A. Ouafi | Pr | Univ Biskra | Examinateur |

Année universitaire : 2024/2025

Université Mohamed Khider de Biskra
Faculté des Sciences et de la Technologie
Département de génie électrique

# MÉMOIRE DE MASTER

Sciences et Technologies
Filière d'Automatique
Spécialité d'Automatique et Informatique Industrielle

Réf. : Entrez la référence du document

---

Classification automatisée de la rétinopathie diabétique basée sur l'apprentissage profond

Le : Mercredi 04 Juin 2025

**Présenté par :** Bouzahar Hadil et Benchabane Imane

**Avis favorable de l'encadreur :**

**Signature Avis favorable du Président du Jury**

**Cachet et signature**

# Résumé

La rétinopathie diabétique (RD) est l'une des principales causes de perte de vision chez les adultes en âge de travailler à travers le monde. Une détection précoce et une classification précise des stades de la RD sont essentielles pour prévenir la cécité. Cependant, le diagnostic manuel réalisé par les ophtalmologues est long, coûteux et sujet à une variabilité inter-observateur. Dans ce contexte, l'apprentissage profond s'impose comme un outil prometteur pour automatiser le dépistage avec efficacité.

Ce mémoire propose une approche de classification automatique de la rétinopathie diabétique basée sur des modèles ensemblistes de deep learning. Deux architectures hybrides ont été développées et comparées : l'une combinant EfficientNet-B3 et DenseNet-169, et l'autre fusionnant EfficientNet-B3 avec un Vision Transformer à base de pooling (PiT-Base). Ces modèles exploitent les avantages des réseaux convolutifs et des Transformers pour améliorer l'extraction des caractéristiques visuelles.

Les expérimentations menées sur le jeu de données APTOS 2019 montrent que les modèles proposés atteignent des performances élevées, surpassant plusieurs méthodes de la littérature récente. Cette étude met également en évidence l'importance du prétraitement des images, de la gestion du déséquilibre des classes, ainsi que du choix rigoureux des métriques d'évaluation telles que l'exactitude, le F1-score et le score Kappa.

Ce travail ouvre des perspectives vers la conception de systèmes de dépistage assistés par intelligence artificielle, robustes, interprétables et intégrables dans des environnements cliniques réels.



***Mots-clés :*** La rétinopathie diabétique*,* Vision Transformer*,* EfficientNet-B3, DenseNet- 69, CNN.

# Abstract

Diabetic retinopathy (DR) is a leading cause of vision loss among working-age adults worldwide. Early detection and classification of DR stages are crucial for timely intervention and prevention of blindness. However, manual diagnosis by ophthalmologists is time-consuming and subject to variability. In this context, deep learning has emerged as a powerful tool to support automated and accurate DR classification.

This thesis presents an ensemble-based deep learning approach for the automated classification of diabetic retinopathy severity from retinal fundus images. Two ensemble strategies were developed and compared: one combining EfficientNet-B3 and DenseNet-169, and the other integrating EfficientNet-B3 with a Pooling-based Vision Transformer (PiT-Base). These hybrid architectures leverage both convolutional and attention-based mechanisms to enhance feature extraction and improve classification performance.

Extensive experiments conducted on the APTOS 2019 dataset demonstrate that the proposed models achieve high accuracy and outperform several recent methods from the literature. The study also highlights the importance of data preprocessing, class imbalance management, and the use of appropriate evaluation metrics such as accuracy, F1-score, and Cohen's kappa.

This work paves the way for more interpretable and robust AI-assisted screening tools, with promising applications in real-world clinical decision support systems.



***Index Terms***—Diabetic retinopathy, Deep learning, Vision Transformer, EfficientNet-B3, DenseNet-169**,** CNN.

## الملخص

تُعدّ اعتلالات الشبكية الناتجة عن داء السكري (RD) من الأسباب الرئيسية لفقدان البصر لدى البالغين في سنّ العمل على مستوى العالم. إنّ الكشف المبكر والتصنيف الدقيق لمراحل هذا المرض أمران ضروريان للوقاية من العمى. ومع ذلك، فإنّ التشخيص اليدوي الذي يُجريه أطباء العيون يستغرق وقتًا طويلاً، ويُعدّ مكلفًا ومعرّضًا لتفاوت في التقدير بين الأطباء. في هذا السياق، يبرز التعلم العميق كأداة واعدة لأتمتة عملية الكشف بكفاءة.

يقترح هذا المذكرة مقاربة لتصنيف اعتلال الشبكية السكري بشكل تلقائي بالاعتماد على نماذج Ensembles من التعلم العميق. تم تطوير ومقارنة بنيتين هجينتين: الأولى تجمع بين EfficientNet-B3 وDenseNet-169، والثانية تدمج EfficientNet-B3مع Transformer بصري يعتمد على التجميع .(PiT-Base) تستفيد هذه النماذج من مزايا الشبكات الالتفافية و Transformersلتحسين عملية استخراج الخصائص البصرية.

تُظهر التجارب التي أُجريت على مجموعة بيانات APTOS 2019 أن النماذج المقترحة تحقق أداءً عاليًا، متفوّقةً على العديد من الأساليب الحديثة في الأدبيات. كما تُبرز هذه الدراسة أهمية المعالجة المسبقة للصور، والتعامل مع عدم توازن الفئات، بالإضافة إلى ضرورة اختيار دقيق لمقاييس التقييم مثل الدقة، وF1-score، ومعامل كابا (Kappa) .

يفتح هذا العمل آفاقًا نحو تصميم أنظمة كشف مدعومة بالذكاء الاصطناعي تكون قوية، قابلة للتفسير، وقابلة للدمج في البيئات السريرية الواقعية.



**الكلمات المفتاحية :**اعتلال الشبكية السكري، Vision Transformer، EfficientNet-B3، DenseNet- 169، الشبكات العصبية الالتفافية (CNN) .

# Remerciements

Nous tenons à remercier **Dr Messaoudi Abdelhamid** d'avoir dirigé nos travaux de thèse et de son soutien durant tous nos années de thèse.

Nous remercions tout particulièrement Dr. **Boukhari Djamel Eddine**, Chercheur dans CRSTRA, de nous avoir accueilli dans son équipe, et de son soutien moral au cours de nos travaux de thèse.

Nous remercions Pr. **Ouafi Abdelkrim** pour son aide précieuse et d'avoir accepté de juger ce travail.

Nous remercions Dr. **Rezig Mohamed** qui me fait l'honneur de présider ce jury.

Nous remercions tout particulièrement Mr. **Zidi Fadi Abdeladhim**, de son soutien moral au cours de nos travaux de thèse.

Nous tenons également à remercier nos collègues du département d'automatique et tous nos amis.

Nous tenons à remercier du fond du cœur tous les enseignants qui nous ont accompagnés, soutenus et transmis leur savoir tout au long de notre parcours universitaire. Grâce à votre engagement, votre patience et votre passion pour l'enseignement, nous avons pu grandir, apprendre et évoluer jusqu'à ce stade important de notre vie.

Merci pour votre dévouement et pour avoir cru en nous, même quand nous doutions.

*Dédicace*

*À ma chère maman,*

*Pour ton amour inépuisable, ta patience dans mes moments de doute et ta manière unique de calmer mes inquiétudes. Ta présence a été mon refuge tout au long de ce parcours.*

*À mon cher papa,*

*Merci pour ta présence discrète mais rassurante, ton regard bienveillant et tes conseils toujours justes au bon moment.*

*À ma sœur et à mon frère,*

*Merci pour vos attentions, votre humour spontané et votre capacité à alléger les journées difficiles. Vous avez été ma source d'énergie et de réconfort.*

*À ma petite nièce Layane,*

*Ton innocence et tes sourires ont apporté douceur et lumière à mes journées. Tu es une joie à part entière.*

*À mes chères amies Yamina et Mounira,*

*Pour votre amitié sincère, vos encouragements constants et vos messages qui arrivaient toujours au bon moment.*

*À Imane, ma binôme,*
*Merci pour ta collaboration sincère, ton sérieux, ton soutien constant et ta bonne humeur. Nous avons traversé cette aventure ensemble, et ton implication a été précieuse du début à la fin.*
*À mes deux petits compagnons à quatre pattes, Bissou et Léo, Pour leur tendresse, leur présence apaisante et les instants de bonheur qu'ils m'ont offerts.*
*À moi-même,*
*Pour avoir persévéré malgré les obstacles, pour être restée fidèle à mes objectifs, et pour ne pas avoir baissé les bras. Cette réussite, je me la dois aussi. Merci à celle que je suis devenue.*

*Hadil*

*Dédicace*

*À ma chère famille,*

*À mes parents,*

*Pour votre amour inconditionnel, vos sacrifices et votre soutien indéfectible et vos sacrifices qui m'ont permis d'arriver jusqu'ici. Vous êtes ma plus grande source de motivation.*

*À mes sœurs et à mon frère,*
*Merci pour votre présence, votre affection et vos encouragements constants.*

*À mes chères copines Rania, Maissa, Malak et Mounira*

*Merci pour votre amitié sincère, votre présence précieuse, vos encouragements et votre bienveillance constante tout au long de ce parcours.*

*À ma binôme Hadil,*

*Merci pour ton engagement, ta persévérance et ta collaboration précieuse tout au long de ce travail. Ce projet n'aurait pas été le même sans toi. Ta rigueur, ton énergie et ton esprit d'équipe ont grandement contribué à cette réussite.*

*Merci à vous tous d'avoir été ma force, mon refuge et ma motivation.*

*Cette thèse vous est dédiée, avec tout mon amour et ma gratitude.*

*À moi-même.*

*Pour chaque nuit blanche, chaque moment de doute, chaque page réécrite.*

*Pour la persévérance face aux obstacles, et la foi en mes capacités lorsque les choses semblaient impossibles.*

*Ce travail est la preuve que la discipline, la patience et la passion mènent toujours quelque part.*

*Je me rends hommage pour n'avoir jamais abandonné.*

*Imane*

# Table Des matières

# Liste des figures

## Chapitre 1 : Etat de l'art des méthodes de la rétinopathie diabétique



## Chapitre 2 : Les techniques d'apprentissage appliquées



## Chapitre 3: Résultats et discussions

# Liste des tables

## Chapitre 1 : Etat de l'art des méthodes de la rétinopathie diabétique



## Chapitre 3 : Résultats et Discussions

# Introduction générale

La rétinopathie diabétique (RD) représente l'une des complications les plus graves et les plus fréquentes du diabète sucré, pouvant mener à une perte irréversible de la vision si elle n'est pas diagnostiquée et traitée à temps. Avec l'augmentation mondiale du nombre de personnes atteintes de diabète, notamment dans les pays à revenu faible et intermédiaire, le besoin de solutions de dépistage automatisées, fiables et accessibles devient de plus en plus pressant. Face à la pénurie de spécialistes en ophtalmologie dans certaines régions et à la difficulté d'accès aux soins, les systèmes d'aide au diagnostic basés sur l'intelligence artificielle (IA) offrent une alternative prometteuse pour alléger la charge clinique et améliorer la prévention des formes graves de RD.

Parmi les avancées technologiques récentes, les techniques d'apprentissage profond, en particulier les réseaux de neurones convolutifs (CNN), ont démontré une efficacité remarquable dans l'analyse d'images médicales. En parallèle, les modèles basés sur l'attention tels que les Vision Transformers (ViT) ont élargi les perspectives en capturant des relations globales dans les images, ce qui est particulièrement pertinent dans le contexte d'une pathologie multifocale comme la RD. Toutefois, malgré ces avancées, de nombreux défis persistent : déséquilibre des classes, complexité des lésions rétiniennes, qualité variable des images, ou encore manque d'interprétabilité des modèles.

C'est dans ce contexte que s'inscrit ce mémoire, qui vise à proposer une méthode de classification automatique de la rétinopathie diabétique reposant sur des architectures ensemblistes de modèles profonds pré-entraînés. Deux configurations ont été explorées : une première combinant DenseNet-169 et EfficientNet-B3, exploitant la complémentarité des CNN traditionnels, et une seconde intégrante EfficientNet-B3 avec un Transformer visuel PiT-Base, tirant parti de la fusion entre convolution locale et attention globale. L'étude s'appuie sur la base de données APTOS 2019, reconnu pour sa diversité et ses annotations cliniques précises.

Ce travail s'inscrit dans une démarche méthodologique rigoureuse, intégrant un prétraitement adapté aux images médicales, une gestion du déséquilibre des classes, une formation optimisée (précision mixte, callbacks personnalisés), et une évaluation multicritère (accuracy, F1-score, Kappa, matrice de confusion). En analysant et comparant les résultats

obtenus avec ceux de l'état de l'art, ce mémoire entend démontrer la pertinence des approches hybrides pour améliorer le diagnostic automatisé des maladies ophtalmologiques.

Cette thèse est structurée comme suit :

Chapitre 1 : Revue de la littérature sur la rétinopathie diabétique, les approches classiques de classification et les avancées récentes en deep learning.

Chapitre 2 : Présentation de la méthodologie utilisée, incluant la description des modèles et des techniques d'apprentissage appliquées.

Chapitre 3 : Expérimentation et analyse des résultats obtenus. Discussion et comparaison avec d'autres approches existantes.

Nous terminons cette thèse par une conclusion générale et les perspectives envisagées concernant notre travail.

# **Chapitre 1 :** Etat de l'art des méthodes de la rétinopathie diabétique

**Résumé**

Nous nous limiterons, dans ce chapitre à la présentation des principales méthodes de la rétinopathie diabétique proposées dans la littérature. On commence par décrire les caractéristiques principales d'une rétinopathie diabétique. Ensuite un bref rappel sur les méthodes de la rétinopathie diabétique les plus utilisées est souligné. La grande partie de ce chapitre sera consacrée aux méthodes de la rétinopathie diabétique basées sur l'apprentissage profond, plus spécifiquement les réseaux de neurones convolutifs.

# 1 Introduction

L'expression état de l'art désigne l'ensemble des connaissances, méthodes, techniques et outils existants à un moment donné dans un domaine de recherche ou d'application. Dans le cadre de ce mémoire, il s'agit de faire le point sur les approches développées pour la détection et la classification de la rétinopathie diabétique (RD), en s'appuyant à la fois sur les techniques classiques de traitement d'image et sur les avancées récentes de l'intelligence artificielle, notamment le deep learning.

La rétinopathie diabétique est une complication grave du diabète sucré qui touche la rétine et peut entraîner une perte irréversible de la vision si elle n'est pas détectée à temps. Face à l'augmentation constante du nombre de personnes atteintes de diabète à l'échelle mondiale, le besoin de systèmes de dépistage automatiques, précis et rapides devient crucial pour alléger la charge des spécialistes, surtout dans les régions où les ophtalmologistes sont peu nombreux [1].

Analyser l'état de l'art permet non seulement de mieux comprendre les fondements des approches existantes, mais aussi d'identifier leurs limites, leurs besoins en données, leur complexité computationnelle, ainsi que leur niveau de maturité clinique. Cela permet ensuite de dégager des pistes pour améliorer ou proposer de nouvelles méthodes plus adaptées aux exigences actuelles en termes de précision, d'efficacité et d'interprétabilité.

Ainsi, ce chapitre se propose d'examiner successivement :

- L'anatomie de l'œil et les manifestations de la RD,
- Les techniques d'imagerie utilisées,
- Les approches classiques de traitement d'image,
- Les techniques modernes basées sur le deep learning,

Cette analyse comparative posera les bases nécessaires à la compréhension des choix méthodologiques retenus dans la suite du mémoire.

# 2 Rétinopathie diabétique

## 2.1 Définition et contexte médical

La rétinopathie diabétique (RD) est une complication microvasculaire du diabète sucré, qui affecte la rétine — la membrane sensible à la lumière située au fond de l'œil. Elle résulte de dommages progressifs causés aux petits vaisseaux sanguins rétiniens à la suite d'une hyperglycémie chronique. Avec le temps, ces dommages peuvent entraîner des fuites vasculaires, des hémorragies, un œdème maculaire et la prolifération anormale de nouveaux vaisseaux (néovascularisation), menaçant gravement la vision du patient [2].

La RD représente aujourd'hui l'une des principales causes de cécité évitable chez les adultes en âge de travailler. Selon l'Organisation mondiale de la santé (OMS), des millions de personnes dans le monde souffrent de formes modérées à sévères de cette affection. La détection précoce est donc un enjeu de santé publique majeur, surtout dans les pays où l'accès à des soins ophtalmologiques spécialisés est limité.

## A. Anatomie de l'œil et de la rétine

Comprendre l'impact de la RD nécessite une brève révision de l'anatomie de l'œil :

- La rétine est constituée de plusieurs couches de cellules, dont les photorécepteurs (cônes et bâtonnets) et les cellules ganglionnaires.
- La macula, au centre de la rétine, est responsable de la vision centrale fine.
- La fovéa, située au centre de la macula, offre la plus grande acuité visuelle.
- La vascularisation rétinienne est assurée par des capillaires fragiles, particulièrement vulnérables aux effets du diabète.

### L'anatomie du l'œil

L'œil est l'organe qui permet la vision. La lumière le traverse en franchissant la cornée, première lentille transparente, puis la pupille, couronnée de l'iris, et le cristallin. Elle passe ensuite le vitré, pour terminer sa course sur la rétine, souvent comparée à une pellicule photographique dont le centre se nomme la macula. La lumière y est transformée en signal électrique, message envoyé au cerveau par l'intermédiaire du nerf optique, où l'image sera interprétée. L'œil se situe dans l'orbite. Il est protégé par les paupières et la conjonctive [3].

Le globe oculaire se compose de trois couches distinctes. De la plus superficielle à la plus profonde, on retrouve :

La couche fibreuse est constituée de la sclère et de la cornée. La sclère est une couche opaque qui entoure les cinq sixièmes postérieurs du globe oculaire. La cornée est une couche transparente qui est antérieurement continue avec la sclère, occupant le sixième antérieur de l'œil [4].

La couche vasculaire, également connue sous le nom d'uvée ou tractus uvéal. Elle est composée de trois parties qui sont continues entre elles. De l'arrière à l'avant, ce sont la choroïde, le corps ciliaire et l'iris [4].

La couche nerveuse, également connue sous le nom de rétine. C'est la couche la plus interne du globe oculaire. La rétine elle-même est divisée en deux couches : une couche externe pigmentée et une couche neurosensorielle interne.

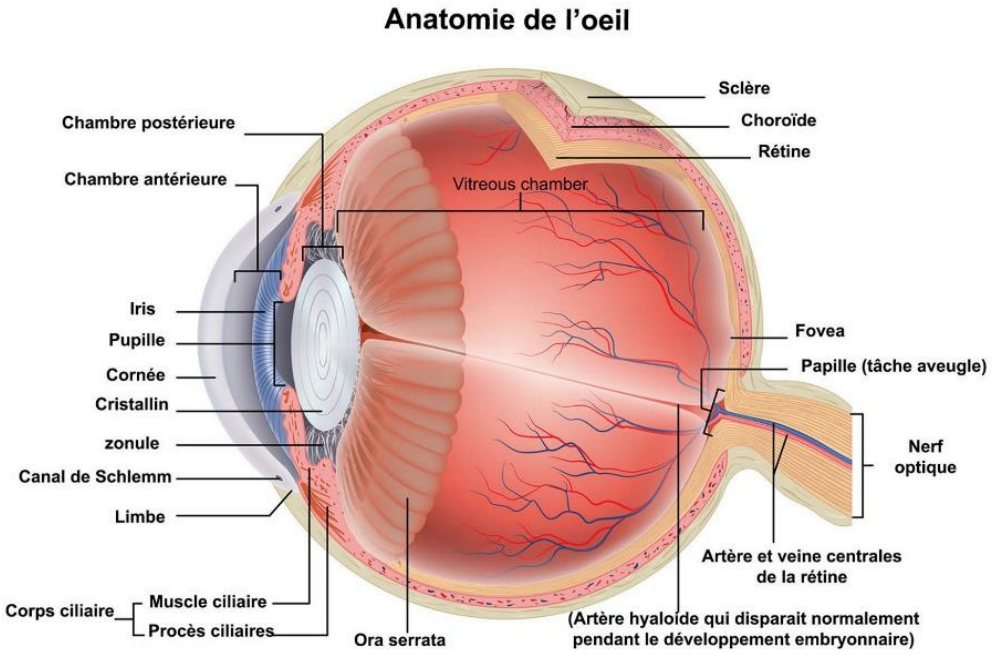


Figure 1. 1. Anatomie de l'œil [5].

## La rétine

La rétine est une fine membrane sensorielle essentielle, de couleur rosée, transparente et bien vascularisée, qui tapisse l'intérieur du globe oculaire depuis la papille optique jusqu'à l'ora-serrata [6]. Elle joue un rôle crucial dans la transformation des stimuli lumineux en signaux nerveux, permettant ainsi la vision. En contact avec l'humeur vitrée à l'avant et la choroïde à l'arrière, elle est composée de plusieurs couches de cellules, notamment les photorécepteurs (cônes et bâtonnets), qui captent la lumière et transmettent l'information au cerveau via le nerf optique. Une compréhension approfondie de sa structure et de son fonctionnement est essentielle pour appréhender les mécanismes de la vision et les pathologies associées.

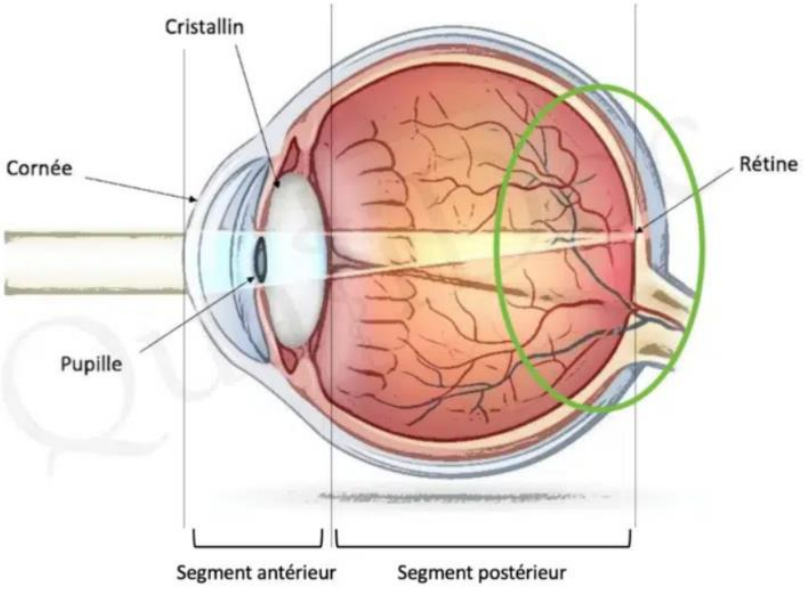


Figure 1.2. La rétine tapisse le fond d'œil [6].

Il existe trois zones particulières :

**La macula :** zone centrale de la rétine. la fovéa : dépression centrale de la macula caractérisée par une densité importante de cônes où l'acuité visuelle est à son maximum. la papille optique : zone d'émergence du nerf optique dépourvue de photorécepteurs.

