

Ministère de l'enseignement Supérieur et de la recherche Scientifique
وزارة التعليم العالي والبحث العلمي

Badji Mokhtar Annaba University
Université Badji Mokhtar – Annaba
Faculté des Sciences de l'Ingéniorat



جامعة باجي مختار – عنابة

كلية علوم الهندسة

قسم الإعلام الآلي

Département d'informatique

Thèse

Présentée pour obtenir le diplôme de

Doctorat

Spécialité : Informatique

Filière : Informatique

Par :

TOUAHRI Radia

Thème :

L'Apprentissage Profond pour la Classification et l'Interprétation d'Images

Membres de jury :

Nom et Prénom	Grade	Etablissement	Qualité
MEROUANI Hayet Farida	Prof.	Université Badji Mokhtar -Annaba	Président
AZIZI Nabiha	Prof.	Université Badji Mokhtar -Annaba	Rapporteur
HAMMAMI Nacer Eddine	MCA	Université Arabie Saoudite	Co-rapporteur
SARI Toufik	Prof.	Université Badji Mokhtar -Annaba	Examineur
MAHNANE Lamia	MCA	Université Badji Mokhtar -Annaba	Examineur
MAAZOUZI Faiz	MCA	Université Med-Cherif Messaadia –Souk Ahras	Examineur

Dédicace

Je dédie cette thèse à
Mes chers parents,
Mon mari Farid & mes enfants
Takī Eddine & Fakher Eddine
mes frères, mes beaux frères & mes belles sœurs,
Et toute ma famille, surtout tante Malīka
Mes amis et mes collègues
Mon encadreur et mon co-encadreur
tous mes enseignants tout au long de mes études,
À tous ceux qui ont participé de près ou de loin à la
réalisation de ce travail.

Remerciement

Avant tout, je remercie ALLAH le tout puissant, de m'avoir donné la santé, le courage ainsi qu'une grande volonté pour aboutir ce travail.

Je tiens à remercier ma chère directrice de thèse « Mme. Azizi Nabihia » de m'avoir conseillé judicieusement, orienté, encouragé et de m'apporter son attention tout au long de ce travail avec patience et disponibilité et pour la confiance qu'elle m'a accordée.

J'exprime ma gratitude à mon co-encadreur « Mr. Nacer Eddine HAMMAMI » pour ses conseils et orientations.

Mes vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont portés à notre recherche en acceptant d'examiner ce travail et de l'enrichir par leurs propositions.

Je remercie mon mari qui grâce à son aide j'ai pu surmonter tous les obstacles.

Je remercie ma belle sœur l'ophtalmologue TALAI Wafa pour ces éclaircissements sur le domaine médical

Enfin, Je remercie chaleureusement ma famille, tous mes amis et collègues qui m'ont toujours soutenu et encouragé au cours de la réalisation de cette thèse.

A tous ceux qui m'ont été une source d'aide ou de motivation.

Merci infiniment.

المخلص

"التعلم العميق لتصنيف الصور وتفسيرها"

في الوقت الحاضر، أصبح الباحثون مهتمين كثيراً باستخدام أساليب التعلم العميق لتحليل صور الشبكية نظراً لفعاليتها في حل المشكلات المعقدة. الميزة الرئيسية لهذه الطرق هي قدرتها على الاستخراج التلقائي للميزات على عكس الطرق التقليدية. تتطلب أساليب حجماً كبيراً من البيانات لضبط المعلمات الدقيقة لشبكة وتجنب مشكلة الإفراط في التعلم، على الرغم من المشاكل المتعلقة بجمع البيانات والتعليق على هذه الصور، التي يعد مهمة صعبة يمكن أن تستغرق وقتاً طويلاً مع تكييف مجهود كبير من الخبراء.

من أجل تحسين أساليب التعلم العميق المخصصة لتصنيف صور شبكية العين ولتجنب مشكلة الإفراط في التعلم التي يسببها اختيار بنية المصنف، اقترح الباحثون في هذا المجال بعض الأعمال استخدام تقنيات التعلم المنقولة والتنظيم وطرق التعلم الجماعي. تركز هذه الأطروحة على مساهمتين أساسيتين تهدفان إلى تحسين وزيادة فعالية تصنيف صور شبكية العين،

أولاً، اقترحنا منهجية التحليل الاجمالي تعتمد على تعدد الوسائط لدمج المصنفات حيث يعتمد على أربعة أنماط للصورة (الملونة، ثنائية اللون، الرمادية و مرشح غابور) لذلك اقترحنا بنية موثوقة لشبكة "سي ان ان" لتشخيص الجلوكوما، حيث أن ترتيب طبقات هذه الشبكة والتكيف مع معلوماتها الدقيقة يعد موضوعاً للبحث. في المقام الثاني، اقترحنا نهجاً محلياً يعتمد على مرحلتين من التجزئة (التجزئة المسبقة والتجزئة) تهدف أولاً إلى تحديد موقع القرص البصري وقطعه للتركيز على منطقة مجال الاهتمام، ثم اعتمدنا على التجزئة الدلالية باستخدام بنية "يونات" الأنيقة من نوع تشفير وضد التشفير من أجل تقسيم هذه المنطقة بشكل أفضل، بمجرد إنشائها، تخضع هذه الصور إلى شبكة "سي ن ن" أخرى لتصنيف محتمل.

وفقاً للنتائج التجريبية للمساهمة الأولى، أكدنا أن استخدام طرق متعددة الوسائط يؤثر بشكل إيجابي على كفاءة التصنيف بفضل إدخال طريقة النسيج (مرشح غابور) والتي أعطت نتائج أفضل. في نفس السياق، تقدم المساهمة الثانية القائمة على تعاقب مرحلتين من تجزئة منطقة القرص البصري تجزئة دقيقة لمجال الاهتمام مما يؤدي لاحقاً إلى زيادة أداء التصنيف الذي يعد الهدف الرئيسي لعملنا.

كلمات مفتاحية: الشبكات العصبية التلافيفية (سي ن ن)، التصنيف، التجزئة الدلالية، التعلم العميق، التصنيف

متعدد الوسائط، التشفير وضد التشفير.

Résumé

« L'Apprentissage Profond pour la Classification et l'Interprétation d'Images »

Ces dernières années, l'utilisation des méthodes de l'apprentissage profond ou Deep Learning « *DL* » pour l'analyse des images rétiniennes attire de plus en plus l'attention de la communauté des chercheurs en apprentissage automatique; cela se fait grâce à leur efficacité dans la résolution des problèmes complexes. L'avantage majeur de ces modèles est leur capacité d'extraire automatiquement les caractéristiques représentant l'image traitée contrairement aux méthodes classiques. Les méthodes *DL*, et plus particulièrement les réseaux de neurones convolutifs (*CNNs*), nécessitent un grand volume de données pour ajuster équitablement les hyper-paramètres du modèle profond de classification et éviter le problème du sur-apprentissage.

Afin d'optimiser les méthodes *DL* dédiées à la classification des images rétiniennes, certains travaux ont proposé l'utilisation des techniques d'apprentissage par transfert, par régularisation et par des méthodes ensemblistes.

Dans cette thèse, deux contributions ont été proposées et bien décrites pour améliorer et augmenter le taux de classification des images rétiniennes.

Dans un premier lieu, une approche globale multimodale basée sur quatre modalités d'images (couleur, binaire, niveau de gris, et intégrant les filtre de Gabor) est proposée. Le choix de l'architecture adapté du *CNN adopté* pour le diagnostic de la maladie du glaucome et la prise en considération de l'ordre des couches et l'adaptation des hyper-paramètres est un sujet de recherche.

Dans un second lieu, nous avons proposé une approche locale basée sur deux étapes de segmentation (pre-segmentation et segmentation) visant d'abord à localiser et couper le disque optique pour se focaliser sur la zone d'intérêt, ensuite une segmentation sémantique utilisant l'architecture *U-Net* est adoptée pour l'extraction de cette zone. Une fois générée, elle sera l'entrée d'un autre modèle de classification de type *CNN* pour une éventuelle classification.

Mots clés : Réseau de Neurones Convolutifs, Classification, Segmentation sémantique, Apprentissage profond, Multimodalité, Encodeur-décodeur.

Abstract

« Deep Learning for Image Classification and Interpretation »

Nowadays, researchers are becoming very interested in using deep learning "DL" methods for retinal image analysis due to their effectiveness in solving complex problems. The major advantage of these methods is their ability for automatic feature extraction unlike classical methods. DL methods require a large volume of data to fairly adjust the hyper-parameters of a CNN and avoid the problem of over-learning, despite the limitations of collecting data and annotating these images which is a difficult task and can take a huge amount of time with a large workload devoted by experts.

In order to optimize the DL methods dedicated to the classification of retinal images and to avoid the problem of over-learning which is caused by the choice of CNN architectures, some works have proposed the use of transferred learning techniques, regularization and ensemble learning methods.

In this thesis, we have proposed two contributions to improve and increase the classification rate of retinal images.

First, we proposed a global multimodal approach based on four image modalities (color, binary, gray scale, gabor), from which we adopted a reliable architecture of a CNN for glaucoma diagnosis, where the order of layers and the adaptation of hyper-parameters is a subject of research.

In a second place, we proposed a local approach based on two segmentation stages (pre-segmentation and segmentation) aiming first to locate and cut the optical disc to focus on the area of interest (ROI) then we have approached a semantic segmentation using the elegant U-Net architecture of encoder-decoder type in order to better segment this area, once generated it will be the input of another CNN for a possible classification.

According to the experimental results of the first contribution, we confirmed that the use of multimodal methods positively influences the classification efficiency thanks to the introduction of the texture modality (gabor filter) which gives better results. In the same context the contribution two based on the succession of two segmentation stages of the optical disc area offers a refined segmentation of the area of interest which subsequently increases the performance of the classification which is the goal of our work.

Key words: convolutional neural network, classification, Semantic segmentation, Deep learning, multimodality, encoder-decoder.

Liste des abréviations

DL	Deep Learning
ML	Machine learning
IA	Intelligence artificielle
CNNs	Réseaux de Neurones Convolutionnels
R-CNN	réseaux neuronaux convolutionnels récurrents
RIM-ONE	Base de données rétinienne pour l'évaluation du nerf optique
RNA	Réseau de Neurones Artificiels
SAD	Système d'aide au diagnostic
SADe	Système d'aide à la détection
SADx	Système d'aide au diagnostic
SVM	Séparateurs à Vaste Marge
KNN	K-Nearest Neighbor
KPPV	Méthode des k plus proches voisins
CT	Computed Tomography
TDM	Tomodensitométrie
IRM	Imagerie par résonance magnétique
M/EEG	Magnéto/Electro-encéphalographie
RNM	Rétinographies non mydriatique
DMLA	Dégénérescence maculaire liée à l'âge
OCT	cohérence optique Tomographie
ROC	Fonction d'efficacité du récepteur
AUC	L'air sous la courbe
GPU	Processeur graphique
CPU	unité centrale de traitement
DA	Data Augmentation
TansorFlow	bibliothèque open source de Machine Learning
RNA	Réseau de Neurones Artificiels
Keras	bibliothèque open source écrite en python
MLP	Multi Layers Perceptron
CONV	La couche convolutive
ReLU	Unité Linéaire Rectifiée
ELU	Exponential Linear Unit
LRN	Local Response Normalisation
YOLO	You Only Look Once
ROI	Region of interest
FCN	fully convolutional network
NEI	National Eye Institute
OD	disque optique
CO	cupule optique
RVB	Rouge vert bleu
IoU	Intersection over Union
ADAM	adaptive learning rate optimization algorithm

Sommaire

المخلص	II
Résumé.....	III
Abstract	IV
Liste des abréviations	V
Sommaire	VI
Liste des figures	X
Liste des tableaux	XIII
Introduction Générale.....	1
Chapitre I: Intelligence artificielle pour le diagnostic médical	
I. Introduction.....	8
II. Intelligence artificielle et diagnostic médical	8
III. Le diagnostic médical.....	8
IV. Système d'aide au diagnostic médical (SAD).....	9
V. Les types des systèmes d'aide au diagnostic médical :.....	10
VI. Processus général d'un système d'aide au diagnostic (SAD).....	11
VI.1. Acquisition et numérisation	12
VI.2. Prétraitement	13
VI.3. Extraction des caractéristiques	13
VI.4. La classification des images médicales	14
VI.4.1. Méthodes supervisées	14
VI.4.2. Méthode non supervisée	14
VII. L'imagerie médicale :.....	14
VII.1. Définition d'une image médicale :	15
VII.2. Les moyens physiques des différents procédés utilisés en imagerie médicale.....	15
VII.2.1. Radiographie (rayons-X) :	15
VII.2.2. La mammographie (Computed Tomography (CT)) :	16
VII.2.3. Scanographie (Tomodensitométrie (TDM)) :	17
VII.2.4. Ultrasons (Ultrasonographie ou échographie).....	18
VII.2.5. Imagerie nucléaire (scintigraphie ou «gammagraphie»)	18
VII.2.6. Imagerie par résonance magnétique (IRM).....	19
VII.2.7. Magnéto/Electro-encéphalographie (M/EEG).....	20
VII.2.8. Rétinographies (mydriatique et non mydriatique RNM).....	21

VIII.	Diagnostic des images rétinienne	21
VIII.1.	Description de l'œil humain :	21
VIII.2.	Anatomie de l'œil humain :	22
VIII.3.	Différents types de maladies de la rétine	23
VIII.4.	Le fond d'œil humain	25
IX.	Les travaux connexes des systèmes d'aide au diagnostic médical	26
X.	Les Bases de données des images rétinienne à accès public	28
XI.	L'augmentation des données (data augmentation)	30
XII.	Conclusion	33
Chapitre II: Les techniques de l'IA et l'apprentissage profond		
I.	Introduction	35
II.	C'est quoi l'Intelligence Artificielle (IA) ?	36
III.	L'évolution des réseaux de neurones :	36
IV.	Les réseaux de neurones	37
IV.1.	Le neurone biologique	37
IV.2.	Neurone formel :	37
IV.3.	Réseau de Neurones Artificiels	38
V.	L'apprentissage Automatique	38
V.1.	L'apprentissage supervisé (Supervised Learning) :	39
V.2.	L'apprentissage non supervisé:	39
V.3.	L'apprentissage par renforcement	40
VI.	Historique d L'apprentissage profond	40
VII.	Domaines d'application de l'apprentissage profond :	40
VIII.	Les réseaux de neurones convolutifs	41
VIII.1.	La phase d'extraction des caractéristiques	42
VIII.2.	La phase de classification de modèle :	42
VIII.3.	Architecture et fonctionnement du CNN :	42
VIII.4.	Les différentes couches de CNN :	43
VIII.4.1.	La couche convolutive (CONV)	44
VIII.4.2.	La couche d'agrégation (pooling) :	48
VIII.4.3.	La couche entièrement connectée :	49
VIII.5.	Ajustement des hyperparamètres :	50
IX.	Apprentissage profond et imagerie médicale	50
X.	L'interprétation des images	51

X.1.	La Classification.....	51
X.1.1.	Les architectures de classification les plus connues.....	51
X.1.2.	Exemples des travaux de classification profonde.....	57
X.2.	La Détection et localisation des objets.....	57
X.2.1.	Les architectures de détection les plus connues	58
X.2.2.	Exemples des travaux de détection.....	60
X.3.	La Segmentation.....	60
X.3.1.	Les architectures de segmentation les plus connues.....	60
XI.	Conclusion :.....	64
Chapitre III: Approche globale de classification multimodale de la maladie de glaucome basée sur les classifieurs CNNs et la représentation de texture (Filtre de gabor)		
I.	Contexte de la contribution.....	66
II.	Introduction et travaux connexes.....	67
III.	Concepts de base :.....	69
III.1.	La Multimodalité.....	69
III.1.1.	Fusion multimodale	70
III.1.2.	Intérêt de la fusion multimodale.....	70
III.2.	Classification d'ensemble (Ensemble Learning).....	70
III.2.1.	Pourquoi combiner des classifieurs ?	71
III.2.2.	Les stratégies de combinaison	71
III.3.	Filtre de Gabor	73
IV.	La méthode proposée	74
IV.1	Prétraitement	75
IV.2	La structure β plus –CNN.....	76
IV.2.1.	Phase d'apprentissage:	76
IV.2.2.	Phase de test:	78
V.	Description du jeu de données	79
VI.	Résultats expérimentaux	80
VI.1	Les mesures d'évaluation.....	80
VI.2	Analyses empiriques des paramètres de β plus –CNN.....	82
VI.2.1.	L'impact du nombre d'époques et du nombre d'itérations:	82
VI.2.2.	L'impact de la fonction d'activation.....	83
VI.2.3.	Les paramètres choisis de notre CNN	84
VI.2.4.	Les résultats de la phase d'apprentissage du système globale β plus –CNN	85

VI.2.5. Les performances obtenues	85
VI.2.6. Les résultats de la classification du système global β plus -CNN	86
VI.2.7. Les performances obtenues en phase de test	87
VI.2.8. Comparaison des résultats	88
VI.2.9. Fusion parallèle (agrégation somme / produit).....	88
VI.2.10. Analyse des résultats par la courbe ROC.....	89
VII. Conclusion.....	92
Chapitre IV: Une approche locale combinant deux phases de segmentation du disque optique basées U-NET pour le diagnostique du glaucome.....	95
I. Objectif et motivation	95
II. Introduction.....	95
III. Motivation de la segmentation/détection automatique des points de repère de la rétine (Retina land marks) :	97
IV. Quels sont les points de repère à segmenter dans le fond d'œil de la rétine (Retina land marks) ?	98
IV.1. Segmentation du disque optique:	98
IV.2. Segmentation de la macula et de la fovéa:	99
IV.3. Segmentation vasculaire (veines et artères):	100
V. Système proposé :	101
V.1. Etapes 1 : La segmentation.....	102
V.1.1. La pré-segmentation :	102
V.1.2. Localisation et coupure de l'emplacement du disque optique.....	102
V.1.3. La segmentation avec U-Net	107
V.2. Étape 2 : classification avec l'architecture CNN	110
V.2.1. Architecture CNN:.....	112
V.3. Description de l'ensemble de données REFFUGE	112
V.4. Résultats et discussion:.....	113
V.4.1. Mesures de performances:	113
V.4.2. Implémentation.....	115
V.4.3. Résultats de l'étape de segmentation:	115
V.4.4. Résultats de l'étape de classification:.....	116
VI. Conclusion.....	119
Perspectives.....	122
Productions scientifiques.....	123
Bibliographie :	124

Liste des Figures

Figure 1.1: processus générale du système d'aide au diagnostic médical	12
Figure 1.2 : Radiographie (rayons-X) [15].....	16
Figure 1.3 : La mammographie [16]	17
Figure 1.4: Scanographie [15].....	17
Figure 1.5: Radiologie Ultrasons (Ultrasonographie ou échographie) [16].....	18
Figure 1.6:Exemple de la scintigraphie cérébrale	19
Figure 1.7: Diagnostique par résonance magnétique (IRM) [15]	19
Figure 1.8:Imagerie par magnéto-et électro-encéphalographie (MEG/EEG) [17]	20
Figure 1.9: Imagerie du système nerveux central[17].....	20
Figure 1.10: La Rrétinographie Non Mydriatique (RNM) [18].....	21
Figure 1.11: Coupe transversale et vue de fond de l'œil [19].....	22
Figure 1.12:Anatomie de l'œil [19].....	22
Figure 1.13: Maladie de glaucome [20]	25
Figure 1.14 : Échantillons d'images de fond d'œil normal(a) et anormal(b) [34].....	26
Figure 1.15 : Flip horizontal (a)et flip vertical(b) [36]	31
Figure 1.16 : effet de Luminosité [36]	31
Figure 1.17 : effet de changement de canal [36].....	31
Figure 1.18 : Recadrage [36].....	32
Figure 1.19: effet de Rotation [36]	32
Figure 1.20 : Décalage à gauche [36].....	32
Figure 1.21 : effet de zoom [37].....	33
Figure 2.1 : Schéma simplifié d'un neurone biologique [39].....	37
Figure .2.2 : Comparaison entre neurone biologique et neurone formel [39].....	38
Figure 2.3 : architecture global d'un CNN[34]	42
Figure 2.4 : La couche de convolution[68]	44
Figure 2.5 : Profondeur d'un filtre[67].....	45
Figure 2.6 : Le stride d'un filtre [67]	46
Figure 2.7 : Représentation de la fonction ReLU [68].....	46
Figure 2.8 : Représentation de la fonction tangente hyperbolique.....	47
Figure 2.9 : Représentation de la fonction sigmoïde.....	47
Figure 2.10 : La fonction d'activation Identity [69].....	48

Figure 2.11 :La couche entièrement connectée [68]	49
Figure 2.12 : Les trois taches essentielles de l'interprétation d'images.....	51
Figure 2.13 : L'architecture du réseau LeNet [53].....	52
Figure 2.14: L'architecture du réseau AlexNet [77]	53
Figure 2.17 :La structure d'un module d'Inception	55
Figure 2.18: L'architecture du réseau Inception	55
Figure 2.19 – La structure des blocs résiduels [82].....	56
Figure 2.20 : l'architecture du réseau ResNet (34 couches) [82].....	56
Figure 2.21: L'architecture du réseau DenseNet [83]	57
Figure 2.22 – L'architecture de Régions avec réseaux de neurones convolutif (R-CNN).....	58
Figure 2.23 – l'architecture du réseau Fast R-CNN.....	59
Figure 2.24 – L'architecture de détection des objets par YOLO	59
Figure 2.25: L'architecture du réseau fully convolutional network.....	61
Figure 2.26 :L'architecture du réseau U-Net	62
Figure 2.27 : L'architecture du réseau Seg-Net	63
Figure 3.1 : Combinaison séquentielle de classifieurs [127]	72
Figure 3.2 : Combinaison parallèle de classifieurs[127].....	72
Figure 3.3 : Combinaison hybride de classifieurs	72
Figure 3.4 (a) Image de fonds d'œil en niveaux de gris tirée de l'ensemble de données RimOne et (b) traitement de la même image à l'aide du filtre Gabor.....	73
Figure 3.5. Organigramme proposé β plus –CNN	74
Figure 3.6: Résultats des images filtrées par Gabor avec différentes orientations et fréquences .	76
Figure.3.7 :Les quatre modalités (a) couleur,(b) binaire, (c) niveau de gris,(d) filtre de Gabor...	76
Figure.3.8. phase d'apprentissage	78
Figure 3.9 : phase de test.....	79
Figure 3.10 :Exemples d'images numériques du fond de l'œil:.....	79
(a, b) glaucome, (c, d) normal [133].	79
Figure 3.11 : Matrice de confusion	80
Figure 3.12 Matrice de Confusion de la phase d'apprentissage.....	85
Figure 3.13. Score du modèle de couleur par rapport à l'itération	86
Figure 3.14. Apprendre des informations sur le modèle Couleur	86
Figure 3.15: Matrice de Confusion de phase de test pour les quatre modèles	87
Figure 3.16: Confusion matrices for: (a) Sum aggregation method, (b) Product aggregation method.....	88
Figure 3.17 : La courbe ROC et les valeurs AUC pour les quatre classificateurs	90

Figure 3.18 La courbe ROC et les valeurs AUC pour la méthode de somme et d'agrégation	90
Figure 4.1: image de fond d'œil montrant le disque optique et la cupule optique.....	99
Figure 4.2: image de fond montrant Macula et fovéa	100
Figure4.3: image de fond montrant les vaisseaux sanguins	101
Figure 4.4 : L'approche locale et l'approche globale.....	102
figure 4.5: décomposition de l'image de fond d'œil en trois canaux rouge, vert, bleu.....	103
figure 4.6: localisation et coupure de la zone lumineuse selon les paramtres (Px,Py, Lx,Ly)....	104
Figure 4.7 : la phase de pré-segmentation.....	105
Figure 4.8 exemple de localisation et coupure de disque optique.....	106
Figure 4.9: L'architecture U-Net.....	107
Figure 4.10 : la phase de segmentation par U-net	109
Figure 4.11 : exemple d'images d'entrée et de sortie de l'étape de segmentation (phase de test)	110
Figure 4.12 : la phase de classification	111
Figure 4.13: Les techniques d'augmentation de données utilisées dans notre approche.....	112
Figure 4.14: image originale et son annotation correspondante à partir de l'ensemble de données Reffuge	113
Figure 4.15: les résultats de performance de la modalité Image originale.....	117
Figure 4.16: les résultats de performance des mages originaux prédites	117
Figure 4. 17: les résultats de performance des images de la zone coupée «DO» prédites (après segmentation)	118

Liste des tableaux

Tableau 1.1 travaux connexes des systèmes d'aide au diagnostic médical	27
Tableau 1.2 Résumé des bases de données d'images du fond d'œil accessibles au public :	28
Tableau 2.1 – Les domaines d'application des CNNs	41
Tableau 2.2: Présentation des types de pooling	49
Tableau 2.3 : Exemples des travaux de classification profonde	57
Tableau 2.4 : Exemples des travaux de détection profonde	60
Tableau 2.5 : Exemples des travaux de segmentation profonde	63
Tableau 3.1 : Les performances des quatre modalités avec une époque et une itération	82
Tableau 3.2 : Les performances des quatre modalités avec 30 époques et 10 itérations	82
Tableau 3.3 : Les performances des quatre modalités avec 60 époques et 20 itérations	83
Tableau 3.4: Les performances des quatre modalités avec 30 époques et 10 itérations	83
Tableau 3.5 Paramètres choisis de notre CNN	84
Tableau 3.6: Résultats obtenus des différentes métriques de la méthode proposée en phase d'apprentissage	85
Tableau 3.7 : Résultats obtenus des différentes métriques de la méthode proposée en phase de test.	87
Tableau 3.8 Résultats obtenus des différentes métriques après les méthodes d'agrégation somme/ produit.	88
Tableau 3.9 Les performances des quatre classificateurs.	89
Tableau 4.1: les paramètres choisis de notre CNN	112
Tableau 4.2: les performances de segmentation basées sur les images originales	115
Tableau 4.3: les performances de segmentation basées sur les images pré-segmentées (zones DO coupées)	116

Introduction Générale

Durant les dernières décennies, le développement des systèmes d'aide au diagnostic dans le domaine médical a été perçu avec beaucoup d'intérêt par les médecins.

Un système d'aide au diagnostic médical permet d'identifier des structures utilisées afin d'apporter une aide précieuse aux médecins. Les fonctions de diagnostic et de suivi s'appuient généralement sur un modèle de fonctionnement normal et/ou défaillant. Lorsque cette connaissance est minimale et insuffisante, les méthodes de reconnaissances des formes constituent une solution efficace pour construire un modèle du système.

Aujourd'hui, le développement des systèmes automatiques pour le traitement d'images médicales, permettant de fournir un second avis ou décision médicale aux médecins est un objectif qui a été poursuivi par les chercheurs depuis les premiers travaux sur ce domaine.

Diagnostiquer des maladies d'une façon complètement automatique pouvait sembler irréalisable il y a encore quelques années, mais les progrès brillants des techniques avancées de l'apprentissage artificiel ainsi que l'avènement du "big data" (grandes bases de données disponibles) ont changé à faire grand pas dans le diagnostic. Toutefois, il n'existe à ce jour aucun algorithme capable de combattre avec la finesse d'analyse humaine dans la séquence de tâches que constitue l'interprétation des images médicales. Il s'agit d'un processus séquentiel consistant à détecter des anomalies dans les images médicales, à les caractériser et parfois à quantifier leur évolution.

La limite principale dont souffrent les systèmes d'aide au diagnostic, dans l'interprétation des images médicales est la capacité de trouver les caractéristiques pertinentes représentant les images d'entrées. Comme les bases médicales sont toujours évolutives avec des nouveaux cas pour une même maladie, ces systèmes d'aide aux diagnostics médicaux (SAD) ne peuvent garantir que ces caractéristiques déjà sélectionnées restent les plus performantes.

Récemment, la révolution se produit par l'explosion d'un nouveau concept qu'est l'apprentissage profond, ce dernier est un ensemble de méthodes d'apprentissage automatique proposé afin de dégager les limites citées ci-dessus, à cause de sa modélisation permettant un haut niveau d'abstraction des données grâce à des architectures détachées de différentes transformations non linéaires. Parmi ces différents algorithmes, les réseaux de neurones convolutifs (les CNNs), le modèle le plus exploité, qui ont une méthodologie ressemblante aux méthodes traditionnelles d'apprentissage supervisé, mais avec plus de fiabilité, ils reçoivent des images en

entrée, détectent les caractéristiques pertinentes d'une manière automatique et assure la classification par son modèle neuronal.

Nous investiguons dans cette thèse différentes architectures de l'apprentissage profond qui ont connu un grand succès ces dernières années afin d'augmenter la performance du diagnostic.

En effet, l'apprentissage profond est un ensemble de méthodes avancées d'intelligence artificielle qui ont apporté des contributions importantes où l'efficacité des systèmes de prise de décision dans le domaine médical a été très encourageante.

I. Description de la problématique de la thèse

Le glaucome est considéré comme la deuxième maladie oculaire responsable de la cécité dans le monde. Il est chronique et irréversible de plus il ne montre aucun symptôme précoce, le problème une fois qu'on prend conscience de sa présence, une grande partie de la vue peut être perdue. Afin d'aider les ophtalmologistes à améliorer la détection précoce des anomalies dans l'œil et d'éviter qu'ils tombent dans des décisions hésitantes, les systèmes d'aide au diagnostic (SAD) sont exploités. Dans notre cas, nous nous concentrerons sur le domaine de vision par ordinateur qui constitue le noyau essentiel des systèmes automatiques pour l'interprétation des images médicales englobant les trois tâches essentielles : la classification, la détection et la segmentation.

Ces dernière années, et grâce aux architectures d'apprentissage profond, les SADs sont orientés vers l'utilisation des méthodes de machine learning avancées telles que les réseaux de neurones convolutifs. Comme les méthodes d'apprentissage automatique nécessitent un prétraitement et l'extraction manuelle des caractéristiques (handcrafted features) et exige la présence des experts pour décider des attributs discriminants. L'avantages de ces modèles est de bénéficier de l'extraction automatiques des caractéristiques les plus pertinentes et de la standardisation du système.

Malgré les avantages de DL et le succès des architectures CNNs dans différents domaines, ils ont plusieurs défis à soulever surtout ceux liés aux problèmes de sous-apprentissage provoqué par les volumes limités de données et leurs nécessités en vitesse de calcul et en capacité de stockage.

Il faut noter que malgré l'existence des bases de données publiques, le déroulement des algorithmes basés DL nécessitent toujours plus d'investissement. Un des chalenges traitant les images médicales est l'annotation de données déployées qui reste une tâche difficile et couteuse, en particulier dans les cas de recherche des zones d'intérêt (par segmentation) nécessitant une attention énorme de la part des experts.

Afin de palier à ces limites, beaucoup d'efforts ont été fournis dans l'état de l'art pour adapter les méthodes DL, en proposant différentes architectures CNNs dédiées à la classification des images et d'autres de type encodeur-décodeur assurant la segmentation sémantique.

La fusion multimodale comprend deux catégories importantes : la fusion au niveau de caractéristiques et la fusion au niveau de classifieurs. Son principal objectif est de renforcer l'ensemble de caractéristiques provenant de diverses sources et bénéficier des diversités de différentes modalités en faisant coopérer plusieurs décideurs.

Notre objectif est de contribuer à l'interprétation des images médicales en abordant deux axes : une approche globale multimodale, où nous proposons une architecture ensembliste se composant des réseaux convolutifs (CNNs), dont l'ordre des couches (couche convolution, couche pooling et la couche entièrement connectée) et l'adaptation des hyper-paramètres est un sujet de recherche. Et une approche locale visant à localiser le disque optique pour se focaliser sur la zone d'intérêt (ROI) en proposant une technique basée sur l'architecture élégante U-Net de type encodeur-décodeur afin de segmenter cette zone, une fois générée elle sera l'entrée d'un réseau convolutif pour une éventuelle classification.

Nous voulons à travers ce travail **analyser le** comportement et l'efficacité des méthodes d'apprentissage profond pour la classification et la segmentation des images rétiniennes afin de générer des systèmes robustes pour l'aide à la décision **médicale**.

Pour atteindre ces objectifs, le système proposé doit prendre en considération plusieurs critères:

- Le choix d'une architecture optimale basée DL pour la résolution des tâches d'interprétation d'images afin d'éviter les problèmes de sous-apprentissage.
- Le choix des hyper-paramètres adéquats.
- Le choix des paramètres optimaux de chaque modèle; ces derniers seront sélectionnés après plusieurs passages de la phase d'apprentissage.
- Le choix convenable de la base d'images.

L'objectif principal de cette recherche sera d'étudier le rendement des approches basées sur l'apprentissage profond en utilisant différentes architectures, tout en proposant des architectures multimodales et des techniques d'augmentation de données, dans le but d'aboutir aux objectifs mentionnés ci-dessous:

- Développer et mettre au point une approche globale pour la classification des images rétiniennes.

- **Introduire** la multimodalité (Ensemble Learning) pour bénéficier de la diversité apportée par chacune des **sources d'images manipulées par le système** dans le but d'améliorer les performances de système proposé.
- Exploiter les techniques de l'augmentation de données.
- Développer et évaluer une approche locale faisant appel à l'architecture U-Net (encodeur-décodeur) qui englobe deux tâches d'interprétation d'images: la segmentation et la classification des images rétinienne.

II. Contribution

Nous proposons à travers la présente thèse deux contributions principales. Le but de la première contribution est de présenter le processus global de la classification de la maladie de glaucome en se basant sur une architecture CNN bien choisie. D'autre part, la deuxième contribution propose une approche locale qui combine deux phases de segmentation (localisation de disque optique puis sa segmentation basée sur l'architecture U-net) les images résultantes subies une classification.

- La première contribution porte sur une approche globale multimodale de classification des images. Afin de montrer l'impact d'utiliser la représentation de la texture (le filtre de Gabor) comme une source complémentaires de notre système d'aide au diagnostic médical et de montrer le rôle de fusion des classifieurs issus de plusieurs sources d'informations. Notre système **β plus-CNN** proposé est dédié à la classification des images de fond d'œil en deux classes: "Glaucome" et "Non-Glaucome" avec quatre visions de l'image en entrée (couleurs, binaires, niveaux de gris et les images obtenues par le filtrage de Gabor). Nous avons utilisé la base de données «RimOne» pour la validation de cette contribution.

Nous nous sommes concentrés sur le problème le plus fréquemment mentionné dans le domaine de l'apprentissage profond, à savoir l'absence de quantité suffisante de données.

- La seconde contribution repose sur une approche locale incrémentale à deux phases pour améliorer les performances du système de diagnostic assisté par ordinateur basée sur l'architecture améliorée du modèle U-Net nommée ***DSC-U-Net (deep segmentation classification U-Net)***; ce système combine deux étapes; la segmentation et la classification. D'abord, la localisation et la coupure de la zone du disque optique de l'image de fond d'œil de la rétine est réalisée dans le but de se focaliser sur la zone d'intérêt (ROI) ; puis l'utilisation de l'architecture U-Net pour une segmentation affinée

des zones (CO) et (DO) est adoptée. Un modèle de classification basé CNN est par la suite analysé pour classer le disque segmenté en glaucome et non glaucome.

Notre approche est validée par l'ensemble des images de la base de données «REFFUGE»

Ce choix d'architectures est justifié car il permet de bénéficier de deux avantages fondamentaux des modèles d'apprentissage profond déployés qui se caractérisent d'une part par la génération automatique et adaptative de caractéristiques et d'autre part, la capacité à réaliser une segmentation sémantique .

III. Organisation de la thèse

Le contenu de cette thèse est organisé en quatre chapitres, détaillé comme suit :

Chapitre 1 : *Intelligence artificielle pour le diagnostic médical*

Dans ce chapitre, le concept d'intelligence artificielle dédiée au diagnostic médical est présenté. Nous discutons par la suite le diagnostic médical ainsi que les systèmes d'aide au diagnostic médical (SAD). Nous illustrons les moyens physiques de l'imagerie médicale et plus particulièrement les concepts liés à l'œil humain, son anatomie, les différents types de maladies pouvant toucher la rétine tout en se concentrant sur la maladie du glaucome qui est le sujet de notre étude. Les bases d'image publiques utilisées pour l'analyse automatique des maladies oculaires sont présentées. Le chapitre se clôture par une section des travaux connexes résumant les principaux systèmes d'aide au diagnostic médical existant dans la littérature.

Chapitre 2 : *Les techniques d'IA et l'apprentissage profond*

Nous introduisons dans ce chapitre l'historique des réseaux de neurones ainsi que leurs évolutions; nous étalons sur la présentation des deux sous domaines de l'Intelligence Artificielle : l'apprentissage automatique (ML) et l'apprentissage profond (DL) tout en spécifiant leurs domaines d'application. L'architecture générale des réseaux de neurones convolutifs (CNN) comportant les différentes couches qui le composent est bien définie. Ce chapitre se clôture par la présentation des principales tâches de l'interprétation d'images (classification, détection et localisation, segmentation), suivie par une illustration des architectures CNNs proposées dans la littérature dans le domaine de DL.

