République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

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

Mémoire de Master en Informatique

Options : Intelligence Artificielle et Systèmes d'Information Avancés

Thème

Proposition d'une Architecture Vision Transformers à double branche pour le diagnostic automatisé de la rétinopathie diabétique

---

Présenté par :

MADI Farid (IA)
DJERROUDI Bouzid (SIA)

Encadrant : Mme. S. Ait Kaci Azzou, Université de Béjaïa

Co-Encadrant : Mme. YAICI Malika, Université de Béjaïa

Jury :

Président :	M. BOUCHEBBAH Fatah, Université de Béjaïa.
Examinateur :	M. ACHROUFENE Achour, Université de Béjaïa.
Examinateur :	Mme. BOULAHROUZ Djamila, Université de Béjaïa.
Examinateur :	Mme. ALOUI Soraya, Université de Béjaïa.

Année universitaire 2024/2025

Remerciements

Nous tenons tout d'abord à exprimer notre profonde gratitude à Dieu Tout-Puissant pour la santé, la patience et la persévérance qui nous ont permis de mener à bien ce modeste travail.

Nous exprimons notre profonde reconnaissance à nos directrices de mémoire, Mme Samira AIT KACI AZZOU et Mme YAICI Malika, pour la qualité de leur encadrement, la pertinence de leurs orientations, la richesse de leurs conseils ainsi que leur disponibilité constante tout au long de la réalisation de ce mémoire.

Nous tenons également à remercier tous les membres du jury pour avoir accepté d'examiner et d'évaluer ce travail.

Nous adressons également nos remerciements à l'ensemble de l'équipe pédagogique du département d'informatique, pour la qualité de l'enseignement dispensé et pour les compétences qu'ils nous ont transmises au cours de notre formation.

Enfin, nous remercions nos familles et nos amis pour leur soutien indéfectible, leurs encouragements constants et leur patience tout au long de ce parcours académique.

Dédicace

Nous dédions ce travail à nos chers parents,

pour leurs encouragements précieux et leur soutien.

À nos familles,

À nos professeurs et enseignants,

Et à toutes les personnes qui, de près ou de loin,
ont contribué à la réalisation de ce travail.

Farid Madi
Bouzid Djerroudi

Table des matières

Liste des figures **6**

Liste des tableaux **7**

Liste des acronymes **8**

Introduction Générale **9**

1 Rétinopathie diabétique **11**
I Introduction . . . 11
II Définition de la rétinopathie diabétique . . . 13
II.1 Types de la rétinopathie diabétique . . . 14
II.1.1 Rétinopathie non proliférative . . . 14
II.1.2 Rétinopathie proliférative . . . 14
III Signes et Symptômes de la Rétinopathie Diabétique . . . 16
III.1 Symptômes Généraux aux Stades Avancés . . . 16
III.2 Symptômes Spécifiques selon le Type de RD . . . 17
III.2.1 Rétinopathie Non Proliférante . . . 17
III.2.2 Rétinopathie Proliférante (RDP) . . . 17
IV Causes de la rétinopathie diabétique . . . 18
V Stades de la rétinopathie diabétique . . . 19
VI Méthodes de Diagnostic de la Rétinopathie Diabétique . . . 21
VII Traitement de la rétinopathie diabétique . . . 21
VII.1 Photocoagulation au laser . . . 21
VII.2 Chirurgie de la rétine(Vitrectomie) . . . 21
VII.3 Thérapies médicamenteuses . . . 22
VII.4 Thérapie cellulaire . . . 22
VIII Prévention de la Rétinopathie Diabétique . . . 22
VIII.1 Contrôle Métabolique Rigoureux . . . 22
VIII.2 Surveillance Ophtalmologique . . . 22
IX Complications de la rétinopathie diabétique . . . 23
X Conclusion . . . 23

2 Vision Transformers & Etat de l'art **25**
I Introduction . . . 25
II Introduction aux Transformers . . . 25
II.1 Architecture des Transformers . . . 26
II.2 Du NLP à la vision . . . 27

II.2.1 Principes de base des Vision Transformers (ViT) . . . 27
II.3 Architecture des Vision Transformers . . . 27
II.3.1 Mécanisme d'attention dans les Vision Transformers . 28
II.4 Applications des ViT dans la vision par ordinateur . . . 29
III Bases de Données et Métriques d'Évaluation des Performances . . . 30
III.1 Les jeux de données pour la détection de la rétinopathie diabétique 30
III.2 Métriques d'évaluation des performances . . . 31
III.3 Revue de littérature . . . 34
IV Conclusion . . . 38

3 Conception & Réalisation 41
I Introduction . . . 41
II Méthodologie . . . 41
II.1 Préparation des données . . . 42
II.1.1 Dataset utilisé . . . 42
II.1.2 Prétraitement des données . . . 43
II.1.3 Augmentation des données . . . 45
II.2 Classification de la RD en cinq classes . . . 46
II.2.1 Approche ViT avec une seule branche . . . 46
II.2.2 Approche ViT avec deux branches . . . 55
II.3 Comparaison des modèles . . . 60
II.4 comparaison avec les autre modéles . . . 60
III Conclusion . . . 62

Conclusion Générale 63

Bibliographie 64

Liste des figures

1.1 Carte mondiale montrant la prévalence et le nombre de cas de rétinopathie diabétique (RD) par régions du monde selon la Fédération Internationale du Diabète en 2020. AFR = Afrique ; EUR = Europe ; MENA = Moyen-Orient et Afrique du Nord ; NAC = Amérique du Nord et Caraïbes ; SACA = Amérique du Sud et centrale ; SEA = Asie du Sud-Est ; WP = Pacifique occidental[20]. 12
1.2 Images de rétine normale et de rétinopathie diabétique[22]. 14
1.3 Images de rétine normale et de rétinopathie diabétique[21]. 15
1.4 a Vision normale b Vision floue due à la rétinopathie diabétique.[21]. . 17
1.5 Représentation comparée d'images oculaires normales et atteintes de rétinopathie diabétique, illustrant les vaisseaux sanguins(blood vessels)[21] 19
1.6 Stades de la rétinopathie diabétique, a image rétinienne normale, b image rétinienne avec rétinopathie diabétique légère, c image rétinienne avec rétinopathie diabétique modérée, d image rétinienne avec rétinopathie diabétique proliférante[21]. 20

2.1 Structure de l'encodeur et du décodeur [29] 26
2.2 Division en patchs de taille fixe 28
2.3 Architecture Vision Transformers [27] 28
2.4 Matrice de confusion pour une classification à trois classes. Les éléments diagonaux représentent les prédictions correctes, tandis que les éléments hors diagonale indiquent les erreurs de classification. L'intensité de la couleur reflète le nombre d'échantillons. Ce format permet d'identifier les points forts (ex. : la classe A présente peu d'erreurs) et les faiblesses du modèle (ex. : la classe C est souvent confondue avec la classe B)..[41] 34

3.1 Pipeline du processus de classification de la rétinopathie diabétique . . 42
3.2 Distribution des classes du dataset APTOS 2019. 43
3.3 gaussian blur et circle crop 44
3.4 normalization 44
3.5 augmenatations 45
3.6 VIT 47
3.7 Perte & Accuracy – ViT-SS 49
3.8 F1-score & sensibilité du ViT-SS 50
3.9 Matrice de confusion – ViT-SS 51
3.10 Perte & Accuracy – ViT-LS 52
3.11 F1-score & Sensibilité – ViT-LS 52

3.12 Matrice de confusion – ViT-LS 54
3.13 Architecture de VIT-DR 56
3.14 Perte & Accuracy – ViT-DR 57
3.15 F1-score & Sensibilité – ViT-DR 58
3.16 Matrice de confusion – ViT-DR 59

Liste des tableaux

1.1 Projections du nombre de cas (en millions) de rétinopathie diabétique (DR),et rétinopathie diabétique menaçant la vision (VTDR) pour 2030 et 2045, par région[20]. 13

2.1 Comparaison de méthodes utilisant les Vision Transformers pour la classification de la rétinopathie diabétique. 39

3.1 Hyperparamètres des modèles ViT-SS et ViT-LS 47
3.2 Hyperparamètres d'entraînement des modèles ViT-SS, ViT-LS et ViT-DR 48
3.3 Performances globales du modèle VIT-SS 49
3.4 Les performances par classe du model VIT-SS 50
3.5 Performances globales du modèle VIT-LS 52
3.6 Les performances par classe du model VIT-LS 53
3.7 Comparaison des performances globales des modèles ViT-SS et ViT-LS 55
3.8 Performances globales du modèle ViT-DR 57
3.9 Performance par classe du modèle VIT-DR 58
3.10 Comparaison des performances globales des trois modèles 60
3.11 Comparaison des performances des différents modèles 61

Liste des acronymes

— **ViT** : Vision Transformer
— **CNN** : Convolutional Neural Network
— **IA** : Intelligence Artificielle
— **ML** : Machine Learning
— **DL** : Deep Learning
— **RD** : Rétinopathie Diabétique
— **ViT-SS** : Vision Transformer – Small Scale
— **ViT-LS** : Vision Transformer – Large Scale
— **ViT-DR** : Vision Transformer – Dual Resolution

Introduction Générale

Les atteintes oculaires dues au diabète, en particulier la rétinopathie diabétique, constituent l'une des principales causes de déficience visuelle dans le monde. Cette pathologie représente un défi majeur pour la santé publique, car elle évolue de manière silencieuse et peut entraîner une perte de vision irréversible si elle n'est pas détectée à temps. Un dépistage précoce et un suivi régulier sont donc essentiels pour limiter les complications et améliorer la qualité de vie des patients [1].

Cependant, le diagnostic de la rétinopathie diabétique demeure complexe. Il nécessite une expertise médicale approfondie et repose sur l'interprétation d'images de fond d'œil, une tâche qui peut être rendue difficile par la diversité des lésions observées et par une certaine subjectivité dans leur évaluation. Cette complexité accentue la nécessité de recourir à des outils technologiques capables d'assister les professionnels de santé et de fiabiliser le processus de détection.

Dans ce contexte, les avancées en intelligence artificielle (IA) se présentent comme une alternative prometteuse [49]. Les réseaux de neurones convolutifs (CNNs), qui constituent la référence en analyse d'images médicales, ont montré leur efficacité dans la détection et la classification de la rétinopathie diabétique [48]. Toutefois, ces approches présentent plusieurs limites : elles nécessitent généralement de vastes ensembles de données annotées pour éviter le surapprentissage, souffrent d'une dépendance à la qualité des images de fond d'œil, et restent coûteuses en ressources computationnelles. De plus, leur capacité à capturer des relations globales à longue portée demeure limitée, car elles s'appuient principalement sur des filtres locaux.

Pour surmonter certaines de ces contraintes, le transfert d'apprentissage constitue une approche efficace [50]. Il permet de réutiliser des modèles préalablement entraînés sur de grands ensembles d'images génériques et de les adapter à la tâche spécifique du dépistage de la rétinopathie diabétique. Cette stratégie réduit significativement le besoin en grandes bases de données annotées, accélère la convergence de l'apprentissage et améliore la performance des modèles même en présence de données médicales limitées.

En complément, les Vision Transformers (ViTs) [27] se présentent comme une solution prometteuse. Grâce à leurs mécanismes d'attention, ils offrent la possibilité de modéliser des relations contextuelles à grande échelle tout en captant des détails subtils de l'image. Contrairement aux CNNs, qui tendent à perdre une partie de l'information spatiale à travers les opérations de convolution et de pooling, les ViTs préservent explicitement la structure spatiale des images grâce à l'utilisation des position embeddings. Cette capacité d'intégrer une vision globale tout en maintenant la cohérence spatiale ouvre de nouvelles perspectives pour améliorer la précision, la robustesse et la généralisabilité du diagnostic automatisé de la rétinopathie diabétique.

Dans ce travail, nous proposons le développement de trois modèles basés sur les

Vision Transformers. Les deux premiers different selon l'échantillonage de l'image en 16 ou 32 patchs, nommés : ViT-SS, construit à partir de l'architecture ViT16 et ViT-LS, dérivé de l'architecture ViT32. Le modèle hybride est la combinaison du ViT-SS et du ViT-LS. Il exploite simultanément et en parallèle les deux architecture ViT16 et ViT32 afin de capter des informations visuelles complémentaires.

Notre objectif principal est de concevoir une approche performante pour le diagnostic automatique de la rétinopathie diabétique à partir d'images de fond d'œil, en évaluant la contribution individuelle des architectures ViT et l'apport d'une combinaison hybride.

La suite de ce mémoire est organisée comme suit : le Chapitre 1 présente un aperçu de la rétinopathie diabétique et de ses principales complications. Le Chapitre 2 introduit les concepts fondamentaux des Vision Transformers et dresse un état de l'art des méthodes d'apprentissage profond appliquées à cette problématique. Le Chapitre 3 expose nos modèles proposés, leur conception, ainsi que les résultats expérimentaux obtenus. Enfin, une conclusion générale viendra clore ce mémoire en ouvrant des perspectives de recherche futures.

Chapitre 1

Rétinopathie diabétique

I Introduction

La rétinopathie diabétique (RD) constitue l'une des principales complications du diabète sucré et représente une menace majeure pour la santé visuelle à l'échelle mondiale. Selon les dernières estimations et projections mondiales [20], la RD touche des millions de personnes, avec des disparités marquées selon les régions géographiques, comme illustré dans la figure 1.1.

Les données de prévalence indiquent que certaines régions comme l'Afrique (AFR, 35,90 %), l'Amérique du Nord et les Caraïbes (NAC, 33,30 %) ou encore le Moyen-Orient et l'Afrique du Nord (MENA, 32,90 %) sont particulièrement touchées. Ces taux élevés de prévalence traduisent une proportion importante de personnes diabétiques souffrant de complications rétiniennes, souvent en raison de facteurs tels qu'un dépistage insuffisant, un accès limité aux soins spécialisés, un manque de sensibilisation des patients, ou encore des contraintes socio-économiques et organisationnelles des systèmes de santé[4].

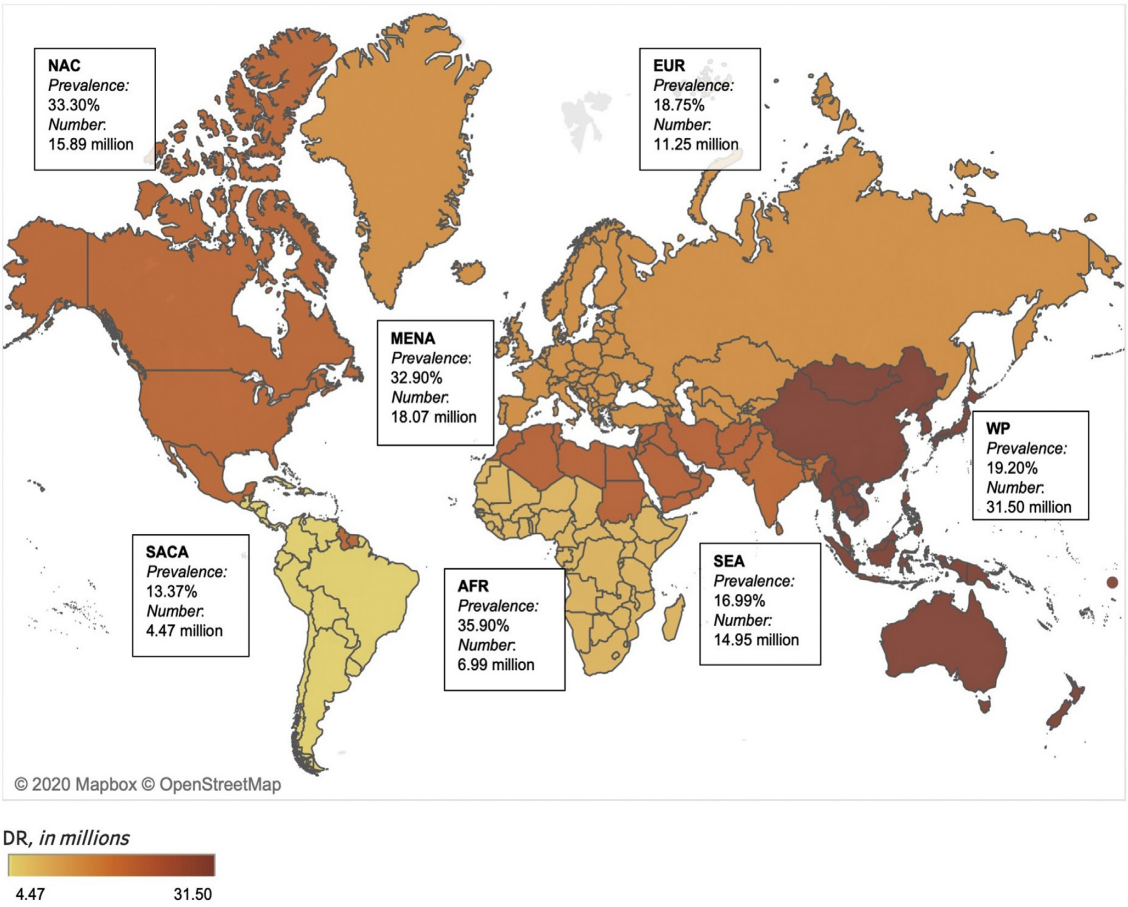


FIGURE 1.1 – Carte mondiale montrant la prévalence et le nombre de cas de rétinopathie diabétique (RD) par régions du monde selon la Fédération Internationale du Diabète en 2020. AFR = Afrique ; EUR = Europe ; MENA = Moyen-Orient et Afrique du Nord ; NAC = Amérique du Nord et Caraïbes ; SACA = Amérique du Sud et centrale ; SEA = Asie du Sud-Est ; WP = Pacifique occidental[20].

Les chiffres absolus sont tout aussi alarmants : en 2020, la région du Pacifique occidental (WP) comptait à elle seule plus de 31 millions de cas, suivie du Moyen-Orient et Afrique du Nord (MENA, 18 millions) et de l'Amérique du Nord et des Caraïbes (15,89 millions). Ces données suggèrent une charge de morbidité importante, particulièrement concentrée dans les régions à forte population diabétique.

Les projections mondiales renforcent cette inquiétude. D'après les estimations pour 2030 et 2045(voire le tableau 1.1) :

— Le nombre de personnes atteintes de RD devrait passer de **129,84 millions en 2030** à **160,50 millions en 2045**.

— Le nombre de cas menaçant directement la vision (VTDR) pourrait dépasser **44 millions d'ici 2045**.

— Des régions comme l'Asie du Sud-Est et le Pacifique occidental sont en voie de devenir des foyers épidémiques majeurs, avec des hausses anticipées importantes.

Ces dynamiques soulignent l'urgence de stratégies de dépistage précoce, de classification selon la sévérité de la maladie, et d'interventions thérapeutiques adaptées. D'autant plus que la RD affecte principalement la population active (20–64 ans), contribuant

ainsi non seulement à la perte de la vision, mais aussi à une baisse de la productivité économique[3].

TABLE 1.1 – Projections du nombre de cas (en millions) de rétinopathie diabétique (DR),et rétinopathie diabétique menaçant la vision (VTDR) pour 2030 et 2045, par région[20].

Région	**DR**		**VTDR**	
	2030	2045	2030	2045
Asie du Sud-Est (SEA)	19.62 (16.18–23.37)	26.06 (21.64–31.08)	4.13 (2.79–5.82)	5.48 (3.76–7.77)
Afrique	10.29 (8.36–12.28)	16.93 (13.91–20.19)	4.16 (2.86–5.75)	6.84 (4.78–9.44)
Europe	12.46 (9.03–16.50)	12.89 (9.23–17.17)	3.64 (3.04–4.30)	3.76 (3.13–4.45)
Moy. Orient & Afrique du Nord (MENA)	25.05 (19.75–30.79)	35.47 (27.98–43.66)	6.36 (3.86–9.77)	9.01 (5.48–13.89)
Am. Nord & Caraïbes (NAC)	18.75 (14.15–23.83)	21.11 (15.95–26.83)	4.46 (2.98–6.36)	5.02 (3.36–7.16)
Am. Sud & Centrale (SACA)	5.69 (2.44–10.82)	6.96 (3.00–12.99)	2.37 (1.66–3.30)	2.90 (2.04–3.97)
Pacifique Occidental (WP)	37.98 (27.76–50.13)	41.08 (30.03–54.40)	10.93 (8.89–13.32)	11.81 (9.58–14.43)
Global	**129.84 (115.30–145.60)**	**160.50 (143.70–178.60)**	**36.05 (31.63–41.15)**	**44.82 (39.20–51.33)**

II Définition de la rétinopathie diabétique

La rétinopathie diabétique (RD), également appelée maladie oculaire diabétique, est une affection qui endommage la rétine de l'œil en raison du diabète sucré. La RD affecte les vaisseaux sanguins de l'œil, provoquant des fuites de liquide et des hémorragies qui altèrent la vision(voire figure 1.2). La cécité peut survenir dans les deux yeux si la maladie n'est pas traitée[5].