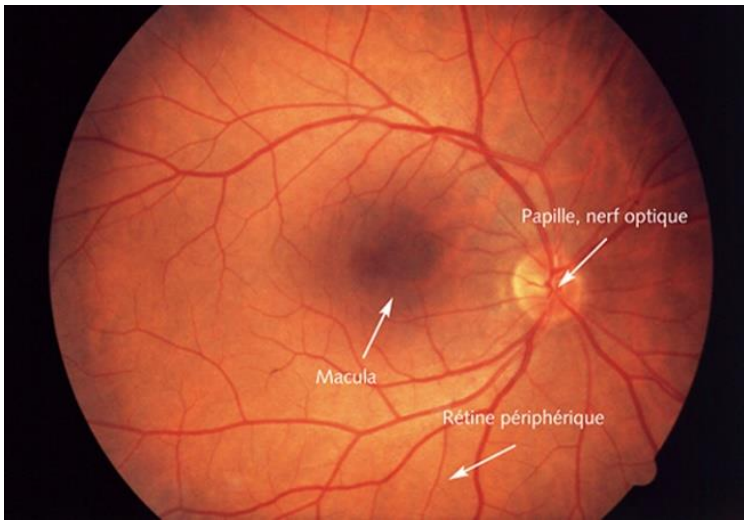


Figure 1.3. La macula de l'œil normal [7].

La rétine est constituée de deux tissus :

**La couche neurosensorielle :** couche composée de cônes et bâtonnets : photorécepteurs qui captent les signaux lumineux et les transforment en signaux électrochimiques [8].

**L'épithélium pigmentaire qui a quatre grands rôles :** un rôle d'écran, un rôle d'échanges, un rôle dans le métabolisme de la vitamine A et un rôle dans la phagocytose des articles externes des photorécepteurs [9].

D'un point de vue histologique, on distingue 10 couches, qui sont de l'extérieur vers l'intérieur :

- L'épithélium pigmentaire.
- La couche des photorécepteurs comprenant les cônes qui sont responsables de la vision centrale et des couleurs et les bâtonnets qui sont plutôt responsables de la vision périphérique et nocturne.
- La Membrane Limitante Externe.
- La Couche Nucléaire Externe.
- La Couche Plexiforme Externe.
- La Couche Nucléaire Interne.
- La Couche Plexiforme Interne.
- La Couche Des Cellules Ganglionnaires.
- La Couche Des Fibres Optiques.
- La Membrane Limitante Interne.

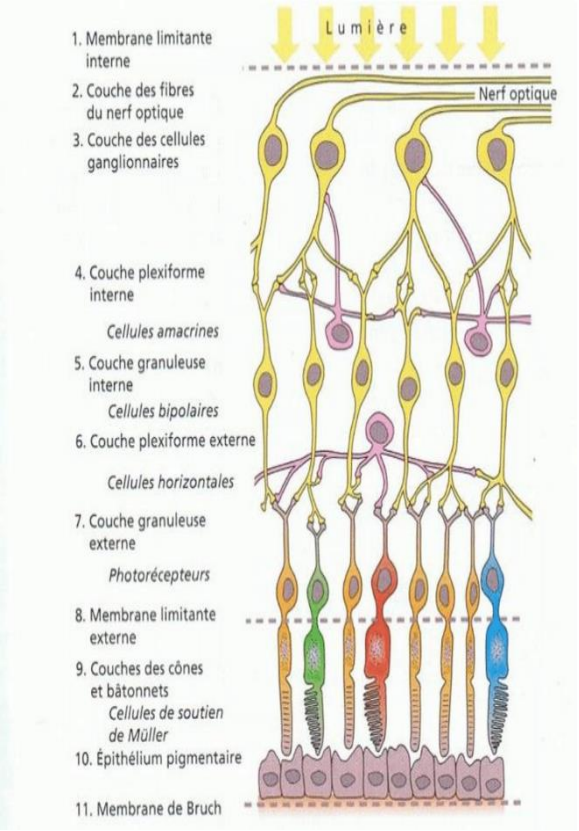


Figure 1.4. L'organisation schématique de la rétine [9].

La rétinopathie diabétique (atteinte des yeux : œil et rétine) est une grave complication du diabète qui touche 50 % des patients diabétiques de type 2 [10]. Elle représente une atteinte du complexe neurovasculaire de la rétine et constitue une microangiopathie rétinienne, précédée d'une atteinte neuronale rétinienne [11].

Les complications de la rétinopathie diabétique sont liées à l'hyperglycémie qui altère les vaisseaux de petit calibre de la rétine, appelés capillaires, entraînant des lésions visibles au fond d'œil. L'hyperglycémie et l'altération des capillaires rétiniens génèrent un manque d'oxygénation du tissu rétinien par raréfaction des petits vaisseaux. De façon réactionnelle, des anomalies apparaissent, allant des micro-anévrismes (dilatation des capillaires) aux hémorragies rétiniennes. Les vaisseaux de la rétine se déforment, puis des vaisseaux anormaux appelés « néovaisseaux » se forment, caractérisant une forme sévère de rétinopathie appelée « proliférante » [12].

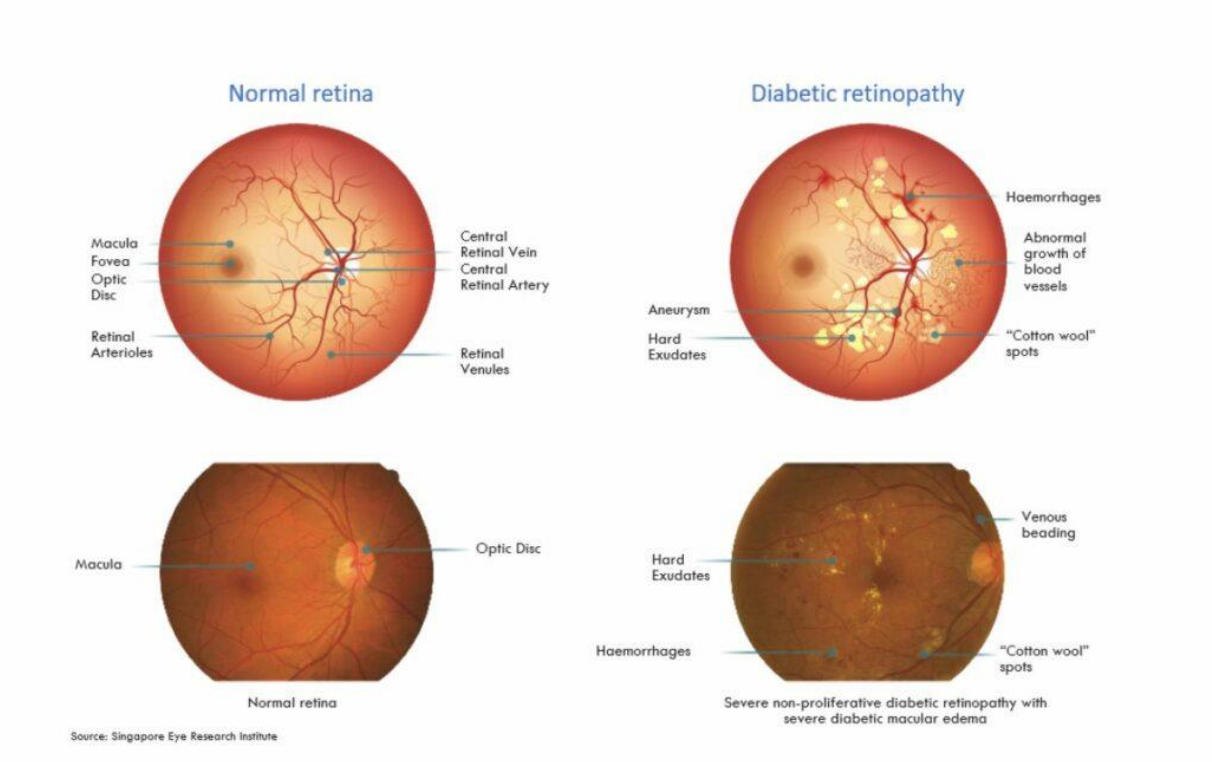


Figure 1.5. Comparaison entre une rétine normale et une rétinopathie diabétique [13].

## B. Manifestations cliniques

Les signes visibles de la rétinopathie diabétique, détectables sur les images du fond d'œil, incluent :

- Microanévrismes : petites poches de sang dues à une dilatation capillaire.
- Hémorragies rétiniennes : fuites sanguines localisées.
- Exsudats durs : dépôts lipidiques brillants.
- Œdème maculaire : accumulation de liquide dans la macula.
- Exsudats cotonneux : zones blanchâtres dues à une ischémie nerveuse.
- Néovascularisation : prolifération anormale de vaisseaux sanguins dans les stades avancés.

## C. Classification des stades de la RD

La classification standard de la RD, notamment selon l'International Clinical Diabetic Retinopathy Disease Severity Scale, repose sur la sévérité des lésions observées :

- Stade 0 : absence de rétinopathie.
- Stade 1 (RD non proliférante légère) : présence de quelques microanévrismes.

- Stade 2 (modérée) : hémorragies et exsudats plus nombreux.
- Stade 3 (sévère) : signes d'occlusion capillaire extensive.
- Stade 4 (RD proliférante) : apparition de néovaisseaux avec risque élevé de cécité.

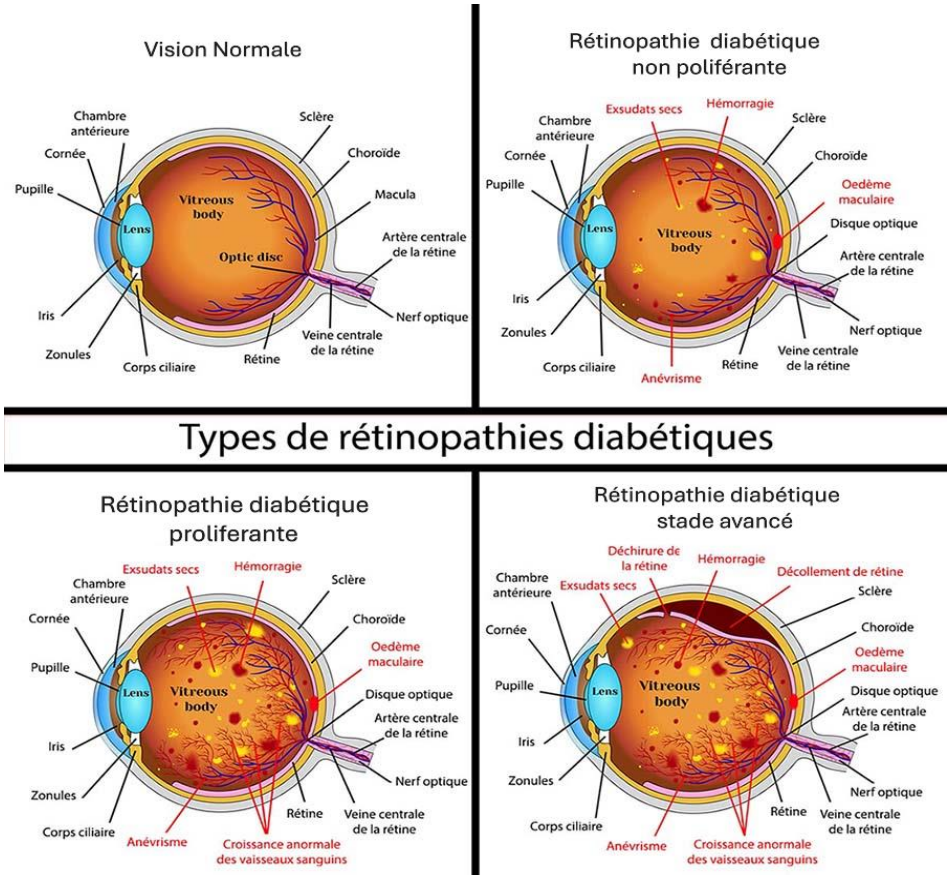


Figure 1.6. Les types de la rétinopathie diabétique [14].

Ce classement est essentiel dans le cadre de l'automatisation de la détection, car il constitue une base pour l'entraînement et l'évaluation des algorithmes.

Tableau 1.1 : Résumé des stades et évolution de la RD.

| Stade | Caractéristiques principales | Risque de progression |
|---|---|---|
| RDNP Légère | Micro-anévrismes | Faible, nécessite une surveillance |
| RDNP Modérée | Hémorragies rétiniennes, nodules cotonneux | Élevé, surtout en présence d'œdème maculaire |
| RDNP Sévère | Occlusions capillaires, ischémie, anomalies veineuses | Très élevé, progression vers RD proliférante |
| RD Proliférante | Néovaisseaux, hémorragies intravitréennes, risque de décollement de rétine | Urgence médicale, risque de cécité permanent |

## 2.2 Techniques classiques de détection et de classification de la rétinopathie diabétique

Avant l'avènement des méthodes basées sur l'apprentissage automatique, la détection de la rétinopathie diabétique reposait essentiellement sur l'expertise humaine et des techniques d'analyse d'image classiques [15]. Ces approches, bien que limitées en précision et en scalabilité, ont jeté les bases des systèmes automatisés modernes. On peut distinguer deux grandes familles de méthodes traditionnelles : les approches basées sur le traitement d'image et les méthodes utilisant des classificateurs standards.

### A. Prétraitement des images

Les images du fond d'œil, obtenues par rétinographie, présentent souvent des problèmes liés à l'éclairage, au contraste ou à la présence de bruit. Le prétraitement vise à corriger ces défauts afin d'améliorer la détection des lésions :

- Correction d'illumination : normalisation de la luminosité pour rendre les images comparables.
- Amélioration du contraste : techniques comme l'égalisation d'histogramme ou CLAHE (Contrast Limited Adaptive Histogram Equalization).
- Filtrage spatial : suppression du bruit par des filtres gaussiens, médians ou passe-bas.

### B. Segmentation des structures rétiniennes

La segmentation permet d'isoler les éléments clés de la rétine, facilitant ainsi l'analyse :

- Détection de la papille optique : généralement la région la plus lumineuse de l'image.
- Segmentation de la macula : cruciale pour évaluer un éventuel œdème maculaire.
- Extraction du réseau vasculaire : à l'aide de filtres morphologiques ou de transformées (par exemple Gabor, Frangi).

### C. Extraction de caractéristiques

Une fois les régions d'intérêt localisées, des descripteurs sont extraits pour caractériser les lésions :

- Caractéristiques de texture (LBP, co-occurrence de niveaux de gris – GLCM).
- Caractéristiques morphologiques : taille, forme, orientation des anomalies.
- Descripteurs de couleur : histogrammes RVB ou HSV pour détecter les exsudats et hémorragies.

### D. Classification basée sur des algorithmes standards

Les caractéristiques extraites sont ensuite utilisées pour entraîner des classificateurs classiques :

- Support Vector Machines (SVM) : efficace pour séparer différentes classes, avec ou sans noyau.
- K-Nearest Neighbors (KNN) : méthode simple reposant sur la distance euclidienne.
- Arbres de décision et forêts aléatoires : capables de gérer des ensembles de données hétérogènes.
- Réseaux de neurones classiques : à couches entièrement connectées (MLP) dans les premières tentatives.

Ces méthodes nécessitent généralement une ingénierie de caractéristiques poussée, ce qui limite leur adaptabilité à de grandes variations de données réelles.

### E. Limites des approches traditionnelles

Bien que relativement efficaces sur des jeux de données bien contrôlés, ces techniques présentent plusieurs limitations :

- Sensibilité aux variations d'éclairage et de qualité d'image.
- Performance fortement dépendante de la qualité des caractéristiques extraites.
- Manque de généralisation face à des images issues de différentes sources (matériel, patients, conditions de prise de vue).
- Temps de traitement élevé en raison de chaînes de traitement complexes.

Ces limites ont favorisé l'émergence de méthodes basées sur le deep learning, capables d'apprendre automatiquement des représentations discriminantes à partir des données brutes, que nous détaillerons dans la section suivante.

## 3 Approches par deep learning

L'essor du deep learning a transformé le domaine de l'analyse d'images médicales, y compris celui de la rétinopathie diabétique [16]. Contrairement aux techniques classiques qui reposent sur l'extraction manuelle de caractéristiques, les approches par deep learning apprennent automatiquement des représentations discriminantes directement à partir des images, rendant les systèmes plus robustes, précis et généralisables.

## 3.1 Réseaux de neurones convolutifs (CNN)

Les réseaux de neurones convolutifs (Convolutional Neural Networks, CNN) ont révolutionné le domaine de l'analyse d'images médicales en fournissant des outils puissants pour l'extraction automatique de caractéristiques discriminantes à partir d'images. Contrairement aux méthodes traditionnelles nécessitant une ingénierie manuelle des caractéristiques, les CNN apprennent à identifier des motifs pertinents directement à partir des données, ce qui les rend particulièrement adaptés à la classification de pathologies visuelles comme la rétinopathie diabétique (RD) [17].

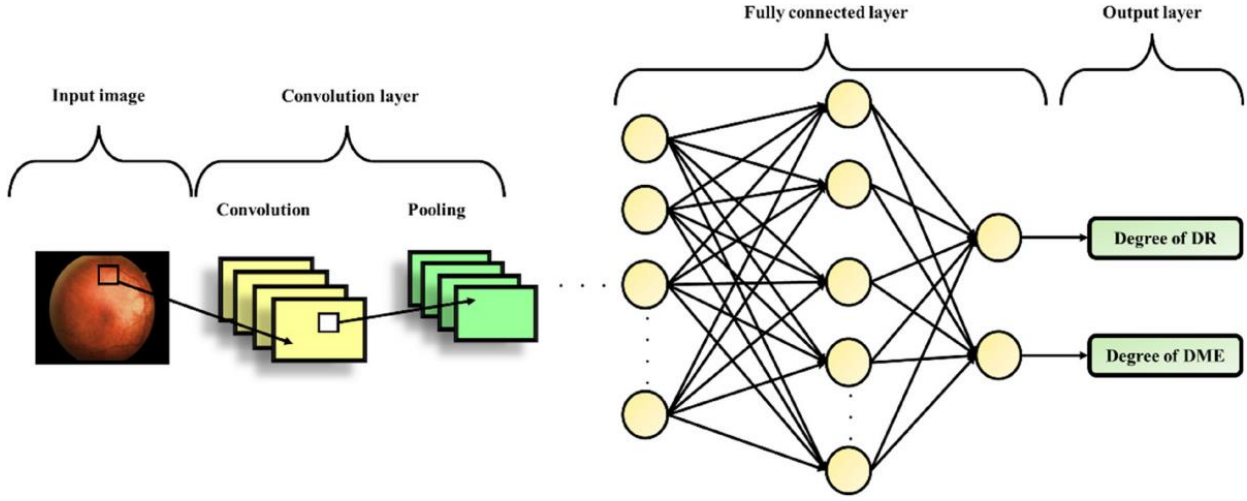


Figure 1.7. Architecture d'un CNN pour la classification.

### A. Principe de fonctionnement

Un CNN est composé de plusieurs types de couches, dont les principales sont :

- **Couches convolutives** : elles appliquent des filtres (ou noyaux) à l'image d'entrée pour extraire des caractéristiques locales (contours, textures, micro-anévrismes, hémorragies, etc.).
- **Couches de pooling** : elles réduisent la dimension spatiale des cartes de caractéristiques, améliorant ainsi l'efficacité computationnelle tout en conservant l'information importante.
- **Couches entièrement connectées (Fully Connected)** : elles interprètent les caractéristiques extraites pour effectuer la classification finale.
- **Fonctions d'activation non-linéaires** (comme ReLU) : elles introduisent de la non-linéarité, permettant au réseau de modéliser des relations complexes.

### B. Application à la Rétinopathie Diabétique

Les CNN sont particulièrement efficaces pour la détection automatique des lésions rétiniennes, notamment [18] :

- Les **micro-anévrismes**, premiers signes de la RD,
- Les **Hémorragies rétiniennes**,
- Les **exsudats**,
- Et les changements vasculaires associés aux stades avancés.

Les modèles CNN peut être entraînés sur des bases de données annotées, comme **EyePACS**, **Messidor** ou **APTOS**, pour apprendre à classer les images selon les différents stades de la RD (de 0 à 4, selon la classification de l'International Clinical Diabetic Retinopathy scale).

### C. Architectures populaires utilisées

Plusieurs architectures CNN ont été explorées dans la littérature [19] :

- **LeNet** : une architecture pionnière, mais trop simple pour des images médicales complexes.
- **AlexNet** et **VGGNet** : introduisent des réseaux plus profonds avec de meilleures performances.
- **ResNet** : introduit des connexions résiduelles permettant d'entraîner des réseaux très profonds sans dégradation des performances.
- **Inception** : utilise des convolutions multi-échelles pour capturer des motifs à différentes résolutions.
- **EfficientNet** : combine efficacité computationnelle et performance en classant les images avec un nombre réduit de paramètres.
- **DenseNet** : efficace pour la propagation de gradient et la réutilisation des caractéristiques.

### D. Performances et limites

Les CNN ont permis d'atteindre des niveaux de précision très élevés dans la classification de la RD, avec des scores d'AUC (Area Under Curve) supérieurs à 0.90 dans plusieurs études.

Cependant, ils présentent encore certaines **limitations** :

- **Besoin de grandes quantités de données annotées**, difficiles à obtenir en imagerie médicale.
- **Sensibilité au déséquilibre de classes** (ex. : peu d'images de RD sévère par rapport à des cas normaux).
- **Manque d'interprétabilité**, ce qui pose des problèmes en contexte clinique.

- **Difficulté de généralisation** à d'autres jeux de données (problème de surapprentissage aux conditions de prise de vue spécifiques).

**E. Vers des CNN améliorés**

Pour pallier ces limites, de nombreuses approches ont été proposées :

- **Data augmentation** pour enrichir le jeu d'entraînement.
- **Techniques de régularisation** (dropout, batch normalization).
- **Transfer learning**, en utilisant des modèles pré-entraînés sur ImageNet.
- **Méthodes d'explicabilité** (Grad-CAM, LIME) pour visualiser les régions les plus influentes dans la décision du modèle.

### 3.2 Apprentissage par transfert (Transfer Learning)

Dans le contexte médical où les bases de données annotées sont souvent limitées, l'apprentissage par transfert s'est révélé particulièrement utile. Il consiste à :

- Réutiliser des modèles préentraînés (ImageNet par exemple).
- Adapter les couches finales pour les tâches spécifiques à la RD.

**Bénéfices :**

- Réduction du temps d'entraînement.
- Meilleure généralisation sur des petits ensembles de données médicales.

### 3.3 Intégration de mécanismes d'attention

Des approches plus récentes incorporent des mécanismes d'attention spatiale ou de canal, permettant au modèle de se concentrer sur les zones critiques :

- CBAM (Convolutional Block Attention Module).
- Squeeze-and-Excitation Networks (SENet).

Cela améliore la détection des lésions diffuses ou peu visibles, notamment dans les premiers stades de la RD.

### 3.4 Entraînement supervisé et annotation

La majorité des modèles nécessitent un grand volume d'annotations médicales, ce qui représente un défi :

- Annotation coûteuse, nécessitant des experts ophtalmologues.
- Déséquilibre des classes (les cas sévères étant moins fréquents).

Pour contrer cela, des méthodes de data augmentation (rotation, zoom, bruit), semi-apprentissage ou apprentissage faiblement supervisé sont souvent utilisées.

