

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement supérieur et de la Recherche scientifique

Université A. Mira-Béjaïa
Faculté des Sciences Exactes
Department D'informatique



Master Recherche

En

Informatique

Option

Système d'information avancés (SIA)

Thème

Diagnostic automatique de la Rétinopathie Diabétique en exploitant les architectures basées sur le Transfer Learning

Made by:

M. BAOUZ Sif Eddine

Devant le jury composé de :

Président : M. Atmani

Examineur : Mme. H. Khaled

Encadrant : Mme. S. Ait Kaci Azzou

Co-Encadrant : Mme. D. Boukredera

Université de Béjaïa.

Université de Béjaïa.

Université de Béjaïa.

Université de Béjaïa.

U. A/M Béjaïa, Juin 2024.

Remerciements

Avant toute chose, il me semble opportun de commencer cette thèse en exprimant ma gratitude, tout d'abord à Dieu pour m'avoir accordé la force et le courage de mener à bien ce modeste travail.

Mes remerciements et ma gratitude vont à mes superviseurs, Mme Samira AIT KACI AZZOU et Mme Djamila BOUKREDERA, qui m'ont soutenus et guidés tout au long de cette expérience professionnelle avec patience et enthousiasme.

Je tien également à remercier les membres du jury d'avoir accepté d'examiner et d'évaluer mon travail.

A tous ceux qui ont contribué, directement ou indirectement, par leurs conseils, leurs encouragements ou leur amitié,

à la réalisation de cet ouvrage, j'exprime ma profonde gratitude.

Pour leurs encouragements, leur soutien moral et la patience dont ils ont fait preuve tout au long de l'année,

je suis sincèrement reconnaissants à tous les membres de ma familles.

Dédicaces

Ce modeste travail est dédié à:

Mes chers parents qui m'ont soutenus et encouragés tout au long de

mon éducation,

Mes sœurs,

Mes professeurs,

Mes amis,

Et à toutes les personnes qui nous ont apporté leur aide.

Sif eddine

Table des matières

Liste des figures	5
Liste des tableaux	7
Liste des acronymes	8
Introduction Générale	9
1 Rétinopathie Diabétique	11
I Introduction	11
II Physiopathologie de la rétinopathie diabétique	11
II.1 Changements physiologiques dans l'œil causés par le diabète	11
II.2 Stades de la rétinopathie diabétique	11
III Stades de la RD	12
III.1 Symptômes communs de la rétinopathie diabétique	14
III.2 Méthodes de diagnostic	14
III.2.1 Examen du fond d'œil	14
III.2.2 Tomographie par cohérence optique (OCT) et l'angiographie par OCT	14
IV Traitement et Gestion	15
V Importance de la détection précoce de la RD	15
VI Conclusion	16
2 État de l'art: Transfert learning & RD	17
I Introduction	17
II Introduction aux Transformers	17
II.1 Architecture des Transformers	17
II.2 Du traitement du langage à la vision	18
II.2.1 Les principes de base des Vision Transformers (ViT)	18
II.3 Architecture des Vision Transformers	19
II.3.1 Mécanisme d'auto-attention dans le contexte des images	19
II.4 Pré-entraînement et fine-tuning des ViT	20
II.5 Applications des ViT dans la vision par ordinateur	20
III Yolo	21
III.1 Historique de YOLO	21
III.2 Architecture Yolo	21
III.3 Domaine d'application	22
IV Architecture Resnet50	23
V Apprentissage par transfert et fine-tuning	23

VI	Métriques de performances	24
VI.1	Accuracy	24
VI.2	Précision	25
VI.3	Rappel	25
VI.4	F1 Score	25
VII	État de l’art	25
VIII	Conclusion	27
3	Conception & Réalisation	31
I	Introduction	31
II	Méthodologie	31
II.1	Préparation des données	32
II.1.1	Dataset	32
II.1.2	Prétraitement d’images	33
II.1.3	Fine Tuning	34
II.2	Algorithme	34
II.3	Détection de la RD	35
II.3.1	Approche basée ViT	36
II.3.2	Approche basée Yolo	39
II.3.3	Approche Hybride	42
II.3.4	Comparaison des modèles	45
II.3.5	Comparaison avec d’autres travaux	45
II.4	Classification de la RD en 3 classes	46
II.4.1	Architectures proposées	47
II.5	Classification de la RD en 5 classes	56
III	Conclusion	64
	Conclusion	65
	BIBLIOGRAPHIE	66

Liste des figures

1.1	Physiopathologie de la rétinopathie diabétique	12
1.2	Les différents stades de la RD	13
1.3	Examen du fond d’œil	14
1.4	Tomographie par cohérence optique (OCT) et l’angiographie par OCT	15
2.1	Structure de l’encodeur et du décodeur	18
2.2	Division en patches de taille fixe	19
2.3	Architecture Vision Transformer	20
2.4	Architecture CNN de YOLO	22
2.5	Architecture Resnet50	23
2.6	Diagramme comparatif des processus d’apprentissage entre l’apprentissage automatique conventionnel et l’apprentissage par transfert	24
3.1	Architecture Proposée	32
3.2	Dataset Aptos	33
3.3	Phases de prétraitement	34
3.4	Normalisation	34
3.5	Division des classes du dataset	36
3.6	Architecture ViRd pour la détection de la RD basée Vit	37
3.7	Matrice de confusion de ViRd	39
3.8	Graphes d’exactitude et de perte (ViT)	39
3.9	Architecture RDYol pour la détection de la RD	40
3.10	Matrice de confusion RDYol	41
3.11	Graphe d’exactitude et de perte pour le RDYol	42
3.12	Architecture hybride ReVi-RD pour la détection de la RD	43
3.13	Matrice de confusion de ReVi-RD	44
3.14	Graphes d’exactitude et de perte ReVi-RD	44
3.15	Comparaison des métriques les des différents modèles.	45
3.16	Comparaison des métriques des différents travaux.	46
3.17	Divisions Aptos	47
3.18	Architecture ViR-3C proposée pour la classification de la RD en 3 classes	48
3.19	Matrice de confusion pour la ViR-3C	49
3.20	Graphes d’exactitude et de perte ViR-3C	49
3.21	Architecture RD-Yol3C pour la détection de la RD	50
3.22	Matrice de confusion pour la RD-Yol3C	51
3.23	Graphe d’exactitude et de perte pour le RD-Yol3C	52
3.24	Architecture ReVi-3C pour la détection de la RD 3C	53
3.25	Matrice de confusion pour la ReVi-3C	54
3.26	Graphes d’exactitude et de perte (ReVi-3C)	54

3.27	Comparaison des résultats des modèles	55
3.28	Comparaison avec les travaux antérieurs	56
3.29	Distribution des classes Aptos 5C	56
3.30	Architecture ViR5C	57
3.31	Matrice de confusion pour le ViR5C	58
3.32	Graphes d'exactitude et de perte ViR5C	59
3.33	Architecture Yolo 5C (RD-Yol5C)	59
3.34	Matrice de confusion pour le RD-Yol5C	60
3.35	Graphe d'exactitude et de perte pour le RD-Yol5C	61
3.36	Architecture ReVi-5C	62
3.37	Matrice de confusion pour le ReVi-5C	63
3.38	Graphes d'exactitude et de perte ReVi-5C	63

Liste des tableaux

2.1	Long table caption.	28
3.1	Hyperparameters du fine tuning	38
3.2	Résultats ViRd	38
3.3	Résultats RDYol	41
3.4	Résultats ReVi-RD	43
3.5	Comparaison des résultats	46
3.6	Résultats ViR-3C	49
3.7	Résultats Yolo	51
3.8	Résultats ReVi-3C	54
3.9	Comparaison des résultats 3C	55
3.10	Hyperparameters du fine tuning	57
3.11	Résultats ViR5C	58
3.12	Résultats RD-Yol5C	60
3.13	Résultats ReVi-5C	62
3.14	Comparaison des résultats 5C	64
3.15	Comparaison des résultats 5C	64

Liste des acronymes

ViT Vision Transformer

YOLO Yoo look only once

IA Intelligence artificielle

ML Machine learning

DL Deep learning

RD Rétinopathie diabétique

ViRD Vision Transformer Retino Detection

RDYol Retino detection based Yolo

ReVi-RD Resnet Vision Transformer Retino Detection

Introduction Générale

La rétinopathie diabétique (RD) est une cause importante de déficience visuelle, ce qui souligne la nécessité cruciale d'un dépistage précoce et d'une intervention en temps utile pour éviter la détérioration visuelle [31, 66].

La Fédération Internationale du Diabète (FID) a estimé que la population mondiale atteinte de diabète sucré était de 463 millions en 2019 et qu'elle devrait atteindre 700 millions d'ici 2045 [50]. En tant que complication la plus courante et spécifique du diabète sucré, la rétinopathie diabétique est également l'une des principales causes de cécité [3]. Par conséquent, le diagnostic précoce de la RD est essentiel pour un traitement efficace et la prévention d'une éventuelle perte de vision [45].

La rétinopathie diabétique représente une lourde charge pour les spécialistes de la santé, pouvant parfois entraîner des diagnostics erronés. La détection précoce de la RD revêt donc une importance capitale, permettant de prévenir les complications visuelles, de réduire les coûts élevés des traitements et d'alléger la charge de travail des professionnels de la santé.

La surcharge des médecins et la nécessité d'un diagnostic précoce rendent impérative l'utilisation des techniques de l'intelligence artificielle (IA). Les méthodes d'apprentissage automatique (ML) et d'apprentissage profond (deep learning) offrent des solutions prometteuses pour améliorer la précision et l'efficacité du diagnostic de la RD [45]. Toutefois, ces méthodes présentent certains inconvénients, notamment la nécessité de grandes quantités de données annotées pour l'entraînement des modèles, la complexité computationnelle, et le risque de surajustement (overfitting).

C'est ici que l'apprentissage par transfert (transfer learning) devient crucial. L'apprentissage par transfert permet de surmonter certaines de ces limitations en utilisant des modèles pré-entraînés sur des tâches similaires et en les adaptant à des tâches spécifiques avec des données limitées. Cette approche peut réduire le besoin de grandes quantités de données annotées et améliorer la généralisation des modèles, rendant ainsi l'IA plus accessible et efficace pour le diagnostic de la RD.

Ces dernières années, les techniques d'apprentissage profond ont montré des résultats prometteurs dans l'amélioration du diagnostic de la RD [47]. Le domaine du transfert learning offre un large éventail d'architectures performantes, parmi lesquelles YOLO et les Vision Transformers. Ces modèles tirent parti du concept d'apprentissage par transfert, qui implique l'utilisation de modèles pré-entraînés pour extraire des caractéristiques et affiner certaines couches afin de répondre aux exigences uniques de cette tâche de diagnostic spécifique [23].

Notre Objectif vise à développer des modèles d'apprentissage profond performants pour le diagnostic automatique de la RD à partir d'images de fond d'œil. Nous proposons pour cela d'explorer deux architectures distinctes :

- YOLOv8: Cette architecture est basée sur le modèle YOLOv8, reconnu pour son efficacité dans des tâches de détection d'objets en vision par ordinateur. Nous proposons d'adapter YOLOv8 pour détecter et classifier les anomalies rétiniques caractéristiques de la RD.

- Vision Transformers (ViTs): Ces architectures, basées sur des mécanismes d'attention, ont récemment surpassé les CNNs dans diverses tâches de vision par ordinateur. Nous proposons d'exploiter les capacités des ViTs pour capturer des dépendances à longue portée entre les pixels des images rétinienne et identifier des patterns discriminants pour le diagnostic de la RD. En plus nous prposons une nouvelle architecture hybride ReVi, Combinaison du modèle PANC proposé par [1] ViTs: Cette architecture hybride que nous proposons, combine les avantages des CNNs et des ViTs.

Nous pouvons résumer notre apport comme suit :

- Nous avons regrouper les classes du dataset pour une classification en 3 classes, car la détection précoce est plus importante que la classification de la pathologie en plusieurs stades en raison de l'absence des symptômes de la maladie à ses débuts, chose qui n'a pas été faite auparavant dans le cas de la classification en utilisant les Vits, et Yolov8.
- Adapter l'architecture Yolov8 pour la détection, et la classification de la RD en 3 et 5 classes, chose qui n'a pas été faite dans la littérature.
- Adapter l'architecture Vit pour la détection, et la classification de la RD en 3 et 5 classes.
- Développer une nouvelle architecture Hybride ReVi, combinaison de PANC ($N = 2$, $N = 3$, $N = 5$) modèle proposé dans[1] et Vit pour la détection et la classification de la rétinopathie diabétique en 3 et 5 classes.
- Les résultats obtenues sont prometteurs pour un diagnostique automatique.

Cette thèse est structurée comme suit : Le chapitre 1 présente un aperçu de la Rétinopathie Diabétique (RD) et de ses complications. Le chapitre 2 aborde les concepts d'apprentissage par transfert, en mettant l'accent sur les transformateurs de vision et YOLO, et fournit un état de l'art des différents travaux réalisés dans le cadre du diagnostic de la rétinopathie diabétique utilisant les vision transformers et YOLO. Le détail de nos modèles proposés, ainsi que les résultats obtenus, seront présentés dans le chapitre 3. Nous concluons par une discussion et des perspectives.

Chapitre 1

Rétinopathie Diabétique

I Introduction

La rétinopathie diabétique (RD) est une complication du diabète et une cause majeure de cécité. Elle survient lorsque le diabète endommage les minuscules vaisseaux sanguins à l'intérieur de la rétine, le tissu sensible à la lumière à l'arrière de l'œil [17]. Une rétine saine est nécessaire pour une bonne vision [17].

Le diabète sucré est l'un des troubles métaboliques les plus fréquemment rencontrés, altérant la vie saine à travers le globe, avec des complications secondaires allant de légères à sévères. Le diabète sucré est associé à des complications secondaires, dont la rétinopathie diabétique, qui endommage la rétine et peut conduire à une perte de vision [3].

Dans ce chapitre, nous explorerons la physiopathologie de la RD, ses symptômes, son diagnostic et son traitement. Nous soulignerons également l'importance de la détection précoce et de la gestion du diabète pour prévenir la progression de la RD.

II Physiopathologie de la rétinopathie diabétique

La rétinopathie diabétique (RD) est reconnue depuis longtemps comme une maladie microvasculaire. Le diagnostic de la RD repose sur la détection de lésions microvasculaires [57]. Les changements physiologiques dans l'œil voir figure 1.1 causés par le diabète sont à l'origine de la RD [57].

II.1 Changements physiologiques dans l'œil causés par le diabète

- Le diabète sucré provoque des changements dans les petits vaisseaux sanguins de la rétine, le tissu sensible à la lumière à l'arrière de l'œil [57].
- Ces changements peuvent entraîner une augmentation de la perméabilité vasculaire et une occlusion capillaire, qui sont deux observations principales dans la vasculature rétinienne [57].

II.2 Stades de la rétinopathie diabétique

- Au cours du stade précoce de la RD, appelé rétinopathie diabétique non proliférative (NPDR), des pathologies rétiniennes telles que les microanévrismes, les hémorragies et

les exsudats durs peuvent être détectés par photographie du fond d'œil. À ce stade, les patients peuvent être asymptomatiques.

- La rétinopathie diabétique proliférante (PDR), un stade plus avancé de la RD, est caractérisée par la néovascularisation. Pendant ce stade, les patients peuvent subir une perte de vision sévère lorsque les nouveaux vaisseaux anormaux saignent dans le vitré (hémorragie vitréenne) ou lorsqu'un décollement rétinien de traction est présent [57].

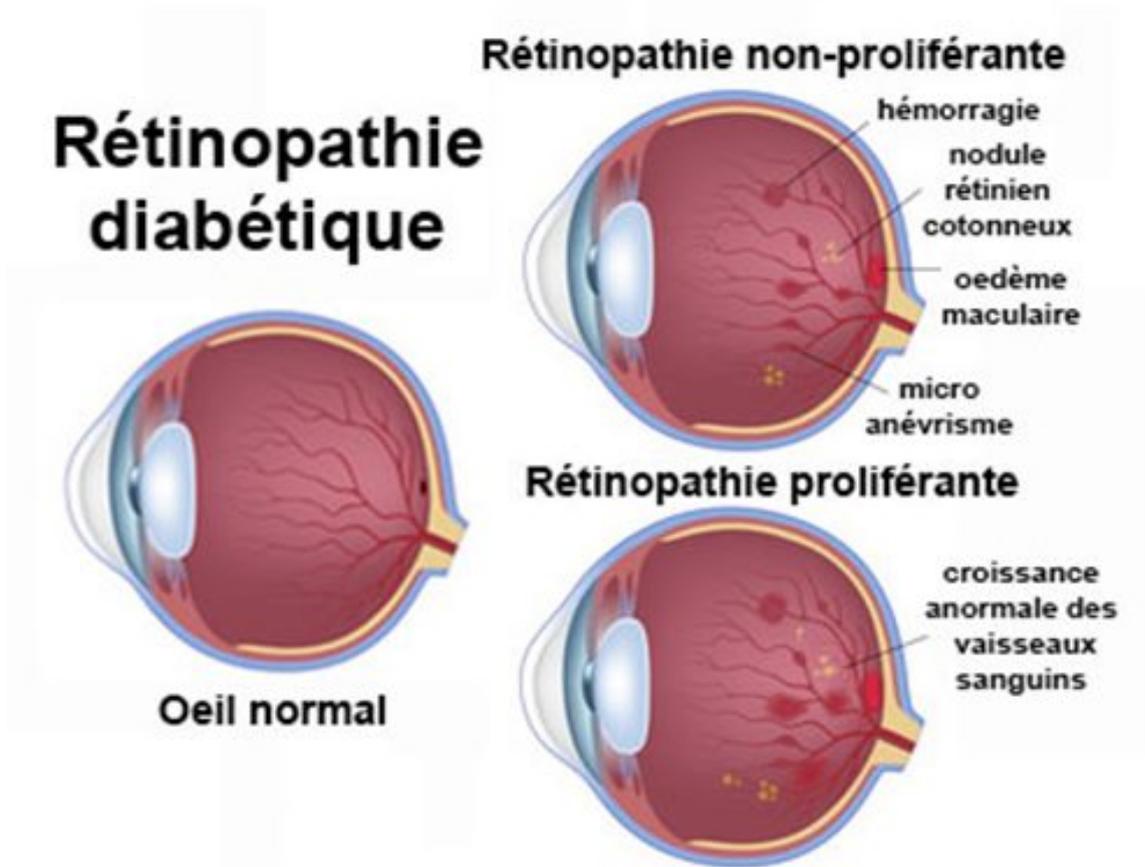


Figure 1.1: Physiopathologie de la rétinopathie diabétique [41]

III Stades de la RD

La RD peut être classée en quatre niveaux : rétinopathie diabétique non proliférante (NPDR) légère, modérée et sévère, et rétinopathie diabétique proliférante (PDR), comme le montre le figure 1.2

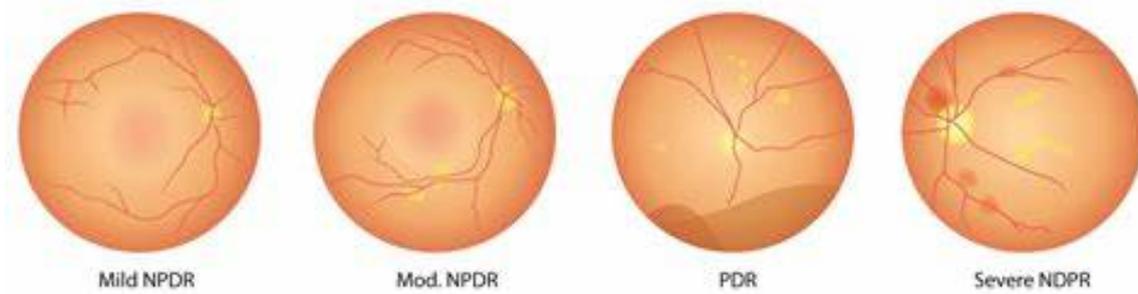


Figure 1.2: Les différents stades de la RD [52]

- **Stade 1 : Rétinopathie diabétique non proliférante légère**

Il s'agit du stade le plus précoce de la rétinopathie diabétique, caractérisé par de minuscules gonflements/bulles dans les vaisseaux sanguins de la rétine. Ces zones de gonflement sont appelées microanévrismes. Ces microanévrismes peuvent entraîner la fuite de petites quantités de liquide dans la rétine, ce qui déclenche un gonflement de la macula, c'est-à-dire de la partie arrière de la rétine. Malgré cela, il n'y a généralement pas de symptômes clairs indiquant qu'il y a un problème. [52]

- **Stade 2 : Rétinopathie diabétique non proliférante modérée**

À ce stade, les minuscules vaisseaux sanguins se gonflent davantage, bloquant la circulation sanguine vers la rétine et l'empêchant d'être correctement alimentée. Ce stade ne provoque des signes visibles que s'il y a une accumulation de sang et d'autres liquides dans la macula, ce qui rend la vision floue. [52]

- **Stade 3 : Rétinopathie diabétique non proliférante sévère**

Au cours de cette phase, une grande partie des vaisseaux sanguins de la rétine se bloque, entraînant une diminution significative du flux sanguin dans cette zone. Le manque de sang déclenche un signal à l'organisme qui commence à faire croître de nouveaux vaisseaux sanguins dans la rétine. [52]

Ces nouveaux vaisseaux sanguins sont extrêmement fins et fragiles et provoquent un gonflement de la rétine, ce qui se traduit par une vision sensiblement floue, des taches sombres et même des zones de perte de vision. Si ces vaisseaux s'infiltrèrent dans la macula, une perte soudaine et permanente de la vision peut survenir. À ce stade, le risque de perte irréversible de la vision est élevé. [52]

- **Stade 4 : Rétinopathie diabétique proliférante**

À ce stade avancé de la maladie, de nouveaux vaisseaux sanguins continuent de se développer dans la rétine. Ces vaisseaux sanguins, qui sont minces, faibles et susceptibles de saigner, provoquent la formation de tissu cicatriciel à l'intérieur de l'œil. Ce tissu cicatriciel peut éloigner la rétine de l'arrière de l'œil, provoquant un décollement de la rétine. Un décollement de la rétine se traduit généralement par une vision floue, un champ de vision réduit, voire une cécité permanente. [52]

Dans cette section nous présentons les différents symptômes de la RD ainsi que les différentes techniques de son diagnostic.

III.1 Symptômes communs de la rétinopathie diabétique

La rétinopathie diabétique (RD) est souvent asymptomatique aux premiers stades de la maladie. Cependant, à mesure que la maladie progresse, les patients peuvent commencer à remarquer des symptômes tels que :

- La vision floue.
- Les tâches flottantes.
- La perte de vision.
- La cécité

III.2 Méthodes de diagnostic

Pour pouvoir traiter la RD, il faut savoir la diagnostiquer. Dans ce qui suit, nous présenterons les techniques utilisées par les ophtalmologues pour déterminer l'existence de la pathologie.

III.2.1 Examen du fond d'œil

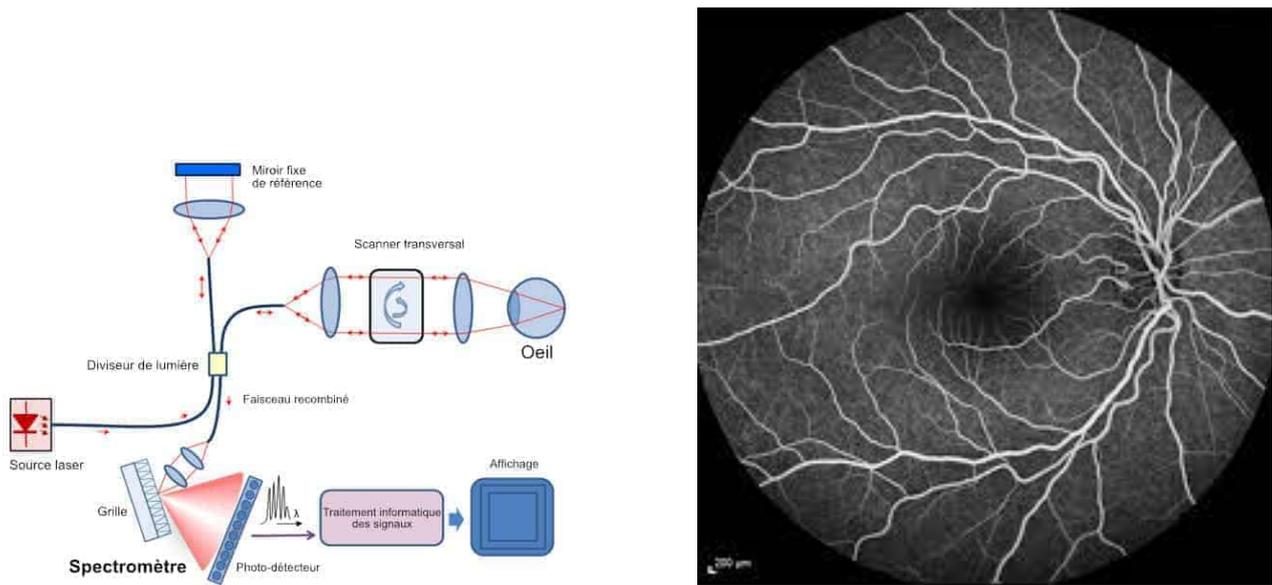
L'examen du fond d'œil est le principal outil de diagnostic. Il permet de visualiser les lésions de la rétine et de détecter les stades de la maladie [46]. la figure 1.3 illustre ce dernier.