Chapitre 3 : *Approche globale de classification multimodale de la maladie de glaucome basée sur les classifieurs CNNs et la représentation de texture (filtre de Gabor)*

Ce chapitre décrit la première contribution de cette thèse basée sur la proposition d'une approche globale multimodale de classification des images rétiniennes. Cette approche vise à analyser

l'impact d'introduire la représentation du texture (le filtre de Gabor) comme une source complémentaires à notre système d'aide au diagnostic médical ainsi que le rôle de fusion des classifieurs issues de plusieurs sources d'informations. Notre système fait appel à quatre modalités de l'image en entrée ou chacun des modèles de classification CNN sera responsable d'apprendre la maladie par sa vision liée à la nature de la source utilisée (couleurs, binaires, niveaux de gris et les images passées par le filtrage de Gabor. La base de données «RimOne» a été utilisée pour la validation de cette approche.

Chapitre 4 : *Une approche locale combinant la segmentation et la classification du disque optique en deux étapes basée sur l'architecture U-NET pour le diagnostic du glaucome.*

Ce chapitre détaille la deuxième contribution de cette thèse, qui repose sur une approche locale du système de diagnostic intégrant l'architecture U-Net. Ce système combine deux étapes : la segmentation et la classification. Après la localisation et la coupure de la zone du disque optique (ROI) de l'image de fond d'œil de la rétine par un module de pré segmentation, une architecture U-Net est adoptée pour une segmentation affinée de cette zone. Ensuite, un modèle de classification basé CNN est utilisé pour classer le disque segmenté en glaucome ou non glaucome. Notre approche est validée par l'ensemble de données de la base «REFFUGE».

Cette thèse s'achève par une conclusion générale et des perspectives envisagées pour mener le parcours de ce travail vers horizons futurs.

Chapitre I : Intelligence Artificielle pour le Diagnostic Médical

I. Introduction

Le diagnostic assisté par ordinateur (SAD) est un domaine très important de la reconnaissance des formes, avec l'apparition des big-data; des milliers des bases de données des images médicales sont offertes pour améliorer le fonctionnement de ces systèmes, des outils informatiques et des techniques de traitement et d'analyse d'images, visant à aider les médecins à prendre leurs décisions médicales. Comme l'interprétation des données est souvent coûteuse et difficile à réaliser, les systèmes SAD donnent un deuxième avis au médecin. Ces systèmes automatiques sont prometteurs dans la détection des cas suspects où ils ont prouvé des bons résultats dans les décisions finales de diagnostic de différentes maladies. Dans ce chapitre nous allons définir les concepts généraux des systèmes d'aide au diagnostic médical ainsi que les phases de son processus. Etant donné que notre travail se focalise sur les images de la rétine, nous incluons dans ce chapitre une présentation des notions de l'imagerie médicale. Ensuite, nous illustrons l'anatomie de l'œil humain, tout en basant sur la maladie de glaucome qui est l'axe principal de notre recherche. Nous regroupons dans un tableau bien détaillé les bases de données des images rétiniennes à accès public et les techniques les plus utilisées pour l'augmentation de données. Nous clôturons ce chapitre par un état de l'art des travaux d'actualité sur les SAD d'une manière générale.

II. Intelligence artificielle et diagnostic médical

Les dernières statistiques confirment que l'intelligence artificielle pourrait donner aux cliniciens un deuxième avis avec un minimum de temps et plus de précision. La nouvelle tendance consiste à intégrer l'intelligence artificielle pour **interpréter l'imagerie médicale**. Ce domaine repose sur les techniques de Machine Learning et de Deep-Learning.

Pour créer un algorithme d'interprétation d'images médicales, il faut d'abord récolter des connaissances sur ces images, sur leurs structures anatomiques, leurs méthodes d'acquisition et comment ces dernières seront utilisées dans les systèmes d'aide au diagnostic médical. A la fin les algorithmes sélectionnent les caractéristiques et apprennent à les classifier. Dans cette section nous allons éclaircir ces deux points et résumer les travaux connexes menés dans la littérature avant d'entamer la partie pratique de cette thèse.

III. Le diagnostic médical

Le mot «**diagnostic** » provient du grec διάγνωση, *diágnosi*; cette action permet d'acquérir la connaissance à travers les symptômes observables. Sa définition forcément introduit la notion de

catégories ou classes de diagnostics préexistantes, d'où le résultat peut appartenir à une classe ou à une autre [1].

Donc, un diagnostic médical représente l'action d'affecter le nom d'une ou plusieurs maladies à des déclarations présentées par le patient et qui peuvent être des symptômes ou des signes [2].

Le diagnostic médical a été défini par Jean-Charles Sournia [3] comme suit: «*Démarche intellectuelle par laquelle une personne d'une profession médicale identifie la maladie d'une autre personne soumise à son examen, à partir des symptômes et des signes que cette dernière présente, et à l'aide d'éventuelles investigations complémentaires*». C'est pour ça, le diagnostic médical est considéré comme une tâche vraiment critique car il repose sur la capacité de raisonnement du médecin et de sa compétence pour décider. En effet, plus la possibilité d'erreur est diminuée plus le diagnostic est bon.

Tout processus de diagnostic médical passe par les étapes suivantes:

- Tout d'abord, le médecin reçoit le patient et lui pose des questions sur les symptômes qu'il éprouve. A partir de ses symptômes, il formule des hypothèses de diagnostic initiales. L'interrogatoire, le recueil des symptômes et le passé de la maladie chez le patient portent le nom de « signes fonctionnels ».
- La deuxième étape est l'examen clinique basant sur des moyens simples : auscultation, palpation, prendre des mesures et utilisation de quelques appareils dans le but de recueillir des signes physiques. Ce test est toujours accompagné par les questions de médecin qui lui permet d'augmenter la part de confiance en certaines hypothèses et la diminuer pour d'autres.
- La troisième étape est la demande d'examens complémentaires (des analyses, des radios, ou des scanners....) et leurs interprétation en confrontation avec les données précédentes de l'examen physique qui permet au médecin de prendre une décision finale.[1]

Les médecins trouvent toujours des difficultés au niveau de chaque étape de diagnostic ce qui nous a amené à la conception et le développement d'un système d'aide au diagnostic médical pour assister les médecins à formuler leurs décisions.

IV. Système d'aide au diagnostic médical (SAD)

Le système d'aide au diagnostic ou le système assisté par ordinateur (SAD) a été défini par Kawamoto et al [4] comme suit :

« *We defined a clinical decision support system as any electronic or non-electronic system designed to aid directly in clinical decision making, in which characteristics of individual patients are used to generate patient-specific assessments or recommendations that are then presented to clinicians for consideration* ».

Cette définition confirme l'importance d'ajouter un système d'aide à la décision clinique qu'il soit électronique ou non, conçu pour aider directement à l'enrichissement de l'examen clinique, dans lequel les caractéristiques de chaque patient sont utilisées pour générer des informations et des recommandations spécifiques au patient et qui seront par la suite présentées aux cliniciens pour prendre des décisions.

D'autre part, Kong et al [5] et Stojkovska et al [6] ont défini le système d'aide au diagnostic médical comme suit: « *Software designed to be a direct aid to clinical decision-making, in which the characteristics of an individual patient are matched to a computerized clinical knowledge base and patient specific assessments or recommendations are then presented to the clinician or the patient for a decision* ».

D'après cette définition le système d'aide au diagnostic médical est un logiciel dédié pour aider à la prise de décision du médecin, dans lequel les symptômes d'un patient correspondent à une base de connaissances cliniques informatisée qui transforme une évaluation de l'état d'un patient au clinicien pour décision.

Aussi la prise de décision d'un diagnostic devient plus au moins facile grâce aux systèmes d'informations, telle que la possibilité d'accéder aux bases de données publiques et les dossiers informatisés.

Donc, le but principal des systèmes d'aide à la décision est d'assister le médecin, soit en le remplaçant ou en imitant son raisonnement humain. Parmi les exemples des systèmes d'aide au diagnostic on peut citer : *Les systèmes experts, les systèmes d'apprentissage, les systèmes de fouille de données, les systèmes d'indexation de recherche d'images, les systèmes de raisonnement à base de cas et les systèmes de raisonnement par classification.*

V. Les types des systèmes d'aide au diagnostic médical :

Le but principal d'intégration des techniques de reconnaissances des formes dans le processus d'interprétation de l'imagerie radiologique est d'assurer la précision dans les résultats. Les systèmes d'aide au diagnostic médical sont séparés en deux grandes catégories:

- Des systèmes d'aide à la détection (SADe)

- Des systèmes d'aide au diagnostic (SADx).

- ***Système d'aide à la détection (SADe)***:

Selon Burhenne et al [7], Les systèmes **SADe** sont des applications informatiques qui aident les radiologues à localiser les zones suspectes dans les images radiologiques, où les radiologues sont les responsables à donner leurs interprétations.

- ***Système d'aide au diagnostic (SADx)*** :

Ce sont des outils informatiques dont le but est de fournir aux radiologues en temps et lieux utiles les informations nécessaires pour caractériser les zones suspectes (lésions ex : malignité vs bénignité) qui sont déjà identifiées par un radiologue ou un modèle SADe afin de décrire la situation clinique d'un patient.

VI. Processus général d'un système d'aide au diagnostic (SAD)

Le but principal des systèmes SADs est de fournir aux cliniciens des applications informatisées pour faire le diagnostic automatique de l'image médicale en moins temps et avec plus de précision, cela démontre une amélioration remarquable de la qualité des soins et la santé des patients au niveau des hôpitaux [8]. Cet accroissement continu de ce genre des systèmes nécessite le bon choix des techniques de traitement d'image et de l'apprentissage automatique.

Dans cette section, nous présentons le processus général du système SAD qui peut suivre les étapes suivantes :

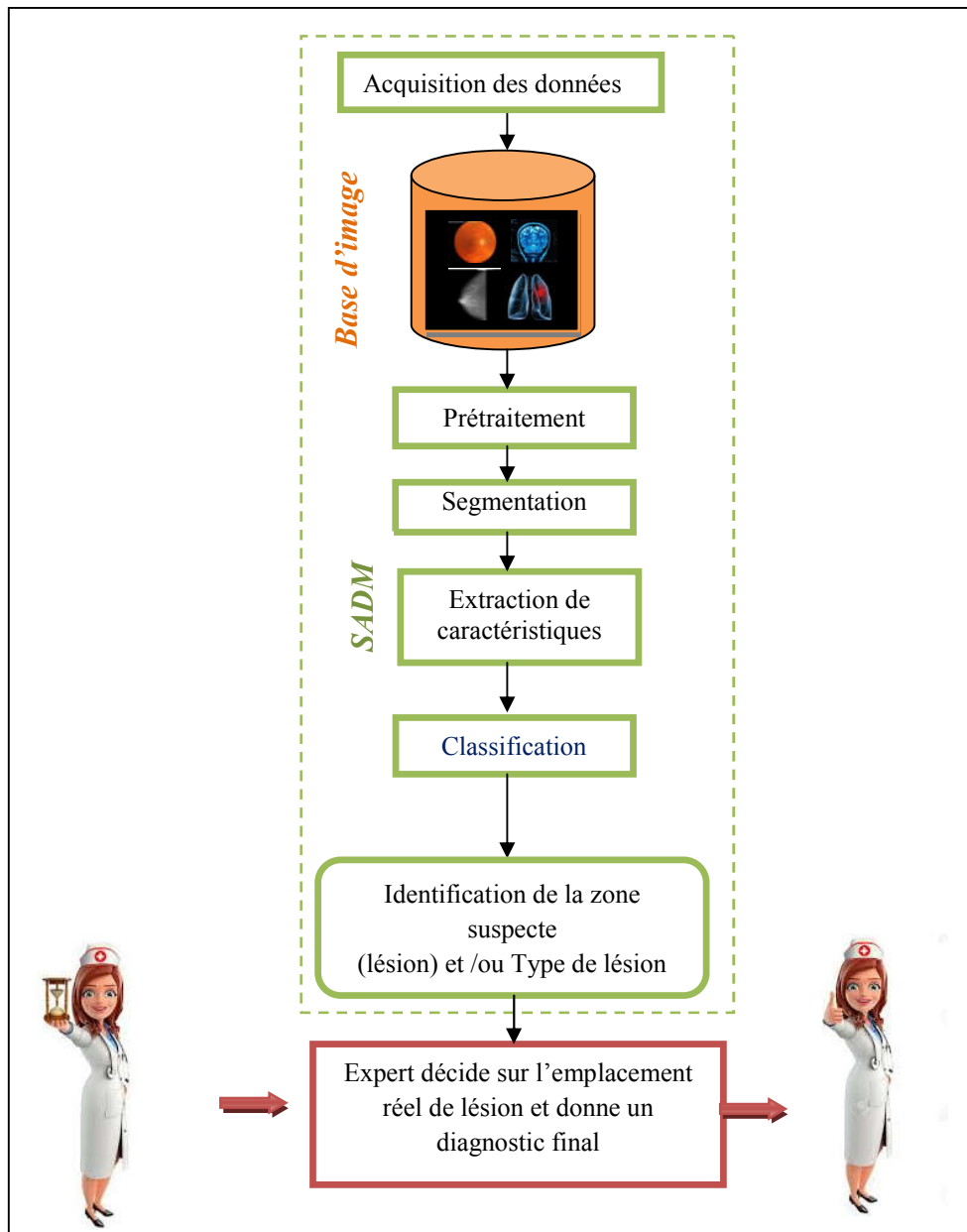


Figure 1.1: processus générale du système d'aide au diagnostic médical [8]

VI.1. Acquisition et numérisation

Plusieurs outils d'acquisition de données peuvent être intégrés au sein des SAD afin de permettre au clinicien d'obtenir sa décision en utilisant les données médicales récoltées des patients de manière structurée et automatisée[9] Ces outils sont basés sur deux champs principaux :

- les formulaires et les questionnaires électroniques,
- les techniques d'imagerie médicale et les capteurs connectés

Dans le domaine de traitement d'image, l'acquisition des données est la phase la plus importante dans toute chaîne de conception d'un système SAD, où chaque image nécessite certaines transformations qui la rendra lisible et manipulable.

VI.2. Prétraitement

Plusieurs facteurs tels que le contraste, la luminosité ou le bruit présent dans les images acquises peuvent influencer négativement et dégrader les performances d'un SAD qui amène le clinicien à une mauvaise interprétation de diagnostic [10]. De ce fait, le prétraitement est toujours mis en œuvre dans ce processus. Le prétraitement d'images joue un rôle important dans l'amélioration des images médicales pour obtenir les résultats idéaux dans les autres étapes du SAD, telles que la segmentation et l'extraction de caractéristiques, il regroupe toutes les techniques visant à améliorer la qualité d'une image médicale originale et simplifier le processus pour lui rendre plus fiable. Parmi les techniques de prétraitement :

- **La restauration:** c'est l'opération qui permet de recréer une image par l'élimination des défauts de l'image source. Toute en gardant une grande ressemblance entre l'image résultant et l'image acquise.
- **L'amélioration:** C'est la technique qui permet de modifier l'image source pour quelle soit plus déterminée de l'œil humain, de telle façon les composants d'une scène devenant bien séparer.
- **Compression:** cette technique agit sur la taille de l'image dans le but de la réduire pour simplifier son traitement et son stockage.

VI.3. Extraction des caractéristiques

C'est l'ensemble de caractéristiques (descripteurs) qui doit être extrait à partir de l'image avant de passer à la classification et l'interprétation de l'image dans le but de réduire le volume de données. L'objectif principal de l'extraction de caractéristiques est d'identifier les meilleurs descripteurs afin de discriminer entre les classes. Cette phase est une phase primordiale dans le processus de SAD car le bon choix des caractéristiques a une influence directe sur :

- La taille de la mémoire.
- La précision de la classification.
- Le coût de la classification.
- La robustesse du système SAD.

Dans les méthodes classiques d'apprentissage automatique telles que (SVM, KNN, ...) les deux phases d'extraction des caractéristiques et la classification sont séparées mais dans les méthodes récente d'apprentissage profond telle que les CNNs, l'extraction et la classification sont regroupés comme une boîte noire où l'extraction des caractéristiques pertinentes est assurée d'une manière automatique.

VI.4. La classification des images médicales

La classification consiste à regrouper les images dans une ou plusieurs classes homogènes (catégorie). Où chaque classe rassemble des images ayant des caractéristiques aussi similaires que possible et les classes obtenues doivent être aussi distinctes que possible [11].

VI.4.1. Méthodes supervisées

La classification supervisée est un traitement basé sur la connaissance préalable des classes présentes dans l'observation. Par conséquent, il a besoin d'une phase d'apprentissage. Le but de cette classification est de concevoir une machine capable d'affecter toute observation inconnue à une classe déjà connue.

VI.4.2. Méthode non supervisée

Le but de la classification non supervisée est de détecter des groupes ou des catégories dans la base d'apprentissage sans aucune connaissance au préalable des classes. Par conséquent, il nécessite moins d'entrées initiales.

VII. L'imagerie médicale :

Le diagnostic médical a été facilité par l'introduction de nouvelles technologies et principalement l'imagerie médicale et ses outils; qu'ils soient électroniques ou informatiques pour mesurer et évaluer le corps humain dans le but de détecter des maladies au stade le plus précoce que possible.

De nos jours, l'imagerie médicale est utile tout au long du cycle de diagnostic; du dépistage jusqu'à l'évaluation [12], et elle a bénéficié de progrès technologiques rapides et continus qui ont multiplié la puissance des machines.

Au contraire que pensent la plupart des gens, l'imagerie médicale n'est pas seulement une photographie de l'organe étudié mais aussi une représentation visuelle des caractéristiques physiques ou chimiques particulières des tissus, ce qui nécessite l'utilisation d'un appareillage spécifique.

L'axe principal de l'imagerie médicale est l'image, c'est pour cela dans le domaine de traitement d'images on doit passer par la notion de l'image médicale et l'image numérique :

VII.1. Définition d'une image médicale :

Le mot "image" dérivant du latin " imago" désigne la reproduction visuelle d'un objet par différents outils (dessin, photographie, peinture,...). Selon MEZIANE [13], l'image médicale est la représentation des variations spatiales et temporelles d'une ou de plusieurs propriétés physiques d'un sujet sous investigation, si la variation temporelle de la propriété est prise en considération on parle de l'image dynamique sinon c'est l'image statique.

• C'est quoi l'image numérique ?

L'image numérique est un ensemble régulier d'éléments appelés « pixels » (contraction du terme anglo-saxon « picture elements ») et elle-même généralement appelée image « bitmap » (contraction du terme anglo-saxon « bits mapped»). La représentation la plus courante est celle d'une trame régulière de points selon deux axes orthogonaux qui forment une représentation sous forme de matrice, mais d'autres représentations existent comme l'échantillonnage selon une trame hexagonale ou même triangulaire [14].

En résumé, Le terme d'image numérique désigne, dans son sens le plus général, toute image qui a été acquise, traitée et sauvegardée sous une forme codée représentable par des nombres (valeurs numériques de 0 et des 1).

Pour être impliqué dans le domaine de l'imagerie médicale, il faudrait avoir des connaissances sur l'anatomie humaine, la physiologie et la pathologie et aussi les techniques de positionnement et de projection.

Comme notre thèse se concentre sur les images médicales, dans ce qui suit on va aborder les systèmes d'aide au diagnostic médicaux basés images.

VII.2. Les moyens physiques des différents procédés utilisés en imagerie médicale.

VII.2.1. Radiographie (rayons-X) :

La radiographie est un examen d'imagerie médicale qui utilise les rayons x pour visualiser l'os (arthrose, rhumatisme, fracture, tumeur, le thorax et l'abdomen) [15]. Lors de l'imagerie avec des rayons X, une installation radiologique est nécessaire pour obtenir les informations du corps entier sous différents angles. Lors de l'imagerie avec des rayons X, un faisceau de rayons X produit par un tube à rayonnement traversant le corps. Sur sa trajectoire, des l'énergie du faisceau de rayons X sont absorbées partiellement provoquant une atténuation du faisceau. Sur le côté opposé du corps, des détecteurs (le film capture les rayons atténués) donnant une image clinique en 2D. En

tomodensitométrie, le tube et les détecteurs tournent autour du corps de telle façon plusieurs images peuvent être acquises, résultant une vue en 3D.

En radiographie utilisant des rayons X l'injection de substances pour le contraste est nécessaire des fois. Les images obtenues sont des projections des organes et sur un plan. Généralement, rayon X est utilisé pour les os car ils sont les plus visibles.



Figure 1.2 : Radiographie (rayons-X) [15]

VII.2.2. La mammographie (Computed Tomography (CT)) :

La mammographie (appelée aussi mastographie) est une **radiographie du sein**, qui permet d'obtenir des images de l'intérieur du sein et de détecter les anomalies, parmi eux les cancers qui est la cause la plus fréquente de décès des femmes [16].

La radiographie des seins considérée comme l'examen le plus efficace pour dépister le cancer le plus tôt possible même en absence de ses symptômes. Il sert à visualiser des structures à l'intérieur du sein et de détecter les petites lésions difficiles à repérer par la palpation.

C'est un examen d'une durée de quelques minutes qui est réalisé par une technologue qui met la patiente face à l'appareil de mammographie et place un sein entre deux plaques qui réalisent une compression graduelle et contrôlée sur le sein, la compression, même s'il elle est gênante, elle est nécessaire pour obtenir des images de bonne qualité.

Dans la majorité des cas, deux clichés sont effectués par sein, un de face et l'autre en oblique.

Les films sont ensuite développés, par des développeuses automatiques conçues spécialement à la mammographie, ils sont un peu lentes afin d'obtenir une qualité d'image radiographique idéale),



Figure 1.3 : La mammographie [16]

VII.2.3. Scanographie (Tomodensitométrie (TDM)) :

Les tomodensitomètres (TDM) (ou les scanners) inventés en 1970. Ils sont souvent utilisés dans les urgences, car ils fournissent rapidement des informations détaillées, chose très utile pour prendre des décisions thérapeutiques convenables[15].

Le plus important dans TDM est le portique - un cadre rotatif circulaire avec un tube à rayons X monté d'un côté et des détecteurs de l'autre côté. Les faisceaux de rayons X en forme d'un éventail est créé lorsque le cadre tournant fait pivoter le tube à rayons X autour du patient. Les scanners produisent des images beaucoup plus détaillées que celles des rayons X classique, surtout dans le cas des vaisseaux sanguins et des tissus mous.



Figure 1.4: Scanographie [15]

VII.2.4. Ultrasons (Ultrasonographie ou échographie)

Ce type d'imagerie permet d'obtenir des images du corps par l'usage d'ondes sonores de hautes fréquences. Les images sont obtenues en temps réel. Ce procédé n'implique aucune irradiation.

Ultrasonographie est très utile pour examiner les organes tels que le foie, la vésicule, le pancréas, des reins et la vessie. Du fait qu'elles sont acquises dynamiquement, les images peuvent montrer les mouvements des différents organes internes. Le flux des vaisseaux sanguins peut être détecté par une échographie Doppler [15].

Les appareils échographiques sont composés d'un ordinateur et un écran, plus des sondes qui sont appliquées à la surface du corps en utilisant un gel spécial.



Figure 1.5: Radiologie Ultrasons (Ultrasonographie ou échographie) [16]

VII.2.5. Imagerie nucléaire (scintigraphie ou «gammagraphie»)

Le principe de l'imagerie nucléaire est d'injecter par voie intraveineuse une molécule marquée avec un isotope radioactif à vie brève afin de suivre, par détection externe, le fonctionnement d'un organe. Détection de photons émis par le traceur.

Cette technique consiste à capter le reflet des milliers de points, où le traceur s'est fixé dans l'organe et la radioactivité temporaire du tissu est détectée par une gamma-caméra, qui explore la zone d'intérêt [15].

Il faut noter que la dose doit être contrôlée de manière à minimiser les risques de l'irradiation.

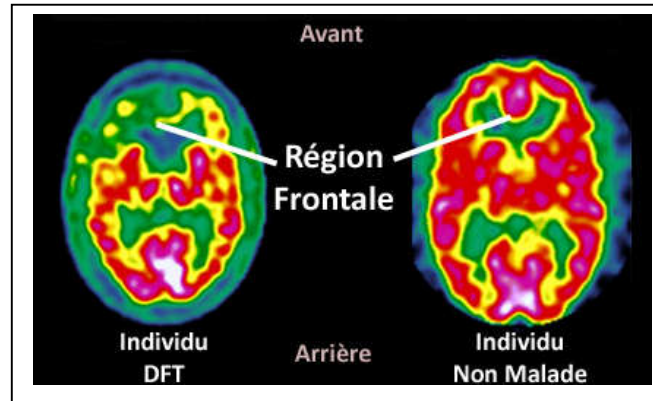


Figure 1.6: Exemple de la scintigraphie cérébrale [15]

VII.2.6. Imagerie par résonance magnétique (IRM)

L'imagerie par résonance magnétique (IRM) est un examen qui permet d'obtenir des vues en deux ou trois dimensions de l'intérieur du corps. Il donne des informations sur des lésions qui ne sont pas visibles sur les radiographies standards, l'échographie ou le scanner [15].

Cette technique explore l'organisme humain en se basant sur les modifications d'orientations (la résonance ou vibration) de certains atomes (hydrogène par exemple) soumis à un champ magnétique très intense. Elle est très utile en radiologie pour le diagnostic des pathologies cérébrales, abdominales et cardiovasculaires.



Figure 1.7: Diagnostic par résonance magnétique (IRM) [15]

VII.2.7. Magnéto/Electro-encéphalographie (M/EEG)

Le cerveau humain durant son travail génère des champs magnétiques issus de l'activité électrique des neurones. La magnétoencéphalographie (MEG) est une technique qui mesure ces champs magnétiques [17].



Figure 1.8: Imagerie par magnéto-et électro-encéphalographie (MEG/EEG) [17]

Elle n'est pas un outil de clinique répandu mais elle est utilisée dans certains diagnostics pré-opératoires en épilepsie.

Récemment, la MEG s'est utilisée pour le diagnostic précoce et le suivi de certaines pathologies comme la maladie de Parkinson ou la maladie d'Alzheimer, Elle est utilisée aussi, pour assister les neuro-chirurgiens dans certaines stratégies d'intervention chirurgicales.

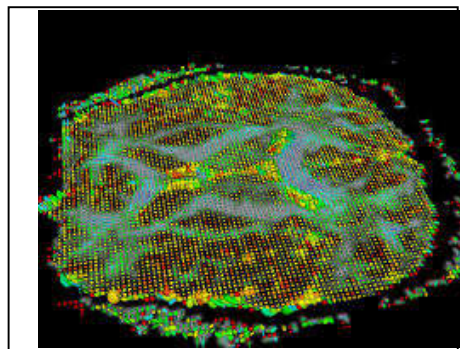


Figure 1.9: Imagerie du système nerveux central [17]

VII.2.8. Rétinographies (mydriatique et non mydriatique RNM)

La rétino-graphie utilise un appareillage spécial appelé « *le rétino-graphe* » ou bien aussi « *fundus camera* », pour photographier le fond d'œil tout ce qui est en arrière de l'iris et du cristallin y compris la rétine, la papille optique et la macula.

On utilise le rétino-graphe pour vérifier les modifications douteuses du fond d'œil, tel que le décollement de la rétine, la rétino-pathies diabétique, l'affections vasculaires, l'inflammation ou autres maladies oculaires telles que le glaucome [18].

Si l'examen à l'aide de rétino-graphe est réalisé sans dilatation pupillaire il est appelé Non Mydriatique (Rétino-graphie Non Mydriatique RNM) sinon l'examen est RM (Rétino-graphie Mydriatique)



Figure 1.10: La Rétino-graphie Non Mydriatique (RNM) [18]

VIII. Diagnostic des images rétinienne

VIII.1. Description de l'œil humain :

La complexité de la vision humaine, vient du fait que l'organe sensoriel (la rétine) permet de recevoir un signal électromagnétique qui est la lumière et lui transmette à un signal dit biologique qui est une dépolarisation axonale transmissible au cerveau à travers le nerf optique;

Ce signal est traduit par le cerveau, au niveau du cortex visuel et renvoie l'image traitée ce qui permet l'interprétation d'environnement autour de nous [19].

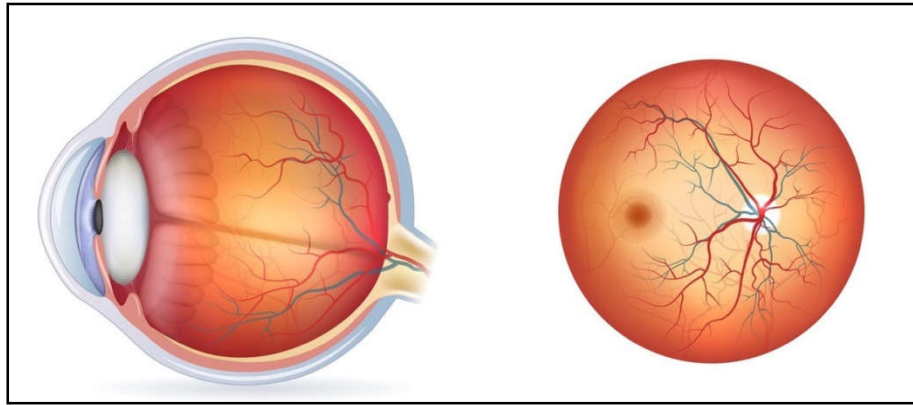


Figure 1.11: Coupe transversale et vue de fond de l'œil [19]

VIII.2. Anatomie de l'œil humain :

L'œil est une boule d'un diamètre de 25mm environ, et qui pèse approximativement 8 grammes.

Il est composé de plusieurs organes:

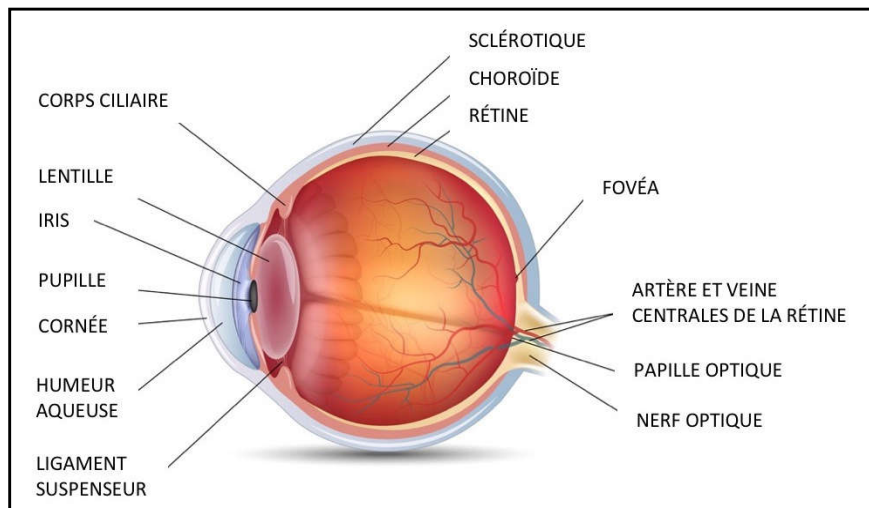


Figure .12: Anatomie de l'œil [19]

- **Le corps ciliaire :**

C'est une partie située en antérieure de la choroïde, sur laquelle le cristallin est attaché par des fibres appelées la zonule de Zinn. Son rôle est la sécrétion d'humeur aqueuse et l'accommodation de la vision.

- **L'iris :**

Est un tissu en forme de disque perforé au centre par la **pupille**, c'est la partie colorée de l'œil dont la couleur dépend de l'épaisseur des lamelles de l'épithélium de l'œil c'est-à-dire- plus les lamelles sont fines l'iris est clair et le plus les lamelles sont épaisses la couleur devient plus foncée.

L'Iris est toujours influencée par la lumière ce qui déclenche un mécanisme physiologique de contraction ou dilatation ce réflexe permet à l'iris de contrôler la quantité de la lumière (si la lumière est forte, la pupille est contractée (myosis), si la lumière est faible la pupille devient grande afin de capter un maximum de lumière (mydriase)).

- **La cornée :**

Est un tissu transparent situé sur la partie antérieure de l'œil, il permet de transmettre la lumière au cristallin et à la rétine. Il est composé de 5 couches (Epithélium, Membrane de Bowman, Stroma, Membrane de Descemet, Endothélium), il ne contient aucun vaisseau ce qui explique son non vieillissement mais il contient beaucoup de nerfs ce qui explique sa très grande sensibilité, et le "réflexe cornéen" qui se manifeste par une fermeture brusque des paupières pour protéger les yeux en cas de toucher de la cornée. Elle se nourrit en permanence par les larmes et l'humeur aqueuse.

- **L'humeur aqueuse :**

C'est une sorte de liquide transparent dont le rôle est d'apporter les nutriments à la cornée et au cristallin. Ce liquide sert à maintenir la pression intra-oculaire stable et la forme de globe oculaire.

- **La sclérotique, ou sclère :**

C'est le blanc de l'œil elle est très résistante.

- **La choroïde :**

C'est le tissu du globe oculaire, il est plein de vaisseaux, c'est la membrane nutritive de l'œil.

- **La rétine :**

C'est une membrane mince qui recouvre la surface interne du globe oculaire elle contient beaucoup de vaisseaux. Elle est sensible à la lumière par sa composition de photorécepteurs appelés « cônes » et « bâtonnets » qui convertissent la lumière reçue en signaux électriques transmissibles par les neurones qu'ils lui sont reliés. Au centre de la rétine on trouve la **macula** et la **fovéa**.

- **Le nerf optique :**

Appelé aussi le deuxième nerf crânien, il commence de la **papille optique** et permet d'envoyer les informations visuelles reçus de la rétine vers le cerveau.

VIII.3. Différents types de maladies de la rétine

Vue les anomalies survenant sur la rétine, de nombreuses pathologies peuvent atteindre cette zone telle que: DMLA, rétinopathie diabétique, rétinites pigmentaires et glaucome [20].

- **Dégénérescence maculaire liée à l'âge**

L'infection par cette maladie est due à un problème dans la macula (la zone centrale de la rétine qui transmet l'essentiel de l'information visuelle au cerveau). Si le patient arrive à un stade avancé, une tache aveugle apparaîtra au centre du champ de vision. Cette maladie se présente sous deux formes: la forme «atrophique » et la forme «exsudative ». Les facteurs de risques de cette maladie est l'âge, les problèmes génétiques, le tabagisme et même l'exposition au soleil.

- **Rétinopathie diabétique**

Le diabète peut affecter la santé des yeux car la concentration de sucre dans les petits vaisseaux sanguins de la rétine peut entraîner des complications cela est appelé rétinopathie diabétique. L'hypoxie (insuffisance d'oxygène) peut fragiliser les vaisseaux sanguins et les déchirerai ce qui peut provoquer des micro-saignements qui entraînent un véritable décollement de la rétine. Un diagnostic tardif peut causer une perte de vision.

- **Maladies héréditaires de la rétine**

Plusieurs maladies héréditaires rares peuvent toucher la rétine tels que :

- Rétinite pigmentaire
- amaurose congénitale de Leber ou maladie de Stargardt

- **Maladie de Glaucome**

Le glaucome est une maladie oculaire très grave qui peut endommager le nerf optique. La principale cause de cette maladie est la pression intraoculaire élevée dans l'œil causée par la quantité de liquide excessive et un mauvais fonctionnement du système de drainage. Ce fluide interne s'écoule généralement à travers un tissu appelé maillage trabéculaire [20].

Malgré que le glaucome peut survenir à tout âge, mais selon les statistiques de l'OMS (l'Organisation Mondial de la Santé), le glaucome est l'une des principales causes de cécité chez les personnes âgées de plus de 60 ans [20].

C'est une maladie sévère car la majorité de ses formes ne présentent aucun signe avant-coureur, autrement dit, le patient ne remarque jamais de changement dans la vision avant que son état n'atteigne un stade avancé, où la perte de vision n'est pas récupérable. C'est pour cela il est conseillé de passer régulièrement des examens des fond d'œil afin d'avoir un diagnostic précoce et traiter le glaucome de manière appropriée [5].