## 3.5 Approches modernes : Vision Transformers et Vision Mamba

Avec les progrès récents en apprentissage automatique, de nouvelles architectures ont émergé, capables de surpasser les CNN dans certaines tâches de vision par ordinateur. Parmi elles, les Vision Transformers (ViT) [20] et les architectures séquentielles avancées comme Vision Mamba marquent un tournant dans l'analyse d'images médicales, y compris pour la détection et la classification de la rétinopathie diabétique (RD).

### A. Vision Transformers (ViT)

Inspirés des modèles Transformers utilisés en traitement du langage naturel (NLP), les Vision Transformers ont été adaptés pour la vision en découpant une image en patchs (petites régions) et en les traitant comme une séquence de « mots visuels » [21].

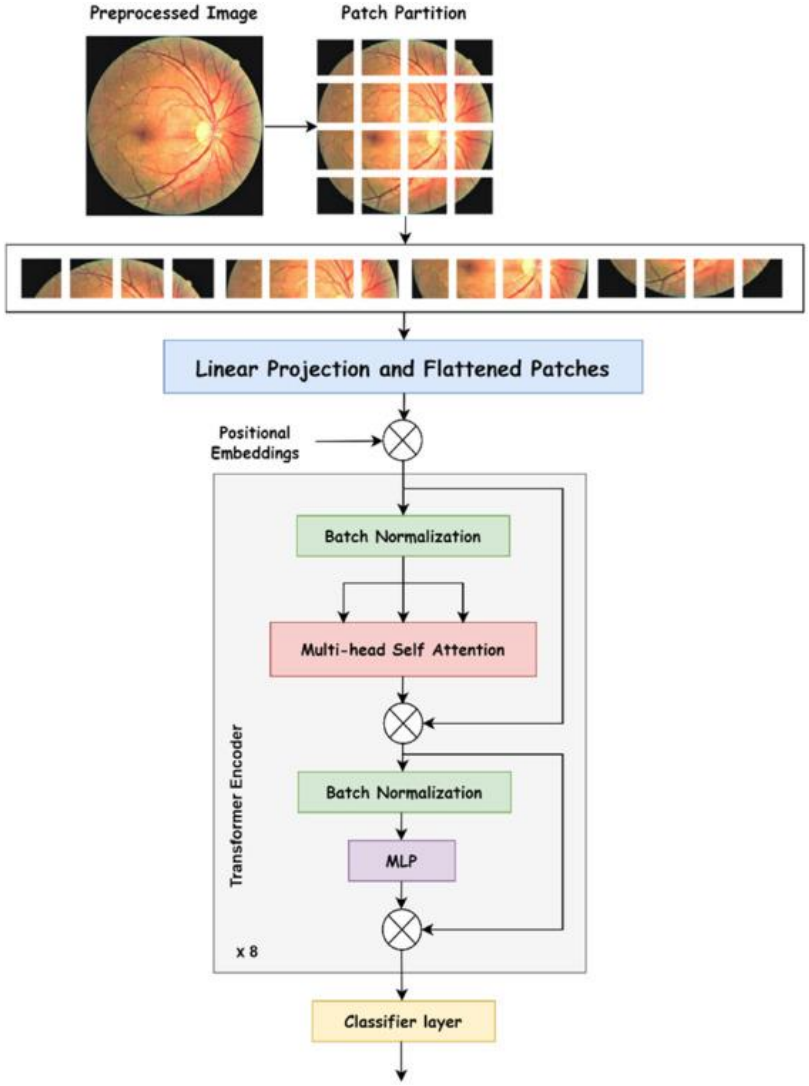


Figure 1.8. Architecture d'un Vision Transformers pour la classification.

- L'image est divisée en patchs (ex. 16×16 pixels).

- Chaque patch est encodé en vecteur, auquel on ajoute une position.
- Un encodeur Transformer traite cette séquence pour extraire les dépendances globales.
- Apprentissage de relations globales : capable de capturer des dépendances longue distance dans l'image (par exemple, entre deux lésions éloignées).
- Flexibilité architecturale : facile à adapter pour la classification, la détection ou la segmentation.
- Classification fine des stades de la RD.
- Localisation explicite des régions anormales via les cartes d'attention.

**Exemples d'architectures :** ViT (Vision Transformer de base) [22]. Swin Transformer : version hiérarchique avec attention locale. DETR (DEtection TRansformer) : pour la détection d'objets (lésions)...etc [23].

**B. Vision Mamba**

Vision Mamba est une architecture plus récente qui s'appuie sur les modèles séquentiels d'état (State Space Models – SSM) [24]. Contrairement aux Transformers qui utilisent de l'attention globale, Mamba offre un mécanisme efficace de traitement séquentiel tout en conservant la capacité de modéliser des dépendances à longue portée [25].

- Moins coûteux en calcul que les Transformers pour des séquences longues.
- Excellente capacité de modélisation contextuelle sur de grandes images médicales.
- Mamba apprend à sélectionner dynamiquement des régions importantes, sans coût attentionnel.
- Traitement rapide et précis d'images haute résolution.
- Capacité à modéliser les séquences spatiales complexes dans la rétine.
- Meilleure scalabilité que les Transformers classiques pour les grandes bases de données ophtalmologiques.

**C. Comparaison des approches CNN, ViT et Mamba**

**Tableau 1.2:** Comparaison des approches CNN, ViT et Mamba.

| Caractéristiques | CNN | Vision Transformer (ViT) | Vision Mamba |
|---|---|---|---|
| Capture du contexte global | Limitée | Très bonne | Excellente |
| Coût de calcul | Faible à modéré | Élevé (attention quadratique) | Faible (traitement séquentiel) |
| Interprétabilité | Moyenne (Grad-CAM) | Bonne (cartes d'attention) | Bonne (sélection dynamique) |
| Performances sur la RD | Élevées | Très élevées (avec préentraînement) | Prometteuses (en cours d'exploration) |

Vision Transformers ont permis de franchir un cap dans l'analyse des relations spatiales globales dans les images de fond d'œil, tandis que les modèles comme Vision Mamba offrent une alternative efficace, moins coûteuse et potentiellement plus adaptée à des déploiements en milieu clinique [26]. Ces nouvelles approches représentent l'avant-garde de la recherche en détection automatisée de la rétinopathie diabétique.

## 4 Limites des approches actuelles

Bien que les méthodes traditionnelles et modernes aient permis d'importants progrès dans la détection automatisée de la rétinopathie diabétique (RD), plusieurs limites persistent, freinant leur déploiement à grande échelle, notamment en milieu clinique réel. Ces contraintes peuvent être regroupées en plusieurs catégories :

- Hétérogénéité des images : Les images de fond d'œil varient fortement en fonction des capteurs utilisés, des conditions d'éclairage, et des spécificités anatomiques des patients (teinte du fond d'œil, mydriase, etc.)
- Images de faible qualité : Des artefacts, un flou ou une mauvaise exposition peuvent gêner l'extraction automatique des caractéristiques.
- Déséquilibre des classes : Les stades avancés de la RD sont souvent sous-représentés, ce qui biaise l'entraînement des modèles et diminue leur capacité à détecter les cas sévères.
- Boîte noire : Les modèles d'apprentissage profond (CNN, ViT, Mamba) sont souvent perçus comme des boîtes noires, rendant difficile leur adoption par les professionnels de santé.
- Manque d'explicabilité : Les outils comme Grad-CAM ou les cartes d'attention apportent des pistes, mais ne sont pas toujours suffisants pour expliquer précisément une décision.
- Annotations coûteuses : L'entraînement des modèles nécessite des annotations de haute qualité par des ophtalmologistes expérimentés, ce qui représente un coût important en temps et en ressources.
- Variabilité inter-experts : Même entre spécialistes, le diagnostic peut varier, introduisant une incertitude dans les labels utilisés pour l'apprentissage.
- Sur-apprentissage (overfitting) : Les modèles entraînés sur des bases de données spécifiques (ex. EyePACS, Messidor) peuvent ne pas bien généraliser à d'autres populations ou à d'autres appareils de capture.

- Manque de robustesse : Une légère perturbation de l'image (flou, bruit, rotation) peut altérer significativement les performances du modèle.
- Modèles lourds : Les Vision Transformers et certains CNN profonds nécessitent une puissance de calcul élevée (GPU), ce qui limite leur intégration dans des dispositifs médicaux portables ou dans des contextes à ressources limitées.
- Temps de traitement : Certains modèles peuvent être lents à l'inférence, ce qui complique leur usage en temps réel dans les cliniques ou centres de dépistage.

Le tableau comparatif des travaux de recherche sur la classification de la rétinopathie diabétique (RD) met en évidence plusieurs aspects clés, notamment les ensembles de données utilisés, les architectures de classification employées, les méthodes de validation et les performances atteintes en termes de métriques [27]. Les travaux recensés exploitent divers ensembles de données publics et privés, parmi lesquels on retrouve EyePACS, Kaggle, APTOS, MESSIDOR, DDR et IDRID. EyePACS et MESSIDOR sont les plus fréquemment utilisés, en raison de leur taille et de leur diversité [28]. L'utilisation de ensembles de données privés peut limiter la reproductibilité des résultats et la comparaison directe avec d'autres travaux.

**Tableau 1.3 :** Compilation des approches de la classification de la rétinopathie diabétique (RD) développées au cours des cinq dernières années.

| Référence | Ensemble de données | Technique de classification | Métrique | Année |
|---|---|---|---|---|
| [27] | EyePACS | ResNet 50 | 97.87% | 2023 |
| [28] | Kaggle | ResNet152 | 94.40% | 2023 |
| [29] | APTOS | Hybrid<br>a combination of VGG16 and XGBoost Classifier | 79.50% | 2023 |
| [30] | APTOS | Google Net | 97% | 2022 |
| [31] | APTOS | CNN | 94.44% | 2022 |
| [32] | APTOS 2019 Blindness | Local binary patterns (LBP)+ResNet-based system | 96.36% | 2021 |
| [33] | DDR and EyePACS | Deep neural network and Inception-ResNet-v2-based framework | 93.33% | 2021 |
| [34] | MESSIDOR | Deep multi-task DR grading (Deep MT-DR) model | 88.7% | 2021 |
| [35] | DDR and EyePACS | Deep multi-task DR grading (Deep MT-DR) model | 88.7% | 2021 |
| [36] | Private dataset | GAN+ViT | 85.7% | 2021 |
| [37] | APTOS, RFMiD | ViT (MIL) | 85.5% | 2021 |
| [38] | Custom-developed at Hospital of Zhejiang University School of | Stochastic gradient descent (SGD) and DenseNet-based approach | 96.53% | 2020 |

| | | | | |
|---|---|---|---|---|
| | Medicine from August 2016 to October 2018 | | | |
| [39] | MESSIDOR | Histogram equalization and contrast limited adaptive histogram equalization + CNN | 97% | 2020 |
| [40] | MESSIDOR | Firefly-principal component analysis and deep neural network-based model | 96% | 2020 |
| [41] | MESSIDOR | DCNN and synergic network (SN)-based model | 99.28% | 2020 |
| [42] | MESSIDOR | Hyper-parameter-tuning Inception-V4 (HPTI-V4) and feed-forward artificial neural network (ANN)-based model | 99.49% | 2020 |
| [43] | MESSIDOR and custom-developed at Beijing Tongren Eye Center | ResNet and gradient-weighted class activation mapping (Grad-CAM)-based multi-label model | 93.9% | 2020 |
| [44] | Messidor-2 | Custom CNN | 91% | 2020 |
| [45] | Messidor | Custom CNN | 92.6% | 2020 |
| [46] | Messidor, IDRiD | ResNet50 | 92.6% | 2020 |
| [47] | IDRID | Custom CNN | 91% | 2019 |
| [48] | DDR | ResNet | 82.84% | 2019 |
| [49] | Private dataset | Multi CNN | 96.5% | 2019 |
| [50] | EyePACS | VGGNet | 95.68% | 2018 |
| [51] | EyePACS | Inception V3 | 63.23% | 2018 |
| [52] | EyePACS | Custom CNN | 86.10% | 2018 |

Les modèles de deep learning dominent les approches de classification, avec une prédilection pour les réseaux de neurones convolutifs (CNN) et leurs variantes. Les architectures les plus couramment utilisées incluent :

- ResNet (ResNet-50, ResNet-101, ResNet-152) : Utilisée pour ses capacités d'apprentissage profond, elle montre de bonnes performances avec des précisions souvent supérieures à 90 %.
- VGGNet : Bien qu'efficace, elle tend à être moins performante que ResNet en raison de sa profondeur plus réduite.
- GoogleNet (Inception-V3, Inception-ResNet-v2) : Utilisée pour sa capacité à extraire des caractéristiques complexes.
- Hybrid CNN + autres techniques : Certains travaux intègrent des approches hybrides, combinant CNN avec des classificateurs comme XGBoost, des modèles basés sur l'attention (ViT), ou des techniques d'extraction de caractéristiques comme PCA.

La validation des modèles est un aspect crucial, bien que certaines publications n'explicitent pas clairement leurs protocoles. Les méthodes de validation courantes comprennent :

- Validation croisée (k-fold cross-validation) : Fréquemment utilisée pour réduire le surapprentissage.
- Ensembles d'entraînement/test (évaluation sur des ensembles de données indépendants) : Permet d'évaluer la généralisation du modèle.
- Grad-CAM et autres méthodes interprétables : Certaines approches utilisent des techniques comme Grad-CAM pour visualiser les régions d'intérêt des modèles.

Les performances des modèles sont mesurées principalement par l'exactitude (accuracy), mais d'autres métriques comme la précision, le rappel et le score F1 sont parfois considérés. Les meilleures performances (>97%) sont obtenues avec des architectures comme ResNet-50, GoogleNet, et des modèles hybrides (VGG16+XGBoost). Les performances intermédiaires (85-95%) sont souvent observées avec des modèles comme ViT et certaines configurations de CNN. Les performances plus faibles (<80%) concernent principalement les modèles moins profonds ou ceux appliqués à des tâches multi-label complexes.

Augmentation des performances via des architectures plus profondes et hybrides : Les combinaisons de CNN avec des algorithmes d'optimisation ou des approches hybrides (CNN + XGBoost, CNN + PCA) montrent un potentiel intéressant. Interprétabilité et explication des décisions des modèles : L'intégration de techniques d'explication est cruciale pour l'acceptation clinique. Harmonisation des méthodes de validation : L'absence de standardisation complique la comparaison entre les études. Utilisation de plus grands ensembles de données diversifiés : La généralisation reste un enjeu, surtout lorsque les données proviennent de sources limitées.

L'analyse des différents travaux montre une nette préférence pour les architectures CNN profondes, notamment ResNet et GoogleNet. Les approches hybrides et l'utilisation de nouvelles techniques comme ViT ou Grad-CAM ouvrent des perspectives intéressantes. Toutefois, des efforts sont encore nécessaires pour garantir une évaluation rigoureuse et standardisée des performances des modèles.

Malgré leur potentiel, les approches actuelles souffrent encore de limitations majeures en termes de généralisation, d'explicabilité, et de faisabilité clinique. Une recherche active est en cours pour répondre à ces défis, notamment par l'intégration de techniques d'apprentissage auto-supervisé, de mécanismes d'explicabilité plus avancés, et de modèles plus légers adaptés à l'usage réel.

# 5 Conclusion

Ce chapitre a présenté un état de l'art complet des approches utilisées pour la détection et la classification de la rétinopathie diabétique (RD). Nous avons commencé par une exploration de l'anatomie de l'œil et de la structure de la rétine, afin de mieux comprendre les régions anatomiques ciblées par les algorithmes de diagnostic. Ensuite, la classification des stades de la RD a été détaillée, fournissant le cadre clinique fondamental sur lequel se basent les systèmes de classification automatique.

Nous avons ensuite examiné les méthodes traditionnelles de traitement d'images et d'apprentissage automatique, qui ont constitué les premières tentatives de diagnostic assisté par ordinateur. Toutefois, ces approches ont progressivement laissé place aux méthodes modernes de deep learning, notamment les réseaux de neurones convolutifs (CNN), qui ont montré de nettes améliorations en termes de précision.

L'émergence deep learning et des modèles récents a ouvert de nouvelles possibilités en capturant des dépendances globales et temporelles dans les images, augmentant ainsi la capacité des modèles à détecter les signes subtils de la maladie. Malgré leurs performances impressionnantes, ces modèles présentent encore des limites importantes, notamment en termes de besoin en données, d'interprétabilité et de robustesse.

Enfin, nous avons identifié plusieurs pistes d'amélioration prometteuses pour surmonter ces défis : enrichissement des données, amélioration de la généralisation des modèles, renforcement de l'explicabilité, et optimisation pour un usage clinique réel. Ces perspectives ouvrent la voie à des solutions plus fiables, accessibles et intégrables dans les systèmes de santé.

Ce socle théorique est essentiel pour motiver le développement de notre propre approche, présentée dans le chapitre suivant, qui s'appuie sur ces constats pour proposer une méthode optimisée de détection de la RD.

# **Chapitre 2 :** Les techniques d'apprentissage appliquées

**Résumé**

Dans ce chapitre, nous avons détaillé la méthodologie employée pour développer et évaluer un système de classification de la rétinopathie diabétique basé sur un ensemble de modèles deep learning. Nous détaillons l'architecture de deux ensembles distincts : l'un combinant EfficientNet-B3 et DenseNet-169, et l'autre EfficientNet-B3 et PiT (Pooling-based Vision Transformer).

# 1 Introduction

Ce chapitre présente en détail la méthodologie adoptée pour développer un système de classification automatique de la rétinopathie diabétique à partir d'images de fond d'œil. L'objectif principal est d'exploiter la complémentarité de plusieurs modèles de deep learning afin d'améliorer la précision de la détection et la robustesse du système [42].

La démarche suit plusieurs étapes successives. Tout d'abord, nous décrivons le jeu de données utilisé, APTOS 2019 Blindness Detection, issu d'une compétition organisée sur la plateforme Kaggle, qui fournit des images annotées selon les cinq stades cliniques de la rétinopathie diabétique. Ensuite, un prétraitement des images est réalisé pour normaliser les conditions d'entrée, corriger les effets d'éclairage, et renforcer la qualité visuelle des régions rétiniennes. Des techniques d'augmentation de données sont également mises en œuvre pour enrichir le jeu d'entraînement et améliorer la capacité de généralisation des modèles [52, 51].

L'architecture d'ensemble constitue une approche puissante en apprentissage profond, visant à combiner plusieurs modèles pour améliorer les performances globales d'un système de classification. Contrairement à un modèle unique, une architecture d'ensemble exploite la complémentarité entre différents réseaux de neurones afin de capter des représentations variées et riches des données. Cette stratégie permet de réduire la variance, améliorer la robustesse et atténuer les erreurs de généralisation, ce qui est particulièrement important dans des contextes complexes tels que la détection de la rétinopathie diabétique.

Dans ce travail, nous avons conçu deux architectures d'ensemble reposant sur des réseaux profonds performants : DenseNet-169, EfficientNet-B3, et PiT-Base. Chaque architecture propose une manière différente de fusionner les prédictions issues de plusieurs modèles entraînés en parallèle sur des images de fond d'œil. Le premier algorithme combine DenseNet-169 et EfficientNet-B3 pour tirer parti de la connectivité dense et de l'efficacité paramétrique, tandis que le second fusionne EfficientNet-B3 avec un transformeur visuel (PiT-Base) pour enrichir l'apprentissage par des mécanismes d'attention.

Ce chapitre décrit les fondements théoriques de chaque composant, la motivation de leur sélection, ainsi que l'architecture complète de l'ensemble mis en œuvre. L'objectif est de démontrer comment l'intégration intelligente de plusieurs réseaux peut améliorer les performances de classification dans le cadre du dépistage automatique de la rétinopathie diabétique.

# 2 Architecture de l'ensemble

L'architecture d'ensemble repose sur une stratégie de parallélisation, où les différents réseaux (DenseNet-169, EfficientNet-B3, et PiT-Base selon l'algorithme) sont entraînés indépendamment à partir des mêmes données d'entrée. Les prédictions issues de chaque réseau sont ensuite fusionnées à l'aide d'une méthode d'agrégation, typiquement une moyenne pondérée ou une couche dense de fusion.

Cette approche permet de combiner les points forts de chaque architecture : la connectivité dense de DenseNet, l'efficacité paramétrique d'EfficientNet et les capacités d'attention globale de PiT. L'objectif est d'obtenir une représentation plus complète et plus discriminante de l'image, conduisant à une amélioration significative des performances de classification.

## 2.1 Ensemble EfficientNet-B3 et DenseNet-169

L'architecture proposée repose sur une combinaison de deux réseaux profonds performants, EfficientNet-B3 et DenseNet-169, agissant en parallèle pour extraire des caractéristiques complémentaires à partir des images de rétine. Chaque réseau produit une prédiction indépendante, fusionnée en fin de traitement par une stratégie d'agrégation par fusion (cf. figure 2.1).

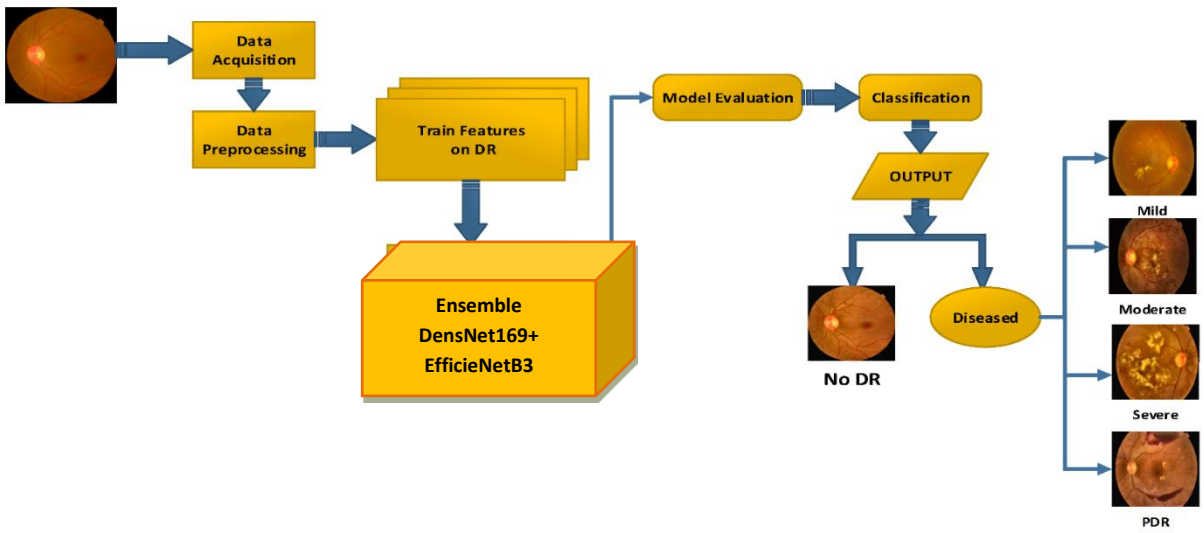


Figure 2.1 : Architecture proposée pour classification de la rétinopathie diabétique.