Figure 1.3: Examen du fond d'œil
Examen du fond d'œil [25]

III.2.2 Tomographie par cohérence optique (OCT) et l'angiographie par OCT

La tomographie par cohérence optique (OCT) voir figure 1.4a et l'angiographie par OCT voir figure 1.4b sont également utilisées pour le diagnostic et le suivi de la RD. Ces techniques d'imagerie non invasives permettent une évaluation détaillée de la rétine et du nerf optique, et peuvent aider à détecter les changements précoces de la RD avant l'apparition des symptômes [46].



(a) Tomographie en cohérence optique [26]

(b) Angiographie OCT [2]

Figure 1.4: Tomographie par cohérence optique (OCT) et l'angiographie par OCT

IV Traitement et Gestion

Le traitement de la rétinopathie diabétique (RD) reste un défi. Le traitement le plus efficace pour la RD est le contrôle des niveaux de glucose dans le sang.[54].

Les cas plus avancés nécessitent d'autres traitements comme :

- Traitement au laser.
- Une thérapie anti-VEGF.
- Des stéroïdes.
- Une vitrectomie.

V Importance de la détection précoce de la RD

La rétinopathie diabétique est souvent asymptomatique au premier stade, ce qui rend difficile son diagnostic sans examen médical. Cela signifie que les patients ne présentent souvent pas de symptômes visuels jusqu'à ce que la lésion soit à un stade avancé [43, 37]. L'absence de symptômes au début de la maladie rend la détection précoce encore plus que nécessaire .

Les complications de la rétinopathie diabétique, comme la perte de vision et la cécité, peuvent entraîner des coûts élevés pour les systèmes de santé et des pertes de productivité pour les individus et les sociétés. [33] Une solution proposée est le recours à l'intelligence artificielle (IA) pour aider les médecins dans la détection précoce de la rétinopathie diabétique. En effet, les réseaux de neurones, en particulier, ont montré une grande efficacité dans la classification automatique de la rétinopathie diabétique [43, 33, 37].

VI Conclusion

Dans ce chapitre, nous avons vu les différents aspects de la rétinopathie diabétique. Nous avons exploré les différents stades de cette maladie, ses facteurs de risque ainsi que les différentes options de traitement disponibles. Cependant, le dépistage précoce demeure un défi majeur, car en plus des examens qui sont coûteux, les régions reculées souffrent du manque d'ophtalmologistes formés.

C'est là que l'intelligence artificielle et plus précisément les réseaux de neurones peuvent apporter une solution prometteuse. Dans le prochain chapitre, nous découvrirons comment ces systèmes d'apprentissage automatique avancés permettent de détecter de manière fiable la rétinopathie diabétique à partir d'images du fond d'œil.

Chapitre 2

État de l'art: Transfert learning & RD

I Introduction

Au cours de la dernière décennie, l'apprentissage automatique et l'apprentissage profond ont révolutionné la détection de la RD, offrant des moyens efficaces et efficaces pour le diagnostic précoce [47].

Des algorithmes d'apprentissage profond ont été utilisés dans des systèmes de diagnostic assisté par ordinateur qui pourraient être utiles pour diagnostiquer rapidement la RD [47]. Des modèles tels que les réseaux de neurones convolutifs (CNN) ont été largement utilisés pour détecter la RD à partir d'images du fond d'œil et identifier la gravité de la maladie [9]. Cependant, malgré leurs succès, ces approches présentent des inconvénients. Par exemple, elles nécessitent de grandes quantités de données d'entraînement étiquetées, ce qui peut être difficile et coûteux à obtenir [42],

Pour y remédier à ce problème on utilise l'apprentissage par transfert qui est une technique d'apprentissage en profondeur qui permet de surmonter certaines de ces limitations. Il utilise des modèles pré-entraînés sur de grands ensembles de données et les adapte à de nouvelles tâches avec des ensembles de données plus petits [32].

Dans ce chapitre, nous allons explorer trois différents modèles basés sur le transfert learning pour la détection de la RD : le Transformateur de Vision (ViT), You Only Look Once (YOLO) et une combinaison de ResNet et ViT.

II Introduction aux Transformers

Les Transformers ont été introduits en 2017 par l'article de Vaswani et al. [55] dans le cadre du traitement du langage naturel (NLP). Leur architecture repose entièrement sur des mécanismes d'attention, contrairement aux réseaux de neurones convolutifs (CNN) et récurrents (RNN) qui étaient précédemment utilisés pour de nombreuses tâches de NLP.

Les Transformers ont rapidement gagné en popularité grâce à leurs performances supérieures dans diverses tâches de NLP, comme la traduction automatique, la modélisation de langage et les systèmes de questions-réponses. Leur succès a conduit à l'émergence de modèles pré-entraînés de grande taille, tels que BERT [10], GPT [38, 7], et plus récemment PaLM [8] et Claude [4], qui ont démontré des capacités impressionnantes dans de nombreux domaines.

II.1 Architecture des Transformers

Le cœur de l'architecture Transformer repose sur le mécanisme d'auto-attention, qui

permet de capturer les dépendances entre les différents éléments d'une séquence. Pour une séquence donnée, l'auto-attention calcule une représentation pour chaque élément en tenant compte des informations provenant de tous les autres éléments [55]. Cela permet de mieux modéliser les relations à longue distance, un défi majeur pour les réseaux récurrents classiques.

Les Transformers sont composés de deux parties principales : l'encodeur et le décodeur voir figure 2.1. L'encodeur traite la séquence d'entrée et produit une représentation encodée. Le décodeur, quant à lui, génère la séquence de sortie à partir de la représentation encodée et d'un décalage dans la séquence de sortie (pour les tâches de génération de texte, par exemple) [55].

Les deux parties principales des transformers à savoir l'encodeur et le décodeur sont constitués d'une succession de couches d'auto-attention et de couches de feedforward (réseaux de neurones entièrement connectés).

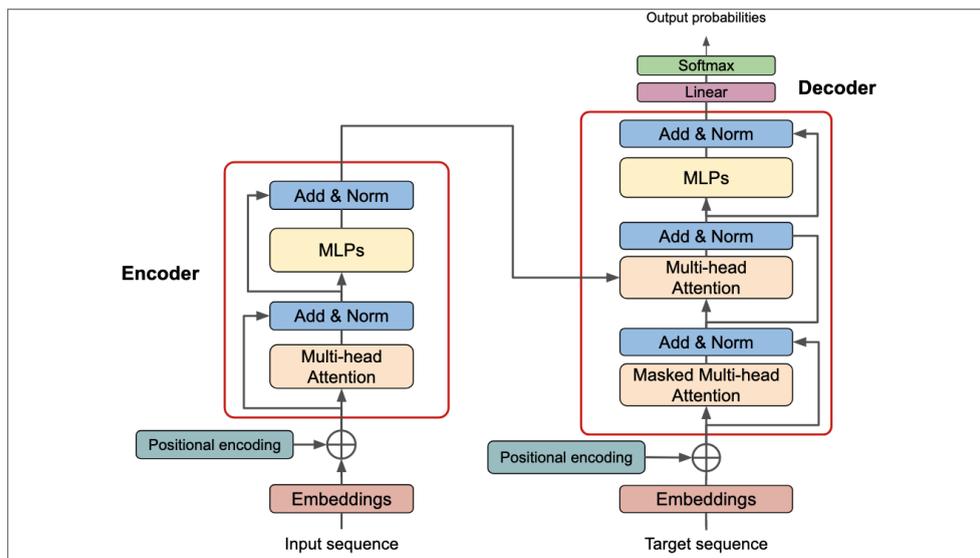


Figure 2.1: Structure de l'encodeur et du décodeur [55]

Le mécanisme d'auto-attention permet aux Transformers de modéliser efficacement les dépendances à longue distance dans les séquences, ce qui est crucial pour de nombreuses tâches de NLP comme la traduction automatique, les systèmes de questions-réponses et la modélisation de langage [55, 10]. Contrairement aux réseaux récurrents qui traitent les séquences de manière séquentielle, l'auto-attention considère toute la séquence en parallèle, ce qui améliore l'efficacité et permet un meilleur apprentissage des dépendances à longue portée

II.2 Du traitement du langage à la vision

Bien que les Transformers aient initialement été conçus pour le traitement du langage naturel, leur capacité à modéliser efficacement les relations à longue distance dans les données séquentielles les a rendus attrayants pour d'autres domaines, notamment la vision par ordinateur. Les chercheurs ont adapté l'architecture des Transformers pour traiter les données d'images, donnant naissance aux Vision Transformers (ViT) [12].

II.2.1 Les principes de base des Vision Transformers (ViT)

Les ViT fonctionnent en divisant l'image d'entrée en une séquence de patches d'image comme illustré dans la figure 2.2, qui sont ensuite traités par l'architecture Transformer stan-

ard. Un encodeur ViT prend en entrée cette séquence de patches et produit une représentation encodée de l'image, qui peut être utilisée pour diverses tâches de vision par ordinateur comme la classification d'images, la détection d'objets et la segmentation sémantique [12].

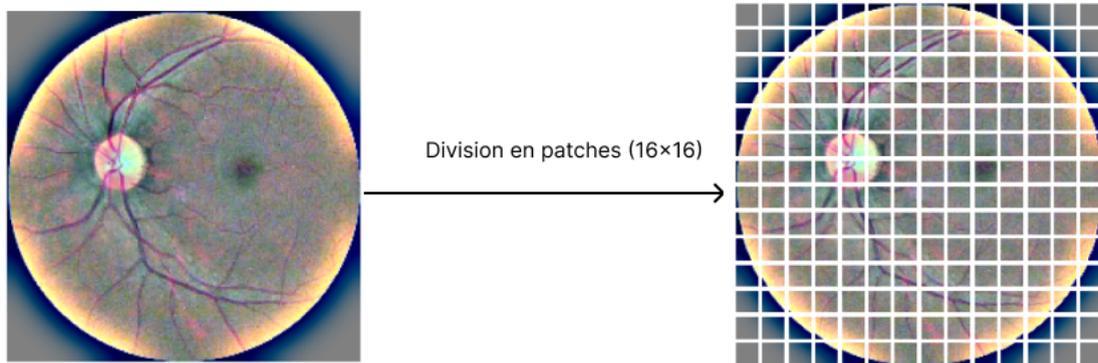


Figure 2.2: Division en patches de taille fixe

II.3 Architecture des Vision Transformers

La première étape du traitement d'une image par un ViT consiste à diviser l'image en une séquence de patches d'image voir figure 2.2.

Le transformer standard reçoit en entrée une séquence 1D d'encastres de jetons (token embeddings). Pour traiter les images 2D, l'image $x \in \mathbb{R}^{H \times W \times C}$ est remodelée en une séquence de patches 2D aplatis $x_p \in \mathbb{R}^{N \times (P^2 \cdot C)}$, où (H, W) est la résolution de l'image originale, C est le nombre de canaux, (P, P) est la résolution de chaque patch d'image, et $N = HW/P^2$ est le nombre résultant de patches, qui sert également de longueur effective de la séquence d'entrée pour le transformateur. [12]

Les encastres de position (Position embeddings) sont ajoutés aux encastres de patches (Patch embeddings) pour conserver les informations de position. [12]

Un embeddings de classification supplémentaire est également ajouté à la séquence pour permettre la tâche de classification d'images. [12]

II.3.1 Mécanisme d'auto-attention dans le contexte des images

Comme dans l'architecture Transformer originale pour le NLP, le mécanisme d'auto-attention est au cœur du ViT. Cependant, dans le contexte des images, l'auto-attention capture les relations entre les différents patches d'image plutôt qu'entre les mots ou les sous-mots [22]. Cela permet au ViT de modéliser efficacement les dépendances à longue distance dans les images.

La figure 2.3 illustre le l'architecture des ViT

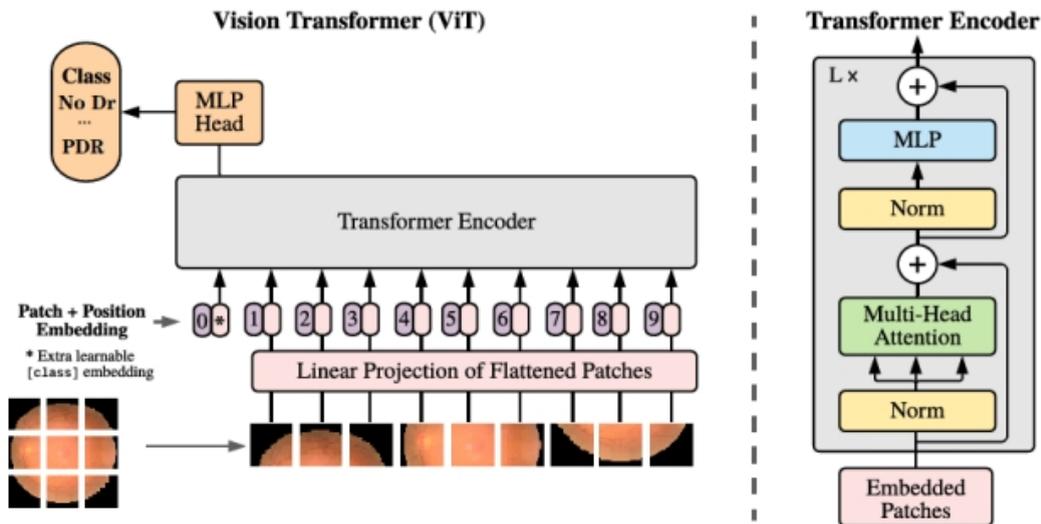


Figure 2.3: Architecture Vision Transformer [60]

II.4 Pré-entraînement et fine-tuning des ViT

Comme pour les modèles de langue naturelle, le pré-entraînement des ViT sur de grandes quantités de données suivi d'un fine-tuning sur des tâches spécifiques s'est avéré crucial pour obtenir des performances élevées [12, 6].

Le fine-tuning des ViT pré-entraînés sur des tâches spécifiques, comme la classification d'images, la détection d'objets ou la segmentation sémantique, permet de transférer les connaissances acquises lors du pré-entraînement et d'obtenir des performances compétitives avec un ensemble de données d'entraînement limité.

II.5 Applications des ViT dans la vision par ordinateur

Les Vision Transformers (ViT) ont des applications étendues dans divers domaines. Ils sont largement utilisés dans des tâches de reconnaissance d'images telles que la détection d'objets, la segmentation d'images, et la classification d'images

- **Agriculture** : Les ViT sont utilisés pour la détection des maladies des plantes, ce qui est crucial pour la santé et la croissance des plantes [29, 30]. Ils sont également utilisés pour la classification des mauvaises herbes et des cultures. [29, 30]
- **Véhicules autonomes** : Les ViT sont utilisés pour la segmentation, la détection d'objets, de piétons, de voies et plus encore. Ils jouent un rôle crucial dans le traitement visuel dynamique en temps réel, un élément essentiel pour la navigation des véhicules autonomes en toute sécurité. [27, 28]
- **Surveillance** : Les ViT sont utilisés pour la détection d'objets en temps réel dans des conditions météorologiques défavorables. Ils sont également utilisés pour la reconnaissance d'actions en temps réel dans une solution basée sur le cloud [20, 48].
- **Divertissement** : Les ViT sont utilisés pour la reconnaissance d'actions vidéo en temps réel dans une solution basée sur le cloud [20, 48]. Ils sont également utilisés pour la génération d'images et la réparation d'images [20, 48].

III Yolo

Le modèle YOLO (You Only Look Once) est une méthode de détection d'objets révolutionnaire qui a changé la façon dont les systèmes de vision par ordinateur perçoivent les images [63]. Contrairement aux approches traditionnelles de détection d'objets qui segmentent une image en plusieurs régions et effectuent la détection sur chaque région individuellement, YOLO adopte une approche holistique et examine l'image dans son ensemble lors de la détection des objets [40].

III.1 Historique de YOLO

- **YOLOv1 (Juin 2016)** : Le modèle YOLO original a été introduit par Joseph Redmon, Santosh Divvala, Ross Girshick, et Ali Farhadi [14]. Il a révolutionné la détection d'objets en la traitant comme un problème de régression, permettant une détection d'objets en temps réel [14].
- **YOLOv2 (Décembre 2016)** : Aussi connu sous le nom de YOLO9000, cette version a introduit l'utilisation de la normalisation par lots dans toutes ses couches convolutionnelles, améliorant ainsi la stabilité et les performances du modèle [14]. YOLOv2 a également introduit l'utilisation de boîtes d'ancrage (empruntées à Faster R-CNN) pour une prédiction plus précise de la forme et de la taille des objets [14].
- **YOLOv3 (Avril 2018)** : Cette version a apporté des améliorations significatives en termes de précision et de vitesse par rapport à YOLOv2 [14].
- **YOLOv4 et au-delà** : Les versions ultérieures de YOLO, jusqu'à YOLOv8, ont continué à améliorer la précision et la vitesse de détection d'objets. De plus, des variantes de YOLO ont été développées, comme YOLO-NAS et YOLO avec Transformers, qui ont apporté des innovations supplémentaires dans la détection d'objets [51].

III.2 Architecture Yolo

L'algorithme YOLO prend une image en entrée et utilise ensuite un simple réseau neuronal convolutionnel profond pour détecter des objets dans l'image. L'architecture du modèle CNN qui constitue l'épine dorsale de YOLO est présentée dans la figure 2.4 [63].

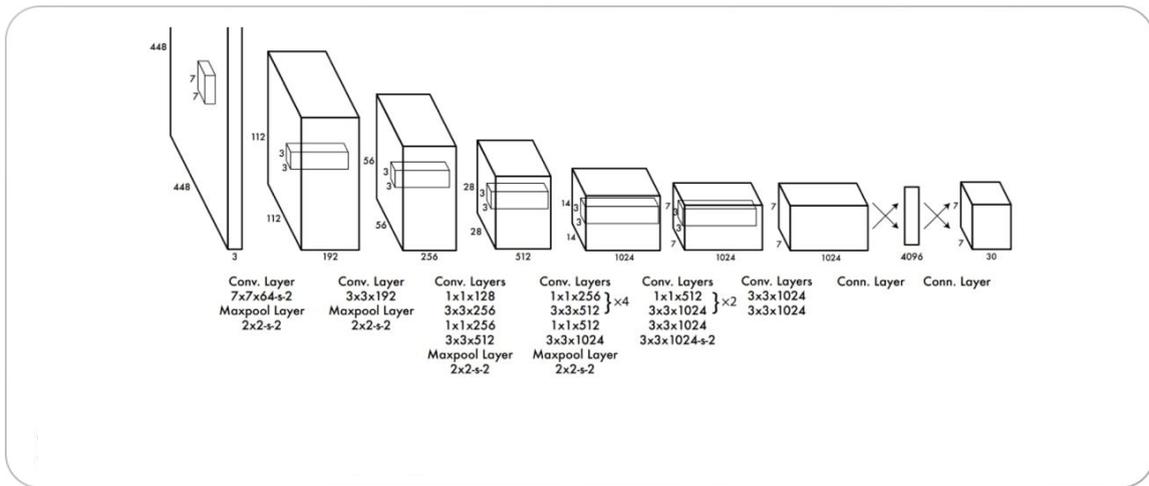


Figure 2.4: Architecture CNN de YOLO [63]

III.3 Domaine d'application

Les capacités de YOLO vont au-delà de la détection d'objets, le rendant polyvalent dans diverses applications et domaines. Voici quelques-uns des principaux domaines d'application:

- Applications en temps réel:** La vitesse de YOLO est un de ses avantages les plus significatifs. En effet, il traite les images à un taux de 45 images par seconde, ce qui en fait un choix de premier plan pour les applications en temps réel [53]. Cette vitesse, combinée à sa haute précision de détection, fait de YOLO un candidat idéal pour les applications nécessitant un traitement en temps réel.[62]
- Systèmes de surveillance:** La capacité de YOLO à détecter des objets en temps réel le rend très adapté pour les systèmes de surveillance [56]. Il peut identifier et suivre des objets, des personnes et des activités, fournissant des informations précieuses pour la sécurité et la surveillance.[56]
- Véhicules autonomes:** Dans le domaine des véhicules autonomes, YOLO peut être utilisé pour la détection d'objets en temps réel, ce qui est crucial pour la navigation et la sécurité [56]. Il peut identifier d'autres véhicules, des piétons, des panneaux de signalisation et d'autres objets pertinents, aidant le système autonome à prendre des décisions éclairées [56].
- Villes intelligentes:** Les villes intelligentes peuvent tirer parti de YOLO pour diverses applications, y compris la gestion du trafic, la surveillance des foules et la maintenance de l'infrastructure [56]. En fournissant une détection d'objets en temps réel et précise, YOLO peut contribuer à une gestion plus optimisée des villes intelligentes.[56]
- Santé:** Dans le domaine de la santé, YOLO peut être utilisé pour l'analyse d'images médicales, aidant à la détection et au diagnostic de diverses conditions. Par exemple, il peut être utilisé pour détecter des anomalies dans les images de radiographie ou les scans IRM [56].

IV Architecture Resnet50

ResNet50 est une architecture de réseau neuronal convolutionnel profond qui a été développée par Microsoft Research en 2015 [24]. Il s'agit d'une variante de l'architecture populaire ResNet (pour "Residual Network"). Le "50" dans le nom fait référence au nombre de couches du réseau, qui est de 50 couches de profondeur [24].

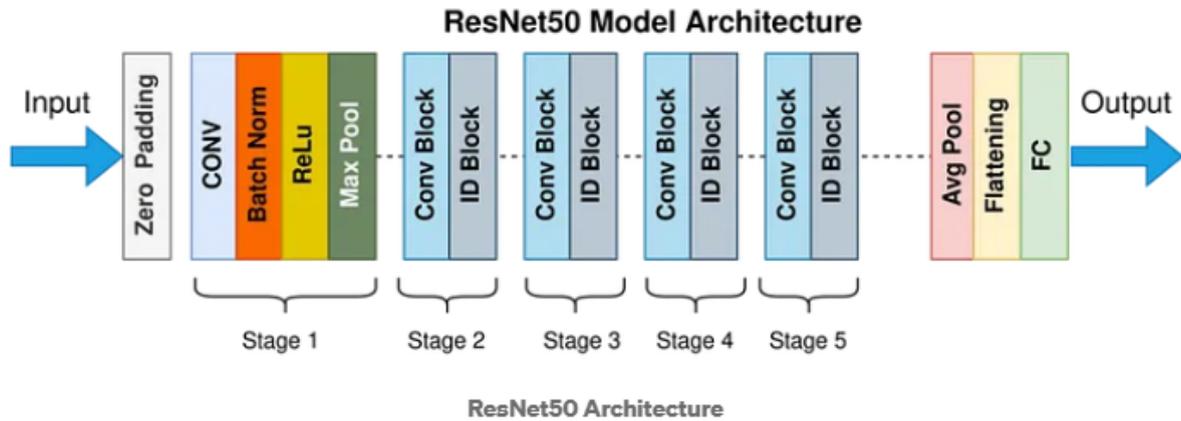


Figure 2.5: Architecture Resnet50 [24]

L'architecture de ResNet50 est divisée en quatre parties principales [24]:

- **Les couches convolutionnelles :** Elles sont responsables de l'extraction des caractéristiques de l'image d'entrée, comme les bords, les textures et les formes. [24]
- **Le bloc d'identité :** C'est un bloc simple qui passe l'entrée à travers une série de couches convolutionnelles et ajoute l'entrée à la sortie. [24]
- **Le bloc convolutionnel :** Il est similaire au bloc d'identité, mais avec l'ajout d'une couche convolutionnelle de 1x1 qui est utilisée pour réduire le nombre de filtres avant la couche convolutionnelle de 3x3. [24]
- **Les couches entièrement connectées :** Ces couches sont responsables de la classification finale. [24]

V Apprentissage par transfert et fine-tuning

L'apprentissage par transfert est une technique d'apprentissage en profondeur où un modèle formé sur une tâche est réutilisé comme point de départ pour un modèle sur une deuxième tâche [16].