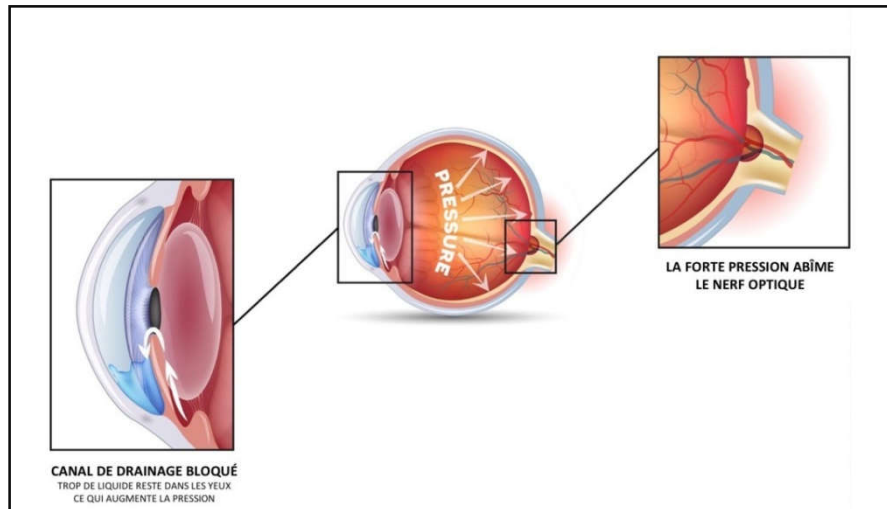


Figure 1.13: Maladie de glaucome [20]

Il existe deux types principaux de glaucome :

- **Le glaucome à angle ouvert**

Ce type est considéré comme la forme la plus courante car il se produit en raison d'une mauvaise circulation de l'humeur aqueuse (le fluide contenu dans le globe oculaire) à travers les membranes oculaires. Cette interférence augmente progressivement la pression intraoculaire et endommage le nerf optique. Ce déséquilibre peut survenir pendant une longue période et le patient peut soudainement perdre la vue avant de savoir qu'il y a un problème avec ses yeux.

- **Le glaucome à angle fermé**, il apparaît lorsque le drainage des fluides devient impossible à cause du rapprochement de l'iris et la cornée. Il est considéré comme une maladie chronique. L'obstruction peut survenir en quelques semaines ou années.

Nous nous intéressons dans ce travail aux systèmes d'aide aux diagnostics basés sur les nouvelles approches d'apprentissage profond pour la classification des images médicales de fond d'œil en glaucome et non glaucome.

VIII.4. Le fond d'œil humain

La rétine est la couche de tissu sensible à la lumière qui capture l'image du monde visuel tout comme le film dans une caméra. Le cercle noir visible au centre de nos yeux s'appelle pupille et en son centre là est un petit trou appelé iris. La lumière passe à travers l'iris, qui forme une image de la scène sur la rétine en initiant une cascade d'événements chimiques et électriques. Ces signaux

d'événements sont envoyés à divers centres du cerveau à travers les nerfs optiques. Le cerveau comprend enfin ces signaux et l'image se forme dans notre esprit. Les images numériques de la rétine sont capturées à l'aide de deux techniques: la photographie du fond d'œil et la cohérence optique Tomographie (OCT). La photographie du fond d'œil consiste à capturer l'image de la partie arrière de l'œil, c'est-à-dire couche de tissu rétinien qui capte la lumière. Ce tissu rétinien est également appelé fond de l'œil et ces termes sont interchangeables. Désormais, nous appellerons les images de la couche de tissu rétinien des images rétiniennes. La pupille est généralement dilatée (agrandie) avant de capturer les images de la rétine (du tissu du fond d'œil). Cela permet plus de lumière pour passer à travers l'iris et donne des images de la rétine avec un SNR (L'indice global d'affaiblissement) plus élevé [21].

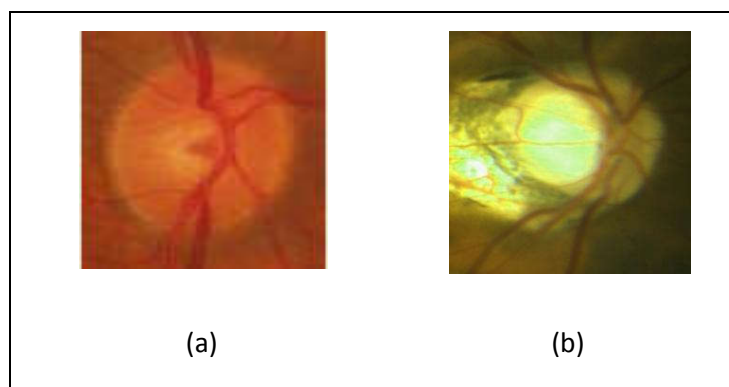


Figure 1.14 : Échantillons d'images de fond d'œil normal(a) et anormal(b) [34].

IX. Les travaux connexes des systèmes d'aide au diagnostic médical

Nous voulons à travers le tableau suivant illustrer les principaux travaux qui ont été proposés comme des systèmes d'aide au diagnostic médical (SADx) dédiés aux maladies de fond d'œil. Cette étude a pour objectif d'analyser, d'une manière générale, les méthodes classiques de classification (ANN, k-NN, SVM...), ainsi les nouvelles techniques d'apprentissage profond qui se basent sur l'extraction automatiques des caractéristiques les plus pertinentes (les CNNs).

La dernière colonne de ce tableau illustre les résultats en termes de la précision (accuracy (Acc) et de l'air sous la courbe ROC (AUC)

Tableau 1.1 travaux connexes des systèmes d'aide au diagnostic médical

Références	L'année	BDD	Architecture	Résultats
Mat et al. [21]	2008	EGPS	ANN	Acc = 87.5%
Nor et al. [22]	2014	Collected from the Ophthalmology department of KMC, Manipal, India.	SVM, Naive bayes	Acc= 84,72%
Tan et al. [23]	2014	ORIGA dataset	SVM	Acc= 95%
Sing et al. [24]	2015	collected from the Venue Eye Research Center, New Delhi, India.	Random forest, ANN, k-NN, SVM	Random forest and ANN: Acc= 94.75% k-NN, SVM: Acc=94.75%
Che et al. [25]	2015	ORIGA and SCES	CNN(6 layers)	ORIGA_AUC=0.831 SCES_AUC=0.887
Mah et al. [26]	2017	private data set	LS-SVM (RBF)	Acc = 94.79 %
Zil et al. [27]	2017	DRISHTI-GS data set	CNN	Acc =92,3%
Ben et al [28]	2017	RIM-ONE	Multimodal CNN classifier	Cnn1:Acc = 0.956 Cnn2:Acc = 0. 969
Al-band et al. [29]	2017	RIM-ONE	CNN-SVM	Acc = 88.2 %
Acharya et al. [30].	2017	Manipal, India.	KNN	Acc = 95.7 %
Rag et al [31]	2018	Kasturba Medical College, Manipal, india	CNN(18-layers)	Acc = 98.13%
Chai and al. [32]	2018	private data set	MB-NN	Acc = 91,51%
Tou et al[33]	2018	RIM-ONE	CNN(18-layers)	Acc =97,26%
Tou et al[34]	2019	RIM-ONE	Multimodal CNN classifier	CNN1:Acc =95.56% CNN2:Acc =90.37% CNN3:Acc =77.78% CNN4:Acc =91.85% CNN-Sum:Acc =96.6% CNN-Prod:Acc =97.8%

Nous pouvons remarquer que de nombreux travaux de recherche sont concentrés sur l'analyse des images du fond rétinien pour diagnostiquer la maladie du glaucome. Nous voyons également que les méthodes d'apprentissage profond basées CNN sont très bien sollicitées grâce à leurs capacités de générer des bons résultats.

Aujourd'hui, les chercheurs ont penchés vers l'utilisation des techniques d'apprentissage profond pour la classification et la segmentation des images et particulièrement pour les images médicales afin de surmonter les problèmes d'extraction manuelle des meilleurs caractéristiques et pour assurer le diagnostic automatique des maladies en moins de temps, moins de couts et avec plus de précision.

X. Les Bases de données des images rétinienne à accès public

La disponibilité de bases de données d'images accessibles au public ouvre la porte devant le développement de stratégies de conception, la mise en œuvre et le test des algorithmes pour analyser la morphologie de la rétine et même pour diagnostiquer d'autres maladies. L'innovation dans les nouveaux paradigmes de l'apprentissage automatique tels que l'apprentissage profond, nécessitent une quantité importante de données à traiter. Cette nécessité a conduit plusieurs groupes de recherche à faire leurs propres ensembles de données d'images rétinienne accessibles au public. Cette tendance aboutie à la génération de bases de données avec une grande variété de signes pathologiques en haute résolution avec la segmentation des composants rétinien (le disque optique, les vaisseaux sanguins et la macula) accompagnée des instructions pour faciliter l'accès à l'utilisateur. Cela permet de comparer les performances des algorithmes au sein de la même image de fond.

Nous avons résumé les principales caractéristiques des différentes bases de données de la rétine dans le tableau suivant:

Tableau 1.2 Résumé des bases de données d'images du fond d'œil accessibles au public :

Le nom de la base	Nombre des images	Système de caméra de fond	Domaine d'utilisation
DRIVE (Utrecht Database) – Digital Retinal Images for Vessel Extraction	40	Canon CR5 (Canon Corp., Tokyo, Japan)	Segmentation des vaisseaux sanguins
REVIEW – Retinal Vessel Image Set for Estimation of Widths	16	HRIS – Canon 60 UV VDIS – Zeiss fundus camera CLRIS – Zeiss FF 450 (Carl Zeiss AG, Oberkochen, Germany) KPIS – Canon 60 UV (Canon Corp.)	Segmentation des vaisseaux sanguins
CHASE_DB1 – Child Heart and Health Study in England	28	Nidek NM-200-D (Nidek Co., Ltd., Gamagori, Japan)	Segmentation des vaisseaux sanguins
HRF – High Resolution Fundus Image Database	45	Canon CF-60 Uvi Canon EOS 20D (Canon Corp.)	Segmentation des vaisseaux sanguins- Détection de disque optique (OD)
STARE – Structured Analysis of the Retina	81	TopCon TRV-50 (Topcon Corp., Tokyo, Japan)	Segmentation des vaisseaux sanguins – Artery/Vein Labeling
VICAVR	58	TopCon camera NW-100 (Topcon Corp.)	Calibre des vaisseaux – Étiquetage Artère / Veine
INSPIRE- AVR	40	Zeiss Fundus Camera (Carl Zeiss AG)	étiquetage artère / veine Segmentation OD
ORIGA(-light) Online Retinal Fundus Image Database for Glaucoma Analysis	650		segmentation OD et OC et Cup-to-Disc Ratio (CDR)

BioImLab	60	TopCon TRC-50 (Topcon Corp.)	Évaluation de la tortuosité des vaisseaux
DRIONS-DB – Digital Retinal Images for Optic Nerve Segmentation Database	110	35 mm Film Color Fundus Camera Digitized using HP-PhotoSmart-S20 scanner (HP, Palo Alto, CA, USA)	Segmentation de OD
RIM-ONE	159	Nidek AFC-210 (Nidek Co., Ltd.) Canon EOS 5D (Canon Corp.)	Segmentation OD and OC
ImageRet	219	Digital Fundus Camera – Settings Unknown	La rétinopathie diabétique (DR) Lesions Grading
MESSIDOR	1200	TopCon TRC NW6 (Topcon Corp.)	La rétinopathie diabétique (DR) Macula Edema
E-Ophtha	463	Telemedical Network Various Cameras used	Classement des lésions DR
ARIA – Automated Retinal Image Analysis	450	Zeiss FF450+ Fundus Camera (Carl Zeiss AG)	Segmentation de Limite de la tête du nerf optique
Reffuge	1200	Zeiss Visucam 500 Canon CR-2	Classification du glaucome Segmentation du OD et OC Localisation de la fovéa
Drishti-GS	101	----	segmentation OD et OC

XI. L'augmentation des données (data augmentation)

Les réseaux de neurones convolutifs fonctionnent parfaitement dans de nombreuses tâches de vision par ordinateur. Cependant, ces réseaux s'appuient fortement sur un grand jeu de donnée (big data) pour éviter le problème du sous-apprentissage. D'une autre façon, plus qu'on a une grande quantité de données, plus l'efficacité globale du système sera améliorée. Malheureusement, de nombreux domaines d'application n'ont pas accès aux méga données. autrement dit, malgré l'existence des bases de données publiques (Tableau 1.2), la quantité des ensembles de données est limitée, ce qui ne fourni pas un meilleur apprentissage. C'est pour cela des nouvelles techniques d'augmentation de données sont mises en œuvre pour élargir la qualité et la quantité des données.

L'augmentation des données pour l'apprentissage comprend la génération de données artificielles pour enrichir la diversité, dans le cadre de l'amélioration de ses performances [35]. L'application des techniques d'augmentation de données devient très simple avec l'utilisation des classes prédéfinies dans TensorFlow et Keras.

- **Les transformations géométriques**

Différentes techniques d'augmentations basées sur des transformations géométriques sont reconnues par leur facilité de mise en œuvre, afin de fournir une base utile pour l'apprentissage profond.

Nous décrivons également diverses améliorations géométriques dans le cadre de la «sécurité» applicative. La sécurité des méthodes d'augmentation de donnée fait référence à sa probabilité de conservez l'étiquette de l'image après le traitement. Par exemple, la rotation et le flip sont généralement sans danger pour l'ensemble ImageNet (tels que les chats et les chiens), mais pas pour les tâches de reconnaissance de texte numérique (telles que 6 et 9). C'est pour cela le défi de développer des politiques d'augmentation généralisables est mis au point. Pour garder la référence de l'étiquette [36]. Dans la section suivante en va présenter quelques techniques géométriques d'augmentation de données illustrées par des exemples:

- **Flipping (Retournement)**

Le basculement sur l'axe horizontal est beaucoup plus courant que le basculement à l'axe vertical. Cette augmentation est l'une des plus faciles à mettre en œuvre et s'est avérée utile sur des ensembles de données tels que CIFAR-10 et ImageNet. Et moins utile pour MNIST ou SVHN, car elle ne préserve pas les étiquettes.



Figure 1.15 : Flip horizontal (a) et flip vertical (b) [36]

- **Espace colorimétrique**

Les données d'images numériques sont généralement codées sous forme de tenseur de taille (hauteur \times largeur \times canal de couleur). Augmenter l'espace des canaux de couleur est une autre stratégie très pratique dans l'augmentation de données. Un ajout de couleur très simple consiste à isoler un seul canal de couleur, tel que R, V ou B. Une image peut être rapidement convertie en son représentation dans un canal de couleur en isolant cette matrice et en ajoutant 2 matrices nulles des autres canaux de couleur. De plus, les valeurs RVB peuvent être facilement manipulées avec des opérations matricielles simples pour augmenter ou diminuer la luminosité de l'image [36].

Exemple : la luminosité et Changement de canal



Figure 1.16 : effet de Luminosité [36]



Figure 1.17 : effet de changement de canal [36]

- **Recadrage (Cropping)**

Le recadrage des images est un traitement très pratique pour l'augmentation de données avec des dimensions de hauteur et de largeur en recadrant un patch central de chaque image.

Exemple :



Figure 1.18 : Recadrage [36]

- **Rotation**

Les augmentations de rotation se font en faisant pivoter l'image à droite ou à gauche sur un axe entre 1° et 359° . La sécurité des augmentations de rotation est fortement déterminée par le paramètre de degré de rotation.



Figure 1.19: effet de Rotation [36]

- **Le décalage (*Translation*)**

Décalage des images vers la gauche, la droite, le haut ou le bas peut être une transformation très utile pour éviter de se concentrer sur un seul emplacement de données dans l'image.



Figure 1.20 : Décalage à gauche [36]

- **L'Injection de bruit**

L'injection de bruit a été testée la première fois par Moreno-Barea et al [37], cette technique consiste à injecter une matrice de valeurs aléatoires généralement tirées d'une distribution gaussienne. L'ajout de bruit aux images peut aider les CNN à apprendre des fonctionnalités plus robustes.

- **Le Zoom**

Zoomer une image permet de bénéficier mieux de ces caractéristiques et préserver la référence de l'étiquette.



Figure 1.21 : effet de zoom [37]

XII. Conclusion

Au cours des cinquante dernières années, les SADM ont été développés par vague dans le but d'automatiser le processus de diagnostic de plusieurs maladies. Malgré que les différentes techniques d'imagerie sont de plus en plus robustes et rapides telles que l'échographie, la mammographie et l'IRM, la phase de diagnostic et l'identification des données provenant de l'image sont encore sous la responsabilité des médecins et le personnel médical. L'appariation des méthodes de reconnaissance d'organe/tissus par ordinateur peuvent faciliter les taches de staffe médical et permet d'avoir des résultats en moins de temps et avec plus de précisions.

Vu que les techniques de l'Intelligence artificielle sont très impliquées dans la prise de décision de diagnostic de la maladie du glaucome, dans ce cas le chapitre suivant va introduire les outils de l'IA et nous nous pencherons plus en détail sur l'apprentissage automatique et particulièrement les techniques d'apprentissage profond qui se voit prometteuse dans le domaine d'imagerie médicale.

***Chapitre II : Les techniques d'IA et
l'Apprentissage Profond***

I. Introduction

L'Intelligence Artificielle IA consiste à mettre en œuvre des systèmes informatiques pouvant imiter l'intelligence humaine et même de prendre des décisions.

Le terme Apprentissage automatique ou Machine Learning (ML) date du milieu du 20^{ème} siècle. En 1950, le mathématicien britannique *Alan Turing* a proposé la première machine capable d'apprendre et de simuler l'intelligence réelle. Depuis, plus d'un demi-siècle, les techniques d'IA sont devenues pluridisciplinaires et reconnues un développement considérablement dans le but de créer des algorithmes capables d'apprendre et de s'améliorer de manière autonome. Ce qui fait l'apparition de l'apprentissage profond **qui est un algorithme de haut niveau d'abstraction qui permet de modéliser des grands ensembles de données.**

Récemment, les ordinateurs modernes ont fait de grands progrès sur le plan matériel et logiciel, en plus un développement élevé des technologies en terme d'acquisition d'images (microscopes, caméras, capteurs) et de stockage engendrent des bases de données riches en information, qui facilite la recherche et augmentes les taux de réussite des systèmes tout en diminuant l'intervention humaine.

Les réseaux d'apprentissage profond (Deep Learning) dérive de l'apprentissage automatique qui été largement utilisé dans différents domaines quelque soit leur degrés de complexité grâce à son avantage d'extraction automatique des caractéristiques la chose pour laquelle il a prouvé son efficacité par rapport aux méthodes d'apprentissage classique.

L'objectif principal de notre thèse est l'analyse de comportement de Deep Learning pour améliorer l'interprétation des images rétiniennes en utilisant deux approches différentes basées CNN ; une approche globale et une approche locale qui seront respectivement détaillées dans les chapitres 3 et 4, raison pour laquelle ce chapitre est consacré à l'illustration des principaux concepts de l'apprentissage profond en commençant par le classifieur noyaux qui est le réseau de neurones, passant par l'apprentissage profond et ses domaines d'application, ensuite nous allons présenter et détailler les réseaux de neurones convolutifs et leurs architectures dont nous allons concentrer sur la manière d'ajustement des hyperparamètres. Nous citons aussi les fonctions d'activation.

Dans la dernière partie, Nous allons clarifier le domaine d'interprétation des images qui englobe trois tâches essentielles (Classification, détection, segmentation) où nous avons analysé quelques architectures connues en classification, en détection des objets et en segmentation dont nous allons mis en revue les travaux récents des chercheurs qui sont présents dans différentes architectures dans ce domaine.

II. Concept d'Intelligence Artificielle (IA)

Le terme **Intelligence Artificielle** (IA, ou AI en anglais pour Artificial Intelligence) est créé par le mathématicien **John McCarthy** en 1956 [38]. Dans le but d'élaborer des algorithmes capables de prendre en charge les tâches habituellement de l'être humain basant sur l'apprentissage.

Selon le dictionnaire français, l'intelligence artificielle est un ensemble de théories et de techniques mises en œuvre en vue de réaliser un système informatique qui simule l'intelligence humaine. Autrement dit, l'IA est un domaine très vaste qui s'intéresse à la construction des machines intelligentes capables d'effectuer des tâches qui nécessitent généralement une intelligence humaine. Un système est qualifié intelligent, s'il est capable d'apprendre et d'exécuter une tâche liée au processus auquel il a été affecté sans aucune intervention humaine et avec une grande précision. L'intelligence artificielle a été utilisée dans une variété de domaines tels que : Finance et banques, Militaire, Médecine, Renseignement policier, Droit, Logistique, Transports, Industrie et Robotique.

III. L'évolution des réseaux de neurones :

Depuis leur apparition, les réseaux de neurones ont connu une évolution remarquable jusqu'à ce jour. Les fondateurs des réseaux de neurones sont les neurologues **Warren McCulloch** et **Walter Pitts** après leurs travaux publiés dans le célèbre article : " What the frog's eye tells to the frog's brain". Où ils ont présenté un modèle mathématique simplifié de neurone biologique souvent appelé neurone formel.

En 1949, **Donald Hebb** a introduit une règle qui permet de modifier les valeurs des coefficients synaptiques en fonction des unités qui les relient. Cette règle est très utilisée dans le modèle actuel elle est même appelée « règle de Hebb ».

Ainsi, Le premier système artificiel de perceptron qui a la possibilité d'apprendre fut introduit en 1957 par **Franck Rosenblatt**.

L'année 1982, marque l'apparition du premier modèle de réseau complètement récurrent qui a été réalisé par le physicien **John Joseph Hopfield**, permettant de résoudre la limitation des réseaux de neurone envers les problèmes non-linéaires.

En 1986, pour le traitement des événements non-linéaires une approche de perceptron multicouches a été introduite par **Rumelhart**. Par la suite **Werbos**, **Yann le Cun**, **Marvin Minsky** ont proposé des systèmes semblables au perceptron de Rumelhart et qui sont tous basés sur la rétropropagation du gradient [39].

IV. Les réseaux de neurones

IV.1. Le neurone biologique

Le neurone est un type particulier de cellule du corps humain. La combinaison de ces neurones avec les cellules gliales constitue le tissu nerveux. Les neurones représentent l'unité fonctionnelle du système nerveux, il y a environ neuf fois plus de cellules gliales que de neurones dans le cerveau, son rôle est de faciliter la connexion des neurones et assurer leur alimentation avec un grand soutien. Les neurones sont les éléments de base du système nerveux central (cerveau) [39], le cerveau humain comprend environ 100 milliards de neurones, où ses derniers assurent la transmission d'un signal (influx nerveux). Elles sont composées de trois parties (Figure 2.1), le corps cellulaire, les synapses et l'axone :

- Corps (Somme) : Elle contient le noyau du neurone, c'est là où se déroulent toutes les activités nécessaires à la survie de la cellule.
- Axone: son rôle principal est d'assurer la communication avec les autres neurones.
- Synapse : C'est une jonction entre les terminaisons axonales et les autres cellules. L'impulsion nerveuse est une manifestation de la communication intercellulaire.

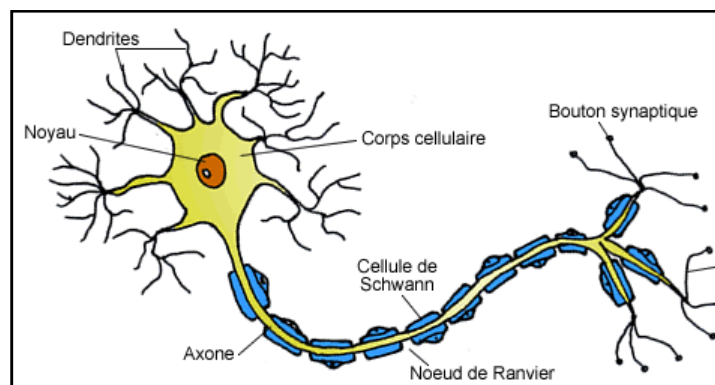


Figure 2.1 : Schéma simplifié d'un neurone biologique [39].

IV.2. Neurone formel :

Un neurone formel est un modèle mathématique non linéaire, paramétré, à des valeurs bornées. Il est inspiré de son équivalent biologique et composé des signaux d'entrées (représentant les dendrites) et d'un signal de sortie (représentant l'axone) ainsi d'une fonction d'activation (représentant le soma).

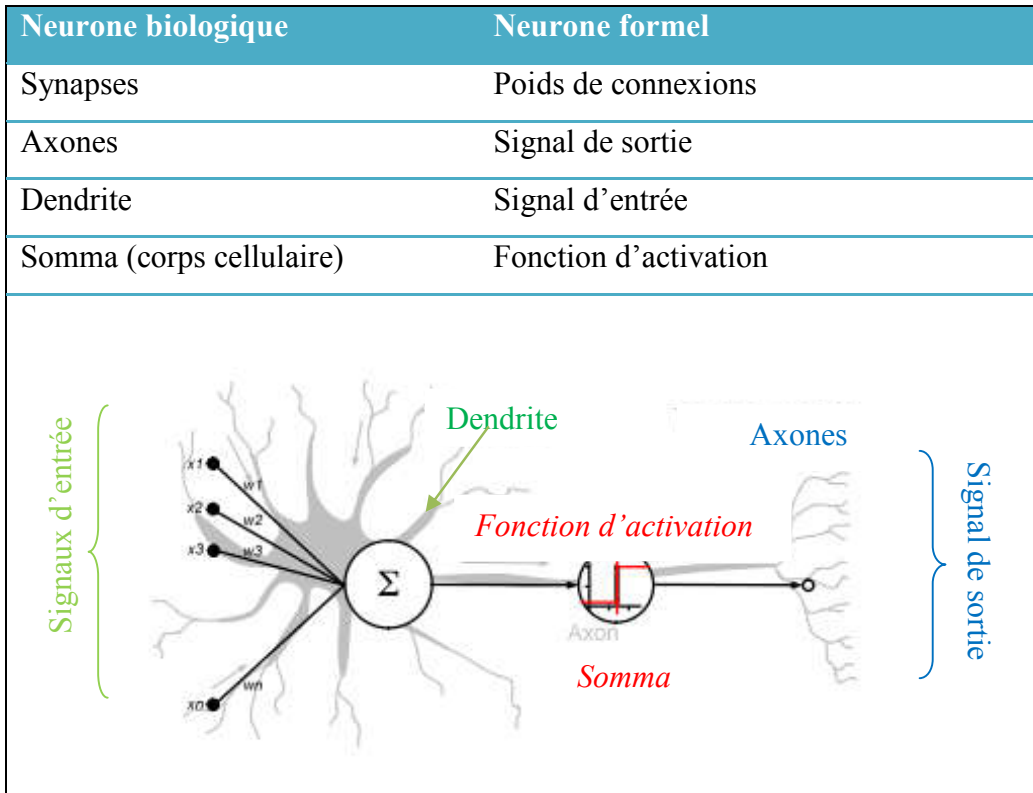


Figure .2.2 : Comparaison entre neurone biologique et neurone formel [39]

Le neurone traite ses entrées et communique la sortie qui peut être le résultat final du système ou l'une des entrées d'un autre neurone ce qui forme un réseau neuronal.

Chaque neurone est défini par trois caractéristiques principales: son état, ses connexions avec d'autres neurones et sa fonction d'activation

IV.3. Réseau de Neurones Artificiels

Un Réseau de Neurones Artificiels (RNA) est un ensemble de neurones formels organisés en couches fonctionnant en parallèle, Où chaque neurone artificiel est un processeur élémentaire qui reçoit un nombre variable d'entrées en provenance d'autres neurones. Chaque entrée possède un poids ' w ' qui représente la force de la connexion. Chaque neurone est doté d'une sortie unique, qui peut être utilisé ensuite pour alimenter un nombre variable de neurones avals [40].

V. L'apprentissage Automatique

L'apprentissage est un concept basé sur plusieurs mécanismes qui nous permet l'acquisition de connaissances et leurs réutiliser ultérieurement. Pour nous, les êtres humains, notre apprentissage se fait quotidiennement tout au long de notre vie par la collecte d'informations de l'environnement avec nos cinq sens, ainsi les expériences acquises de la répétition des événements dans la vie, et de notre jugement.

Pour faciliter nos recherches, nous avons créé l'ordinateur, qui permet de résoudre en quelques seconds des calculs qui nous prendraient des années de travail [41].

L'apprentissage automatique, qu'il soit *supervisée* ou non *supervisée*, est un sous domaine de l'intelligence artificielle (IA) son objectif est de faire entraîner des modèles pour résoudre des problèmes à partir d'un ensemble de données [42]. Les algorithmes d'apprentissage automatique ont fourni des solutions très encourageantes à des problèmes complexes dans différents domaines : reconnaissance de la parole, traduction du langage naturel, véhicules autonome...etc.

Selon l'inventeur American de machine learning: Arthur Samuel [42], qui a défini l'apprentissage automatique en 1959, comme un outil qui permet à une machine d'avoir des connaissances sans être explicitement programmée et sans aucune intervention humaine. Après, en 1998, un autre chercheur américain au nom de Tom Mitchell [43] a donné une autre définition de l'apprentissage c'est quand la performance d'une machine augmente avec de nouvelles expériences on dit que cette machine apprend.

Il existe trois principaux paradigmes d'apprentissage automatique :

V.1. L'apprentissage supervisé (Supervised Learning) :

Consiste d'entraîner la machine avec des exemples x **étiquetés** pour avoir une relation avec leurs sorties désirées y selon la fonction ($f: x \rightarrow y$).

Quelques algorithmes d'apprentissage supervisé :

- Machine à vecteurs de support (SVM)
- Réseau de neurones (RNA)
- Méthode des k plus proches voisins (KPP)
- Arbre de décision
- Classification naïve bayésienne...

V.2. L'apprentissage non supervisé:

L'objectif principal de l'apprentissage **non supervisé** est de découvrir des modèles cachés dans les données d'où les exemples x ne sont pas **étiquetés**.

Quelques algorithmes d'apprentissage non supervisé :

- Clustering (segmentation, regroupement) : construire des classes automatiquement en fonction des exemples disponibles
- Règles d'association : analyser les relations entre les variables ou détecter des associations
- Réduction de dimensions

V.3. *L'apprentissage par renforcement*

Les entrées sont un ensemble de rétroactions provenant d'un environnement dynamique auquel l'algorithme est confronté. Le but est d'atteindre un objectif prédéterminé. C'est la méthode la plus populaire en robotique.

VI. *Historique d L'apprentissage profond*

D'après Yann LeCun [40], « *L'apprentissage profond ou le Deep Learning permet aux modèles de calcul composé de plusieurs couches d'apprendre des représentations de données avec plusieurs niveaux d'abstraction* ». Autrement dit, c'est une technique d'apprentissage automatique qui fait apprendre l'ordinateur ce que l'être humain le fait naturellement (apprendre par l'exemple) [7]. d'où les algorithmes d'apprentissage profond ont attiré récemment beaucoup d'attention car ils abouti à des résultats qui étaient irréalisables auparavant dans différents domaines de traitement d'images tel que : le domaine de détection d'objets [46], reconnaissance de visage [47], détection précoce de différentes maladies [48,49,50 ,51 ,52,53], tout en profitant des données massives (big data) et de la capacité de calcul phénoménale notamment grâce aux GPU et CPU [54].

l'apprentissage profond est un modèle informatique très puissant car il est capable d'effectuer des tâches de classification directement à partir des images, du son ou texte, comme entrée du système sous forme de matrice, ce qui donne l'avantage principal de ces modèles qui est l'extraction automatique des données brutes [43] .

VII. *Domaines d'application de l'apprentissage profond :*

Un model issu d'un apprentissage profond peut servir dans des situations réelles telle que :

- Reconnaissance des visages;
- Identification selon les mouvements de la marche;
- lecture des chiffres et des textes manuscrits;
- Reconnaissance et analyses des discours;
- Traduction des langues;
- Contrôler les voitures autonomes;
- Les comportements des robots (position/map estimation) ;
- Bioinformatique ;
- E-commerce (détection de fraude) ;
- Analyses financière
- Médecine (diagnostic des maladies) ;
- Web (les réseaux sociaux) ;

Le tableau suivant montre quelques applications qui ont été proposées comme modèles de classification des images basée CNN.

Tableau 2.1 – Les domaines d'application des CNNs .

Référence	Domaine d'application
[55]	Classification des caractères manuscrits
[56]	Classification des vidéos à grande échelle
[57]	Analyse des images médicales
[58]	Classification des images hyperspectrales
[59]	Classification des images histologiques du cancer du sein
[60]	Identification des plantes
[61]	Classification à partir des images radiologiques
[62]	Reconnaissance de piétons
[63]	Reconnaissance de visages
[64]	Classification d'iris
[65]	Classification des véhicules

VIII. Les réseaux de neurones convolutifs

L'appariation des réseaux de neurones convolutifs crée une évolution massive dans le domaine des réseaux de neurones, ce sont deux réseaux presque similaires car ils inspirent leur fonctionnement du comportement du cerveau.

Les réseaux de neurones convolutifs (CNNs) donnent des meilleurs résultats dans différents domaines surtout celui de la vision par ordinateur et notamment la classification des images étiquetées [66].

La principale motivation derrière l'émergence des CNNs dans les scénarios d'apprentissage profond été de surmonter les limites des réseaux de neurones traditionnels. Lorsqu'ils sont utilisés dans des domaines comme la classification d'images, les réseaux neuronaux traditionnels entièrement connectés souffrent de problème majeur d'extraction manuelle des caractéristiques, par contre les CNNs apportent de nouvelles idées permettant l'extraction automatique à travers plusieurs couches d'abstraction qui contribuent à améliorer l'efficacité des réseaux de neurones profonds [67].

Le réseau de neurones convolutif appelé aussi CNN (en anglais «Convolutional Neural Network») est une nouvelle architecture d'apprentissage profond inventée par LeCun et al. [53,54]. Elle est modélisé sous formes multicouches hiérarchique. Il se compose généralement de trois types de couches: la couche d'entrée (convolution layer), couche(s) cachée(s) (hidden layer) et la couche de sortie (fully connected layer). Ces couches sont organisées selon leurs fonctionnement en deux parties distinctes : la première partie est la phase d'extraction des caractéristiques et la deuxième partie est la phase de classification qui correspond à un modèle **MLP classique (Multi Layers Perceptron)**.

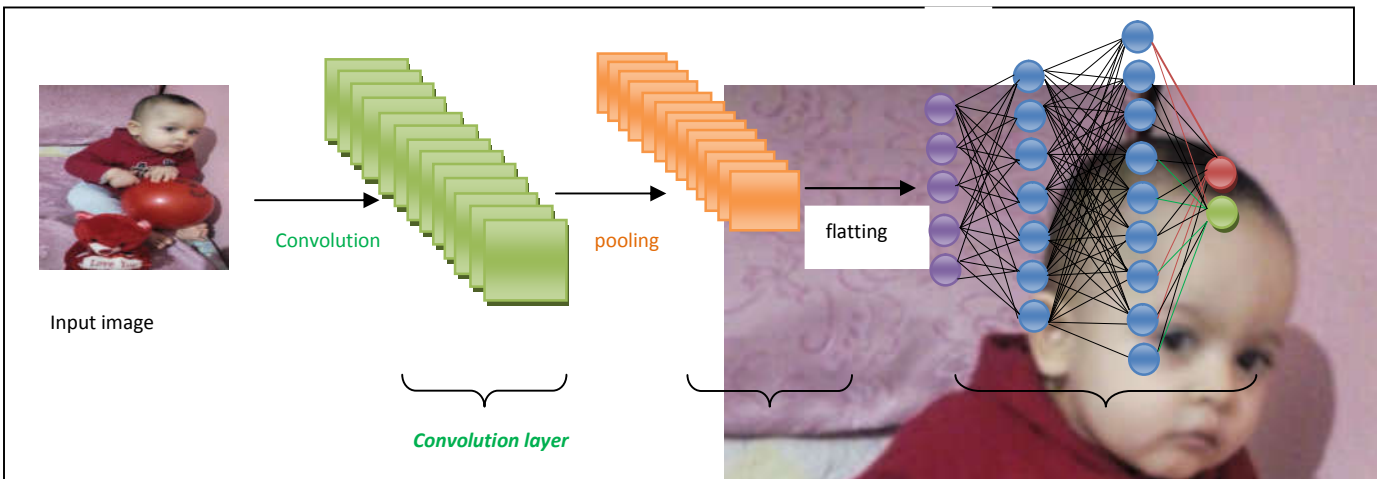


Figure 2.3 : architecture global d

VIII.1. La phase d'extraction des caractéristiques

La première phase d'un CNN est l'extraction des caractéristiques. Cette phase est divisée en deux couches (la couche convolution + la couche pooling). La couche convolution extrait des caractéristiques les plus pertinentes des images à l'aide d'une succession de filtres 'F' de taille impaire, créant des cartes de convolution, qui seront passées à la couche pooling pour réduire la dimension des cartes de caractéristiques de taille plus petite.