### A. DenseNet-169

DenseNet-169 est un réseau de neurones convolutifs profond caractérisé par des connexions denses entre les couches, où chaque couche reçoit en entrée les sorties de toutes les

couches précédentes. Cette architecture favorise la réutilisation des caractéristiques et améliore la propagation des gradients.

DenseNet a été introduit par Gao Huang et al. (CVPR 2017) pour pallier deux problèmes récurrents dans les très grands réseaux profonds :

- Vanishing gradients : à mesure que la profondeur augmente, les gradients se diluent et la phase de rétropropagation devient instable.
- Redondance des caractéristiques : les couches profondes apprennent souvent des représentations très similaires à celles des couches précédentes, gaspillant des paramètres.
- Les auteurs ont proposé de connecter chaque couche à toutes les couches antérieures par une concaténation de leurs sorties, ce qui :
- Assure un flux direct de gradient depuis la couche de sortie jusqu'aux premières couches, facilitant l'entraînement.

**Structure principale:**

- **Entrée :** Image de taille 224×224×3.
- **Convolution initiale :** Convolution 7×7 avec stride 2, suivie d'un max pooling 3×3 avec stride 2.
- **Blocs denses :** Quatre blocs denses avec respectivement 6, 12, 32 et 32 couches.
- **Couches de transition :** Entre les blocs denses, des couches de transition composées d'une convolution 1×1 suivie d'un average pooling 2×2 avec stride 2.
- **Activation :** Fonction ReLU après chaque convolution.
- **Sortie :** Après le dernier bloc dense, application d'un global average pooling, suivi d'un dropout (taux de 0,5), puis d'une couche dense de 1 024 neurones avec activation ReLU, et enfin d'une couche de sortie softmax à 5 neurones pour la classification.

**B. EfficientNet-B3**

EfficientNet-B3 est une architecture optimisée utilisant une méthode de mise à l'échelle composée (compound scaling) qui ajuste uniformément la profondeur, la largeur et la résolution du réseau. Elle intègre des blocs MBConv avec des connexions résiduelles inversées et des mécanismes de squeeze-and-excitation.

EfficientNet-B3 est le plus performant de la famille EfficientNet proposée par Mingxing Tan et Quoc V. Le en 2019. Son architecture repose sur le principe du compound scaling, qui ajuste simultanément la profondeur, la largeur et la résolution d'entrée d'un réseau de manière systématique, contrairement aux approches traditionnelles qui ne varient qu'un seul facteur.

EfficientNet-B3 est basé sur une version améliorée de MobileNetV2, avec les ajouts suivants :

- MBConv (Mobile Inverted Bottleneck Convolution) : bloc de base utilisant des convolutions depthwise separable avec expansion et projection.
- Squeeze-and-Excitation (SE) : modules SE intégrés dans chaque bloc MBConv pour renforcer la modélisation des dépendances entre canaux.

**Structure principale:**

- **Entrée:** Image de taille 224×224×3.
- **Convolution initiale :** Convolution 3×3 avec stride 2, suivie d'une activation Swish.
- **Blocs MBConv :** Série de blocs MBConv avec différentes configurations (expansion, kernel size, stride, squeeze-and-excitation).
- **Activation :** Fonction Swish dans les blocs MBConv.
- **Sortie :** Après les blocs MBConv, application d'un global average pooling, suivi d'un dropout (taux de 0,5), puis d'une couche dense de 1 024 neurones avec activation ReLU, et enfin d'une couche de sortie softmax à 5 neurones pour la classification.

**C. Fusion de DenseNet-169 et EfficientNet-B3**

L'architecture ensembliste combine les sorties des deux modèles pour améliorer la performance de classification.

**Processus de fusion:**

1. **Entrée commune :** Image de taille 224×224×3, adaptée pour DenseNet-169.
2. **Traitement parallèle :** L'image est traitée simultanément par DenseNet-169 et EfficientNet-B3 (avec redimensionnement à 224×224×3 pour ce dernier).
3. **Extraction des caractéristiques :** Chaque modèle extrait ses propres caractéristiques via ses couches respectives.
4. **Fusion des caractéristiques :** Les vecteurs de caractéristiques obtenus sont concaténés.
5. **Classification finale :** Le vecteur fusionné passe par un dropout (taux de 0,5), une couche dense de 1 024 neurones avec activation ReLU, puis une couche de sortie softmax à 5 neurones pour la classification.

Cette fusion permet de tirer parti des forces de chaque architecture, améliorant ainsi la robustesse et la précision du modèle final.

L'algorithme proposé (cf. algorithme 1) pour l'entraînement et la prédiction repose sur une architecture ensembliste combinant deux modèles profonds préentraînés : DenseNet-169 et EfficientNet-B3. L'objectif est de tirer parti de leurs forces complémentaires en matière

d'extraction de caractéristiques pour améliorer la classification des stades de la rétinopathie diabétique.

Le processus commence par la préparation des deux modèles de base, chacun tronqué avant leur couche de classification initiale. Une couche de global average pooling, suivie d'un dropout et d'un perceptron de 1 024 neurones avec activation ReLU, est ajoutée à la sortie de chaque modèle. Les deux branches reçoivent la même image d'entrée, prétraitée et redimensionnée aux formats requis (224×224 pour DenseNet, 224×224 pour EfficientNet).

Les représentations extraites par les deux branches sont ensuite concaténées pour former un vecteur de caractéristiques fusionnées. Ce vecteur est à nouveau régularisé par un dropout avant d'être traité par une couche dense de 1 024 neurones. La sortie finale est une couche softmax à 5 neurones représentant les cinq classes de la maladie.

Le modèle est ensuite compilé avec l'optimiseur Adam, une fonction de perte catégorique (cross-entropy) et des métriques standard (accuracy). L'entraînement est réalisé en utilisant des générateurs de données avec augmentation, incluant des transformations géométriques aléatoires pour améliorer la généralisation. Des callbacks tels que l'early stopping, la réduction dynamique du taux d'apprentissage et la sauvegarde des meilleurs poids assurent un entraînement stable et optimal.

Enfin, le modèle entraîné est évalué sur les données de validation et de test à l'aide de métriques multicritères, incluant l'exactitude, le F1-score et le score Kappa quadratique. (Cf annexe lien de code Algorithme 1)

**Algorithme 1 : Entraînement et prédiction de l'ensemble proposé**

**Entrées:**

- X_train, y_train : données d'entraînement
- X_val, y_val : données de validation
- Model_DenseNet, Model_EfficientNet : modèles pré-entraînés
- nb_classes : nombre de classes (ici 5)
- epochs, batch_size, learning_rate: hyperparamètres
- callbacks: early stopping, ReduceLROnPlateau, checkpoints

**Sorties:**

- Model_ensemble : modèle entraîné
- y_pred : prédictions sur les nouvelles données

**Étapes:**

1. **Préparation des modèles de base**
    - Charger DenseNet-169 et EfficientNet-B3 avec les poids ImageNet
    - Supprimer leurs couches de sortie respectives
    - Ajouter pour chacun:
        - GlobalAveragePooling2D
        - Dropout(rate=0.5)
        - Dense (1024, activation='relu')
2. **Fusion des caractéristiques**
    - Entrée commune Input (shape= (224, 224, 3))
    - Redimensionner une copie à 224×224×3 pour EfficientNet-B3
    - Appliquer chaque sous-modèle sur son entrée respective
    - Concaténer les sorties des deux branches
    - Appliquer:
        - Dropout(rate=0.5)
        - Dense (1024, activation='relu')
        - Dense (nb_classes, activation='softmax')
3. **Compilation de l'ensemble**
    - Définir Model_ensemble avec l'entrée commune et la sortie fusionnée
    - Compiler avec:
        - optimizer = Adam(learning_rate)
        - loss = categorical_crossentropy
        - metrics = ['accuracy']
4. **Prétraitement et augmentation**
    - Appliquer le pipeline de transformation :
        - crop, redimensionnement, normalisation, amélioration du contraste
        - ImageDataGenerator avec flips, zoom, rotation
5. **Entraînement**
    - Utiliser les générateurs train_generator, val_generator
    - Appeler:

```
Model_ensemble.fit (
  train_generator,
```

```
    validation_data=val_generator,
    epochs=epochs,
    batch_size=batch_size,
    callbacks=callbacks
)
```

6. **Sauvegarde du meilleur modèle**
    - Sauvegarder les poids avec le meilleur val_loss ou meilleure Kappa
7. **Prédiction**
    - Charger les meilleurs poids du modèle
    - Appliquer les mêmes prétraitements à X_test
    - Calculer:

```
y_pred = Model_ensemble.predict(X_test)
y_pred_classes = argmax (y_pred, axis=1)
```

8. **Évaluation**
    - Calculer accuracy, F1-score, cohen_kappa, matrice de confusion

L'approche ensembliste mise en œuvre dans ce projet permet de combiner efficacement la puissance de deux modèles réputés pour leur capacité à extraire des caractéristiques discriminantes à différentes échelles. DenseNet-169 excelle dans la réutilisation des informations locales via ses connexions denses, tandis qu'EfficientNet-B3 exploite une mise à l'échelle équilibrée de la profondeur, de la largeur et de la résolution pour capturer des patterns globaux avec efficacité.

En fusionnant leurs sorties dans un cadre commun, l'ensemble tire profit de la complémentarité structurelle et sémantique entre les deux modèles. Cette synergie se traduit par une meilleure robustesse face aux variations des images rétiniennes, une réduction du risque de sur-apprentissage et une amélioration globale de la performance sur l'ensemble des classes. Cette architecture démontre ainsi tout son intérêt pour une tâche médicale exigeante comme la classification fine des stades de la rétinopathie diabétique.

## 2.2 Combinaison de EfficientNet-B3 et PiT-Base-224

Cette approche combine les avantages de deux paradigmes fondamentaux en vision par ordinateur : les réseaux convolutifs (CNNs) pour l'extraction de caractéristiques locales, et les

Transformers visuels pour la modélisation globale et contextuelle. EfficientNet-B3 est performant pour détecter des détails fins (microanévrismes, exsudats, hémorragies) grâce à ses blocs MBConv efficaces. PiT-Base (Pooling-based Vision Transformer) capture des relations longues distances et une vue d'ensemble de l'image, ce qui est essentiel pour évaluer la progression globale de la rétinopathie. La combinaison vise à réduire les erreurs de confusion inter-classes (modérée vs sévère, etc.) en enrichissant la représentation. **Vision hiérarchique** du PiT : meilleure généralisation sur données médicales limitées, par rapport au ViT standard. **Complémentarité spatiale** : CNN pour structures locales (textures, vaisseaux), Transformer pour disposition et corrélation spatiale à long terme. **Eviter le surapprentissage** grâce au double Dropout et au fine-tuning progressif.

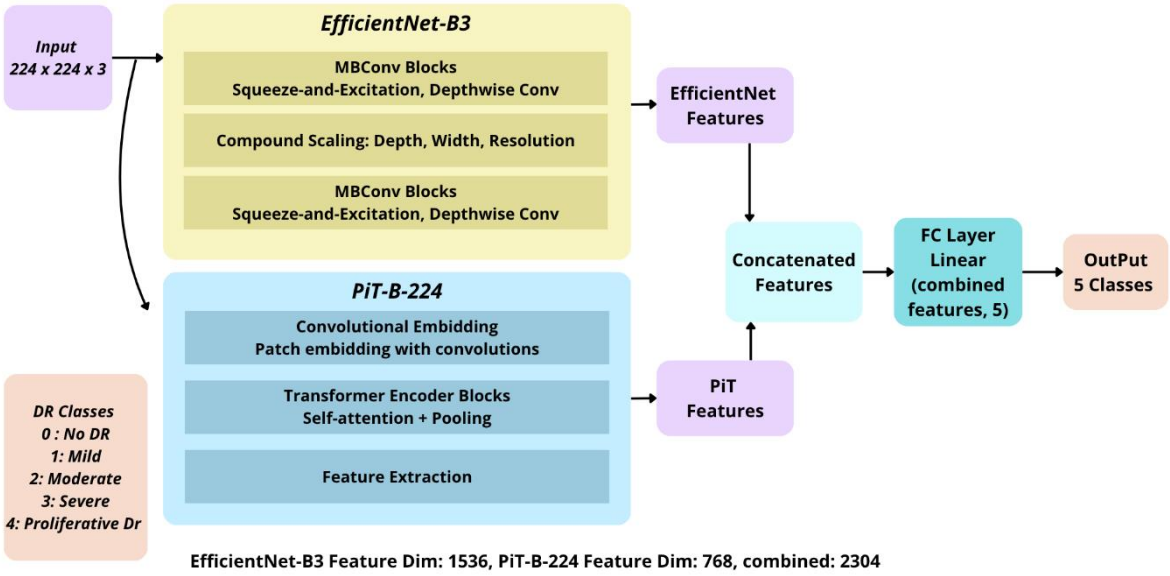


Figure 2.2 : Architecture proposée Combine EfficientNetB3 and PiT-Base pour la classification de la rétinopathie diabétique.

## A. EfficientNet-B3

EfficientNet-B3 est une architecture convolutive optimisée pour l'efficacité en termes de performances et de nombre de paramètres. Elle repose sur le concept de mise à l'échelle composée (compound scaling), ajustant de façon coordonnée la profondeur, la largeur et la résolution des images d'entrée. Dans notre système:

- Le modèle est pré-entraîné sur ImageNet.
- Les images sont redimensionnées à 224×224×3.
- Une couche de global average pooling extrait un vecteur de caractéristiques.

- Un dropout de 0,5 est appliqué pour régulariser.
- Le vecteur résultant est transmis pour fusion avec PiT.

**B. PiT-Base (Pooling-based Vision Transformer)**

PiT est une variante des Vision Transformers intégrant des opérations de **pooling hiérarchique** pour simuler l'effet de réduction de dimension des CNNs. Il permet de mieux gérer la complexité tout en préservant les bénéfices de la modélisation globale de la vision transformer.

- Le modèle PiT-Base-224 est utilisé avec des poids pré-entraînés.
- Les images sont également redimensionnées à 224×224×3.
- Un embeddings linéaire transforme l'image en séquences de patchs.
- Des blocs d'attention avec pooling hiérarchique extraient des représentations globales.
- Un vecteur de caractéristiques est obtenu après pooling final et normalisation.

**C. Fusion des Caractéristiques**

Les vecteurs extraits par EfficientNet-B3 et PiT sont concaténés :

- Une couche de dropout (0,5) est appliquée à la fusion.
- Une couche dense de 1 024 neurones avec ReLU est utilisée.
- Une couche softmax à 5 neurones effectue la classification multiclasse.

Cette architecture hybride vise à exploiter les forces complémentaires des CNNs (caractéristiques locales) et des Transformers (relations globales). (Cf l'annexe lien de code Algorithme 2).

**Algorithme 2 : Entraînement et Prédiction – Combinaison EfficientNet-B3 & PiT-Base**

**Entrées**:

- Jeu d'images annotées $D = \{(x_i, y_i)\}_{i=1}^{N}$
- Modèles pré-entraînés EfficientNet-B3 $M_1$ et PiT-Base $M_2$
- Taux d'apprentissage $\alpha$, nombre d'époques $E$, taille de lot $B$
- Fonction de coût $\mathcal{L}$ (par exemple, entropie croisée)

**Sorties**:

- Modèle entraîné : $f(x) = Softmax(W[\phi_1(x)||\phi_2(x)])$

**Prétraitement des images**

1. Appliquer CLAHE (Contrast Limited Adaptive Histogram Equalization).

2. Convertir en niveaux de gris si nécessaire.
3. Recadrer dynamiquement les images autour du contenu principal (fond supprimé).
4. Redimensionner à 224×224 \times 224×224 pixels.
5. Normaliser les images selon les statistiques d'ImageNet.

**Étapes d'Entraînement**

**Pour chaque époque e = 1 à E faire :**

1. **Pour** chaque lot $\{(x_b, y_b)\}_{b=1}^{B} \subset D_{train}$ faire :
   - Extraire les représentations locales : $\phi_1(x_b) = M_1(x_b)$
   - Extraire les représentations globales : $\phi_2(x_b) = M_2(x_b)$
   - Concaténer les deux : $h_b = [\phi_1(x_b)||\phi_2(x_b)]$
   - Prédire les probabilités : $\hat{y}_b = Softmax(Wh_b + b)$
   - Calculer la perte : $\mathcal{L}_b = \mathcal{L}(\hat{y}_b, y_b)$
   - Propager le gradient et mettre à jour les poids : Descente de gradient stochastique.

**Étapes de Prédiction**

**Pour** chaque image $x \in D_{train}$:

1. Appliquer les mêmes étapes de prétraitement.
2. Calculer $\phi_1(x)\ et\ \phi_2$.
3. Fusionner les deux vecteurs : $h = [\phi_1(x)||\phi_2(x)]$
4. Prédire la classe : $\hat{y} = \arg\max\ Softmax(Wh + b)$

Cet algorithme vise à combiner deux architectures de réseaux de neurones convolutifs avancées, EfficientNet-B3 et PiT-Base (Progressive Tokenization Transformer), afin d'exploiter à la fois les caractéristiques locales (via CNN) et globales (via ViT) des images rétiniennes pour la classification de la rétinopathie diabétique.

Le processus commence par un prétraitement rigoureux des images, incluant le contraste adaptatif (CLAHE), la suppression de l'arrière-plan, et le redimensionnement standardisé, garantissant une entrée cohérente pour les deux modèles. Ensuite, chaque image est propagée à travers EfficientNet-B3, qui extrait des caractéristiques locales riches basées sur des

convolutions optimisées en profondeur, et PiT-Base, qui extrait des caractéristiques globales basées sur des auto-attentions.

Les vecteurs de caractéristiques issus des deux modèles sont concaténés afin de constituer une représentation hybride de l'image. Cette représentation est ensuite transmise à un classifieur linéaire (une couche entièrement connectée suivie d'une fonction Softmax) qui produit les probabilités de chaque classe. Durant l'entraînement, seuls les poids de cette couche de classification sont mis à jour, tandis que les modèles EfficientNet-B3 et PiT-Base restent gelés pour bénéficier de leur pré-entraînement sur ImageNet.

L'entraînement utilise une fonction de perte standard telle que l'entropie croisée, optimisée via descente de gradient, avec l'aide potentielle d'optimisations comme l'apprentissage mixte (AMP) pour réduire l'utilisation de mémoire.

Cette approche de fusion permet de tirer parti de la complémentarité entre la sensibilité spatiale des CNNs et la vision holistique des Transformers, améliorant ainsi la robustesse et la précision du modèle final pour la détection automatique de la rétinopathie diabétique.

# 3 Conclusion

Dans ce chapitre, nous avons développé deux algorithmes de la classification de rétinopathie diabétique. Les deux approches présentées exploitent la complémentarité entre modèles convolutifs (DenseNet, EfficientNet) et architectures plus récentes (Vision Transformer, PiT). L'architecture DenseNet-EfficientNet mise sur la densité des connexions et l'efficacité paramétrique, tandis que l'ensemble EfficientNet-PiT combine la précision locale des CNNs avec la vision globale et contextuelle des Transformers.

Grâce à une stratégie de fusion de caractéristiques, des techniques de prétraitement robustes, et des mécanismes d'entraînement adaptés (poids de classes, précision mixte, callbacks), ces deux systèmes offrent une grande robustesse face à la variabilité des images médicales et à l'équilibre des classes. L'approche EfficientNet + PiT en particulier permet de capturer des patterns visuels complexes en profondeur, ce qui en fait un choix pertinent pour la classification fine des stades de rétinopathie diabétique.

# Chapitre 3 : Résultats et Discussions

**Résumé**

Dans ce chapitre, nous présentons les expériences réalisées pour évaluer les performances des modèles proposées pour la classification de la rétinopathie diabétique. Nous décrivons les paramètres d'entraînement, les métriques de performance utilisées, ainsi que l'analyse des résultats obtenus.

# 1 Introduction

Dans ce chapitre, nous présentons les expériences réalisées pour évaluer les performances du modèle proposé pour la classification de la rétinopathie diabétique. Les sections suivantes décrivent les stratégies d'entraînement, incluant le choix des hyperparamètres, des méthodes de régularisation, ainsi que les protocoles de validation croisée et d'évaluation externe. Enfin, nous présentons les métriques d'évaluation choisies pour mesurer la performance du système, et nous précisons les outils logiciels et matériels utilisés pour les expérimentations.

Cette méthodologie vise à offrir une solution fiable, reproductible et extensible pour la classification automatique de la rétinopathie diabétique, en s'appuyant sur les meilleures pratiques du deep learning appliqué à l'imagerie médicale.

# 2 Environnement Expérimental

Le développement, l'entraînement et l'évaluation du système de classification ont été réalisés dans un environnement technique adapté aux exigences du deep learning, notamment en termes de calcul GPU, de manipulation d'images haute résolution et de reproductibilité des expériences [53].

## 2.1 Langage et Framework

- Langage : Python 3.9 Choisi pour sa richesse en bibliothèques scientifiques, sa syntaxe claire et son écosystème adapté au deep learning.

- Framework principal : PyTorch 1.12 PyTorch offre une grande flexibilité pour la définition de modèles personnalisés, un support GPU efficace via CUDA, ainsi qu'une intégration fluide avec des outils de visualisation (TensorBoard, matplotlib).

- Bibliothèque de haut niveau : FastAI Utilisée pour faciliter les étapes de data loading, d'augmentation, de scheduling de learning rate et pour accélérer le prototypage de modèles. FastAI est particulièrement bien adapté aux workflows sur PyTorch [54].

## 2.2 Traitement et visualisation

- OpenCV : pour les opérations de traitement d'images (redimensionnement, centrage, histogramme).

- Pandas et NumPy : manipulation des annotations, structures de données et calculs numériques.

- Matplotlib et Seaborn : visualisation des courbes d'apprentissage, matrices de confusion, histogrammes de distribution des classes.

### 2.3 Évaluation et métriques

- Scikit-learn : calcul des F1-scores, AUC multiclasses, et génération de matrices de confusion.

- TorchMetrics : métriques différentiables compatibles avec PyTorch pour intégration dans les phases d'entraînement.

### 2.4 Infrastructure matérielle

- GPU : NVIDIA Tesla T4 Tesla /V100 (mémoire 16 Go) Permet l'entraînement efficace de modèles lourds comme EfficientNet-B3 sur des images 512×512, avec traitement par lots raisonnable (batch size = 16).

### 2.5 Environnement de calcul

- Plateforme cloud : Google Colab.

- Système d'exploitation : Ubuntu 20.04.

- Environnement virtuel : gestion via conda ou pipenv pour assurer la reproductibilité des expériences.

### 2.6 Reproductibilité et gestion des expériences

- Seeds fixés : pour NumPy, PyTorch et Python (seed = 42).

- Sauvegarde des checkpoints modèles : meilleure validation sur ensemble de validation (early stopping).

- Logging : sauvegarde automatique des courbes de perte et d'accuracy via TensorBoard et CSV.

# 3 Base de Données Utilisés

## Description de l'APTOS 2019 (Asia Pacific Tele-Ophthalmology Society 2019 Blindness Detection Dataset)

Le jeu de données APTOS 2019 Blindness Detection a été publié dans le cadre d'une compétition organisée par l'Asian Pacific Tele-Ophthalmology Society (APTOS) en collaboration avec la plateforme Kaggle. Il vise à promouvoir le développement d'outils d'intelligence artificielle pour le dépistage de la rétinopathie diabétique à partir d'images de fond d'œil [61].