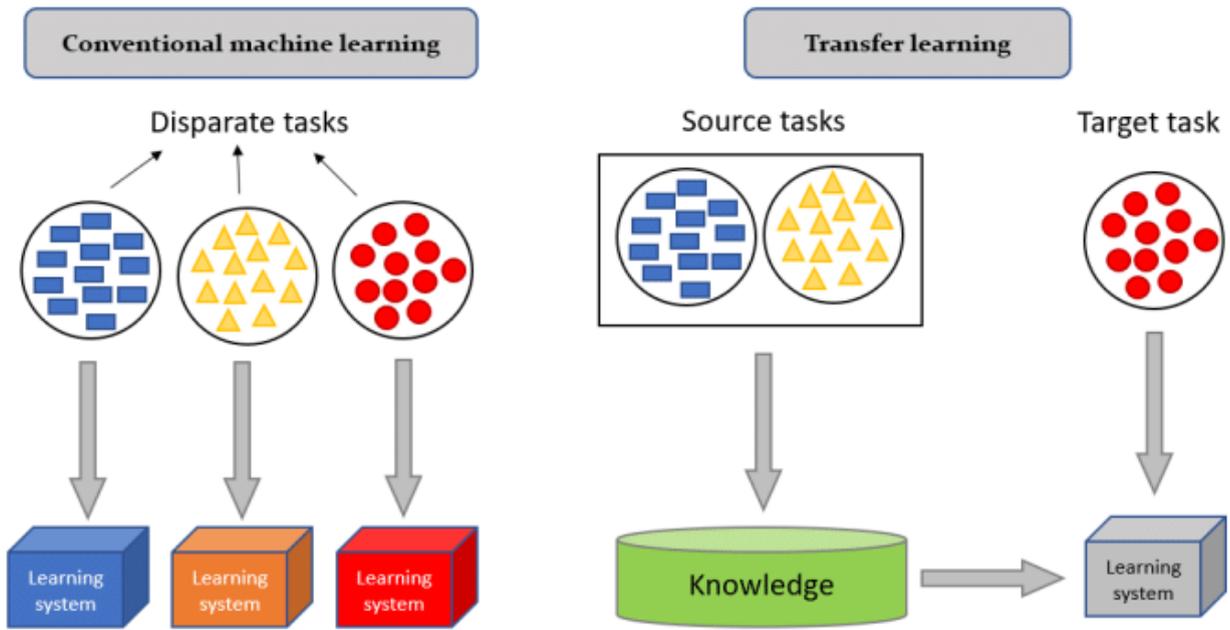


Figure 2.6: Diagramme comparatif des processus d'apprentissage entre l'apprentissage automatique conventionnel et l'apprentissage par transfert [49]

Le fine-tuning est une technique qui implique la mise à jour et l'ajustement des poids des modèles pré-entraînés sur un nouvel ensemble de données spécifique à une tâche [64]. Cela permet aux modèles de s'adapter à la nouvelle tâche tout en continuant à bénéficier des connaissances acquises précédemment, comme le montre la figure 2.6. Lors du fine-tuning, les poids des modèles pré-entraînés sont ajustés sur la base des gradients calculés à partir du nouvel ensemble de données, en prenant soin d'éviter tout surajustement [64]. Lors du fine-tuning des modèles pré-entraînés, divers hyperparamètres peuvent être ajustés pour obtenir des performances optimales. Ces hyperparamètres comprennent notamment les fonctions d'activation, les taux d'abandon (dropout rates), les taux d'apprentissage (learning rate) et la taille des lots (batch sizes).

VI Métriques de performances

Les métriques de performance sont essentielles pour évaluer et comparer les modèles de machine learning. Elles permettent de mesurer la qualité des prédictions et d'identifier les points forts et les faiblesses des algorithmes. Cette section présente les principales métriques utilisées pour les problèmes de classification.

VI.1 Accuracy

L'accuracy (exactitude) sert de mesure globale de l'exactitude du modèle dans la prédiction des étiquettes de classe correctes. Elle quantifie le pourcentage d'images correctement classées par rapport au nombre total d'images de l'ensemble de données de validation.

$$\text{Accuracy} = \frac{\text{Vrai Positives} + \text{Faux Négatives}}{\text{Vrai Positives} + \text{Faux Positives} + \text{Vrai Négatives} + \text{Faux Négatives}} \quad (2.1)$$

VI.2 Précision

La précision, quant à elle, se concentre sur la capacité du modèle à identifier correctement les cas positifs, tels que la détection de la présence de la RD dans les images rétinienne. Elle mesure la proportion de vraies prédictions positives sur l'ensemble des prédictions positives faites par le modèle.

$$\text{Précision} = \frac{\text{Vrai Positives}}{\text{Vrai Positives} + \text{Faux Positives}} \quad (2.2)$$

VI.3 Rappel

En outre, le rappel évalue la capacité du modèle à capturer toutes les instances positives, ce qui indique sa capacité à détecter correctement la présence de DR. Il calcule la proportion de vraies prédictions positives par rapport à toutes les instances positives réelles présentes dans l'ensemble de données d'évaluation.

$$\text{Rappel} = \frac{\text{Vrai Positives}}{\text{Vrai Positives} + \text{Faux Négatives}} \quad (2.3)$$

VI.4 F1 Score

Le score F1 combine la précision et le rappel, fournissant une moyenne harmonique qui prend en compte à la fois les faux positifs et les faux négatifs. Cette mesure offre une évaluation équilibrée des performances du modèle, en tenant compte à la fois de la capacité à identifier correctement les instances positives et à minimiser les erreurs de classification.

$$\text{F1-score} = \frac{2 \times \text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \quad (2.4)$$

VII État de l'art

Dans cette partie, nous présenterons une vue d'ensemble des différents travaux relatifs à la détection et à la classification de la RD.

1. Dans l'article de recherche *Vision Transformer Model for Predicting the Severity of Diabetic Retinopathy in Fundus Photography-Based Retina Images* [34], les auteurs ont utilisé l'ensemble de données FGADR, qui se compose de 1 842 images de fond d'œil.

L'approche adoptée dans l'étude a consisté à adapter le modèle Vision Transformer (ViT) spécifiquement pour la classification automatisée de la rétinopathie diabétique. Les chercheurs ont optimisé un ensemble de couches denses pour entraîner le classificateur à prédire la gravité de la rétinopathie diabétique.

L'étude a évalué les performances du modèle ViT à l'aide de diverses mesures, notamment le score F1, l'accuracy, le rappel et la spécificité. Les résultats obtenus de l'étude (F1-score, accuracy, le rappel) de 82,5%, 82,5% et 82,5%, respectivement.

2. Dans une autre étude intitulée *Classification of diabetic retinopathy severity in fundus images using the vision transformer and residual attention* [15] utilise l'ensemble de données DDR pour la classification de la RD. Cet ensemble de données comprend 13 673 images de fond d'œil. En outre, l'étude évalue également le modèle sur l'ensemble de données IDRiD. L'approche proposée s'appuie sur un vision transformer (ViT) et un

mécanisme d'attention résiduelle. La méthode se compose de deux modules principaux : le bloc d'extraction des caractéristiques (FEB) et le bloc de prédiction de la classification (GPB).

- Feature Extraction Block (FEB)
- Grading Prediction Block (GPB)

Les performances du modèle proposé sont évaluées à l'aide de plusieurs mesures, notamment la précision, le rappel, la spécificité, l'accuracy et le score F1. Les résultats obtenus de l'étude (spécificité, sensibilité, accuracy) de 82,45%, 81,40% et 82,35%, respectivement.

3. L'article intitulé *Diabetic Retinopathy Classification Using Swin Transformer with Multi Wavelet* [11] a utilisé l'ensemble de données Kaggle APTOS 2019. Les auteurs ont proposé une nouvelle technique qui combine la transformée multi-Wavelet avec le transformateur de Swin pour la classification automatique des niveaux de progression de la rétinopathie diabétique. Les auteurs ont évalué les performances de leur approche à l'aide de plusieurs métriques :

- Pour une classification binaire (Pas de DR vs. DR) :
 - (a) Accuracy: 97%
 - (b) F1 score: 98,73%
 - (c) Sensitivité: 95,20%
 - (d) Spécificité: 95,20%
- Pour la classification multiclasse (5 classes) :
 - (a) Accuracy: 82%
 - (b) F1 score: 90,22%
 - (c) Sensitivité: 46,26%
 - (d) Spécificité: 80,17%

4. Dans l'étude menée par Mohammed Oulhadj et ses collègues [36], les chercheurs se sont concentrés sur la détection automatique de la rétinopathie diabétique en utilisant des techniques avancées d'apprentissage profond. La recherche a utilisé quatre ensembles de données clés : APTOS, EyePACS, Messidor-2 et DDR.

L'approche proposée dans cette étude implique une nouvelle étape de prétraitement d'ensemble qui combine les méthodes Patch Local Ternary (PLT) et Contrast Limited Adaptive Histogram Equalization (CLAHE). En outre, un modèle d'apprentissage profond a été développé en affinant le vision transformer (ViT) avec un réseau de capsules modifié.

Les résultats obtenus à partir des expériences à travers tous les ensembles de données sont les suivants:

- Aptos:
 - (a) Accuracy: 88.18%
 - (b) Précision: 80%
 - (c) Rappel: 76%
 - (d) F1-Score: 78%

- DDR:
 - (a) Accuracy: 80.36%
 - (b) Précision: 59%
 - (c) Rappel: 55%
 - (d) F1-Score: 56%
- Messidor-2:
 - (a) Accuracy: 87.78%
 - (b) Précision: 87%
 - (c) Rappel: 87%
 - (d) F1-Score: 87%
- EyePACS:
 - (a) Accuracy: 78.64%
 - (b) Précision: 54%
 - (c) Rappel: 42%
 - (d) F1-Score: 46%

VIII Conclusion

Dans ce chapitre, nous avons défini l'apprentissage par transfert et discuté de ses avantages. Nous avons également exploré les différentes architectures des modèles d'apprentissage profond utilisés dans la détection et la classification de la RD. Ensuite, nous avons présenté les études les plus récentes sur le classement de la rétinopathie diabétique.

Ces études ont prouvé que l'emploi de grandes quantités de données, l'utilisation de paramètres optimaux et les méthodes d'apprentissage par transfert peuvent aboutir à des résultats très intéressants. Toutefois, il est primordial de tenir compte des défis spécifiques liés au déséquilibre des ensembles de données médicales lorsqu'on utilise ces techniques.

Notre objectif est de concevoir trois modèles différents en utilisant les visions transformers, l'algorithme Yolo, et enfin une combinaison du modèle Resnet50 et du ViT pour classer la RD.

Dans le prochain chapitre, nous examinerons trois approches de modèle basées sur l'apprentissage profond et l'apprentissage par transfert, en nous appuyant sur les enseignements de ces études.

En utilisant cette approche, nous serons en mesure de détecter et de classer de manière efficace la RD.

Table 2.1: Long table caption.

Beginning of Table		
Auteurs/Année	Méthode	Performance
Nazih et al[34] 2023	ViT modifié	FGADR Accuracy: 82,5% F1-Score: 82,5% Rappel: 82,5%
Gu et al[15] 2023	Vision Rrans- former et un mécanisme d'attention résiduelle	DDR Spécificité: 82,45% Sensitivité: 81,40% Accuracy: 82,35%
Dihin et al[11] 2023	Combiner Swin Trans- former Et multi- Wavelet	APTOS 2019[19] Binaire: Accuracy: 97% Sensitivité: 95,20% Spécificité: 95,20% F1-Score: 98,73%
		5-classes: Accuracy: 88.14% Precision: 88% Recall: 88% F1-Score: 88,2%

Continuation of Table 2.1

Oulhadj et al[36] 2024	Vision trans- former avec un réseau de capsules modifié	APTOS, EyePACS, Messidor-2 et DDR	Aptos: Accuracy: Précision: Rappel: F1-Score: DDR: Accuracy: Précision: Recall: F1-Score: Messidor-2: Accuracy: Précision: Rappel: F1-Score: EyePACS: Accuracy: Précision: Rappel: F1-Score:	88,18% 80% 76% 78% 80.36% 59% 55% 56% 87,78% 87% 87% 87% 78,64% 54% 42% 46%
---------------------------	---	---	---	--

Continuation of Table 2.1

Yang et al[61] 2024	Vision trans- former avec des masques autoen- codeurs	APTOS, EyePACS, Messidor-2 et, OIA-DDR	APTOS Messidor-2, Accuracy: Sensitivité: Spécificité: AUC:	without OIA-DDR: 90.41% 95.27% 90.78% 97.19%	labels, OIA-DDR:
Wu et al[58] 2021	Vision trans- former	(Non commu- nique)	APTOS labels, OIA-DDR, Accuracy: Sensitivité: Spécificité: AUC:	without Messidor-2, EyePACS: 92.60% 97.30% 92.63% 98.03%	la- Messidor-2, EyePACS:
			Accuracy: Précision: sitivité: Spécificité: AUC:	91.40% 92.80% 92.60% 97.70% 98.60%	Sen-
End of Table					

Chapitre 3

Conception & Réalisation

I Introduction

Notre objectif est de développer un modèle basé sur le transfert learning pour la prédiction et la classification de la rétinopathie diabétique. Afin d'améliorer la prédiction et la classification de la RD, nous présenterons dans ce chapitre différentes architectures basées sur le transfert learning pour la résolution de ce problème tels que Resnet50, Yolo et ViT. Une combinaison de ses architectures est développée et une comparaison des résultats est effectuée.

II Méthodologie

Nous avons adapté et expérimenté différents modèles basés sur le transfert learning tels que ViT(Vision Transformer) et Yolo (You Only Look Once) pour la détection et la classification de la RD en 3 et 5 classes puis nous avons proposé une architecture hybride combinant (PAD2C, PAD3C, PAD5C) [1] qui est une architecture Resnet50 modifiée ou on ajoute des couches supplémentaires (Flatten , Droupout, Dense) et ViT pour la résolution du même problème. Les différentes étapes nécessaires à la résolution du problème voir la figure 3.1 seront développées dans les sections qui suivent.

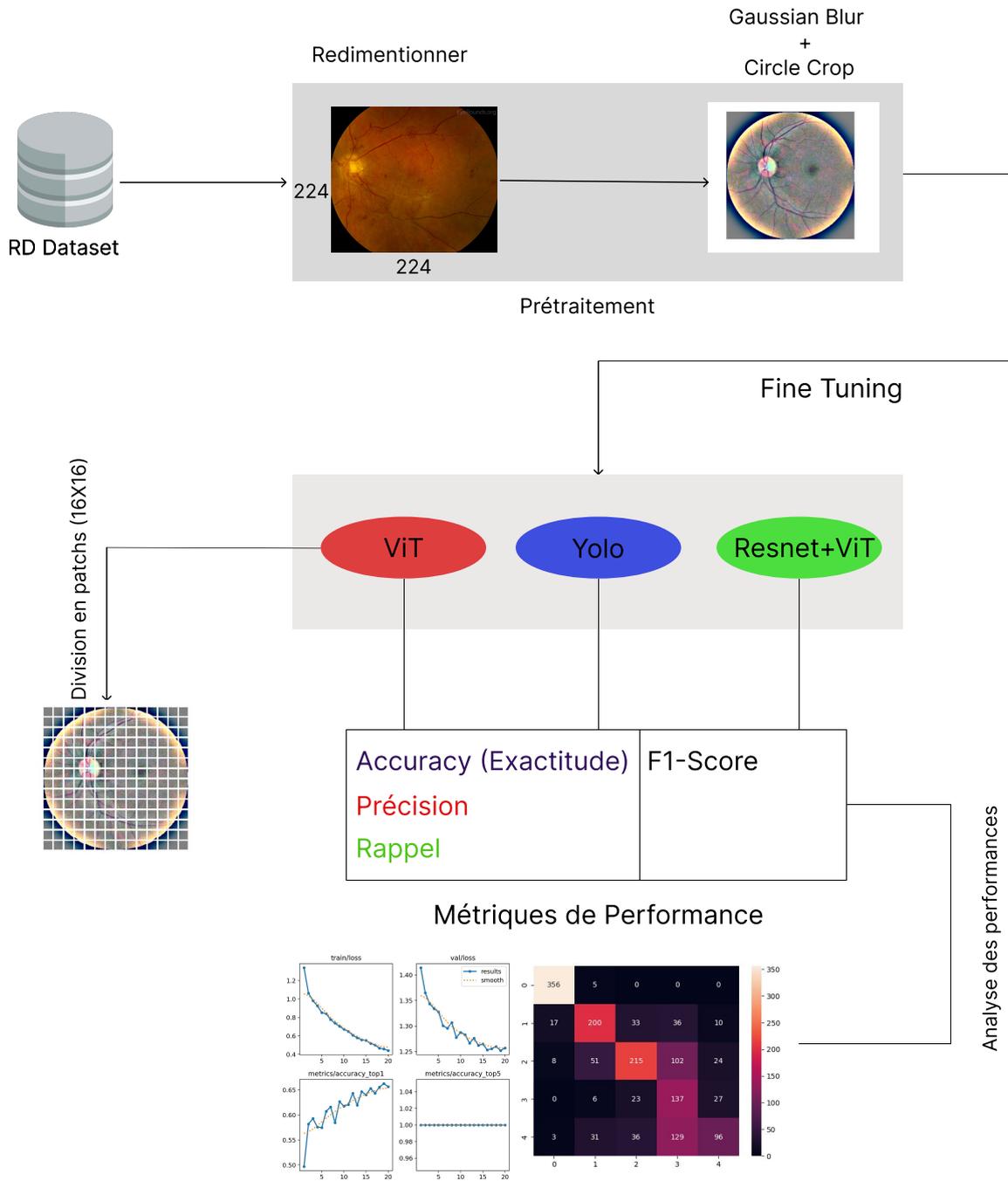


Figure 3.1: Architecture Proposée

II.1 Préparation des données

II.1.1 Dataset

Tous les modèles que nous avons développés ont été entraînés en utilisant la base de données publique APTOS 2019 Blindness [19] Detection, qui contient 3662 images. Cette base de données est divisée en cinq classes : Pas de DR (0), Léger (1), Modéré (2), Sévère (3) et DR proliférante (4). La répartition des données est illustrée dans 3.2.

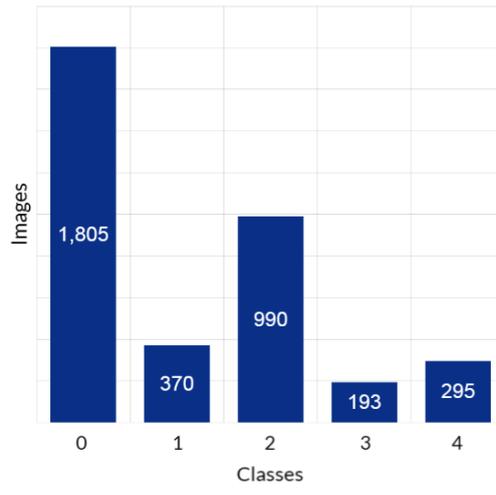


Figure 3.2: Dataset Aptos

II.1.2 Prétraitement d'images

Afin d'améliorer la visibilité des caractéristiques, notamment les vaisseaux sanguins, les lésions rouges et les exsudats, des techniques de prétraitement spécifiques sont choisies en raison de la grande quantité de données disponibles.

2.a Redimensionner

Les tailles des images des ensembles de données varient en fonction des sources et des périodes durant lesquelles elles ont été collectées. Donc, pour garantir des performances optimales pendant le traitement, il a été nécessaire de réduire toutes les images à une taille fixe de 224x224.

2.b Grey Scale & Gaussian Blur

Pour obtenir des résultats améliorés dans le modèle d'entraînement, il est crucial que les images et leurs caractéristiques soient facilement repérables. Ainsi, la première méthode utilisée pour améliorer les fonctionnalités est la conversion des images couleur en niveaux de gris, puis l'utilisation d'un filtre de flou gaussien pour réduire le bruit de l'image. Une approche de masquage pondéré est employée afin d'accentuer les bords de l'image obtenue [5].

Les images montrant des ecchymoses et des rougeurs, en particulier les vaisseaux sanguins, ont été améliorés par rapport aux originaux voir Fig.3.3.

2.c Circle Crop

Une technique de recadrage circulaire a été utilisée pour améliorer les performances lors de l'entraînement du modèle en éliminant les pixels noirs indésirables (l'arrière-plan) et en se concentrant davantage sur la partie de l'image qui représente le fond de l'œil. Pour réaliser un recadrage circulaire, un masque est créé et appliqué à l'image afin de ne garder que les pixels dans le cercle, tout en éliminant les autres [5] voir figure 3.3.

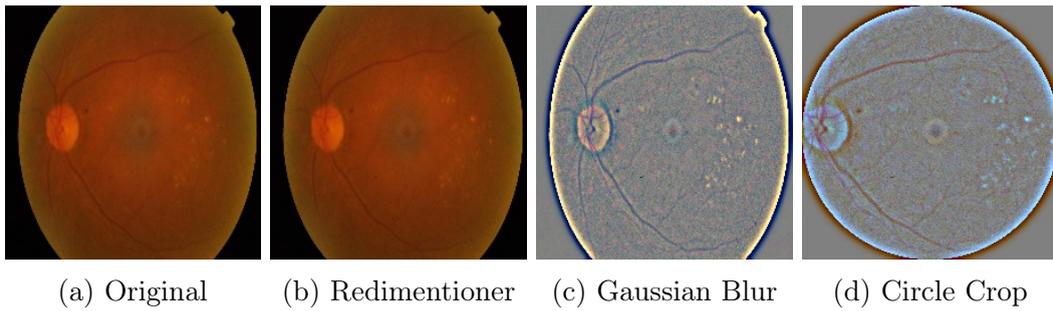


Figure 3.3: Phases de prétraitement

2.d Normalisation

Pour faciliter la procédure d'apprentissage de nos modèles, nous avons employé une technique de normalisation sur les images d'entrée. Le principe consiste à convertir les valeurs des pixels des images de la plage originale de 0 à 255 vers une plage normalisée de 0 à 1, comme le montre la figure 3.4. Ce processus de normalisation garantit que toutes les valeurs de pixels sont mises à l'échelle de manière appropriée, permettant un entraînement plus efficace et stable du modèle. En normalisant les valeurs, nous réduisons l'impact potentiel des plages d'intensité de pixels variables, améliorant ainsi la convergence et les performances de nos modèles.



Figure 3.4: Normalisation

II.1.3 Fine Tuning

Afin d'optimiser les performances des modèles pré-entraînés ViT, YOLO et ResNet+ViT, nous avons effectué un ajustement manuel des paramètres. Nous avons recherché minutieusement les paramètres les plus adaptés à chaque modèle et ensemble de données, y compris la taille d'image, la taille des lots, le taux d'apprentissage, etc.

II.2 Algorithme

Le développement des futures architectures suivront toutes l'algorithme algorithm 1:

Algorithm 1: Traitement, séparation et classification d'images de fond d'œil

```
1: Entrée: Dataset d'images de fond d'œil annotées de 0 à 4
2: Prétraitement des images:
   for each image  $\in$  dataset do
3:
   Redimensionner l'image en (224x224)
4: Appliquer un flou gaussien
5: Appliquer un circle_crop
6:
7: Data Augmentation:
   for each image  $\in$  (3  $\vee$  5) classes do
8:
   Appliquer des techniques d'augmentation de données
9:
10: Séparation du dataset:
11: Diviser le dataset en ensemble d'entraînement et ensemble de test
12: Fine-tuning des modèles:
13: Initialiser Vision Transformer (ViT)
14: Initialiser YOLO
15: Initialiser ResNet+ViT
   for chaque modèle do
16:
   Effectuer le fine-tuning sur l'ensemble d'entraînement
17:
18: Évaluation des performances: for chaque modèle do
19:
   Évaluer les performances sur l'ensemble de test
20: Calculer les métriques de performance
21: Analyser les résultats
22:
23: Sortie: Modèles entraînés et analyse des performances
```

II.3 Détection de la RD

Dans cette section nous abordons la détection de la RD qui consiste à déterminer la présence ou de l'absence de la RD (classification binaire 0/1). pour cela on regroupe les classes 1,2,3 et 4 en classe 1 nous obtenons deux classes : 0 et 1, comme l'illustre la figure 3.5. La méthodologie employée repose sur l'utilisation du dataset APTOS [19].

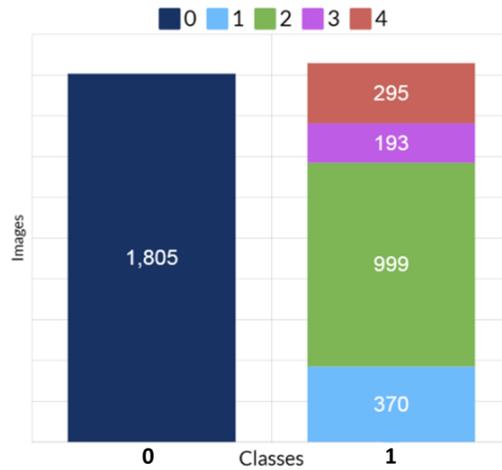


Figure 3.5: Division des classes du dataset

Après le prétraitement, le dataset est divisé en un ensemble d’entraînement et un ensemble de test. L’ensemble d’entraînement est utilisé pour l’apprentissage du modèle, tandis que celui de test servira à évaluer les performances finales du modèle.

Les différentes architectures, les résultats détaillés de ces évaluations ainsi que leurs analyses sont présentés dans les prochaines sections.

II.3.1 Approche basée ViT

Dans cette section, nous présentons notre approche basée sur les Vision Transformers pour la prédiction la rétinopathie diabétique.