VIII.2. La phase de classification de neurones

C'est la dernière phase de CNN ou plus précisément la phase de classification. Elle consiste à combiner les caractéristiques déjà extraites dans la phase d'extraction des caractéristiques) sous forme de vecteur où chacune de ces valeurs doit être renvoyée à



fi
er
me
l'a
ait
x
ma
s et

La **convolution** est un outil mathématique qui agit comme un **filtrage**, où le filtre est une matrice souvent carrées de taille $(n \times n \times q)$ et généralement de taille inférieure aux tailles des images à classer $(n \leq m)$ et q est le nombre de canaux qui peut être le même que r [68].

Le filtre a pour rôle d'extraire certaines caractéristiques d'une image d'entrée. Il va balayer successivement toute l'image. Au début de l'étape de **convolution**, le filtre sera positionné tout en haut à gauche de l'image puis il va se déplacer avec un pas choisi vers la droite et lorsqu'il arrivera au bout de l'image, il se décalera de même **pas** vers le bas et ainsi de suite jusqu'à ce que le **filtre** scanne la totalité de l'image.

Pour chaque position de l'image, les valeurs des deux matrices en superposition (filtre et image) sont multipliées. La valeur résultante est projetée dans une nouvelle matrice. Cette matrice représente une nouvelle image appelé carte de caractéristiques de taille $(m-n + 1)$. L'opération de convolution est présentée par l'équation suivante :

$$S = \sum_{k=0}^{M'} \sum_{l=0}^{N'} \sum_{i=0}^R \sum_{j=0}^C I_{(i+k,(j+l))} F_{(R+i-1),(C-j+1)} \quad (2.1)$$

Avec :

I : Matrice de pixels représentant l'image

F : Matrice de filtre de convolution à appliquer sur l'image

S : résultante de produite de convolution (feature map)

R : nombre de ligne de la matrice filtre

C : Nombre de colonne de la matrice filtre

$M' : \frac{M-R}{P} + 1$, avec M est le nombre de lignes d la matrice d'images et P le nombre de pas du filtre suivant la ligne et colonne

$N' : \frac{N-R}{P} + 1$, avec N est le nombre de lignes d la matrice d'images

Chaque carte de caractéristiques est alors sous-échantillonnée de façon typique avec une couche « pooling » qui va **réduire sa dimension**.

Une fois que les étapes de convolutions sont réalisées, les caractéristiques acquises en sortie sont injectées sous forme d'un long vecteur comme données d'entrée dans un réseau neuronal classique qui nous a permet de classer n'importe quelle image.

Nous détaillons dans ce qui suit les couches de CNN :

VIII.4. Les différentes couches de CNN :

Un CNN est composé des couches suivantes :

- Couche de convolution,
- Couche pooling,
- Couche entièrement connectée,

VIII.4.1. La couche convolutive (CONV)

La couche convolutive (en anglais *convolution layer*), C'est la couche la plus importante d'un CNN car c'est elle qui fait l'extraction automatique des caractéristiques à travers le déroulement de l'apprentissage des fonctionnalités des modèles d'après les images qu'elle reçoit en entrée.

Elle effectue des opérations de convolution avec des filtres qui scannent l'entrée de n'importe quelle modèle pour avoir des matrices des caractéristiques d'une taille plus petite en comparant à la taille des matrices d'entrée c'est-à-dire faire "glisser" des filtres de taille impaire (3×3 , 5×5 ...) sur l'image d'entrée et de calculer le produit de convolution entre le filtre et chaque portion de l'image balayée[68].

Pour chaque paire (image, filtre) on obtient une carte d'activation, ou *feature map*, qui nous montre où s'arrangent les *caractéristiques* dans l'image

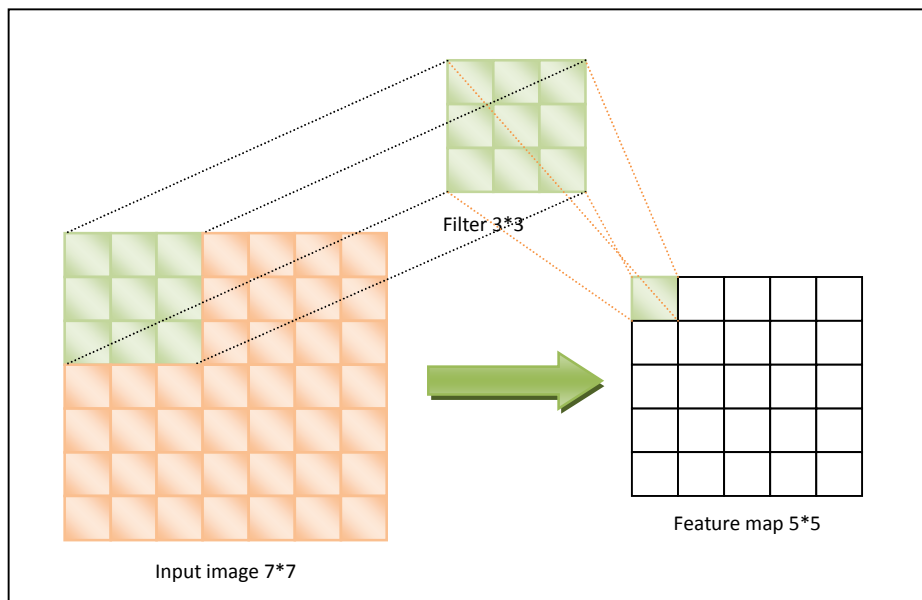


Figure 2.4 : La couche de convolution[68]

L'opération de convolution est caractérisée par deux principaux avantages :

- Permettre de diminuer le nombre de poids du réseau de neurone.
- Permettre l'identification des caractéristiques d'une manière automatique contrairement aux méthodes traditionnelles.

- **Les hyper-paramètres de la couche convolutive :**

Le but principal de la couche convolutive est de détecter les caractéristiques pertinentes dans des images ou dans des scènes : telles que des bords, des lignes, des taches de couleur et autres éléments visuels, Où les filtres peuvent les détectées. Les hyper-paramètres de la couche convolutive, qui spécifier l'architecture de CNN [67], dont les plus essentiels sont :

- Nombre de filtres
- La profondeur
- Le pas ou le stride
- La marge ou le padding
- Les fonctions d'activation

- **Profondeur de filtre :** c'est le nombre de filtres appliqués à une image donnée, c'est-à-dire un filtre de taille ($F \times F \times C$) qui effectue des convolutions sur une image d'entrée de taille $I \times I \times C$ produit un *feature map* de sortie de taille $S \times S \times 1$. Si on applique K filtres sur une image de taille $I \times I \times C$ on obtient un feature map de taille $S \times S \times K$ où K est le nombre de filtres ou bien sa profondeur.

Remarque : 'C' , est le nombre de canaux varie selon la couleur de l'image.

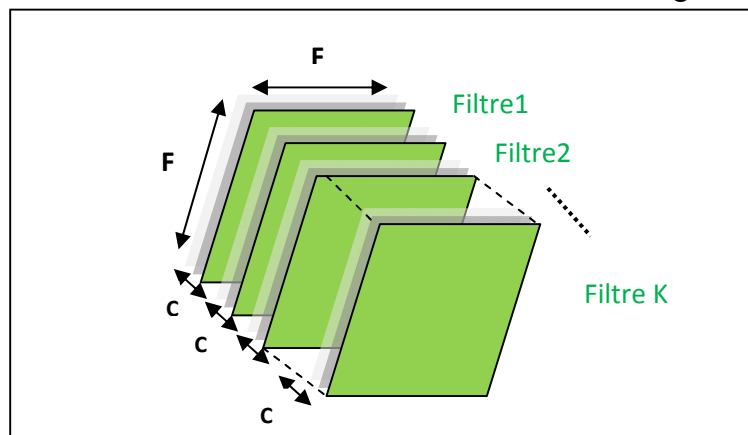


Figure 2.5 : Profondeur d'un filtre[67]

- **Le pas ou la stride :** Dans le contexte d'une opération de convolution ou de pooling, le pas ou la stride 'S' d'une opération de convolution est un paramètre qui désigne le nombre de pixels de déplacement d'un filtre sur l'image d'entrée après chaque opération.

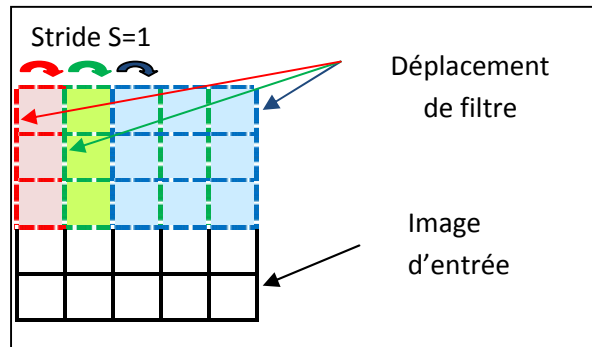


Figure 2.6 : Le stride d'un filtre [67]

- **La marge (à 0) ou zéro padding :** le zero padding est une technique qui consiste à ajouter des '0' à chaque côté des frontières de l'entrée. parfois, il est pratique de mettre ces zéros pour permettre le contrôle de la dimension spatiale du volume de sortie. Cette technique peut être réalisée manuellement, ou automatiquement.

- **Les fonctions d'activation**

La fonction d'activation (ou fonction de transfert) sert à convertir le résultat de la somme pondérée des entrées d'un neurone en une valeur de sortie, cette conversion s'effectue par un calcul de l'état du neurone en introduisant une non-linéarité dans le fonctionnement du neurone.

- **La fonction d'activation Relu**

Les CNNs ont la particularité d'introduire sur des fonctions non linéaires pour signaler une identification distincte des caractéristiques probables sur chaque couche cachée. Parmi ces fonctions, la fonction d'activation d'unité linéaire rectifiée «ReLU» en anglais Rectified Linear Units proposée en 2000 par Nair & Hinton, c'est une fonction mathématique qui s'applique en général après la couche de convolution : $f(x)=\max(0,x)$ (2.2),

qui a pour but d'augmenter les propriétés non linéaires de réseau[68], qui va améliorer l'efficacité du traitement entre les couches de CNN. Cette fonction possède deux autres variantes **Leaky ReLU**, **ELU** le déroulement de **ReLU** sera présenter comme suit :

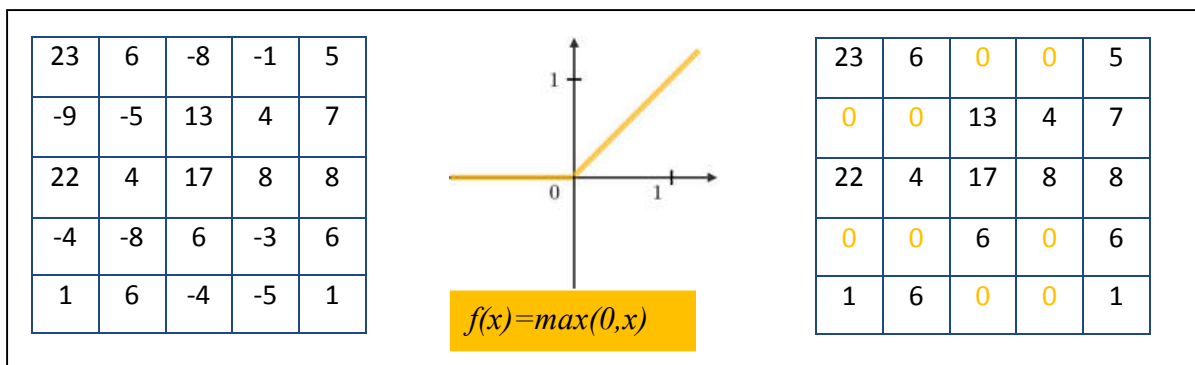


Figure 2.7 : Représentation de la fonction ReLU [68]

D'autres fonctions sont également utilisées pour augmenter la non-linéarité comme :

- La correction par tangente hyperbolique:

$$f(x) = \tanh(x) \quad (2.3)$$

et la tangente hyperbolique saturante :

$$f(x) = |\tanh(x)| \quad (2.4)$$

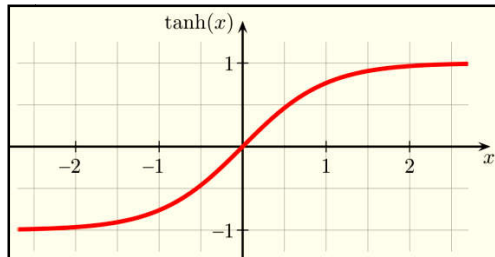


Figure 2.8 : Représentation de la fonction tangente hyperbolique

- **La correction par la fonction sigmoïde** : le choix de la fonction d'activation pour la couche de sortie doit correspondre à la fonction de coût qui est choisie en fonction de la tâche que le modèle entraîné à réaliser. Dans le cas de classification binaire, la fonction d'activation *sigmoïde* notée

$$F(x) = \frac{1}{1+e^{-x}} \quad (2.5)$$

est généralement adoptée et ses valeurs sont pris sur l'intervalle $[0, 1]$ dans le cas où la prédiction suit une loi probabiliste [69]. La fonction sigmoïde donne souvent des meilleurs résultats de prédiction. Tant que x converge vers un grand nombre positif, la fonction *sigmoïde* tendra vers 1 .

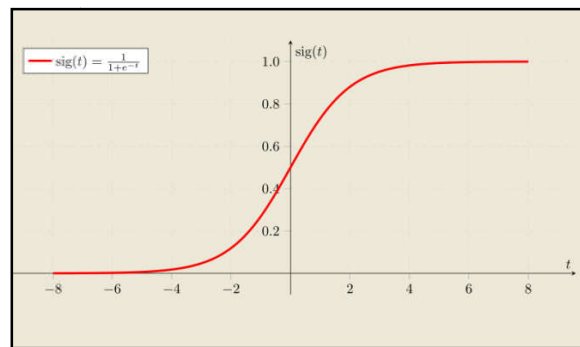


Figure 2.9 : Représentation de la fonction sigmoïde [69]

- **La fonction softmax**: appelée aussi fonction logistique, elle est favorable dans les multi-classes, appliquée toujours à la fin de réseau, définie comme suit :

$$\text{Softmax} = \frac{e^{x_i}}{\sum_{j=1}^J e^{x_j}} \quad (2.6)$$

Son rôle est d'associer le vecteur de caractéristiques obtenu par la couche pooling à un vecteur de probabilité de somme 1.

- **La fonction d'activation Identity :**

Identity, c'est une fonction d'activation linéaire (**linear activation function**). Elle peut rendre un réseau neurone multicouches équivalent à un réseau neuronal mono-couche. Son équation est la suivante : $f(x) = x$ (2.7)

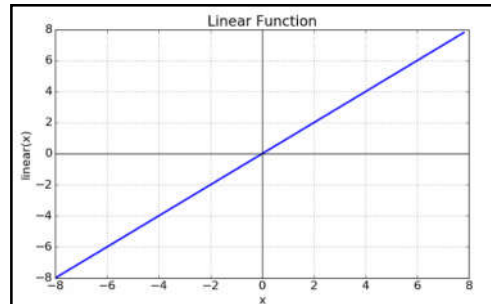


Figure 2.10 : La fonction d'activation Identity [69].

- **Les avantages de la fonction ReLu**

Malgré l'existence d'autres fonctions d'activation de la non-linéarité, La fonction ReLU est la plus répandue dans les CNNs grâce à ses avantages:

- Assure une optimisation facile par l'élimination des valeurs négatives et réduire le sur-apprentissage.
- Fonction très simple au calcul qui fournit des réponses optimales et qui permet la réduction des problèmes de disparition de gradient qui rend le réseau incapable d'optimiser ses paramètres [69].
- Renvoie, un gradient constant pour une entrée grande permettant ainsi d'apprendre plus rapidement (surtout pour les réseaux de grande profondeur).
- Soulève le problème de la sigmoïde et la tangente hyperbolique qui ont l'inconvénient de renvoyer des gradients très petits lorsque la valeur absolue de l'entrée est grande. Ce qui rend la mise à jour des paramètres de réseau difficile lors de la phase d'apprentissage.

VIII.4.2. La couche d'Agrégation (pooling) :

La couche pooling (en anglais pooling layer) c'est la deuxième couche de réseaux de neurone convolutif qui est considérée comme une spécificité de ces réseaux. Le pooling est une étape de sous-échantillonnage des cartes de caractéristiques générées par la couche de convolution tout en réduisant leurs dimensions mais en préservant leurs caractéristiques importantes avec un grand gain de temps de calcul[68].

Les deux types de pooling les plus populaires sont: le Max Pooling et l’Average pooling. Soit on extrait uniquement la valeur la plus élevée (**max pooling**) dans carte de caractéristique, soit on fait la moyenne des valeurs de la zone (**average pooling**), ces deux opérations seront bien détailler dans le tableau suivant :

Tableau 2.2: Présentation des types de pooling

Type	Max pooling	Average pooling
Définition	consiste à sélectionner la valeur maximale sur une fenêtre de calcul.	permet de sélectionner la moyenne des valeurs sur une fenêtre de calcul.
avantage	efficace lorsque on veut détecter des signaux forts	efficace lorsque on souhaite détecter des signaux faibles
Exemple		
Utilisation	C'est le plus utiliser dans la plus part architecteur du CNN	Utilisé dans l'architecture Le-Net

VIII.4.3. La couche entièrement connectée :

En anglais (*fully connected layer*) c'est la dernière couche de CNN qui se situe après une succession de multiples couches de convolutions et pooling. Comme son nom l'indique elle est entièrement connectée à la sortie de la couche précédente qui génère les cartes des caractéristiques qui seront aplatis par la suite dans un vecteurs de caractéristiques où chacune de ces valeurs est totalement connecter par tous les neurones du réseau régulier de cette couche pour donner les prédictions désirées en sortie.

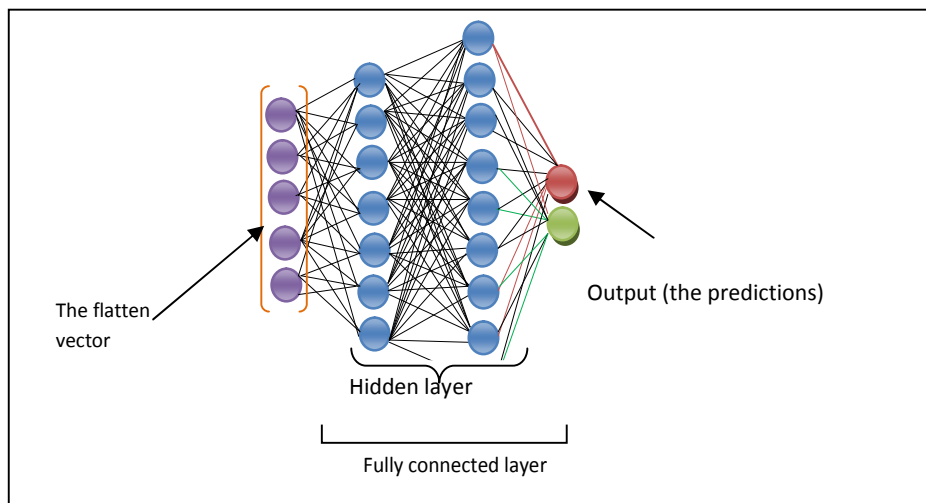


Figure 2.11 :La couche entièrement connectée [68]

VIII.5. Ajustement des hyperparamètres :

Le choix des paramètres d'un CNN joue un rôle important dans l'augmentation des performances de l'architecture des CNNs, en plus des paramètres habituelles d'un MLP standard (les règles pour les taux d'apprentissage et des constantes de régularisation), les CNNs possèdent autres hyperparamètres telle que le nombre de filtres, leur taille et la forme du pooling.

- **Nombre de filtres:** à mesure que la taille des images intermédiaires diminue avec la profondeur du traitement, les couches proches de l'entrée ont tendance à avoir moins de filtres, tandis que les couches plus proches de la sortie peuvent en avoir plus.
- **Taille du filtre:** la forme du filtre varie considérablement dans la littérature. Ils sont généralement sélectionnés en fonction de l'ensemble de données. Les meilleurs résultats sur les images MNIST (28x28) sont généralement de l'ordre de 5x5 dans la première couche, tandis que les ensembles de données d'images naturelles (généralement des centaines de pixels) ont tendance à utiliser une couche de filtre de 12x12 ou même 15x15. Le défi consiste donc à trouver le bon niveau de granularité afin de créer une abstraction adaptée à chaque situation appropriée.
- **Forme du pooling :** Les valeurs typiques sont 2x2. Si le volume d'entrée est grand on peut utiliser un pooling 4x4 dans les premières couches. Toutefois, le choix de formes plus grandes va considérablement réduire la dimension du signal ce qui entraînera une perte de l'information.

IX. Apprentissage profond et imagerie médicale

L'acquisition des images internes du corps humain peut être réalisée à travers différentes techniques telles que : les imageries par résonance magnétique (IRM), les ultrasons, les scanners, les tomographies d'émission monophonique, ...etc.

Aujourd'hui, les chercheurs utilisent les représentations numériques des parties du corps humain sous forme d'images ou son pour faire des diagnostics précoces des maladies. Ces diagnostics prennent beaucoup de temps aux médecins et ce sont très coûteux pour les patients. C'est pour cela, ils ont créé des systèmes automatiques d'analyses et de traitements d'images médicales [70].

Au départ, les opérations de classification et de segmentation ont été réalisées avec beaucoup de techniques d'apprentissage automatique (SVM, KNN, ...) mais malheureusement leurs réalisation reste encore difficile à cause de problème d'extraction manuelle des caractéristiques « *feature engenering* » .

L'émergence d'apprentissage profond fait la tendance par son utilisation dans l'analyse des images médicales vue les résultats remarquables réalisés dans différents domaines. Grâce à ses performances et sa capacité de la généralisation sur des données diverses et surtout sa possibilité de surmonter les problèmes des anciennes techniques celle d'automatiser la phase d'extraction des caractéristiques dites pertinentes.

C'est la raison pour laquelle, l'intérêt s'est attiré soudainement les méthodes d'apprentissage en profondeur plus que les méthodes d'apprentissage automatique traditionnelles.

X. L'interprétation des images

L'interprétation des images dans l'apprentissage profond regroupe trois tâches essentielles : la classification, la localisation / détection et la segmentation, dans cette section nous allons analyser les architectures les plus connues de ces tâches :

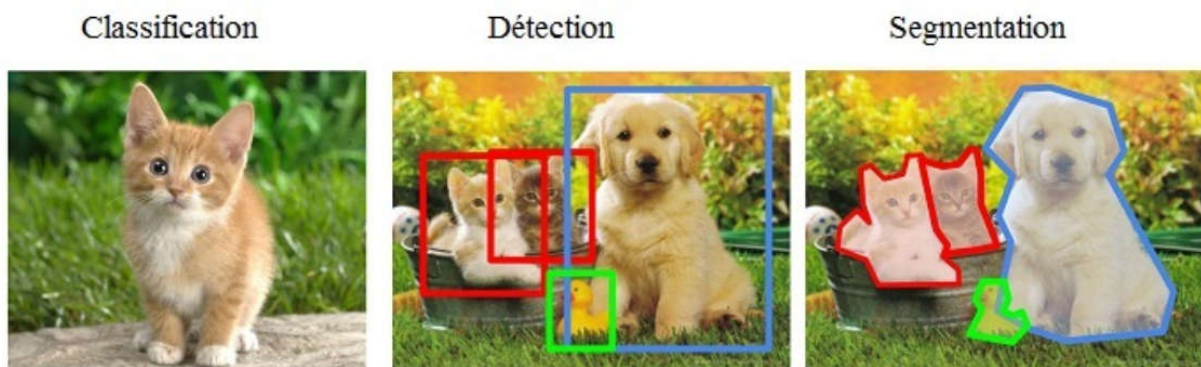


Figure 2.12 : Les trois tâches essentielles de l'interprétation d'images

X.1. La Classification

La classification automatique d'images est une application de la reconnaissance des formes. Elle s'agit d'attribuer une image à une classe à l'aide d'un système de classification automatisé afin de faciliter les études. Au fil des années et avec la révolution de la performance du matériel informatique, beaucoup d'approches ont été développées dans différents domaines d'activités tels que : le domaine militaire [71], domaine médical [59, 73]; domaine d'agriculture [74], domaine de reconnaissance d'écriture manuscrite pour les chèques, les codes postaux [75], Classification d'écriture arabe [76].

X.1.1. Les architectures de classification les plus connues

- **LeNet:**

LeNet est la première architecture de type CNN proposée par LeCun *et al* [53] en 1998 pour la classification supervisée. La figure 2.13 illustre la structure du réseau LeNet. Cette architecture est composée de 7 couches (3 couches de convolution, 2 couches de Avg-pooling et 2 couches

entièrement connectées). Avec une taille des filtres des couches de convolution de 5×5 et les entrées sont des images de taille 32×32 .

Les couches *convolution* et *pooling* sont composées d'un certain nombre de cartes de caractéristiques d'une taille définie. Plus qu'on rapproche de la sortie de ce réseau plus la taille des images est réduite par rapport aux premières couches. Dans cette architecture l'apprentissage est basé sur la rétropropagation. L'implémentation de cette architecture a été testée la première fois sur la base de MNIST de classification des chiffres manuscrits qui contient 60 000 images pour l'apprentissage et 10 000 images pour le test, où l'algorithme d'apprentissage profond LeNet a montré une efficacité remarquable en comparant par les anciennes architectures telles que SVM et KNN.

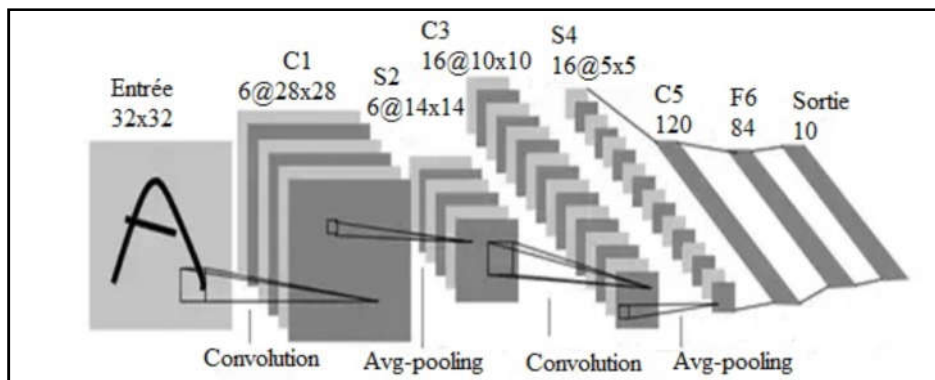


Figure 2.13 : L'architecture du réseau LeNet [53]

- **AlexNet**

L'apparition du réseau AlexNet [77] en 2012 était un point important dans l'histoire de l'apprentissage profond surtout après la compétition ImageNet de reconnaissance visuelle à grande échelle (ILSVRC). Dans cette compétition, **AlexNet** a été proposé avec un jeu de données de la base ImageNet composé de 1000 catégories, 1.2 million images d'apprentissage, 50 000 images de validation, et 150 000 images de test où il a obtenu le meilleur taux d'erreur. L'architecture de ce réseau (figure 2.14) est semblable à l'architecture LeNet mais plus profonde. AlexNet est composé de 5 couches de convolution et 3 couches entièrement connectées. La première couche de convolution utilise 96 filtres de taille 11×11 , mais les autres couches convolutives sont basées sur des filtres de taille 5×5 et 3×3 . La première, la deuxième et la cinquième couche de convolution sont suivies par des couches de max-pooling et les deux premières couches sont suivies par une opération de normalisation (local response normalisation (LRN)). La fonction LRN est implémentée pour améliorer la généralisation et la non linéarité du réseau, elle est appliquée sur les résultats de la fonction *d'activation ReLu*, dans le but d'accélérer

l'apprentissage comme la *fonction tanh* proposée dans l'architecture LeNet, ce qui simplifie la complexité temporelle pour le cas des grands modèles qui sont entraînés sur des jeux de données immenses.

Krizhevsky a exploité les techniques d'augmentation de données et de régularisation par abandon (Dropout) pour éviter le problème de sur-apprentissage qui intervient à cause du nombre élevé des paramètres (60millions),

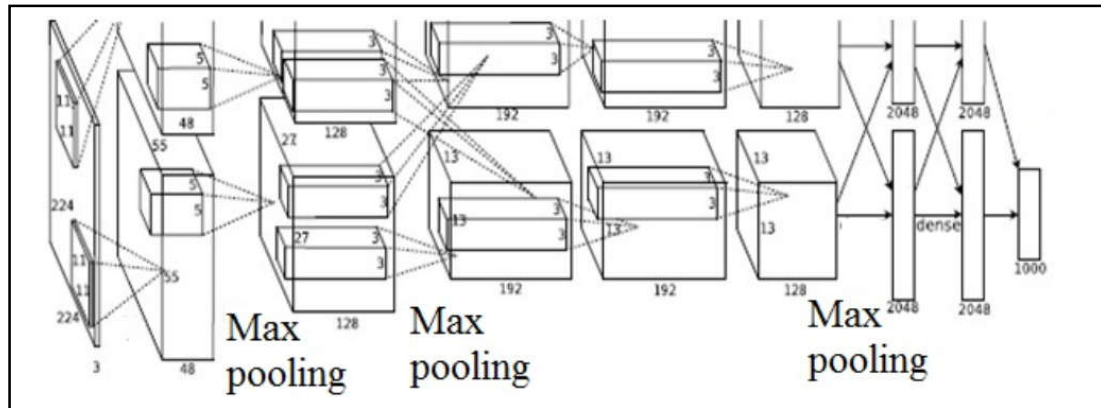


Figure 2.14: L'architecture du réseau AlexNet [77]

- **ZFNet**

En 2014, et après une étude détaillée de comportement de AlexNet, Zeiler & Fergus ont proposé une nouvelle structure **ZFNet**[78], qui permet la visualisation du contenu des couches intermédiaires par un réseau déconvolutif multicouche (Deconvnet) [78] Le réseau Deconvnet est lié à chaque couche du réseau CNN qui permet de projeter les cartes des caractéristiques internes aux entrées afin de visualiser leur contenu.

Le réseau Deconvnet est composé de trois opérations : unpooling, rectification et filtrage, où la méthode unpooling est l'opposé de l'opération pooling d'un CNN. Elle permet de restaurer le contenu des cartes des caractéristiques avant l'opération de pooling. L'opération rectification est basée sur la fonction d'activation ReLu qui permet d'éliminer les valeurs négatives des cartes des caractéristiques. Le filtrage est l'opération inverse d'une convolution basée sur la version transposée des filtres utilisés dans l'opération de convolution. Cette opération applique les filtres transposés sur les cartes des caractéristiques pour obtenir la couche de convolution précédente. Ces trois opérations sont répétées d'une manière successive sur les cartes des caractéristiques internes jusqu'à atteindre l'ensemble des pixels en entrée. Afin de visualiser le contenu des cartes des caractéristiques.

Cette proposition a permis d'améliorer la performance du réseau AlexNet par 1.7 % qui a encouragé les chercheurs à utiliser les techniques de visualisation dans l'ajustement des hyperparamètres.

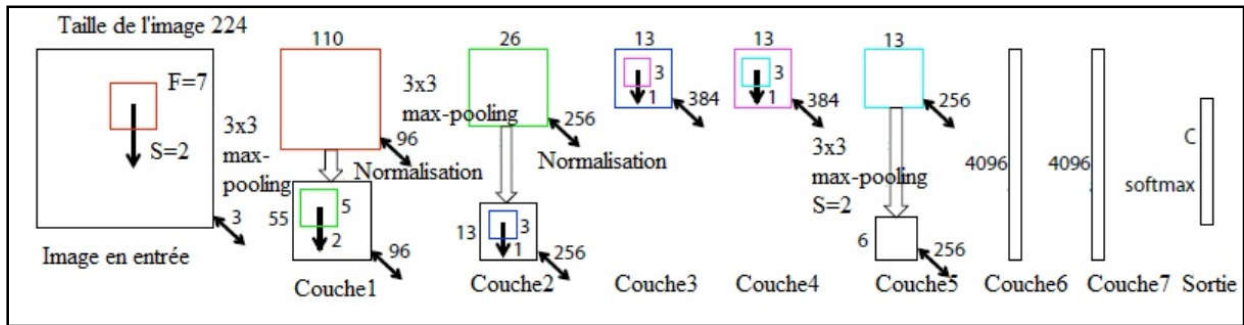


Figure 2.15 : l'architecture du réseau ZFNET [78]

- VGGNet

VGGNet [79], est un réseau convolutif proposé en 2014 avec une configuration profonde similaire à celle d'AlexNet où les couches de VGGNet peuvent varier de 16 à 19 couches. Cette architecture est basée sur la technique de stabilisation structurelle pour contrôler le nombre de ses paramètres afin de diminuer les risques de sur-apprentissage, aussi, le réseau VGGNet recommande de diminuer la taille des filtres afin de réduire le nombre des paramètres. Par exemple, l'utilisation de 2 couches de filtre 3x3 est meilleur que l'utilisation d'une seule couche de filtre 5x5 car en utilise moins de paramètres : Le nombre d'un paramètre est réduit de 28% $((25-18) / 25)$. C'est-à-dire :

$$\text{Nombre de paramètres de 2 couches de filtre } 3 \times 3: 2 \times 3 \times 3 = 18$$

$$\text{Nombre de paramètres de 1 couche de filtre } 5 \times 5: 1 \times 5 \times 5 = 25$$

En résumé, l'amélioration de la non-linéarité du réseau est réalisée grâce aux fonctions d'activation (ReLU) supplémentaires.

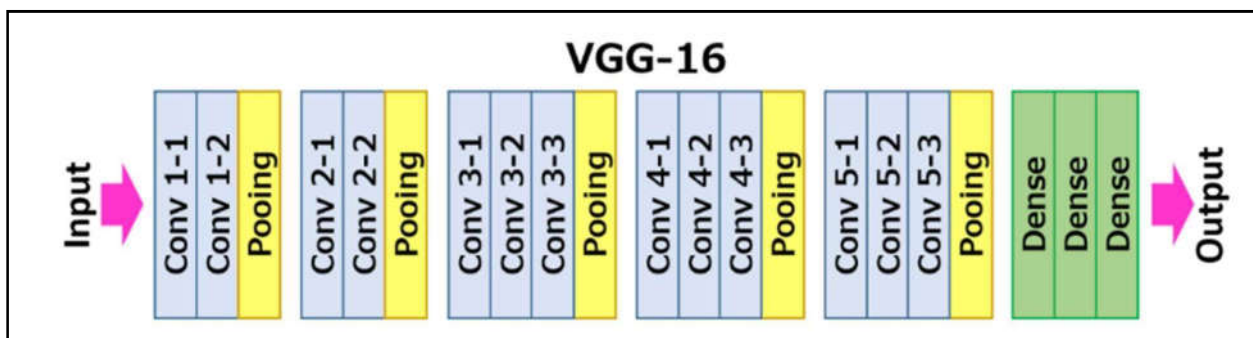


Figure 2.16 : l'architecture du réseau VGGNet [79]

- Inception

Les experts en apprentissage profond recommandent généralement d'augmenter la taille du réseau en fonction de la largeur (nombre de paramètres dans chaque couche) et de la profondeur (nombre

de couches) pour améliorer les performances du réseau. Depuis 2015, et pour éviter les grands problèmes des réseaux profonds qui sont le risque de sur-apprentissage et l'exigence de capacité de stockage et de calcul, beaucoup de recherches sont tournées vers les architectures partiellement connectées.

Le réseau Inception [80] est un réseau neuronal convolutif utilisant les modules Inception. Ces modules permettent l'optimisation des couches convolutives classiques. Le module initial introduit des connexions partielles à l'intérieur de la couche convolutive pour réduire sa dimension. La Figure 2.17 illustre la structure du module Inception.

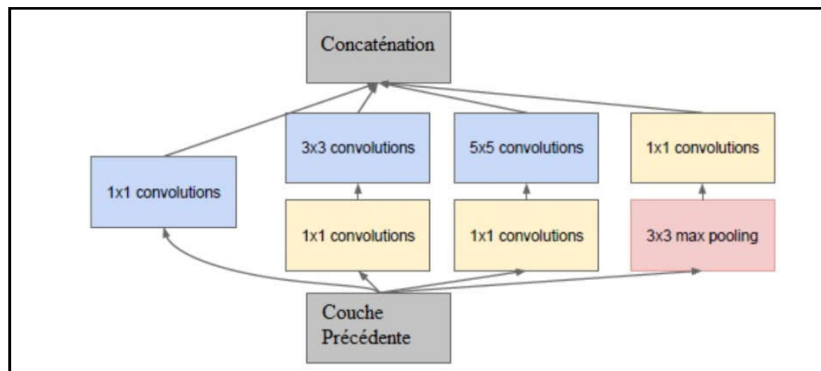


Figure 2.17 : La structure d'un module d'Inception

On peut voir que la structure de module d'Inception [80] Utilise des filtres de taille variable (1×1 , 3×3 et 5×5) et applique-les sur la même couche convolutive pour éviter les problèmes d'alignement du patch. Ensuite, les cartes d'entités générées sont rassemblées pour former la couche convolutive suivante. Afin de résoudre le problème de l'augmentation rapide de la profondeur de la couche convolutive due à la cascade de cartes de caractéristiques, le filtre de la taille 1×1 a été introduit avant les filtres de taille 3×3 et 5×5 . Ces filtres réduisent la profondeur de la couche convolutive avant d'appliquer d'autres filtres et améliorent la non-linéarité grâce à la fonction d'activation ReLu.