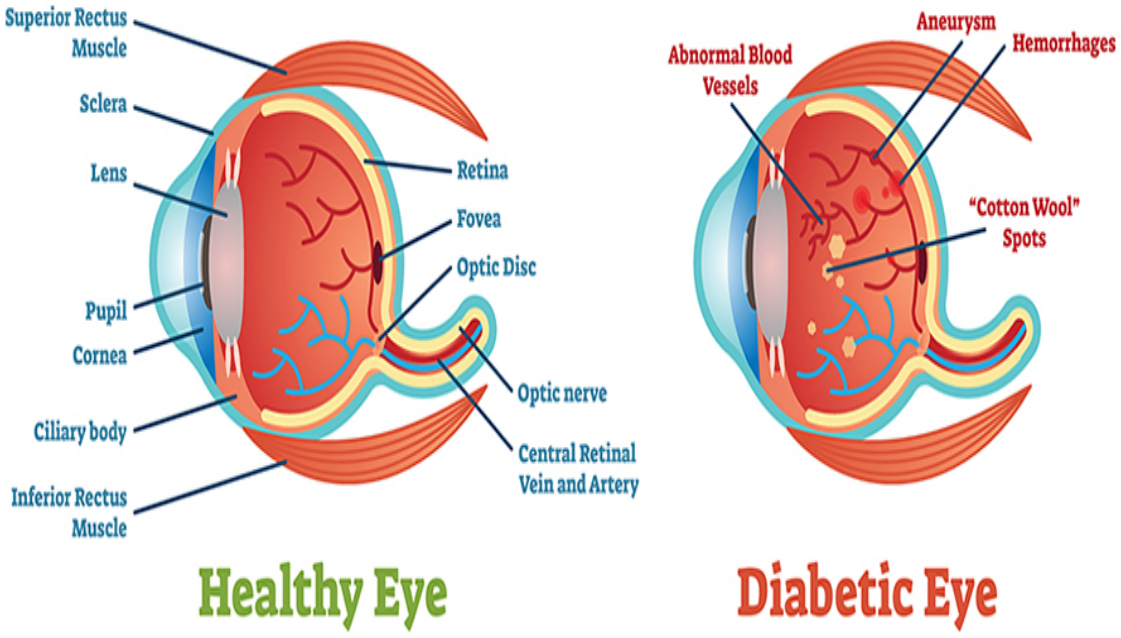


FIGURE 1.2 – Images de rétine normale et de rétinopathie diabétique[22].

II.1 Types de la rétinopathie diabétique

La rétinopathie diabétique est classée en deux types(voire la figure 1.3) : la rétinopathie non proliférante et la rétinopathie proliférante[6].

II.1.1 Rétinopathie non proliférative

On désigne par rétinopathie de fond, ou rétinopathie non proliférante, un stade préliminaire de la rétinopathie diabétique, caractérisé par l'absence de formation de nouveaux vaisseaux sanguins dans l'œil. À ce stade, les petits vaisseaux de la rétine subissent des altérations, entraînant des fuites de sang et de liquide. Ces fuites peuvent provoquer un œdème au niveau de la macula, la partie centrale de la rétine responsable de la vision fine. Cette forme de rétinopathie est généralement classée selon sa gravité en trois niveaux : légère, modérée ou sévère[6].

II.1.2 Rétinopathie proliférative

Ce stade représente la forme la plus avancée de la rétinopathie diabétique, dans laquelle la rétine subit des dommages graves, pouvant aller jusqu'à sa destruction. À ce niveau, des vaisseaux sanguins anormaux et fragiles se développent en remplacement de la rétine altérée. Ces nouveaux vaisseaux, instables, présentent un risque élevé

d'hémorragie, pouvant entraîner une perte soudaine et sévère de la vision.[6].

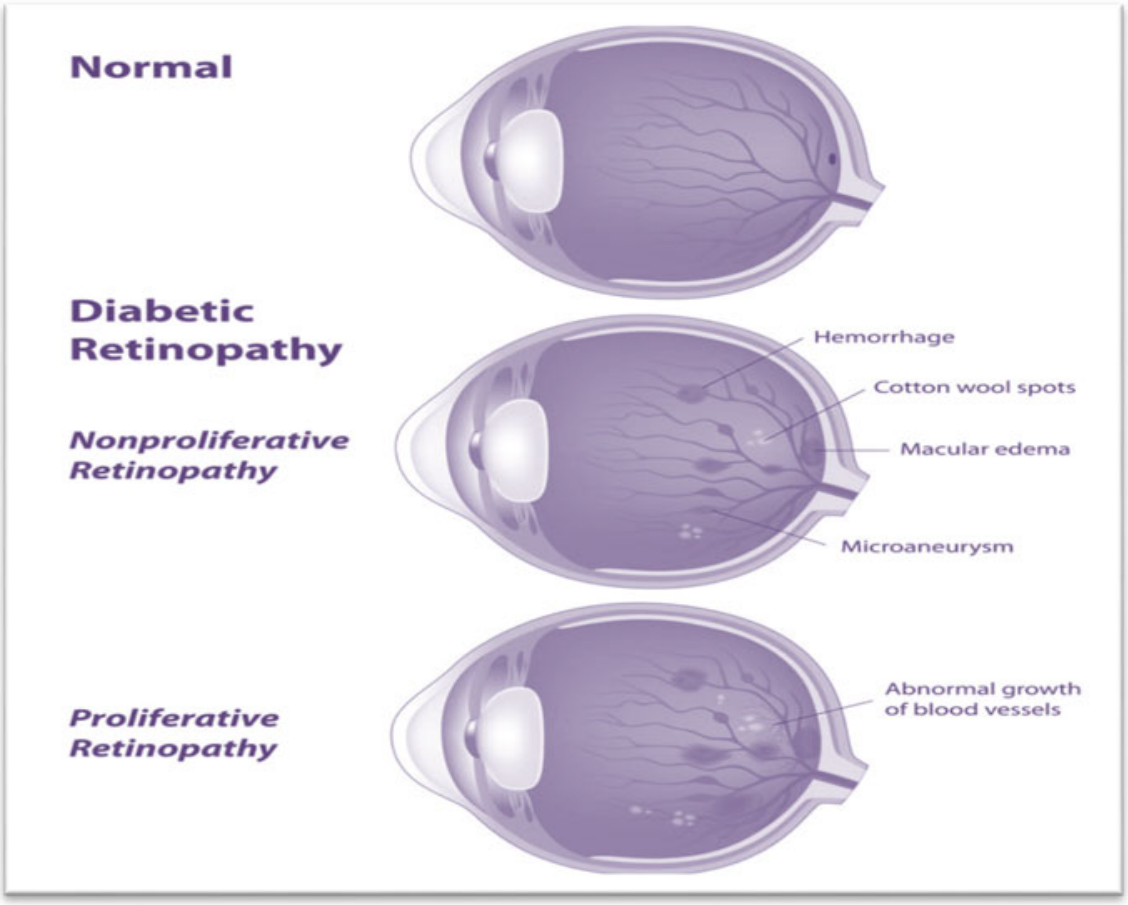


FIGURE 1.3 – Images de rétine normale et de rétinopathie diabétique[21].

III Signes et Symptômes de la Rétinopathie Diabétique

La rétinopathie diabétique, À ses débuts elle est généralement asymptomatique, ce qui rend son dépistage précoce impossible sans un examen ophtalmologique spécialisé. Les symptômes n'apparaissent souvent qu'aux stades avancés, lorsque les lésions rétiniennes deviennent sévères [2].

III.1 Symptômes Généraux aux Stades Avancés

- **Vision floue ou fluctuante(voire la figure 1.4)** : Due à des fuites de liquide (œdème maculaire) ou à des hémorragies intrarétiniennes [17].
- **Taches sombres** ("mouches volantes" ou myodésopsies) : Causées par des saignements intravitréens [17].
- **Perte soudaine de vision** : Peut survenir en cas d'hémorragie vitréenne massive ou de décollement de rétine [17].
- **Difficultés à percevoir les contrastes ou les couleurs** : Liées à l'atteinte de la macula, zone centrale de la rétine [17].
- **Vision déformée (métamorphopsies)** : Les lignes droites apparaissent courbées, signe d'un œdème maculaire [17].

FIGURE 1.4 – a Vision normale b Vision floue due à la rétinopathie diabétique.[21].

III.2 Symptômes Spécifiques selon le Type de RD

Les symptômes spécifiques peuvent se classer en deux catégories [1] :

III.2.1 Rétinopathie Non Proliférante

Ce type correspond à un stade précoce de la maladie. Dans la plupart des cas, il n'y a pas de symptôme visible pour le patient. Les signes sont détectés uniquement lors d'un examen ophtalmologique.

- **Micro-hémorragies rétiniennes** : Petites taches de sang dans la rétine, visibles seulement au fond d'œil[18].
- **Nodules cotonneux** : Zones blanches sur la rétine, qui indiquent un manque d'oxygène localisé (ischémie)[18].

III.2.2 Rétinopathie Proliférante (RDP)

Cette forme avancée de la maladie s'accompagne de symptômes visuels, car la rétine est gravement endommagée.

- **Corps flottants** ou "toiles d'araignée" : Apparition de taches mobiles dans le champ de vision, causées par des saignements issus de vaisseaux sanguins anormaux [19].

— **Flashs lumineux** : Sensation d'éclairs visuels, signe que la rétine est tirée ou déformée par des cicatrices internes [19].
— **Cécité brutale(perte brutale de la vision)** : Peut survenir soudainement si une hémorragie importante ou un décollement de la rétine se produit [19].

IV Causes de la rétinopathie diabétique

La rétinopathie diabétique résulte principalement de l'effet prolongé du diabète sur la microcirculation rétinienne [5]. Trois mécanismes pathologiques principaux interviennent dans son développement :

— **Excès de sucre dans le sang** (hyperglycémie chronique) : Lorsque le taux de sucre reste élevé pendant longtemps, cela fragilise les petits vaisseaux sanguins de la rétine (la partie de l'œil qui capte la lumière). Ces vaisseaux deviennent poreux, ce qui laisse passer du liquide. Cela peut provoquer un gonflement au centre de la vision (œdème maculaire), entraînant une vision floue ou déformée [8].
— **Affaiblissement et déformation des vaisseaux de l'œil** : Avec le temps, les parois des vaisseaux de la rétine peuvent se déformer et former de petites poches appelées microanévrismes, qui peuvent fuir ou saigner. Plus la maladie progresse, plus il y a un risque de développement de nouveaux vaisseaux anormaux, fragiles, qui augmentent le risque de saignement interne et de décollement de la rétine [8].
— **Manque d'oxygène dans la rétine** (ischémie rétinienne) : Quand certains vaisseaux se bouchent, la rétine ne reçoit plus assez d'oxygène. En réponse, l'œil produit une substance qui fait pousser de nouveaux vaisseaux sanguins anormaux(voire la figure 1.5). Mais ces nouveaux vaisseaux sont très fragiles et peuvent aggraver la situation, en causant notamment des hémorragies [8].

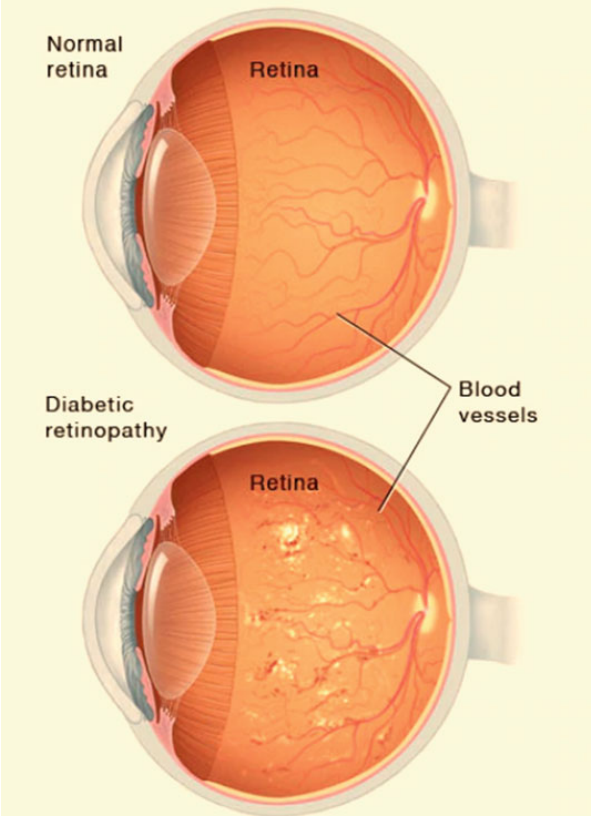


FIGURE 1.5 – Représentation comparée d'images oculaires normales et atteintes de rétinopathie diabétique, illustrant les vaisseaux sanguins(blood vessels)[21]

.

V Stades de la rétinopathie diabétique

La rétinopathie diabétique se divise en quatre stades principaux, qui reflètent la gravité et l'évolution des lésions rétiniennes :

1. **Rétinopathie non proliférante légère** : Ce stade initial est principalement caractérisé par la présence de microanévrismes, qui correspondent à de petites dilatations des capillaires rétiniens . Ces microanévrismes provoquent des fuites de liquide dans la rétine, entraînant de légères altérations de la microcirculation oculaire [7].
2. **Rétinopathie non proliférante modérée** : À ce stade intermédiaire, les lésions s'intensifient : les vaisseaux sanguins sont davantage endommagés, et la structure normale de la rétine commence à se déformer. Bien qu'aucun saignement important ne soit encore visible, ces anomalies peuvent favoriser l'apparition d'un œdème maculaire diabétique (OMD). Cet œdème se manifeste par un gonflement localisé au niveau de la macula, la zone centrale de la rétine responsable de la vision [7].
3. **Rétinopathie non proliférante sévère** : À ce troisième stade, une grande partie des vaisseaux sanguins de la rétine est obstruée, ce qui entraîne une insuffisance du flux sanguin vers les tissus rétiniens. Face à ce déficit en oxygène, l'œil tente de compenser en stimulant la production de facteurs de croissance, condui-

sant à la formation de nouveaux vaisseaux sanguins anormaux . Ces néovaisseaux, cependant, sont fragiles et ne fonctionnent pas correctement, augmentant le risque de complications [7].

4. **Rétinopathie diabétique proliférante** : Ce stade représente la phase la plus grave de la maladie. De nombreux néovaisseaux se développent au sein de la rétine ainsi que dans le corps vitré, le gel transparent qui remplit l'intérieur de l'œil. Ces nouveaux vaisseaux sont extrêmement fragiles et susceptibles de provoquer des hémorragies intraoculaires, qui peuvent entraîner une perte de vision rapide et sévère. Par ailleurs, la prolifération de tissus cicatriciels peut provoquer des tractions sur la rétine, favorisant un décollement rétinien. Ces dommages irréversibles peuvent aboutir à une cécité permanente si aucun traitement n'est administré [7].

Pour illustrer ces stades, des images représentatives des différents niveaux de gravité de la rétinopathie diabétique sont présentées à la figure 1.6, facilitant ainsi la compréhension des modifications rétiniennes au cours de la progression de la maladie.

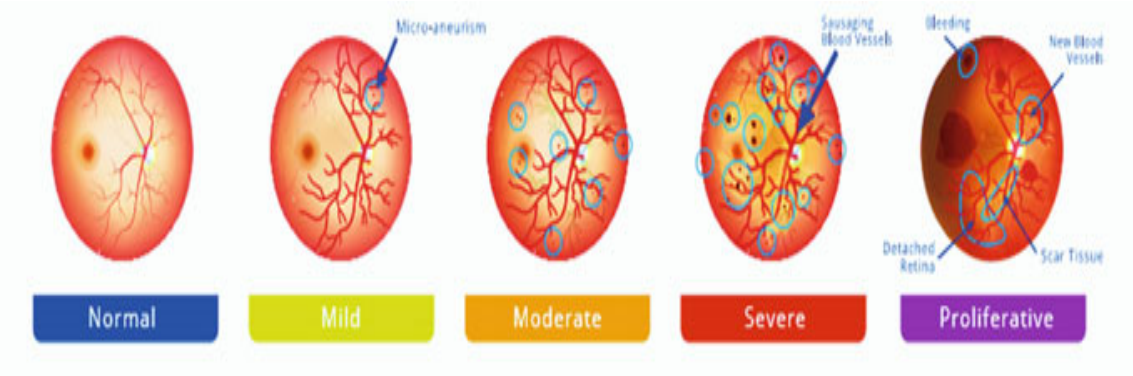


FIGURE 1.6 – Stades de la rétinopathie diabétique, a image rétinienne normale, b image rétinienne avec rétinopathie diabétique légère, c image rétinienne avec rétinopathie diabétique modérée, d image rétinienne avec rétinopathie diabétique proliférante[21].

VI Méthodes de Diagnostic de la Rétinopathie Diabétique

Plusieurs méthodes de diagnostic permettent de détecter précocement les lésions rétiniennes et d'évaluer leur sévérité :

- Examen du Fond d'Œil avec Dilatation Pupillaire : L'examen du fond d'œil après dilatation pupillaire est la méthode de référence pour le dépistage de la rétinopathie diabétique [8]. Il permet d'observer directement les anomalies vasculaires telles que :
  - Microanévrismes
  - Hémorragies rétiniennes
  - Nodules cotonneux
  - Néovaisseaux
- Photographies du Fond d'Œil (Rétinographie) :
  - **Rétinographie non mydriatique** : Méthode moins invasive utilisant des caméras spéciales[1].
  - **Angiographie à la fluorescéine** : Visualisation des fuites vasculaires[1].
- Tomographie par Cohérence Optique (OCT) : Examen non invasif pour détecter l'œdème maculaire diabétique . L'OCT-Angiographie permet de visualiser les anomalies microvasculaires sans injection[8].
- Autres Méthodes : Télédépistage et Intelligence Artificielle[8].

VII Traitement de la rétinopathie diabétique

Le traitement de la rétinopathie diabétique dépend principalement du type, de la sévérité et du risque évolutif de la maladie. Lorsqu'elle est détectée à un stade précoce, la rétinopathie peut souvent être maîtrisée efficacement grâce à un suivi ophtalmologique régulier associé à une gestion stricte du taux de glucose sanguin.

VII.1 Photocoagulation au laser

Parmi les options thérapeutiques, la photocoagulation au laser est l'une des méthodes les plus utilisées. Ce traitement vise à réduire les fuites des vaisseaux sanguins rétiniens et à prévenir la progression de la maladie, notamment en limitant les risques de perte de vision [8]. La procédure comporte plusieurs phases :

- **Laser diffus** (Photocoagulation en grille) : consiste à appliquer des impacts laser sur de multiples zones de la rétine, brûlant les petits trous ou zones endommagées afin de réduire le risque de cécité [1].
- **Laser focalisé**(Photocoagulation focale) : cible spécifiquement les vaisseaux sanguins qui fuient, particulièrement dans la zone maculaire, afin de traiter l'œdème maculaire diabétique [1].

VII.2 Chirurgie de la rétine(Vitrectomie)

La vitrectomie est une intervention chirurgicale plus invasive, consistant à retirer le gel vitréen lorsque des tissus cicatriciels ou des hémorragies gênent la vision [8]. Cette

procédure est particulièrement indiquée pour :

- Hémorragies vitréennes persistantes(Saignements persistants à l'intérieur de l'œil).
- Décollement de rétine tractionnel
- Membranes épirétiniennes (Présence de membranes qui gênent la vision).

VII.3 Thérapies médicamenteuses

En complément, des injections intraoculaires sont utilisées :

- **Corticostéroïdes** (dexaméthasone, triamcinolone) : pour réduire l'inflammation [8].
- **Agents anti-VEGF** (aflibercept, ranibizumab) : pour inhiber la croissance de nouveaux vaisseaux anormaux(néovascularisation) [8].

VII.4 Thérapie cellulaire

Basée sur l'injection de cellules souches dans la rétine, cette approche est actuellement en phase expérimentale [16].

Avec un traitement adapté, les chances de stabilisation ou d'amélioration de la vision atteignent environ 95%, bien que ces interventions ne constituent pas une guérison définitive [17]. Elles visent avant tout à ralentir la progression de la maladie et à limiter les complications graves.

VIII Prévention de la Rétinopathie Diabétique

La prévention de la rétinopathie diabétique repose sur deux piliers complémentaires : le contrôle métabolique strict et la surveillance ophtalmologique régulière .

VIII.1 Contrôle Métabolique Rigoureux

Une gestion efficace du diabète permet de ralentir l'apparition ou l'aggravation de la rétinopathie. Cela inclut :

- Glycémie(taux de sucre dans le sang) : Un contrôle regulier de la glycémie (taux de sucre dans le sang) est primordial [10] et le taux d'hémoglobine glyquée (HbA1c) ne doit pas dépassé 7%. [9].
- Pression Artérielle : elle ne doit pas dépasser 140/90 mmHg [11].
- Mode de Vie : Il faut une Activité physique[12] et une Nutrition équilibrée[12].

VIII.2 Surveillance Ophtalmologique

Le dépistage précoce permet de traiter les lésions avant qu'elles n'affectent la vision. Le dépistage se fait soit par un Examen annuel du fond d'œil avec dilatation pupillaire [13] ou un OCT maculaire [14].

IX Complications de la rétinopathie diabétique

Si la rétinopathie diabétique n'est pas détectée à temps, elle peut entraîner des complications graves qui peuvent altérer fortement ou définitivement la vision[5]. Voici les principales complications, classées selon leur fréquence et leur gravité.