2.1 Architecture ViRd basé sur les Vit

Nous avons proposé une architecture nommée **ViRd** (Vision transformer Retino Detection) Vision Transformer (ViT) pour effectuer une classification binaire. L’image d’entrée est d’abord divisée en patches $(224 \times 224) / (16 \times 16) = 256$. Ensuite, ces derniers sont traités de manière séquentielle à l’aide d’encodeurs Transformer. Enfin, la tâche de classification binaire est réalisée. L’architecture **ViRd** est illustrée dans la figure 3.6.

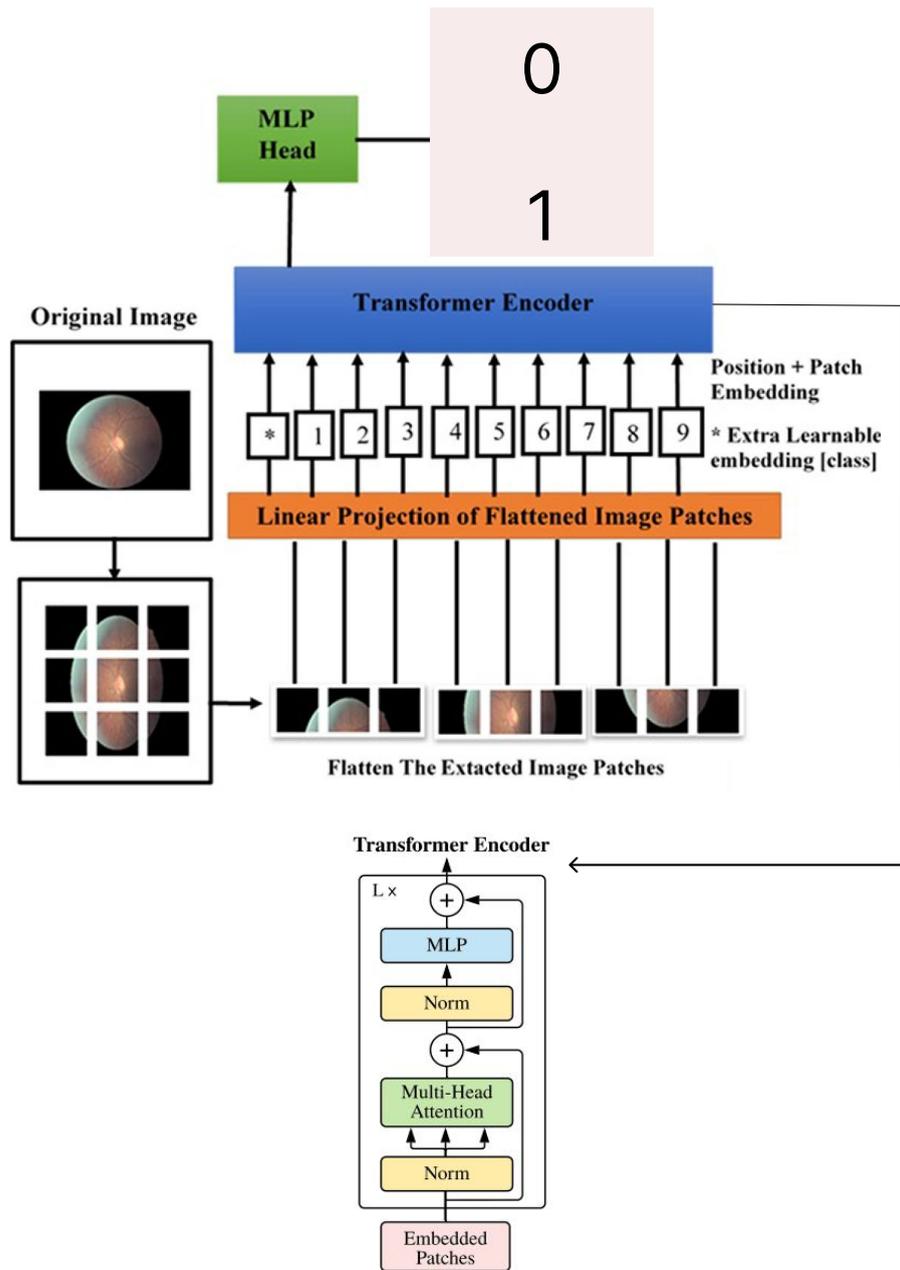


Figure 3.6: Architecture **ViRd** pour la détection de la RD basée ViT

Comme tous les modèles de transfert learning, le ViT est pré-entraîné sur un grand ensemble de données d'images génériques, avant d'être transféré à notre tâche spécifique de classification de la rétinopathie. Nous avons alors procédé à une phase de fine-tuning sur les données APTOS, et adapté les poids du modèle pré-entraîné à notre problème.

Dans le cadre de ViT (Vision Transformer), nous avons examiné différentes tailles d'images d'entrée, dimensions de patches et profondeurs de modèle afin d'optimiser les performances.

Les paramètres sélectionnés après ajustement sont illustrés dans le tableau 3.1

Pendant l'entraînement, nous avons suivi l'évolution des métriques de performance telles que la précision, le rappel et le score F1 sur un ensemble de validation.

Une fois l'entraînement terminé, nous avons évalué les performances finales du modèle sur l'ensemble de test indépendant. Les résultats obtenus montrent que notre approche à base de Vision Transformers atteint des niveaux de précision compétitifs par rapport à d'autres

méthodes de l'état de l'art pour la classification de la rétinopathie diabétique.

Table 3.1: Hyperparameters du fine tuning

Paramètre	Valeur
Validation	20% Du dataset
Test	30% De validation
Taille de l'image	224x224
Taille des patches (ViT)	16x16
Train Batch Size	32
Test Batch Size	64
warmup steps	500
Warmup learning rate	0.00001
Epochs	15
Learning rate	0.00002
Wieht decay	0.01

2.1.a Résultats

Pour évaluer les capacités de détection de la RD du modèle ViT, nous avons effectué une expérimentation avec l'ensemble de données Aptos. Un ensemble de test comprenant 733 images a été choisi pour évaluer les modèles.

Comme le montre le tableau 3.2 les résultats utilisant notre méthode sur l'ensemble de données APTOS ont atteint une précision impressionnante de 97,73% en utilisant le modèle ViT.

Table 3.2: Résultats ViRd

		ViRd	
		Class 0	Class 1
APTOS	Accuracy	0.9773	
	Precision	0.97	0.98
	Recall	0.98	0.97
	F1_score	0.98	0.98

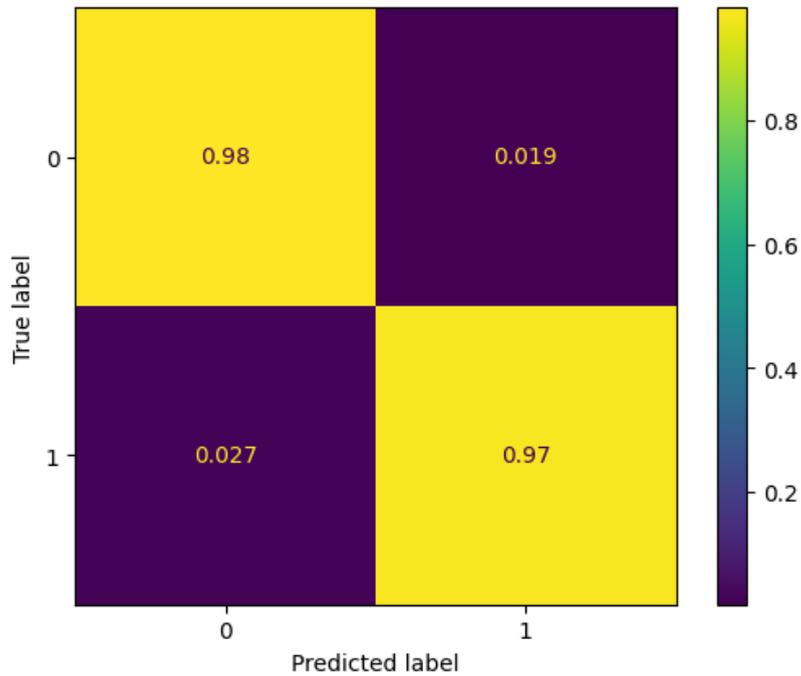
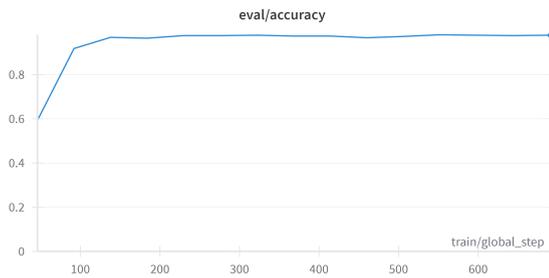
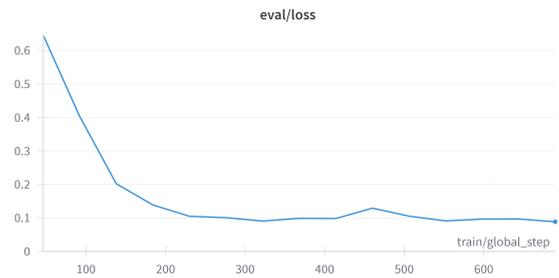


Figure 3.7: Matrice de confusion de ViRd



(a) Graphes d'exactitude



(b) Graphe de perte

Figure 3.8: Graphes d'exactitude et de perte (ViT)

La matrice de confusion 3.7 illustre les prédictions faites par notre modèle ViRd par rapport aux étiquettes réelles des données de test.

Les résultats obtenus pour la détection de la rétinopathie diabétique sont remarquables.

Ces résultats démontrent non seulement la robustesse et la fiabilité du modèle avec une accuracy de 97,73% et une précision de 97,72%, mais aussi son potentiel à être un outil précieux dans le dépistage de la rétinopathie diabétique. La faible proportion de personnes malades classées comme non malades est particulièrement impressionnante et souligne l'efficacité du modèle à minimiser les erreurs critiques, garantissant ainsi un diagnostic précis et fiable pour la majorité des patients.

II.3.2 Approche basée Yolo

2.2 Architecture de détection de la RD basée sur Yolo

Nous avons proposé une architecture nommée **RDYol** (Retino detection based Yolo) basé sur le modèle préentraîné YOLO pour la détection de la RD. Dans l'architecture **RDYol**

l'image d'entrée est divisée en une grille de cellules, et pour chaque cellule, elle prédit des boîtes englobantes et les probabilités de classe pour les objets qui y sont détectés. Son architecture est composée de couches convolutionnelles, de couches de regroupement (pooling) et de couches entièrement connectées. La figure 3.9 illustre l'architecture du modèle **RDYol** proposée.

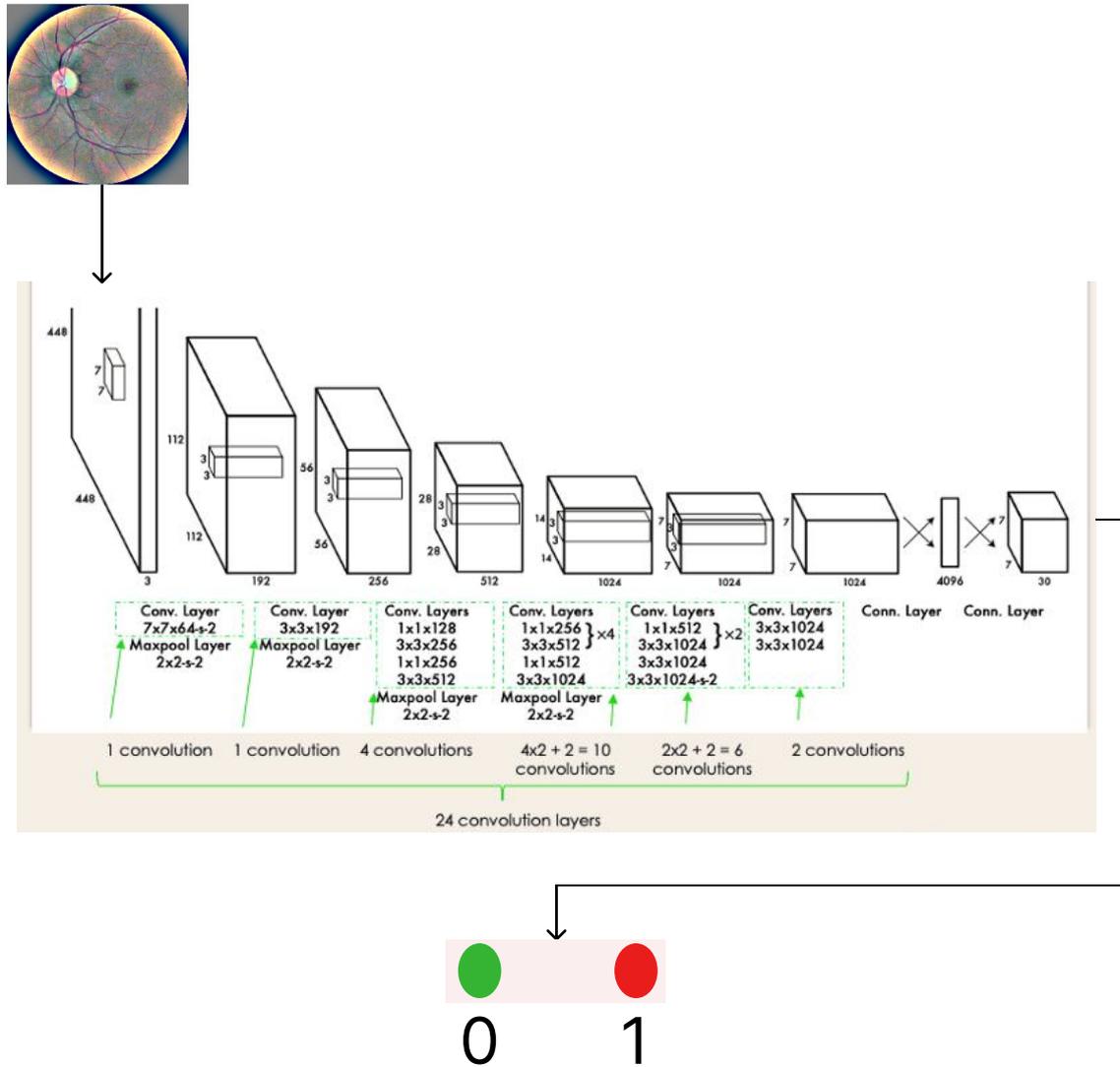


Figure 3.9: Architecture **RDYol** pour la détection de la RD

2.2.a Résultats

Pour évaluer les capacités de détection de la RD du modèle **RDYol**, nous avons effectué une expérience avec l'ensemble de données Aptos. Un ensemble de tests comprenant 733 images a été choisi pour évaluer le modèles.

Comme le montre le tableau 3.3 les résultats obtenues ont atteint une précision impressionnante de 97,95% .

Table 3.3: Résultats RDYol

		RDYol	
		Class 0	Class 1
APTOS	Accuracy	0.9795	
	Precision	0.98	0.98
	Recall	0.98	0.98
	F1_score	0.98	0.98

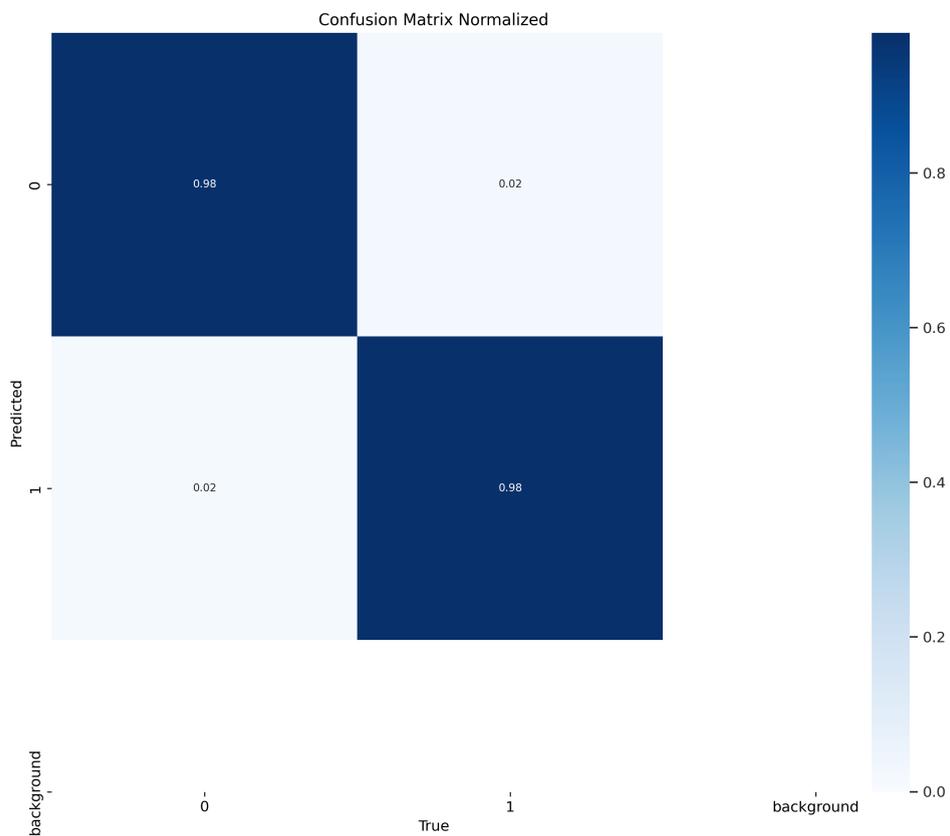


Figure 3.10: Matrice de confusion **RDYol**

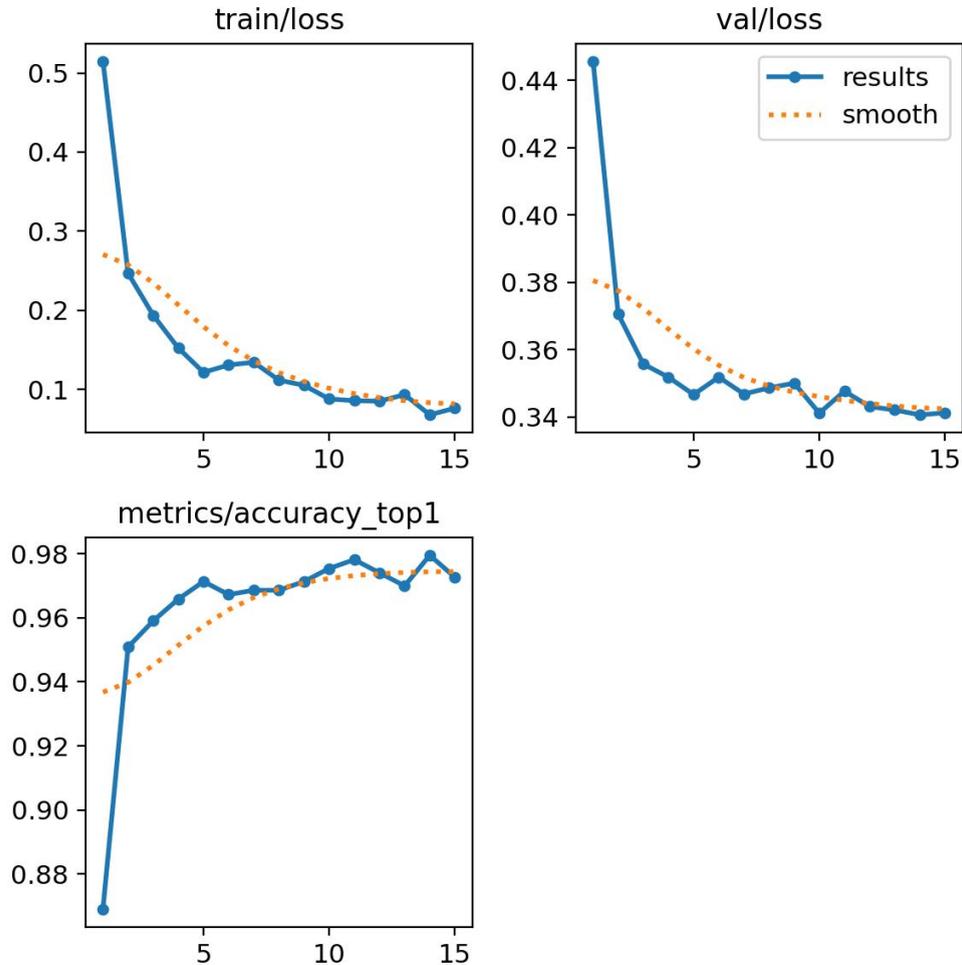


Figure 3.11: Graphe d'exactitude et de perte pour le RDYol

La matrice de confusion 3.10 illustre les prédictions faites par notre modèle par rapport aux étiquettes réelles des données de test.

Le modèle montre une performance élevée avec une précision (accuracy) globale de 97.95%. La précision et le rappel sont également très élevés, respectivement à environ 98% et 98%, ce qui indique que le modèle est très bon pour classifier les patient positifs et négatifs.

II.3.3 Approche Hybride

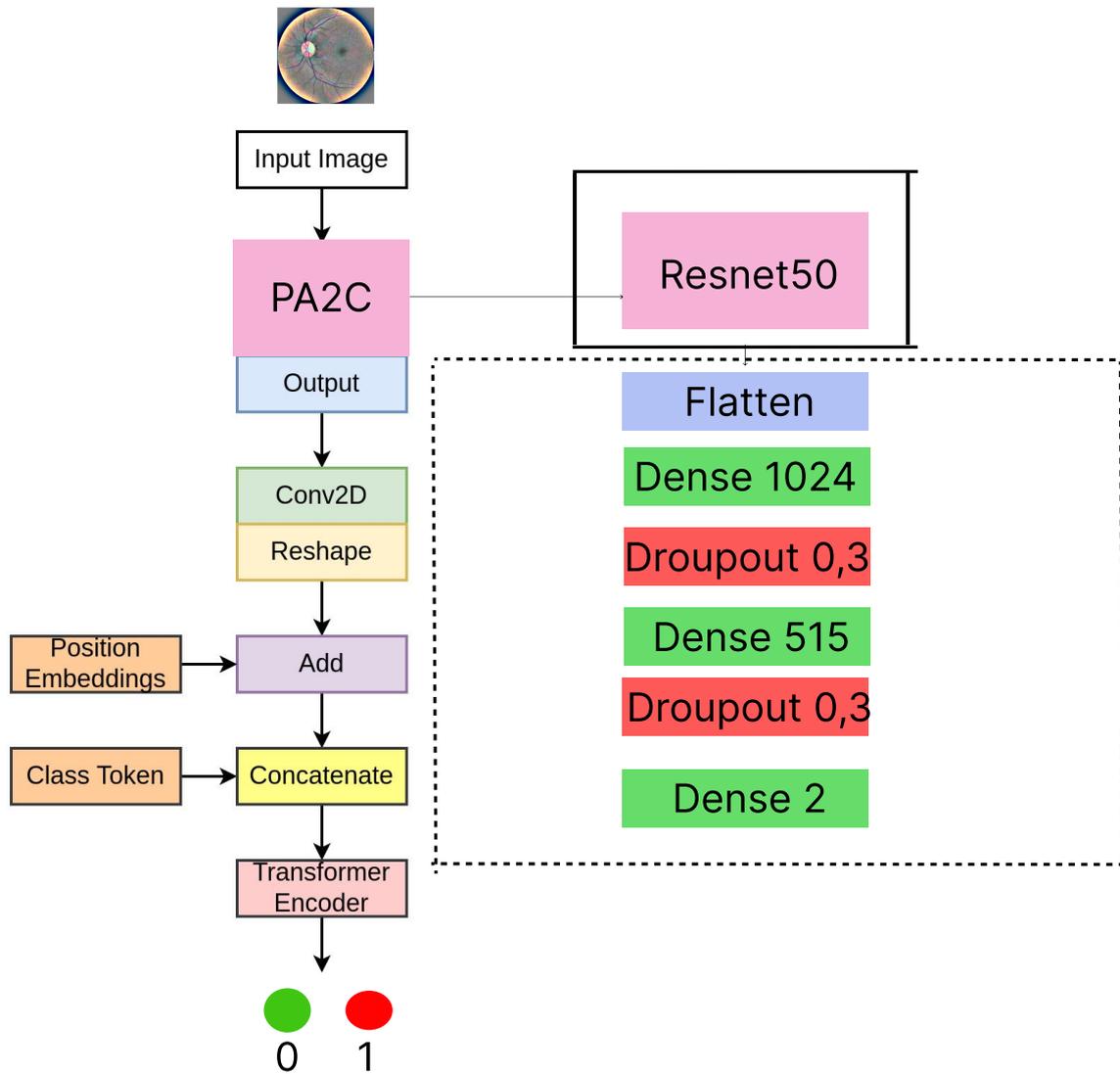
Nous proposons un modèle hybride en combinant un réseau de neurones convolutionnels (PA2C) [1] et un transformateur de vision (ViT) pour effectuer la classification binaire (0 ou 1) de la RD.