L'ensemble comprend 3 662 images couleur de la rétine, capturées à l'aide de caméras de fond d'œil sous différents éclairages et résolutions. Chaque image est étiquetée manuellement par des spécialistes selon cinq niveaux de sévérité de la rétinopathie diabétique, définis selon la classification de l'International Clinical Diabetic Retinopathy (ICDR) :

0 - Aucune rétinopathie : pas d'anomalies visibles.

1 - Rétinopathie légère non proliférante : présence de microanévrismes.

2 - Rétinopathie modérée non proliférante : présence d'hémorragies ou d'exsudats.

3 - Rétinopathie sévère non proliférante : nombreuses anomalies vasculaires sans néovascularisation.

4 - Rétinopathie proliférante : néovascularisation, hémorragies du vitré, risques de cécité.

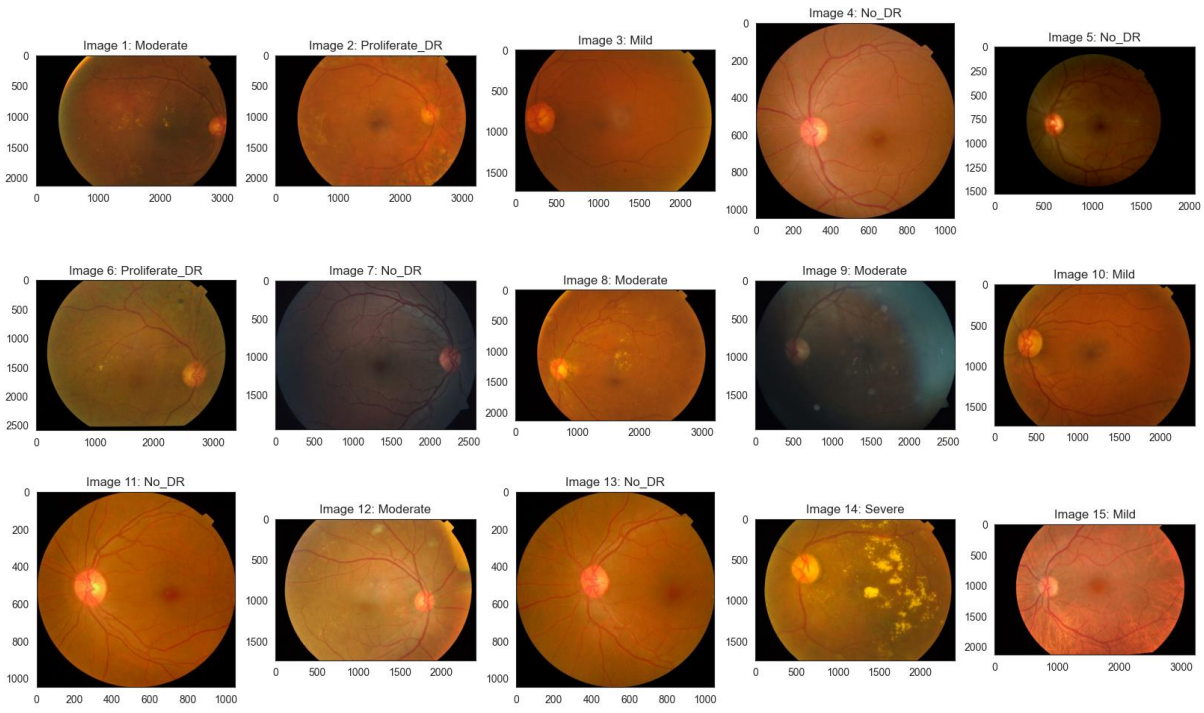


Figure 3.1: Asia Pacific Tele-Ophthalmology Society 2019 Blindness Detection Dataset.

Les images sont de qualité variable, certaines comportant du bruit, des zones floues ou un faible contraste. Cela reflète la diversité des conditions réelles de dépistage et représente un défi important pour les algorithmes d'apprentissage automatique. Le déséquilibre entre les classes (classe 0 surreprésentée) constitue également un enjeu, qui doit être pris en compte via des techniques de pondération ou d'augmentation ciblée [62].

- Source : Collecté par l'Asia Pacific Tele-Ophthalmology Society (APTOS).
- Objectif : Détecter et classifier la sévérité de la rétinopathie diabétique.
- Type de données : Images de fond d'œil (photographies rétiniennes).
- Format des images : Fichiers au format JPEG avec des tailles variables.

L'ensemble de données contient 3 662 images de rétines étiquetées selon cinq niveaux de sévérité de la rétinopathie diabétique, selon la classification de l'International Clinical Diabetic Retinopathy Scale [63].

**Tableau 3.1 :** les classes de l'ensemble de données.

| Classe | Niveau de sévérité | Nombre d'images |
|---|---|---|
| 0 | Aucune rétinopathie diabétique (No DR) | 1 805 |
| 1 | RD non Proliférante légère | 370 |
| 2 | RD non Proliférante modérée | 999 |
| 3 | RD non Proliférante sévère | 193 |
| 4 | RD Proliférante | 295 |

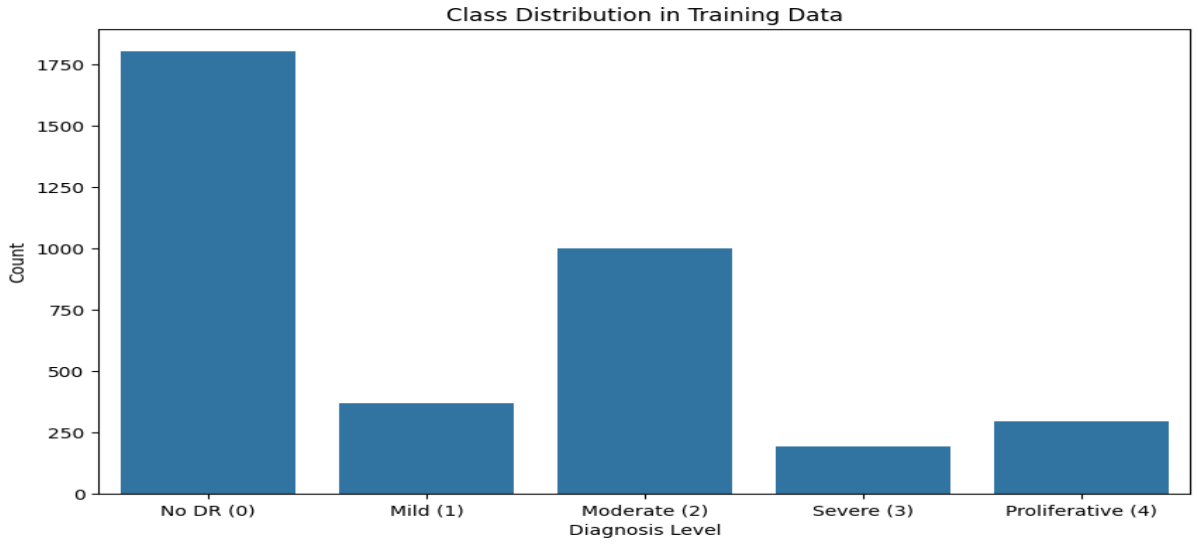


Figure 3.2: Distribution de Dataset.

Le déséquilibre des classes est un défi majeur pour la classification.

Pour utiliser efficacement les images dans des modèles de deep learning, certaines étapes de prétraitement sont nécessaires :

- Redimensionnement : Normalisation des tailles d'image (souvent 224x224, 256x256 ou 512x512 pixels).
- Suppression des artefacts : Réduction du bruit, correction des couleurs et normalisation de l'intensité lumineuse.
- Ajustements de contraste : Utilisation de techniques comme l'égalisation d'histogramme CLAHE.
- Augmentation de données : Rotation, miroir, zoom, pour équilibrer les classes sous-représentées.

Plusieurs architectures de deep learning ont été utilisées pour classifier les images de ce dataset, notamment :

- CNN classiques (ResNet50, InceptionV3, VGG16...etc.).
- Approches hybrides combinant CNN et modèles traditionnels (XGBoost, SVM).
- Transformers en vision (ViT - Vision Transformer).

L'APTOS 2019 est l'un de base de données les plus populaires pour la détection de la rétinopathie diabétique et continue d'être utilisé pour l'évaluation des modèles d'IA en ophtalmologie [64].

Cette base de données constitue un benchmark de référence pour la détection automatique de la rétinopathie diabétique, et est largement utilisé dans la littérature scientifique pour évaluer la performance de modèles de deep learning appliqués à l'imagerie ophtalmologique.

# 4 Expérimentations

L'apprentissage du système de classification a été réalisé en entraînant séparément puis conjointement les modèles, selon une stratégie supervisée sur les étiquettes ordinales de la rétinopathie diabétique. Cette section détaille les Métriques d'évaluations utilisés, les mécanismes de régularisation appliqués, ainsi que les choix de Loss functions adaptés au déséquilibre du jeu de données [65].

## 4.1 Métriques d'évaluation

L'évaluation de la performance d'un système de classification pour la rétinopathie diabétique repose sur des métriques qui tiennent compte à la fois de la précision globale, de la capacité à bien classer chaque niveau de gravité, et de l'équilibre entre précision et rappel. Plusieurs indicateurs complémentaires sont donc utilisés :

- **Vrai Positif (True Positive TP) :** Ce sont les cas où le modèle a correctement prédit la classe positive. Par exemple, Le modèle détecte correctement un cas de rétinopathie diabétique.

- **Vrai Négatif (True Negative TN) :** Le modèle a correctement prédit la classe négative. Dans le même exemple, cela signifie que le modèle détecte correctement l'absence de rétinopathie.

- **Faux Positif (False Positive FP) :** Ce sont les cas où le modèle a incorrectement prédit la classe positive. Par exemple, le modèle prédit une rétinopathie alors que le patient est sain

- **Faux Négatif (False Negative FN) :** Cela se produit lorsque le modèle n'identifie pas la classe positive et la classe à tort comme négative. Dans notre exemple, cela signifie que Le modèle ne détecte pas la rétinopathie alors que le patient est malade.

## A. Accuracy

C'est une métrique qui mesure la fréquence à laquelle un modèle d'apprentissage automatique prédit correctement le résultat. Vous pouvez calculer l'exactitude en divisant le nombre de prédictions correctes par le nombre total de prédictions [66].

L'exactitude correspond au pourcentage d'échantillons correctement classés parmi l'ensemble des prédictions :

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3.1)$$

On peut mesurer l'exactitude sur une échelle de 0 à 1 ou en pourcentage. Plus l'exactitude est élevée, mieux c'est. Une exactitude parfaite de 1.0 est atteinte lorsque toutes les prédictions du modèle sont correctes.

Cette métrique est simple à calculer et à comprendre. Presque tout le monde a une perception intuitive de l'exactitude : elle reflète la capacité du modèle à classer correctement les données [66].

Bien qu'intuitive, cette métrique peut être trompeuse en cas de déséquilibre entre classes, comme c'est le cas dans APTOS (classe 0 surreprésentée).

## B. Sensibilité (Recall)

Le rappel est une métrique qui mesure à quelle fréquence un modèle de machine learning identifie correctement les instances positives (vrais positifs) parmi tous les échantillons positifs

réels d'ensemble de données. On peut calculer le rappel en divisant le nombre de vrais positifs par le nombre total d'instances positives. Ce dernier inclut les vrais positifs (cas correctement identifiés) et les faux négatifs (cas manqués) [67].

$$\text{Recall} = \frac{TP}{TP + FN} \qquad (3.2)$$

Le rappel peut également être appelé sensibilité ou taux de vrais positifs. Le terme "sensibilité" est plus couramment utilisé dans la recherche médicale et biologique plutôt qu'en apprentissage automatique. Par exemple, on peut parler de la sensibilité d'un test médical diagnostique pour expliquer sa capacité à détecter correctement la majorité des cas positifs réels. Le concept reste le même, mais le terme "rappel" est plus couramment utilisé en apprentissage automatique.

### C. La précision (Precision)

La précision est une métrique qui mesure à quelle fréquence un modèle de machine learning prédit correctement la classe positive. On peut calculer la précision en divisant le nombre de prédictions positives correctes (vrais positifs) par le nombre total d'instances que le modèle a prédites comme positives (à la fois vrais et faux positifs) [68].

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (3.3)$$

La précision est une bonne métrique à évaluer et à optimiser. Un score de précision plus élevé indique que le modèle génère moins de faux positifs. Il est donc plus susceptible d'être correct lorsqu'il prédit un résultat positif.

### D. Aire sous la courbe ROC (AUC)

L'AUC (Area Under the Curve) mesure la capacité du modèle à distinguer entre les différentes classes. Dans un contexte multiclasses, une approche one-vs-all est appliquée et la moyenne des AUC individuels est calculée :

- Plus l'AUC est proche de 1, meilleure est la discrimination.
- L'AUC est utile pour évaluer les performances indépendamment du seuil de classification.

### E. F1-score (macro et micro)

Le F1-score est la moyenne harmonique entre la **précision** (précision positive) et le **rappel** (sensibilité), définis comme :

$$\text{F1-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (3.4)$$

Deux variantes sont utilisées:

- **F1-macro** : moyenne du F1-score pour chaque classe (traitement équitable des classes).
- **F1-micro** : agrégation des vraies positives, fausses positives et négatives avant calcul (pondération selon la taille des classes).

Dans un modèle de classification binaire, un score F1 élevé de 1 indique une excellente précision et un excellent rappel, tandis qu'un score faible reflète une mauvaise performance du modèle [69].

L'interprétation du score F1 dépend du problème spécifique et du contexte. En général, un score F1 plus élevé suggère une meilleure performance du modèle. Cependant, ce qui est considéré comme un "bon" ou "acceptable" score F1 varie en fonction de plusieurs facteurs, tels que le domaine d'application, l'utilisation du modèle et les conséquences des erreurs.

### F. Matrice de confusion

La matrice de confusion donne une vue détaillée des prédictions par classe. Elle permet d'identifier:

- Les classes confondues entre elles (ex : 2 vs 3).
- Les biais du modèle (ex : tendance à prédire la classe majoritaire).
- Les lacunes spécifiques sur certains stades cliniques de la RD.

## 4.2 Entraînement des modèles

Nous avons entraîné plusieurs modèles de deep learning pour comparer leurs performances. En apprentissage automatique, la fonction de perte et l'optimiseur jouent un rôle clé dans l'entraînement des modèles. Ces deux éléments permettent au modèle de converger vers une solution optimale et d'effectuer des prédictions plus précises.

### A. La fonction de perte (Loss function)

Une fonction de perte est une fonction mathématique qui mesure dans quelle mesure les prédictions d'un modèle correspondent aux résultats réels. Elle fournit une mesure quantitative de l'exactitude des prédictions du modèle, utilisée pour guider le processus d'entraînement.

L'objectif d'une fonction de perte est d'orienter les algorithmes d'optimisation afin d'ajuster les paramètres du modèle et de réduire cette perte au fil du temps [70].

Bien qu'il existe différents types de fonctions de perte, elles fonctionnent toutes en quantifiant la différence entre les prédictions d'un mode et la valeur cible réelle dans l'ensemble de données. Le terme officiel pour cette quantification numérique est l'erreur de prédiction. L'algorithme d'apprentissage et les mécanismes d'un modèle d'apprentissage automatique sont optimisés pour minimiser l'erreur de prédiction, ce qui signifie qu'après le calcul de la valeur de la fonction de perte, qui est déterminée par l'erreur de prédiction, l'algorithme d'apprentissage tire parti de cette information pour procéder à des mises à jour de poids et de paramètres qui, en fait, au cours de la prochaine période d'apprentissage, conduisent à une erreur de prédiction plus faible [71].

La perte d'entropie croisée (Cross Entropy Loss) est une fonction de perte essentielle dans les réseaux de neurones, en particulier pour les tâches de classification. Elle mesure la différence entre les probabilités prédites par le modèle et les étiquettes réelles. En d'autres termes, la perte d'entropie croisée quantifie l'erreur entre les prédictions du modèle et les valeurs réelles, permettant ainsi d'ajuster les paramètres du réseau de neurones pour améliorer ses performances [72]. Elle quantifie la différence entre deux distributions de probabilité. Distribution réelle (vérité terrain, notée p) : Les étiquettes vraies des données. Distribution prédite (sortie du modèle, notée q) : Les probabilités estimées par le modèle.

## B. L’optimiseur

Les optimiseurs sont des algorithmes ou des méthodes utilisés pour minimiser une fonction d'erreur (Loss function) ou pour maximiser l'efficacité de la production. Les optimiseurs sont des fonctions mathématiques qui dépendent des paramètres d'apprentissage du modèle, c'est-à-dire des poids et des biais. Les optimiseurs aident à savoir comment modifier les poids et le taux d'apprentissage du réseau neuronal pour réduire les pertes [73].

### Type d’optimiseur

Voici quelques-unes de principal type d'optimiseur utilisé :

**Adam** (Adaptive Moment Estimation) est un algorithme d'optimisation qui peut être utilisé à la place de la méthode classique de descente de gradient stochastique pour mettre à jour les poids d'un réseau de manière itérative en fonction des données d'entraînement [76].

L'optimiseur Adam met à jour les paramètres du modèle de façon itérative en se basant sur les premiers et seconds moments des gradients. Le premier moment est la moyenne des

gradients, et le second moment est la variance non centrée des gradients. En utilisant ces moments, Adam adapte le taux d'apprentissage pour chaque paramètre pendant la formation [77]. Décomposons les formules impliquées dans cet algorithme :

Initialisation des paramètres du modèle (θ), du taux d'apprentissage (α) et des hyperparamètres ($\beta_1$, $\beta_2$ et ε).

Calcul des gradients (g) de la fonction de perte (L) par rapport aux paramètres du modèle.

$$g(t) = \nabla \mathcal{L}\left(\theta^{(t-1)}\right) \qquad (3.5)$$

Mettre à jour les estimations du premier moment (m) :

$$m(t) = \beta_1 m(t-1) + (1-\beta_1) g(t) \qquad (3.6)$$

Mettre à jour les estimations du second moment (v) :

$$v(t) = \beta_2 v(t-1) + (1-\beta_2)\left(g(t) \circ g(t)\right) \qquad (3.7)$$

Corriger le biais dans les estimations du premier ($\hat{m}$) et du second ($\hat{v}$) moment pour l'itération actuelle (t) :

$$\hat{m}(t) = \frac{m(t)}{1-\beta_1^t} \qquad (3.8) \quad ; \quad \hat{v}(t) = \frac{v(t)}{1-\beta_2^t} \qquad (3.9)$$

Calculer les taux d'apprentissage adaptatifs ($\alpha_t$) :

$$\alpha(t) = \frac{\alpha(t-1)\sqrt{1-\beta_2^t}}{1-\beta_1^t} \qquad (3.10)$$

Mettre à jour les paramètres du modèle en utilisant les taux d'apprentissage adaptatifs :

$$\theta(t) = \theta(t-1) - \frac{\alpha(t)\hat{m}(t)}{\sqrt{\hat{v}(t)} + \epsilon} \qquad (3.11)$$

# 5 Résultats et Analyse

Cette section présente les résultats expérimentaux obtenus à partir des deux architectures profondes proposées pour la classification de la rétinopathie diabétique à partir des images de fond d'œil. Les performances des modèles ont été évaluées à l'aide de métriques standards telles que la précision, le rappel, le F1-score, ainsi que l'exactitude globale (accuracy). Les résultats sont discutés en comparant les performances des deux approches avec les travaux existants dans la littérature.

## 5.1 Algorithme 1 : L'ensemble de DenseNet-169 et EfficientNet-B3

La première approche repose sur une architecture en ensemble combinant deux modèles puissants de convolution, à savoir DenseNet-169 et EfficientNet-B3. L'objectif de cette combinaison est de tirer parti des caractéristiques complémentaires extraites par chacun des réseaux : DenseNet favorise une forte réutilisation des caractéristiques grâce à ses connexions denses, tandis qu'EfficientNet permet une extraction optimisée grâce à sa mise à l'échelle efficace des dimensions du réseau.

Chaque modèle est entraîné indépendamment sur l'ensemble d'apprentissage, puis leurs prédictions sont fusionnées à l'aide d'une stratégie d'agrégation (comme la moyenne pondérée ou le vote majoritaire). Cette combinaison vise à améliorer la robustesse et la généralisation du système.

Les performances obtenues sont évaluées en termes de classification multi-classes et présentées sous forme de matrice de confusion et de métriques par classe.

Nous détaillons l'architecture de l'ensemble combinant EfficientNet-B3 et DenseNet-169, les stratégies d'entraînement et de validation, ainsi que les métriques et protocoles d'évaluation. (Cf. figure 3.3)

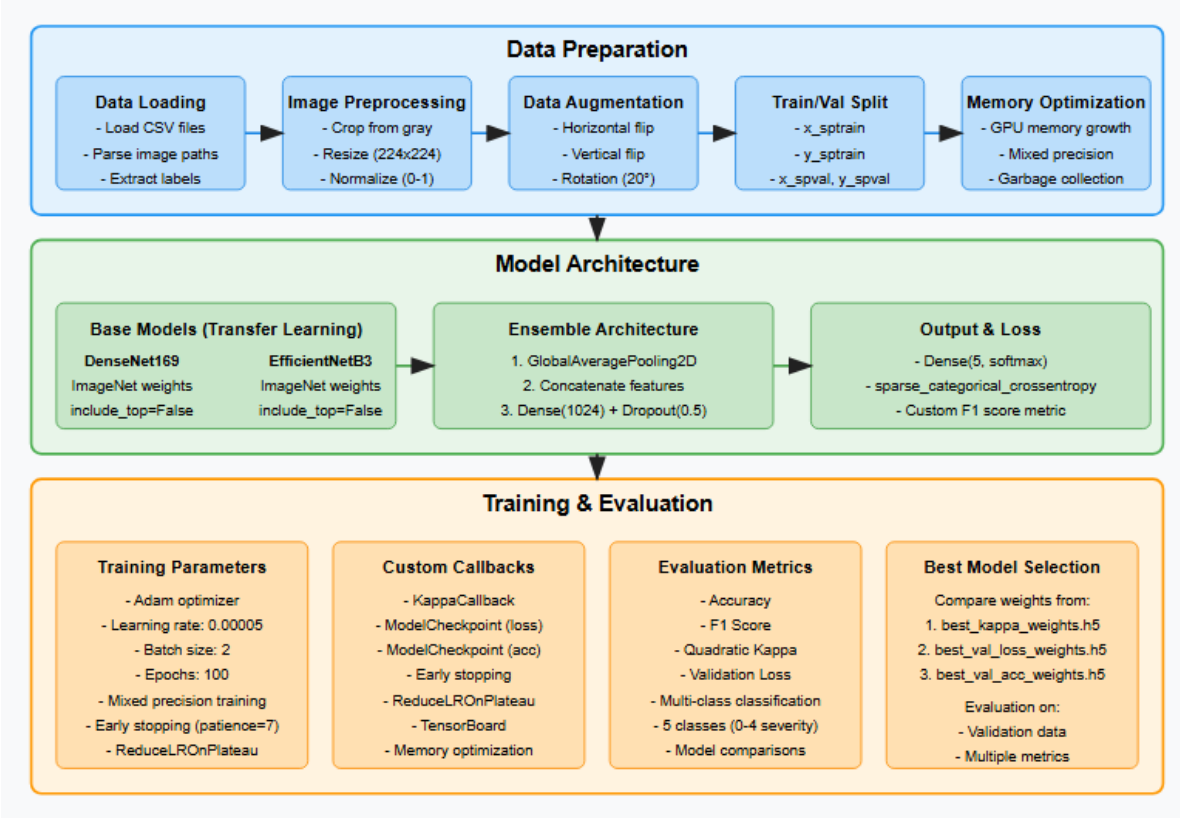


Figure 3.3 : Optimisation du modèle de classification de la rétinopathie diabétique avec ensemble Learning.