1. **Œdème maculaire diabétique (OMD)(Gonflement de la zone centrale de la rétine)** : L'OMD est la complication la plus fréquente. Il s'agit d'un gonflement de la partie centrale de la rétine (la macula), provoqué par une fuite de liquide des petits vaisseaux sanguins. Cela entraîne une vision floue ou déformée. Sans traitement, cela peut provoquer une perte permanente de la vision centrale dans 25 % des cas [5].
2. **Hémorragie du vitré(Saignement à l'intérieur de l'œil)** : Dans les formes avancées de la rétinopathie diabétique, des vaisseaux anormaux peuvent apparaître. Ils sont très fragiles et peuvent saigner dans le vitré, une substance gélatineuse à l'intérieur de l'œil. Ce saignement provoque des taches sombres dans le champ de vision ou une perte soudaine de vision. Le sang peut disparaître naturellement en quelques mois, mais une intervention chirurgicale (vitrectomie) est parfois nécessaire [5].
3. **Décollement de la rétine (tractionnel)** : Chez environ 5 à 10 % des personnes atteintes de rétinopathie proliférante, la formation de tissus cicatriciels peut tirer sur la rétine et provoquer son décollement. Cela se manifeste par des éclairs lumineux et des zones sombres dans le champ visuel. Un traitement chirurgical rapide est indispensable pour éviter la cécité [5].
4. **Glaucome néovasculaire(Nouveaux vaisseaux anormaux)** : Lorsque l'œil manque d'oxygène, il produit une substance qui favorise la croissance de nouveaux vaisseaux sanguins anormaux dans l'iris. Ces vaisseaux bloquent la circulation normale des liquides de l'œil, augmentant ainsi la pression intraoculaire. Cela provoque une douleur intense et une perte rapide de la vision. Le traitement repose sur des injections urgentes (anti-VEGF) et parfois une chirurgie (trabéculectomie ou valve) [5].
5. **Cécité(perte de vision complète)** : Environ 5 % des personnes diabétiques deviennent aveugles après 30 ans de maladie. Le risque augmente en cas de forme avancée de la maladie, surtout s'il y a un glaucome ou un œdème dans les deux yeux. Des examens réguliers de la vue peuvent prévenir ces situations graves [5].

X Conclusion

Ce chapitre a dressé un panorama complet de la rétinopathie diabétique, en détaillant ses mécanismes physiopathologiques, ses stades d'évolution, ses symptômes, et l'arsenal thérapeutique disponible pour la prendre en charge. Il est clair que le pronostic visuel des patients repose essentiellement sur un dépistage le plus précoce possible et un suivi ophtalmologique régulier.

Cependant, comme nous l'avons évoqué, le défi du dépistage précoce est de taille. Les examens spécialisés, bien qu'efficaces, peuvent être coûteux et difficilement accessibles,

en particulier dans les régions reculées qui souffrent d'une pénurie criante d'ophtalmologistes. Cet obstacle majeur à la prévention de la cécité diabétique à l'échelle mondiale nécessite des solutions innovantes, scalable et efficaces.

C'est précisément pour relever ce défi que l'intelligence artificielle et les réseaux de neurones profonds émergent comme une solution des plus prometteuses. Dans le chapitre suivant, nous allons explorer comment ces technologies de pointe sont capables d'analyser automatiquement les images rétiniennes (rétinographies) pour détecter, avec une précision remarquable, les signes précoces de la rétinopathie diabétique. Cette révolution technologique a le potentiel de démocratiser l'accès au dépistage, d'alléger la charge des professionnels de santé et, in fine, de prévenir des milliers de cas de cécité évitables.

Chapitre 2

Vision Transformers & Etat de l'art

I Introduction

La rétinopathie diabétique (RD) constitue l'une des principales causes de cécité évitable dans le monde, touchant un nombre croissant de patients atteints de diabète [24]. Le dépistage précoce de cette maladie est crucial afin d'éviter des complications graves, et repose principalement sur l'analyse des images du fond d'œil. Ces dernières années, l'apprentissage profond a profondément transformé ce domaine, en particulier grâce aux réseaux de neurones convolutifs (CNN), qui ont montré une efficacité notable dans la détection automatique et la classification des stades de la RD [25].

Cependant, malgré ces avancées, les CNN présentent plusieurs limites : ils nécessitent de larges ensembles de données annotées, souvent difficiles à obtenir dans le domaine médical, et leur structure de convolution limite leur capacité à capturer les relations globales entre différentes régions de l'image [26]. Ces contraintes posent un défi important pour le développement de systèmes robustes et généralisables en imagerie médicale.

Pour répondre à ces problématiques, une nouvelle architecture inspirée des transformateurs utilisés en traitement du langage naturel a été introduite : le Vision Transformer (ViT) [27]. Contrairement aux CNN, les ViT divisent une image en patchs et appliquent un mécanisme d'attention qui permet de modéliser efficacement les dépendances à longue portée entre ces patchs. Cette approche offre des représentations plus riches et flexibles, particulièrement adaptées à la complexité des images médicales. En outre, combinés à l'apprentissage par transfert, les ViT permettent de contourner la contrainte de données limitées, en tirant parti de modèles pré-entraînés sur de grands ensembles d'images [28].

Dans ce chapitre, nous présentons les fondements théoriques des Vision Transformers ainsi que leurs principes de fonctionnement en vision par ordinateur. Nous exposons ensuite l'état de l'art des approches appliquant les ViTs à la détection de la rétinopathie diabétique.

II Introduction aux Transformers

En 2017, Vaswani et ses collègues ont introduit l'architecture Transformer dans l'article « Attention Is All You Need » [29]. Contrairement aux approches antérieures basées sur des réseaux de neurones convolutifs (CNN) ou récurrents (RNN), le Trans-

former s'appuie entièrement sur un mécanisme d'auto-attention, qui permet à chaque position d'une séquence de puiser de l'information dans l'ensemble des autres positions pour construire sa représentation. L'encodeur transforme la séquence d'entrée en vecteurs enrichis grâce à une succession de couches d'auto-attention multi-têtes et de réseaux feed-forward, tandis que le décodeur génère la séquence de sortie en combinant cette représentation encodée avec un décalage des jetons cibles [29]. En traitant toutes les positions en parallèle, l'architecture résout le problème de l'apprentissage séquentiel des RNN et capture efficacement les dépendances à longue portée, essentielles pour des tâches de traduction automatique, de modélisation de langage et de questions-réponses[33].

II.1 Architecture des Transformers

Les Transformers ont été présentés pour la première fois en 2017 dans l'article « Attention Is All You Need » de Vaswani et al., qui a proposé de remplacer à la fois les convolutions et la récursivité par un mécanisme d'auto-attention global capable de traiter en parallèle tous les éléments d'une séquence . À chaque position de la séquence, ce mécanisme calcule trois vecteurs (Query, Key, Value) et utilise le produit scalaire mis à l'échelle pour pondérer dynamiquement l'influence de chaque élément sur tous les autres, permettant ainsi de modéliser efficacement les dépendances à longue portée sans les contraintes séquentielles des RNN .L'architecture se compose de deux modules principaux : un encodeur, qui transforme l'entrée en une représentation intermédiaire, et un décodeur, qui génère la sortie en s'appuyant sur cette représentation et sur les décalages de la séquence cible . Chacune de ces parties est organisée en couches alternées d'auto-attention multi-têtes et de réseaux feedforward entièrement connectés, avec des connexions résiduelles et des normalisations LayerNorm pour stabiliser l'apprentissage et accélérer la convergence.. En traitant l'ensemble de la séquence en une seule fois, les Transformers surmontent les limitations des architectures antérieures en matière de parallélisation et de capture de relations distantes, ce qui a favorisé leur adoption massive dans des tâches telles que la traduction automatique, la modélisation de langage et les systèmes de questions-réponses [29](Figure 2.1) .

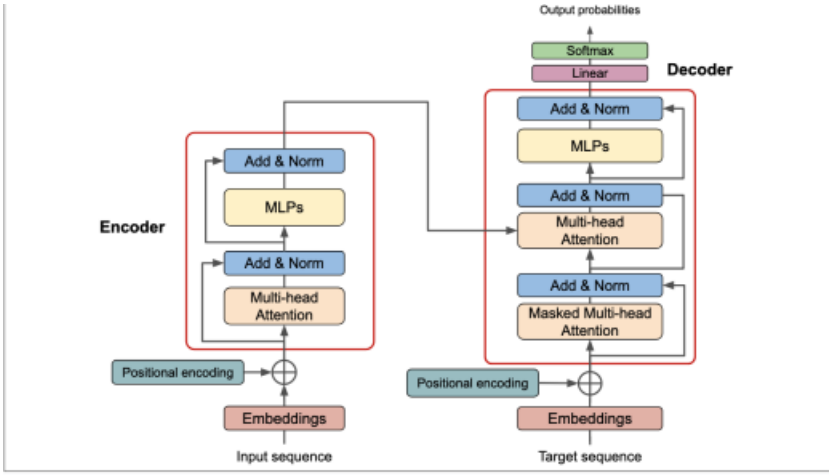


FIGURE 2.1 – Structure de l'encodeur et du décodeur [29]

Les deux parties principales des transfomers 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).

Auto-attention multi-têtes : Chaque tête apprend une manière différente de pondérer les interactions entre éléments d'entrée. Par exemple, dans la traduction, une tête peut se focaliser sur les accords grammaticaux, une autre sur la co-référence. En vision, cela peut correspondre à différents niveaux de détail entre régions de l'image [29].

Feedforward positionnel : Après l'auto-attention, chaque position est traitée individuellement par un réseau de deux couches linéaires séparées par une activation non linéaire (ReLU), permettant d'augmenter la capacité de modélisation locale[29].

II.2 Du NLP à la vision

Les Transformers, conçus à l'origine pour le NLP, ont été transposés à la vision en découpant chaque image en patchs fixes (ex. 16×16), puis en traitant cette séquence de patchs via un encodeur Transformer pur, sans aucune couche convolutionnelle supplémentaire. Dosovitskiy et ses collègues ont montré que, dès lors qu'ils sont pré-entraînés sur d'énormes bases (ImageNet-21k, JFT-300M) puis affinés sur des jeux de taille plus modeste (ImageNet, CIFAR-100, VTAB...), ces Vision Transformers égalent ou surpassent les meilleurs CNN tout en offrant une architecture plus simple [27].

II.2.1 Principes de base des Vision Transformers (ViT)

Les Vision Transformers (ViT) reposent sur un principe très simple : au lieu de traiter l'image dans son ensemble, on la découpe en un maillage de patchs de taille fixe (par exemple 16 × 16 pixels) Chacun de ces patchs est aplatit et projeté linéairement pour former une suite de vecteurs d'entrée, à laquelle on ajoute un encodage de position afin de conserver l'information spatiale. Cette séquence de vecteurs est ensuite injectée dans un encodeur Transformer « classique » : les blocs d'auto-attention multi-têtes permettent à chaque patch de capter le contexte global de l'image, tandis que les couches feed-forward viennent enrichir ces représentations locales. Enfin, la sortie du token de classification (ou l'agrégation de tous les patch-embeddings) fournit une représentation encodée de l'image, prête à être utilisée pour des tâches telles que la classification, la détection d'objets ou la segmentation sémantique [27] .

II.3 Architecture des Vision Transformers

Le traitement d'une image par un Vision Transformer (ViT) commence par la division de l'image en une séquence de petits patchs réguliers, comme illustré dans la figure 2.2.

L'architecture Transformer standard, conçue pour traiter des séquences de jetons (comme des mots), a été adaptée aux images en les convertissant en une séquence de patchs. Pour cela, l'image d'entrée, de dimensions hauteur × largeur × nombre de canaux, est découpée en patchs de taille fixe (par exemple 16 × 16 pixels). Chaque patch est ensuite aplati, c'est-à-dire transformé en un vecteur(voire la figure 2.3), de

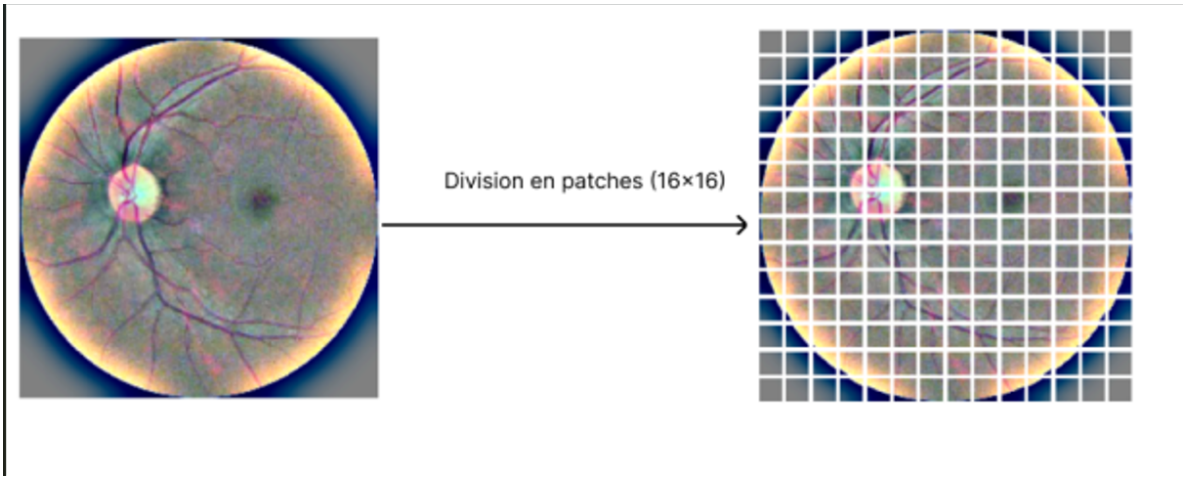


FIGURE 2.2 – Division en patchs de taille fixe

sorte que l'ensemble de ces vecteurs forme la séquence d'entrée du modèle. Le nombre total de patchs détermine la longueur de cette séquence [27] .

À chaque patch est associé un encodage de position, qui est ajouté pour conserver l'information spatiale de l'image. Un vecteur spécial, appelé « embedding de classification », est également inséré au début de la séquence. Ce vecteur jouera un rôle central dans la tâche de classification, en résumant l'information globale extraite de tous les patchs à la sortie du modèle [27] .

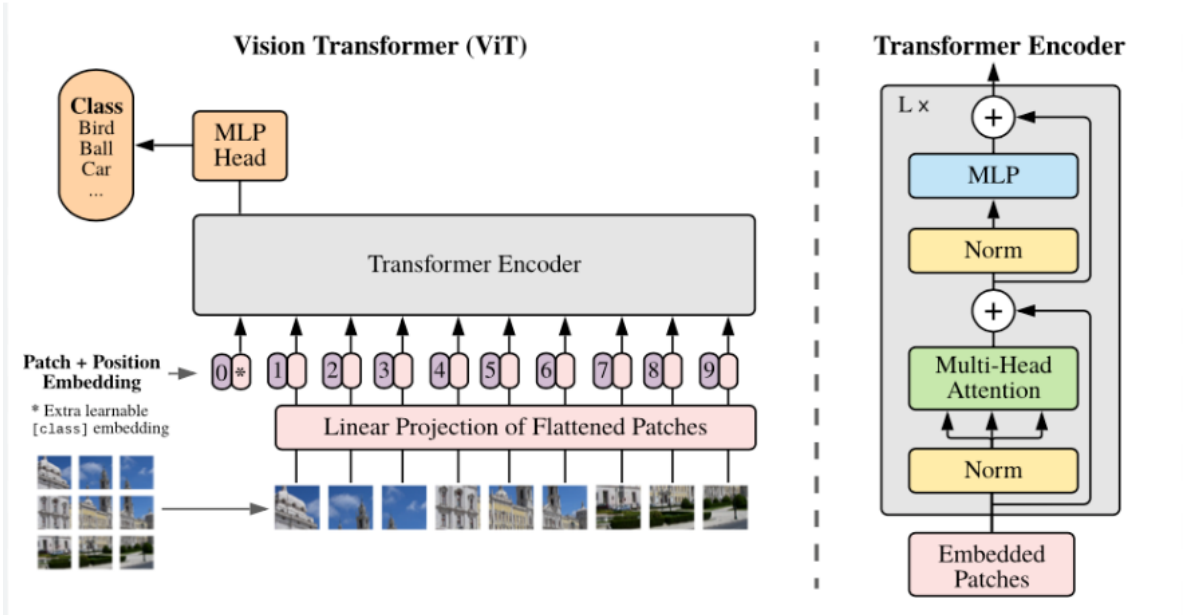


FIGURE 2.3 – Architecture Vision Transformers [27]

II.3.1 Mécanisme d'attention dans les Vision Transformers

Dans les Vision Transformers, les images sont d'abord découpées en petits blocs appelés **patchs**, qui sont ensuite transformés en vecteurs. Ces vecteurs passent ensuite

par le **mécanisme d'attention**, qui permet au modèle de comprendre les relations entre différentes parties de l'image [27].

Le calcul de l'attention se fait selon la formule suivante :

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Pour chaque patch d'image, le modèle calcule trois vecteurs :

Q (requête) : ce que le patch cherche.

K (Clé) : ce que les autres patchs représentent.

V (Valeur) : l'information contenue dans chaque patch.

Le produit QK^T mesure la similarité entre les patchs.Un patch compare ainsi son vecteur Query à tous les vecteurs Key des autres patchs [27].

Division par $\sqrt{d_k}$: Cette normalisation évite que les scores deviennent trop grands, ce qui rendrait la suite du calcul instable [27].

Application de Softmax : C'est ici que l'attention est réellement calculée. La fonction Softmax transforme les scores en poids compris entre 0 et 1, qui indiquent à quel point un patch doit faire attention aux autres. Ces poids forment une distribution de probabilité [27].

Combinaison avec les valeurs (V) : Chaque patch combine ensuite l'information des autres patchs, en leur donnant plus ou moins d'importance selon les poids obtenus avec la Softmax [27].

II.4 Applications des ViT dans la vision par ordinateur

Les ViT ont prouvé leur efficacité dans plusieurs domaines :

- **Agriculture** : identification précoce de maladies foliaires sur des images de plantes haute résolution [30].
- **Véhicules autonomes** : segmentation sémantique des routes et détection multi-classes en temps réel [32].
- **Surveillance** : reconnaissance d'actions humaines et détection d'événements suspects dans des vidéos à faible luminosité [31].
- **Imagerie médicale** : classification de tumeurs mammaires et détection de la rétinopathie diabétique à partir de scans OCT et de fonds d'œil [23].

III Bases de Données et Métriques d'Évaluation des Performances

Dans cette section, nous présentons les principaux jeux de données utilisés pour la détection de la rétinopathie diabétique, ainsi que les métriques d'évaluation permettant de mesurer la performance des modèles.Enfin, nous analysons les travaux récents exploitant les Vision Transformers pour la détection de la rétinopathie diabétique.

III.1 Les jeux de données pour la détection de la rétinopathie diabétique

Dans le domaine du diagnostic médical assisté par intelligence artificielle, plusieurs jeux de données publics sont utilisés pour entraîner et évaluer des modèles de détection automatique de la rétinopathie diabétique à partir d'images du fond d'œil. Les jeux de données les plus courants sont :

APTOS 2019 Blindness Detection [39]

APTOS 2019 est un jeu de données de 3 662 images rétiniennes, annotées par des experts selon cinq niveaux de gravité de la rétinopathie diabétique(0-4), destiné à la classification automatique.

EyePacs [38]

EyePACS est un vaste jeu de données de plus de 88 000 images rétiniennes, annotées par des ophtalmologues selon cinq niveaux de gravité de la rétinopathie diabétique(0-4). Chaque patient est représenté par deux clichés (œil gauche et droit). En raison de la forte variabilité de qualité des images, un prétraitement est généralement nécessaire. Ce jeu est largement utilisé pour l'apprentissage profond à grande échelle appliqué à la classification de la rétinopathie diabétique.

Messidor[35] / Messidor-2[34]

Les jeux de données Messidor et Messidor-2, développés en France par l'INSERM, sont destinés à l'évaluation du diagnostic automatisé en ophtalmologie. Messidor 1 contient 1 200 images de 400 patients, tandis que Messidor 2 regroupe 1 748 images en haute résolution, annotées selon cinq niveaux de gravité de la rétinopathie diabétique(0-4) et l'évaluation de l'œdème maculaire. Ces bases sont des références majeures pour comparer les modèles d'intelligence artificielle en analyse d'images ophtalmiques.

DDR(Diabetic Retinopathy Dataset)[37]

est un jeu de données multicentrique de 13 673 images rétiniennes couleur, annotées en cinq niveaux de gravité de la rétinopathie diabétique(0-4), largement utilisé pour la classification automatique en imagerie médicale.

OIA-DDR (Open-Access Diabetic Retinopathy)[36]

OIA-DDR est une extension du jeu de données DDR, publiée en 2020. Il contient 13 673 images rétiniennes annotées à la fois pour le niveau de rétinopathie diabétique (0–4) et pour la présence de lésions (microanévrismes, hémorragies, exsudats), avec une forte diversité visuelle. Il est utilisé pour la classification, la segmentation et l'apprentissage multi-tâches.