2.3.a Modèle hybride ReVi-RD

Pour procéder à cette classification nous avons choisi d'utiliser les poids du modèle PA2C en supprimant les six dernières couches (Flatten, Dense (1024), Droupout (0.3), Dense (515), Droupout (0.3), Dense (3)) utilisé l'année passée dans la thèse [1] pour les charger dans notre première partie du modèle ReVi-2C (la partie Resnet50).

Nous allons cependant modifier La sortie du réseau ViT en l'adaptant à une classification en 2 classes (0, 1)

L'architecture proposée est illustrée dans la figure 3.12


 Figure 3.12: Architecture hybride **ReVi-RD** pour la détection de la RD

2.3.b Résultats

Pour évaluer les capacités de détection de la RD du modèle **ReVi-RD**, nous avons effectué une expérimentation avec l'ensemble de données d'Aptos. Un ensemble de tests comprenant 220 images a été choisi pour évaluer le modèles.

Comme le montre le tableau 3.4 les résultats utilisant notre méthode sur l'ensemble de données APTOS ont atteint une précision impressionnante de 99,55% en utilisant le modèle Hybride.

 Table 3.4: Résultats **ReVi-RD**

		ReVi-RD	
		Class 0	Class 1
APTOS	Accuracy	0.9955	
	Precision	0.99	1.00
	Recall	1.00	0.99
	F1_score	1.00	1.00

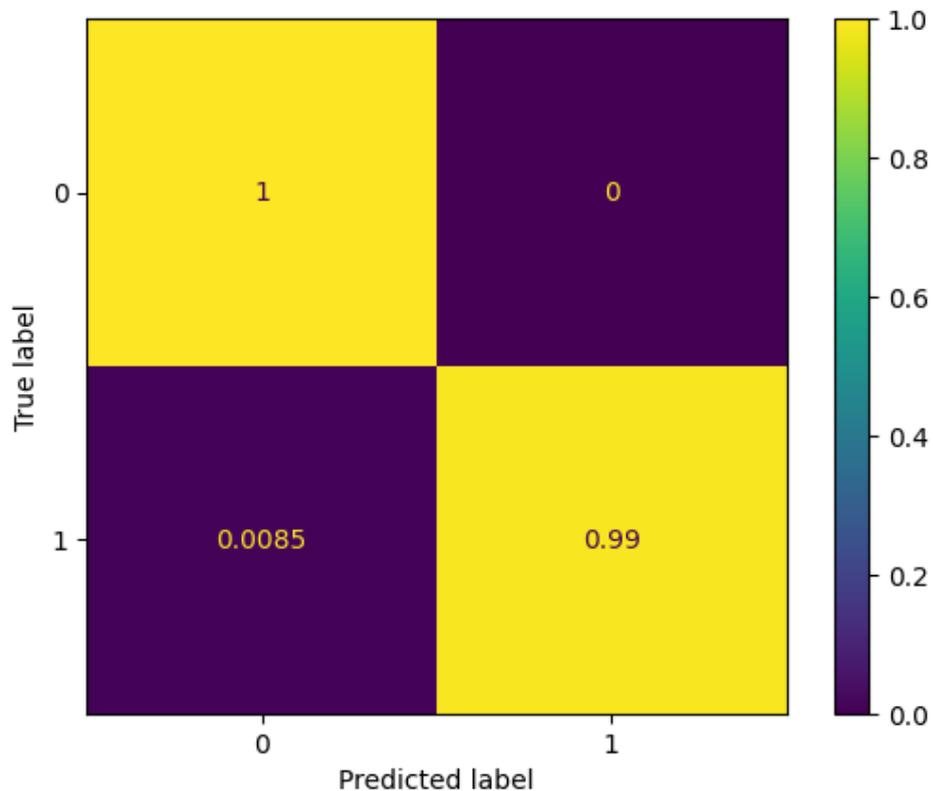
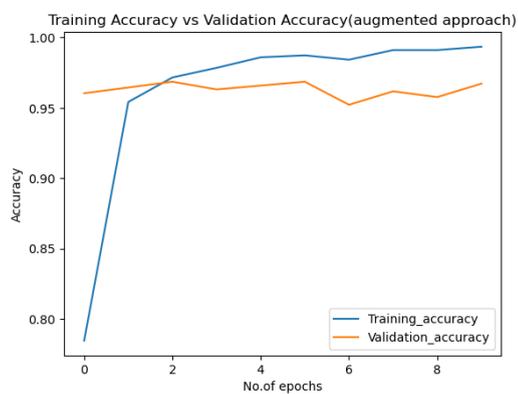
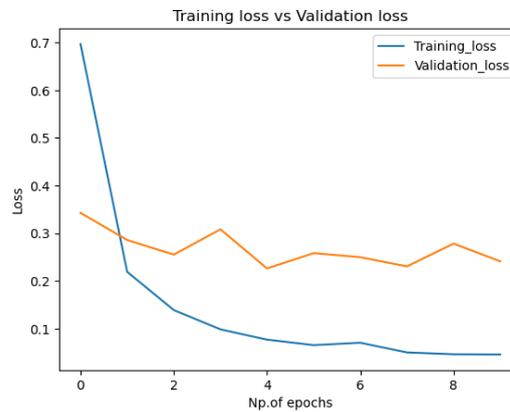


Figure 3.13: Matrice de confusion de **ReVi-RD**



(a) Graphes d'exactitude



(b) Graphe de perte

Figure 3.14: Graphes d'exactitude et de perte **ReVi-RD**

La matrice de confusion 3.13 illustre les prédictions faites par notre modèle par rapport aux étiquettes réelles des données de test.

II.3.4 Comparaison des modèles

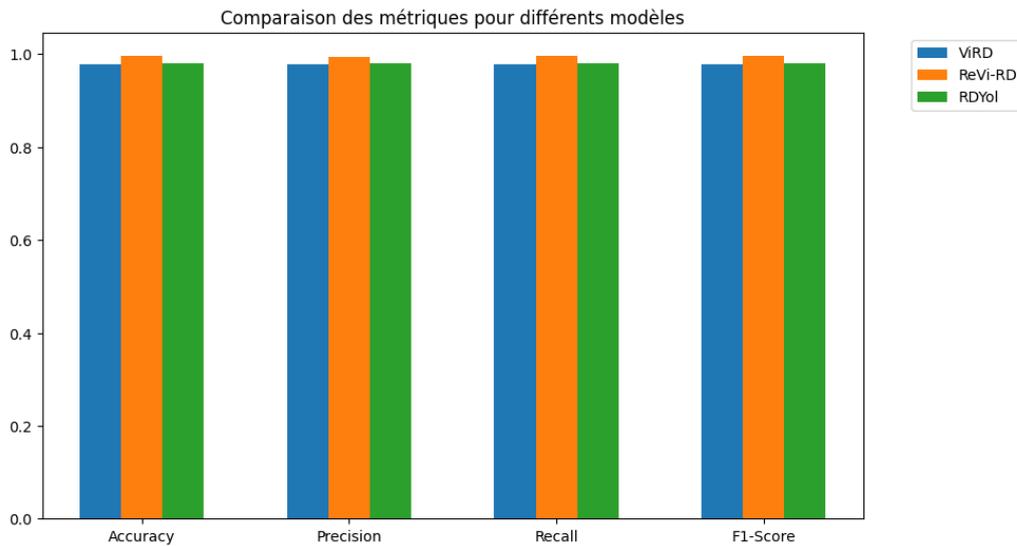


Figure 3.15: Comparaison des métriques les des différents modèles.

Pour comparer les performances des modèles proposés ViRd, RDYOL, et ReVi-RD, nous analysons leurs résultats d'entraînement sur le jeu de données APTOS en se concentrant sur les mesures clés telles que l'exactitude, la précision, le rappel et le F1-score pour les deux classes (Classe 0 et Classe 1).

Le modèle ReVi-RD se distingue par une exactitude de 0,9955, surpassant ainsi RDYol (exactitude de 0,9795) et ViRD (exactitude de 0,9773). Cette performance globale montre que ReVi-RD a une meilleure capacité à classer correctement les échantillons.

En termes de précision, rappel et F1-score, le modèle ReVi-RD maintient des valeurs élevées pour les deux classes, ce qui indique des prédictions très fiables et équilibrées. En comparaison, le modèle RDYOL affiche des scores légèrement inférieurs avec une précision de 0,98 et un rappel de 0,98 pour les deux classes, montrant une légère baisse de performance par rapport à ReVi-RD.

Le modèle ReVi-RD se distingue également par sa constance. Les scores de précision, rappel et F1 sont quasiment identiques pour les deux classes, ce qui montre une performance équilibrée et fiable pour les différentes catégories. En revanche, RDYOL, bien que performant, montre une légère baisse avec des scores de précision et de rappel constants mais inférieurs à ceux de ReVi-RD.

En conclusion, le modèle ReVi-RD présente des performances supérieures par rapport à RDYOL et ViRD sur toutes les mesures évaluées, notamment par sa constance et son équilibre entre les différentes classes. Cela met en évidence la robustesse et la fiabilité de l'approche hybride pour les tâches de classification sur le jeu de données APTOS.

II.3.5 Comparaison avec d'autres travaux

La comparaison avec les autres travaux est illustrée dans le tableau 3.5 et le graphe 3.16

Table 3.5: Comparaison des résultats

metrics	[13]	[59]	[21]	[18]	[39]	[61]	ViRd	RDYOI	ReVi-RD
Accuracy	86%	94.50%	97.77%	98.36%	96.59%	92.05%	97.73%	97.95%	99.55%
Precision	85%	/	100.00%	98.37%	97%	/	97.72%	98%	99.51%
Recall	86%	/	96.00%	98.36%	97%	/	97.73%	98%	99.58%
F1_Score	85%	/	97.90%	98.37%	96.59%	/	97.73%	98%	99.54%

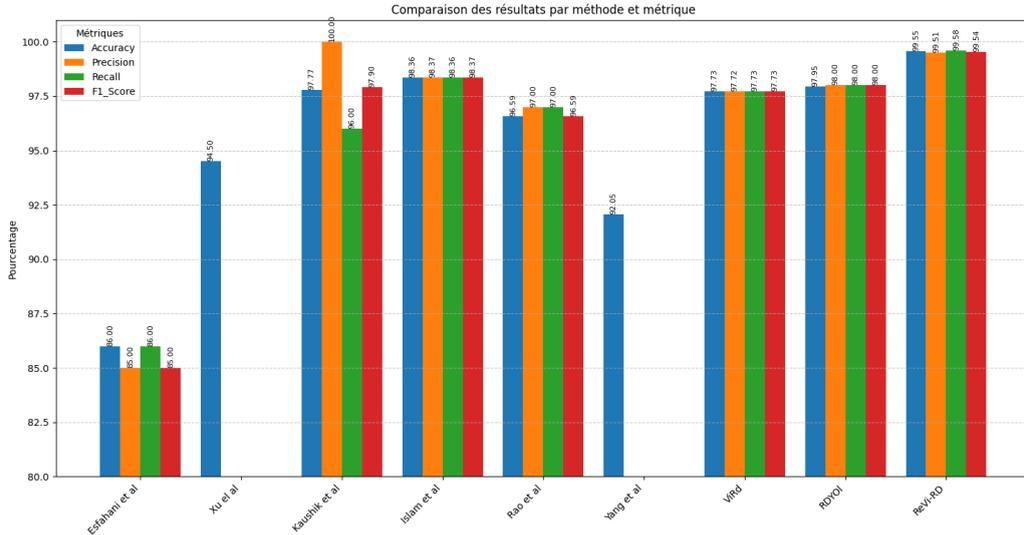


Figure 3.16: Comparaison des métriques des différents travaux.

II.4 Classification de la RD en 3 classes

Dans cette section, nous aborderons la classification de la rétinopathie diabétique en trois classes : absence de rétinopathie, rétinopathie non proliférante et rétinopathie proliférante. Nous présenterons l'architecture des systèmes développés pour cette tâche ainsi que les résultats obtenus.

Puisque l'objectif principal de cette section est de classer la rétinopathie diabétique en trois catégories distinctes, nous avons procédé à une augmentation du dataset à cause du déséquilibre des classes de ce dernier et regroupé les classes précédemment définies. La classe 0 correspond à l'absence de rétinopathie diabétique. La classe 1 regroupe les stades précoces de la rétinopathie, combinant les anciennes classes 1 (rétinopathie diabétique non proliférante légère) et 2 (rétinopathie diabétique non proliférante modérée). Enfin, la classe 2 englobe les stades avancés de la pathologie, en unifiant les anciennes classes 3 (rétinopathie diabétique non proliférante sévère) et 4 (rétinopathie diabétique proliférante). La figure 3.17 illustre ce regroupement des classes pour la nouvelle classification en trois niveaux de sévérité.

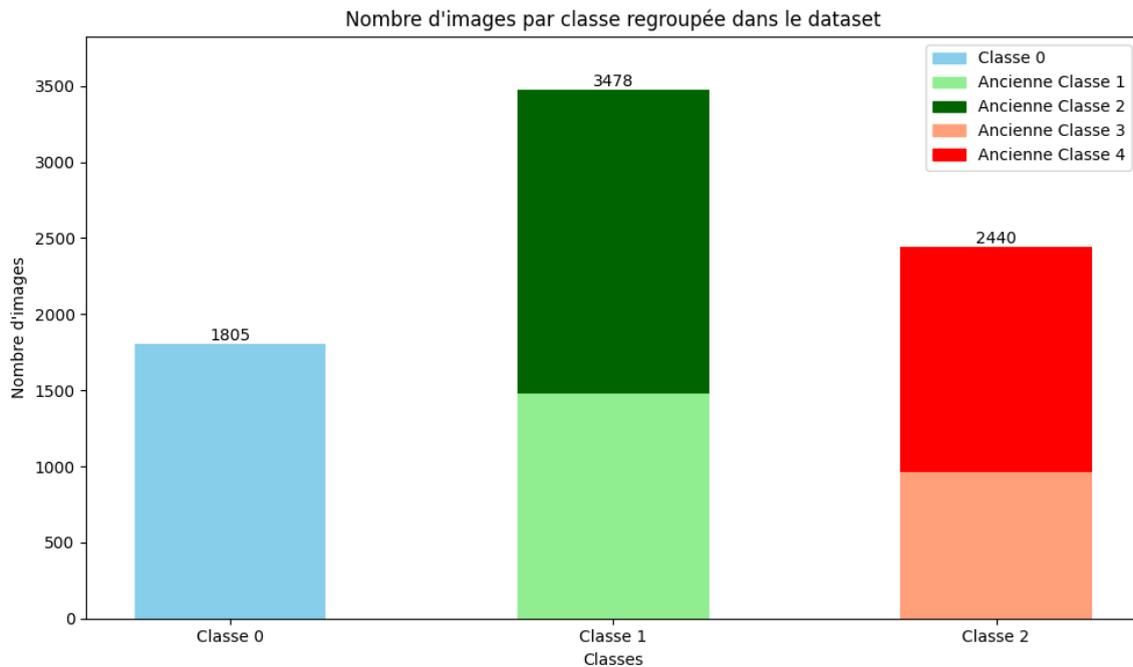


Figure 3.17: Divisions Aptos

II.4.1 Architectures proposées

L'approche proposée vise à classifier la RD en trois classes (0, 1 et 2), après un prétraitement et un fine-tuning, les modèles sont évalués afin de mettre en valeur leurs performances.

Nous proposons 3 différents modèles dont le ViR-3C, RD-Yol3C et un modèle hybride ReVi-3C pour pouvoir comparer leurs différents résultats et performances.

b.1 L'architecture ViR-3C

Nous avons proposé une architecture ViR3C basée sur les Vision Transformer (ViT) pour effectuer une classification en trois classes 0, 1 et 2. L'architecture divise l'image d'entrée en patches (petites images), les traite de manière séquentielle à l'aide d'encodeurs Transformer, et produit une représentation de l'image pour la tâche de classification en trois classes. Son architecture est illustrée à la figure 3.18.

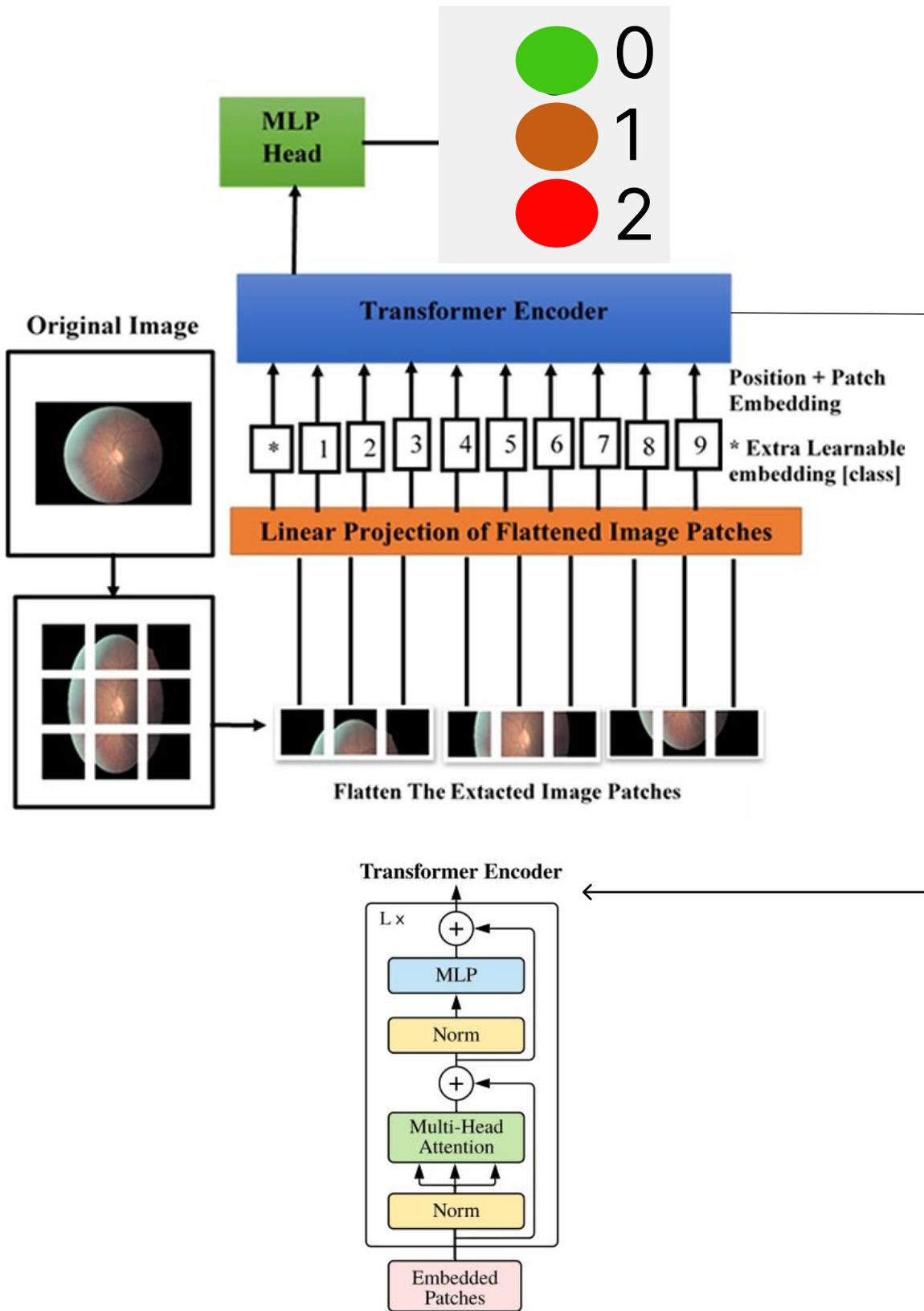


Figure 3.18: Architecture ViR-3C proposée pour la classification de la RD en 3 classes

b.1.1 Résultats

Pour évaluer les capacités de classification de la RD du modèle ViR-3C, nous avons effectué une expérimentation avec l'ensemble de données Aptos. Un ensemble de tests comprenant 733

images a été choisi pour évaluer le modèles.

Comme le montre le tableau 3.6 les résultats utilisant notre méthode sur l'ensemble de données APTOS ont atteint une précision impressionnante de 92,97% .

Table 3.6: Résultats ViR-3C

		ViR3C		
		Class 0	Class 1	Class 2
APTOS	Accuracy	0.9297		
	Precision	0.98	0.91	0.93
	Recall	0.98	0.94	0.88
	F1_score	0.98	0.92	0.90

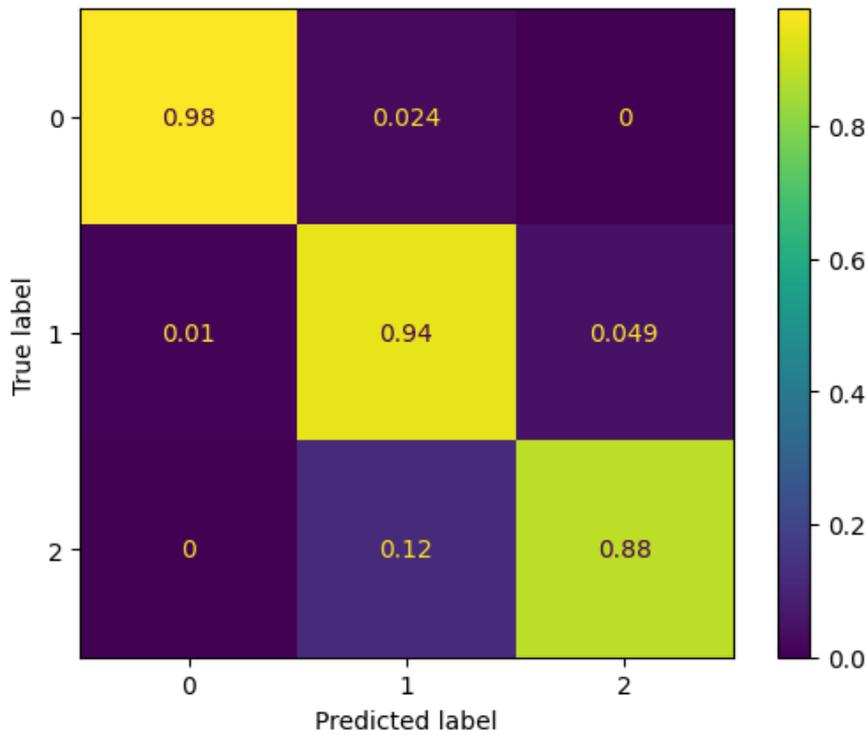
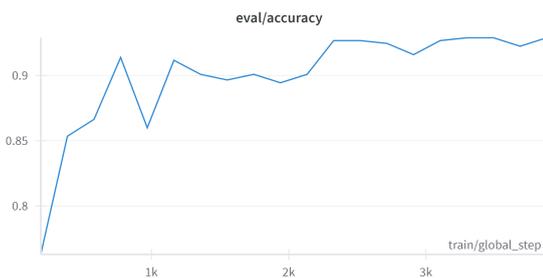
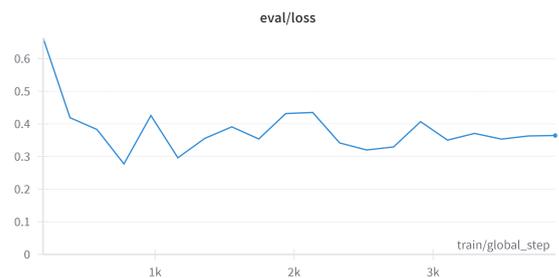


Figure 3.19: Matrice de confusion pour la ViR-3C



(a) Graphes d'exactitude



(b) Graphe de perte

Figure 3.20: Graphes d'exactitude et de perte ViR-3C

Le modèle est très précis pour détecter les patients sans rétinopathie diabétique (RD) avec une précision de 98% pour la classe 0, et une très faible marge d'erreur. Cette capacité est cruciale pour éviter des traitements inutiles pour les patients sans RD.

Le modèle identifie correctement la plupart des cas de RD modérée avec une précision de 91% pour la classe 1. Les erreurs sont réparties entre des faux négatifs (considérés comme sans RD) et des faux positifs (considérés comme RD sévère), ce qui suggère une bonne distinction entre les niveaux de gravité.

Le modèle a une forte sensibilité pour les cas de RD sévère, crucial pour une intervention rapide avec une précision de 93% pour la classe 2. Les erreurs de classification sont relativement faibles, ce qui montre une bonne précision pour les cas graves de RD.

b.2 Architecture RD-Yol3C

Pour le modèle YOLO, nous avons utilisé une architecture de réseau de neurones convolutionnels pour effectuer une tâche de détection d'objets. L'architecture YOLO divise l'image d'entrée en une grille de cellules, et pour chaque cellule, elle prédit des boîtes englobantes et les probabilités de classe pour les objets détectés dans cette cellule. Son architecture est basée sur des couches convolutionnelles, des couches de regroupement (pooling) et des couches entièrement connectées. 3.21 illustre l'architecture du modèle YOLO.