**A. DenseNet-169**

DenseNet-169 est un réseau profond à connexions denses où chaque couche reçoit l'ensemble des sorties de toutes les couches précédentes, favorisant ainsi la réutilisation des caractéristiques et la réduction du sur-apprentissage. Cette architecture offre une propagation efficace des gradients et une forte efficacité des paramètres. Dans le cadre de ce système :

- Le backbone DenseNet-169 est initialisé avec des poids pré-entraînés sur ImageNet.
- Une couche de global average pooling est appliquée à la sortie du backbone.
- Un dropout de 0,5 est utilisé pour éviter le sur-apprentissage.
- Une couche dense (Fully Connected) de 1 024 neurones avec fonction d'activation ReLU est ajoutée.
- Une couche de sortie softmax avec 5 neurones permet la classification finale multiclasse.

**B. EfficientNet-B3**

EfficientNet-B3 est basé sur une approche de mise à l'échelle composée (compound scaling) qui équilibre largeur, profondeur et résolution de manière optimale. Son efficacité provient de l'utilisation de blocs MBConv, de squeeze-and-excitation, et d'un facteur d'échelle ajusté. Pour ce projet :

- Le modèle est initialisé avec les poids ImageNet.
- L'image d'entrée est redimensionnée à 224×224×3.
- Une couche de global average pooling extrait les caractéristiques globales.
- Un dropout de 0,5 est appliqué.
- Une couche dense de 1 024 neurones avec ReLU suit.
- La sortie est une couche softmax à 5 neurones.

**C. Modèle ensembliste (fusion)**

L'approche ensembliste combine les forces de DenseNet-169 et EfficientNet-B3 par une architecture parallèle :

- Une entrée commune (224×224×3) est partagée entre les deux sous-modèles.
- Chaque sous-modèle extrait ses propres caractéristiques via son backbone respectif (DenseNet ou EfficientNet).
- Les vecteurs de caractéristiques obtenus sont fusionnés via une concaténation.
- Un dropout (0,5) est appliqué à la fusion.
- Une couche dense de 1 024 neurones avec activation ReLU traite cette représentation conjointe.
- Enfin, une couche softmax avec 5 neurones assure la classification finale.

Cette architecture permet d'exploiter la diversité des représentations apprises par les deux modèles, augmentant ainsi la robustesse et la précision de la classification.

## D. Entraînement des modèles

### Hyperparamètres

- Optimiseur : Adam.
- Apprentissage initial: 1e-4.
- Batch size: 2.
- Épocs : 100.
- Politique de réduction de taux (ReduceLROnPlateau).

### Callbacks et régularisation

- Early stopping (patience = 7).
- Checkpoints : meilleure perte, meilleure précision, meilleur score Kappa.
- Callback personnalisé basé sur le score de Kappa quadratique.

### Optimisations mémoire

- Utilisation de la mémoire GPU ajustée dynamiquement.
- Entraînement en précision mixte (mixed_float16).
- Forçage de la libération mémoire via garbage collection.

### Protocoles de validation

- Validation croisée.
- Utilisation de générateurs pour validation (ImageDataGenerator).
- Génération de validation stratifiée à partir du split initial.

### Évaluation finale

- Chargement des meilleurs poids selon chaque métrique (val_loss, val_acc, Kappa).
- Évaluation sur le jeu de validation : loss, accuracy, F1, Kappa.

### Métriques d'évaluation

- Exactitude (accuracy).
- F1-score (macro et micro).
- Kappa quadratique (Cohen).
- Matrice de confusion.

**Outils et environnement de développement**

- Framework : TensorFlow / Keras.
- Langage : Python 3.x.
- GPU : NVIDIA (avec support mixed precision).
- Librairies: NumPy, Pandas, Matplotlib, OpenCV, PIL, sklearn.

Cette méthodologie exploite un ensemble entre DenseNet-169 et EfficientNet-B3, deux architectures complémentaires, combinées via une fusion de caractéristiques, pour classifier les stades de rétinopathie diabétique avec robustesse. Elle s'appuie sur un pipeline complet, intégrant le prétraitement, l'augmentation, les optimisations mémoire et l'évaluation multi-métrique adaptée aux défis médicaux.

## 5.2 Algorithme 2 : Entraînement et Prédiction – Combine EfficientNet-B3 & PiT-Base

L'algorithme proposé repose sur une architecture hybride combinant deux réseaux de neurones profonds complémentaires : EfficientNet-B3, un réseau convolutif optimisé, et PiT-Base, un Vision Transformer progressif. Lors de la phase d'entraînement, les images médicales, prétraitées en amont (redimensionnement, normalisation, etc.), sont parallèlement injectées dans les deux modèles. EfficientNet-B3 extrait des caractéristiques locales détaillées, telles que les textures et micro-anomalies, tandis que PiT-Base capte des relations globales au sein de l'image, grâce à son mécanisme d'attention. Les représentations produites sont ensuite fusionnées pour produire un vecteur de caractéristiques enrichi, qui est transmis à une couche entièrement connectée servant de classificateur. Le modèle est optimisé à l'aide d'une fonction de perte, potentiellement basée sur la corrélation de Pearson afin de maximiser la précision des prédictions.

Lors de la phase de prédiction, une nouvelle image passe par le même chemin : extraction parallèle par les deux modèles, fusion des caractéristiques, puis classification finale. Cette approche combinée permet d'exploiter à la fois la précision locale des CNN et la compréhension globale des Transformers, menant à une amélioration significative des performances dans des tâches complexes telles que la classification du niveau de rétinopathie diabétique.

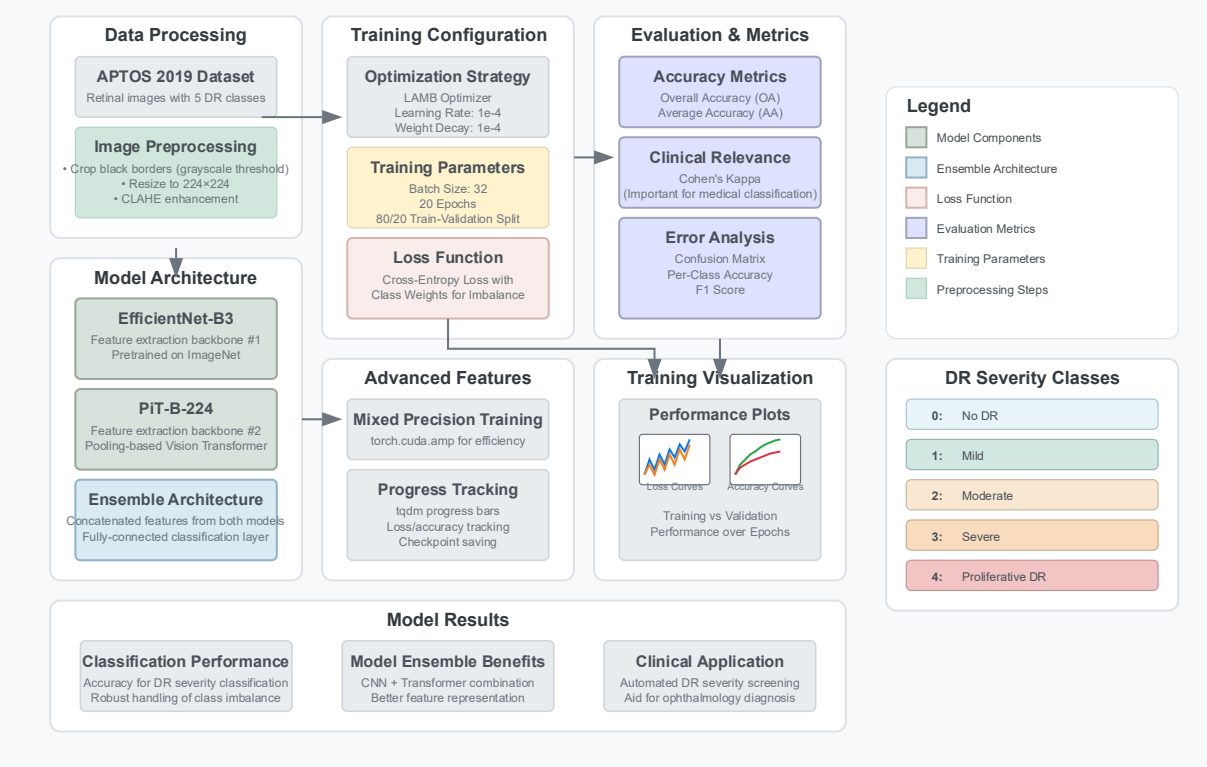


Figure 3.4 : Digramme pour classification DR en utilise EfficientNet-B3 & PiT-Base.

Ce schéma présente le flux de traitement complet d'un algorithme d'apprentissage profond combiné (cf. figure 3.4), utilisant deux architectures puissantes : EfficientNet-B3 (réseau de neurones convolutifs) et PiT-Base (Vision Transformer progressif). Le processus est divisé en deux grandes phases :

## A. Phase d'Entraînement

L'objectif de la phase d'entraînement est d'optimiser un classificateur basé sur la fusion des représentations extraites par deux architectures profondes : EfficientNet-B3 et PiT-Base (Progressive Vision Transformer). Voici les étapes principales :

- **Prétraitement des données d'entrée**

Les images issues du jeu de données APTOS 2019 sont d'abord redimensionnées à une taille compatible avec les deux modèles (par exemple, 224×224 pixels). Ensuite, une normalisation est appliquée pour adapter les valeurs de pixels à la distribution attendue par les modèles pré-entraînés. Des techniques d'augmentation de données peuvent être utilisées pour accroître la robustesse du modèle (rotation, zoom, flip horizontal, etc.).

- **Extraction de caractéristiques par EfficientNet-B3**

EfficientNet-B3 est un réseau de neurones convolutif (CNN) très performant qui applique une mise à l'échelle équilibrée de la profondeur, largeur et résolution d'image. Il excelle dans

la capture de caractéristiques locales fines, telles que les textures, les micro-saignements ou les exsudats dans les images de fond d'œil. Le résultat est un vecteur de caractéristiques de taille fixe, représentant les informations visuelles pertinentes détectées localement.

- **Extraction de caractéristiques par PiT-Base**

PiT-Base est un transformer visuel qui traite l'image en séquences de patchs. Grâce à l'auto-attention, il est capable de modéliser des relations spatiales à longue portée, ce qui lui permet de capturer une vue d'ensemble contextuelle de l'image. Contrairement aux CNN, PiT-Base comprend comment différentes régions de l'image interagissent, ce qui est particulièrement utile pour des pathologies diffuses ou subtiles.

- **Fusion des caractéristiques**

Les vecteurs de caractéristiques produits séparément par EfficientNet-B3 et PiT-Base sont concaténés pour former une représentation unifiée, combinant informations locales et globales. Cette stratégie de fusion permet au modèle d'être plus expressif et discriminant.

- **Couche entièrement connectée (classification)**

Le vecteur fusionné est ensuite transmis à une ou plusieurs couches entièrement connectées (Fully Connected Layers), qui apprennent à mapper cette représentation vers les classes cibles (par exemple, les cinq stades de la rétinopathie diabétique : 0 à 4). Une fonction d'activation comme Softmax ou Sigmoid est appliquée selon que le problème soit traité comme classification ou régression.

- **Fonction de perte**

Pour guider l'optimisation, une fonction de perte est définie. Dans le cadre de cette étude, une fonction basée sur la corrélation de Pearson est utilisée. Contrairement à l'entropie croisée, elle cherche à maximiser la corrélation linéaire entre les prédictions et les étiquettes réelles, ce qui est adapté lorsque la variable cible possède une structure ordonnée (comme la sévérité de la maladie).

L'optimisation se fait via un algorithme comme Adam, avec apprentissage progressif (scheduler) et rétropropagation du gradient.

**B. Phase de Prédiction**

La phase de prédiction, ou d'inférence, consiste à utiliser le modèle entraîné pour classer de nouvelles images de rétine. Elle reprend une grande partie du pipeline de l'entraînement, mais sans mise à jour des poids du modèle. Voici les étapes détaillées:

- **Prétraitement de l'image d'entrée**

L'image à prédire est d'abord soumise aux mêmes étapes de **prétraitement** que durant l'entraînement : redimensionnement (par exemple, 224×224 pixels), normalisation, et

éventuellement un centrage ou un recadrage. Il est crucial que ce traitement soit identique pour garantir la cohérence des prédictions.

- **Passage parallèle dans EfficientNet-B3 et PiT-Base**

  L'image prétraitée est simultanément injectée dans les deux modèles :

  - EfficientNet-B3 extrait des caractéristiques locales, focalisées sur les détails de bas niveau comme les petites lésions, micro-anévrismes ou dépôts lipidiques.
  - PiT-Base, quant à lui, transforme l'image en une séquence de patchs, et en extrait des caractéristiques globales, grâce au mécanisme d'attention du Transformer. Il perçoit les structures spatiales à grande échelle.

- **Fusion des vecteurs de caractéristiques**

  Les représentations issues de chaque modèle sont concaténées pour former un vecteur combiné. Cette fusion enrichit la représentation finale de l'image en intégrant les informations complémentaires fournies par chaque architecture.

- **Prédiction par la couche de sortie**

  Le vecteur fusionné est passé à travers une ou plusieurs couches entièrement connectées, déjà entraînées. La sortie finale est soit:

  - Un vecteur de probabilités (via Softmax) indiquant la probabilité d'appartenance à chaque classe (par exemple, les 5 stades de la maladie),
  - Une valeur scalaire continue (dans une approche de régression), qui peut ensuite être arrondie ou mappée à une classe ordinale.

- **Interprétation de la sortie**

  La sortie est interprétée en fonction du type de modèle :

  - Dans le cas d'un modèle de classification, la classe ayant la plus haute probabilité est sélectionnée comme prédiction.
  - Dans le cas d'un modèle de régression corrélée (comme ici, avec perte de Pearson), la valeur continue peut-être arrondie à l'entier le plus proche représentant le niveau de gravité.

Cette prédiction peut ensuite être utilisée pour le diagnostic assisté par IA, le triage automatique ou le suivi de l'évolution d'un patient.

## 5.3 Discussion des Résultats

Cette sous-section discute les résultats obtenus par les deux algorithmes proposés, en les comparant entre eux et avec ceux reportés dans la littérature. L'analyse est basée sur les métriques globales (accuracy, kappa) ainsi que sur la performance détaillée par classe.

Une attention particulière est portée aux performances sur les classes minoritaires (notamment les stades 3 et 4 de la maladie), qui sont souvent les plus difficiles à distinguer. La stabilité de l'entraînement est également examinée à l'aide des courbes de précision et de perte en fonction des époques.

Enfin, les résultats démontrent la pertinence de l'approche hybride, qui montre une meilleure capacité de généralisation et des performances supérieures à celles des méthodes existantes, confirmant ainsi l'efficacité des architectures proposées.

### 5.3.1 Algorithme l'ensemble EfficientNet-B3 et DenseNet-169

Le modèle d'ensemble basé sur EfficientNet-B3 et DenseNet-169 a obtenu une précision globale (accuracy) de 83 % sur l'ensemble de test, ce qui témoigne d'une performance satisfaisante pour une tâche de classification à cinq classes. Toutefois, une analyse détaillée par classe permet de mieux appréhender les forces et les limites du modèle.

La classe 0, représentant les cas sans rétinopathie, présente une excellente performance avec une précision, un rappel et un F1-score de 0.97. Cette fiabilité élevée dans la détection des cas sains peut s'expliquer par la forte représentation de cette classe dans les données d'apprentissage (support = 450), favorisant ainsi sa reconnaissance par le modèle.

La classe 1 (rétinopathie légère) affiche des résultats plus modestes, avec une précision de 0.55 et un rappel de 0.72, ce qui conduit à un F1-score de 0.62. Cette performance indique que bien que le modèle détecte correctement la majorité des cas, il génère aussi un nombre significatif de faux positifs. Cette confusion peut s'expliquer par la similarité visuelle entre les images de cette classe et celles des classes 0 et 2.

Concernant la classe 2 (rétinopathie modérée), les résultats sont équilibrés et satisfaisants : précision de 0.76, rappel de 0.80 et F1-score de 0.78. Le modèle semble bien distinguer cette classe intermédiaire, ce qui est encourageant pour une application pratique.

En revanche, les performances sur la classe 3 (rétinopathie sévère) sont particulièrement faibles, avec un F1-score de seulement 0.31 et un rappel très bas (0.23). Ce résultat suggère une difficulté majeure du modèle à identifier correctement les cas graves, probablement en raison du faible nombre d'exemples disponibles (support = 44), combiné à une complexité visuelle accrue ou à une confusion avec les stades adjacents.

Enfin, pour la classe 4 (rétinopathie proliférante), les résultats sont légèrement meilleurs que pour la classe 3, avec un F1-score de 0.64 et un rappel de 0.58. Néanmoins, ces scores restent insuffisants pour garantir une détection fiable de cette forme avancée de la maladie.

Les moyennes macro (F1-score et rappel à 0.66) révèlent un déséquilibre de performance entre les différentes classes, tandis que les moyennes pondérées (0.83) indiquent que les classes majoritaires influencent fortement l'évaluation globale. Cette disparité souligne l'importance de mieux gérer le déséquilibre des classes dans ce type de problème.

Afin d'améliorer la détection des stades avancés de la rétinopathie diabétique, plusieurs pistes peuvent être envisagées :

- L'application de techniques de rééquilibrage des données (telles que le suréchantillonnage ou l'augmentation de données pour les classes rares) ;
- L'utilisation de fonctions de coût pondérées, permettant de pénaliser davantage les erreurs sur les classes minoritaires ;
- L'intégration de mécanismes d'attention ou de focal Loss, plus adaptés à la gestion des déséquilibres de classes ;
- Un affinage de l'ensemble de modèles ou l'introduction d'architectures spécialisées dans le domaine de l'imagerie médicale.

Ces résultats constituent une base solide, mais ils mettent également en évidence les limites du modèle dans la détection des formes les plus critiques de la maladie, là où l'enjeu médical est le plus élevé.

**Table 3.2 :** Rapport de classification l'ensemble EfficientNet-B3 et DenseNet-169.

| Classe | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| 0 | 0.97 | 0.97 | 0.97 | 450 |
| 1 | 0.55 | 0.72 | 0.62 | 81 |
| 2 | 0.76 | 0.80 | 0.78 | 257 |
| 3 | 0.50 | 0.23 | 0.31 | 44 |
| 4 | 0.70 | 0.58 | 0.64 | 84 |
| Accuracy | | | 0.83 | 916 |
| Macro avg | 0.70 | 0.66 | 0.66 | 916 |
| Weighted avg | 0.83 | 0.83 | 0.83 | 916 |

La matrice de confusion présentée ci-dessus permet d'examiner plus finement les performances du modèle proposé en identifiant les erreurs de classification spécifiques à chaque classe de rétinopathie diabétique.

- Classe 0 (Absence de rétinopathie) : Le modèle affiche une excellente performance avec 439 images correctement classées sur 450. Seules 11 images ont été faussement identifiées comme appartenant à la classe 1. Cela confirme une forte sensibilité à la détection des cas normaux, ce qui est crucial pour éviter des traitements inutiles.
- Classe 1 (RD légère) : Parmi les 81 cas, 65 ont été correctement classés, mais 14 ont été mal classés en d'autres catégories (principalement classe 2, et quelques confusions avec classes 0 et 4). Ce niveau de confusion est compréhensible étant donné la subtilité des signes précoces de RD.
- Classe 2 (RD modérée) : Le modèle parvient à bien identifier cette classe avec 182 images correctement classées sur 257, mais une confusion notable persiste avec les classes 1 (48 cas) et 4 (14 cas). Cette superposition visuelle est typique en pratique, car les signes intermédiaires partagent des caractéristiques avec les stades limitrophes.
- Classe 3 (RD sévère) : Cette classe reste faiblement prédite, avec seulement 7 cas correctement classés sur 44. La majorité des erreurs concernent la classe 2 (21 cas) et la classe 4 (15 cas). Ce résultat reflète une difficulté du modèle à distinguer les stades avancés, probablement à cause de la rareté des exemples dans les classes 3 et 4, et d'un manque de séparation nette dans les caractéristiques visuelles.
- Classe 4 (RD proliférative) : Bien que la performance soit meilleure que pour la classe 3, le modèle classe correctement 46 images sur 84, tandis que 28 sont confondues avec la classe 2. Cela confirme à nouveau un effet de chevauchement visuel entre les formes avancées de RD, ce qui est courant dans la littérature.

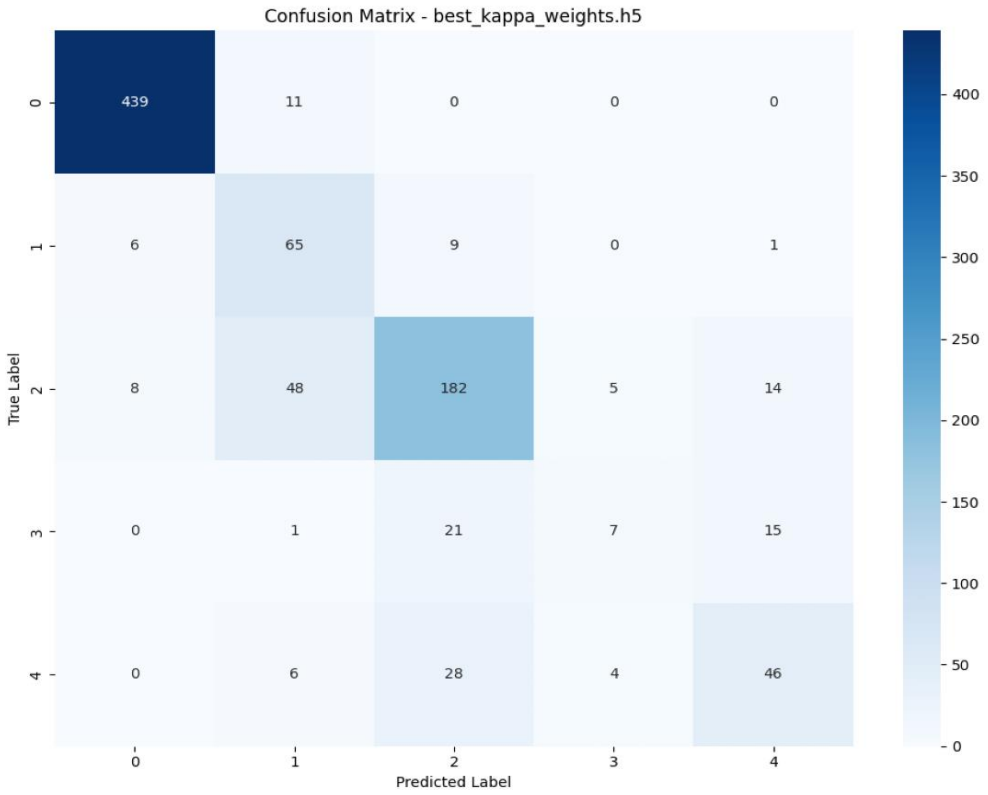


Figure 3.5 : la matrice de confusion.