III.2 Métriques d'évaluation des performances

Les métriques d'évaluation permettent de mesurer la performance d'un modèle de classification en quantifiant la qualité de ses prédictions.Elles servent à comparer différents modèles entre eux et à identifier leurs forces et faiblesses selon le contexte d'utilisation. Certaines métriques évaluent la justesse globale des prédictions, tandis que d'autres se concentrent sur la capacité du modèle à bien distinguer les différentes classes, en particulier lorsque les données sont déséquilibrées. Leur utilisation est donc essentielle pour juger de la fiabilité et de la pertinence d'un modèle en pratique :

— Précision Globale (Accuracy) L'exactitude représente la proportion d'exemples correctement classés par le modèle par rapport à l'ensemble des prédictions réalisées. Elle est définie par la relation suivante [41] :

$$\text{Accuracy} = \frac{VP + VN}{VP + VN + FP + FN} \tag{2.1}$$

où :

— VP : Vrais Positifs,Cas où le modèle prédit correctement la classe positive,par exemple en diagnostic médical un Patient malade correctement identifié
— VN : Vrais Négatifs,Cas où le modèle prédit correctement la classe négative par exemple en finance Transaction légitime non signalée comme fraude
— FP : Faux Positifs ,Cas où le modèle prédit à tort la classe positive par exemple en diagnostic médical Patient sain déclaré malade
— FN : Faux Négatifs,Cas où le modèle rate une instance positive par exemple en cybersécurité Attaque réelle non détectée

Interprétation :

— Métrique intuitive mais potentiellement trompeuse pour les ensembles déséquilibrés
— Utile pour une première évaluation globale
— Peut masquer des problèmes spécifiques à certaines classes
— Valeur idéale : 1 (100% de prédictions correctes)

— Précision par Classe (Precision)
La précision mesure la fiabilité des prédictions positives [41] :

$$\text{Précision} = \frac{\text{V P}}{\text{V P} + \text{F P}} \tag{2.2}$$

Interprétation :

— Mesure la qualité des prédictions positives
— Critique lorsque les faux positifs sont coûteux (ex : diagnostics médicaux)
— Sensible au déséquilibre de classes
— Valeur idéale : 1 (aucun faux positif)

— Rappel (Recall ou Sensibilité)
Le rappel mesure la capacité à détecter les vrais positifs[41] :

$$\text{Rappel} = \frac{\text{V P}}{\text{V P} + \text{F N}} \tag{2.3}$$

Interprétation :
— Mesure l'exhaustivité des prédictions positives
— Crucial lorsque les faux négatifs sont critiques (ex : détection de fraudes)
— Indique si le modèle ignore des instances importantes
— Valeur idéale : 1 (aucun faux négatif)

— Aire sous la courbe ROC (AUC)
L'AUC mesure la capacité d'un modèle à distinguer entre les classes positive et négative en évaluant la performance à tous les seuils de classification possibles. Elle est définie par l'aire sous la courbe ROC (Receiver Operating Characteristic), qui trace le taux de vrais positifs (TPR) contre le taux de faux positifs (FPR) pour différents seuils[41].

$$\text{AUC} = \int_0^1 \text{ROC}(t), dt \tag{2.4}$$

où :
— TPR (Sensibilité) : Proportion de positifs correctement identifiés (ex. : patients diabétiques correctement diagnostiqués).
— FPR (1 - Spécificité) : Proportion de négatifs incorrectement classés comme positifs (ex. : patients sains déclarés diabétiques).

Interprétation :
— Valeur comprise entre 0.5 (performance aléatoire) et 1 (discrimination parfaite).
— Utile pour comparer des modèles indépendamment du seuil de classification.
— Adaptée aux problèmes de classification binaire et aux ensembles déséquilibrés.
— Exemple : une AUC élevée indiquerait une excellente capacité à distinguer les patients atteints de rétinopathie diabétique des non-atteints.

— Spécificité
La spécificité mesure la capacité d'un modèle à identifier correctement les cas négatifs. Elle est définie comme la proportion de vrais négatifs parmi tous les cas réellement négatifs [41] :

$$\text{Spécificité} = \frac{VN}{VN + FP} \tag{2.5}$$

où :
— VN : Vrais négatifs (ex. : patients sans rétinopathie correctement classés).
— FP : Faux positifs (ex. : patients sains incorrectement diagnostiqués avec la rétinopathie).

Interprétation :
— Valeur idéale : 1 (aucun faux positif).
— Critique dans les domaines où les faux positifs ont des conséquences graves (ex. : diagnostics médicaux erronés).

— Souvent utilisée avec la sensibilité pour évaluer les compromis (trade-offs).

— Accord de Kappa (Kappa)
Le coefficient Kappa de Cohen mesure l'accord entre les prédictions du modèle et les vérités terrain, en corrigeant l'accord dû au hasard. Il est défini par [41] :

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{2.6}$$

où :
— p_o : Proportion observée d'accords (exactitude).
— p_e : Proportion attendue d'accords aléatoires.
Interprétation :
— Valeur comprise entre -1 (désaccord total) et 1 (accord parfait).
— Utile pour évaluer la fiabilité d'un modèle dans des tâches subjectives (ex. : annotation médicale).
— Un $\kappa > 0.6$ est généralement considéré comme acceptable.

— Score F1
Le score F1 est une mesure harmonique entre la précision (*precision*) et le rappel (*recall*). Il est défini comme suit [41] :

$$\text{F1-score} = 2 \cdot \frac{\text{Précision} \cdot \text{Rappel}}{\text{Précision} + \text{Rappel}} \tag{2.7}$$

Interprétation :
— Moyenne harmonique entre précision et rappel
— Particulièrement utile pour les ensembles déséquilibrés
— Métrique robuste pour comparer des modèles
— Pénalise les déséquilibres entre précision et rappel
— Valeur idéale : 1 (parfaite précision et rappel)

— Matrice de Confusion comme métrique d'évaluation

La matrice de confusion est un outil fondamental pour évaluer les performances des modèles de classification. Contrairement aux simples métriques de précision, elle fournit une analyse plus nuancée des types d'erreurs commises par le classifieur[41].

— Définition Mathématique :
Pour un problème de classification à C classes, la matrice de confusion M est une matrice carrée de dimension $C \times C$ où :

$$M_{i,j} = \text{Nombre d'instances de classe } i \text{ prédites comme classe } j \tag{2.8}$$

Les éléments diagonaux $M_{i,i}$ représentent les prédictions correctes, tandis que les éléments non diagonaux $M_{i,j}$ (avec $i \neq j$) représentent les erreurs de classification,voire.
Interprétation :
— Les classes bien classées (valeurs élevées sur la diagonale)
— Les confusions entre classes (valeurs hors diagonale élevées)
— Les biais éventuels du classifieur
— Les classes difficiles à distinguer

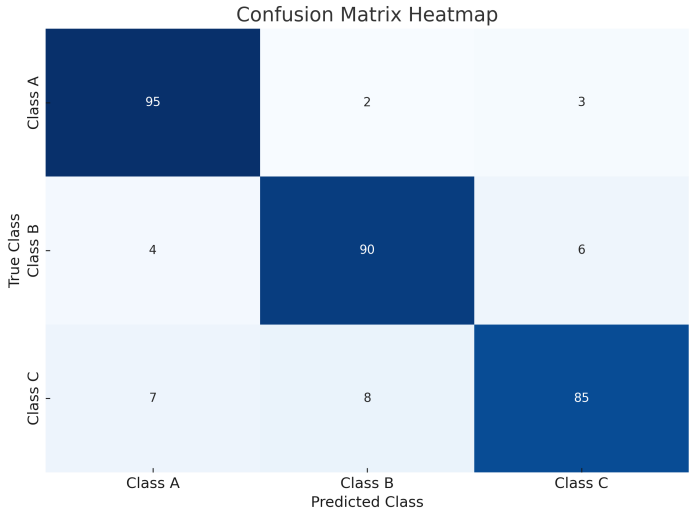


FIGURE 2.4 – Matrice de confusion pour une classification à trois classes. Les éléments diagonaux représentent les prédictions correctes, tandis que les éléments hors diagonale indiquent les erreurs de classification. L'intensité de la couleur reflète le nombre d'échantillons. Ce format permet d'identifier les points forts (ex. : la classe A présente peu d'erreurs) et les faiblesses du modèle (ex. : la classe C est souvent confondue avec la classe B)..[41]

III.3 Revue de littérature

Dans le cadre de la détection automatique de la rétinopathie diabétique à partir d'images du fond d'œil, les architectures basées sur les *Vision Transformers (ViTs)* ont récemment suscité un intérêt croissant[23]. Contrairement aux réseaux de neurones convolutifs traditionnels (CNN), les ViTs exploitent des mécanismes d'attention globale pour capturer des relations spatiales à longue distance[27].

Plusieurs travaux récents ont démontré l'efficacité des Vision Transformers dans la classification de la rétinopathie diabétique[23], en particulier grâce à leur capacité à traiter des images haute résolution tout en préservant les structures visuelles importantes.

Dans cette optique, nous présentons ci-dessous un ensemble de contributions scientifiques qui illustrent comment les Vision Transformers ont été adaptés et optimisés pour améliorer la détection automatique de la rétinopathie diabétique.

1. Yao et al[42] : L'article « FunSwin : A deep learning method to analysis diabetic retinopathy grade and macular edema risk based on fundus images » propose une méthode d'apprentissage profond appelée FunSwin pour analyser le grade de la rétinopathie diabétique (DR) à partir d'images du fond d'œil. FunSwin est une méthode de deep learning basée sur le Swin Transformer, conçue pour analyser automatiquement les images du fond d'œil afin de classer les stades de rétinopathie diabétique. L'architecture combine trois composants principaux : (1) un backbone Swin Transformer qui extrait des caractéristiques hiérarchiques via des mécanismes d'attention locale et globale, (2) une couche de Global Average Pooling pour réduire la dimensionnalité, et (3) une tête de classification linéaire.

Dataset Utilisé : MESSIDOR

Résultats : Pour la classification binaire, le modèle a atteint une accuracy de 0,9062, un F1-score de 0,9366, une sensibilité de 0,9376 et une spécificité de 0,8171. En revanche, pour la classification multiclasse, les performances obtenues sont une accuracy de 0,8412, un F1-score de 0,8400, une sensibilité de 0,8154 et une spécificité de 0,9413.

2. Gu et al[40] : L'article « Classification of Diabetic Retinopathy Severity in Fundus Images Using the Vision Transformer and Residual Attention » propose un modèle basé sur deux modules principaux pour la classification de la sévérité de la rétinopathie diabétique (DR) à partir d'images du fond d'œil. Le premier module, Feature Extraction Block (FEB), utilise un Vision Transformer (ViT) pour extraire des caractéristiques fines en divisant l'image en patches, en ajoutant des embeddings de position et en traitant les séquences via un encodeur transformer multi-têtes. Le second module, Grading Prediction Block (GPB), repose sur une attention résiduelle spécifique aux classes pathologiques (CSRA, pour Class-Specific Residual Attention) — un mécanisme qui génère des caractéristiques discriminantes par classe via une attention spatiale et une fusion résiduelle — pour classer les stades de la DR. Le CSRA calcule des scores d'attention spatiale pour chaque classe, fusionne les caractéristiques locales et globales, et produit une prédiction finale.

   Dataset Utilisé : DDR

   Résultats : Le modèle a atteint une accuracy de 0,8235, une AUC de 0,9018, une sensibilité de 0,8140 et une spécificité de 0,8245.

3. Yang et al[43] : L'article « vision transformer with masked autoencoders for referable diabetic retinopathy classification based on large-size retina image » propose une nouvelle méthode pour la classification de la rétinopathie diabétique référable (rDR), en utilisant un modèle Vision Transformer (ViT) pré-entraîné sur un large ensemble d'images rétiniennes de haute résolution. Cette étude propose une architecture qui combine un Vision Transformer (ViT) avec des autoencodeurs masqués (MAE) pour la classification de la rétinopathie diabétique référable. Les MAE reposent sur un mécanisme d'apprentissage auto-supervisé où une portion significative de l'image (50 %-75 %) est masquée aléatoirement, contraignant le modèle à reconstruire les parties occultées. Cette approche permet d'extraire des caractéristiques visuelles robustes sans recourir à des annotations manuelles. Le processus opère en divisant l'ensemble des données en patches non chevauchants. Durant la phase de prétraitement, l'encodeur ViT ne traite que les patches visibles tandis qu'un décodeur léger tente de restituer l'image complète à partir des représentations latentes et des patches masqués. Cette reconstruction forcée améliore la compréhension par le modèle des relations spatiales et des structures anatomiques. Après cette étape d'apprentissage auto-supervisé, seule la partie encodeur du système est conservée et affinée pour la classification supervisée.

Dataset Utilisé : Les auteurs ont utilisé plusieurs datasets publics :
- **Pré-entraînement :**
  - APTOS
  - EyePACS
  - Messidor-2
  - OIA-DDR
- **Fine-tuning et test :**
  - APTOS
- **Trois sous-ensembles :**
  - Dataset1 : 17 349 images (APTOS + Messidor-2 + OIA-DDR)
  - Dataset2 : 106 051 images (Dataset1 + EyePACS)
  - Dataset3 : 3 662 images (APTOS)
  - **ViT (pré-entraîné sur Dataset2 puis testé sur Dataset3)**

Résultats : Pour la classification binaire, le modèle a atteint une accuracy de 0,9342, une AUC de 0,9825, une sensibilité de 0,9662 et une spécificité de 0,9539

4. Yuanyuan et al[44] : L'article « STMF-DRNet : A multi-branch fine-grained classification model for diabetic retinopathy using Swin-TransformerV2 » propose STMF-DRNet, un modèle avancé pour la classification fine des stades de la rétinopathie diabétique (RD) à partir d'images du fond d'œil. Basé sur un Swin-Transformer, le modèle intègre une architecture multi-branches pour extraire des caractéristiques à différentes échelles (globales, locales et fines), combinée à des mécanismes d'attention hybrides (CBAM) pour cibler les lésions pertinentes. tandis qu'un mécanisme d'attention par catégorie fusionne intelligemment les caractéristiques pour une classification finale.

   Datasets Utilisés :
   - DDR
   - EyePACS
   - APTOS-2019
   - Dataset clinique : collecté localement, utilisé pour tester la capacité de généralisation du modèle.

   Résultats : Pour le jeu de données DDR, le modèle a obtenu une accuracy de 0,879, un F1-score de 0,891, un Kappa de 0,890, une sensibilité de 0,816 et une spécificité de 0,849. Avec le jeu de données EyePACS, les performances atteignent une accuracy de 0,863, un F1-score de 0,877, un Kappa de 0,841, une sensibilité de 0,808 et une spécificité de 0,896. Concernant APTOS-2019, le modèle présente une accuracy de 0,855, un F1-score de 0,859, un Kappa de 0,872, une sensibilité de 0,810 et une spécificité de 0,904. Enfin, avec les données cliniques, les résultats obtenus sont une accuracy de 0,77, un F1-score de 0,77, un Kappa de 0,877, une sensibilité de 0,77 et une spécificité de 0,942.

5. Touati et al[45]

   L'article « DRCCT : Enhancing Diabetic Retinopathy Classification with a Compact Convolutional Transformer » propose une architecture hybride nom-

mée DRCCT, appliquée à la classification de la rétinopathie diabétique à partir d'images du fond d'œil. Cette méthode associe des composants de réseaux convolutifs (CNN) et de Transformers dans le but de traiter simultanément les informations locales et globales de l'image. L'architecture débute par une étape de tokenisation convolutive, où des couches CNN génèrent des représentations locales à l'aide de filtres convolutifs de tailles croissantes (16, 32, 64, 128), remplaçant le découpage en patchs fixes utilisé dans les ViT classiques. Ces représentations sont ensuite traitées par un encodeur Transformer comprenant une attention multi-têtes, des connexions résiduelles et une régularisation par stochastic depth,forçant ainsi le modèle à apprendre des représentations plus robustes et à réduire le surapprentissage. L'agrégation des informations issues de la séquence de tokens est réalisée à l'aide d'un sequence pooling basé sur l'attention, en remplacement d'un pooling moyenné. Une couche de classification dense est enfin utilisée pour prédire l'un des cinq niveaux de gravité de la rétinopathie diabétique.

Dataset Utilisé : APTOS

Résultats : Le modèle a atteint une accuracy de 0,9693, une sensibilité de 0,9889 et un F1-score de 0,973.

* Discussion des résultats rapportés dans la littérature
L'analyse du tableau 2.1 met en évidence plusieurs tendances importantes. Tout d'abord, les architectures basées uniquement sur le Vision Transformer (ViT), comme dans les travaux de Yao et al[42] 2022, Gu et al[40] 2023 et Yang et al[43] 2024, ont permis d'obtenir des résultats compétitifs, avec des accuracies allant de 0,8235 à 0,9342 et des AUC atteignant jusqu'à 0,9825. Toutefois, ces approches présentent des limites récurrentes, notamment la dépendance à la qualité des données, l'exigence computationnelle élevée et le risque de sur-ajustement.

En revanche, les approches hybrides combinant ViT et CNN, comme celles de Yuanyuan et al[44] 2025 et Touati et al[45] (2025), semblent offrir une amélioration significative des performances. Par exemple, les résultats de Touati et al[45] 2025 sur le jeu de données APTOS atteignent une accuracy de 0,9693 et une sensibilité remarquable de 0,9889, ce qui constitue l'une des meilleures performances reportées. Néanmoins, ces méthodes exigent une puissance de calcul considérable et demeurent sensibles à la qualité des images, ce qui peut limiter leur déploiement en contexte clinique réel.

IV Conclusion

Dans ce chapitre, nous avons introduit les fondements des Transformers et montré comment leur adaptation à la vision, à travers les Vision Transformers (ViT), a ouvert de nouvelles perspectives dans le traitement d'images médicales. Nous avons détaillé leur architecture, en insistant sur le rôle central du mécanisme d'attention, et présenté leurs principales applications en vision par ordinateur. Ensuite, nous avons dressé l'état de l'art relatif à la détection automatique de la rétinopathie diabétique (RD), en abordant les jeux de données disponibles, les métriques d'évaluation couramment utilisées ainsi que les contributions récentes exploitant les ViTs pour cette tâche. Ces travaux ont confirmé le potentiel de ces modèles pour améliorer la précision et la robustesse du diagnostic assisté par l'intelligence artificielle, tout en mettant en évidence plusieurs défis persistants : le déséquilibre des données utilisées, le coût computationnel élevé, le risque de sur-apprentissage, la forte dépendance à la qualité des images et la difficulté d'applicabilité en temps réel. Dans le chapitre suivant, nous examinerons trois modèles basés sur les Vision Transformers, dont le potentiel réside dans leur capacité à mieux exploiter les représentations visuelles complexes,afin de proposer une détection plus robuste et efficace de la rétinopathie diabétique.

Début du tableau 2.1

Auteurs/Année	Dataset	Architecture	Résultats	Limites
Yao et al[42] 2022	MESSIDOR[35]	Vision Transformer (ViT)	**Classification binaire :** Accuracy : 0,9062 F1-score : 0,9366 Sensibilité : 0,9376 Spécificité : 0,8171 **Classification multiclasse :** Accuracy : 0,8412 F1-score : 0,8400 Sensibilité : 0,8154 Spécificité : 0,9413	Validation limitée Coût computationnel non discuté
Gu et al[40] 2023	DDR[37]	Vision Transformer (ViT)	Accuracy : 0,8235 AUC : 0,9018 Sensibilité : 0,8140 Spécificité : 0,8245	Déséquilibre des données utilisées Complexité et coût computationnel important
Yang et al[43] 2024	APTOS [39]	Vision Transformer (ViT)	Accuracy : 0,9342 AUC : 0,9825 Sensibilité : 0,9662 Spécificité : 0,9539	Exigences computationnelles élevées Risque de sur-ajustement (overfitting)

TABLE 2.1 – Comparaison de méthodes utilisant les Vision Transformers pour la classification de la rétinopathie diabétique.

Yuanyuan et al[44] 2025	DDR[37], EyePACS[38], APTOS[39], Données cliniques	ViT + CNN	**DDR :** Accuracy : 0,879, F1-Score : 0,891, Kappa : 0,890, Sensibilité : 0,816, Spécificité : 0,849 **EyePACS :** Accuracy : 0,863, F1-Score : 0,877, Kappa : 0,841, Sensibilité : 0,808, Spécificité : 0,896 **APTOS :** Accuracy : 0,855, F1-Score : 0,859, Kappa : 0,872, Sensibilité : 0,810, Spécificité : 0,904 **Données cliniques :** Accuracy : 0,77, F1-Score : 0,77, Kappa : 0,877, Sensibilité : 0,77, Spécificité : 0,942	Besoins en données et calcul Difficulté d'applicabilité en temps réel
Touati et al[45] 2025	APTOS[39]	ViT + CNN	Accuracy : 0,9693 Sensibilité : 0,9889 F1-Score : 0.973	Dépendance à la qualité des images Exigences computationnelles élevées

Fin du tableau 2.1

Chapitre 3

Conception & Réalisation

I Introduction

L'objectif dans ce chapitre est de concevoir un modèle de classification de la rétinopathie diabétique reposant sur les Vision Transformers. Dans cette optique, plusieurs variantes de modèles basés sur les Vision Transformers sont étudiées, notamment ViT16, ViT32 ainsi qu'un modèle hybride combinant les deux. Une analyse comparative des performances obtenues est ensuite réalisée afin d'évaluer l'efficacité de chaque approche dans le cadre de la prédiction et de la classification de la rétinopathie diabétique.

II Méthodologie

Dans le cadre de cette étude, nous avons exploré l'utilisation de différentes architectures de Vision Transformers appliquées à la classification de la rétinopathie diabétique en plusieurs niveaux de gravité. Deux variantes principales ont d'abord été considérées : ViT16, reposant sur une taille de patch de 16×16, et ViT32, basé sur des patchs de 32×32.

Par la suite, nous avons proposé une architecture hybride combinant les sorties issues de ViT16 et de ViT32, dans l'objectif d'exploiter la complémentarité entre les deux représentations et d'améliorer la capacité de discrimination du modèle.