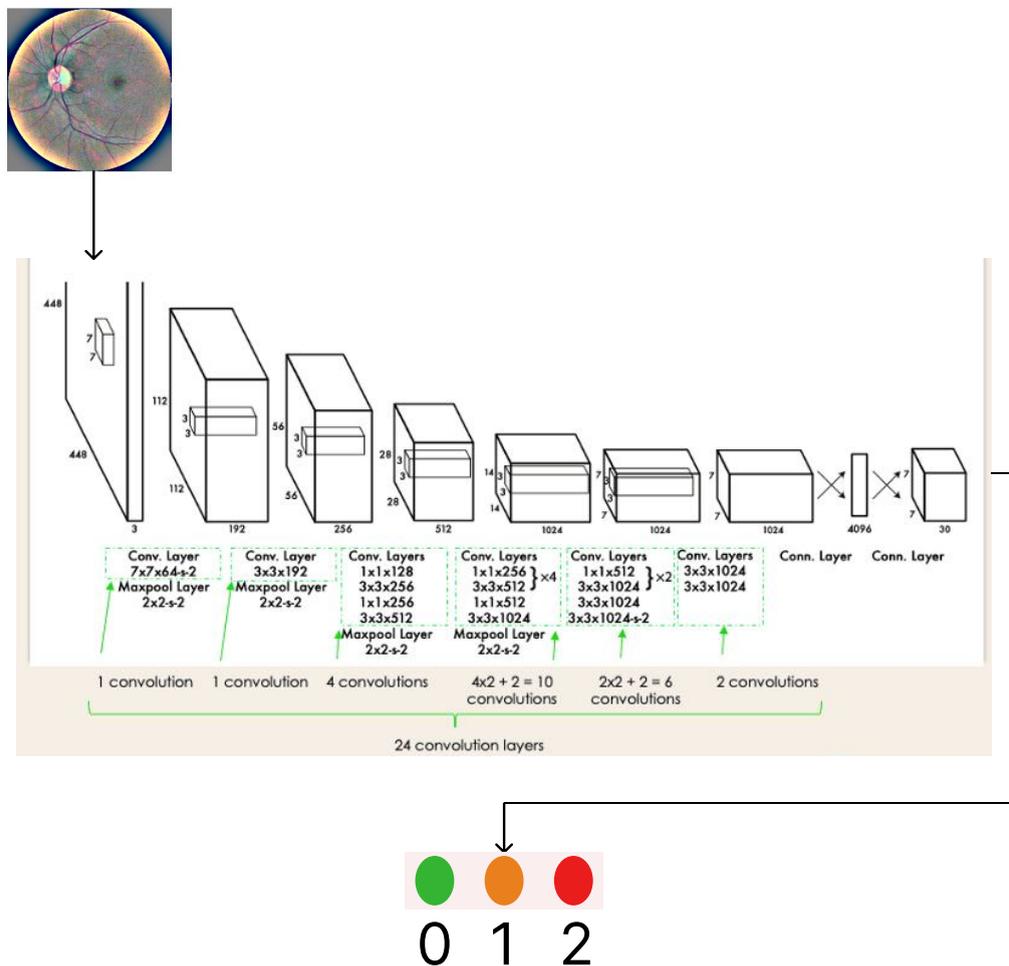


Figure 3.21: Architecture RD-Yol3C pour la détection de la RD

b.2.1 Résultats

Pour évaluer les capacités de détection de la RD du modèle Yolo, nous avons effectué une expérience avec l'ensemble de données d'Aptos. Un ensemble de tests a été choisi pour évaluer le modèles.

Comme le montre le tableau 3.7 les résultats utilisant notre méthode sur l'ensemble de données APTOS ont atteint une précision impressionnante de 91,40% en utilisant le modèle ViT.

Table 3.7: Résultats Yolo

		Yolo		
		Class 0	Class 1	Class 2
APTOS	Accuracy	0.9140		
	Precision	0.96	0.86	0.88
	Recall	0.97	0.87	0.84
	F1_score	0.96	0.86	0.86

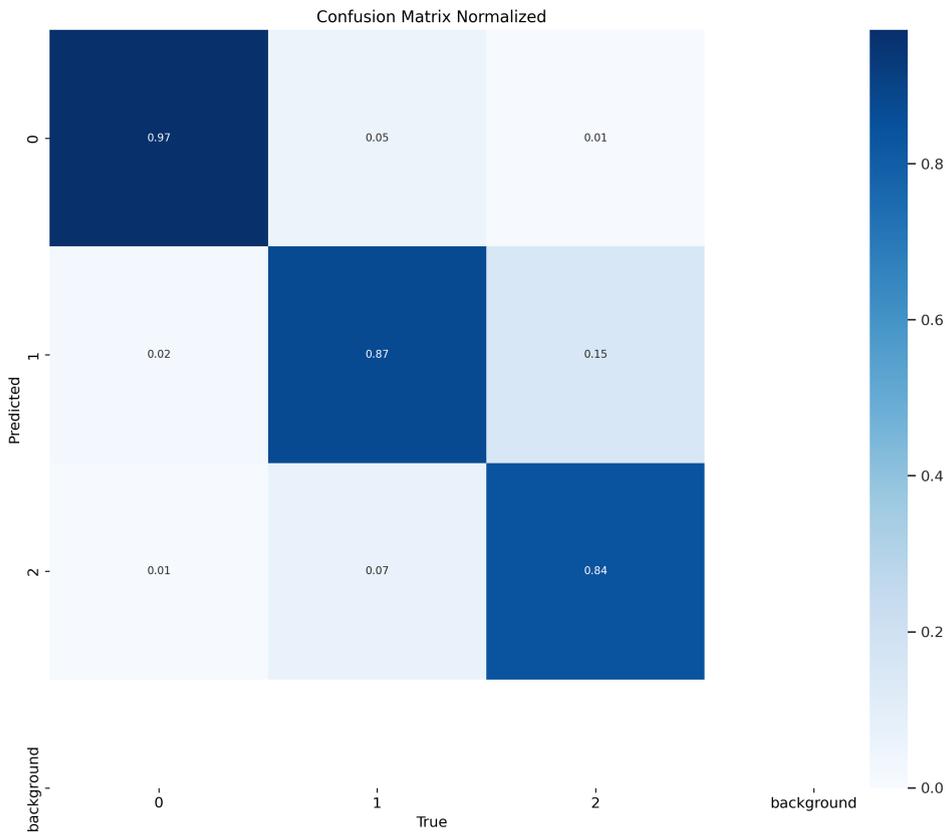


Figure 3.22: Matrice de confusion pour la RD-Yol3C

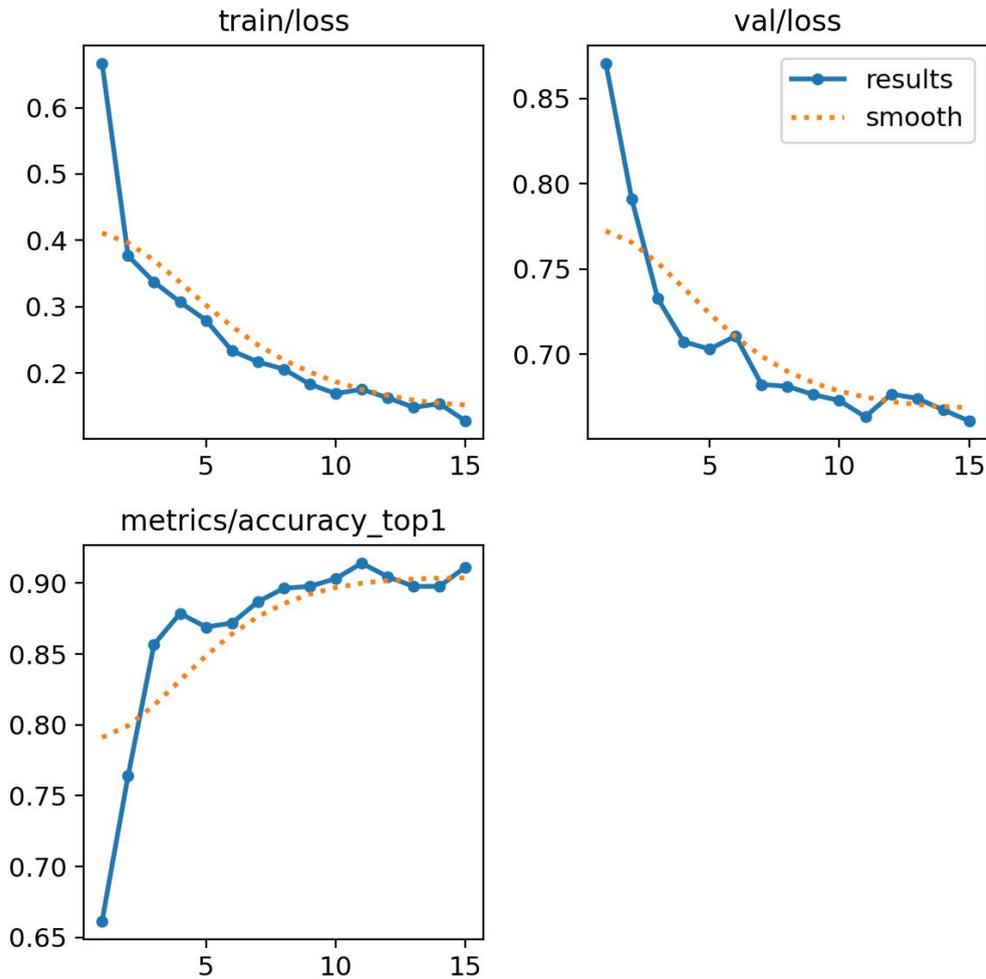


Figure 3.23: Graphe d'exactitude et de perte pour le RD-Yol3C

Le modèle présenté affiche d'excellentes performances, démontrées par les métriques élevées obtenues lors de l'évaluation. Les résultats montrent que ce modèle est non seulement précis mais aussi particulièrement efficace pour détecter les différents stades de la rétinopathie diabétique.

En examinant les métriques par classe, on observe une excellente performance pour la classe 0 (absence de rétinopathie) avec une précision de 96% et un rappel de 97%. Cela signifie que le modèle est extrêmement fiable pour identifier correctement les patients sans RD, ce qui est crucial pour éviter les faux positifs et réduire les examens inutiles.

Le modèle montre une capacité robuste à identifier les premiers signes de RD, comme indiqué par les scores élevés de précision et de rappel pour la classe 1. Un rappel de 87% pour la classe 1 signifie que le modèle est capable de détecter la majorité des cas précoces de RD, permettant ainsi une intervention médicale rapide et efficace.

En résumé, le modèle de classification présenté ici est non seulement performant dans la distinction entre les différents stades de RD mais excelle particulièrement dans la détection précoce. Avec une accuracy globale de 91.4% et des scores de précision et de rappel élevés, il constitue un outil précieux pour les professionnels de santé, facilitant une prise en charge rapide et améliorant ainsi les perspectives de traitement et de prévention des complications de la rétinopathie diabétique.

b.3 ReVi-3C

Pour procéder à cette classification nous avons choisi d'utiliser les poids du modèle PA3C en supprimant les six dernières couches (Flatten, Dense (1024), Droupout (0.3), Dense (515), Droupout (0.3), Dense (3)) utilisé l'année passée dans la thèse [1] pour les charger dans notre première partie du modèle ReVi-3C (la partie Resnet50).

Nous allons cependant modifié La sortie du réseau ViT en l'adaptant à une classification en 3 classes (0, 1, 2)

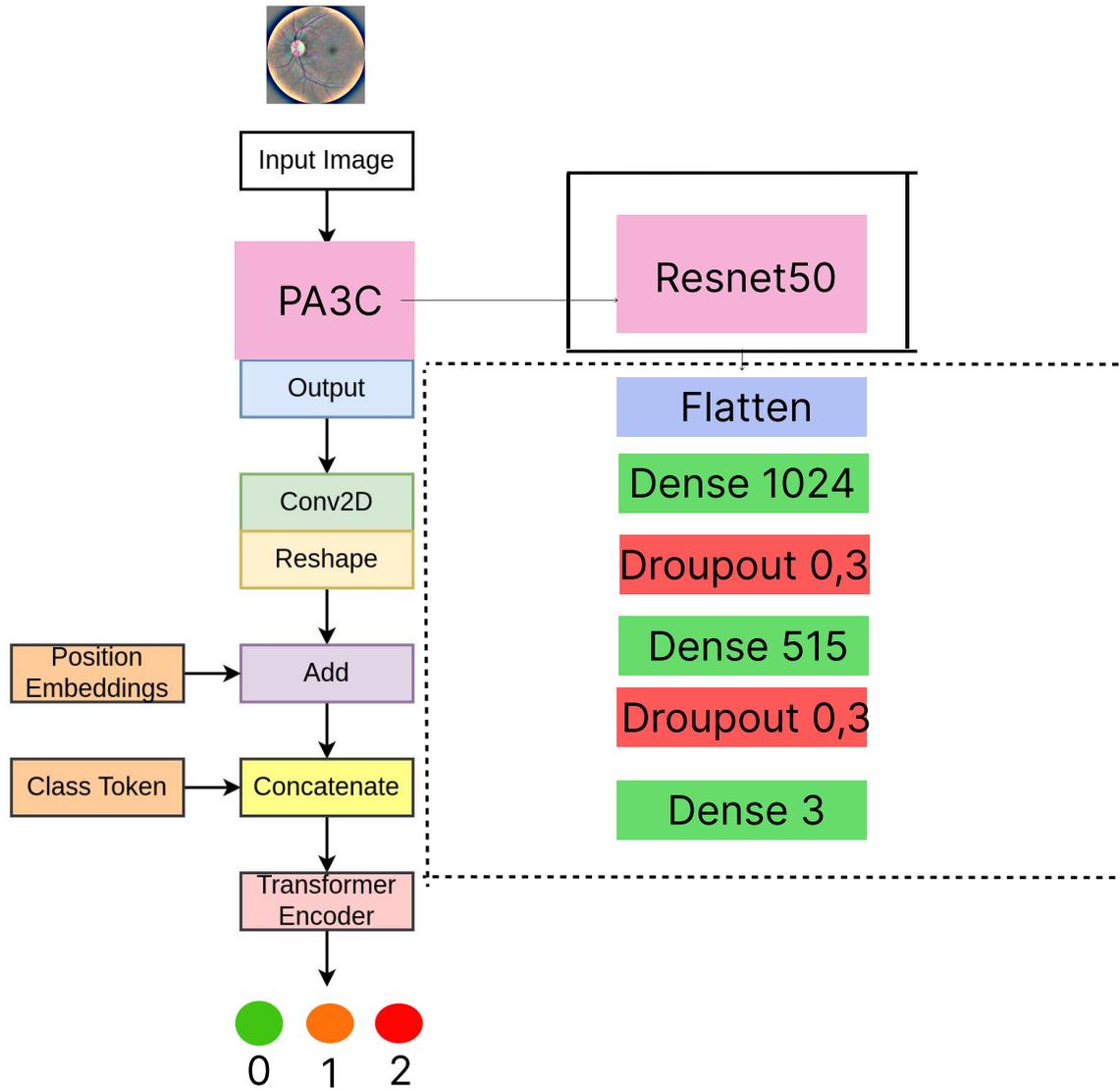


Figure 3.24: Architecture ReVi-3C pour la détection de la RD 3C

b.2.1 Résultats

Comme le montre le tableau 3.8 les résultats utilisant notre méthode sur l'ensemble de données APTOS ont atteint une précision impressionnante de 93,99% en utilisant le modèle ViT.

Table 3.8: Résultats ReVi-3C

		ReVi-3C		
		Class 0	Class 1	Class 2
APTOS	Accuracy	0.9826		
	Precision	1.00	0.98	0.98
	Recall	0.99	0.99	0.97
	F1_score	1.00	0.98	0.97

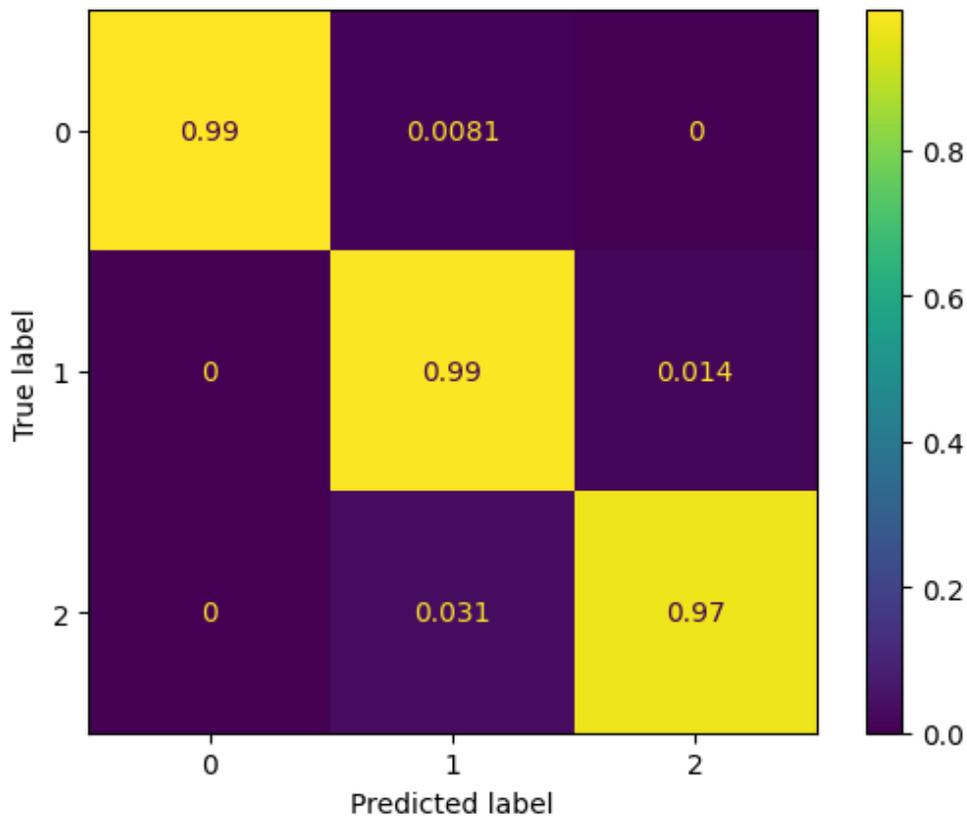
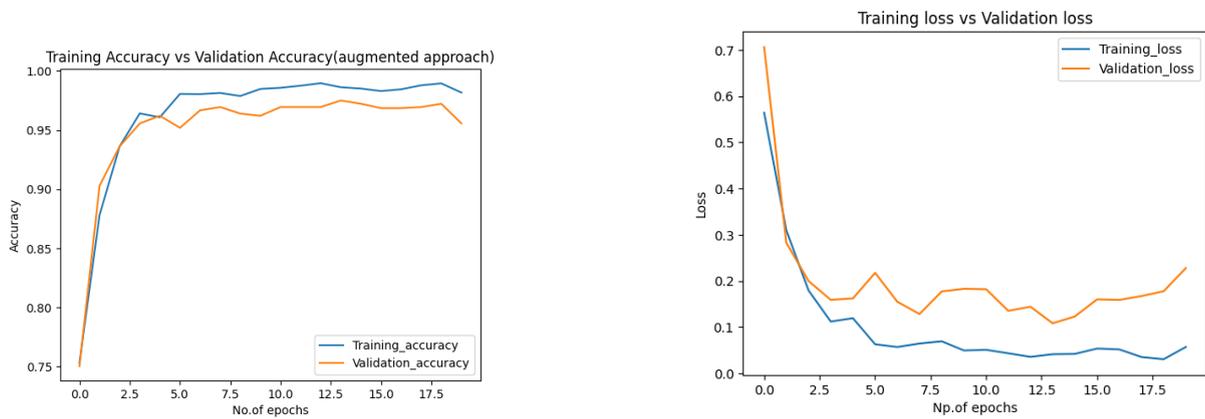


Figure 3.25: Matrice de confusion pour la ReVi-3C



(a) Graphes d'exactitude

(b) Graphe de perte

Figure 3.26: Graphes d'exactitude et de perte (ReVi-3C)

Le modèle est très précis pour détecter les patients sans rétinopathie diabétique (RD), avec une précision de 100% pour la classe 0

Le modèle parvient à identifier la majorité des cas de RD modérée avec une précision de 98% pour la classe 1.

Le modèle montre une forte sensibilité pour les cas de RD sévère avec une précision de 98% pour la classe 2, ce qui est crucial pour une intervention rapide.

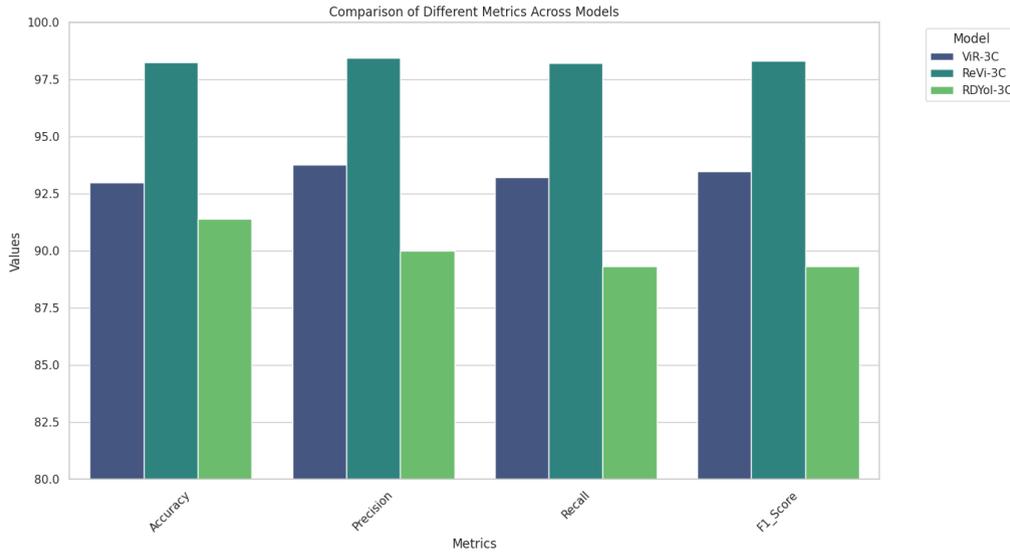


Figure 3.27: Comparaison des résultats des modèles

Table 3.9: Comparaison des résultats 3C

metrics	[65]	[39]	[44]	ViR-3C	ReVi-RD	RD-Yol3C
Accuracy	85%	88.14%	89%	92.97%	98.26%	91.40%
Precision	/	88%	/	93.77%	98.43%	90%
Recall	/	88%	/	93.22%	98.21%	89.33%
F1_Score	/	88.2%	/	93.46%	98.32%	89.33%

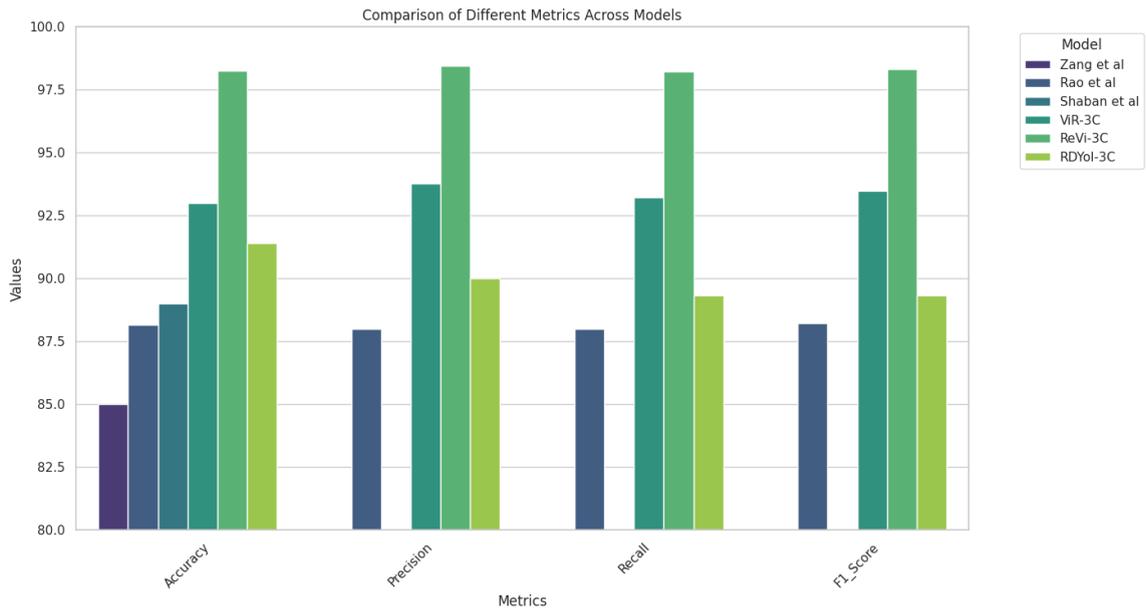


Figure 3.28: Comparaison avec les travaux antérieurs

II.5 Classification de la RD en 5 classes

Dans cette section nous abordons la classification de la RD en 5 classes, 0, 1, 2, 3 et 4

a. Dataset

Nous utilisons le dataset Aptos augmenté, la figure 3.29 illustre la distribution des classes après augmentation.