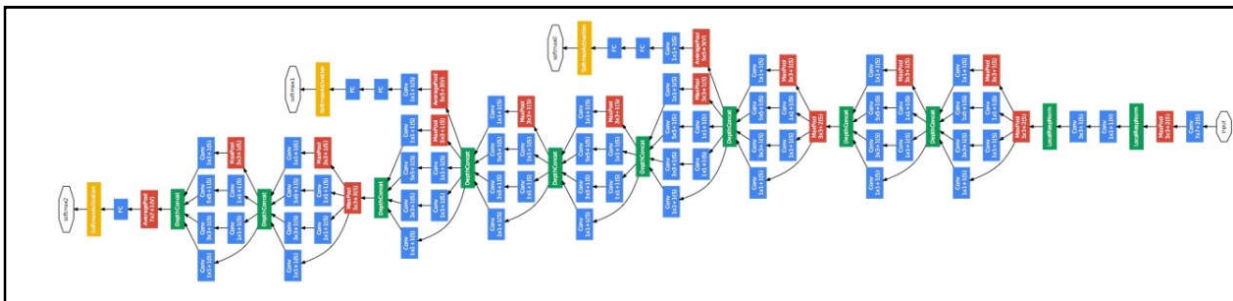


Figure 2.18: L'architecture du réseau Inception [80].

Toujours dans le cadre d'optimisation de la complexité de calcul Szegedy *et al.*, ont proposé en 2015 **InceptionV2** et **InceptionV3** [81], qui est des versions améliorées du réseau Inception.

- **ResNet**

He et al ont proposé en 2016 un réseau neuronal convolutif appelé ResNet [82] basé sur le bloc résiduel, ils visent à résoudre le problème de dégradation du gradient, qui survient généralement dans les réseaux très profonds. La précision de ces réseaux commence à atteindre la saturation, puis elle se dégrade rapidement en raison de la diminution de la valeur du gradient. Dans l'architecture ResNet, ils ont introduit des blocs résiduels sous forme de jonctions entre la sortie de la couche précédente et la sortie de la couche courante. La formule de ces jonctions est présentée par l'équation 2.4 pour réaliser l'addition, elle est suivie d'une projection linéaire Ws afin d'obtenir des dimensions équivalentes (x et $F(x)$).

$$y = F(x) + x \tag{2.4}$$

L'étude comparative de He *et al.*, d'un CNN avec et sans blocs résiduels montre que les réseaux les plus profonds de Res-Net sont les plus performants

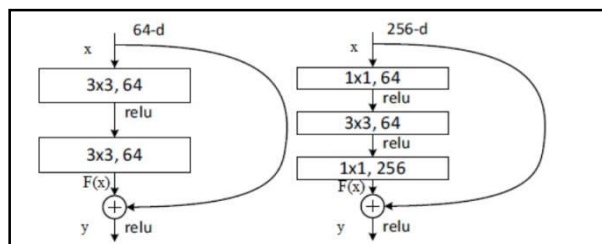


Figure 2.19 – La structure des blocs résiduels [82]

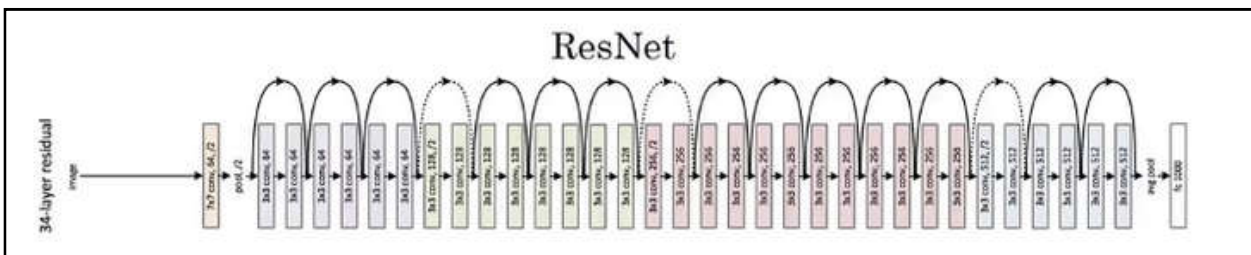


Figure 2.20 : l'architecture du réseau ResNet (34 couches) [82]

- **DenseNet**

DenseNet[83] est un réseau de neurones convolutif proposé en 2017, basé sur des connexions denses entre les couches convolutives. Selon la Figure 2.21, il est constitué d'une collection de blocs denses liés par une couche de transition. Chaque bloc dense contient un ensemble de couches

convolutives, où chaque couche est connectée à toutes les autres couches qui la suivent à condition qu'elle appartienne au même bloc. Au contraire d'un CNN classique, chaque couche reçoit comme entrée la carte des caractéristiques de la couche qui la précède. Les connexions dans le bloc dense établissent un lien direct entre le gradient de la fonction de coût et les entrées originales. De plus, ils peuvent améliorer la régularisation, réduisant ainsi le sur-apprentissage et la dégradation de gradient.

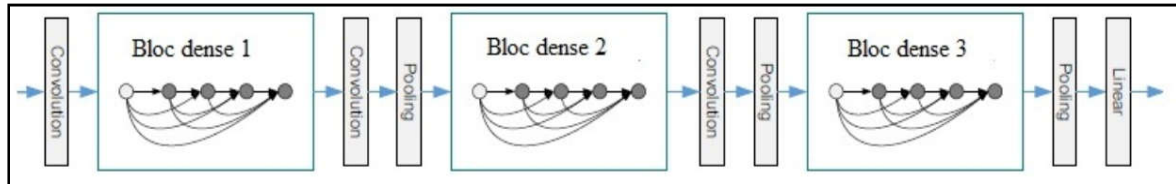


Figure 2.21: L'architecture du réseau DenseNet [83]

D'après les exemples cités en haut, on peut dire que toutes les architectures appliquées sur la base d'images « ImageNet » ont réalisé des résultats remarquables grâce à sa richesse et sa taille. Nous avons vu une variété d'architectures profondes, dont l'objectif concurrentiel est de réduire le nombre total de paramètres et d'accélérer le temps d'exécution, bien sûr, pour réduire le taux d'erreur. La concurrence pour ces architectures a toujours été l'optimisation pour éliminer les problèmes de sur-apprentissage et de dégradation de gradient.

X.1.2. Exemples des travaux de classification profonde

Voici quelques exemples des travaux réalisés pour la classification des images médicales basant sur différentes architectures des réseaux de neurones convolutifs :

Tableau 2.3 : Exemples des travaux de classification profonde

Référence	Type de Réseau	Type de diagnostic
[84]	Variante de GoogLeNet	Classification à haut débit des radiographies
[85]	CNN	Diagnostic d'Alzheimer
[59]	Variante de AlexNet	Classification de cancer du sein
[86]	CNN	La classification des tissus épithélial (EP) et stroma (ST)
[87]	CNN	Classification de cancer du sein
[88]	ChexNet	Classification des images de radiographie pulmonaire

X.2. La Détection et localisation des objets

La détection consiste à localiser des éléments ou des organes à partir de la surface de l'image, ce qui est considéré comme difficile par rapport à la classification car il faut localiser des informations

plus précises. Habituellement, la zone d'intérêt sera entourée par un contour, cadre ou cercle, ou mieux encore, une courbe (bounding box). En imagerie médicale, cette opération de détection pose de grands défis pour les raisons suivantes: sa rapidité, la corrélation du cadre de contour et la détection de petits éléments.

X.2.1. Les architectures de détection les plus connues

- **R-CNN**

C'est le premier modèle d'apprentissage profond dédié pour la détection des objets, proposé en 2014 par Girshick et al [89]. Cette méthode est une combinaison entre les méthodes de proposition des régions (regions proposal) et les CNN. la structure R-CNN est composée de 3 modules: extraction des régions d'intérêt, extraction des caractéristiques et classification. Le premier module est basé sur l'algorithme de recherche sélective, qui fait la sélection de divers patches de l'image susceptibles de contenir des objets. Ensuite, ces régions sont ajustées pour obtenir une taille conforme à l'entrée de CNN. Le réseau R-CNN utilise aussi le modèle Alexnet [77] pré-entraîné précédemment sur la base ImageNet, et suivi par un fine-tuning sur la base d'apprentissage cible PASCAL. A la fin, ces vecteurs sont fournis à l'algorithme SVM pour finir la classification et mettre les objets dans des cadres grâce à l'algorithme bounding-box regressors.

Malgré que l'apparition de R-CNN a fait un grand changement dans le domaine de détection d'objet grâce à son efficacité de fonctionnement mais il a connu un échec en raison du manque d'optimisation du temps d'apprentissage et de test.

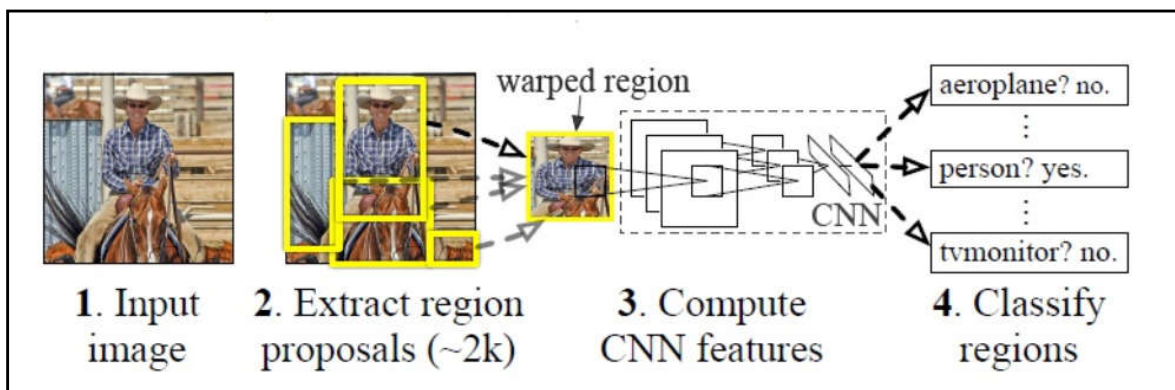


Figure 2.22 – L'architecture de Régions avec réseaux de neurones convolutif (R-CNN) [89]

- **Fast R-CNN**

Afin de surmonter les problèmes d'optimisation du temps d'apprentissage et de test dans la version R-CNN, Girshick a proposé la version améliorée Fast R-CNN [90]. Le principe général du réseau Fast R-CNN est de transmettre l'image d'entrée au réseau CNN pour concevoir la carte des caractéristiques en sortie (l'extraction des caractéristiques et la classification de bout en bout sont effectuées par un seul algorithme). Ensuite, identifiez les zones suggérées dans ces cartes et

redimensionne-les via la couche de fusion RoI-pooling. Ensuite, le vecteur généré passe à travers les couches entièrement connectées. Enfin, l'étape de classification amène à régler le cadre de sélection par le bounding-box regressor en utilisant softmax pour réajuster le cadre de sélection. On peut dire que Faster-RCNN est rapide en phase de test, car l'étape de classification est presque en temps réel, il est donc plus facile à entraîner, plus précis et plus rapide.

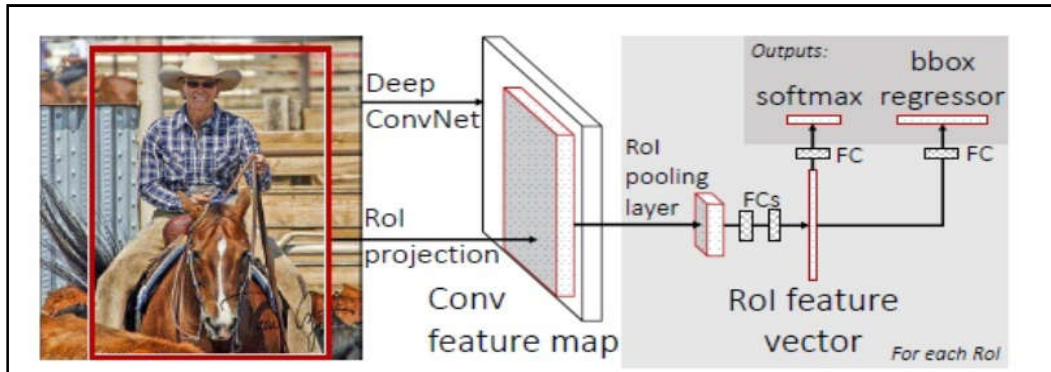


Figure 2.23 – l'architecture du réseau Fast R-CNN [90].

- YOLO (en anglais You Only Look Once)

YOLO [91], c'est un autre algorithme de détection d'objet, en comparant par d'autres architectures mentionnées précédemment, cette méthode présente des comportements différents. Ce niveau de détection est un simple problème de régression dans lequel la prédiction de catégorie et la sélection d'objets sont effectuées en parallèle.

Le principal avantage de l'apprentissage et de la prédiction depuis l'image entière et que le taux de faux positifs soit minime. Le fonctionnement de YOLO commence par diviser l'image d'origine en grilles $S \times S$, où chaque grille prédit B images sélectionnées et la probabilité que l'objet appartienne à une catégorie différente C. Le cadre sélectionné est caractérisé par 5 paramètres: 4 coordonnées (x, y, h, w) et le score de confiance P, qui est la probabilité d'appartenance de l'objet au cadre et sa précision. En raison qu'il traite 25 images par seconde YOLO est très efficace.

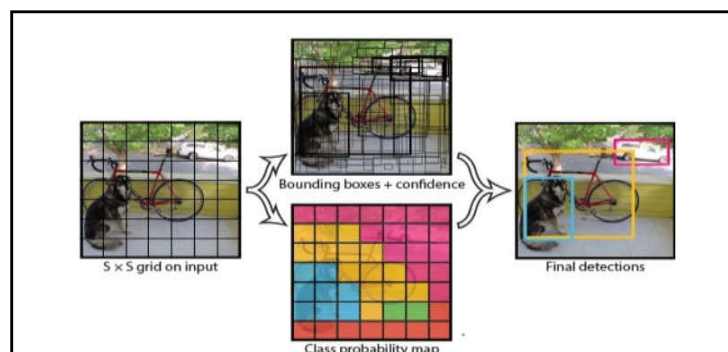


Figure 2.24 – L'architecture de détection des objets par YOLO[91].

Le but principal des différentes architectures proposées pour la détection d'images est de fournir une structure permettant la détection d'objets en temps réel. Par exemple, YOLO a pu accélérer le temps de détection en intégrant la phase générée par RoI dans le réseau.

X.2.2. Exemples des travaux de détection

Dans le tableau suivant nous citons quelque travaux de détection dans des images médicales se basant sur les réseaux de neurones convolutifs :

Tableau 2.4 : Exemples des travaux de détection profonde

Référence	Type de Réseau	Type de diagnostic
[92]	R-CNN	La détection des lésions malignes ou bénignes
[93]	Variante de Faster R-CNN	La détection des nodules pulmonaires
[94]	CNN	La détection des points de repère sur la surface du fémur distal
[95]	CNN	La détection des microsaignements cérébraux
[96]	CNN	La détection des noyaux

X.3. La Segmentation

La segmentation est une tâche de base dans les applications de diagnostic assisté par ordinateur car le diagnostic des pathologies représente des situations vraiment critiques où le résultat de la segmentation peut remplacer l'avis de médecin dont l'erreur ne sera jamais autorisée.

La segmentation sémantique est une brique essentielle dans la compréhension d'une scène. Elle sert à classifier l'ensemble de pixels d'une image en attribuant une classe à chaque pixel. Avec l'apparition des techniques d'apprentissage profond, les CNN peuvent jouer le rôle des classifieurs de pixels. Dans ce cas le réseau prend en entrée des parties de l'image et classifie le pixel de centre, cela est répétée pour tous les pixels de l'image, où chacun est considéré comme le centre du segment déduit. Le problème de cette répétition et qu'elle consomme un temps énorme ce qui limite utilisation des CNN dans les applications qui travaillent en temps réel. De ce fait, la création d'autres architectures de segmentation plus robustes est nécessaire.

X.3.1. Les architectures de segmentation les plus connues

- **Fully convolutional networks(FCN)**

C'est un réseau très efficace pour la segmentation sémantique de nombreux types d'images : multimédia [97], aériennes [98], ou médicales [99]. Le réseau FCN [100] est une architecture de bout en bout dédié pour la segmentation, il est capable de réaliser la phase de segmentation en un seul passage contrairement à un CNN. Ce réseau est composé de deux principales phases: le sous-échantillonnage et le ré-échantillonnage. Une fois les informations sémantiques et contextuelles sont captées dans la phase de sous-échantillonnage, elles seront ré-échantillonnées pour rétablir les informations spatiales. Les réseaux FCN n'utilisent pas les couches entièrement connectées pour ne

pas fixer la taille des entrées et gérer les entrées de taille variante. Ainsi que les couches entièrement connectées sont remplacées par des couches de convolution pour produire des cartes spatiales. Qui seront passées à des couches de déconvolution pour restaurer la taille en entrée et produire des sorties classifiées par pixel.

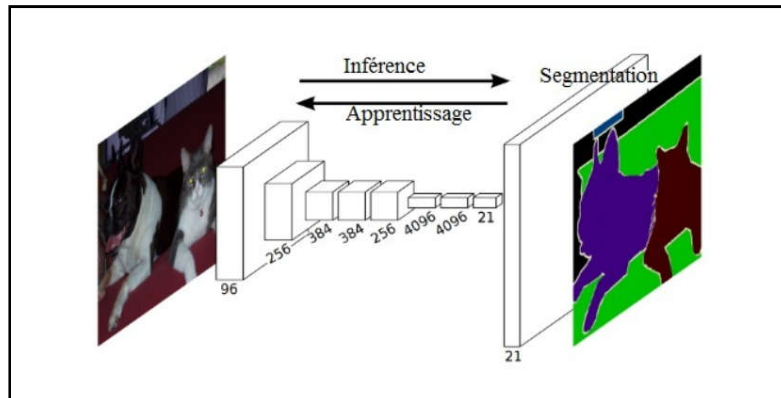


Figure 2.25: L'architecture du réseau fully convolutional network [100].

- **Les architectures Encodeur-décodeur**

La majorité des réseaux de convolution sont utilisés dans les tâches de classification des images, où la sortie est une prédiction. Néanmoins, cette sortie peut être un objet localisé ou une forme segmentée utilisant d'autres variantes des réseaux de l'apprentissage profond. Récemment, de nombreuses études sont basées sur les architectures *Encodeur-décodeur* pour améliorer leurs systèmes d'aide au diagnostic médical.

La croissance de l'utilisation des approches encodeur-décodeur a également été considérable pour segmenter les images médicales de différentes maladies telles que: les images biomédicale [101], les vaisseaux sanguins intracrâniens, vaisseaux rétiniens, contours de cellules et poumon [102], les lésions cutanées du psoriasis [103], les tumeurs cérébrales [104,105], les lésions cutanées dans les images de dermatoscopie [106], les images radiologiques médicales [107].

Le plus connu dans la segmentation des images médicales dont l'architecture encodeur-décodeur est 'U-Net', qui sera l'axe de notre deuxième contribution.

- *U-Net*

U-Net est le premier réseau de genre encodeur-décodeur dédié pour la segmentation, il a été créé par Ronneberger et al [101] sous forme de la lettre ‘U’, qui se déroule facilement avec une succession des couches, une fois on met une image à l’entrée du U-Net on obtient l’image segmentée à la sortie.

“U-Net” est une architecture qui contient deux chemins symétriques: un chemin de contraction (côté gauche) suivi par un chemin expansif (côté droit), d’où le chemin de contraction suit parfaitement la même architecture d’un CNN.

Le U-Net se compose de 5 niveaux, chaque image couleur de taille (x*x* 3) à l’entrée du premier niveau va approfondir à travers les autres niveaux à gauche et suit la forme ‘U’ jusqu’à la sortie de U-Net à droite. Cette architecture sera bien détaillée dans la contribution 2.

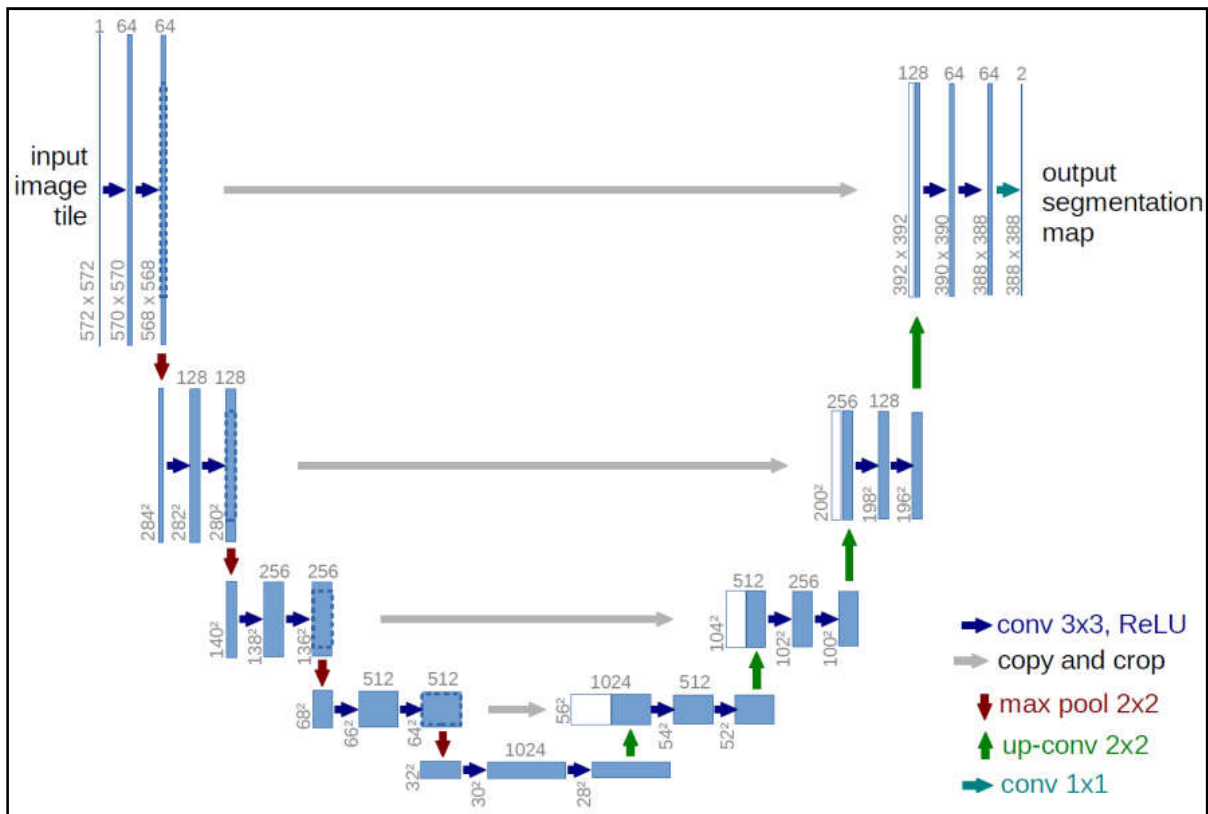


Figure 2.26 :L’ architecture du réseau U-Net [101].

- **SegNet**

SegNet[108] est un autre modèle encodeur-décodeur proposé en 2017. Conçu pour la segmentation sémantique par pixels. Son architecture totalement convolutive est inspirée du modèle VGG-16, où l'encodeur comprend typologiquement 13 couches convolutives du VGG16 qui produit des cartes d'activation qui sont par la suite sur-échantillonnées et projetées dans l'espace sémantique par le décodeur dont le rôle est de mapper les cartes de caractéristiques de basse résolution de codeurs avec les cartes de caractéristiques de résolution d'entrée totale.

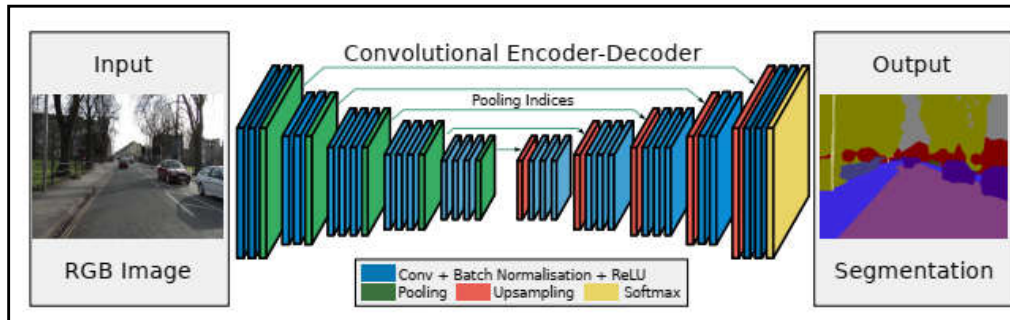


Figure 2.27 : L'architecture du réseau Seg-Net [108].

L'intérêt majeur de proposer des architectures dédiées pour la segmentation sémantique, est de segmenter les objets en temps réel comme le réseau FCN, Qui peut effectuer la segmentation dans la phase de prédiction en un seul passage. Aussi l'apparition des architectures encodeur -décodeur (U-Net, Seg-Net...) a révolutionné ce domaine car ces réseaux assurent une segmentation plus précise avec moins d'images d'entraînement. Par exemple U-net est capable de segmenter une image 512 * 512 en moins d'une seconde sur un GPU.

En plus des architectures encodeur-décodeur citées précédemment, il existe d'autres architectures connues comme : M-Net , PSPNet,... etc, nous illustrons ces modèles dans le tableau suivant :

Tableau 2.5 : Exemples des travaux de segmentation profonde

Référence	Type de Réseau	Type de diagnostic
[101]	U-net	La segmentation des structures neuronales
[109]	CNN	La segmentation du gliome dans le cerveau
[110]	FCN	La segmentation des tissus
[107]	XNet	Segmentation des images de différentes parties du corps :pieds, de genoux et de têtes
[111]	U-Net ++	Segmentation des tissus dans Noyaux cellulaires, Polype du côlon, Le foie, et les Nodules pulmonaire
[105]	architecture hybrid pour SegNet et U-Net (U-SEGNET)	segmentation des tissus de cerveau
[112]	DCNN	La segmentation des noyaux
[103]	variante de U-Net(PsLSNet)	Segmentation de psoriasis
[104]	Architecture modifiée de SegNet	segmentation des tissus de cerveau

XI. Conclusion :

Dans ce chapitre, nous avons clarifié les principaux concepts de l'intelligence artificielle. Nous avons parlé des réseaux de neurones, du biologique à l'artificiel en passant par le formel. Nous avons défini le domaine d'apprentissage automatique en survolant ses techniques.

Ensuite, nous avons passé à l'apprentissage profond qui est la tendance de l'Intelligence Artificielle ces dernières années et qui fournit des algorithmes puissants notamment dans le domaine médical, raison pour laquelle nous avons illustré ses domaines d'application.

Ensuite, nous avons détaillé la structure générale d'un réseau CNN qui est l'architecture de base de l'apprentissage profond et nous nous sommes concentrés sur la manière d'ajustement des hyperparamètres de ce dernier. A la fin, nous avons défini l'interprétation des images médicales qui englobe trois tâches essentielles (Classification, détection, segmentation) où nous avons analysé quelques architectures connues en classification, en détection des objets et en segmentation accompagné d'un état de l'art de ces dernières.

**Chapitre III : Approche Globale de
Classification Multimodale de la Maladie de
Glaucome basée sur les Classifieurs CNNs et
la Représentation de Texture (filtre de Gabor)**

I. Contexte de la contribution

L'objectif de notre thèse est l'interprétation des images médicales où on s'intéressait aux images rétiniennes pour donner un deuxième avis aux ophtalmologistes afin d'apporter un apport à notre pays.

Pour assurer l'interprétation, le problème peut être vu en deux façons, d'une part, de niveau global en essayant de proposer un système d'aide au diagnostic médical basé sur la totalité d'image, et d'autre part, d'une manière locale, c'est-à-dire on s'intéresse à la zone d'intérêt (ROI) qui contient l'anomalie (la zone suspecte) .

Ce travail s'inscrit dans le but d'améliorer le diagnostic précoce de la maladie de glaucome. Il s'agit de développer un système assistant le médecin dans son raisonnement en vue d'identifier un diagnostic avant la complication de l'état de l'œil, pour réaliser ce système, nous avons utilisé les réseaux de neurones convolutifs dans le but de concevoir des modèles pour automatiser l'extraction des caractéristiques pertinentes à partir des images de fond d'œil.

Notre travail est divisé en deux contributions qui seront détaillées chacune dans un chapitre dans le but d'améliorer la performance du système d'aide au diagnostic médical. La première contribution, se concentre sur une approche globale multimodale basée CNN qui permet de classifier directement les images de la rétine et la deuxième c'est une approche locale qui doit localiser la zone suspecte, dans notre cas c'est le disque optique, qui doit être segmenté par un modèle U-Net, une fois générée elle sera l'entrée d'un autre CNN pour la classification des images de fond d'œil de la rétine en deux classes : Glaucome et Non-Glaucome

Le nombre de personnes diagnostiquées par la maladie de glaucome augmente considérablement au cours des dernières années. Selon les dernières statistiques du National Eye Institute (NEI), le glaucome sera considéré comme la deuxième maladie oculaire menant à la cécité dans le monde. Il peut être causé par un groupe de maladies, y compris le diabète, migraines, hypertension et mauvaise circulation sanguine [113]. Les personnes de plus de 40 ans peuvent développer un glaucome, mais il peut survenir à tout moment d'âge, affectant même les petits enfants. Habituellement, Le glaucome peut surprendre les patients. Il ne présente pas de symptômes précoces, mais lorsqu'on prend conscience de sa présence, une grande partie de la vue peut être perdue. Par conséquent, il est chronique et irréversible, il peut conduire à la cécité s'il ne se traite pas tôt [114].

Généralement, le glaucome survient lorsque le liquide (l'humeur aqueuse) s'accumule dans la partie avant de l'œil. Cela augmente la pression et endommage le nerf optique [115]. La forme la plus courante du glaucome est celle d'angle ouvert [113],

Les systèmes de diagnostic assistés par ordinateur basés sur les images ont déjà soutenus les médecins dans divers domaines à garantir leurs décisions en augmentant la prévention contre les maladies et traiter les patients dès les premiers stades avec moins d'argent et moins du temps [116].

II. Travaux connexes

Récemment, La classification du glaucome a subi un excellent développement, notamment en ce qui concerne l'utilisation des paradigmes d'apprentissage profond. En effet, beaucoup d'architectures ont été créées dans le but d'améliorer la robustesse de la classification et permet aux médecins de prendre leurs décisions en temps réel.

L'apprentissage profond a atteint l'état de l'art dans différents domaines de diagnostic médical en raison de ses performances élevées grâce à son avantage majeur d'extraction automatique des caractéristiques discriminantes qui facilite la phase de classification, plusieurs approches ont révélé des records en utilisant des réseaux de neurones convolutifs profonds (CNNs) dans le contexte du diagnostic de la maladie de glaucome, où le nombre de couches et le paramétrage de réseaux peut faire la différence dans le classement de performance de ces architectures.

Che et al. [25] ont proposé une architecture CNN pour la détection du glaucome avec six couches d'apprentissage: quatre couches convolutives et deux couches entièrement connectées. Ils ont basé sur la segmentation de la région d'intérêt (ROI) pour améliorer les performances du diagnostic du glaucome. Les ensembles de données ORIGA et SCES sont utilisés dans la phase de validation de l'approche proposée. Ces résultats s'appuyés sur le critère de performance de l'aire sous la courbe ROC, varie entre 83,1% et 88,7% respectivement des ensembles de données ORIGA et SCES.

Doshi et al.[50] ont mis au point une architecture CNN pour diagnostiquer la rétinopathie diabétique avec 10 couches de convolution (7 max-pooling et 3 couches entièrement connectées). Utilisant les techniques d'augmentation de données pour améliorer la tâche de classification. Ils ont utilisé une unité de processeur graphique (GPU) de haute gamme sur l'ensemble de données Kaggle, où ils ont obtenu une précision de 75%. Aussi, Choi et al [117], ont proposé une structure CNN pour la classification automatique de la rétinopathie diabétique. Ils ont utilisé des images de fond de couleur de jeu de données fourni par Kaggle et ils ont obtenu une précision de 94,5%.

GÓMEZ-VALVERDE et al. [118] ont proposé différentes architectures de réseaux de neurones convolutifs (CNNs) pour illustrer l'influence de certains facteurs pertinents sur la performance de CNN. Leurs études ont été basé sur: la taille de l'ensemble de données, l'architecture et l'utilisation de différentes fonctions d'activations. Après une grande comparaison de plusieurs structures CNNs, ils ont optés pour l'architecture d'apprentissage basée sur le transfert de données avec VGG19 appliquée sur trois ensembles de données différents (2313 images de fonds de glaucome), cette approche atteint une AUC de 94%.

Raghavendra et al.[31] ont proposé un nouveau système d'aide au diagnostic basé CNN et le classifieur ADL pour classer les images du fond d'œil en glaucome et non glaucome, ce réseau est constitué de dix-huit couches afin d'atteindre une précision maximale de 98,13%.

De nombreuses recherches sont orientées vers les architectures multimodales dans le but de bénéficier de l'avantage de plusieurs classifieurs en même temps, car quelle que soit la performance d'un système individuel elle est inférieure à la performance d'un système multimodale. Benzebouchi et al.[28] ont proposé une approche multimodale de réseaux de neurones convolutifs pour diagnostiquer le glaucome, ils ont utilisé des données diverses des images du fond rétinien divisées en deux types: des images en couleur et des images en binaires. L'architecture de ce réseau contient 2 couches convolutives, 2 couche max-pooling et une couche entièrement connectée. Ils ont réalisé des résultats satisfaisants avec AUC d'environ 98% et 97,8%. Xu et al. [51] ont créé un multi-catégoriel CNN appelé MatConvNet pour détecter plusieurs maladies rétinienne. Ils ont utilisé la base d'image Retina (STARE). Qui contient 10 catégories de ces maladies. Ils ont rapporté une précision de 36,7%, 0,225 kappa et 0,053 RCI dans 10 maladies rétinienne.

Récemment, plusieurs architectures CNN visent à introduire la représentation de texture «filtre de Gabor»: afin d'atteindre une grande précision, les chercheurs ont proposé de nombreuses contributions fondées sur la représentation de texture spécialement basé sur le filtre Gabor et les architectures hybrides de CNN pour améliorer leurs performances dans de nombreux domaines; Mahiba et al. [119] ont proposé un système de classification des lésions multi-classes d'images rétinienne appelé MCNN. Ils ont utilisé des caractéristiques combinées de couleurs et de textures (filtre Gabor) appliquées à un ensemble de données privé de 550 images de la rétine. Qui donne une précision de 98,41%. Molaei et al.[120] Ont utilisé un CNN basé sur les filtres de Gabor pour la classification et la segmentation d'images. Ils ont testé différents jeux de données (des images cardiaques IRM pour la tâche de segmentation) et (CIFAR10, MNIST, FashionMNIST et NotMNIST pour la tâche de classification) ; ils ont atteint une précision élevée de 97,8%. Lui et al. [121] ont proposé un nouveau modèle profond, nommé réseaux convolutifs hybrides de Gabor (HGCM) qui utilisaient alternativement une entrée binarisée et filtre binarisé gabor (BGF) appliqué sur le jeu de données MNIST.

Dans cette contribution et afin d'améliorer la robustesse et la sensibilité de la classification de cette dangereuse maladie, nous avons intégré une méthodologie de caractérisation basée sur les techniques de traitement d'image. Pour fournir plus de données multimodales à l'entrée de chaque classifieurs, qui est présenté par quatre types de descripteur: couleur, binaire, niveau de gris et Gabor dont le but d'extraire plus d'informations des images du

fond rétinien issu de jeu de données RIM-ONE pour détecter correctement le nerf optique afin d'obtenir un meilleur taux de classification.

Alors, l'objectif de notre travail est d'analyser d'une part l'importance de la texture comme modalité supplémentaire à la décision et d'autre part l'impact de la fusion des classifieurs à base CNNs entraînés par plusieurs sources d'informations.