La méthodologie adoptée s'articule autour de plusieurs étapes essentielles : prétraitement des images (nettoyage, redimensionnement et normalisation)(voire figure 3.1), construction des modèles (ViT16, ViT32 et le modèle hybride), entraînement sur l'ensemble des données annotées, puis évaluation des performances à l'aide de différentes métriques de classification. Ces étapes, qui structurent notre approche expérimentale, seront présentées en détail dans les sections suivantes.

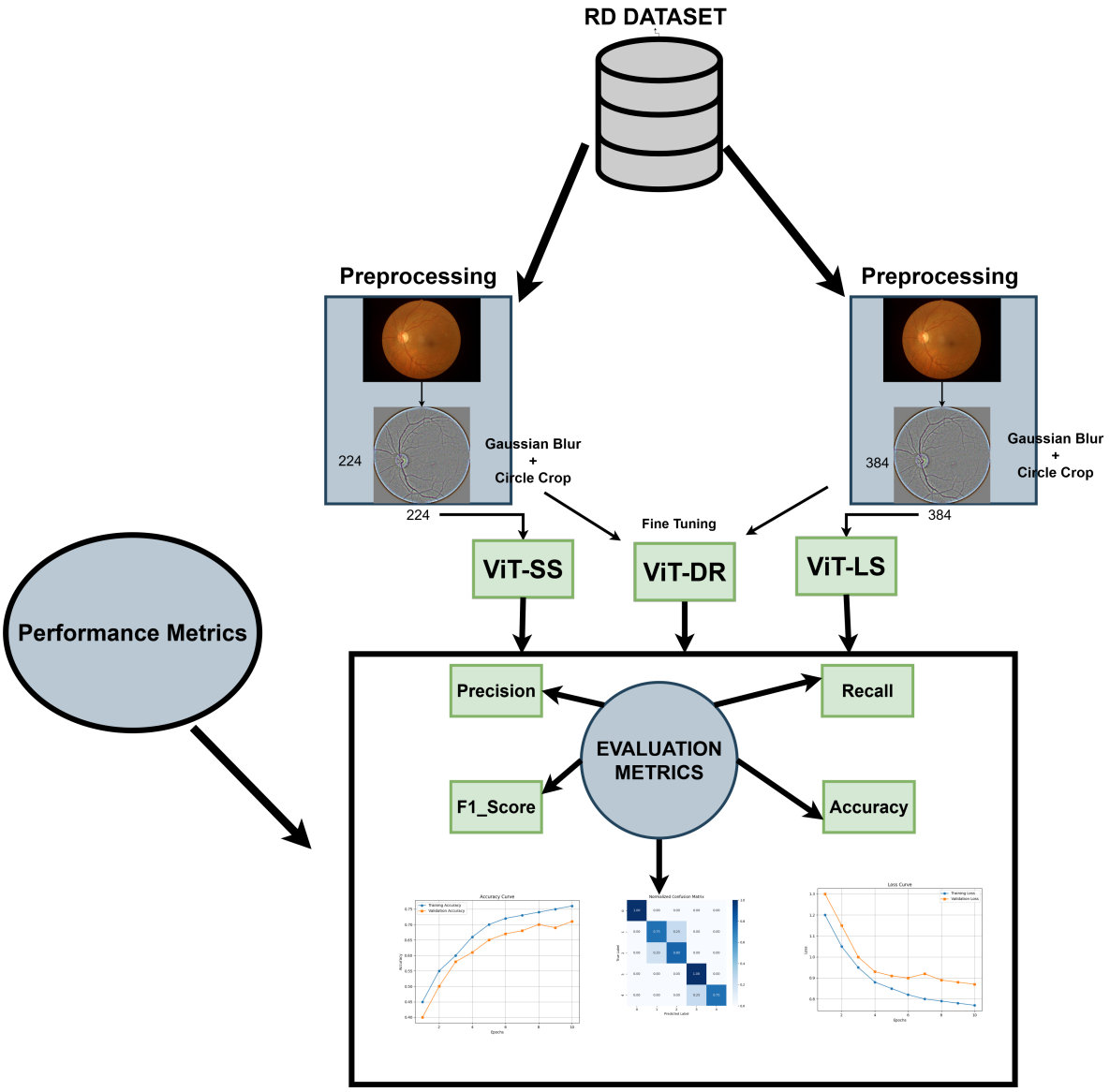


FIGURE 3.1 – Pipeline du processus de classification de la rétinopathie diabétique

II.1 Préparation des données

II.1.1 Dataset utilisé

Nous avons utilisé le dataset APTOS 2019, constitué d'images de fond d'œil,reconnu pour sa qualité et sa large utilisation dans la littérature. Afin d'enrichir l'ensemble d'apprentissage, nous avons appliqué des techniques d'augmentation. La figure 3.2 illustre la distribution des classes du dataset APTOS 2019.

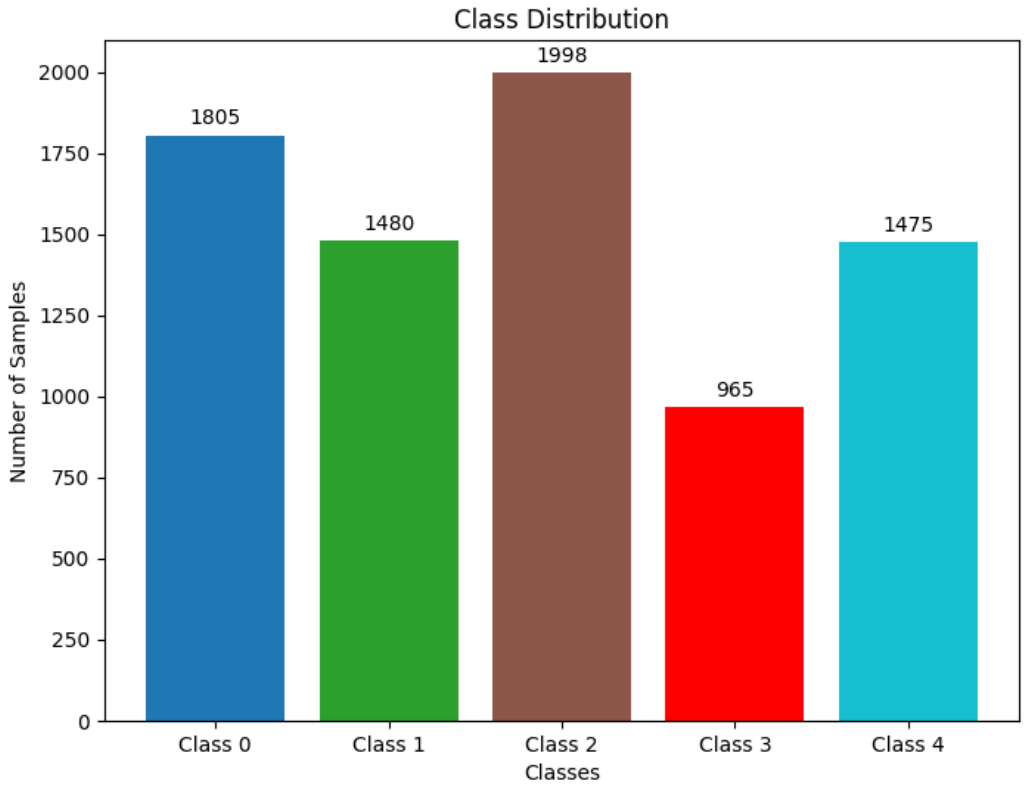


FIGURE 3.2 – Distribution des classes du dataset APTOS 2019.

II.1.2 Prétraitement des données

Les différentes étapes de prétraitement que nous avons réalisées sont :

Redimensionnement des images Toutes les images ont été redimensionnées à une taille fixe (224x224 pour ViT16 et 384x384 pour ViT32) afin d'uniformiser les données d'entrée et d'assurer un traitement optimal tout en réduisant le coût computationnel.

Grey Scale & Gaussian Blur Afin d'optimiser les performances du modèle d'apprentissage, il est essentiel que les images et leurs traits distinctifs soient aisément identifiables. Pour ce faire, la première technique appliquée consiste à convertir les images en niveaux de gris, suivie de l'application d'un filtre gaussien afin de réduire le bruit et de clarifier les caractéristiques[46].

Circle Crop Un recadrage circulaire a été appliqué afin d'éliminer l'arrière-plan noir indésirable et de concentrer le modèle uniquement sur la zone pertinente de l'image[46], à savoir le fond de l'œil(voire figure 3.3).

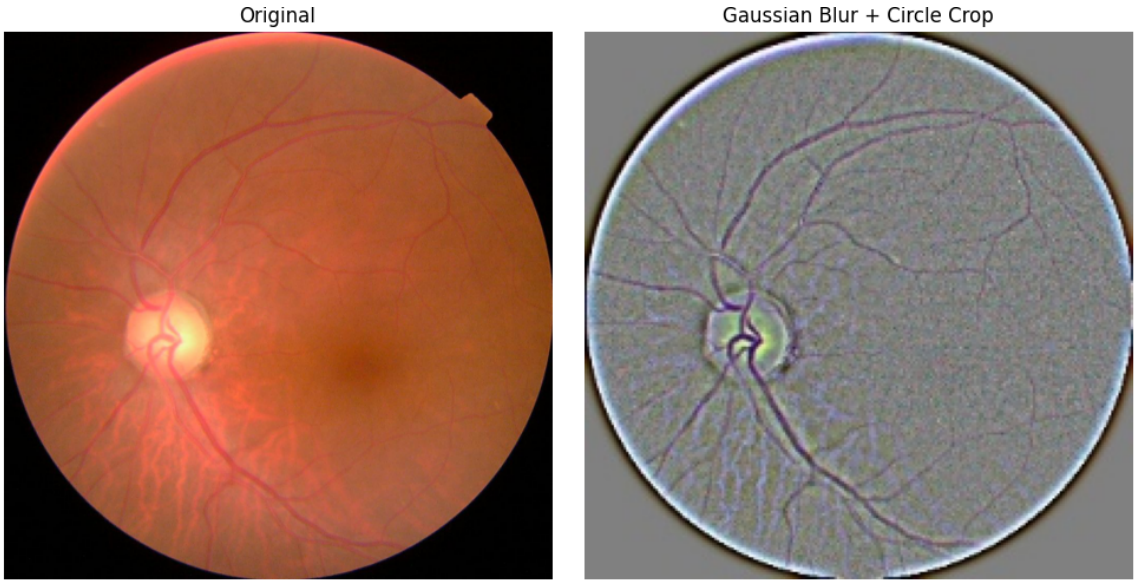


FIGURE 3.3 – gaussian blur et circle crop

Normalisation Les images ont été normalisées en transformant les valeurs de pixels de l'intervalle [0, 255] vers [0, 1]. Cette étape rend les données plus homogènes, ce qui facilite la convergence et améliore la stabilité ainsi que les performances du modèle(voire figure 3.4).

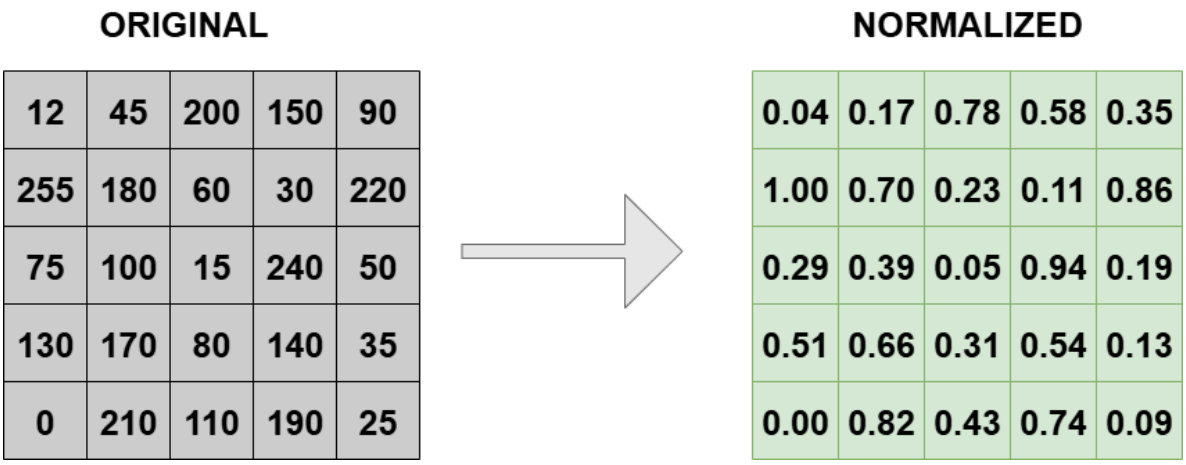


FIGURE 3.4 – normalization

II.1.3 Augmentation des données

Nous avons utilisé des techniques d'augmentation de données pour élargir la base de données et fournir des images supplémentaires des différentes stades de la RD(voire figure 3.5). Les differetransformations appliquées aux images sont :

Random Horizontal Flip Une symétrie horizontale aléatoire a été appliquée afin de diversifier les données d'entraînement et de rendre le modèle plus robuste aux variations d'orientation de l'image[47].

Random Rotation (±10°) Une rotation aléatoire limitée à ±10° a été utilisée pour introduire de légères variations angulaires et améliorer la capacité du modèle à généraliser sur des images prises sous différents angles[47].

Random Color Jitter (Brightness, Contrast, Saturation) Des ajustements aléatoires de la luminosité, du contraste et de la saturation ont été effectués pour simuler des variations d'éclairage et renforcer la robustesse du modèle face aux conditions d'acquisition différentes[47].

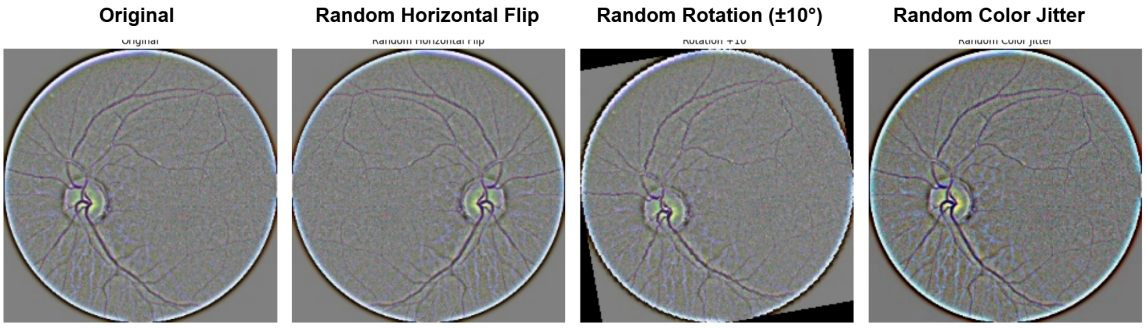


FIGURE 3.5 – augmenatations

II.2 Classification de la RD en cinq classes

Dans cette section, nous présentons les différents modèles utilisés pour la classification de la rétinopathie diabétique en cinq classes. Les différentes architectures mises en œuvre, ainsi que les résultats obtenus et leur analyse, seront exposés dans les sections suivantes.

II.2.1 Approche ViT avec une seule branche

Dans cette partie, nous allons proposer deux modèles, nommés ViT-SS et ViT-LS(ViT-SS basé sur ViT16 et ViT-LS basé sur ViT32), qui s'appuient sur une architecture Vision Transformer (ViT) pré-entraînée, adaptée ici à une tâche de classification multi-classes portant sur cinq niveaux de sévérité de la rétinopathie diabétique. Ces classes sont définies comme suit :

- Classe 0 : Pas de rétinopathie (No DR),
- Classe 1 : Rétinopathie légère,
- Classe 2 : Rétinopathie modérée,
- Classe 3 : Rétinopathie sévère,
- Classe 4 : Rétinopathie proliférante.

Le pipeline d'apprentissage suit l'enchainement suivant(voire figure 3.6) :

1. Un prétraitement des images est d'abord effectué , qui consiste à redimensionner chaque image à une résolution de 224×224 pixels pour le modèle ViT-SS et à 384×384 pixels pour le modèle ViT-LS. Les images ainsi obtenues sont ensuite découpées en patches non superposés de taille 16×16 pour ViT-SS et 32×32 pour ViT-LS.
2. Ces patches sont ensuite vectorisés à l'aide d'un embedding linéaire, formant une séquence de tokens.
3. Un token de classification spécial est ajouté à cette séquence, et un encodage de position est appliqué pour préserver l'information spatiale. La séquence ainsi obtenue est transmise à l'encodeur ViT, qui agit ici comme un extracteur de caractéristiques profondes à travers des mécanismes d'auto-attention multi-têtes.
4. La représentation générée par l'encodeur du ViT est ensuite transmise à un classifieur composé de couches entièrement connectées, dont la sortie est une distribution de probabilité sur les cinq classes cibles.

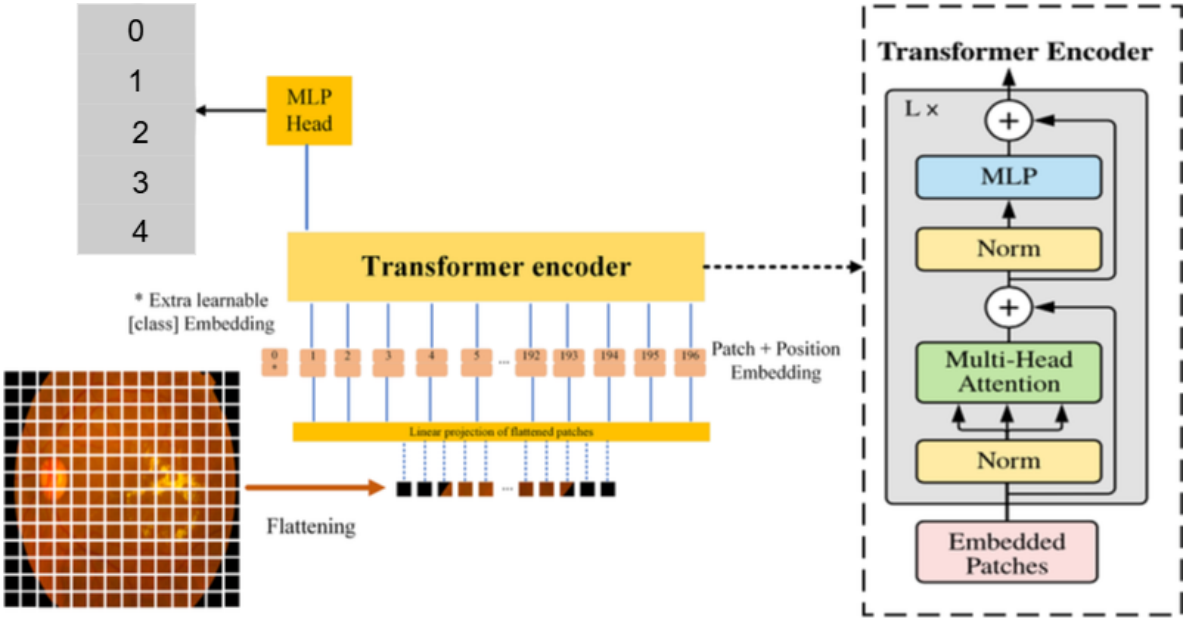


FIGURE 3.6 – VIT

Les principaux hyperparamètres des modèles ViT-SS et ViT-LS utilisés dans cette étude sont résumés dans le tableau 3.1 :

TABLE 3.1 – Hyperparamètres des modèles ViT-SS et ViT-LS

Hyperparamètre	**ViT-SS**	**ViT-LS**
Taille de l'image	224 × 224	384 × 384
Taille des patches	16 × 16	32 × 32
Longueur de séquence (tokens)	197 (196 + 1 [CLS])	145 (144 + 1 [CLS])
Dimension cachée (*hidden size*)	768	
Profondeur (couches)	12	
Nombre de têtes d'attention	12	
Taille intermédiaire du MLP	3072	
Taux de dropout (attn & MLP)	≈ 0,1	

— **Résultats** L'évaluation des performances a été réalisée à l'aide de métriques rigoureuses : accuracy, F1-score, sensibilité (recall) et spécificité, appliquées sur l'ensemble de validation. Une évaluation finale a également été menée sur un jeu de test indépendant, afin d'évaluer la capacité de généralisation du modèle après l'optimisation complète.

Afin d'optimiser l'entraînement, différents hyperparamètres ont été ajustés empiriquement à travers plusieurs essais. Les valeurs retenues se sont révélées être un compromis efficace entre stabilité de l'optimisation, rapidité de convergence et capacité de généralisation. En particulier, nous avons utilisé une taille de batch

de 32, un taux d'apprentissage initial de 2×10^{-5} avec un schéma de *warmup* sur 500 pas, ainsi qu'une pénalisation L2 (weight decay) de 0,01 afin de limiter le surapprentissage. L'arrêt prématuré avec une patience de 3 a permis de contrôler la durée de l'entraînement, aboutissant à 16 époques pour le modèle ViT-SS et 10 époques pour le modèle ViT-LS. Le tableau 3.2 résume l'ensemble de ces hyperparamètres.