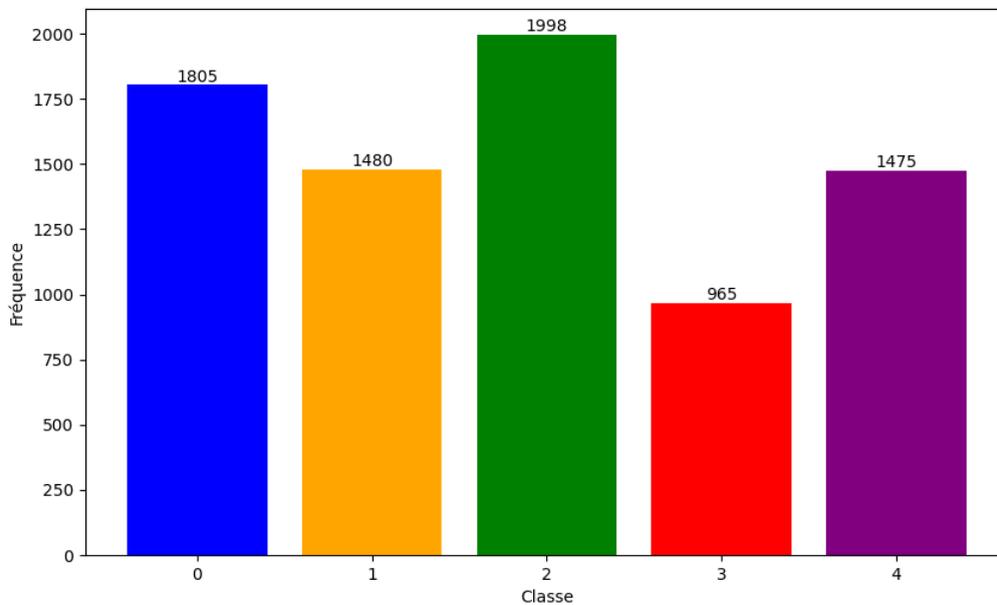


Figure 3.29: Distribution des classes Aptos 5C

Afin d'améliorer les performances des modèles, nous avons modifié les hyperparamètres du fine-tuning. Le tableau 3.10 illustre ces changements

Table 3.10: Hyperparameters du fine tuning

Paramètre	Valeur
Validation	20% Du dataset
Test	30% De validation
Taille de l'image	224x224
Taille des patches	16x16
Train Batch Size	32
Test Batch Size	64
warmup steps	500
Warmup learning rate	0.00001
Epochs	20
Learning rate	0.00002
Weight decay	0.01

b. Architectures Proposées

L'approche proposée vise à classifier la RD en 5 classes (0, 1, 2, 3 et 4) après avoir augmenté l'ensemble des données Aptos. Nous procédons au fine tuning des modèles pour évaluer leurs performances. Notre approche consiste à entraîner 3 différents modèles dont le ViT, Yolo et Resnet+ViT et ainsi comparer leurs différents résultats et performances

b.1 ViR5C

Pour la classification en 5 classes, nous allons utiliser la même architecture des ViT déjà vue dans les sections précédentes (Classification en deux et trois classes).

Nous allons cependant modifier la dernière couche MLP en l'adaptant à une classification en 5 classes.

L'architecture proposée pour cette classification est illustrée dans la figure 3.30

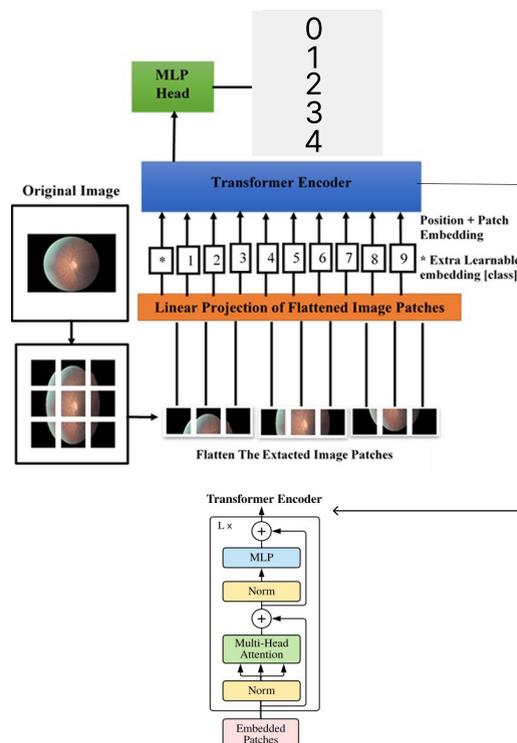


Figure 3.30: Architecture ViR5C

b.1.1 Résultats

Pour évaluer les performances du ViR5C à bien classifier la RD en 5 grades, nous avons utilisé un échantillon de test qui comporte 1545 images annotées.

La matrice de confusion 3.31 ainsi que les métriques 3.11 montrent les performances du modèle ViR5C à bien classifier la RD.

Table 3.11: Résultats ViR5C

		ViT				
		Class 0	Class 1	Class 2	Class 3	Class 4
APTOS	Accuracy	0.8733				
	Precision	0.97	0.87	0.82	0.85	0.85
	Recall	0.98	0.89	0.88	0.71	0.82
	F1_score	0.98	0.88	0.85	0.77	0.84

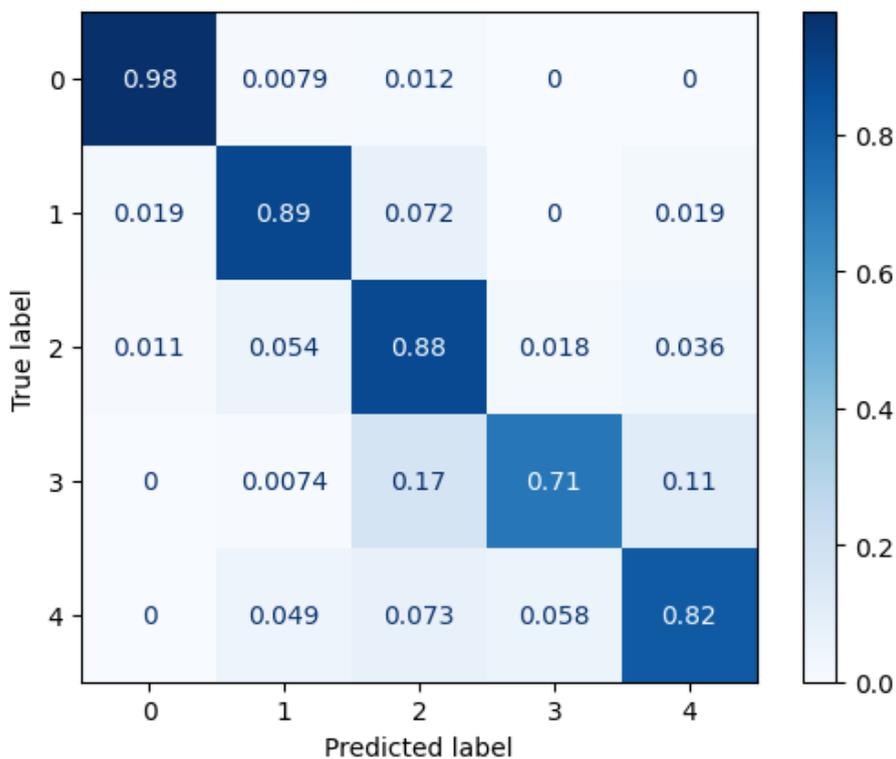
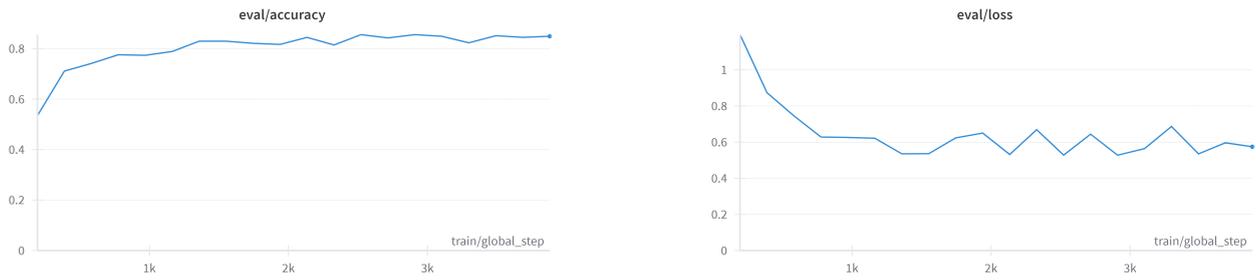


Figure 3.31: Matrice de confusion pour le ViR5C



(a) Graphes d'exactitude

(b) Graphe de perte

Figure 3.32: Graphes d'exactitude et de perte ViR5C

b.2 RD-Yol5C

Pour la classification en 5 classes, nous allons utiliser la même architecture des Yolo déjà vue dans les sections précédentes (Classification en deux et trois classes).

Nous allons cependant modifier La sortie du réseau en l'adaptant à une classification en 5 classes.

L'architecture proposée pour cette classification est illustrée dans la figure 3.33

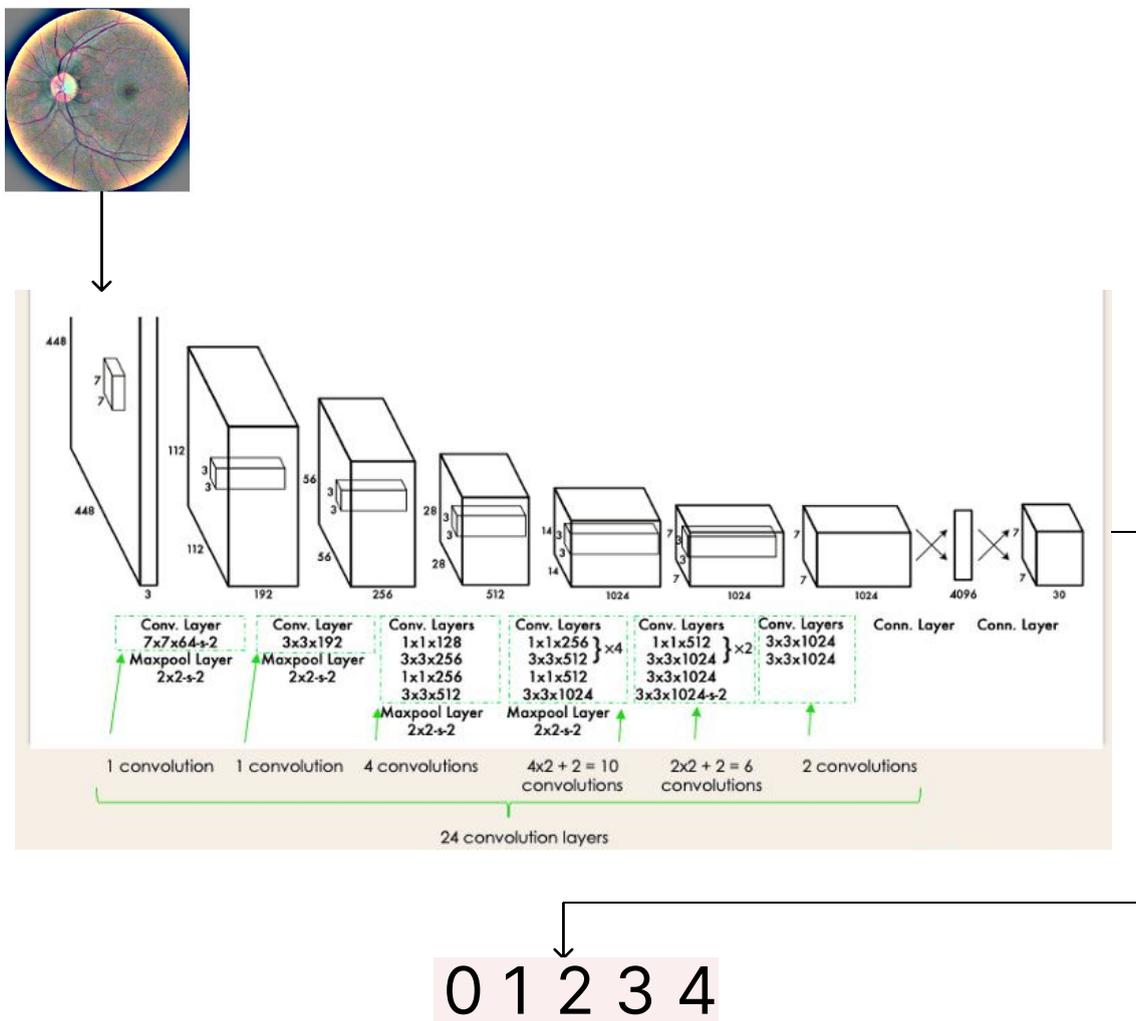


Figure 3.33: Architecture Yolo 5C (RD-Yol5C)

b.1.1 Résultats

Pour évaluer les performances du modèle Yolo à bien classifier la RD en 5 grades, nous avons utilisé un échantillon de test.

La matrice de confusion 3.34 ainsi que les métriques 3.12 montrent les performances du modèles ViT à bien classifier la RD.

Table 3.12: Résultats RD-Yol5C

		RD-Yol5C				
		Class 0	Class 1	Class 2	Class 3	Class 4
APTOS	Accuracy	0.663				
	Précision	0.97	0.65	0.52	0.46	0.64
	Recall	0.98	0.69	0.79	0.21	0.38
	F1_score	0.97	0.66	0.63	0.29	0.48

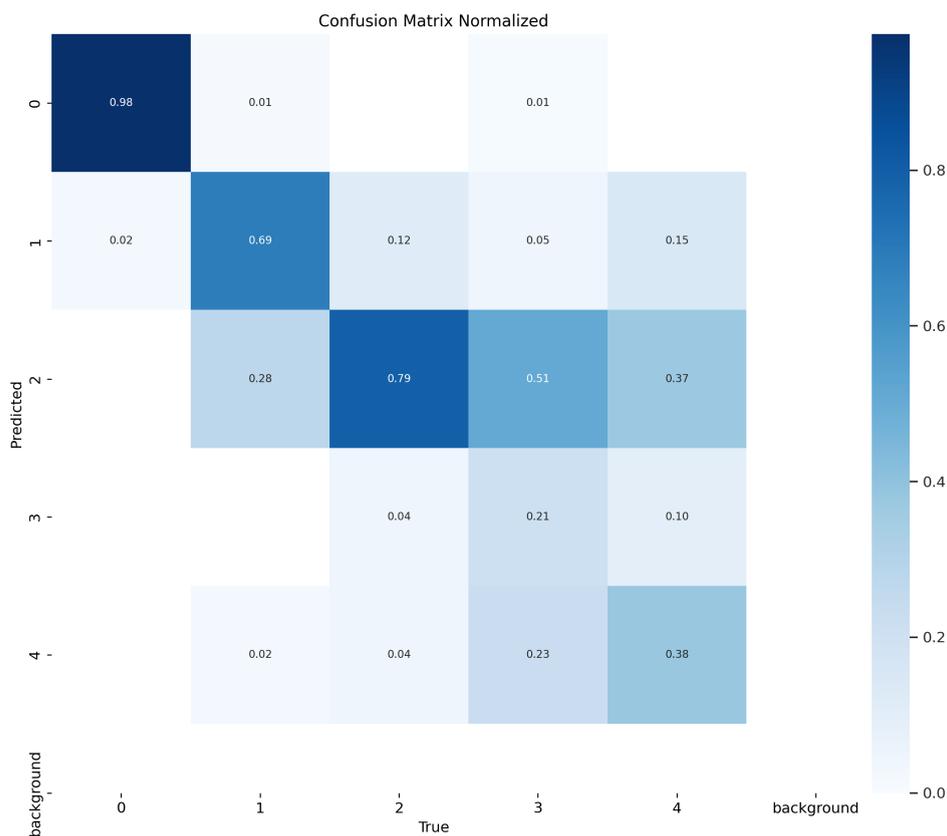


Figure 3.34: Matrice de confusion pour le RD-Yol5C

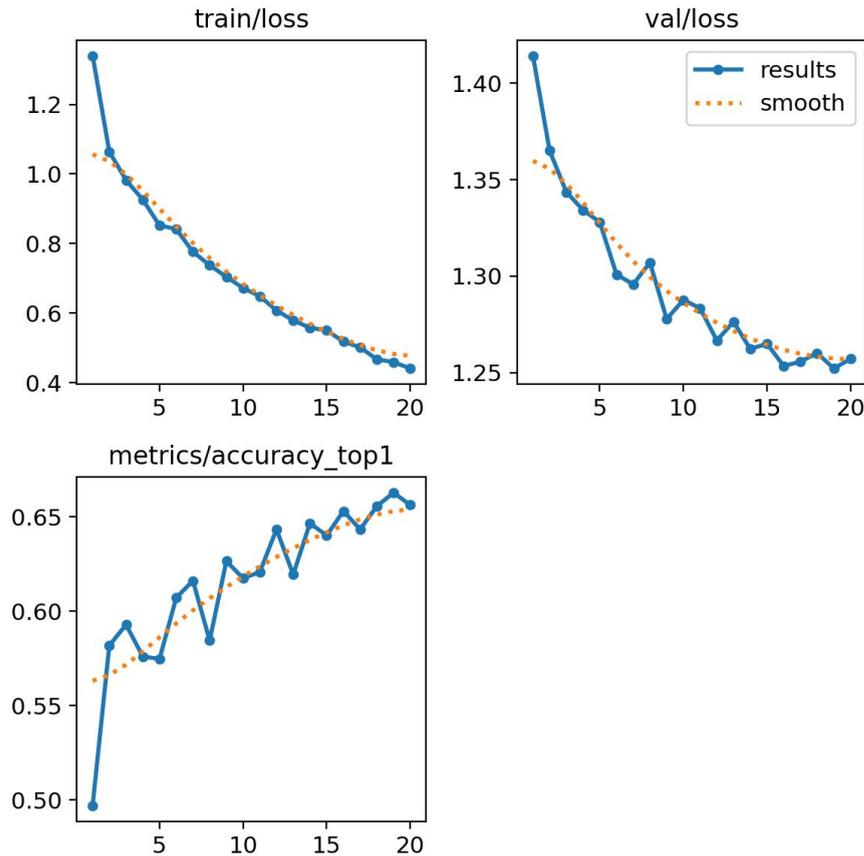


Figure 3.35: Graphe d'exactitude et de perte pour le RD-Yol5C

b.3 ReVi-5C

Pour la classification en 5 classes, nous allons utiliser l'architecture PA5C [1]

Pour procéder à cette classification, nous avons choisi d'utiliser les poids du modèle PA5C en supprimant les six dernières couches (Flatten, Dropout, Dense) utilisées l'année passée dans la thèse [1] pour les charger dans notre première partie du modèle ReVi-5C (la partie Resnet50)

Nous allons cependant modifier La sortie du réseau ViT en l'adaptant à une classification en 5 classes. (0, 1, 2, 3, 4)

L'architecture proposée pour cette classification est illustrée dans la figure 3.36

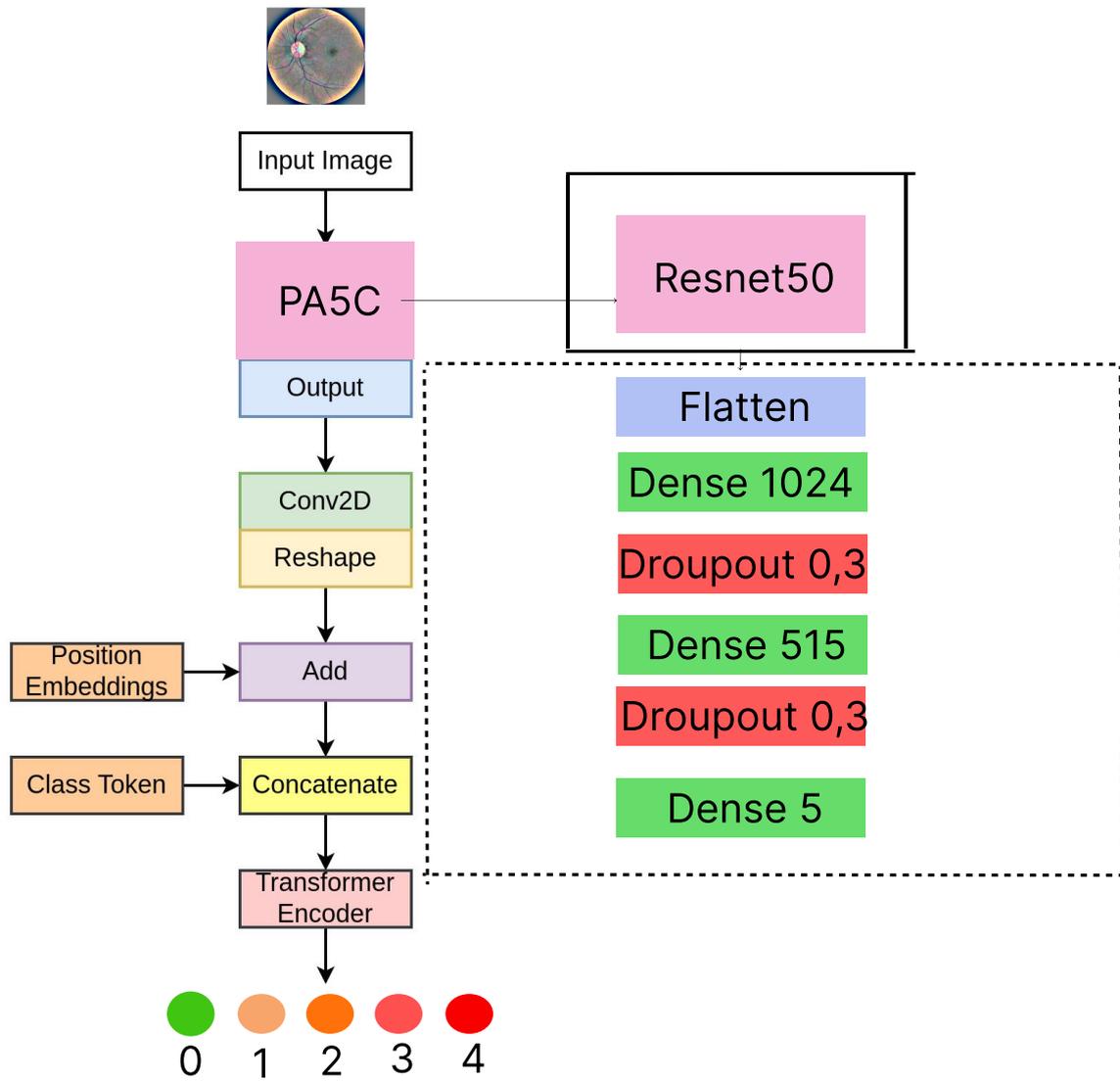


Figure 3.36: Architecture ReVi-5C

b.1.1 Résultats

Pour évaluer les performances du modèle ReVi-5C à bien classifier la RD en 5 grades, nous avons utilisé un échantillon de test qui comporte 1545 images annotées.

La matrice de confusion 3.37 ainsi que les métriques 3.13 montrent les performances du modèles ReVi-5C à bien classifier la RD.