Notre contribution multimodale basée CNN pour la classification du glaucome se résume dans les points suivants :

- Cette étude s'appuie sur quatre modalités relatives aux images d'entrée de chaque CNN, les quarts descripteurs sont : couleur, binaire, niveau de gris et Gabor. Chacune de ces modalités fournit certains types d'informations supplémentaires au système qui ne peut pas être déduites ou obtenues par d'autres. Cette information supplémentaires appelée en termes mathématiques « la diversité ».
- Nous voulons également, analyser l'importance de la texture en utilisant « Gabor » comme une source importante de représentation des données médicales. (Gabor est le descripteur de texture *le plus utilisé dans le traitement d'images*)
- Montrer le rôle de fusion des classifieurs issues de plusieurs sources d'informations.

Ce travail propose un module de fusion appelée « somme et produit » pour combiner les résultats de quatre classifieurs et profiter de la complémentarité qui peut exister entre ces classifieurs.

Comme notre travail se base sur la multimodalité et l'apprentissage d'ensemble (l'ensemble learning), en va détailler dans ce qui suit ces deux termes pour mieux comprendre le contexte de notre approche ;

III. Concepts de base :

III.1. La Multimodalité

Ce terme désigne l'utilisation de plusieurs modalités de nature différentes pour effectuer la même tâche. Donc, une modalité peut être considérée comme un concept flou pouvant être représenté de différentes façons.

Au sens sémiotique, *une modalité* est une technique spéciale utilisée pour coder des informations en se référant à un certain type d'informations stockées dans un format spécifique pour présenter les informations [122].

Une modalité est une forme concrète et spécifique à un mode de communication qu'il soit auditif, spatial, gestuel, ou visuel. Un problème de recherche est qualifié de multimodal si il est capable d'intégrer une ou plusieurs de ces modes.

«*La complémentarité* » est une propriété indispensable de la multimodalité, car chaque modalité apporte à l'ensemble fusionné un certain type de valeur ajoutée qui ne peut être obtenue à partir d'aucune des autres modalités du système. Cette valeur est appelée, au terme mathématique, «*la diversité* », elle permet de réduire le nombre de degrés de liberté dans le système en fournissant des contraintes qui assurent l'unicité, la performance, la robustesse et la prise de décision [123].

III.1.1. Fusion multimodale

La fusion multimodale présente le processus de combinaison de multiples sources de données provenant de différentes modalités afin de déduire de meilleurs résultats qui ne peuvent être obtenus par aucune des sources uniques. De telle façon, quelque soit l'efficacité individuelle d'une modalité, elle ne donne jamais des résultats supérieure/précise comparant par une fusion de t'elle modalité avec d'autres. La figure 3.1 montre un système de fusion multimodale typique.

III.1.2. Intérêt de la fusion multimodale

Il y a, en effet, un grand intérêt à appliquer la fusion multimodale dans les applications. C'est surtout pour obtenir une image et une vue d'ensemble plus unifiées du système afin d'améliorer la prise de décision. Dans les points suivants, nous résumons les principales raisons qui poussent les utilisateurs à préférer les paradigmes multimodaux :

- Plus on ajoute une source complémentaires au système plus on lui offre une représentation complète.
- La fusion de sources multimodales augmente la robustesse d'un système et réduit le taux d'erreur [124].
- Dans les phases pratiques, la fusion réduit toujours les sources non fiables [124].
- L'efficacité de la reconnaissance est influencée par le bruit dans les données, mais le fait d'avoir des sources multiples permet de diminuer son effet [125].
- le système à une seule modalité est moins performant en lui comparant à celui multimodal.
- Améliorer la performance d'un système à modalité unique en agissant sur le réglage de l'extraction des caractéristiques ou certaines autres étapes de classifieur n'est jamais fiable car aucune source n'est parfaite mais la combinaison de plusieurs sources peut augmenter cette fiabilité [125, 126].

III.2. Classification d'ensemble (Ensemble Learning)

La classification d'ensemble (*en anglais Ensemble Learning*) est un concept très utile dans ces derniers temps [127], il fait référence à la combinaison de plusieurs classifieurs, Récemment, il a été reconnu comme une voie de recherche importante afin de bénéficier de la complémentarité qui peut exister entre l'ensemble de classifieurs. Par conséquent, les décisions des classificateurs

combinés sont généralement plus précises que les résultats des classifieurs individuels qui les composent.

Alors, la classification d'ensemble est une combinaison des décisions de différents classifieurs individuels pour classer de nouveaux exemples avec plus de précision.

Le développement de ce concept de combinaison de classifieurs a été toujours basé sur les quatre points suivants:

1. Type d'informations de la sortie de chaque classifieur à fusionner.
2. La structure de la combinaison.
3. Le type et les caractéristiques des classifieurs à fusionner.
4. Les règles de combinaison utilisées pour fusionner les décisions d'ensemble de classifieurs.

III.2.1. Pourquoi combiner des classifieurs ?

- l'absence d'un meilleur classifieur capable de traiter toute distribution de données d'apprentissage.
- Il est très difficile à un classifieur individuel de discriminer correctement un ensemble important de classes.
- Simplifier la procédure de réglage de classifieur et minimiser les taux d'erreurs.
- Exploiter les performances de chacun des classifieurs.
- Profiter de la complémentarité qui peut exister entre les différents classifieurs.
- Assurer une meilleure précision du système, car la combinaison des résultats de plusieurs classifieurs donne une décision plus fiable.
- L'efficacité, en général, un problème complexe devient plus simple si l'en décompose en plusieurs sous-problèmes qui sont plus faciles à comprendre et à résoudre.

III.2.2. Les stratégies de combinaison

Afin de construire un système multi-classifieurs basé sur la combinaison de différentes sorties de ces derniers dans le but d'obtenir une décision finale unique et efficace, trois topologies de combinaison sont introduites et mises en œuvre: combinaison séquentielle, combinaison parallèle et combinaison hybride

a. combinaison séquentielle :

La combinaison séquentielle ou la topologie en série, est déroulée en étapes décisionnelles consécutives pour diminuer progressivement le nombre de classes potentielles. Autrement dit, chaque phase contient un classifieur unique qui prend en considération la réponse donnée par le classifieur de la phase qu'elle précède (c'est-à-dire la sortie de la classification générée par un

classifieur est utilisée comme entrée dans le classifieur suivant afin de traiter les rejets ou confirmer la décision). Ce processus, rend un problème complexe beaucoup plus simple.

Généralement, pour ce genre de combinaison l'ordre d'emplacement des classifieurs devient très important, surtout à propos de choix du premier classifieur qui doit être choisi judicieusement afin d'éviter au maximum la propagation d'une mauvaise réponse.

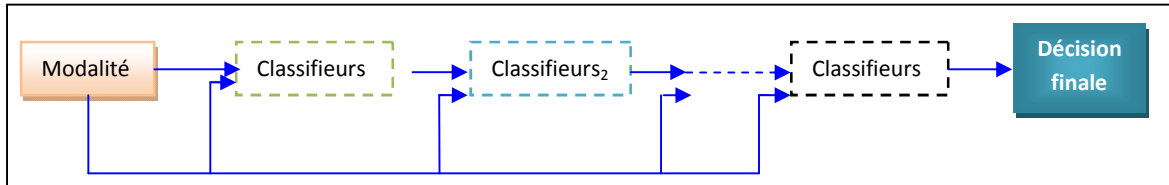


Figure 3.1 : Combinaison séquentielle de classifieurs [127]

b. La combinaison parallèle :

C'est la topologie la plus utilisée à cause de sa simplicité de mise en œuvre ainsi de ses résultats impressionnants. Dans cette stratégie l'ordre dans lequel les classifieurs sont exécutés n'a aucune importance. Le déroulement de cette topologie est basé sur les classifieurs (de base) qui fonctionnent indépendamment les uns des autres (en parallèle) pour prédire l'étiquette d'un objet inconnu (modalité). plus tard, leurs décisions individuelles sont combinées afin de donner à une décision unique (figure 3.2).

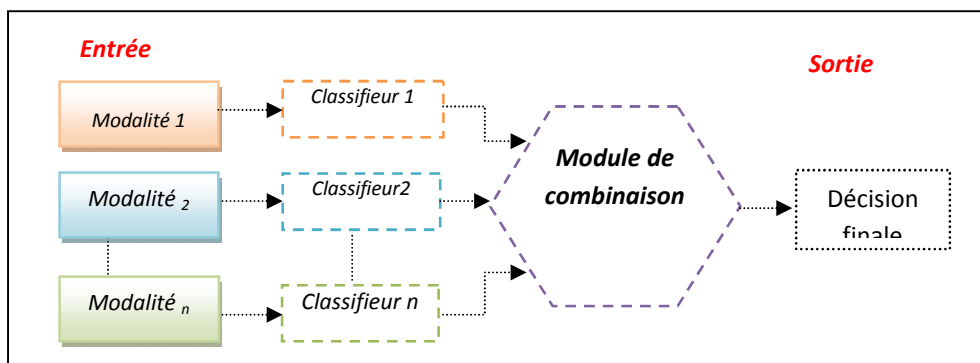


Figure 3.2 : Combinaison parallèle de classifieurs[127]

c. La combinaison hybride :

Elle est basée sur la fusion de deux topologies précédemment citées (séquentielles et parallèles) pour bénéficier pleinement et simultanément des points forts de ces deux topologies.

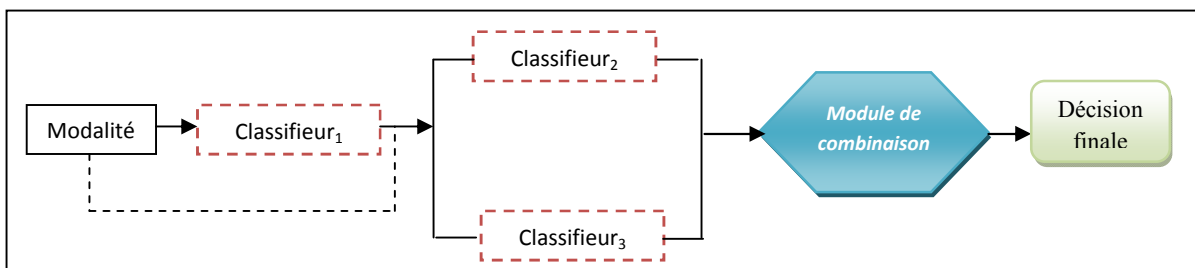


Figure 3.3 : Combinaison hybride de classifieurs[127]

III.3. Filtre de Gabor

Le filtre de Gabor a été inventé par le physicien britannique Dennis Gabor (1900-1979), c'est un filtre linéaire largement utilisé dans les applications du traitement visuel humain, car il a l'avantage d'être paramétré en fréquence et en orientation pour en extraire des traits significatifs utiles. Pendant le processus de convolution, le filtre de Gabor donne également de bons résultats dans la détection de la fréquence et de l'orientation de la texture [128,129].

Une fonction de Gabor dans l'espace du domaine 2-D est une fonction gaussienne (ou enveloppe) modulée par un signal sinusoïdal (également appelé porteuse) [130]

Gabor représentant par l'équation suivante:

$$h(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right]\right\} \cos(2\pi f_0 x) \quad (3.1)$$

Où $x = x \cos \theta + y \sin \theta$ and $y = -x \sin \theta + y \cos \theta$
avec

(f_0): radial centre frequency

(θ): Orientation.

σ_x, σ_y : écarts types

Le filtre de Gabor serait représenté par deux gaussiens espacés symétriquement dans le domaine de fréquence spatiale correspondant comme suit :

$$H(u, v) = \exp\left\{-2\pi^2[(u - f_0)^2\sigma_x^2 + v^2\sigma_y^2]\right\} + \exp\left\{-2\pi^2[(u + f_0)^2\sigma_x^2 + v^2\sigma_y^2]\right\} \quad (3.2)$$

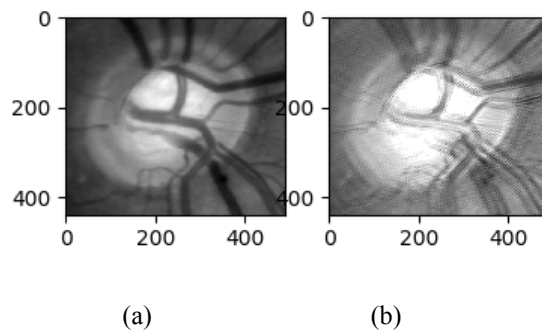


Figure 3.4 (a) Image de fonds d'œil en niveaux de gris tirée de l'ensemble de données RimOne et (b) traitement de la même image à l'aide du filtre Gabor [34]

IV. La méthode proposée

L'objectif de notre travail est la conception d'une approche globale de classification, appelée β plus-CNN. Le modèle β plus-CNN classifie les images de fond d'œil en deux classes: "Glaucome" et "Non-Glaucome" avec quatre visions de l'image en entrée (couleurs, binaires, niveaux de gris et les images passées par le filtrage de Gabor). Nous avons utilisé la base de données «RimOne», qui est un ensemble de données gratuit et indexé pour valider notre système.

Notre contribution montre l'impact d'utiliser la représentation du texture (le filtre de Gabor) comme une source complémentaires de notre système d'aide au diagnostic médical et montrer le rôle de fusion des classifieurs issues de plusieurs sources d'informations. Nous nous sommes concentrés sur le problème le plus fréquemment mentionné dans le domaine de l'apprentissage profond, à savoir l'absence de quantité suffisante de données.

Après des recherches approfondies sur les hyper-paramètres des couches CNN et après plusieurs tests empiriques, nous avons réussi à définir l'architecture β plus -CNN de notre système de classification des images de fond d'œil de la rétine.

Les principales phases de notre système peuvent être résumées par le schéma suivant :

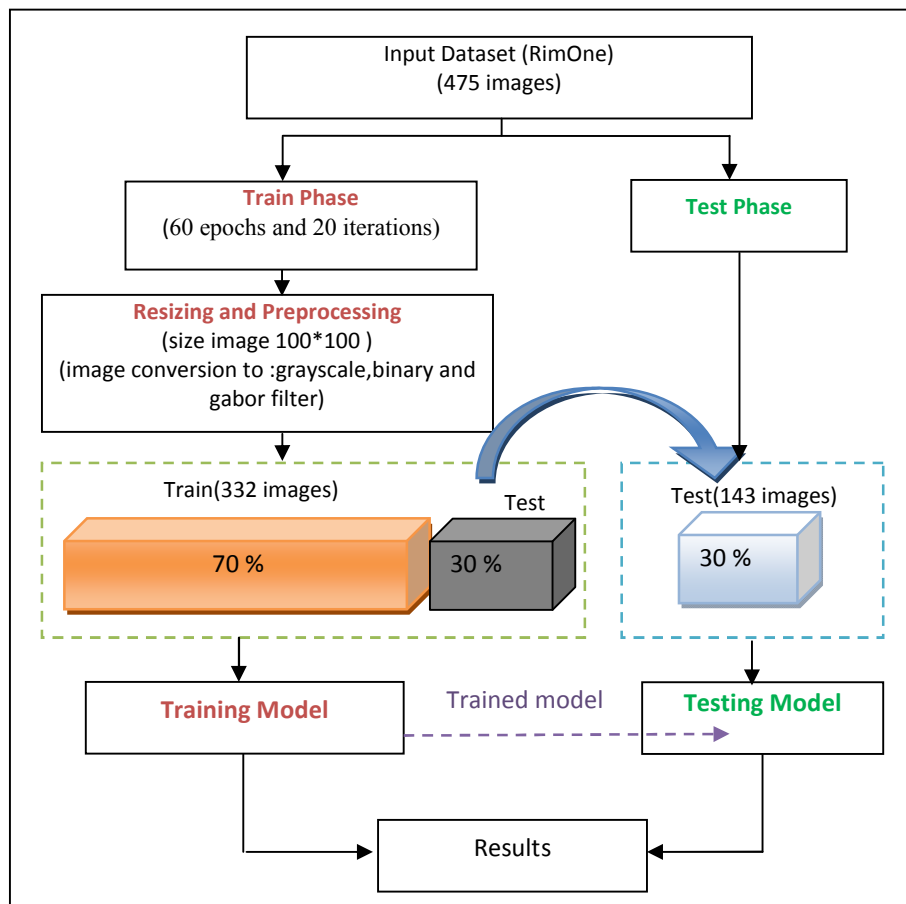


Figure 3.5. Organigramme proposé β plus -CNN [34]

IV.1 Prétraitement

Le prétraitement représente une phase fondamentale dans des nombreux systèmes de reconnaissance des formes. Dans notre cas, nous avons utilisé la base "RIM-ONE" pour valider notre système. En effet, chaque image de l'ensemble de données a été redimensionnée à 100x100 pixels. Afin d'avoir des meilleures performances, nous avons renforcé le système par la combinaison (collaboration) de quatre modalités.

La phase préliminaire de prétraitement consiste à obtenir les modalités complémentaires des images couleurs originales qui constituent d'autres points de vue intéressantes de notre système. Il s'agit notamment de la transformation en binaire, de la transformation en niveau de gris et du passage par le filtre de Gabor, nous illustrons ces opérations comme suit:

a. La Conversion en binaire

Pour la conversion en binaire nous avons appliqué l'algorithme standard reconnu par ses performances de binarisation "OTSU" [131]. La binarisation permet de réduire le nombre des classes de pixels en des pixels noirs et des pixels blancs. Les images résultantes sont bien claires, sans bruit et montre spécifiquement un meilleur contraste entre les structures rétiniennes.

b. La conversion en niveaux de gris :

Pour la conversion en niveaux de gris, nous avons appliqué l'algorithme le plus simple basé sur la fonction appliquée à chaque pixel p ,

$$p = \frac{r + g + b}{3} \quad (3.3)$$

Autrement dit ; convertir une image couleur en niveaux de gris, est de remplacées les trois valeurs représentant le rouge, le vert et le bleu par une valeur unique.

c. La transformation avec le filtre de Gabor :

Pour la transformation avec le filtre de Gabor nous avons appliqué l'algorithme de Gabor [132], où nous avons appliqué différents variantes de filtre sur nos images d'entrée avec plusieurs paramètres d'orientations (30 °, 45 ° et 60 °) et avec certaines fréquences (1, 2, 3, 4, 5). Donc une image peut être changée par 15 filtres en 15 images différentes, nous pouvons voir la différence entre ces images après avoir procédé aux filtres (figure.3.6). Suite à une observation attentive de toutes les images résultantes nous avons sélectionné celles passées par le filtre de Gabor avec une orientation de 45° et une fréquence de 5, en raison de l'émergence de vaisseaux sanguins et de nerf optique dans ces images.

Les images ci-dessous montrent que le nerf optique est clairement visible après l'utilisation du filtre de Gabor.

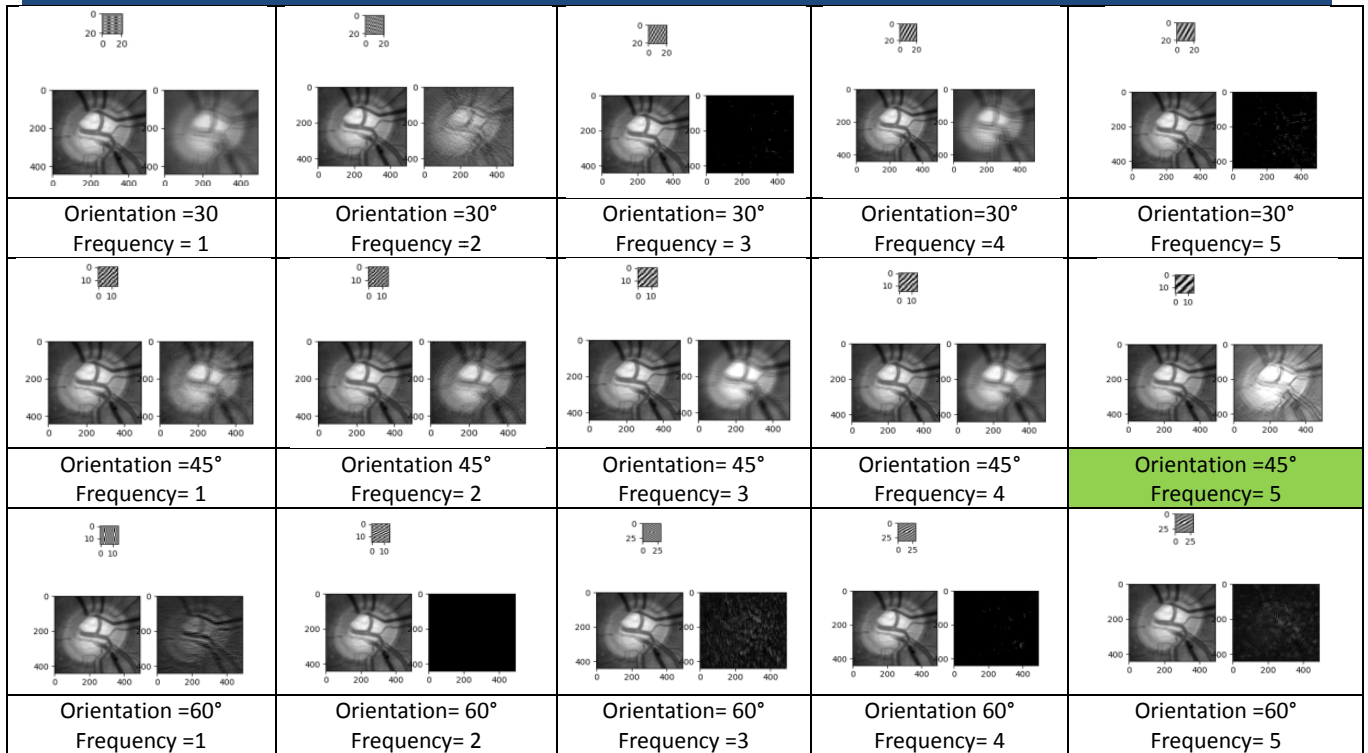
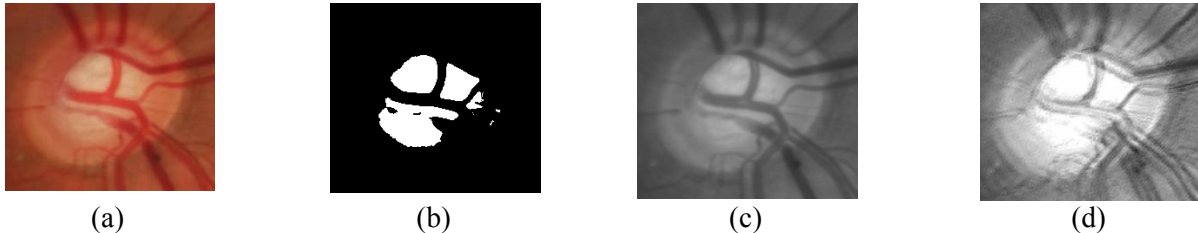


Figure 3.6: Résultats des images filtrées par Gabor avec différentes orientations et fréquences [34]

Voici un exemple de chacune de différentes modalités résultantes du prétraitement, qui seront l'entrée de notre système β plus –CNN



(a) (b) (c) (d)
Figure.3.7 : Les quatre modalités (a) couleur, (b) binaire, (c) niveau de gris, (d) filtre de Gabor [34]

IV.2 La structure β plus –CNN

Nous avons conçu un système d'aide au diagnostic médical basée CNN utilisant la base RimOne représentée par quatre visions : Couleur, binaire, niveaux de gris et Gabor. Notre système a été structuré en deux principales phases (apprentissage et test).

IV.2.1. Phase d'apprentissage:

Cette étape comprend la construction d'un modèle d'apprentissage. Avant de discuter de l'apprentissage de CNN, Nous devons modéliser son architecture en spécifiant de nombreux paramètres. La construction d'une architecture CNN n'est pas une tâche facile; Il s'agit d'un processus long et répétitif qui nécessite des ajustements des hyper-paramètres du réseau. En effet, nous devons d'abord configurer plusieurs couches et choisir soigneusement leurs types, puis tester et mesurer les performances du modèle. Si les résultats obtenus sont satisfaisants on

sauvegarde le modèle sinon on ajuste le modèle en ajoutant, éliminant ou modifiant des paramètres pour obtenir une architecture qui fournit des résultats acceptables. Par conséquent, plusieurs facteurs peuvent influencer sur la performance d'un CNN également :

- Le nombre des filtres, le nombre de pixels dans chaque filtre, pour les couches de convolution
- La taille de la fenêtre, la valeur de stride (le pas), Pour chaque couche de pooling
- Le nombre de neurones cachés pour la couche entièrement connectée

Les réseaux convolutifs sont généralement composés de trois types de couches : Convolution, Sous-échantillonnage généralement de type Max-pooling et la couche entièrement connectés.

Dans notre cas, et après plusieurs tests empiriques, nous choisissons la structure de notre modèle ***β plus –CNN*** qui a été inspiré de l'architecture Alex-net, il comporte deux couches de convolutions dont chacune est suivie par une couche pooling. La sortie de la couche de pooling associée à la deuxième couche de convolution sera l'entrée de la couche entièrement connectée. Ce réseau sera formé pour prédire deux classes (glaucome et non glaucome) comme décision finale. Notre système est réellement constitué de quatre classifieurs, En utilisant la méthode de fusion (somme et produit) pour combiner les sorties de ces derniers. Toutes ces étapes sont détaillées dans la figure 3.8.

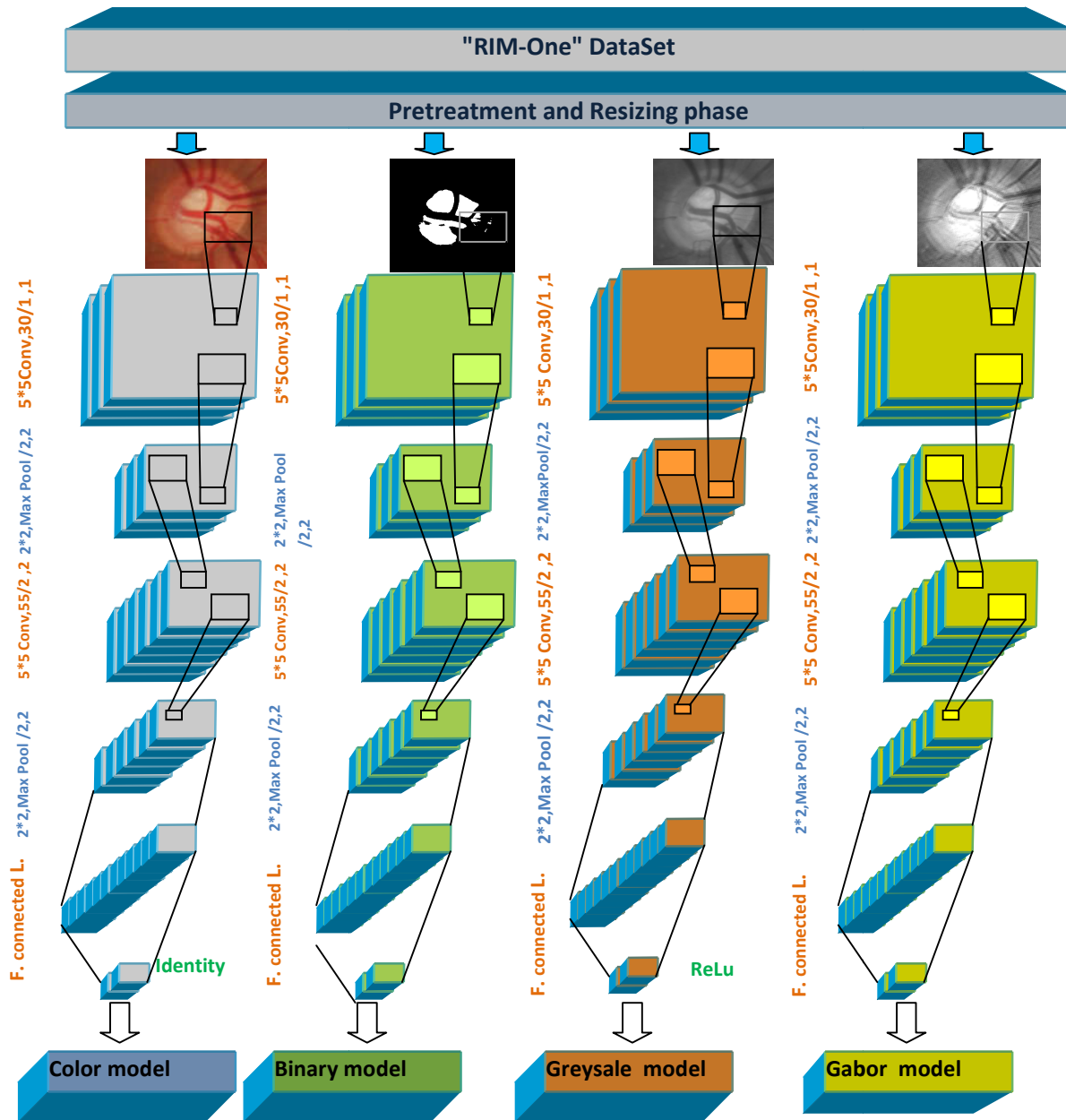


Figure.3.8. phase d'apprentissage [34]

IV.2.2. Phase de test:

Notre jeu de données Rim-One a été partitionné au 70% des images pour la phase d'apprentissage et 30% pour la phase de test. Nous laissons quelques images non étiquetées pour la phase de test. Une fois les quatre modèles sont générés pendant la phase d'apprentissage (couleur, binaire, niveau de gris et filtre de Gabor), le système global sera capable de diagnostiquer n'importe quelle image à l'entrée. Nous combinons les sorties des quatre classifieurs CNN en utilisant la fonction de fusion "Somme et Produit". Où le diagnostic sera attribué à la classe avec la probabilité globale la plus élevée, comme indiqué dans la figure 3.9.

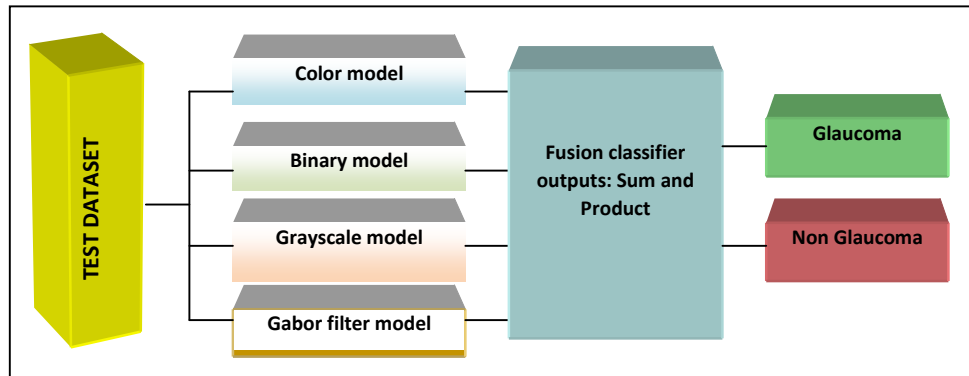


Figure 3.9 : phase de test [34]

V. Description du jeu de données

«*RimOne*» est une base de données d'images rétiniennes ouvertes pour l'évaluation du nerf optique. Elle a été référencée dans de nombreuses études concernant le système de diagnostic assisté par ordinateur telle que les algorithmes de segmentation de la tête du nerf optique, le diagnostic du glaucome et l'interprétation des images de la rétine [133].

L'ensemble de données **RimOne** est l'un des ensembles de données publiques les plus connus dans le domaine de l'interprétation du glaucome, en raison de sa richesse en quantité et qualité d'images. Toutes les images ont été collectées à différents âges et à différents niveaux d'éclairage dans la photographie du fond de l'œil. Il est exclusivement axé sur la segmentation ONH, il dispose d'images de haute résolution (169) et de segmentations de référence manuelles de chacun (5).

Trois hôpitaux ont contribué au développement de cette base de données: l'hôpital universitaire des Canaries, l'hôpital clinique San Carlos et l'hôpital universitaire Miguel Servet .

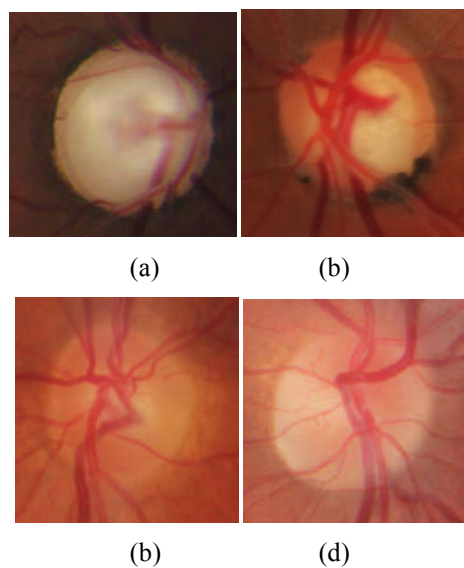


Figure 3.10 :Exemples d'images numériques du fond de l'œil:

(a, b) glaucome, (c, d) normal [133].

VI. Résultats expérimentaux

Afin d'évaluer la méthode de classification multimodale proposée pour la maladie du glaucome. Le modèle CNN proposé est exécuté à l'aide de la bibliothèque de programmation Deeplearning4j, qui a été conçu pour Java et nécessite Java Virtual Machine (JVM) et des framework [134].

Les résultats obtenus seront bien détaillées dans la section suivante.

VI.1 Les mesures d'évaluation

L'évaluation des performances de n'importe quel système de classification est une étape très importante parce que ces performances reflètent la fiabilité du système proposé. Dans les applications médicales et plus particulièrement celle à deux classes, le taux de bonne classification et d'erreur globale ne suffisent pas pour juger les performances.

Dans ce travail, nous avons analysé le comportement du classifieur CNN pour évaluer l'efficacité du diagnostic du glaucome en appliquant plusieurs mesures de performance [135]. Les mesures adoptées sont définies comme suit :

- **Matrice de confusion**

La matrice de confusion ou tableau de contingence est un outil utilisé pour mesurer les performances d'un modèle d'apprentissage automatique en vérifiant spécifiquement l'exactitude de son problème de prédiction et de classification par rapport à la situation réelle. C'est une matrice de taille égale au nombre de classes qui permet la génération des performances de chaque classe prise individuellement.

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Figure 3. 11 : Matrice de confusion

- **Taux de bonne classification (Accuracy)**

Taux de bonne classification (**Accuracy**) est l'indicateur le plus naturel et le plus évident permettant d'évaluer les performances d'un système de classification. Cette valeur simple à calculer, détecte la fréquence à laquelle notre modèle est correct:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.4)$$

Où:

- TP: Les vrais Positifs présents les échantillons de cas de glaucome avec un test positif.
- TN: Les vrais négatifs présentent les échantillons de cas non glaucome avec un test négatif.
- FP: Les faux positifs présentent les échantillons de cas non glaucome avec un test positif.
- FN: Les faux négatifs présentent les échantillons de cas de glaucome avec un test négatif

- **La sensibilité:**

C'est le nombre de vrais positifs divisé par le nombre de vrais positifs et le nombre de faux négatifs :

$$sensibilité = \frac{TP}{TP+FN} \quad (3.5)$$

- **La spécificité:**

C'est le nombre de vrais négatifs divisé par le nombre de vrais positifs et le nombre de faux négatifs:

$$Specificity = \frac{TN}{TN+FP} \quad (3.6)$$

- **La précision :**

C'est le nombre de vrais positifs divisé par le nombre de vrais positifs et de faux positifs:

$$PPV = \frac{TP}{TP+FP} \quad (3.7)$$

- **F1-score:**

C'est une moyenne harmonique de précommeision et de rappel, elle est considérée comme parfaite à 1 et est une défaillance totale à 0

$$F1 = \frac{((1+\beta^2)*prec*rec)}{((\beta^2*prec)+rec)} \quad (3.8)$$

Le paramètre β permet de pondérer la précision ou le rappel et vaut généralement 1.

VI.2 Analyses empiriques des paramètres de β plus –CNN

Pour concevoir un système de classification puissant basé CNN. Nous avons associé à chaque modalité, plusieurs tests empiriques pour définir les paramètres de chaque classifieur afin de générer de meilleures performances.

Dans ce travail, nous avons étudié :

- l'impact de la fonction d'activation
- le nombre d'époques
- le nombre d'itérations

VI.2.1. L'impact du nombre d'époques et du nombre d'itérations:

Avant de commencer notre apprentissage, le choix du nombre d'époques et d'itérations est très important. Nous avons remarqué qu'un faible nombre d'époques peut conduire à des mauvais résultats et qu'un grand nombre peut conduire à un sur-apprentissage. Nous pouvons définir quelques résultats de nos expériences:

- **Avec une époque et une itération**

Tableau 3.1 : Les performances des quatre modalités avec une époque et une itération

	Accuracy	PPV	Sensitivité	F1 score
Modèle Couleur	51,38%	56,06%	54,84%	58,82%
Modèle Binaire	57,81%	71,45%	51,87%	8,16%
Modèle Niveau de gris	50,70%	50,70%	50%	1,63%
Modèle filtre de Gabor	56,87%	78,30%	5071%	2,82%

- **Avec 30 époques et 10 itérations**

Tableau 3.2 : Les performances des quatre modalités avec 30 époques et 10 itérations

	Accuracy	PPV	Sensitivité	F1 score
Modèle Couleur	43,75%	43,75%	50%	60,87%
Modèle Binaire	99,06%	99,01%	99,09%	98,93%
Modèle Niveau de gris	79,72%	84,37%	79,98%	82,69%
Modèle filtre de Gabor	56,25%	56,25%	50%	65,01%

VI.2.2. L'impact de la fonction d'activation.