TABLE 3.2 – Hyperparamètres d'entraînement des modèles ViT-SS, ViT-LS et ViT-DR

Hyperparamètre	**ViT-SS**	**ViT-LS**	**ViT-DR**
Nombre d'epochs	16	10	13
Taille du batch (train / eval)	32 / 32		
Taux d'apprentissage (*learning rate*)	2×10^{-5}		
Nombre de pas de *warmup*	500		
Pénalisation L2 (*weight decay*)	0,01		
Patience (arrêt prématuré)	3		
Validation	20% du dataset		
Test	50% de validation		

L'évaluation des performances des modèles ViT-SS et ViT-LS sur le jeu de données APTOS 2019 augmenté a été conduite à travers deux phases principales : une évaluation intermédiaire sur un ensemble de validation au cours de l'entraînement, suivie d'une évaluation finale sur un ensemble de test indépendant. Le dataset, composé de 7723 images, a été divisé en 80 % pour l'entraînement (6178 images), 10 % pour la validation (772 images) et 10 % pour le test (773 images). L'entraînement des modèles a été réalisé sur la plateforme Kaggle, en exploitant une carte graphique GPU de type P100, permettant d'accélérer le processus d'apprentissage et d'optimiser les performances.

1. **Modèle VIT-SS**

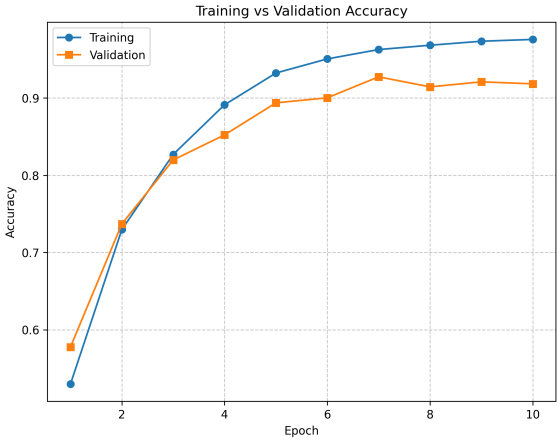
   Comme le montre le Tableau 3.3, le modèle ViT-SS obtient des performances élevées avec une accuracy de 0.9431 et un F1-Score de 0.9430, traduisant un excellent équilibre entre précision et rappel. Sa sensibilité (0.9438) et sa spécificité (0.9857) confirment sa capacité à identifier correctement les cas positifs tout en réduisant les faux positifs. Ces résultats démontrent que le modèle ViT-SS constitue une solution fiable pour la détection de la rétinopathie diabétique.

TABLE 3.3 – Performances globales du modèle VIT-SS

Modèle	Accuracy	F1 Score	Sensitivity	Specificity
VIT-SS	0.9431	0.9430	0.9438	0.9857

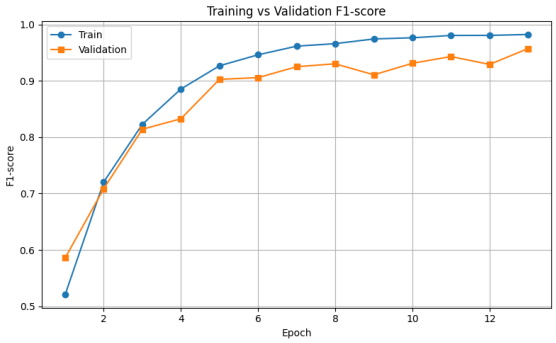
La Figure 3.7 montre que les courbes de perte et d'accuracy se stabilisent après environ 16 époques, traduisant une bonne convergence sans signe de surapprentissage. De plus, la Figure 3.8 illustre une évolution régulière du F1-Score et de la sensibilité. Ces résultats confirment que le modèle ViT-SS apprend de manière cohérente et fiable tout au long de l'entraînement

(a) Perte entraînement/validation – ViT-SS

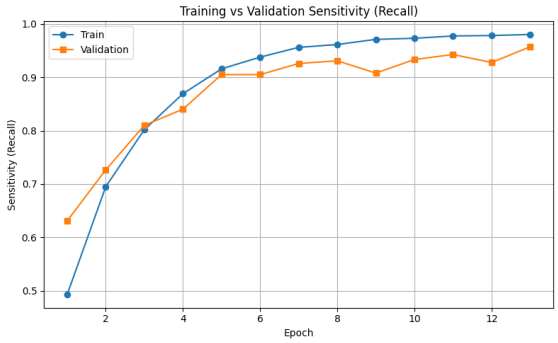


(b) Accuracy entraînement/validation – ViT-SS

FIGURE 3.7 – Perte & Accuracy – ViT-SS



(a) F1-score entraînement/validation – ViT-SS



(b) Sensibilité (Recall) entraînement/validation – ViT-SS

FIGURE 3.8 – F1-score & sensibilité du ViT-SS

TABLE 3.4 – Les performances par classe du model VIT-SS

Class	Precision	Recall (Sens)	Specificity	F1 Score
0	0.9836	0.9945	0.9949	0.9890
1	0.9595	0.9595	0.9904	0.9595
2	0.9196	0.9150	0.9721	0.9173
3	0.8762	0.9583	0.9808	0.9154
4	0.9565	0.8919	0.9904	0.9231

L'évaluation des performances par classe illustré dans le tableau 3.4 pour le modèle VIT-SS montre que la classe 0 obtient les meilleurs résultats, avec une sensibilité de 99,45% et un F1-score de 98,90%, traduisant une excellente capacité à identifier les patients sains. Les classes 1 et 2 présentent également de bonnes performances (F1-scores de 95,95% et 91,73%), bien qu'une légère baisse de sensibilité soit observée pour la classe 2, indiquant quelques confusions avec des stades proches. Pour les classes 3 et 4, les sensibilités élevées (95,83% et 89,19%) suggèrent que le modèle est capable de repérer efficacement les cas graves. Cette capacité à détecter la majorité des cas pathologiques avancés, même au prix de quelques faux positifs, est particulièrement pertinente en pratique clinique, car elle permettrait un triage automatisé fiable et une orientation rapide des patients nécessitant une prise en charge spécialisée, contribuant ainsi à réduire les risques de complications visuelles irréversibles.

La matrice de confusion illustrée dans figure 3.9 révèle que la majorité des erreurs concernent des confusions entre classes adjacentes, en particulier autour des classes 2, 3 et 4. La classe 2 est confondue avec les stades voisins (classes 1 et 3), tandis que les classes 4 et 2 présentent des erreurs notables. Ces confusions suggèrent que le modèle apprend bien la progression des stades, mais peine à distinguer les cas limites, notamment pour les

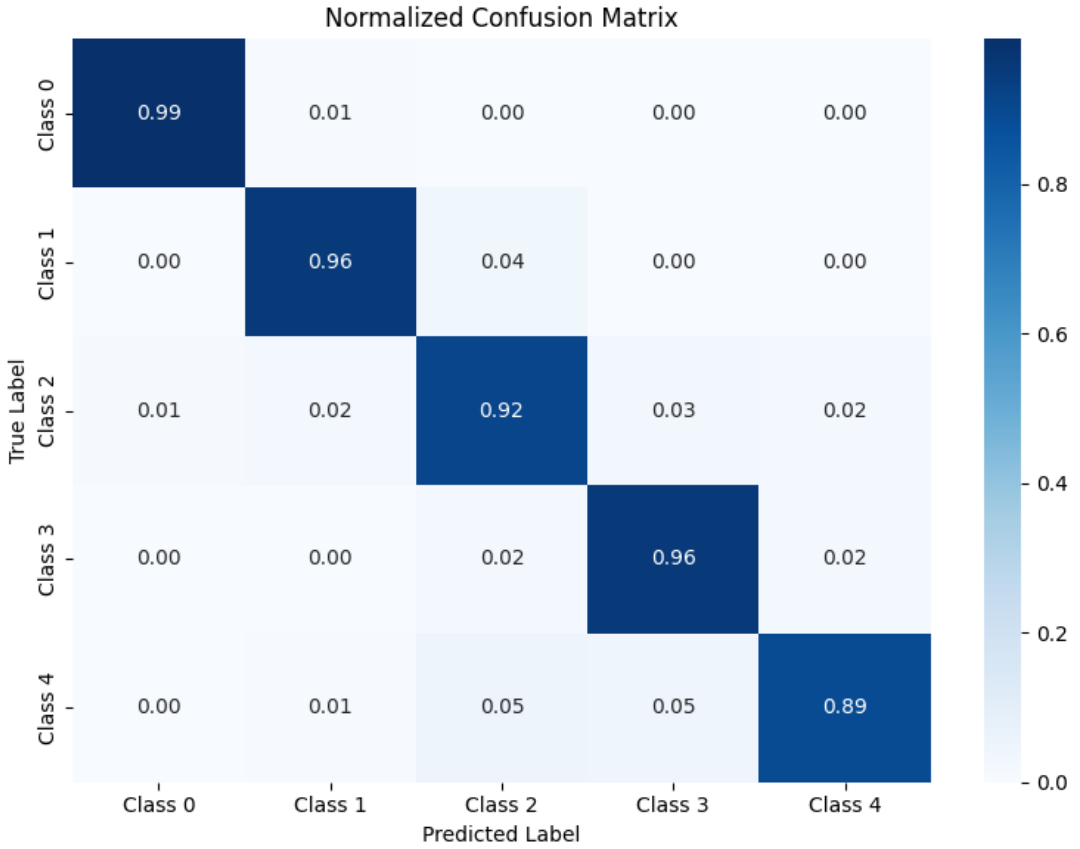


FIGURE 3.9 – Matrice de confusion – ViT-SS

formes avancées. Dans un contexte réel, cela signifie que la majorité des cas pathologiques sont bien détectés, même si certains stades sont légèrement sous-estimés, ce qui reste acceptable dans une logique de dépistage, car cela permettrait de diriger les patients atteints vers un examen spécialisé sans négliger les cas graves.

Les résultats obtenus confirment la capacité du modèle à discriminer efficacement les différents stades de la rétinopathie diabétique, en particulier entre les cas sains ou légers et les formes pathologiques avancées, une distinction essentielle dans une optique de triage automatisé. Toutefois, les erreurs localisées entre les classes intermédiaires et sévères suggèrent des marges d'amélioration.

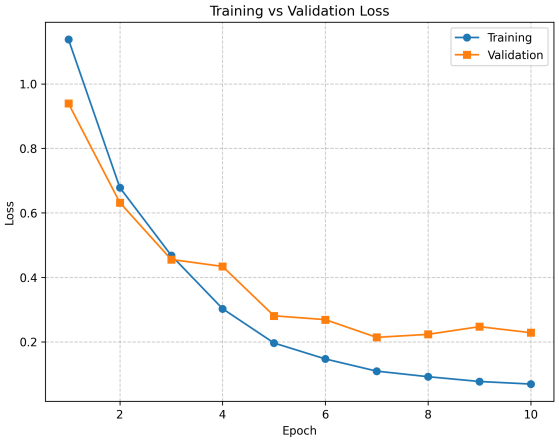
2. **Modèle VIT-LS**

Le Tableau 3.5 montre que le modèle ViT-LS atteint une accuracy de 0.9004 et un F1-Score de 0.8996, traduisant des performances globalement stables. Sa sensibilité (0.8959) reflète sa capacité à détecter correctement la majorité des cas positifs. Ces résultats indiquent que ViT-LS peut être utilisé efficacement pour la classification de la rétinopathie diabétique.

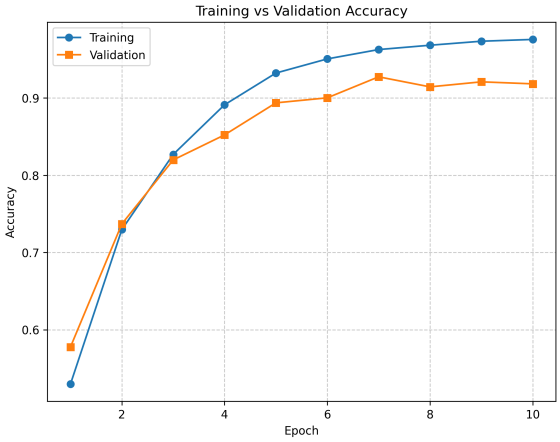
TABLE 3.5 – Performances globales du modèle VIT-LS

Modèle	Accuracy	F1 Score	Sensitivity	Specificity
VIT-LS	0.9004	0.8996	0.8959	0.9750

La Figure 3.10 montre que les courbes de perte et d'accuracy se stabilisent dès la 10 époque, traduisant une convergence correcte du modèle. La Figure 3.11 illustre l'évolution du F1-Score et de la sensibilité.Ces résultats indiquent que le modèle ViT-LS parvient à apprendre de manière régulière sur l'ensemble des données.

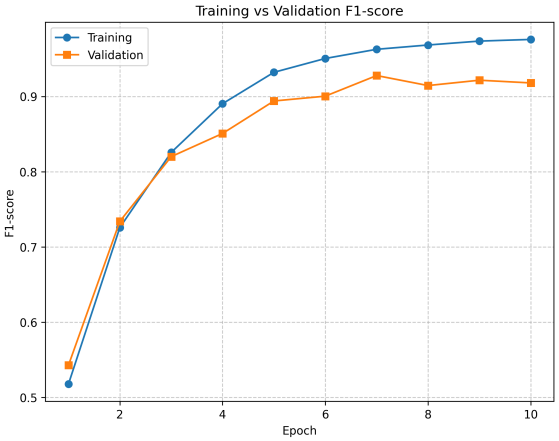


(a) Perte entraînement/validation – ViT-LS

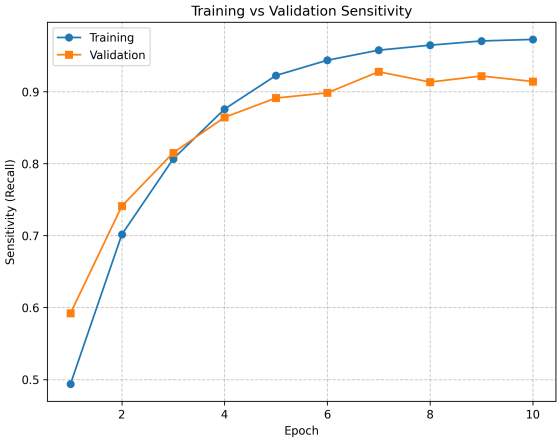


(b) Accuracy entraînement/validation – ViT-LS

FIGURE 3.10 – Perte & Accuracy – ViT-LS



(a) F1-score entraînement/validation – ViT-LS



(b) Sensibilité (Recall) entraînement/validation – ViT-LS

FIGURE 3.11 – F1-score & Sensibilité – ViT-LS

L'analyse des résultats du tableau3.6 met en évidence la robustesse globale du modèle ViT-LS sur l'ensemble des classes. La classe 0 se distingue par des

TABLE 3.6 – Les performances par classe du model VIT-LS

Class	Precision	Recall (Sens)	Specificity	F1 Score
0	0.9728	0.9890	0.9916	0.9808
1	0.8758	0.9527	0.9680	0.9126
2	0.8925	0.8300	0.9651	0.8601
3	0.8587	0.8229	0.9808	0.8404
4	0.8733	0.8851	0.9696	0.8792

performances quasi parfaites (sensibilité de 98,90 % et F1-score de 98,08%), confirmant la capacité du modèle à reconnaître de manière fiable les patients sains et à minimiser les faux positifs. Du côté des classes pathologiques, les résultats sont également solides : la classe 1 atteint un rappel élevé (95,27%), ce qui garantit que la majorité des cas précoces sont correctement détectés, tandis que la classe 2 conserve un bon équilibre entre précision (89,25%) et spécificité (96,51%), malgré une sensibilité légèrement inférieure (83,00%). Les classes 3 et 4, correspondant aux stades sévères et proliférants, affichent des performances encourageantes avec des F1-scores de 84,04% et 87,92% respectivement, démontrant la capacité du modèle à identifier efficacement les cas graves, ce qui est crucial en pratique clinique. Ces résultats montrent que le modèle ne se contente pas de bien détecter l'absence de rétinopathie, mais qu'il assure également une détection pertinente des cas positifs, renforçant son potentiel en tant qu'outil de dépistage fiable pour orienter rapidement les patients nécessitant une prise en charge spécialisée.

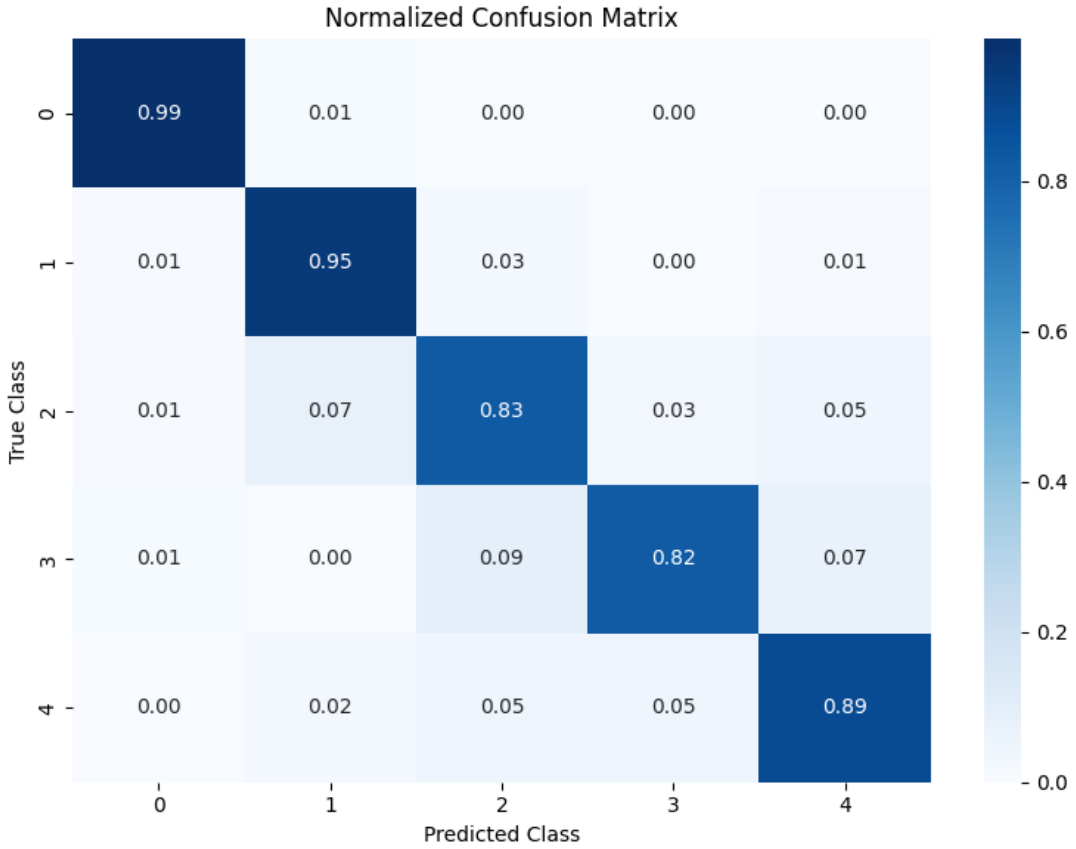


FIGURE 3.12 – Matrice de confusion – ViT-LS

La matrice de confusion de la Figure3.12 montre que le modèle ViT-LS atteint une excellente capacité de distinction pour les classes extrêmes. La classe 0 est reconnue presque parfaitement (99 %), avec très peu de confusions vers d'autres catégories, ce qui garantit une détection fiable des cas sains ou précoces. Du côté opposé, la classe 4 affiche également des performances solides (89%), confirmant la capacité du modèle à identifier les cas les plus avancés de rétinopathie diabétique.

Les erreurs observées concernent principalement les classes intermédiaires (1, 2 et 3), qui présentent des recouvrements visuels plus importants. La classe 2, par exemple, est confondue avec les classes 1 et 3, tandis que la classe 3 est parfois assimilée à la classe 2. Ces confusions suivent une logique de progression de la maladie, traduisant une cohérence dans l'apprentissage du modèle.

En pratique clinique, cette tendance est rassurante : même si le modèle hésite sur la sévérité exacte, il parvient à distinguer de manière fiable les cas pathologiques des cas non pathologiques, ce qui constitue un atout majeur pour le dépistage et le diagnostic automatisé.

— **conclusion**

TABLE 3.7 – Comparaison des performances globales des modèles ViT-SS et ViT-LS

Modèle	Precision	Accuracy	F1 Score	Sensitivity	Specificity
VIT-SS	**0.9391**	**0.9431**	**0.9430**	**0.9438**	**0.9857**
VIT-LS	0.89462	0.9004	0.8996	0.8959	0.9750

L'analyse du tableau 3.7 met en évidence la supériorité de ViT-SS, qui obtient des performances plus élevées sur l'ensemble des métriques. En termes de précision (0.9391 contre 0.8946) et de taux de classification correcte (Accuracy : 0.9431 contre 0.9004), ViT-SS montre une meilleure fiabilité dans la prédiction. Le F1-Score (0.9430 contre 0.8996) et la sensibilité (0.9438 contre 0.8959) confirment cette tendance, indiquant que ViT-SS identifie plus efficacement les cas positifs tout en maintenant un bon équilibre entre rappel et précision. Enfin, la spécificité reste légèrement plus élevée pour ViT-SS (0.9857 contre 0.9750). Dans l'ensemble, ces résultats mettent clairement en évidence la robustesse et la fiabilité de ViT-SS, qui se distingue comme le modèle le plus performant pour la classification de la rétinopathie diabétique par rapport à ViT-LS.

II.2.2 Approche ViT avec deux branches

En complément des modèles ViT-SS et ViT-LS, nous proposons un troisième modèle hybride, noté ViT-DR. Ce modèle exploite simultanément deux encodeurs Vision Transformer fonctionnant en parallèle : le premier traite des patches de taille 16×16 (comme dans ViT-SS), tandis que le second opère sur des patches de 32×32 (comme dans ViT-LS). Chaque encodeur produit un vecteur de représentation global correspondant au *class token*. Les deux vecteurs obtenus sont ensuite concaténés afin de constituer une représentation unifiée, dont la dimension est égale à la somme des tailles des deux *class tokens*. Cette représentation fusionnée est enfin transmise à un classifieur MLP pour effectuer la prédiction finale(voire la figure 3.13).