Table 3.13: Résultats ReVi-5C

		ReVi-5C				
		Class 0	Class 1	Class 2	Class 3	Class 4
APTOS	Accuracy	0.9621				
	Precision	1.00	0.94	0.96	0.96	0.94
	Recall	1.00	0.98	0.96	0.89	0.95
	F1_score	1.00	0.96	0.96	0.92	0.95

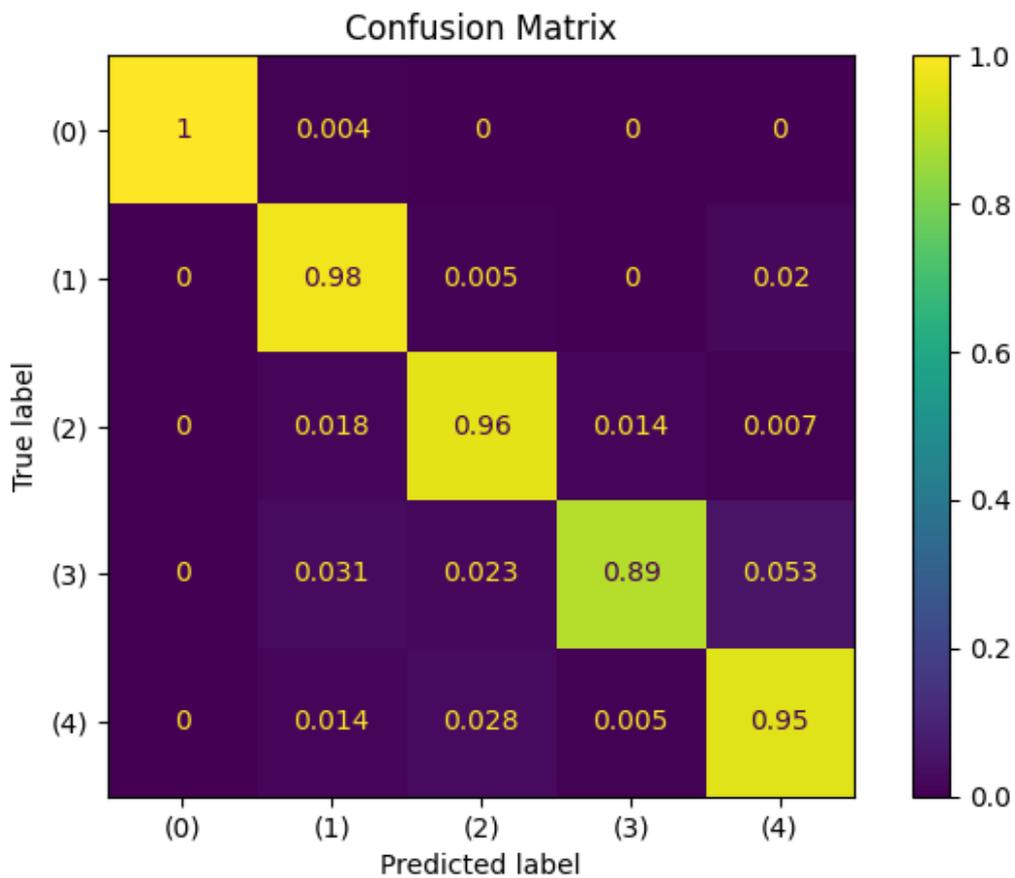
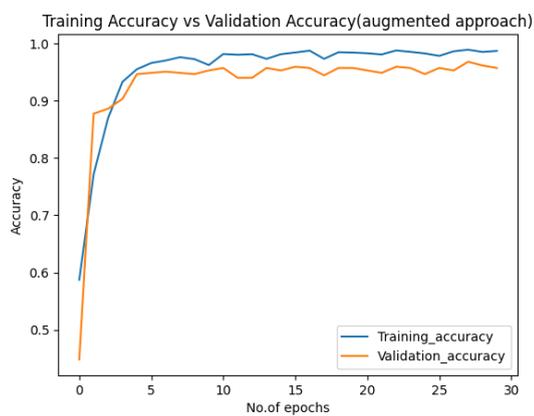
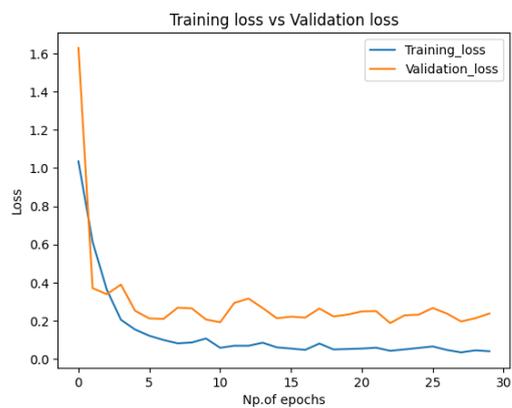


Figure 3.37: Matrice de confusion pour le ReVi-5C



(a) Graphes d'exactitude



(b) Graphe de perte

Figure 3.38: Graphes d'exactitude et de perte ReVi-5C

Table 3.14: Comparaison des résultats 5C

metrics	ViR-5C	ReVi-5C	RD-Yol5C
Accuracy	87.33%	96.21%	66.30%
Precision	87.17%	96.08%	60.60%
Recall	85.66%	95.59%	64.80%
F1_Score	86.26%	95.80%	60.60%

Le tableau 3.14 compare les performances de trois modèles : ViR-5C (Vision Transformer), ReVi-5C et RD-Yol5C, en utilisant les métriques accuracy, precision, recall et F1 score. ReVi-5C se distingue nettement des autres modèles avec une accuracy de 96.21%, une precision de 96.08%, un recall de 95.59% et un F1 score de 95.80%. Ces résultats montrent que ReVi-5C est le modèle le plus performant, surpassant ViR-5C et RD-Yol5C dans toutes les catégories évaluées. ReVi-5C démontre ainsi une grande précision, une meilleure capacité à identifier les instances positives et un excellent équilibre global entre précision et rappel.

Nous avons comparé nos résultats aux travaux antérieurs et notre modèle ReVi-5C semble être bien plus performant que ces derniers, le tableau 3.15 illustre cette comparaison.

Table 3.15: Comparaison des résultats 5C

metrics	[34]	[15]	[11]	[35]	[58]	ViR-5C	ReVi-5C	RD-Yol5C
Accuracy	82.5%	69.90%	82%	85%	91.40%	87.33%	96.21%	66.30%
Precision	82.5%	/	/	80%	92.80%	87.17%	96.08%	64.80%
Recall	82.5%	/	/	70%	92.60%	85.66%	95.59%	60.60%
F1_Score	82.5%	/	90.22%	73%	92.30%	86.26%	95.80%	60.60%

III Conclusion

En somme, ce chapitre a démontré les performances variées de différentes architectures de réseaux de neurones pour la classification de la rétinopathie diabétique (RD) en utilisant des techniques de Vision Transformer (ViT), Combinaison du Resnet50 et ViT, transfer learning etc. L'étude comparative entre les modèles ReVi-5C, RD-Yol5C et Vi-5C a mis en évidence la supériorité de ReVi-5C en termes de précision, de rappel et de score F1. Le modèle ReVi-5C a montré une constance remarquable dans ses performances sur les différentes classes, ce qui souligne sa robustesse et sa fiabilité pour cette tâche spécifique.

La préparation des données, incluant le prétraitement d'images et l'augmentation des données, a joué un rôle crucial dans l'amélioration des performances des modèles. Le dataset APTOS 2019, utilisé pour l'entraînement et le test des modèles, a fourni une base solide pour évaluer les capacités des réseaux proposés.

Les résultats obtenus suggèrent que le modèle ReVi-5C, grâce à son architecture innovante et à sa capacité d'apprentissage profond, est bien adapté à la classification de la rétinopathie diabétique. Cependant, des améliorations peuvent encore être apportées en optimisant davantage les paramètres de l'architecture et en explorant d'autres techniques de transfert learning et de fine-tuning.

Cette analyse pose les bases pour des recherches futures visant à perfectionner les modèles de classification de la RD, avec l'objectif ultime d'améliorer la détection précoce et le diagnostic de cette maladie oculaire grave.

Conclusion Générale

La prévalence croissante de la rétinopathie diabétique dans le monde souligne le besoin urgent de méthodes de diagnostic efficaces et efficaces. Les méthodes traditionnelles, qui reposent sur l'examen manuel des images rétiniennes, sont souvent longues et sujettes à l'erreur humaine.

Les modèles d'apprentissage profond, avec leur capacité à apprendre des modèles complexes et à faire des prédictions précises, ont montré des résultats prometteurs dans l'amélioration du diagnostic de la RD. L'utilisation de l'apprentissage par transfert a encore amélioré les performances de ces modèles, leur permettant de répondre aux exigences uniques du diagnostic de la RD.

Pour résoudre le problème de détection et de classification de la rétinopathie diabétique, nous avons exploité différentes architectures d'apprentissage profond basées sur le transfert learning, en particulier Vision Transformers (ViT), You Only Look Once (YOLO).

Tout d'abord, nous avons proposé une nouvelle architecture basée sur les ViTs pour la détection et la classification de la RD en 3 et 5 classes. Les résultats obtenus sont prometteurs, le cas du 3 classes n'a jamais été abordé, et il est intéressant pour le cas de détection précoce de la RD.

Ensuite, nous avons proposé une architecture basée sur modèle Yolo pour la détection et la classification de la RD en 3 et 5 classes. Les résultats obtenus méritent d'être soulignés. Le modèle Yolo a été utilisé dans la littérature pour les cas de détection d'objet et de segmentation. Nous avons proposé de l'utiliser pour la tâche de classification, ce qui a donné de bons résultats.

Finalement, nous avons développé une architecture hybride ReVi combinant le modèle PANC ($N = 2$, $N = 3$ et $N = 5$) avec le Vision Transformer pour le diagnostic de la rétinopathie diabétique (DR). Cette combinaison a été testée pour les mêmes classifications en 2, 3 et 5 classes. Les résultats obtenus sont exceptionnels, et méritent d'être pris en considération.

Après une étude comparative, notre approche hybride ReVi s'est révélée être une stratégie prometteuse, permettant de tirer parti des points forts de chaque architecture. Le PANC excelle dans l'extraction des caractéristiques locales détaillées, tandis que le Vision Transformer apporte des capacités robustes pour la capture des relations globales et contextuelles. Cette synergie offre une performance améliorée et une meilleure précision dans la détection et la classification de la rétinopathie diabétique.

Ces différentes approches, et en particulier l'approche hybride, ouvrent la voie à des perspectives intéressantes pour des applications cliniques plus efficaces et précises. Des recherches futures pourraient se concentrer sur l'optimisation de ces modèles et leur intégration dans des systèmes de diagnostic assisté par ordinateur, visant à améliorer la détection précoce et le traitement de la rétinopathie diabétique.

BIBLIOGRAPHIE

- [1] A.Akram A. Yacine. “Diabetic Retinopathy classification using transfer learning and GAN”. MA thesis. 2023.
- [2] *ANGIOGRAPHIE*. <https://www.sanotek.com/angiographie/>. Accessed: 2024-06-01.
- [3] Prawej Ansari et al. “Diabetic Retinopathy: An Overview on Mechanisms, Pathophysiology and Pharmacotherapy”. In: *Diabetology* 3 (1 2022), pp. 159–175.
- [4] Anthropic et al. *Constitutional AI: Harmonic Reinforcement Learning for Decentralized Oversight*. <https://www.anthropic.com/constitutional.pdf>. 2023.
- [5] TR Athira and Jyothisha J Nair. “Diabetic Retinopathy Grading From Color Fundus Images: An Autotuned Deep Learning Approach”. In: *Procedia Computer Science* 218 (2023), pp. 1055–1066.
- [6] Hangbo Bao et al. “Beit: Bert pre-training of image transformers”. In: *arXiv preprint arXiv:2106.08254* (2021).
- [7] Tom B Brown et al. “Language models are few-shot learners”. In: *arXiv preprint arXiv:2005.14165* (2020).
- [8] Aakash Chowdhery et al. “Palm: Scaling language modeling with pathways”. In: *arXiv preprint arXiv:2204.02311* (2022).
- [9] Dolly Das, Saroj Kumar Biswas, and Sivaji Bandyopadhyay. “Detection of diabetic retinopathy using convolutional neural networks for feature extraction and classification (DRFEC)”. In: *Multimedia Tools and Applications* 82.19 (2023), pp. 29943–30001.
- [10] Jacob Devlin et al. “BERT: Pre-training of deep bidirectional transformers for language understanding”. In: *arXiv preprint arXiv:1810.04805* (2018).
- [11] Rasha Ali Dihin, Ebtesam AlShemmary, and Waleed Al-Jawher. “Diabetic Retinopathy Classification Using Swin Transformer with Multi Wavelet”. In: *Journal of Kufa for Mathematics and Computer* 10.2 (2023), pp. 167–172.
- [12] Alexey Dosovitskiy et al. “An image is worth 16x16 words: Transformers for image recognition at scale”. In: *arXiv preprint arXiv:2010.11929* (2020).
- [13] Mehdi Torabian Esfahani, Mahsa Ghaderi, and Raheleh Kafiyeh. “Classification of diabetic and normal fundus images using new deep learning method”. In: *Leonardo Electron. J. Pract. Technol* 17.32 (2018), pp. 233–248.
- [14] *From YOLO to YOLOv8: Tracing the Evolution of Object Detection Algorithms*. <https://medium.com/nerd-for-tech/from-yolo-to-yolov8-tracing-the-evolution-of-object-detection-algorithms-eaed9a982ebd>. Accessed: 2024-05-31.
- [15] Zongyun Gu et al. “Classification of diabetic retinopathy severity in fundus images using the vision transformer and residual attention”. In: *Computational Intelligence and Neuroscience* 2023 (2023).

- [16] Mohammadreza Iman, Hamid Reza Arabnia, and Khaled Rasheed. “A Review of Deep Transfer Learning and Recent Advancements”. In: *Technologies* 11.2 (2023). ISSN: 2227-7080. DOI: [10.3390/technologies11020040](https://doi.org/10.3390/technologies11020040). URL: <https://www.mdpi.com/2227-7080/11/2/40>.
- [17] National Eye Institute. “Diabetic Retinopathy”. In: (2024).
- [18] Md Robiul Islam et al. “Applying supervised contrastive learning for the detection of diabetic retinopathy and its severity levels from fundus images”. In: *Computers in Biology and Medicine* 146 (2022), p. 105602.
- [19] Sohier Dane Karthik Maggie. *APTOS 2019 Blindness Detection*. 2019. URL: <https://kaggle.com/competitions/aptos2019-blindness-detection>.
- [20] Rojina Kashefi et al. “Explainability of Vision Transformers: A Comprehensive Review and New Perspectives”. In: *arXiv preprint arXiv:2311.06786* (2023).
- [21] Harshit Kaushik et al. “Diabetic retinopathy diagnosis from fundus images using stacked generalization of deep models”. In: *IEEE Access* 9 (2021), pp. 108276–108292.
- [22] Salman Khan et al. “Transformers in vision: A survey”. In: *ACM computing surveys (CSUR)* 54.10s (2022), pp. 1–41.
- [23] Hee E Kim et al. “Transfer learning for medical image classification: a literature review”. In: *BMC medical imaging* 22.1 (2022), p. 69.
- [24] Nitish Kundu. “Exploring ResNet50: An In-Depth Look at the Model Architecture and Code Implementation”. In: *Medium* (2023). Accessed: 2024-05-31. URL: <https://medium.com/@nitishkundu1993/exploring-resnet50-an-in-depth-look-at-the-model-architecture-and-code-implementation-d8d8fa67e46f>.
- [25] *L'examen du fond d'œil : indications, techniques et déroulement*. https://www.doctissimo.fr/html/sante/bien_voir/articles/16187-examen-fond-oeil.htm. Accessed: 2024-06-01.
- [26] *La tomographie en cohérence optique*. <https://glaucome.tech/index.php/la-tomographie-en-coherence-optique/>. Accessed: 2024-06-01.
- [27] Quoc-Vinh Lai-Dang. *A Survey of Vision Transformers in Autonomous Driving: Current Trends and Future Directions*. 2024. arXiv: [2403.07542 \[cs.CV\]](https://arxiv.org/abs/2403.07542).
- [28] Quoc-Vinh Lai-Dang. “A Survey of Vision Transformers in Autonomous Driving: Current Trends and Future Directions”. In: *arXiv preprint arXiv:2403.07542* (2024).
- [29] Guoqiang Li et al. “PMVT: a lightweight vision transformer for plant disease identification on mobile devices”. In: *Frontiers in Plant Science* 14 (2023), p. 1256773.
- [30] Yuchun Lu et al. “Application of multimodal transformer model in intelligent agricultural disease detection and question-answering systems”. In: *Plants* 13.7 (2024), p. 972.
- [31] Elizabeth A. Lundeen et al. “Prevalence of Diabetic Retinopathy in the US in 2021”. In: *JAMA Ophthalmology* 141.8 (Aug. 2023), pp. 747–754. ISSN: 2168-6165. DOI: [10.1001/jamaophthalmol.2023.2289](https://doi.org/10.1001/jamaophthalmol.2023.2289). eprint: https://jamanetwork.com/journals/jamaophthalmology/articlepdf/2806093/jamaophthalmology_lundeen_2023_oi_230031_1691537643.88835.pdf. URL: <https://doi.org/10.1001/jamaophthalmol.2023.2289>.

- [32] A. M. Mutawa, Shahad Alnajdi, and Sai Sruthi. “Transfer Learning for Diabetic Retinopathy Detection: A Study of Dataset Combination and Model Performance”. In: *Applied Sciences* 13.9 (2023). ISSN: 2076-3417. DOI: [10.3390/app13095685](https://doi.org/10.3390/app13095685). URL: <https://www.mdpi.com/2076-3417/13/9/5685>.
- [33] Huma Naz, Neelu Jyothi Ahuja, and Rahul Nijhawan. “Diabetic retinopathy detection using supervised and unsupervised deep learning: a review study”. In: *Artificial Intelligence Review* 57.5 (2024), pp. 1–66.
- [34] Waleed Nazih et al. “Vision Transformer Model for Predicting the Severity of Diabetic Retinopathy in Fundus Photography-Based Retina Images”. In: *IEEE Access* 11 (2023), pp. 117546–117561.
- [35] Mohammed Oulhadj et al. “Diabetic retinopathy prediction based on deep learning and deformable registration”. In: *Multimedia Tools and Applications* 81.20 (2022), pp. 28709–28727.
- [36] Mohammed Oulhadj et al. “Diabetic retinopathy prediction based on vision transformer and modified capsule network”. In: *Computers in Biology and Medicine* 175 (2024), p. 108523.
- [37] Srikanta Kumar Padhy et al. “Artificial intelligence in diabetic retinopathy: A natural step to the future”. In: *Indian journal of ophthalmology* 67.7 (2019), pp. 1004–1009.
- [38] Alec Radford et al. “Improving language understanding by generative pre-training”. In: (2018).
- [39] Mihir Rao, Michelle Zhu, and Tianyang Wang. “Conversion and implementation of state-of-the-art deep learning algorithms for the classification of diabetic retinopathy”. In: *arXiv preprint arXiv:2010.11692* (2020).
- [40] Joseph Redmon et al. *You Only Look Once: Unified, Real-Time Object Detection*. 2016. arXiv: [1506.02640](https://arxiv.org/abs/1506.02640) [cs.CV].
- [41] *Rétinopathie : symptômes, traitement, définition*. <https://www.docteurcliv.com/maladie/retinopathie.aspx>. Accessed: 2024-06-01.
- [42] Ganeshsree Selvachandran et al. “Developments in the detection of diabetic retinopathy: a state-of-the-art review of computer-aided diagnosis and machine learning methods”. In: *Artificial intelligence review* 56.2 (2023), pp. 915–964.
- [43] Archana Senapati et al. “Artificial intelligence for diabetic retinopathy detection: A systematic review”. In: *Informatics in Medicine Unlocked* (2024), p. 101445.
- [44] Mohamed Shaban et al. “A convolutional neural network for the screening and staging of diabetic retinopathy”. In: *Plos one* 15.6 (2020), e0233514.
- [45] Mohamed R. Shoaib et al. *Deep Learning Innovations in Diagnosing Diabetic Retinopathy: The Potential of Transfer Learning and the DiaCNN Model*. 2024. arXiv: [2401.13990](https://arxiv.org/abs/2401.13990) [eess.IV].
- [46] Stephen H Sinclair and Stanley S Schwartz. “Diabetic retinopathy—an underdiagnosed and undertreated inflammatory, neuro-vascular complication of diabetes”. In: *Frontiers in Endocrinology* 10 (2019), p. 482422.
- [47] Ayoub Skouta et al. “Deep learning for diabetic retinopathy assessments: a literature review”. In: *Multimedia Tools and Applications* 82.27 (2023), pp. 41701–41766.

- [48] Sédrick Stassin et al. “Explainability and Evaluation of Vision Transformers: An In-Depth Experimental Study”. In: *Electronics* 13.1 (2023), p. 175.
- [49] Srikanth Tammina. “Transfer learning using vgg-16 with deep convolutional neural network for classifying images”. In: *International Journal of Scientific and Research Publications (IJSRP)* 9.10 (2019), pp. 143–150.
- [50] Zhen Ling Teo et al. “Global Prevalence of Diabetic Retinopathy and Projection of Burden through 2045”. In: *Ophthalmology* 128 (2021), pp. 1580–1591.
- [51] Juan Terven, Diana-Margarita Córdova-Esparza, and Julio-Alejandro Romero-González. “A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS”. In: *Machine Learning and Knowledge Extraction* 5.4 (Nov. 2023), pp. 1680–1716. ISSN: 2504-4990. DOI: [10.3390/make5040083](https://doi.org/10.3390/make5040083). URL: <http://dx.doi.org/10.3390/make5040083>.
- [52] *The 4 Stages of Diabetic Retinopathy*. <https://www.drsnnyder.org/eye-care-services/eye-disease-management/diabetic-retinopathy/the-4-stages-of-diabetic-retinopathy/>. Accessed: 2024-06-10.
- [53] *The Evolution and Applications of YOLO Object Detection: A Comprehensive Exploration*. <https://medium.com/ubiai-nlp/the-evolution-and-applications-of-yolo-object-detection-a-comprehensive-exploration-9b8dbc0ef2a5>. Accessed: 2024-05-31.
- [54] Yohei Tomita et al. “Updates on the current treatments for diabetic retinopathy and possibility of future oral therapy”. In: *Journal of clinical medicine* 10.20 (2021), p. 4666.
- [55] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).
- [56] Ajantha Vijayakumar and Subramaniaswamy Vairavasundaram. “YOLO-based Object Detection Models: A Review and its Applications”. In: *Multimedia Tools and Applications* (2024), pp. 1–40.
- [57] Wei Wang and Amy CY Lo. “Diabetic retinopathy: pathophysiology and treatments”. In: *International journal of molecular sciences* 19.6 (2018), p. 1816.
- [58] Jianfang Wu et al. “Vision Transformer-based recognition of diabetic retinopathy grade”. In: *Medical Physics* 48.12 (2021), pp. 7850–7863.
- [59] Kele Xu, Dawei Feng, and Haibo Mi. “Deep convolutional neural network-based early automated detection of diabetic retinopathy using fundus image”. In: *Molecules* 22.12 (2017), p. 2054.
- [60] Zongyu Xu, Xuebin Xu, and Zihao Huang. “The Research of Retinopathy Image Recognition Method Based on Vit”. In: *Proceedings of the 2022 5th International Conference on Artificial Intelligence and Pattern Recognition*. AIPR '22. <conf-loc>, <city>Xiamen</city>, <country>China</country>, </conf-loc>: Association for Computing Machinery, 2023, pp. 707–710. ISBN: 9781450396899. DOI: [10.1145/3573942.3574083](https://doi.org/10.1145/3573942.3574083). URL: <https://doi.org/10.1145/3573942.3574083>.
- [61] Yaoming Yang et al. “Vision transformer with masked autoencoders for referable diabetic retinopathy classification based on large-size retina image”. In: *Plos one* 19.3 (2024), e0299265.
- [62] *YOLO Object Detection Explained*. <https://www.datacamp.com/blog/yolo-object-detection-explained>. Accessed: 2024-05-31.

- [63] *YOLO: Algorithm for Object Detection Explained [+Examples]*. <https://www.v7labs.com/blog/yolo-object-detection>. Accessed: 2024-05-31.
- [64] Jason Yosinski et al. “How transferable are features in deep neural networks?” In: *Advances in neural information processing systems* 27 (2014).
- [65] Pengxiao Zang et al. “DcardNet: diabetic retinopathy classification at multiple levels based on structural and angiographic optical coherence tomography”. In: *IEEE Transactions on Biomedical Engineering* 68.6 (2020), pp. 1859–1870.
- [66] Alebachew Ferede Zegeye, Yemataw Zewdu Temachu, and Chilot Kassa Mekonnen. “Prevalence and factors associated with Diabetes retinopathy among type 2 diabetic patients at Northwest Amhara Comprehensive Specialized Hospitals, Northwest Ethiopia 2021”. In: *BMC ophthalmology* 23.1 (2023), p. 9.

ABSTRACT

In response to the global increase in diabetes and the need for early detection of DR, the study explores the use of artificial intelligence to improve the efficiency and accuracy of diagnosis. Three main approaches were developed and evaluated: an architecture based on YOLOv8, adapted for the detection and classification of DR in 3 and 5 classes; another architecture using Vision Transformers (ViTs) for the classification of DR in 3 and 5 classes; and finally, a new hybrid architecture named ReVi for improved performance. The results obtained with these different approaches are promising, particularly with the ReVi hybrid architecture, which has shown exceptional performance. This research opens up perspectives for improving DR diagnosis, with potential applications in computer-assisted diagnostic systems in clinical settings.

Keywords: Diabetic Retinopathy, Vision Transformer, Transfer Learning, Deep Learning, Classification

RÉSUMÉ

Face à l'augmentation mondiale du diabète et à la nécessité d'un dépistage précoce de la RD, l'étude explore l'utilisation de l'intelligence artificielle pour améliorer l'efficacité et la précision du diagnostic. Trois approches principales ont été développées et évaluées en passant par une architecture basée sur YOLOv8, adaptée pour la détection et la classification de la RD en 3 et 5 classes, une autre architecture utilisant les Vision Transformers (ViTs) pour la classification de la RD en 3 et 5 classes, et enfin une nouvelle architecture hybride nommée ReVi pour une performance améliorée. Les résultats obtenus avec ces différentes approches sont prometteurs, notamment avec l'architecture hybride ReVi qui a montré des performances exceptionnelles. Cette recherche ouvre des perspectives pour l'amélioration du diagnostic de la RD, avec des applications potentielles dans les systèmes de diagnostic assisté par ordinateur en milieu clinique.

Mots clés: Rétinopathie Diabétique, Vision Transformer, Apprentissage par transfert, Apprentissage profond, Classification