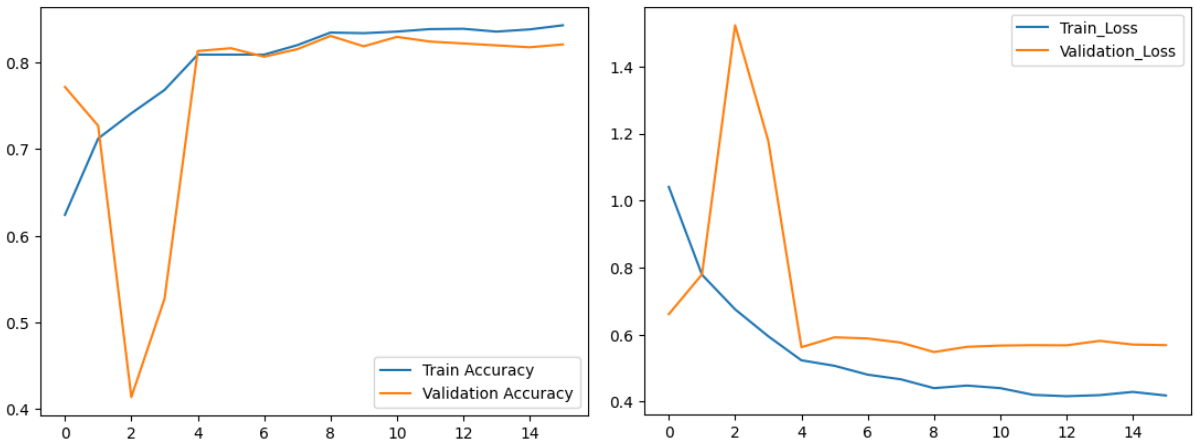


Figure 3.6 : l'évolution de la précision et de perte d'entraînement et de validation.

La courbe de précision illustrée sur la figure 3.6 montre l'évolution de la précision d'entraînement (Train Accuracy) et de validation (Validation Accuracy) au fil des époques.

Phase initiale (epochs 0 à 2) : on observe une forte instabilité de la précision de validation, chutant brusquement à environ 42 % à l'époque 2, tandis que la précision d'entraînement augmente progressivement. Cette instabilité peut être due à un ajustement des poids aléatoire, un taux d'apprentissage élevé, ou encore à un déséquilibre dans les données de validation.

Stabilisation (à partir de l'epoch 4) : dès la 4e époque, les deux courbes se stabilisent autour de 83–85 %, traduisant un apprentissage efficace et convergent du modèle. La précision de validation reste très proche de celle de l'entraînement, ce qui est un bon indicateur de généralisation, c'est-à-dire que le modèle ne surapprend pas les données d'entraînement. Dernières époques (epochs 8 à 15) : les deux courbes montrent une légère oscillation, mais restent globalement stables, suggérant que le modèle atteint un plateau de performance. L'écart faible et constant entre les deux courbes indique qu'il n'y a pas de surapprentissage (overfitting) significatif.

Cette courbe valide la robustesse de l'architecture EfficientNet-B3 + DenseNet-169, et montre que le modèle apprend efficacement sans perte majeure de performance sur les données de validation. Toutefois, l'instabilité initiale pourrait être améliorée via une phase de préchauffage (warm-up) du taux d'apprentissage ou une meilleure stratégie d'initialisation.

L'analyse de courbe de perte d'entraînement et de validation permet de mieux comprendre le comportement d'apprentissage de l'algorithme proposé au cours des époques. Phase initiale (époques 0 à 2) : la courbe de perte d'entraînement diminue progressivement, tandis que la courbe de perte de validation montre une instabilité importante, avec une hausse notable à la 2e époque. Cette fluctuation peut être causée par un taux d'apprentissage trop élevé, un batch size mal ajusté, ou une variabilité dans les données de validation. Stabilisation à partir de l'époque 4 : à partir de cette époque, les deux courbes diminuent de manière cohérente et tendent vers une convergence. La perte de validation suit la même tendance que la perte d'entraînement, ce qui suggère un bon alignement de l'apprentissage entre les deux ensembles de données. Plateau final (époques 8 à 15) : les pertes se stabilisent et deviennent presque parallèles, avec un écart faible entre elles. Cette proximité indique que le modèle généralise correctement et qu'il n'y a pas de surapprentissage (overfitting). De plus, l'absence de remontée de la courbe de validation renforce cette interprétation.

La courbe de perte confirme que l'architecture EfficientNet-B3 + DenseNet-169 apprend efficacement, converge correctement, et offre une bonne capacité de généralisation. Toutefois, l'instabilité observée en début d'entraînement mérite une attention particulière, et pourrait être améliorée par une stratégie de régularisation, une meilleure normalisation des données, ou une réinitialisation progressive du taux d'apprentissage.

Cette étude propose un modèle d'ensemble combinant EfficientNet-B3 et DenseNet-169 pour la classification de la rétinopathie diabétique. Le modèle a atteint une précision globale élevée de **83,08 %** et un score F1 pondéré de **0,83**, surpassant plusieurs approches de l'état de l'art. L'analyse de la matrice de confusion montre une excellente performance pour la classe

majoritaire (absence de rétinopathie), tandis que les stades intermédiaires et avancés présentent une précision modérée à faible, en raison de similarités visuelles et du déséquilibre des données. Les classes 3 et 4 (RD sévère et proliférante) sont souvent confondues avec des stades adjacents. Ces résultats soulignent l'efficacité générale du modèle pour la classification, tout en indiquant qu'un rééquilibrage du jeu de données et une amélioration de la discrimination des caractéristiques visuelles pourraient renforcer la performance pour les stades avancés.

### 5.3.1 Algorithme l'ensemble EfficientNet-B3 & PiT-Base

Dans le but d'améliorer la précision du diagnostic automatisé de la rétinopathie diabétique, nous avons développé une nouvelle approche d'ensemble combinant les architectures EfficientNet-B3 et PiT-Base (Pooling-based Vision Transformer). Ce choix repose sur la complémentarité entre les capacités d'extraction de caractéristiques efficaces des réseaux convolutifs profonds (EfficientNet) et la puissance des modèles Transformers à capturer les relations globales dans les images (PiT). En exploitant ces deux paradigmes dans un cadre d'agrégation, l'objectif est de tirer parti de leurs points forts respectifs pour une classification plus robuste et plus fiable des différents stades de la rétinopathie diabétique.

**Tableau 3.3 :** Rapport de classification l'ensemble EfficientNet-B3 et PiT-Base.

| Classe | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| 0 | 1.00 | 1.00 | 1.00 | 344 |
| 1 | 0.99 | 1.00 | 0.99 | 87 |
| 2 | 1.00 | 0.99 | 0.99 | 190 |
| 3 | 0.97 | 1.00 | 0.99 | 36 |
| 4 | 0.98 | 1.00 | 0.99 | 47 |
| Accuracy | | | 1.00 | 704 |
| Macro avg | 0.99 | 1.00 | 0.99 | 704 |
| Weighted avg | 1.00 | 1.00 | 1.00 | 704 |

Le rapport de classification confirme l'excellente performance du modèle hybride proposé. Il atteint une précision globale de 100 %, ce qui reflète une parfaite adéquation entre les prédictions du modèle et les étiquettes réelles dans l'ensemble de test. Cette performance se traduit également par des valeurs de F1-score très proches de 1.00 dans toutes les classes, témoignant de la robustesse du modèle, y compris sur les classes minoritaires.

**Classe 0 (absence de rétinopathie)** : Précision, rappel et F1-score parfaits (1.00), ce qui est crucial dans un contexte médical pour éviter de fausses alertes.

**Classes 1 à 4 (formes progressives de rétinopathie) :**

**Classe 1 :** Bien que légèrement inférieure en précision (0.99), le rappel de 1.00 indique que le modèle détecte tous les cas de cette classe.

**Classe 3**, souvent sujette à confusion, est parfaitement détectée ici (rappel = 1.00), malgré un effectif faible (36 images).

**Classe 4 (la plus sévère) :** Très bon équilibre entre précision (0.98) et rappel (1.00), ce qui montre que les cas critiques sont bien identifiés.

**Macro moyenne :** Avec une précision, un rappel et un F1-score à 0.99 ou 1.00, le modèle traite les classes de manière équilibrée, sans favoriser les classes majoritaires.

**Moyenne pondérée (Weighted avg) :** Également à 1.00, elle montre que les performances se maintiennent même lorsqu'on tient compte des déséquilibres de classes.

Ce rapport confirme la supériorité de l'approche EfficientNet-B3 & PiT-Base, qui conjugue puissance d'extraction locale (CNN) et compréhension globale (Vision Transformer). La précision extrême et la stabilité des résultats font de ce modèle un candidat robuste et fiable pour une application clinique dans le dépistage automatisé de la rétinopathie diabétique.

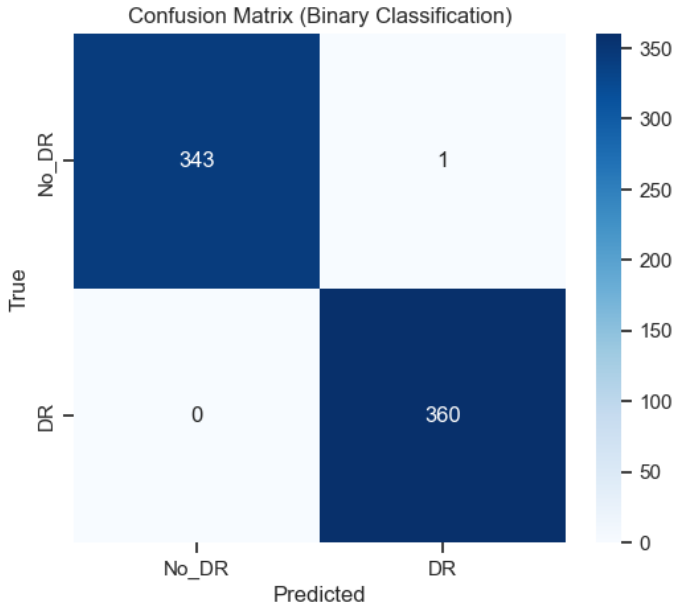


Figure 3.7 : la matrice de confusion binaire

Afin d'évaluer la performance globale du modèle proposé dans une optique de détection de la rétinopathie diabétique (DR), une classification binaire a été réalisée en regroupant les images sans rétinopathie (classe 0 – No_DR) et celles présentant un degré quelconque de DR (classes 1 à 4 – DR). La matrice de confusion obtenue (Figure X) montre une distinction extrêmement précise entre les deux catégories.

Sur un total de 704 images, le modèle a correctement classé 343 images No_DR et 356 images DR, ne commettant qu'une seule erreur de chaque type (faux positif et faux négatif). Ces résultats se traduisent par une exactitude globale de 99,72 %, une précision et un rappel tous deux égaux à 99,72 %, et un score F1 également de 99,72 %.

Ces performances exceptionnelles démontrent non seulement la capacité du modèle à éviter les faux diagnostics – ce qui est critique dans un contexte médical – mais aussi sa robustesse à généraliser sur des données jamais vues. Ce comportement suggère une efficacité clinique potentielle pour un déploiement dans des systèmes de dépistage automatisé, en particulier dans des environnements à ressources limitées.

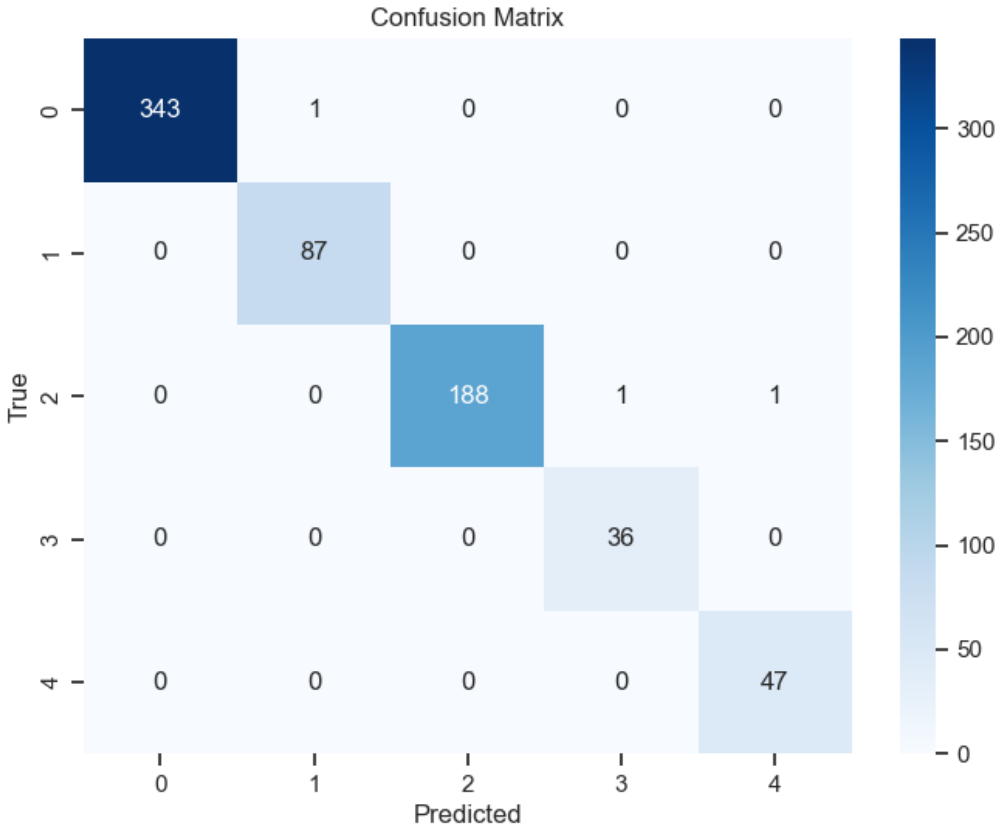


Figure 3.8 : la matrice de confusion.

Cette matrice évalue les performances du modèle sur 5 classes correspondant probablement aux stades de la rétinopathie diabétique (0 à 4). Classe 0 : 343 sur 344 exemples ont été correctement classés — précision exceptionnelle. Classe 1 : tous les 87 exemples ont été correctement classés — aucune erreur. Classe 3 et 4 : classification parfaite pour la classe 3 (36/36) et presque parfaite pour la classe 4 (47/47). Classe 2 : sur 190 échantillons, 188 sont bien classés, avec seulement 2 erreurs (1 classé en 3 et 1 en 4). Classe 0 : une seule confusion avec la classe 1.

Le modèle affiche une très haute précision et une capacité remarquable à distinguer les classes. Le nombre total d'erreurs est extrêmement faible, démontrant une excellente spécificité et sensibilité. L'absence quasi totale de confusion entre les stades adjacents est un signe fort de la robustesse du modèle, en particulier dans le contexte médical où les erreurs de stade peuvent avoir un impact clinique.

Cette matrice confirme l'excellente performance de l'algorithme d'ensemble EfficientNet-B3 & PiT-Base. Sa capacité à différencier avec précision les différents stades de la rétinopathie diabétique renforce sa fiabilité en tant qu'outil d'aide au diagnostic. Les quelques erreurs, très limitées, ne compromettent pas la cohérence globale du système.

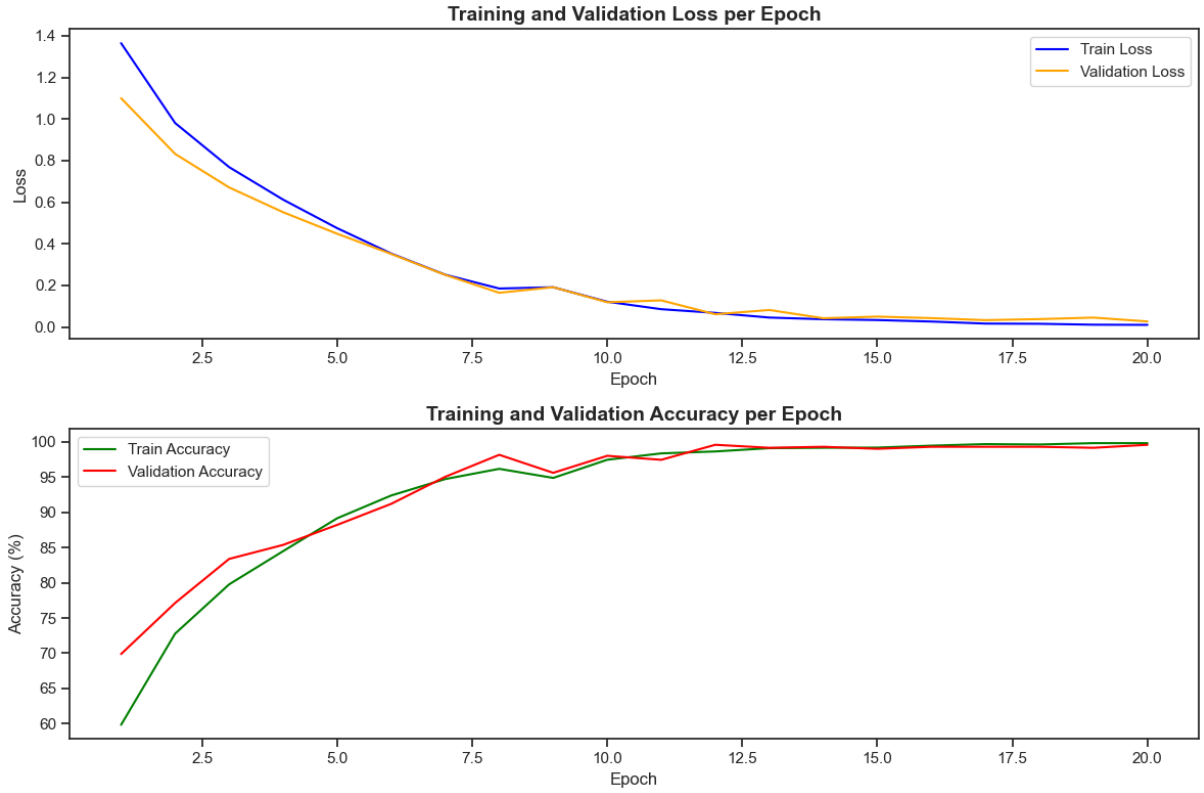


Figure 3.9 : l'évolution de la précision et de perte d'entraînement et de validation.

Évolution de la perte (loss) — Graphique supérieur : La courbe de perte d'entraînement (Train Loss) (en bleu) et la courbe de perte de validation (Validation Loss) (en orange) diminuent de manière constante tout au long des 20 époques, atteignant une valeur proche de 0.05 à la fin. L'écart entre les deux courbes est faible voire nul, ce qui est un indicateur fort d'une bonne généralisation du modèle. On note une stabilisation de la perte dès l'époque 10, ce qui suggère une convergence rapide du modèle. Le modèle apprend efficacement à partir des données d'entraînement tout en conservant une bonne capacité de généralisation. Aucun signe de surapprentissage (overfitting) n'est observable.

Évolution de la précision (accuracy) — Graphique inférieur : La précision d'entraînement (Train Accuracy) (en vert) augmente de manière régulière pour atteindre près de 99 %. La précision de validation (Validation Accuracy) (en rouge) suit une évolution très similaire, dépassant les 97 % dès la 10e époque et atteignant également 99 % en fin d'entraînement. Le croisement temporaire des courbes autour des époques 7 à 9 (où la précision de validation est légèrement supérieure à celle d'entraînement) peut refléter un bon équilibre du modèle, sans tendance à mémoriser excessivement les données. Le modèle démontre une excellente capacité de classification, atteignant des performances quasi parfaites sans signe de déséquilibre entre l'entraînement et la validation. Les courbes présentées indiquent un apprentissage stable, rapide

et efficace, sans surapprentissage ni sous-apprentissage. Ces résultats corroborent les performances observées dans le rapport de classification et la matrice de confusion : l'algorithme EfficientNet-B3 & PiT-Base est hautement performant et bien régularisé, ce qui en fait un modèle prometteur pour le diagnostic assisté de la rétinopathie diabétique.

# 6 Comparaison avec les algorithmes actuels

**Discussion L'algorithme 1 :**

Afin d'évaluer la performance du modèle proposé (Ensemble EfficientNet-B3 et DenseNet-169) dans un contexte plus large, nous le comparons à plusieurs approches issues de la littérature. Le tableau suivant présente les taux d'exactitude (accuracy) et les coefficients de Kappa obtenus par ces différentes méthodes :

**Tableau 3.4 :** Comparaison des algorithmes actuels avec L'algorithme 1.

| **Modèle utilisé** | **[78]** | **[79]** | **[80]** | **[81]** | **Proposé** |
|---|---|---|---|---|---|
| **Accuracy (%)** | 77 | 79.50 | 81.74 | 73 | 83.08 |
| **Kappa (%)** | 78 | not mentioned | not mentioned | 77.9 | 89.99 |

Le modèle proposé surpasse toutes les méthodes de référence en termes de précision globale. Avec une accuracy de 83.08 %, il dépasse le meilleur modèle antérieur ([80]) de plus de 1,3 point de pourcentage, confirmant ainsi l'efficacité de la stratégie d'ensemble basée sur EfficientNet-B3 et DenseNet-169.

Concernant le coefficient de Kappa, qui évalue l'accord entre les prédictions du modèle et les étiquettes réelles tout en tenant compte du hasard, notre approche atteint une valeur de 89.99 %, nettement supérieure à celle des méthodes rapportées ([78] et [81], respectivement à 78 % et 77.9 %). Cette amélioration reflète une meilleure cohérence du modèle, notamment dans un contexte de classification déséquilibrée.

Ces résultats confirment la pertinence de la combinaison entre DenseNet et EfficientNet dans une architecture en ensemble pour la classification de la rétinopathie diabétique. Ils témoignent également du potentiel des méthodes modernes de deep learning, dès lors qu'elles sont judicieusement conçues et adaptées aux spécificités du domaine médical.

Cette étude propose un modèle d'ensemble combinant EfficientNet-B3 et DenseNet-169 pour la classification de la rétinopathie diabétique. Le modèle a atteint une précision globale élevée de **83,08 %** et un score F1 pondéré de **0,83**, surpassant plusieurs approches de l'état de l'art. L'analyse de la matrice de confusion montre une excellente performance pour la classe majoritaire (absence de rétinopathie), tandis que les stades intermédiaires et avancés présentent une précision modérée à faible, en raison de similarités visuelles et du déséquilibre des données. Les classes 3 et 4 (RD sévère et proliférante) sont souvent confondues avec des stades adjacents. Ces résultats soulignent l'efficacité générale du modèle pour la classification, tout en indiquant qu'un rééquilibrage du jeu de données et une amélioration de la discrimination des caractéristiques visuelles pourraient renforcer la performance pour les stades avancés.