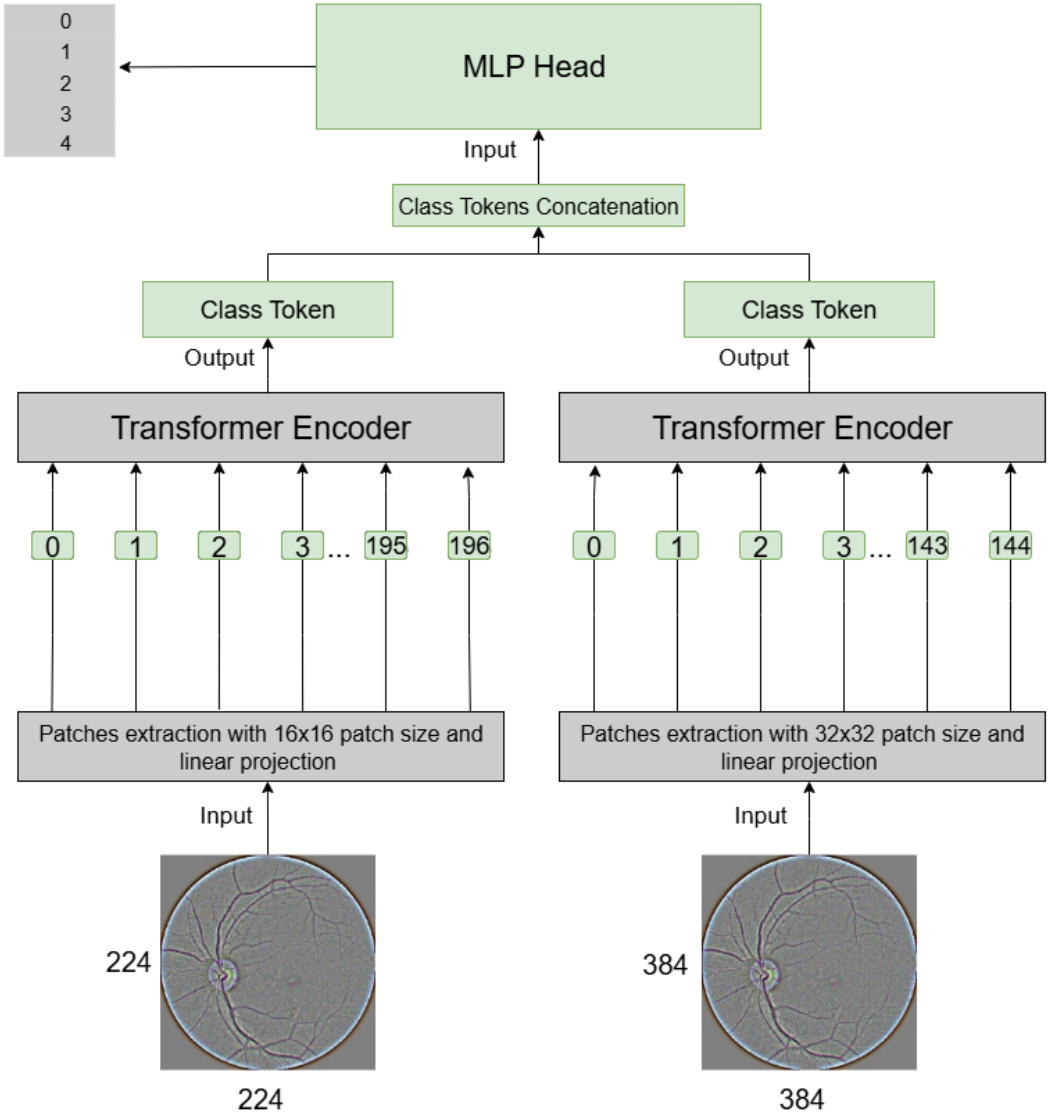


FIGURE 3.13 – Architecture de VIT-DR

Résultats L'évaluation des performances du modèle ViT-DR a été conduite à travers deux phases principales : une évaluation intermédiaire sur un ensemble de validation au cours de l'entraînement, suivie d'une évaluation finale sur un ensemble de test indépendant. Le dataset APTOS 2019 augmenté, composé de 7723 images, a été divisé en 6178 images pour l'entraînement (80 %), 772 images pour la validation (10 %) et 773 images pour le test (10 %). L'entraînement a été effectué sur la plateforme Kaggle, en exploitant une carte graphique GPU de type P100, ce qui a permis d'accélérer le processus d'apprentissage. Le but de cette démarche est de mesurer de manière rigoureuse la capacité du modèle ViT-DR à effectuer une classification multi-classes, correspondant

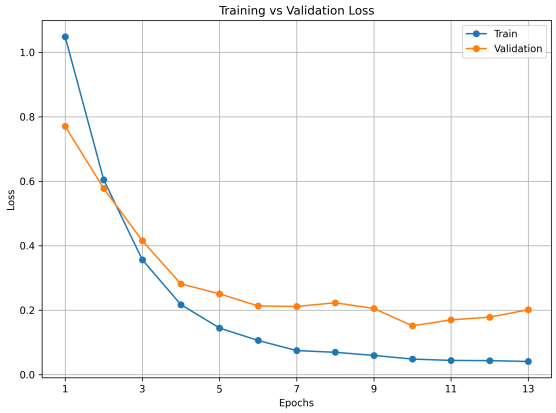
aux différents degrés de sévérité de la rétinopathie diabétique

L'analyse du Tableau 3.8 montre que le modèle ViT-DR affiche des performances solides avec une accuracy de 0.9418 et une sensitivity de 0.9464, ce qui montre sa capacité à détecter efficacement la rétinopathie diabétique. L'équilibre entre les scores F1 et accuracy (0.94) confirme la robustesse du modèle ViT-DR. Ces résultats soulignent que ViT-DR constitue une solution fiable pour l'identification automatique de la RD.

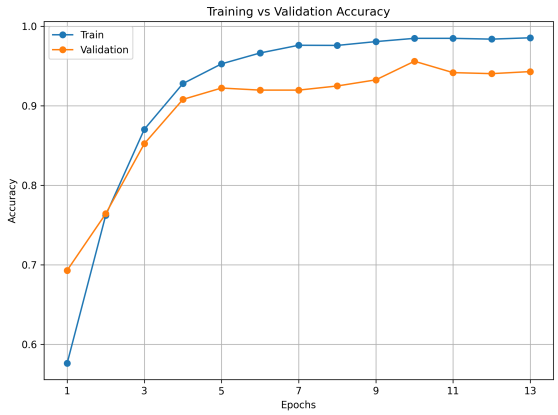
TABLE 3.8 – Performances globales du modèle ViT-DR

Modèle	**Precision**	**Accuracy**	**F1 Score**	**Sensitivity**	**Specificity**
VIT-DR	0.9360	0.9418	0.9414	0.9464	0.9856

La Figure 3.14 montre que les courbes de perte et d'accuracy atteignent une stabilité vers la 13 époque, confirmant une bonne convergence du modèle. La Figure 3.15 illustre l'évolution du F1-Score et de la sensibilité, qui se stabilisent également à ce stade. Ces résultats suggèrent que le modèle ViT-DR est capable d'apprendre efficacement et de détecter la rétinopathie diabétique de manière fiable.

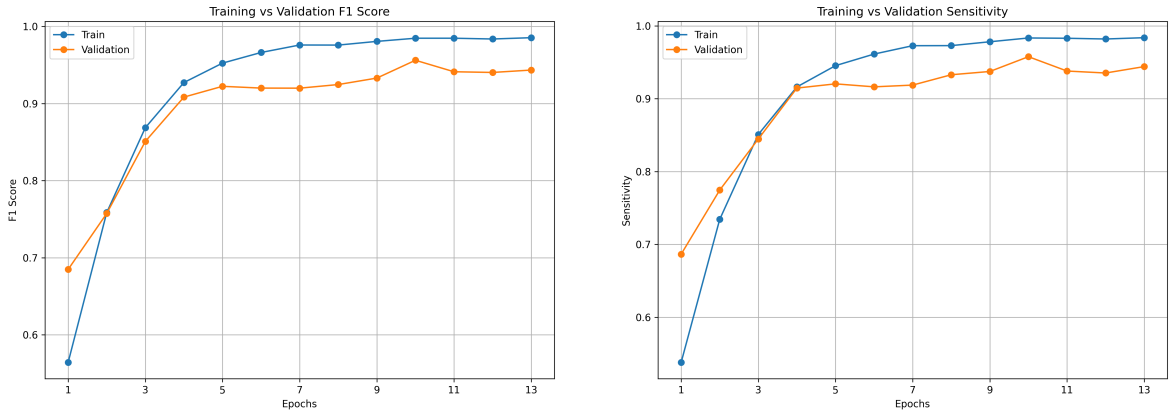


(a) Perte entraînement/validation – ViT-DR



(b) Accuracy entraînement/validation – ViT-DR

FIGURE 3.14 – Perte & Accuracy – ViT-DR



(a) F1-score entraînement/validation – ViT-DR

(b) Sensibilité (Recall) entraînement/validation – ViT-DR

FIGURE 3.15 – F1-score & Sensibilité – ViT-DR

L'analyse du tableau 3.9 met en évidence la solidité globale du modèle sur l'ensemble des classes. La classe 0 se distingue par des performances quasi parfaites (sensibilité de 99,45% et F1-score de 98,90%), traduisant une excellente capacité à identifier les patients sains tout en minimisant les faux positifs.

Pour les classes pathologiques, les résultats restent robustes. La classe 1 atteint un rappel élevé (97,30%), garantissant que la majorité des cas légers sont correctement détectés. La classe 2, bien que présentant une sensibilité plus faible (87,00 %), conserve une précision élevée (95,60 %), ce qui limite les fausses alertes et traduit un bon équilibre global.

Les classes 3 et 4, correspondant aux stades sévères et proliférants, affichent des F1-scores solides (92,54 % et 92,57 %), avec des sensibilités proches de 97 % pour la classe 3 et 93 % pour la classe 4. Ces résultats sont particulièrement encourageants en pratique clinique, car ils démontrent la capacité du modèle à repérer efficacement les cas graves.

TABLE 3.9 – Performance par classe du modèle VIT-DR

Classe	**Precision**	**Recall (Sens)**	**Specificity**	**F1 Score**
0	0.9836	0.9945	0.9949	0.9890
1	0.9290	0.9730	0.9824	0.9505
2	0.9560	0.8700	0.9860	0.9110
3	0.8857	0.9688	0.9823	0.9254
4	0.9257	0.9257	0.9824	0.9257

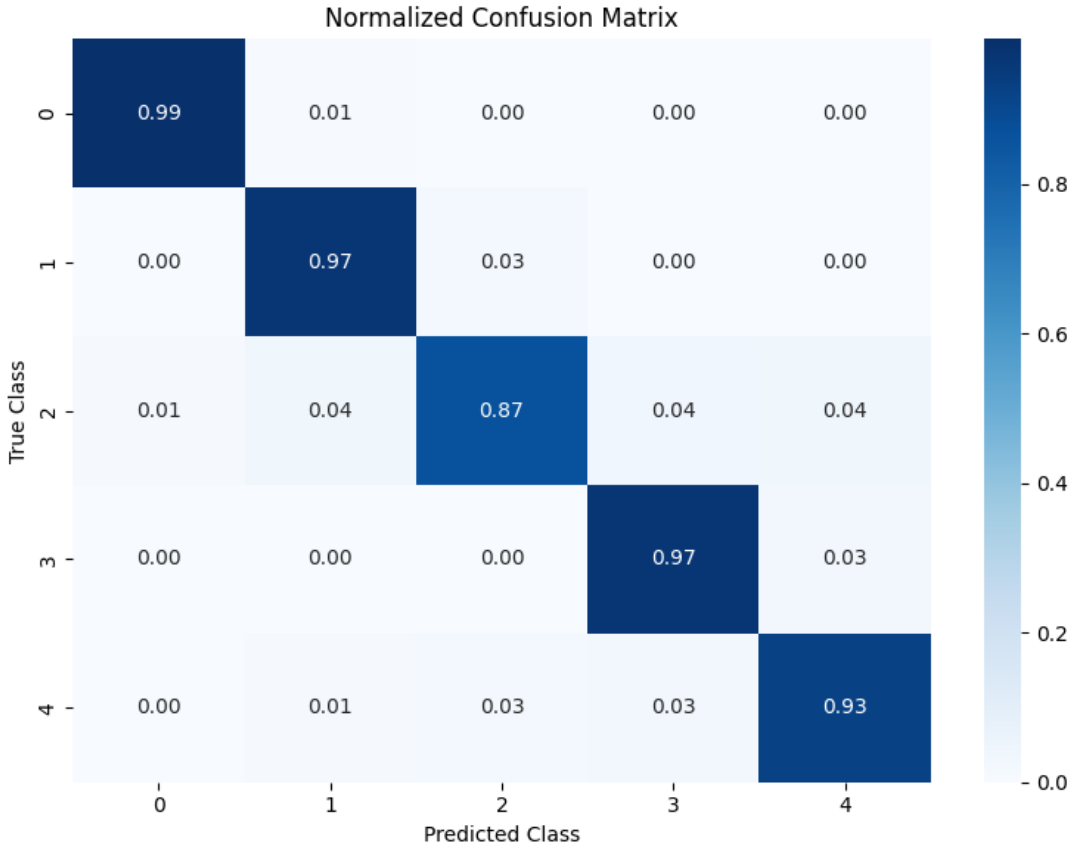


FIGURE 3.16 – Matrice de confusion – ViT-DR

La matrice de confusion de la Figure3.16 montre que le modèle ViT-DR atteint une excellente capacité de distinction pour toutes les classes. La classe 0 est reconnue presque parfaitement (99%), traduisant une très grande fiabilité dans la détection des cas sains. Nous pouvons constater que la classe 1 est detecter à 97%, ceci indique que notre modèle arrive à detecter la RD à ses premiers stades. De même, la classe 3(97%) et la classe 4 (93%) affichent de très bons taux de reconnaissance, confirmant que le modèle parvient à identifier efficacement les stades avancés de la rétinopathie diabétique.

Les erreurs observées concernent principalement la classe 2, qui présente un taux de reconnaissance légèrement plus faible (87 %) et est confondue à la fois avec les classes 1, 3 et 4. Cette tendance indique que le modèle a parfois des difficultés à tracer des frontières claires entre les formes modérées et sévères, ce qui est cohérent avec la continuité clinique de la maladie. En pratique clinique, cette performance reste encourageante.

II.3 Comparaison des modèles

L'analyse comparative des trois modèles (voir tableau3.10) met en évidence que le modéle VIT-SS et le modéle VIT-DR offrent des performances très proches, avec des valeurs d'accuracy (0,9431 et 0,9418) et de F1-score (0,9430 et 0,9414) quasi équivalentes. Le modéle VIT-DR se distingue légèrement par une meilleure sensibilité (0,9464 contre 0,9438), ce qui signifie qu'il détecte un peu plus efficacement les cas positifs, un critère crucial en contexte médical. En revanche, le modéle VIT-SS conserve un léger avantage en termes de spécificité (0,9857 contre 0,9856), traduisant une meilleure capacité à limiter les faux positifs.

Le modéle VIT-LS, affiche des résultats inférieurs sur l'ensemble des métriques (accuracy de 0,9004 et F1-score de 0,8996), ce qui le rend moins compétitif par rapport aux deux autres. Ainsi, bien que les différences entre modéle VIT-SS et VIT-DR soient marginales, on peut considérer que le modéle VIT-DR est le plus équilibré grâce à sa sensibilité supérieure, tandis que le modèle VIT-SS reste plus robuste du point de vue de la spécificité.

En conclusion, pour une utilisation clinique où la détection des cas positifs est prioritaire afin d'éviter les faux négatifs, le modéle VIT-DR apparaît comme le choix le plus pertinent. Toutefois, si l'objectif est de minimiser les fausses alertes dans un cadre de dépistage à grande échelle, le modèle VIT-SS pourrait être privilégié.

TABLE 3.10 – Comparaison des performances globales des trois modèles

Modèle	Precision	Accuracy	F1 Score	Sensitivity	Specificity
VIT-SS	**0.9391**	**0.9431**	**0.9430**	0.9438	**0.9857**
VIT-LS	0.89462	0.9004	0.8996	0.8959	0.9750
VIT-DR	0.9360	0.9418	0.9414	**0.9464**	0.9856

II.4 comparaison avec les autre modéles

Les résultats comparatifs présentés dans le Tableau 3.11 montrent que les trois modèles proposés, basés uniquement sur les Vision Transformers (ViT-SS, ViT-LS et ViT-DR), surpassent la majorité des modèles de l'état de l'art. En particulier, ViT-SS et ViT-DR obtiennent des valeurs élevées d'accuracy (0.9431 et 0.9418 respectivement) et d'AUC (>0.99), traduisant une capacité robuste de classification de la rétinopathie diabétique. Ces scores dépassent largement ceux des modèles [42] ,[40] et [44], et se situent à un niveau comparable au modèle [43].

Cependant, le modèle [45] affiche les meilleurs résultats avec une accuracy de 0.9693 et un F1-score de 0.973, supérieurs à tous les modèles, y compris ViT-SS et ViT-DR. Cette supériorité peut s'expliquer par la combinaison synergique des architectures ViT et CNN. En effet, les CNN sont particulièrement efficaces pour extraire des caractéristiques locales et hiérarchiques, tandis que les ViT capturent les dépendances globales et contextuelles. L'hybridation des deux approches permet donc de bénéficier simultanément d'une représentation fine et globale des images, ce qui améliore significativement les performances.

Le choix des modèles de comparaison repose sur leur représentativité dans la littérature récente. En particulier, [42], [40] et [44] constituent des références majeures utilisées pour la classification de la rétinopathie diabétique, tandis que [43] propose un modèle compétitif basé sur les ViT. Enfin, [45] illustre l'état de l'art le plus avancé grâce à l'hybridation ViT–CNN. Cette sélection assure ainsi une comparaison équilibrée et pertinente avec nos modèles proposés.

TABLE 3.11 – Comparaison des performances des différents modèles

Modèle	**Accuracy**	**F1 Score**	**Sensitivity**	**Specificity**	**Kappa**	**AUC**
[42]	0.8412	0.8400	0.8154	0.9413	–	–
[40]	0.8235	–	0.8140	0.8245	–	0.9018
[43]	0.9342	–	0.9662	0.9539	–	0.9825
[44]	0.8790	0.8910	0.8160	0.8490	0.8900	–
[45]	**0.9693**	**0.973**	**0.9889**	–	–	–
VIT-SS	0.9431	0.9430	0.9438	**0.9857**	**0.9280**	0.9939
VIT-LS	0.9004	0.8996	0.8959	0.9750	0.8739	0.9868
VIT-DR	0.9418	0.9414	0.9464	0.9856	0.9265	**0.9940**

III Conclusion

Dans ce chapitre, nous avons présenté et évalué trois modèles de classification de la rétinopathie diabétique en cinq classes : VIT-SS, VIT-LS et VIT-DR. L'analyse des performances a mis en évidence que les modèles VIT-SS et VIT-DR surpassent clairement VIT-LS, et se distinguent par leurs résultats compétitifs face aux autre travaux présentés dans ce travail. Ces deux modèles démontrent la capacité des architectures Vision Transformer à capturer efficacement les caractéristiques visuelles complexes liées à la rétinopathie diabétique, confirmant leur pertinence pour des tâches de classification médicale.

Au-delà des performances obtenues, cette étude met en avant l'importance de sélectionner soigneusement l'architecture et les paramètres d'entraînement afin d'atteindre une performance stable et fiable. Nos résultats indiquent également que, malgré les avancées, il subsiste des marges d'amélioration qui pourraient renforcer la robustesse et la fiabilité des modèles. Ainsi, ce chapitre confirme la pertinence des modèles Vision Transformer, en particulier VIT-SS et VIT-DR, pour la classification de la rétinopathie diabétique, et ouvre la voie à des développements futurs visant à perfectionner encore davantage leur efficacité et leur applicabilité en pratique clinique.

Conclusion Générale

La rétinopathie diabétique représente aujourd'hui l'une des principales causes de cécité évitable à travers le monde. Face à sa prévalence croissante, le besoin de méthodes de dépistage et de classification automatiques, à la fois fiables et rapides, devient une priorité de santé publique. Les méthodes traditionnelles de diagnostic, basées sur l'examen manuel des images rétiniennes, restent chronophages et exposées à une variabilité inter-observateurs, d'où l'intérêt de l'intégration de modèles d'apprentissage profond dans ce domaine.

Dans ce travail, nous avons exploré le potentiel des Vision Transformers (ViT) pour la classification de la rétinopathie diabétique. Contrairement aux approches classiques reposant uniquement sur les réseaux de neurones convolutifs (CNN), les ViT se distinguent par leur capacité à modéliser efficacement les relations globales et contextuelles présentes dans les images médicales. Trois modèles basés sur cette architecture ont été proposés : ViT-SS, ViT-LS et ViT-DR. Les résultats expérimentaux obtenus démontrent la pertinence de cette approche, avec des performances élevées en termes d'accuracy, de F1-score et d'AUC, positionnant nos modèles parmi les plus compétitifs de l'état de l'art.