Dans notre cas, nous avons essayé d'utiliser la fonction "ReLU" comme fonction d'activation au lieu de la fonction "Identité" pour la modalité couleur et la modalité binaire. Ensuite, nous avons utilisé la fonction d'activation "Identité" au lieu de la fonction d'activation "ReLU" pour les deux autres modalités (filtre de Gabor et niveaux de gris), tout en modifions les paramètres de nombre d'époques et le nombre d'itérations et nous avons obtenu les résultats suivants:

• Avec 60 époques et 20 itérations

Tableau 3.3 : Les performances des quatre modalités avec 60 époques et 20 itérations

	Accuracy	PPV	Sensitivité	F1 score	Function d'activation
Modèle Couleur	63,75%	53,75%	50%	60,87%	ReLU
Modèle Binaire	99,38%	99,30%	99,44%	99,29%	ReLU
Modèle Niveau de gris	49,30%	49,30%	50%	66,04%	Identité
Modèle filtre de Gabor	66,25%	65,95%	60,33%	68,46%	Identité

• Avec 30 époques et 10 itérations

Tableau 3.4: Les performances des quatre modalités avec 30 époques et 10 itérations

	Accuracy	PPV	Sensitivity	F1 score	Function d'activation
Modèle Couleur	53,66%	53,75%	50,22%	60,87%	ReLU
Modèle Binaire	99,38%	99,30%	99,44%	99,29%	ReLU
Modèle Niveau de gris	49,30%	49,30%	50%	66,04%	Identité
Modèle filtre de Gabor	56,25%	56,25%	50%	68,10%	Identité

Après une série des tests empiriques détaillée ci-dessus, les paramètres qui nous ont permis d'obtenir les meilleures performances lors de la phase d'apprentissage sont bien détaillés dans la section suivante.

VI.2.3. Les paramètres choisis du modèle CNN adopté

Pour générer un système de classification robuste basé CNN associés à différentes modalités, nous avons réalisé plusieurs tests empiriques afin de fixer les paramètres de chaque modèle de classification générant les meilleures performances. Les paramètres sélectionnés sont fixés comme suit :

Les couches entièrement connectées prennent les images filtrées de haut niveau et les traduisent en votes. Dans notre cas, nous devons choisir entre deux catégories, "Glaucome" et "Non-Glaucome" ; pour cela, nous avons choisi une couche entièrement connectée avec la fonction d'activation CUBE.

Après une série d'expérimentations détaillée ci-dessus, les paramètres qui nous ont permis d'obtenir les meilleures performances lors de la phase d'apprentissage sont les suivants.

Pour les modèles couleur et binaire, nous avons adopté 60 époques et 20 itérations et le choix de la fonction d'activation « l'identité ». En revanche, pour les modèles niveaux de gris et filtre de Gabor, nous avons opté pour 60 époques et 20 itérations et « ReLU » comme fonction d'activation.

Le tableau ci-dessous illustre les paramètres choisis :

Tableau 3.5 Paramètres choisis de notre CNN.

	Première couche convolutive	deuxième couche convolutive
Nombre de filtres	30	55
Taille de filtre	5*5	5*5
Stride	1,1	1,1
Fonction d'activation	fonction "identité" (pour les deux modèles couleur et binaire) la fonction "ReLU" (pour les deux modèles niveau de gris et gabor).	
Nombre de canaux	valeur = 3 (pour images couleur) valeur =1(pour les images niveau de gris, binaire et filtre de gabore)	
Couche Pooling (MaxPooling)	2*2	2*2
Stride	2,2	2,2

VI.2.4. Les résultats de la phase d'apprentissage du système globale β plus -CNN

Nous avons partitionné l'ensemble de données de la base RimOne qui contient 320 images en 70% pour l'apprentissage et 30% pour le test.

- **Matrice de confusion:**

	Glaucoma	Non Glaucoma		Glaucoma	Non Glaucoma
Glaucoma	139	1		Glaucoma	140
Non Glaucoma	4	176		Non Glaucoma	2
(a)			(b)		
	Glaucoma	Non Glaucoma		Glaucoma	Non Glaucoma
Glaucoma	146	9		Glaucoma	136
Non Glaucoma	17	153		Non Glaucoma	4
(c)			(d)		

Figure 3.12 Matrice de Confusion de la phase d'apprentissage: (a) Modèle couleur, (b) Modèle binaire, (c) Modèle niveau de gris, (d) Modèle filtre Gabor.

En résumé, pendant la phase d'apprentissage: 315 images ont été apprises par le modèle couleur, 318 ont été bien apprises par le modèle binaire, 312 ont été apprises par le modèle de filtre de Gabor et seulement 299 ont été apprises par le modèle en niveaux de gris. Nous pouvons affirmer, avec certitude, que notre système a obtenu d'excellents résultats dans cette phase et les performances obtenues sont :

VI.2.5. Les performances obtenues

Plusieurs mesures de performance dans la phase d'apprentissage ont été utilisées. Le tableau 3.6 résume les valeurs de mesures de la précision, de la sensibilité, la spécificité, la PPV et la VPN pour les quatre modèles CNN ; nous remarquons que les résultats sont élevés dans la phase d'apprentissage dont les meilleurs sont présentés par les modèles couleur et binaire. La figure 3.13 illustre l'évolution du score (erreur moyenne) en fonction du nombre d'itérations et la figure 3.14 présente les informations sur les paramètres d'apprentissage. (Nous avons choisi le modèle couleur comme exemple).

Tableau 3.6: Résultats obtenus des différentes **métriques des modèles** proposés en phase d'apprentissage

	Accuracy	Sensitivity	Specificity	PPV	NPV
Color model CNN1	98.44%	99.29%	97.78%	97.20%	99.44%
Modèle Binaire CNN2	99.38%	100.00%	98.89%	98.59%	100.00%
GrayscaleCNN3	92.00%	94.19%	90.00%	89.57%	94.44%
Modèle filtre Gabor CNN4	97.50%	97.14%	97.78%	97.14%	97.78%

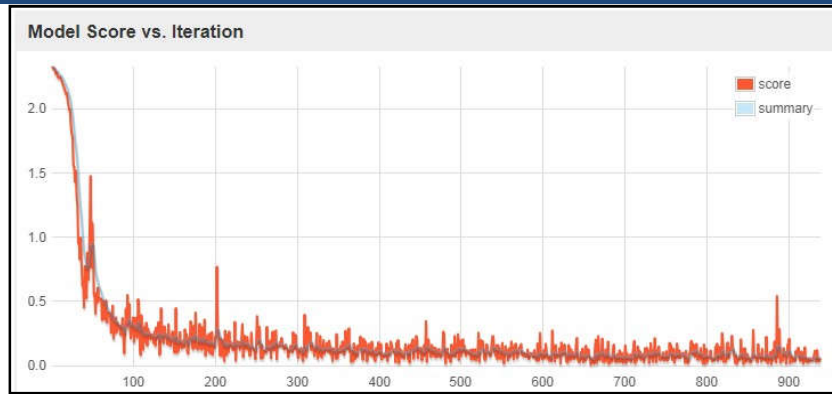


Figure 3.13. Score du modèle de couleur par rapport à l'itération [34]

Model and Training Information	
Model Type	MultiLayerNetwork
Layers	6
Total Parameters	431080
Start Time	
Total Runtime	
Last Update	2018-05-16 13:44:41
Total Parameter Updates	938
Updates/sec	3,28
Examples/sec	104,92

Figure 3.14. Aperçu des paramètres générés utilisant le modèle Couleur [34].

VI.2.6. Les résultats de la classification du système global β plus -CNN

Après avoir défini les paramètres optimaux pour chaque classifieur, nous avons validé le système d'aide au diagnostic en utilisant 135 images pour tester notre ensemble de données. Les résultats obtenus sont résumés comme suit :

• Matrice de confusion de la phase de test

	Glaucoma	Non Glaucoma			
Glaucoma	57	3	Glaucoma	54	6
Non Glaucoma	3	72	Non Glaucoma	7	68
(a)			(b)		
	Glaucoma	Non Glaucoma			
Glaucoma	35	25	Glaucoma	55	5
Non Glaucoma	5	70	Non Glaucoma	6	69
(c)			(d)		

Figure 3.15: Matrice de Confusion de phase de test pour les quatre modèles
 (a) Modèle couleur, (b) Modèle binaire, (c) Modèle niveau de gris, (d) Modèle filtre Gabor.

En résumé, pour la phase de test, 129 images ont été étiquetées avec précision par le modèle couleur, ce qui est le résultat le plus élevé, 124 images ont été étiquetées par modèle de filtre Gabor, 122 par modèle binaire et 105 images ont été étiquetées par le modèle en niveaux de gris.

VI.2.7. Les performances obtenues en phase de test

Le tableau 3.7 résume les résultats obtenus des quatre modèles CNN.

Tableau 3.7 : Résultats obtenus des différentes métriques de la méthode proposée en phase de test.

	Accuracy	Sensibilité	Specificité	PPV	NPV
Modèle Couleur CNN1	95.56%	95.00%	96.00%	95.00%	96.00%
Modèle Binaire CNN2	90.37%	90.00%	90.67%	88.52%	91.89%
Modèle niveau de gris CNN3	77.78%	58.33%	93.33%	87.50%	73.68%
Modèle filtre Gabor CNN4	91.85%	91.67%	92.00%	90.16%	93.24%

VI.2.8. Comparaison des résultats

D'après les résultats obtenus dans le tableau 3.7, nous remarquons que le modèle couleur génère les meilleurs résultats, ainsi, le modèle basé sur le filtre de Gabor fournissant des résultats très satisfaisante. Cela nous confirme l'importance de la texture comme source de représentation.

Pour les modèles individuels et selon le tableau 3.7, nous discernons que les deux CNN 1 et CNN4 représentant respectivement le modèle couleur et le modèle filtre de Gabor ont obtenu les meilleurs performances.

VI.2.9. Fusion parallèle (agrégation Somme / Produit)

Notre objectif principal est de parvenir à un système qui combine les résultats des quatre classificateurs afin que le système global β plus-CNN soit en mesure de diagnostiquer toute entrée d'image rétinienne inconnue

Afin de bénéficier de la diversité existante entre les différents modèles générés dont le choix du meilleurs n'est pas évident, une approche de fusion des sorties de ces derniers est suggérée. Les méthodes d'agrégation « somme » et « produit » ont été employées et ont permis une amélioration de la décision finale par rapports aux différents systèmes prix individuellement ;

En effet, après l'application des méthodes d'agrégation, qui sont la somme et produit, sur les sorties des quatre classificateurs, les matrices de confusion générées sont la suivante.

- **Matrices de confusion**

	Glaucoma	Non Glaucoma			Glaucoma	Non Glaucoma
Glaucoma	58	2		Glaucoma	59	1
Non Glaucoma	3	72		Non Glaucoma	3	72
(a) For the Sum aggregation method			(b) For the Product aggregation method			

Figure 3.16: matrices de Confusion Pour: (a) Méthode aggregation somme, (b) méthode Produit.

Tableau 3.8 Résultats obtenus des différentes métriques après les méthodes d'agrégation somme / produit.

	Accuracy	Sensitivité	Specificité	PPV	NPV
Sum aggregation method	96.30%	96.67%	96.00%	95.08%	97.30%
product aggregation method	97.04%	98.33%	96.00%	95.16%	98.63%

IV.1.1. Analyse des résultats par la courbe ROC

Dans le domaine médical, il est très important d'analyser le comportement des systèmes d'aide au diagnostic médical en utilisant des paramètres locaux tels que la spécificité et la sensibilité au lieu de juger le système sur la base de mesures de performances globales.

La courbe des caractéristiques de fonctionnement du récepteur (ROC), qui est un graphique représentant les performances d'un modèle de reconnaissance pour tous les seuils de classification, est une autre mesure applicable utilisée pour prouver les performances de la classification. Cette courbe trace le taux de vrais positifs en fonction du taux de faux positifs [136].

L'aire sous la courbe ROC (AUC) mesure l'exhaustivité de l'aire bidimensionnelle sous l'ensemble de la courbe ROC (par calculs intégraux) de (0,0) à (1,1). Pour calculer l'AUC, nous utilisons un taux positif réel et taux de faux négatifs

$$TPrate = \frac{TP}{TP+FN} \quad (3.9)$$

$$FP rate = \frac{FP}{FP+TN} \quad (3.10)$$

$$AUC = \frac{(1+TP rate-FP rate)}{2} \quad (3.11)$$

Le tableau 3.9 présente les résultats du taux des meilleures mesures de classification: sensibilité, spécificité ainsi que AUC.

Tableau 3.9 Les performances des quatre classifieurs.

	Sensitivité	spécificité	AUC
Modèle Couleur CNN1	95.00%	96.00%	95,4%
Modèle Binaire CNN2	90.00%	90.67%	99,5%
Modèle niveau de gris CNN3	58.33%	93.33%	75,7%
Modèle filtre de Gabor CNN4	91.67%	92.00%	92,9%
Sum aggregation method	96.67%	96.00%	96,6%
Product aggregation method	98.33%	96.00%	97,3%

Les figures 3.17 et 3.18 illustrent les graphiques générés représentant les courbes ROC associées aux quatre modalités ainsi que les méthodes d'agrégation de somme et produit respectivement. Nous observons que les modèles couleurs et celui basé sur filtres de Gabor ont de meilleures valeurs que les modèles en niveaux de gris et binaires, cependant, nous pouvons déduire que notre système est fiable, car tous les points des quatre courbes sont en haut de la moitié de l'espace ROC (cela signifie au-dessus du dessin de ligne de référence de couleur jaune).

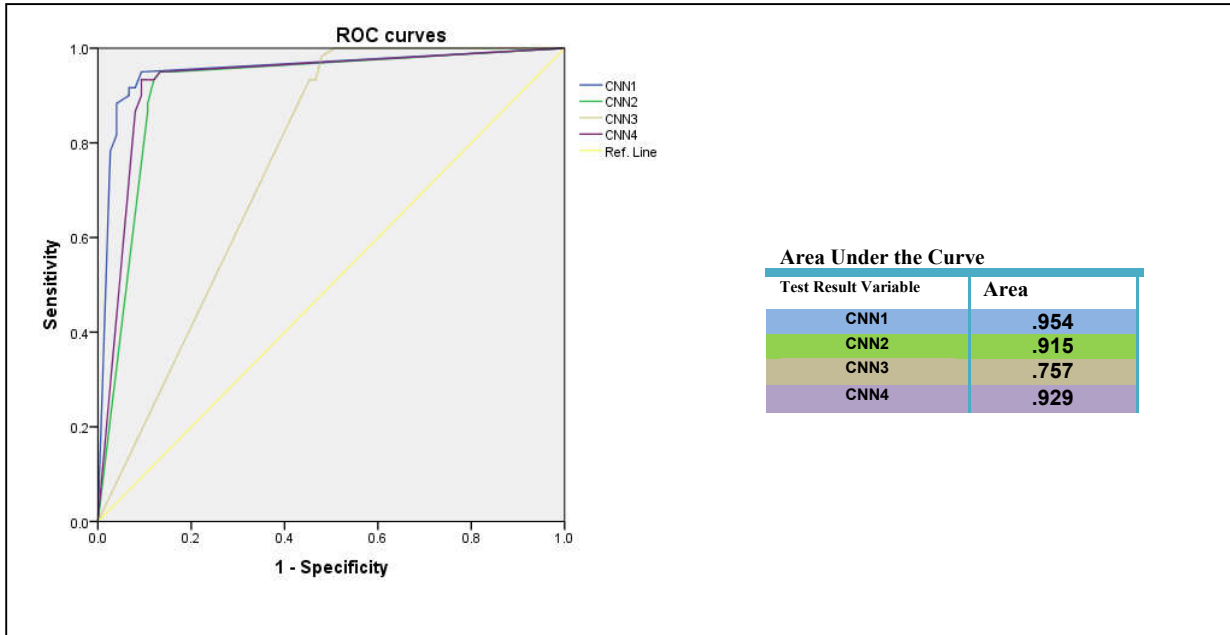


Figure 3.17 : Les courbes ROC et les valeurs AUC pour les quatre classifieurs [34].

À partir des secondes courbes ROC présentées dans la figure 3.18, **représentant les deux méthodes de fusion des différents modèles CNNs** (méthode d'agrégation de somme et méthode d'agrégation de produit), **les résultats sont très satisfaisants car d'une part**, tous les points des courbes respectives sont sur la moitié supérieure au-dessus de la ligne de référence ; **d'autre part**, nous pouvons **conclure** que l'agrégation **par la méthode** « produit » donne de meilleurs résultats selon le tableau 3.9 **en plus de** la figure 3.18.

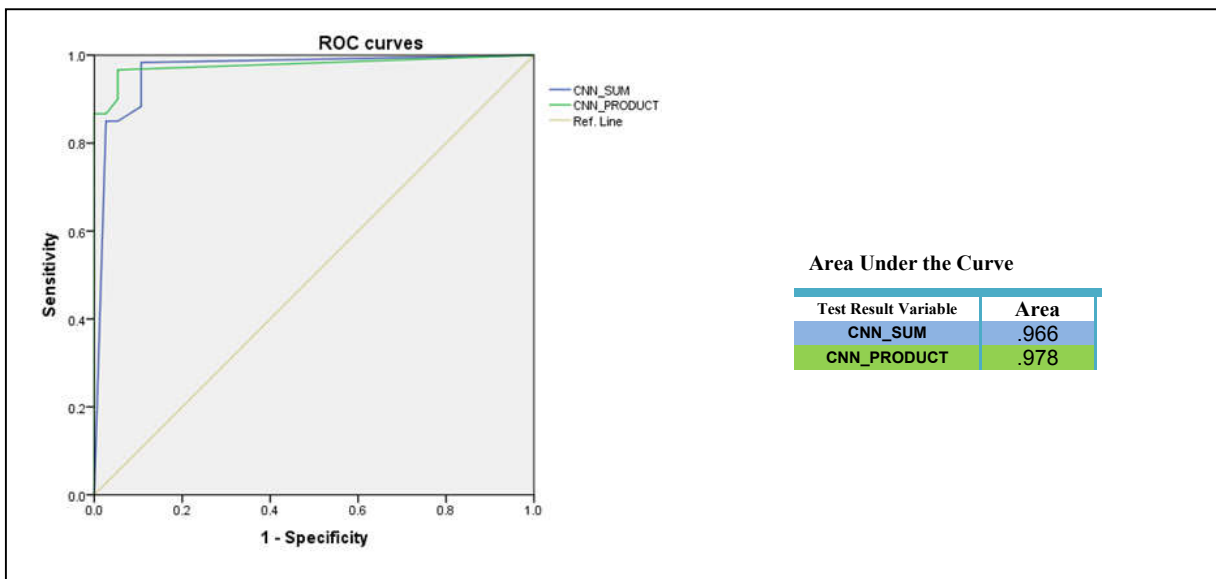


Figure 3.18 : Les courbe ROC et les valeurs AUC pour la méthode de somme et d'agrégation [34].

• Discussion des résultats

De nombreux travaux de recherche se sont concentrés sur l'analyse des images du fond rétinien pour localiser le nerf optique pour mieux diagnostiquer la maladie du glaucome. Aujourd'hui, l'apprentissage en profondeur joue un rôle important dans le système de diagnostic assisté par ordinateur en particulier pour classer automatiquement les images médicales afin de surmonter les problèmes de la longue consultation et la rentabilité des techniques de l'industrie médicale [137].

Nous proposons dans cette première contribution une nouvelle architecture nommée β plus-CNN, un système combinant quatre modèles de classification basé CNN pour diagnostiquer l'image rétinienne. Notre système a permis d'assurer de bonnes performances grâce à plusieurs facteurs :

D'abord, le choix de l'architecture CNN, permettant l'extraction automatique et intelligente des caractéristiques par élimination de l'étape traditionnelle dont le choix était très difficile.

Deuxièmement, nous avons utilisé une base de 475 images en entrée des modèles CNNs, car l'utilisation de plus d'images assure toujours de meilleurs résultats [30, 138]. Troisièmement, nous avons opté pour un redimensionnement de l'image (100x100) pour être adapté à notre structure, car une image de petite taille peut perdre des informations significatives [139].

Quatrièmement, l'idée principale de notre approche basée sur la fusion de différentes modalités d'images rétiniennes, afin de bénéficier de la complémentarité des différents classifieurs y compris la représentation de la texture de Gabor, a donné un plus à notre travail.

D'après les résultats mentionnés dans le tableau 3.9, où nous avons résumé le plus d'études basées sur les architectures CNN pour le diagnostic du glaucome, nous pouvons conclure que les résultats obtenus sont très convaincants et encourageants, indiquant une grande capacité de classification par rapport aux autres études du même domaine.

Notons que selon le tableau 3.7, nous remarquons que les deux CNNs: (1) et (4), représentant les images couleur et les images du filtre de Gabor en entrée (respectivement), ont obtenu **les** meilleurs **taux se basant sur les** différentes métriques de performances comparées aux deux autres CNNs (modèles en niveaux de gris et binaires). Les modèles CNNs basés sur Gabor et la couleur ont atteint une précision accrue d'environ (91,85% et 95,56%). Ces résultats prouvent les informations significatives offertes par les deux modalités correspondantes qui permettent au modèle CNN de mieux détecter la pression appliquée sur le nerf optique.

Nous sommes allés plus loin et nous optons pour la fusion des sorties de ces quatre classifieurs par deux fonctions d'agrégation, à savoir somme et Produit. Où le diagnostic final sera attribué à la classe ayant la probabilité globale la plus élevée.

- Sur la base du tableau 3.8, les résultats de performances obtenus des méthodes d'agrégation (somme et produit) étaient meilleurs **en comparant avec les classifieurs pris individuellement**. Ces résultats prouvent l'intérêt d'utiliser le modèle d'apprentissage d'ensemble (ensemble learning).

- À partir des résultats présentés dans le tableau 3.8, nous pouvons conclure que la méthode d'agrégation de produits a obtenu les meilleurs résultats avec une précision de 97,04% en comparant par la méthode d'agrégation Somme qui a atteint une précision de 96,30%.

Les courbes ROC illustrées à la figure 3.17 et à la figure 3.18 montrent que notre système est vraiment fiable où chacun des points des quatre courbes des quatre CNN se trouvait sur la moitié supérieure de l'espace ROC. On retrouve le même raisonnement si l'on consulte la figure 3.18 où tous les points des deux courbes des modèles somme et produit se situent au-dessus de la ligne de référence (couleur jaune).

V. Conclusion

La première contribution de notre thèse représente une approche globale **nommée β plus-CNN**, c un système d'aide au diagnostic médical pour la classification des images. L'objectif était la conception et la mise en œuvre d'une application pour aider à diagnostiquer la maladie à partir d'images rétinienne, en déterminant la présence ou l'absence de glaucome.

Un système ensembliste basé sur la coopération des différents modèles CNNs ou apprend la maladie à partir d'une source particulière de l'image rétinienne. **Ce choix est justifié par la capacité de ces modèles d'améliorer les taux de reconnaissance de formes et de classification**. C'est un champ de recherche très vaste. **Dans notre cas, nous avons opté pour une approche ensembliste parallèle utilisant quatre modèles CNNs entraînés par quatre sources d'images**: couleur, binaire, niveau de gris et Gabor dans le but d'assurer **une représentation optimale de la maladie**.

Pour valider notre approche, **la base de données publique** RimOne, qui est sollicitée par la majorité des systèmes de classification des maladies du glaucome opérant sur des images rétinienne.

Dans cette étude, nous avons proposé la fusion de plusieurs réseaux de neurones convolutifs pour diagnostiquer le glaucome à l'aide de données multi-sources à partir d'images du fond d'œil. Tout d'abord, les images ont été redimensionnées et prétraitées. Ensuite, nous avons appliqué plusieurs algorithmes sur les images couleur afin de générer plusieurs modalités représentées comme suit: binaire, niveau de gris et filtre de Gabor.

À l'issue de l'analyse des résultats, nous remarquons que le **modèle** couleur et le **modèle basé** Gabor donnent des meilleurs résultats.

En revanche, β plus-CNN a réussi à apprendre et à classer les images rétinienne avec des performances élevées, principalement parce que l'approche d'ensemble learning a obtenu des résultats de prédiction de modèle toujours supérieurs à celles **obtenus par les modèles individuels**.

Ces expériences confirment la supériorité de la qualité des performances de notre modèle proposé par rapport aux autres méthodes de pointe dans le même domaine.

Dans la prochaine contribution nous allons proposer une approche **locale se focalisant sur la zone d'intérêt en intégrant un module de** segmentation sémantique de la zone de disque optique basé sur une architecture encodeur-décodeur (U-net). **Une fois extraites**, une phase de classification des images résultantes en glaucome et non glaucome dans le but de comparer entre les deux approches globale et locale.

**Chapitre IV : Une Approche Locale Combinant
deux Phases de Segmentation du Disque Optique
pour le Diagnostic du Glaucome**

I. Objectif et motivation

Suite aux résultats obtenus dans la première contribution qui ont un aspect global et qui nous ont permis d'assurer la classification des images rétinienne par une architecture fiable basée sur l'apprentissage profond, plus particulièrement les réseaux convolutifs CNNs, où le nombre de couches et les hyper-paramètres ayant été choisis soigneusement après plusieurs études et tests empiriques.

Malgré que les résultats étaient satisfaisants, nous avons voulu se rapprocher plus de la zone d'intérêt (RoI) qui est le disque optique afin de mieux interpréter les images rétinienne. Notre deuxième contribution est la proposition d'un nouveau schéma de segmentation sémantique basée sur l'architecture encodeur-décodeur (U-Net). Une fois générées, les images détectées sont par la suite classifiées. Nous avons validé notre système sur la base des images rétinienne « REFFUGE ». La base d'images « REFFUGE » contient deux catégories d'images qui représentent en réalité un seul échantillon (*ou individu*), où chaque image de fond d'œil est présentée également avec son annotation correspondante. Afin de booster la phase d'apprentissage avec un grand nombre d'images, la base subit une phase d'augmentation de données, pour assurer la diversité de données et d'avoir des meilleures performances du système de détection.

II. Introduction

Une autre façon d'aborder le diagnostic de la maladie du glaucome en se basant sur les images rétinienne est de se concentrer sur la zone porteuse de la maladie (zone d'intérêt) toujours dans le but d'améliorer les performances de classification des images rétinienne. L'objectif de cette contribution est « comment détecter et localiser ces zones (ROI) ? » et « qu'elle est la différence de classer une image globale et une zone d'intérêt ? »

Le processus général de notre système est divisé en deux phases essentielles de segmentation : la première est basée sur la localisation et la coupure de la zone de disque optique de l'image du fond de la rétine. Ensuite, une architecture très élégante d'apprentissage profond qui est le « U-Net » est adaptée pour assurer la segmentation sémantique de disque optique et cupule optique.

Récemment, la plupart des études de diagnostic assistées par ordinateur se sont converties vers les architectures d'apprentissage en profondeur basées CNN, pour surmonter les limites des techniques traditionnelles. L'approche Encodeur-Décodeur est une nouvelle architecture très adoptées dans plusieurs études pour améliorer les performances de la segmentation et de la détection automatique des images médicales.

Le traitement des images du fond rétinien est le moyen de détecter les troubles oculaires à un stade précoce. Le disque optique et cupule optique, sont les caractéristiques les plus appropriées qui peuvent être segmentés automatiquement à partir de l'image de fond d'œil pour analyser l'état pathologique de l'œil [141]. Cependant, l'analyse des résultats basée sur l'indicateur oculaire: Ratio (CDR) facilite la confirmation de la décision finale du diagnostic [142].

De nombreuses études sont basées sur l'architecture U-Net pour améliorer leurs systèmes. L'U-Net est la première architecture encodeur-décodeur dédiée spécialement pour la segmentation des images médicales et a été proposé par Ronneberger et al [101]. Par la suite, beaucoup d'architectures de diagnostic du glaucome sont inspirées de U-Net: Yu et al, [143] ont proposé une nouvelle version de l'architecture U-Net. les auteurs ont utilisé le modèle ResNet-34 comme section d'encodeur de sous-échantillonnage de l'architecture U-Net mais ils ont conservé la section de décodeur sur-échantillonnage de la structure U-Net originale qui était formée par un bloc d'attention de contexte global (GCA) proposé et un bloc de regroupement de pyramides de compression et d'excitation (SEPP). Cette architecture implémentée sur RIGA, DRISHTI-GS, RIM-ONE et sur un ensemble de données privé, où ils ont obtenu des résultats satisfaisants. Dans le même contexte, Sébastopolsky et al [144], ont également modifié l'architecture U-Net pour la segmentation du disque optique et de la cupule et ils ont introduit une modification dans le chemin de contraction qui prend l'architecture typique du réseau de classification (VGG-16) avec moins de filtres dans toutes les couches convolutives, ce qui donne de meilleurs résultats avec moins de temps d'entraînement. Aussi, Fu et al [145], ont proposé M-Net, en tant que réseau profond multi-étiquettes pour la segmentation conjointe du disque optique et de la cupule optique des images du fond d'œil, il s'agit d'un système d'apprentissage en profondeur de bout en bout, où les auteurs appliquent une transformation polaire sur l'image du fond d'œil, ils évaluent ce système sur les ensembles de données ORIGA et SCES et ils ont obtenu des résultats convenables: aires sous courbe (AUC) de 0,85 et 0,90 sur les ensembles de données ORIGA et SCES respectivement. Toujours dans le même sujet, Al-Bander et al [29], ont proposé une nouvelle approche pour la segmentation simultanée de l'OD et de l'OC dans les images du fond de la rétine en combinant un réseau entièrement convolutif avec DenseNet (FC-DenseNet) en utilisant un grand ensemble de données de fond de couleur rétinienne (ORIGA, DRIONS-DB, Drishti-GS, ONHSD et RIM-ONE) afin d'atteindre la supériorité dans leurs résultats.

Le travail le plus récent dans ce domaine est basée sur (REFFUGE) Challenge, proposé par Orlando et al[146], qui repose sur deux tâches de segmentation du disque et cupule optique et de la classification du glaucome en utilisant un grand ensemble de données de 1200 images de fond

d'œil. 12 équipes de travail ont créé un cadre unifié pour comparer différents modèles d'architectures de deep learning. Les résultats ont fait le buzz dans ce challenge.

Le but de cette contribution est de proposer une approche locale qui combine les deux tâches essentielles de l'interprétation d'images qui sont la segmentation et la classification. **Les principales étapes de cette contribution sont les suivantes :**

- Se focaliser sur la zone d'intérêt de fond d'œil de la rétine (disque optique et cupule optique) qui sont les repères les plus utilisés dans les applications d'analyse des images rétinienne.
- Localisation et coupure de la zone de disque optique et cupule optique.
- Bénéficier également des techniques d'augmentation des données pour avoir des meilleures performances.
- Création d'un modèle d'apprentissage basé U-Net qui permettra par la suite de segmenter le disque et cupule optique d'une façon automatique.
- Classification des images résultantes de la phase de segmentation avec une architecture basée CNN.

III. Motivation de la segmentation/détection automatique des points de repère de la rétine (Retina land marks) :

Un diagnostic précoce peut prévenir jusqu'à 98% des pertes de vision sévères chez les personnes comme la rétinopathie diabétique et le glaucome. Des images numériques de la rétine sont utilisées pour le dépistage des maladies ; d'où leur analyse automatisée est désormais cruciale pour traiter des données à grande échelle.

L'Organisation Mondiale de la Santé (OMS) classe l'hypertension parmi les 10 principaux facteurs qui causent les maladies de la rétine dans les pays développés.

De même, le diabète est en hausse dans le monde (plus de 200 millions de personnes), la croissance démographique, le vieillissement et la prévalence de l'obésité, tout ça cause des maladies plus sévères de la rétine qui peuvent conduire les gens facilement à la cécité.

Pour ce qui a été mentionné précédemment, une infrastructure solide et une analyse automatisée sont nécessaires pour diagnostiquer et traiter ces maladies et lutter contre la cécité avec plus d'efficacité et moins de coût.

Récemment, le développement des techniques d'acquisition des images numériques ainsi l'évolution des domaines de programmation informatique permettent d'impressionnantes élévations

de capacités de stockage, de transfert et de quantification des images rétinienne. Cela rend les domaines de recherche d'analyse automatisée efficace et rapide.

La communauté de recherche sur la rétine a fait un travail important dans le développement des algorithmes et des méthodes pour l'analyse des images de la rétine et les diagnostiquer. Ces systèmes d'aide au diagnostic spécifiques à une maladie nécessitent la localisation des points de repère de la rétine : disque optique, macula et les vaisseaux sanguins ainsi leurs formes qui peuvent facilement aider à découvrir les anomalies dans le plus précoce stade.

Par exemple : La localisation de disque optique (OD) peut être utilisée pour améliorer les performances des algorithmes de détection des lésions lumineuses, d'autre part, l'emplacement de OD et la macula aide également à déterminer la gravité de la maladie et la nécessité d'une intervention. Aussi, la forme des les vaisseaux sanguins jouent un rôle important dans la compréhension de l'état médical de la rétine qui aide le médecin à prendre une décision finale.

A la fin, nous pourrons dire que la performance du système d'aide au diagnostic médical est directement proportionnelle à la précision des systèmes de détection de disque optique, de macula et les vaisseaux sanguins [147].

IV. Quels sont les points de repère à segmenter dans le fond d'œil de la rétine (Retina land marks) ?

Le fond de l'œil est segmenté afin de localiser et /ou d'isoler les repères rétinien qui montre les désordres pouvant survenir à l'œil. Les images rétinienne ont trois principaux points de repère visibles: Disque optique (OD), macula et vaisseaux sanguins, qui seront bien détaillés par la suite :

IV.1. Segmentation du disque optique:

Le disque optique est un repère clé dans les images rétinienne, il apparaît comme une structure elliptique brillante dans les images de la rétine, qui montre les changements liés aux maladies y compris, le glaucome et les rétinopathies diabétiques à un stade précoce. L'emplacement et la structure de disque optique sont importants pour deux aspects essentiels: c'est un facteur de confusion dans le contexte de la segmentation des lésions lumineuses (il faut donc le détecter) et sa morphologie peut être directement utilisée pour la détection du glaucome. Le disque optique sert également comme repère afin de localiser d'autres zones caractéristiques telles que la macula et les vaisseaux sanguins.

La majorité des recherches récentes sont tournées vers cette voie [101, 143, 144, 145,29]. Cependant, avec l'introduction des nouvelles approches de l'apprentissage profond et surtout celle basées sur la segmentation sémantique (architecture encoder-decoder)

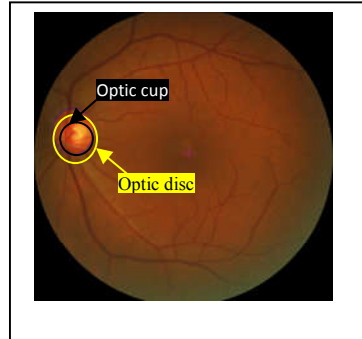


Figure 4.1: image de fond d'œil montrant le disque optique et la cupule optique [147]

IV.2. Segmentation de la macula et de la fovéa:

La macula est la région centrale de la rétine qui mesure environ 3 à 4 mm de diamètre, elle apparaît comme une petite région sombre et ne contient pas de vaisseaux. Au centre de la macula se trouve une petite dépression sous forme d'une zone ronde sombre mesurant environ 1 mm de diamètre appelée fovéa, l'emplacement de la macula peut être utilisé pour juger une gravité d'une anomalie détectée sur la rétine, sa localisation est importante pour détecter les maladies associées telles que l'œdème maculaire [148]. Même s'il existe des difficultés à définir les frontières de la macula, l'identification de la macula est vraiment indispensable car l'émergence de toute structure de maladie dans celle-ci nécessite une attention immédiate pour éviter la perte de vision chez le patient. Plusieurs recherches sont publiées dans ce domaine telles que : Giachetti et al [149], ont proposé une approche basée sur la notion de forme car la macula et le disque optique sont presque circulaires sauf que la macula est sombre et ne contient pas de vaisseaux tandis que le disque optique est brillant et il est intersecté par des vaisseaux. Ensuite, Tewari et al [150], ont proposé une technique de symétrie radiale rapide pour trouver les composants circulaires brillants et sombres dans les images de la rétine.