**Discussion L'algorithme 2 :**

L'algorithme proposé, basé sur l'association d'EfficientNet-B3 et de PiT-Base (Pooling-based Vision Transformer), démontre une performance remarquable dans la classification de la rétinopathie diabétique, surpassant nettement les approches comparées dans la littérature. Il atteint une précision globale de 99,57 % et un indice Kappa de 99,36 %, ce qui reflète à la fois une excellente exactitude et une forte concordance entre les prédictions et les annotations réelles, même en tenant compte du déséquilibre entre les classes.

Les modèles traditionnels comme le ResNet-50 personnalisé et MobileNet V3 affichent une précision inférieure (87 % et 98 %, respectivement) avec une performance notablement plus faible sur les classes minoritaires, notamment la classe 4 (52 % et 95 %). Les modèles avec augmentation des données et test-time augmentation (TTA) atteignent de bonnes performances (jusqu'à 99 % pour la classe 0), mais échouent à offrir une stabilité sur l'ensemble des classes. Le modèle Clustered EfficientNet-B6, bien que performant (92 % de précision), montre une hétérogénéité dans les prédictions de certaines classes (par exemple, seulement 83 % pour la classe 1).

En revanche, le modèle proposé maintient une précision de 99 % ou plus dans toutes les classes, y compris celles qui sont historiquement difficiles à classer (classes 3 et 4). Ce résultat démontre non seulement la robustesse de l'architecture hybride CNN-transformer, mais aussi sa capacité à bien généraliser, en capturant à la fois des caractéristiques locales (via EfficientNet-B3) et des relations globales contextuelles (via PiT). Une meilleure résilience face à la variabilité inter-patients. Une amélioration significative de la sensibilité et spécificité pour les cas critiques.

Une réduction du biais vers les classes majoritaires, souvent observé dans les modèles traditionnels.

**Tableau 3.5 :** Comparaison des algorithmes actuels avec L'algorithme 2.

| Classe | customized ResNet50 [82] | ResNet-50 model and Augmentation and TTA [83] | MobileNet V3 [84] | Clustered EfficientNet-B6 [85] | Modified Inception V3 [86] | Proposé |
|---|---|---|---|---|---|---|
| 0 | 98 | 99 | 92.9 | 87 | 95 | 100 |
| 1 | 60 | 60 | 92 | 83 | 91 | 99 |
| 2 | 83 | 71 | 95 | 97 | 83 | 99 |
| 3 | 58 | 36 | 93.6 | 87 | 95 | 99 |
| 4 | 52 | 15 | 95 | 98 | 89 | 99 |
| Accuracy | 87 | 78.1 | 98 | 92 | 91.64 | 99.57 |
| Kappa | - | 90.3 | 91.1 | - | 87.05 | 99.36 |

L'approche d'ensemble proposée a démontré des performances remarquables, surpassant nettement les méthodes existantes de la littérature tant en termes d'exactitude que de cohérence des prédictions. Avec une précision atteignant 99,57 % et un indice de Kappa de 99,36 %, elle montre une excellente capacité à distinguer les différents stades de la maladie, y compris dans une configuration binaire (No_DR vs DR) où la sensibilité est cruciale. Ces résultats soulignent la pertinence de l'association entre EfficientNet-B3 et PiT-Base, et positionnent cette approche comme une solution prometteuse pour des applications cliniques de dépistage automatisé à grande échelle.

## 7 Conclusion

Ce chapitre a permis d'évaluer et de comparer de manière approfondie les performances des deux approches proposées pour la classification de la rétinopathie diabétique : l'ensemble EfficientNet-B3 & DenseNet-169, et l'ensemble EfficientNet-B3 & PiT-Base. Les résultats expérimentaux ont montré que la seconde approche surpasse largement la première, ainsi que plusieurs méthodes issues de la littérature récente, en atteignant une précision de 99,57 % et un score Kappa de 99,36 %.

L'analyse des matrices de confusion, des courbes d'apprentissage, ainsi que des rapports de classification a mis en évidence une excellente capacité de généralisation du modèle proposé,

avec une sensibilité et une spécificité élevée, y compris dans des scénarios multi-classes et binaires (No_DR vs DR). Ces performances traduisent non seulement la robustesse de la combinaison entre CNN et Vision Transformer, mais aussi l'efficacité des techniques d'entraînement mises en œuvre.

Les résultats obtenus confirment la pertinence de l'approche proposée pour une application réelle dans le domaine du dépistage assisté par intelligence artificielle, tout en ouvrant des perspectives d'optimisation et d'extension pour des systèmes de diagnostic médical automatisés plus complexes.

# Conclusion générale

Ce mémoire a présenté une étude approfondie sur la classification automatique de la rétinopathie diabétique (RD) à l'aide d'approches ensemblistes basées sur l'apprentissage profond. Deux architectures hybrides ont été proposées et évaluées : (1) une combinaison de DenseNet-169 et EfficientNet-B3, et (2) une fusion de EfficientNet-B3 avec PiT-Base (Pooling-based Vision Transformer). Ces deux configurations ont été appliquées au jeu de données APTOS 2019, un ensemble de référence dans le domaine du dépistage de la RD.

Les résultats expérimentaux obtenus ont démontré des performances prometteuses, tant en termes d'exactitude que de métriques médicales robustes telles que le F1-score et le score de Kappa de Cohen. L'approche combinant EfficientNet-B3 et PiT s'est distinguée par une précision exceptionnelle de 99,57 % et une excellente capacité à généraliser sur des classes déséquilibrées. Ces résultats confirment l'intérêt des approches hybrides exploitant la complémentarité entre les réseaux convolutifs (CNN), excellents pour la détection de motifs locaux, et les Transformers visuels, réputés pour leur modélisation des dépendances globales.

Au-delà de la performance brute, ce travail souligne plusieurs facteurs déterminants dans les systèmes d'aide au diagnostic automatisé :

- La qualité du prétraitement des images, notamment la normalisation, le recadrage contextuel et l'amélioration de contraste,
- L'importance de l'augmentation de données pour enrichir la variabilité intra-classe,
- L'utilisation de fonctions de perte pondérées et de techniques d'optimisation de mémoire (entraînement mixte) pour une exploitation efficace des ressources matérielles.

Toutefois, certaines limites subsistent. L'étude s'est appuyée sur un seul jeu de données (APTOS 2019), ce qui pourrait restreindre la généralisation à d'autres contextes cliniques. De plus, bien que performantes, les architectures proposées restent perçues comme des "boîtes noires", limitant leur interprétabilité pour une intégration en milieu médical réel.

Plusieurs pistes d'amélioration se dégagent de ce travail :

- Optimisation dynamique de la fusion des modèles : au lieu d'une simple concaténation des représentations, l'utilisation de mécanismes d'attention ou d'apprentissage de poids adaptatifs pourrait améliorer la qualité de l'ensemble.

- Interprétabilité et explication des décisions : l'intégration de modules explicatifs comme Grad-CAM, LIME ou SHAP serait essentielle pour renforcer la confiance des professionnels de santé.
- Validation croisée multi-centrique : élargir l'étude à d'autres jeux de données cliniques (EyePACS, MESSIDOR, DDR) permettrait de tester la robustesse et la transférabilité des modèles.
- Intégration dans des outils décisionnels : en s'appuyant sur des interfaces conviviales et des API médicales sécurisées, le modèle pourrait être déployé comme assistant intelligent au diagnostic pour les ophtalmologistes.

Ce mémoire ouvre la voie à des applications concrètes de l'intelligence artificielle pour le dépistage précoce de la rétinopathie diabétique. Il démontre que l'approche ensembliste, lorsqu'elle est rigoureusement conçue et évaluée, peut constituer une solution fiable, efficace et extensible dans un contexte de médecine de précision.

# Annexe

## Lien des Codes Algorithmes 1 et 2 :

https://github.com/Imenbnc/Classification-APTOS-2019-collaboration-with-Hadil_Bouzahar.git

# Bibliographie


[1] Pascale Massin, et al. Rétinopathie diabétique. Elsevier Health Sciences, 2025.

[2] Salil Mehta. "Optical Coherence Tomography Angiography Findings Associated with Accelerated Hypertension." Cureus 16.4 (2024).

[3] Nicole Jeanrot, Valérie Ducret, and François Jeanrot. Manuel de strabologie: aspects cliniques et thérapeutiques. Elsevier Health Sciences, 2018.

[4] Ph Panas, and Albert Remy. Anatomie pathologique de l'oeil. Delahaye, 1879.

[5] Bahram Bodaghi, , et al. Rétine et vitré. Elsevier Health Sciences, 2018.

[6] J. Zerbib, et al. "Prise en charge de la DMLA exsudative en France en 2015." Journal Français d'Ophtalmologie 40.9 (2017): 723-730.

[7] Motais, Ernest. Anatomie de l'appareil moteur de l'oeil de l'homme et des vertébrés. Delahaye et Lecrosnier, 1887.

[8] Duke-Elder, Stewart. "Anatomie de L'Oeil." British Medical Journal 2.5468 (1965): 987.

[9] Karan Patel, and Bhavesh Parmar. "Assistive device using computer vision and image processing for visually impaired; review and current status." Disability and Rehabilitation: Assistive Technology 17.3 (2022): 290-297.

[10] P. Massin, and A. Gaudric. "sIntroduction." (2008).

[11] SIDIBE Mohamed Kolé, et al. "Laser argon dans la rétinopathie diabetique au CHU-IOTA: Résultats anatomiques et fonctionnels." Revue SOAO 02-2019 (2019): 53-55.

[12] A. MATHIS, "Rétinopathie diabétique: physiopathologie, diagnostic, évolution et pronostic, traitement." La Revue du praticien (Paris) 43.19 (1993): 2557-2561.

[13] Pierre. AMALRIC, "Historique de la rétinopathie diabétique." Histoire des sciences médicales 27.1-1993 (1993): 69.

[14] Mohamed Amari, et al. "Score calcique coronaire et son rôle dans le dépistage de l'ischémie myocardique silencieuse." journal algérien de médecine 32.1 (2025): 3-7.

[15] Sébastien Dam, et al. "Réseau de neurones convolutifs sur graphes multimodal pour l'étude de la structure et de la fonction cérébrales dans l'anxiété et la dépression à l'adolescence." IABM 2025-3ème édition du Colloque Français d'Intelligence Artificielle en Imagerie Biomédicale. 2025.

[16] Bin Wang, et al. "Explaining deep convolutional neural networks for image classification by evolving local interpretable model-agnostic explanations." IEEE Transactions on Emerging Topics in Computational Intelligence (2025).

[17] Han. Yuan, "Anatomic boundary-aware explanation for convolutional neural networks in

diagnostic radiology." Iradiology 3.1 (2025): 47-60.

[18] Shadman Sakib, et al. "An overview of convolutional neural network: Its architecture and applications." (2019).

[19] Alexey Dosovitskiy, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[20] Kai Han, et al. "A survey on vision transformer." IEEE transactions on pattern analysis and machine intelligence 45.1 (2022): 87-110.

[21] Yaoli Wang, et al. "Vision Transformers for Image Classification: A Comparative Survey." Technologies 13.1 (2025): 32.

[22] Salman Khan, et al. "Transformers in vision: A survey." ACM computing surveys (CSUR) 54.10s (2022): 1-41.

[23] Zhang, Hanwei, et al. "A survey on visual mamba." Applied Sciences 14.13 (2024): 5683.

[24] Liu, Xiao, Chenxu Zhang, and Lei Zhang. "Vision mamba: A comprehensive survey and taxonomy." arXiv preprint arXiv:2405.04404 (2024).

[25] Ali Hatamizadeh, and Jan Kautz. "Mambavision: A hybrid mamba-transformer vision backbone." arXiv preprint arXiv:2407.08083 (2024).

[26] Mahesh S. Patil, et al. "Effective deep learning data augmentation techniques for diabetic retinopathy classification." Procedia Computer Science 218 (2023): 1156-1165.

[27] Tiwalade Modupe Usman, et al. "Diabetic retinopathy detection using principal component analysis multi-label feature extraction and classification." International Journal of Cognitive Computing in Engineering 4 (2023): 78-88.

[28] Cheena Mohanty, et al. "Using deep learning architectures for detection and classification of diabetic retinopathy." Sensors 23.12 (2023): 5726.

[29] Bojia Shi, et al. "GoogLeNet-based Diabetic-retinopathy-detection." 2022 14th International Conference on Advanced Computational Intelligence (ICACI). IEEE, 2022.

[30] R. Yasashvini, et al. "Diabetic retinopathy classification using CNN and hybrid deep convolutional neural networks." Symmetry 14.9 (2022): 1932.

[31] Ramzi Adriman, Kahlil Muchtar, and Novi Maulina. "Performance evaluation of binary classification of diabetic retinopathy through deep learning techniques using texture feature." Procedia Computer Science 179 (2021): 88-94.

[32] Charu Bhardwaj, Shruti Jain, and Meenakshi Sood. "Deep learning–based diabetic retinopathy severity grading system employing quadrant ensemble model." Journal of Digital Imaging 34 (2021): 440-457.

[33] A. Rosline Mary, and P. Kavitha. "Automated Diabetic Retinopathy detection and classification

using stochastic coordinate descent deep learning architectures." Materials Today: Proceedings 80 (2023): 3333-3345.

[34] Sharif Amit Kamran, et al. "Vtgan: Semi-supervised retinal image synthesis and disease prediction using vision transformers." Proceedings of the IEEE/CVF international conference on computer vision. 2021.

[35] Wang, Nannan, et al. "Study on Classification Detection Method of Diabetic Retinopathy Based on SSD." Sensing and Imaging 26.1 (2025): 1-19.

[36] Shuang Yu, et al. "Mil-vt: Multiple instance learning enhanced vision transformer for fundus image classification." Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. Springer International Publishing, 2021.

[37] Xiangji Pan, et al. "Multi-label classification of retinal lesions in diabetic retinopathy for automatic analysis of fundus fluorescein angiography based on deep learning." Graefe's Archive for Clinical and Experimental Ophthalmology 258 (2020): 779-785.

[38] D. Jude Hemanth, Omer Deperlioglu, and Utku Kose. "RETRACTED ARTICLE: An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network." Neural Computing and Applications 32.3 (2020): 707-721.

[39] Thippa Reddy Gadekallu, et al. "Early detection of diabetic retinopathy using PCA-firefly based deep learning model." Electronics 9.2 (2020): 274.

[40] K. Shankar, et al. "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model." Pattern Recognition Letters 133 (2020): 210-216.

[41] K. Shankar, et al. "Hyperparameter tuning deep learning for diabetic retinopathy fundus image classification." IEEE access 8 (2020): 118164-118173.

[42] Shankar, K., Zhang, Y., Liu, Y., Wu, L. and Chen, C.-H. (2020). Hyperparameter Tuning Deep Learning for Diabetic Retinopathy Fundus Image Classification. IEEE Access, 8, pp.118164–118173. doi: https://doi.org/10.1109/access.2020.3005152.

[43] Jiang, Hongyang, et al. "A multi-label deep learning model with interpretable grad-CAM for diabetic retinopathy classification." 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020.

[44] de La Torre, Jordi, Aida Valls, and Domenec Puig. "A deep learning interpretable classifier for diabetic retinopathy disease grading." Neurocomputing 396 (2020): 465-476.

[45] Zhang, Wei, et al. "Automated identification and grading system of diabetic retinopathy using deep neural networks." Knowledge-Based Systems 175 (2019): 12-25.

[46] Li, Xiaomeng, et al. "CANet: cross-disease attention network for joint diabetic retinopathy and

diabetic macular edema grading." IEEE transactions on medical imaging 39.5 (2019): 1483-1493.

[47] Li, Xiaomeng, et al. "CANet: cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading." IEEE transactions on medical imaging 39.5 (2019): 1483-1493.

[48] Zhou, Yi, et al. "Collaborative learning of semi-supervised segmentation and classification for medical images." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

[49] Li, Tao, et al. "Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening." Information Sciences 501 (2019): 511-522.

[50] Wan, Shaohua, Yan Liang, and Yin Zhang. "Deep convolutional neural networks for diabetic retinopathy detection by image classification." Computers & Electrical Engineering 72 (2018): 274-282.

[51] Wang, Xiaoliang, et al. "Diabetic retinopathy stage classification using convolutional neural networks." 2018 IEEE international conference on information reuse and integration (IRI). IEEE, 2018.

[52] Lin, Gen-Min, et al. "Transforming retinal photographs to entropy images in deep learning to improve automated detection for diabetic retinopathy." Journal of ophthalmology 2018.1 (2018): 2159702.

[53] Ashwani K. “What Is Keras and Use Cases of Keras? - DevOpsSchool.com.” 17 Aug. 2023

[54] Google. *TensorFlow: une bibliothèque open source pour le calcul numérique et l'apprentissage automatique*. TensorFlow, https://www.tensorflow.org/

[55] Harris, Charles R., et al. "Array programming with NumPy." Nature 585.7825 (2020): 357-362.

[56] McKinney, Wes. "Data structures for statistical computing in Python." SciPy 445.1 (2010): 51-56.

[57] McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2e éd.). O’Reilly Media.

[58] Wildcodeschool.com. (2025). Définition Google Colaboratory - Qu’est-ce que Google Colab ? | Wild Code School. [online] Available at: https://www.wildcodeschool.com/fr-fr/lexique/google-colaboratory [Accessed 9 Apr. 2025].

[59] NVIDIA, C. "NVIDIA turing GPU architecture whitepaper." (2018).

[60] NVIDIA T4 70W LOW PROFILE PCIe GPU ACCELERATOR. (2018). Available at: https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/tesla-product-literature/T4%20Product%20Brief.pdf.

[61] Niloy Sikder, et al. "Early blindness detection based on retinal images using ensemble learning." 2019 22nd International conference on computer and information technology (ICCIT). IEEE, 2019.

[62] Jia, Zhe, et al. "Dissecting the nvidia turing t4 gpu via microbenchmarking." arXiv preprint arXiv:1903.07486 (2019).

[63] Ng, Darren, et al. "Benchmarking Object Detection Models with Mummy Nuts Datasets." International Symposium on Benchmarking, Measuring and Optimization. Cham: Springer International Publishing, 2022.

[64] Kljucaric, Luke, and Alan D. George. "Deep learning inferencing with high-performance hardware accelerators." ACM Transactions on Intelligent Systems and Technology 14.4 (2023): 1-25. [Accessed 10 Apr. 2025] .

[65] Reuther, Albert, et al. "AI and ML accelerator survey and trends." 2022 IEEE High Performance Extreme Computing Conference (HPEC). IEEE, 2022.

[66] Khairy, Mahmoud, et al. "Simr: Single instruction multiple request processing for energy-efficient data center microservices." 2022 55th IEEE/ACM International Symposium on Microarchitecture (MICRO). IEEE, 2022.

[67] Géron, Aurélien. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. " O'Reilly Media, Inc.", 2022.

[68] Douglass, Michael JJ. "Book Review: Hands-on Machine Learning with Scikit-Learn, Keras, and Tensorflow, by Aurélien Géron: O'Reilly Media, 2019, 600 pp., ISBN: 978-1-492-03264-9." (2020): 1135-1136.

[69] Takaya aito, and Marc Rehmsmeier. "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets." PloS one 10.3 (2015): e0118432.

[70] Pajankar, Ashwin, and Aditya Joshi. Hands-on Machine Learning with Python: Implement Neural Network Solutions with Scikit-learn and PyTorch. Berkeley: apress, 2022.

[71] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.

[72] Shinde, Pramila P., and Seema Shah. "A review of machine learning and deep learning applications." 2018 Fourth international conference on computing communication control and automation (ICCUBEA). IEEE, 2018.

[73] Ballard, Will. Hands-on deep learning for images with TensorFlow: build intelligent computer vision applications using TensorFlow and Keras. Packt Publishing Ltd, 2018.

[74] Rusk, Nicole. "Deep learning." Nature Methods 13.1 (2016): 35-35.

[75] Min, Seonwoo, Byunghan Lee, and Sungroh Yoon. "Deep learning in bioinformatics." Briefings in bioinformatics 18.5 (2017): 851-869.

[76] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[77] Diederik, Kingma. "Adam: A method for stochastic optimization." (No Title) (2014).

[78] O. Dekhil, A. Naglah, M. Shaban, M. Ghazal, F. Taher and A. Elbaz, "Deep Learning Based Method for Computer Aided Diagnosis of Diabetic Retinopathy," 2019 IEEE International Conference on Imaging Systems and Techniques (IST), Abu Dhabi, United Arab Emirates, 2019, pp. 1-4, doi: 10.1109/IST48021.2019.9010333.

[79] Mohanty, C., Mahapatra, S., Acharya, B., Fotios Kokkoras, Gerogiannis, V.C., Ioannis Karamitsos and Kanavos, A. (2023b). Using Deep Learning Architectures for Detection and Classification of Diabetic Retinopathy. 23(12), pp.5726–5726. doi:https://doi.org/10.3390/s23125726.

[80] İ. Akgül, Çağrı Yavuz, Ö. i Yavuz, U. (2023). Deep Learning Based Models for Detection of Diabetic Retinopathy. Tehnički glasnik, 17 (4), 581-587. https://doi.org/10.31803/tg-20220905123827

[81] Ritunsa Mishra, et al. "Diabetic Retinopathy Image Classification And Blind-Ness Detection Using Deep Learning Techniques." Machine Intelligence Research 18.1 (2024): 1056-1077.

[82] Samra Nawazish, Javed, A. and Nazir, T. (2023). Deep Learning Model for the Detection and Classification of Diabetic Retinopathy. Proceedings of 1st International Conference on Computing Technologies, Tools and Applications. [online] Available at: https://www.researchgate.net/publication/372717499_Deep_Learning_Model_for_the_Detection_and_Classification_of_Diabetic_Retinopathy.

[83] Patil, M.S., Chickerur, S., Abhimalya, C., Naik, A., Kumari, N. and Maurya, S. (2023). Effective Deep Learning Data Augmentation Techniques for Diabetic Retinopathy Classification. Procedia Computer Science, [online] 218, pp.1156–1165. Available at: https://www.sciencedirect.com/science/article/pii/S1877050923000947

[84] Sait, W. and Rahaman, A. (2023). A Lightweight Diabetic Retinopathy Detection Model Using a Deep-Learning Technique. Diagnostics, [online] 13(19), p.3120. Available at: https://www.mdpi.com/2075-4418/13/19/3120

[85] Abood, Raghad.H. and Hamad, Ali.H. (2025). View of Optimizing Diabetic Retinopathy Classification with Transfer Learning: a Lightweight Approach Using Model Clustering. [online] Uoitc.edu.iq. Available at: https://ijci.uoitc.edu.iq/index.php/ijci/article/view/533/252 [Accessed 11 May 2025].

[86] Abini, M.A. and Sridevi Sathya Priya (2025). Detection and Classification of Diabetic Retinopathy Using Modified Inception V3. International Journal Bioautomation, [online] 29(1), pp.77–92. Available at: https://www.proquest.com/openview/df59c522984e05157c3ba5c295d9c926/1?cbl=4720765&pq-origsite=gscholar.