Enfin,ce travail ouvre plusieurs perspectives de recherche. L'exploration d'architectures hybrides combinant CNN et Vision Transformers constitue une piste prometteuse, permettant de tirer parti à la fois de l'extraction locale fine et de la modélisation contextuelle globale. De plus, l'élargissement des bases de données et l'intégration de stratégies de prétraitement avancées représentent des leviers essentiels pour accroître la robustesse et la généralisabilité des modèles. Une optimisation plus fine des hyperparamètres, associée à des mécanismes d'attention plus sophistiqués, pourrait également améliorer la détection des structures rétiniennes complexes. Parallèlement, l'adoption de techniques d'augmentation de données diversifiées, telles que la génération d'images synthétiques ou l'apprentissage auto-supervisé, permettrait de pallier les limites liées à la disponibilité des données médicales. À plus long terme, l'intégration de ces approches dans des systèmes de diagnostic assisté offrirait une contribution déterminante au dépistage précoce et à l'amélioration de la prise en charge clinique de la rétinopathie diabétique

Bibliographie

[1] American Academy of Ophthalmology (AAO). *Diabetic Retinopathy Preferred Practice Pattern® Guidelines*. Available from : https://www.aao.org/Assets/811c9cb7-279d-4b3d-9cca-032191e4891c/638749627918470000/diabetic-retinopathy-ppp-pdf

[2] Rodríguez, Amorim. The role of early detection in preventing vision loss from diabetic retinopathy. *Journal of Diabetic Complications Medicine*, 9 :290, 2024. Available from : https://www.hilarispublisher.com/open-access/the-role-of-early-detection-in-preventing-vision-loss-from-diabetic-retinopathy.pdf

[3] Meng, Y., Liu, Y., Duan, R., Liu, B., Lin, Z., Ma, Y., Jiang, L., Qin, Z., & Li, T. Global, Regional, and National Epidemiology of Vision Impairment due to Diabetic Retinopathy Among Working-Age Population, 1990–2021. *Journal of Diabetes*, 17(7) :e70121, 2025. doi :10.1111/1753-0407.70121. PMID : 40660082 ; PMCID : PMC12259346. Available from : https://pmc.ncbi.nlm.nih.gov/articles/PMC12259346/

[4] Piyasena, M. M. P. N., Murthy, G. V. S., Yip, J. L. Y., Gilbert, C., Zuurmond, M., Peto, T., Gordon, I., Hewage, S., & Kamalakannan, S. Systematic review on barriers and enablers for access to diabetic retinopathy screening services in different income settings. *PLoS One*, 14(4) :e0198979, 2019. doi :10.1371/journal.pone.0198979. PMID : 31013274 ; PMCID : PMC6478270. Available from : https://pmc.ncbi.nlm.nih.gov/articles/PMC6478270/

[5] Ansari, P. ; Tabasumma, N. ; Snigdha, N. N. ; Siam, N. H. ; Panduru, R. V. N. R. S. ; Azam, S. ; Hannan, J. M. A. ; Abdel-Wahab, Y. H. A. Diabetic Retinopathy : An Overview on Mechanisms, Pathophysiology and Pharmacotherapy. *Diabetology*, 2022, **3**, 159–175. https ://doi.org/10.3390/diabetology3010011. Available from : https://www.mdpi.com/2673-4540/3/1/11

[6] Wang, W. ; Lo, A. C. Y. Diabetic Retinopathy : Pathophysiology and Treatments. *International Journal of Molecular Sciences*, 2018, **19**(6), 1816. doi :10.3390/ijms19061816. PMID : 29925789 ; PMCID : PMC6032159. Available from : https://www.mdpi.com/1422-0067/19/6/1816 Available from : https://pmc.ncbi.nlm.nih.gov/articles/PMC6032159/

[7] Snyder Eye Care. *The 4 Stages of Diabetic Retinopathy.* Consulté le 31 août 2025. Disponible à l'adresse : https://www.drsnyder.org/eye-care-services/eye-disease-management/diabetic-retinopathy/the-4-stages-of-diabetic-retinopathy

[8] Sinclair, S. H. ; Schwartz, S. S. Diabetic Retinopathy–An Underdiagnosed and Undertreated Inflammatory, Neuro-Vascular Complication of Diabetes. *Frontiers in Endocrinology*, 2019, **10** :843. doi :10.3389/fendo.2019.00843. Available from : https://www.frontiersin.org/articles/10.3389/fendo.2019.00843/full

[9] American Diabetes Association. Standards of Medical Care in Diabetes—2022 Abridged for Primary Care Providers. *Clinical Diabetes*, 2022, **40**(1), 10–38. doi :10.2337/cd22-as01. Available from : https://diabetesjournals.org/clinical/article/40/1/10/138916/Standards-of-Medical-Care-in-Diabetes-2022

[10] Czupryniak, L. ; Barkai, L. ; Bolgarska, S. ; Bronisz, A. ; Broz, J. ; Cypryk, K. ; Honka, M. ; Janez, A. ; Krnic, M. ; Lalic, N. ; Martinka, E. ; Rahelic, D. ; Roman, G. ; Tankova, T. ; Várkonyi, T. ; Wolnik, B. ; Zherdova, N. Self-monitoring of blood glucose in diabetes : from evidence to clinical reality in Central and Eastern Europe–recommendations from the international Central-Eastern European expert group. *Diabetes Technology & Therapeutics*, 2014, **16**(7), 460–475. doi :10.1089/dia.2013.0302. PMID :24716890 ; PMCID :PMC4074758. Available from : https://pmc.ncbi.nlm.nih.gov/articles/PMC4074758/ Available from : https://www.liebertpub.com/doi/10.1089/dia.2013.0302

[11] Aronow, W. S. ; Shamliyan, T. A. Blood pressure targets for hypertension in patients with type 2 diabetes. *Annals of Translational Medicine*, 2018 Jun ; **6**(11) :199. doi :10.21037/atm.2018.04.36. PMID :30023362 ; PMCID :PMC6035980. Available from : https://pmc.ncbi.nlm.nih.gov/articles/PMC6035980/

[12] International Diabetes Federation. *Recommendations For Managing Type 2 Diabetes In Primary Care*, 2017. Available from : https://idf.org/media/uploads/2023/05/attachments-63.pdf

[13] Flaxel, C. J. ; Adelman, R. A. ; Bailey, S. T. ; Fawzi, A. ; Lim, J. I. ; Vemulakonda, G. A. ; Ying, G. Diabetic Retinopathy Preferred Practice Pattern®. *Ophthalmology*, 2020, **127**(1), 66–145. doi :10.1016/j.ophtha.2019.09.025. Available from : https://www.aao.org/Assets/8cd3d9a3-ee7b-4203-8659-68c611b00537/637064008803430000/diabetic-retinopathy-preferred-practice-pattern-2019-pdf

[14] Solomon, S. D. ; Chew, E. ; Duh, E. J. ; Sobrin, L. ; Sun, J. K. ; VanderBeek, B. L. ; Wykoff, C. C. ; Gardner, T. W. Diabetic Retinopathy : A Position Statement by the American Diabetes Association. *Diabetes Care*, 2017, **40**(3), 412–418. doi :10.2337/dc16-2641. Erratum in : *Diabetes Care*, 2017, **40**(6), 809. doi :10.2337/dc17-er06e. Erratum in : *Diabetes Care*, 2017, **40**(9), 1285. doi :10.2337/dc17-er09. PMID :28223445 ; PMCID :PMC5402875. Available from : https://pmc.ncbi.nlm.nih.gov/articles/PMC5402875/

[15] National Institute for Health and Care Excellence (NICE). *Diabetic retinopathy : management and monitoring*. London : NICE ; 2024 Aug 13. (NICE Clinical Guidelines, No. 242). Available from : https://www.ncbi.nlm.nih.gov/books/NBK607261/ Also available from : https://www.nice.org.uk/guidance/ng242

[16] Gaddam, S. ; Periasamy, R. ; Gangaraju, R. Adult Stem Cell Therapeutics in Diabetic Retinopathy. *International Journal of Molecular Sciences*, 2019, **20**(19),

4876. doi :10.3390/ijms20194876. PMID :31575089 ; PMCID :PMC6801872. Available from : https://www.mdpi.com/1422-0067/20/19/4876 Available from : https://pmc.ncbi.nlm.nih.gov/articles/PMC6801872/

[17] Wikipedia contributors. *Diabetic retinopathy — Wikipedia, The Free Encyclopedia*, 2025. Disponible à : https://en.wikipedia.org/w/index.php?title=Diabetic_retinopathy&oldid=1305128267 [En ligne ; consulté le 1-Septembre-2025]

[18] Shukla, U. V. ; Tripathy, K. *Diabetic Retinopathy*. In : StatPearls [Internet]. Treasure Island (FL) : StatPearls Publishing ; 2025 Jan–. Updated 2023 Aug 25. Available from : https://www.ncbi.nlm.nih.gov/books/NBK560805/ Accessed 2025 Sep 1.

[19] Mayo Clinic. *Diabetic Retinopathy – Symptoms & Causes*. Mayo Clinic [Internet]. 2025. Available from : https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/symptoms-causes/syc-20371611 Accessed 2025 Sep 1.

[20] Teo, Z. L. ; Tham, Y. C. ; Yu, M. ; Chee, M. L. ; Rim, T. H. ; Cheung, N. ; Bikbov, M. M. ; Wang, Y. X. ; Tang, Y. ; Lu, Y. ; Wong, I. Y. ; Ting, D. S. W. ; Tan, G. S. W. ; Jonas, J. B. ; Sabanayagam, C. ; Wong, T. Y. ; Cheng, C. Y. Global prevalence of diabetic retinopathy and projection of burden through 2045 : systematic review and meta-analysis. *Ophthalmology*, 2021 Nov ; **128**(11) :1580–1591. doi :10.1016/j.ophtha.2021.04.027. Epub 2021 May 1. PMID :33940045. Available from : https://pubmed.ncbi.nlm.nih.gov/33940045/

[21] Shanthini, A. ; Manogaran, G. ; Vadivu, G. *Deep Convolutional Neural Network for The Prognosis of Diabetic Retinopathy*. Series in BioEngineering. Springer, Singapore ; 2022. ISBN 978-981-19-3876-4. doi :10.1007/978-981-19-3877-1. Available from : https://link.springer.com/book/10.1007/978-981-19-3877-1

[22] American Optometric Association. Diabetic Retinopathy [Internet]. St. Louis : AOA ; c2025. Available from : https://www.aoa.org/healthy-eyes/eye-and-vision-conditions/diabetic-retinopathy Accessed 2025 May 10.

[23] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. Khan, and H. Fu, *Transformers in Medical Imaging : A Survey*, arXiv preprint arXiv :2201.09873, 2022. DOI : 10.48550/arXiv.2201.09873. .

[24] Yau, J. W. Y. ; Rogers, S. L. ; Kawasaki, R. ; Lamoureux, E. L. ; Kowalski, J. W. ; Bek, T. ; Chen, S. J. ; Dekker, J. M. ; Fletcher, A. ; Grauslund, J. ; et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*, 2012, **35**(3) :556–564. doi :10.2337/dc11-1909. PMID :22301125 ; PMCID :PMC3322721. Available from : https://pmc.ncbi.nlm.nih.gov/articles/PMC3322721/

[25] Gulshan, V. ; Peng, L. ; Coram, M. ; Stumpe, M. C. ; Wu, D. ; Narayanaswamy, A. ; Venugopalan, S. ; Widner, K. ; Madams, T. ; Cuadros, J. ; et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 2016, **316**(22) :2402–2410. doi :10.1001/jama.2016.17216. PMID :27898976. Available from : https://jamanetwork.com/journals/jama/fullarticle/2588763

[26] Litjens, G. ; Kooi, T. ; Bejnordi, B. E. ; Setio, A. A. A. ; Ciompi, F. ; Ghafoorian, M. ; van der Laak, J. A. ; van Ginneken, B. ; Sánchez, C. I. A survey on

deep learning in medical image analysis. *Medical Image Analysis*, 2017, **42** :60–88. doi :10.1016/j.media.2017.07.005. PMID :28778026. Available from : https://www.sciencedirect.com/science/article/pii/S1361841517301135

[27] Dosovitskiy, A. ; Beyer, L. ; Kolesnikov, A. ; Weissenborn, D. ; Zhai, X. ; Unterthiner, T. ; Dehghani, M. ; Minderer, M. ; Heigold, G. ; Gelly, S. ; Uszkoreit, J. ; Houlsby, N. An image is worth 16x16 words : Transformers for image recognition at scale. arXiv preprint arXiv :2010.11929, 2020. Available from : https://arxiv.org/abs/2010.11929

[28] Khan, S. ; Naseer, M. ; Hayat, M. ; Zamir, S. W. ; Khan, F. S. ; Shah, M. Transformers in vision : A survey. *ACM Computing Surveys (CSUR)*, 2022, **54**(10s) :1–41. doi :10.1145/3505244. Available from : https://dl.acm.org/doi/10.1145/3505244

[29] Vaswani, A. ; Shazeer, N. ; Parmar, N. ; Uszkoreit, J. ; Jones, L. ; Gomez, A. N. ; Kaiser, Ł. ; Polosukhin, I. Attention Is All You Need. *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, **30**. Available from : https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html Preprint available from : https://arxiv.org/abs/1706.03762

[30] Upadhyay, A. ; Chandel, N. S. ; Singh, K. P. ; et al. Deep learning and computer vision in plant disease detection : a comprehensive review of techniques, models, and trends in precision agriculture. *Artificial Intelligence Review*, 2025, **58** :92. doi :10.1007/s10462-024-11100-x. Available from : https://doi.org/10.1007/s10462-024-11100-x

[31] Ulhaq, A. Dark Transformer : A Video Transformer for Action Recognition in the Dark. *arXiv preprint* arXiv :2407.12805, 2024. Available from : https://arxiv.org/abs/2407.12805

[32] Lai-Dang, Q.-V. A Survey of Vision Transformers in Autonomous Driving : Current Trends and Future Directions. arXiv preprint arXiv :2403.07542, 2024. Available from : https://arxiv.org/abs/2403.07542

[33] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv :1810.04805*, 2019. [Online]. Available : https://arxiv.org/abs/1810.04805

[34] ADCIS. *Messidor-2 : Diabetic Retinopathy Image Dataset*. Disponible sur : https://www.adcis.net/en/third-party/messidor2/

[35] Decencière, E. ; Zhang, X. ; Cazuguel, G. ; Lay, B. ; Cochener, B. ; Trone, C. ; Gain, P. ; Ordonez, R. ; Massin, P. ; Erginay, A. ; Charton, B. ; & Klein, J.C. Feedback on a publicly distributed image database : The Messidor database. *Image Analysis & Stereology* 2014, **33**(3), 231–234. https ://doi.org/10.5566/ias.1155 Dataset disponible à : https://www.adcis.net/en/third-party/messidor/

[36] nkicsl. *OIA-DDR : A general-purpose high-quality dataset for diabetic retinopathy classification, lesion segmentation and lesion detection* [Data set]. GitHub, 2020. Disponible à : https://github.com/nkicsl/DDR-dataset

[37] Li, T. ; Gao, Y. ; Wang, K. ; Guo, S. ; Liu, H. ; & Kang, H. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Information Sciences* 2019, **501**, 511–522. https ://doi.org/10.1016/j.ins.2019.06.011 Dataset disponible à : https://www.kaggle.com/datasets/mariaherrerot/ddrdataset

[38] Kaggle and EyePACS. *Kaggle Diabetic Retinopathy Detection – Diabetic Retinopathy Dataset*, 2015. Disponible sur : https://www.kaggle.com/c/diabetic-retinopathy-detection/data

[39] APTOS. APTOS 2019 Blindness Detection Challenge. *Kaggle*, 2019. https://www.kaggle.com/c/aptos2019-blindness-detection.

[40] Gu Z, Li Y, Wang Z, Kan J, Shu J, Wang Q. Classification of Diabetic Retinopathy Severity in Fundus Images Using the Vision Transformer and Residual Attention. *Computational Intelligence and Neuroscience*. 2023 Jan 3 ;2023 :1305583. doi : 10.1155/2023/1305583. PMID : 36636467 ; PMCID : PMC9831706.

[41] Terven J., Cordova-Esparza D.M., Romero-González J.A. et al. *A comprehensive survey of loss functions and metrics in deep learning. Artificial Intelligence Review*, 58 :195, 2025. doi : https://doi.org/10.1007/s10462-025-11198-7.

[42] Yao Z, Yuan Y, Shi Z, Mao W, Zhu G, Zhang G and Wang Z (2022) FunSwin : A deep learning method to analysis diabetic retinopathy grade and macular edema risk based on fundus images. *Front. Physiol.* 13 :961386. doi:10.3389/fphys.2022.961386.

[43] Yang Y., Cai Z., Qiu S., Xu P. (2024) Vision transformer with masked autoencoders for referable diabetic retinopathy classification based on large-size retina image. *PLoS ONE* 19(3) : e0299265. https://doi.org/10.1371/journal.pone.0299265.

[44] Yuanyuan Liu, Dazhi Yao, Yongwen Ma, Hua Wang, Jinming Wang, Xuefeng Bai, Guang Zeng, Yuejuan Liu, *STMF-DRNet : A multi-branch fine-grained classification model for diabetic retinopathy using Swin-TransformerV2*, Biomedical Signal Processing and Control, Volume 103, 2025, 107352, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2024.107352.

[45] Touati, M. ; Touati, R. ; Nana, L. ; Benzarti, F. ; Ben Yahia, S. *DRCCT : Enhancing Diabetic Retinopathy Classification with a Compact Convolutional Transformer. Big Data Cogn. Comput.* 2025, **9**(1), 9. https://doi.org/10.3390/bdcc9010009.

[46] T. R. Athira et J. N. Jyothisha, "Diabetic Retinopathy Grading From Color Fundus Images : An Autotuned Deep Learning Approach," *Procedia Computer Science*, vol. 218, pp. 1055–1066, 2023. Disponible en ligne : https://www.sciencedirect.com/science/article/pii/S1877050923000856

[47] D. Mok, J. Bum, L. D. Tai, et H. Choo, "Cross Feature Fusion of Fundus Image and Generated Lesion Map for Referable Diabetic Retinopathy Classification," *arXiv preprint arXiv :2411.03618*, 2024. https://arxiv.org/abs/2411.03618

[48] Sebastian A., Elharrouss O., Al-Maadeed S., Almaadeed N. A Survey on Deep-Learning-Based Diabetic Retinopathy Classification. *Diagnostics*, 13(3) :345, Jan. 2023. doi : 10.3390/diagnostics13030345. PMID : 36766451 ; PMCID : PMC9914068. Available online : https://www.mdpi.com/2075-4418/13/3/345.

[49] Alqahtani A.S., Alshareef W.M., Aljadani H.T., et al. The efficacy of artificial intelligence in diabetic retinopathy screening : a systematic review and meta-analysis. *International Journal of Retina and Vitreous*, 11 :48, 2025. doi : 10.1186/s40942-025-00670-9. Available online : https://doi.org/10.1186/s40942-025-00670-9.

[50] Kim H.E., Cosa-Linan A., Santhanam N., et al. Transfer learning for medical image classification : a literature review. *BMC Medical Imaging*, 22 :69, 2022. doi : 10.1186/s12880-022-00793-7. Available online : https://doi.org/10.1186/s12880-022-00793-7.

ABSTRACT

The global rise of diabetes makes early detection of diabetic retinopathy (DR) a pressing public health concern. This study leverages recent advances in Vision Transformers (ViTs) to design models aimed at improving the accuracy and reliability of automated diagnosis. Three approaches were developed: **ViT-SS**, based on ViT16; **ViT-LS**, derived from ViT32; and a hybrid model, **ViT-DR**, which combines the strengths of both. These architectures were applied to the classification of DR into five severity levels using fundus images. Experimental results highlight the relevance of ViTs for this task,with the hybrid ViT-DR model achieving competitive results compared to the individual architectures. This research opens promising perspectives for the integration of computer-aided diagnostic systems, contributing to earlier detection and better clinical management of diabetic retinopathy.

Keywords: *Diabetic Retinopathy, Vision Transformer,Transfer Learning, Deep Learning, Medical Image Classification, Computer-Aided Diagnosis*

RÉSUMÉ

Face à l'augmentation mondiale du diabète et à la nécessité d'un dépistage précoce de la rétinopathie diabétique (RD), cette étude explore l'apport des Vision Transformers (ViTs) pour améliorer la précision et la fiabilité du diagnostic. Trois approches principales ont été développées et évaluées : **ViT-SS**, basé sur ViT16 ; **ViT-LS**, issu de ViT32 ; et un modèle hybride, **ViT-DR**, combinant les deux afin d'exploiter leurs caractéristiques complémentaires. Ces architectures ont été adaptées pour la classification de la RD en cinq classes à partir d'images de fond d'œil. Les résultats obtenus mettent en évidence la pertinence des ViTs dans ce contexte, avec des performances renforcées par le modèle hybride qui présente des résultats compétitifs par rapport aux modèles individuels. Cette recherche ouvre ainsi de nouvelles perspectives pour le développement de systèmes de diagnostic assisté, contribuant à un dépistage plus précoce et à une meilleure prise en charge clinique de la rétinopathie diabétique.

Mots clés : *Rétinopathie diabétique, Vision Transformer,Apprentissage par transfert,Apprentissage profond, Classification d'images médicales, Diagnostic assisté par ordinateur*