En outre, Aquino [151], a proposé une méthodologie qui combine à la fois les connaissances anatomiques antérieures et la segmentation de la macula pour exploiter les avantages des deux approches.

Aussi, le concept d'apprentissage profond, qui a prouvé son potentiel dans de nombreux domaines, est utilisé pour l'identification des maladies maculaires, chose qui a mené Burlina et al [152], à

utiliser les réseaux de neurones profonds pour apprendre efficacement des caractéristiques de l'image pour détecter la dégénérescence maculaire liée à l'âge dans 5000 images.

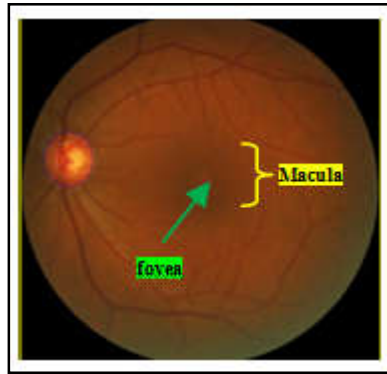


Figure 4.2: image de fond montrant Macula et fovéa[147]

IV.3. Segmentation vasculaire (veines et artères):

Les vaisseaux sanguins apparaissent comme une structure de branches d'arbre sombre, la séparation des vaisseaux sanguins des autres structures de la rétine est une étape de prétraitement clé dans de nombreuses applications d'imagerie rétinienne. L'analyse de ces vaisseaux est importante pour la détection des maladies de la circulation sanguine telles que la rétinopathie diabétique. Au cours de la dernière décennie, les réseaux de neurones convolutifs ont démontrés une robustesse dans ce domaine et beaucoup d'architectures ont été proposées pour assurer des meilleures performances telle que :

Maninis et al [153], sont parmi les premiers qui ont utilisé les CNNs pour la détection vasculaire rétinienne. Cette architecture se compose d'une première branche de 5 couches convolutives (masque 3×3 , pas de 2) dans le but de faire l'extraction des caractéristiques à différentes échelles, qui sont par la suite redimensionnées et concaténées. Par conséquent, le vecteur généré, contient donc la sortie de chacune des couches précédentes, il sera l'entrée de la dernière couche convolutive (de masque 1×1) qui effectue leur combinaison linéaire pour prédire la carte vasculaire. Aussi Fu et al[154], prennent également en charge une approche multi-échelle utilisant l'architecture Deep Vessels. Cette fois la branche de convolutions successives est composée de 4 étages convolutifs chacun étant composé de plusieurs couches convolutives (2 ou 3) et d'une couche de *regroupement* (max-pooling). Chaque étage double le nombre de neurones de sorties (64, 128, 256, 512) mais divise par deux la résolution de ces sorties, donc, chaque étage procède à une échelle supérieure que l'étage précédent. Les caractéristiques générées de ces 4 étages sont redimensionnées à la résolution de l'image initiale et corrélée par une couche exploitant les champs aléatoires conditionnels (Conditional Random Fields, CRF) pour prédire une carte de segmentation. Cette approche donne

de très bons résultats et avec 95% de précision sur DRIVE data set. En 2017, Lahiri et al [155] proposent une approche d'entraînement adversaire, dans laquelle deux réseaux sont entraînés simultanément par deux fonctions de coût opposées, où le premier génère des images semblables aux images d'entraînement et le second tente de discriminer les pixels provenant d'un fond d'œil, d'un vaisseau ou d'une fausse image générée par le premier réseau. En outre, Taibou et al [156] ont proposé un modèle de segmentation des vaisseaux sanguins rétiniens basé sur un réseau de neurones à convolution de type auto-encodeur, appliquée sur les deux bases publiques DRIVE et SATARE.

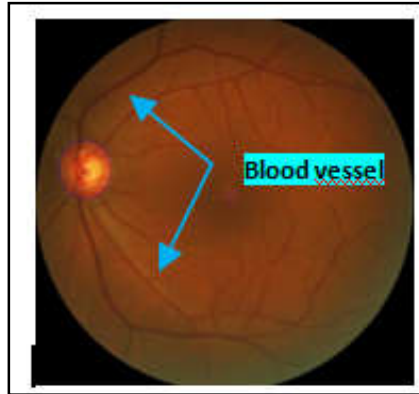


Figure 4.3: image de fond montrant les vaisseaux sanguins [147]

V. Système proposé :

Étant donné que le disque optique et la cupule optique sont les principaux paramètres pour la détection précoce du glaucome, nous présentons dans cette contribution une approche locale du système de diagnostic assisté par ordinateur inspiré des techniques de Deep-Learning et basée sur l'architecture U-Net nommée **DSC-U-Net**. Ce système combine deux phases essentielles de segmentation: la première étant une pré-segmentation basée sur la localisation et la coupure de la zone de disque optique de l'image du fond de la rétine dans le but de se focaliser sur la zone d'intérêt (ROI) pour éviter l'utilisation redondante des sources de calcul et écarter les caractéristiques inutiles. Tandis que la deuxième utilise l'architecture «U-Net» pour faire une segmentation affinée de la zone de disque optique (DO). La classification des images résultantes en glaucome ou non glaucome sera abordée en dernier par une architecture CNN bien choisie.

Afin de comparer les performances de classification, deux systèmes sont mis en œuvre, le premier étant passé en extrayant la région d'intérêt (ROI) des images de fond en localisant et coupant la zone du disque optique pour se concentrer davantage sur cette zone, autrement-dit, une pré-segmentation de ces images avant qu'elles seront segmentées par U-Net, tandis que, le second système passe directement par la segmentation U-Net. Notre approche est appliquée à l'ensemble de données «REFFUGE» qui contient au départ deux types d'images qui représentent réellement un seul

échantillon, où chaque image de fond d'œil est présentée avec son annotation ou le masque correspondant.

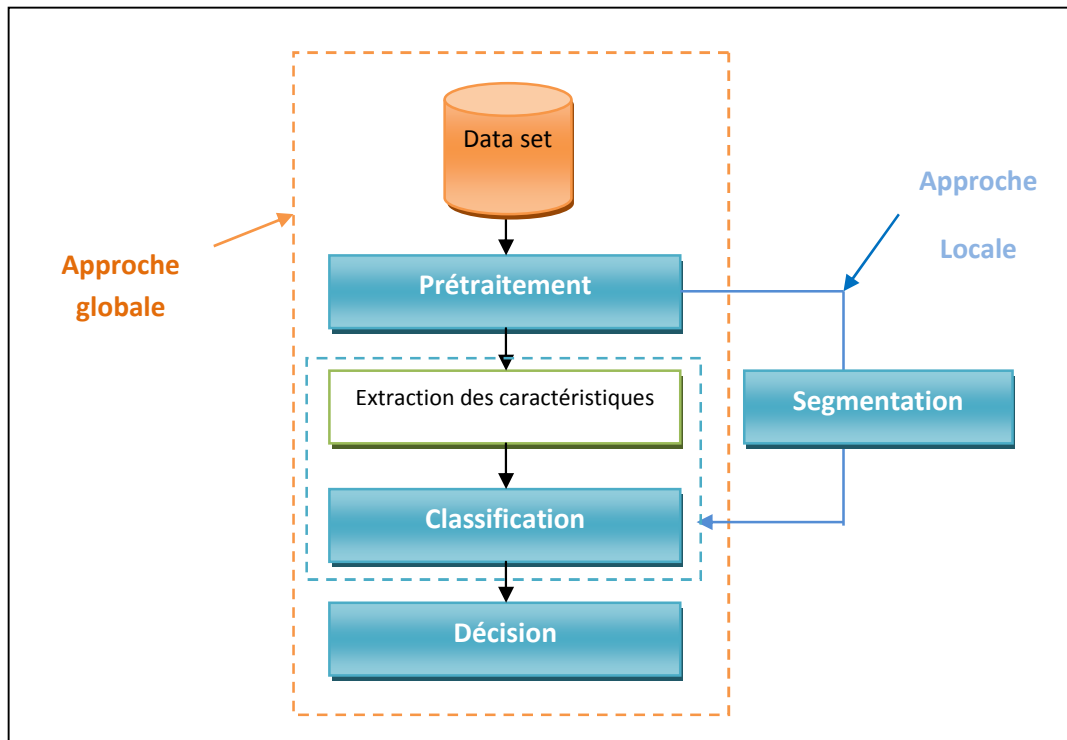


Figure 4.4 : L'approche locale et l'approche globale [162]

V.1. Etape 1 : La segmentation

V.1.1. La pré-segmentation :

Dans le domaine de traitement d'images, le prétraitement joue un rôle important dans l'obtention des résultats idéaux dans tout système d'aide au diagnostic médical (SAD), en particulier, dans les processus de segmentation et l'extraction des caractéristiques. Cette étape vise à améliorer la qualité de l'image médicale originale et à rendre les étapes ultérieures du SAD plus faciles et plus fiables. Cette phase s'effectue après l'acquisition des images et ces opérations sont diverses. Dans notre cas elle comprend la localisation et la coupure de l'emplacement du disque optique.

- **Localisation et coupure de l'emplacement du disque optique**

Dans cette contribution, et pour mieux se focaliser sur la zone d'intérêt (disque optique), nous avons inspiré de la technique de Gilberto Plaza González proposée en 2017 [157], qui permet de détecter la partie du disque optique dans l'image rétinienne.

Notons que le disque contient une cupule optique, qui peut être identifiée par sa luminosité, aussi l'ensemble du disque optique a une géométrie approximativement circulaire et il est plus lumineux

que le reste de la rétine, autrement dit, l'idée de localiser le disque est de trouver cette zone la plus lumineuse de l'image.

Pour localiser et couper la zone d'intérêt (RoI), on doit suivre un processus qui contient plusieurs opérations détaillé comme suit:

- D'abord, ajuster la taille de toutes les images à analyser afin de créer un système rigide et de gérer n'importe quelle taille d'image, les dimensions choisies sont 1024x1024 pixels.
- Décomposer l'image RVB en trois canaux de niveaux de gris pour voir si nous pouvons utiliser les informations fournies par l'un d'entre eux.

D'après l'analyse de ces images, on peut conclure que l'image du canal rouge fournit trop de zones lumineuses et que l'image du canal bleu peut introduire du bruit, dans ce cas nous avons choisi de localiser le disque en utilisant le canal vert.

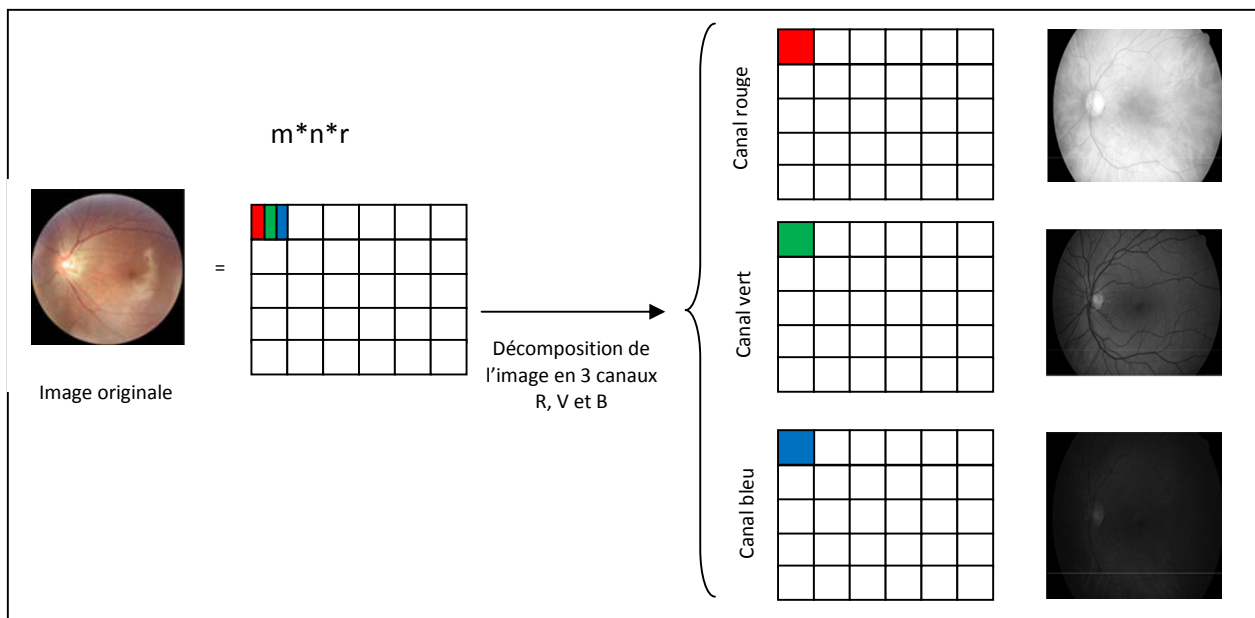


figure 4.5: décomposition de l'image de fond d'œil en trois canaux rouge, vert, bleu [162]

- Eliminer les bruits par l'application d'un filtre gaussien de taille de 15 x 15 pour supprimer tout bruit ou défaut dans l'image. La taille de ce filtre est efficace car si nous augmentons trop cette taille, l'image devient floue.
- Appliquer l'algorithme d'ouverture morphologique des niveaux de gris avec un noyau de 15x15 en utilisant un disque comme élément structurel, ce qui éliminera tout bruit qui pourrait être mal interprété, autrement-dit éviter de chevaucher avec d'autres organes ou formes qui existe dans la surface de l'image (les vaisseau sanguins, petites zones d'intensité lumineuse élevées). l'effet principal de cette algorithme et de faire délatrer le contour de la zone la plus lumineuse (disque optique) pour facilite l'opération de localisation.

- Localiser la zone de disque : comme nous l'avons dit, le choix le plus judicieux est de trouver la zone la plus lumineuse de l'image pour le canal sélectionné qui est le canal vert. Pour cette raison, nous recherchons simplement les valeurs moyennes dont l'intensité est la plus élevée dans un voisinage de 3x3.
- Une fois que nous connaissons l'endroit de la zone la plus lumineuse, nous prenons un point arbitraire dans cette zone et nous élargissons notre champs par le choix d'une fenêtre assez grande de taille 120x120 capable de contenir le disque optique qui est la zone d'intérêt ROI.
- Coupure de la zone ROI : Une fois la zone de disque optique est localisée, nous avons ajouté un module qui fait la sauvegarde des cardinalités de la zone du disque. Ces cardinalités sont:

P_x, P_y : qui représentent d'où commence la partie localisée sur les deux axes x et y

L_x, L_y : représentent la longueur de la zone localisée dans les deux axes x et y

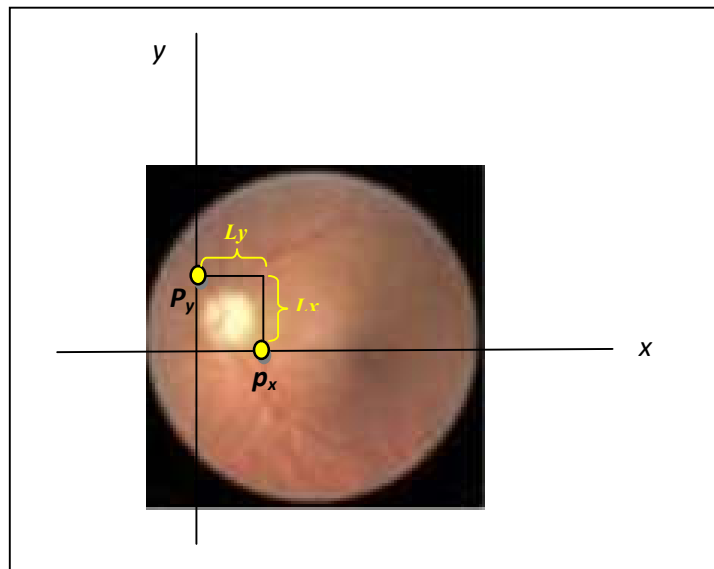


figure 4.6: localisation et coupure de la zone lumineuse selon les paramtres (P_x,P_y, L_x,L_y) [162]

- La coupure d'annotation pour chaque image selon les cardinalités sauvegardées sera alors appliqués . Notons que notre base utilisée REFFUGE est composée de deux sortes d'images, pour chaque image originale, on a une annotation correspondante.

Cette phase est résumée par le schéma suivant:

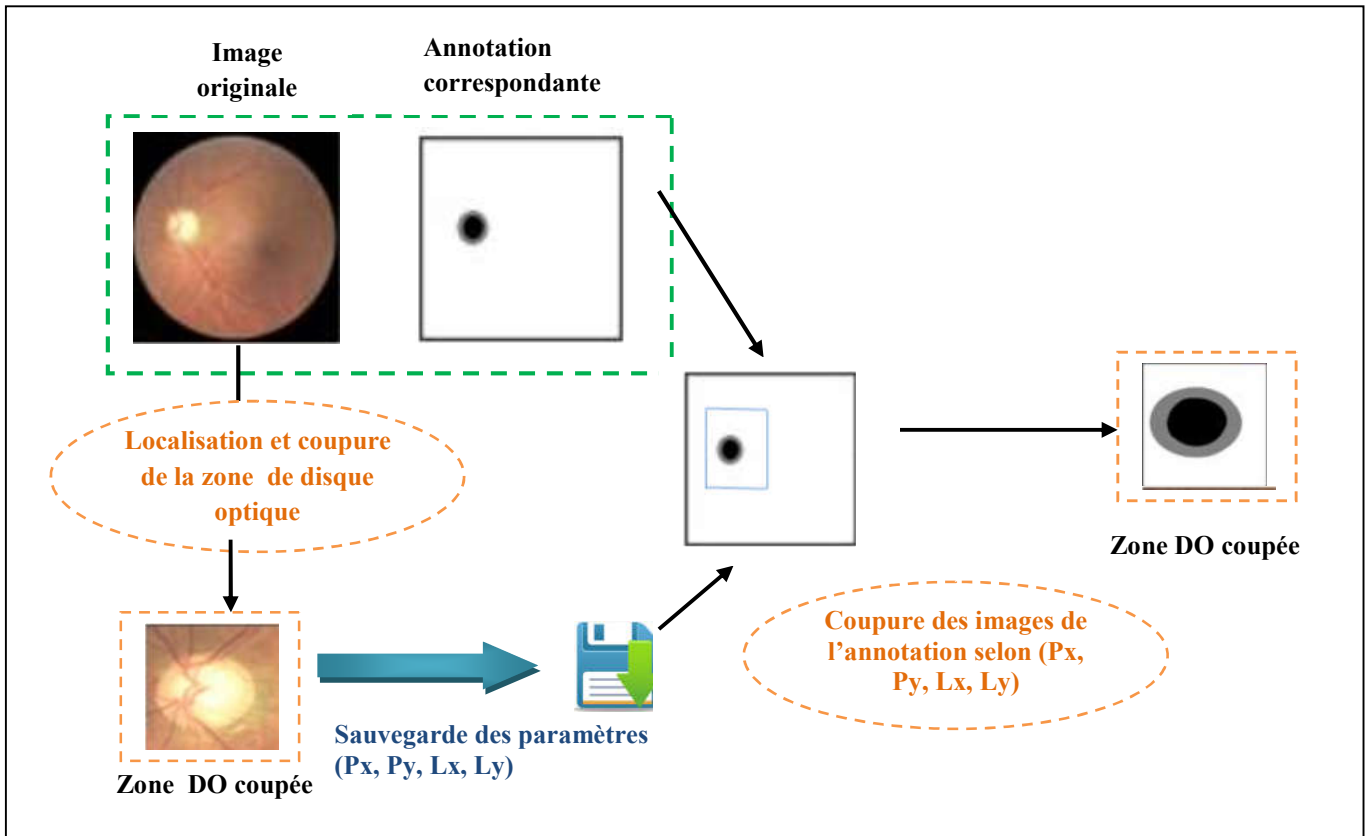


Figure 4.7 : la phase de pré-segmentation [162]

A la sortie de cette phase de pré-segmentation, notre base sera enrichie avec d'autres sortes d'images qui sont plus concentrées sur la zone disque optique, c'est-à-dire on trouve chaque image de zone disque optique coupée « OD » avec son image d'annotation coupée correspondante.

En d'autres termes, nous pouvons trouver quatre modalités d'images du fond d'œil, ces modalités sont:

- L'image originale
- L'annotation correspondante de l'image originale
- L'image pré segmentée avec zone disque optique «DO» coupée
- L'annotation le l'image pré segmentée avec zone de disque optique «DO» coupée

Voici un exemple de la localisation et de la de coupure de la zone du disque optique :

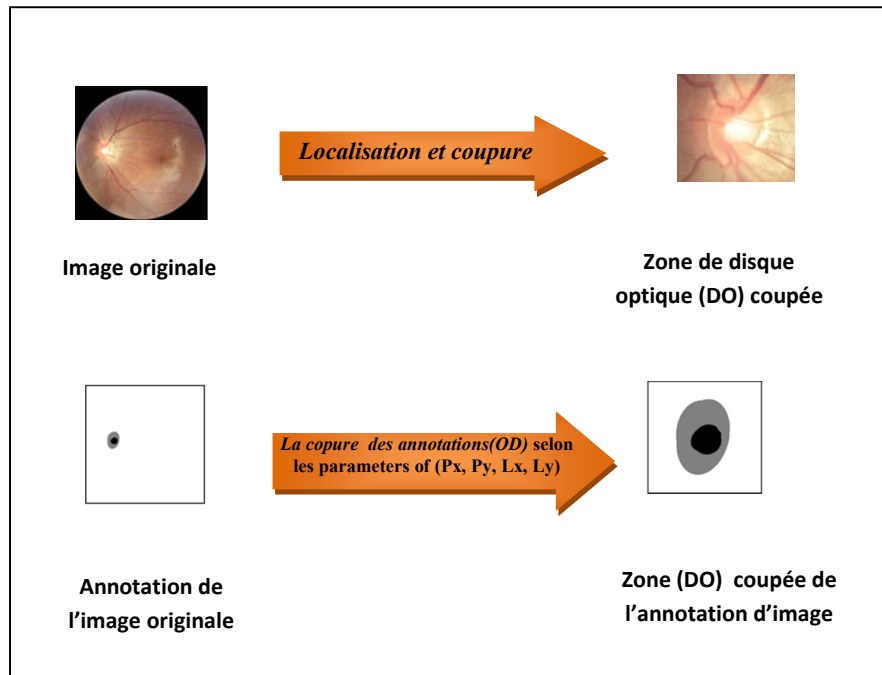


Figure 4.8 Exemple de localisation et coupure de disque optique [162]

V.1.2. La segmentation avec U-Net

Une fois que nous avons la base de données prétraitée focalisée sur la zone d'intérêt, nous abordons la détection automatique du disque optique basée sur l'architecture encodeur-décodeur, connue sous le nom U-Net. La figure suivante présente les détails de l'architecture U-Net adoptée :

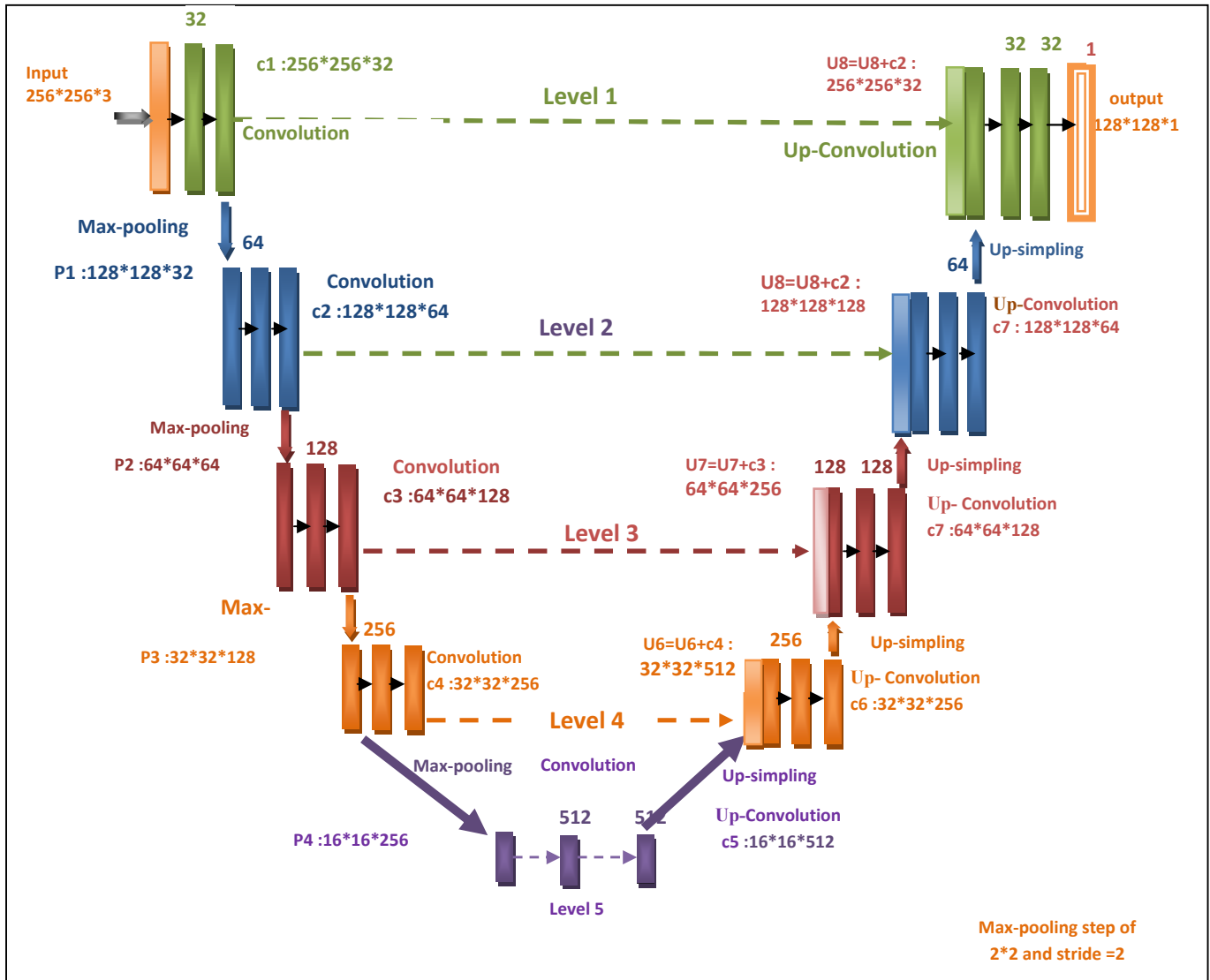


Figure 4.9: L'architecture U-Net [162]

“U-Net” est une architecture qui contient deux chemins symétriques : un chemin de contraction (côté gauche) suivi par un chemin expansif (côté droit). Le chemin de contraction suit parfaitement la même architecture d’un CNN.

Le U-Net adopté se compose de 5 niveaux, chaque image couleur de taille $(x \times x \times 3)$ à l’entrée du premier niveau va s’approfondir à travers les autres niveaux à gauche et suit la forme ‘U’ jusqu’à la sortie de U-Net à droite.

Chaque niveau de U-Net contient la même structure qui est deux couches de convolutions (3x3) avec un nombre variable de filtres qui sera doublé lors de passage d'un niveau à un autre (par exemple si le nombre de filtres est 16 dans le premier niveau il devient 32 dans le deuxième et 64 dans le troisième et ainsi de suite), chaque couche de convolution est suivi par une fonction d'activation ReLu et une couche max-pooling (2x2) avec un pas (stride) de 2, pour diviser la taille des images en deux afin de les réduire.

La fin de la partie gauche sera **systématiquement** le début de la partie droite qui contient **également** 5 niveaux faisant un travail opposé : Les couches convolutions sont remplacées par les couches up-convolution et les couches max-pooling sont remplacées par des couches Up-sampling .

La seule différence entre les deux chemins est le bloc de concaténation des cartes des caractéristiques qui **sauvegardent** les emplacements des informations pour les relocaliser sur les images de **sortie** . La **sortie** de la dernière couche de convolution est une image de type niveau de gris qui a la même taille de l'image d'entrée ($x*x*1$).

Pour faire une segmentation sémantique des images rétiniennes, nous créons tout d'abord un modèle U-NET qui nous permettra de segmenter n'importe quelle image de fond d'œil. Une fois ce modèle est compilé, il sera utilisé dans deux systèmes :

- Le premier système recevra la première catégorie d'images qui sont les images coupées de la zone du disque optique (DO) avec leurs annotations correspondantes
- Le deuxième système recevra la deuxième catégorie d'images qui sont les images originales du fond d'œil avec leurs annotations correspondantes.

L'entraînement des deux systèmes est effectué sur 70% des images de chacune des catégories où nous avons effectué des tests empiriques pour choisir le nombre des époques et d'itérations adéquates.

Au cours de l'apprentissage **du** modèle U-NET, une sauvegarde des meilleurs résultats est procédée selon les métriques DICE et IoU en comparant les images prédites par les annotations correspondantes. (voir les résultats dans la prochaine section)

Une fois que **l'apprentissage des** deux systèmes est assurée, une phase de test avec les 30% de la base des images restantes est effectuée pour évaluer l'aptitude de chacun des deux systèmes générés.

Une comparaison des résultats est réalisée pour évaluer l'effet de la coupure des images pour **assurer une** bonne segmentation.

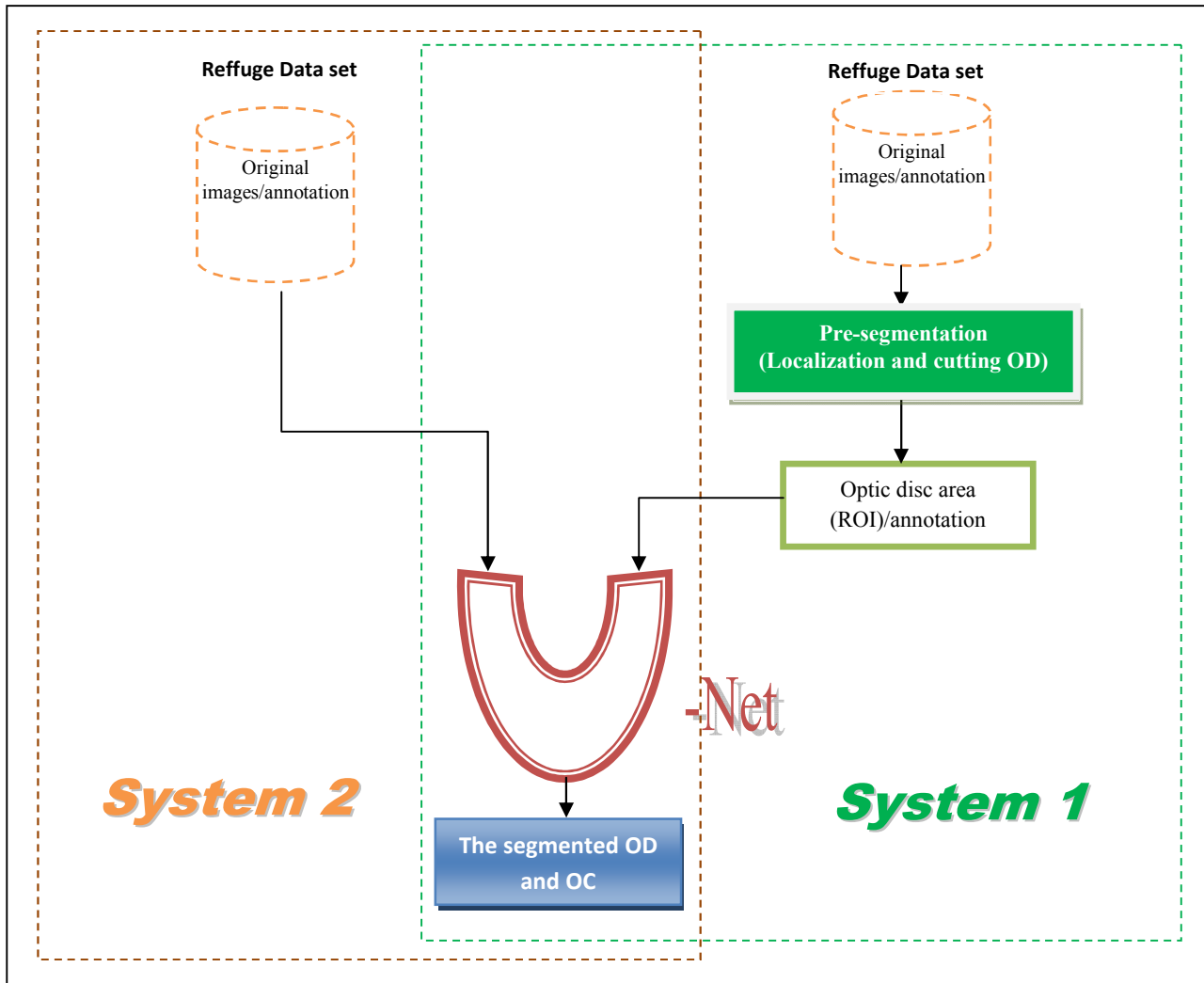


Figure 4.10 : la phase de segmentation par U-net [162]

Le système apprend les images avec leurs annotations (masques) correspondantes dans la phase d'apprentissage ; ce que lui **permettra** de segmenter n'importe quelle image en entrée **en générant** l'annotation prédite du disque optique en sortie.

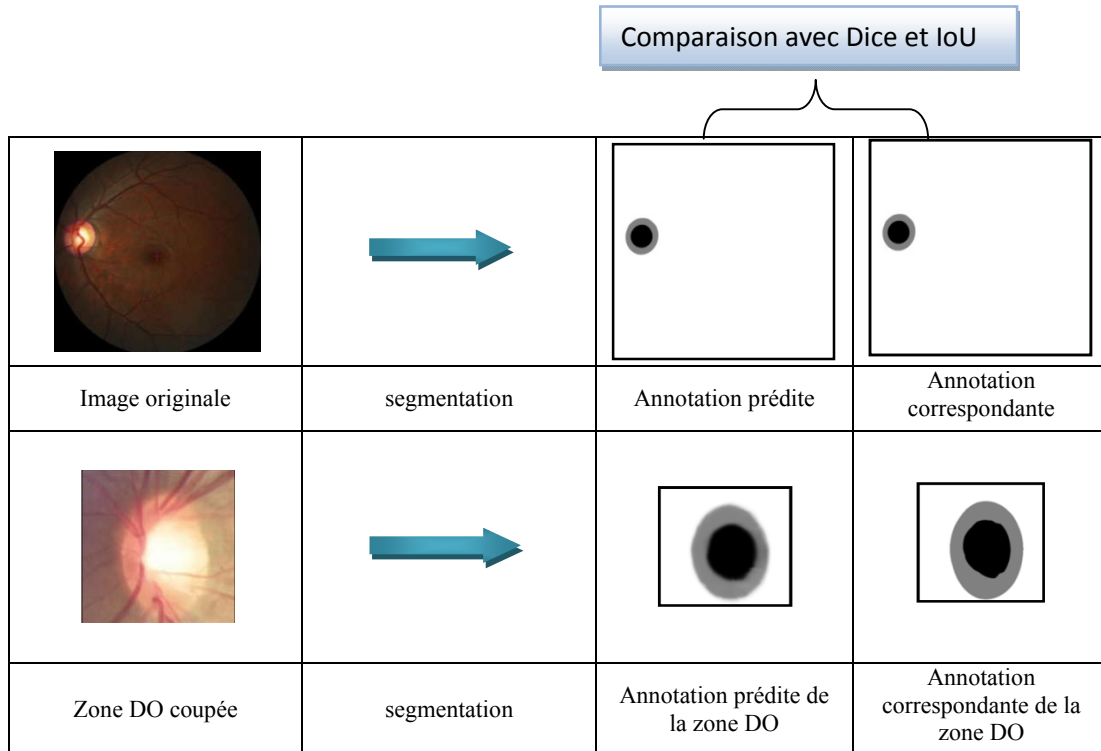


Figure 4.11 : Exemple d'images d'entrée et de sortie de l'étape de segmentation (phase de test) [162]

V.2. Étape 2 : classification avec l'architecture CNN

Afin de souligner l'impact de l'introduction de l'étape de pré-segmentation, nous avons abordé la classification des images résultantes des deux systèmes avec l'étape de pré-segmentation et sans elle.

Cette étape repose sur la classification de trois modalités d'images issues de deux systèmes en glaucome et non glaucome: le premier système **basé sur la base des** images **ayant** subi les deux étapes de segmentation (les images de la zone coupée «DO» prédites) ; tandis que le second système **repose sur les images extraites** par l'étape de segmentation de l'architecture U-Net, (les images originales prédites). Nous avons également effectué la classification des images originales pour **une meilleure analyse des** résultats dans la section de comparaison.

Nous pouvons résumer le schéma de classification comme suit:

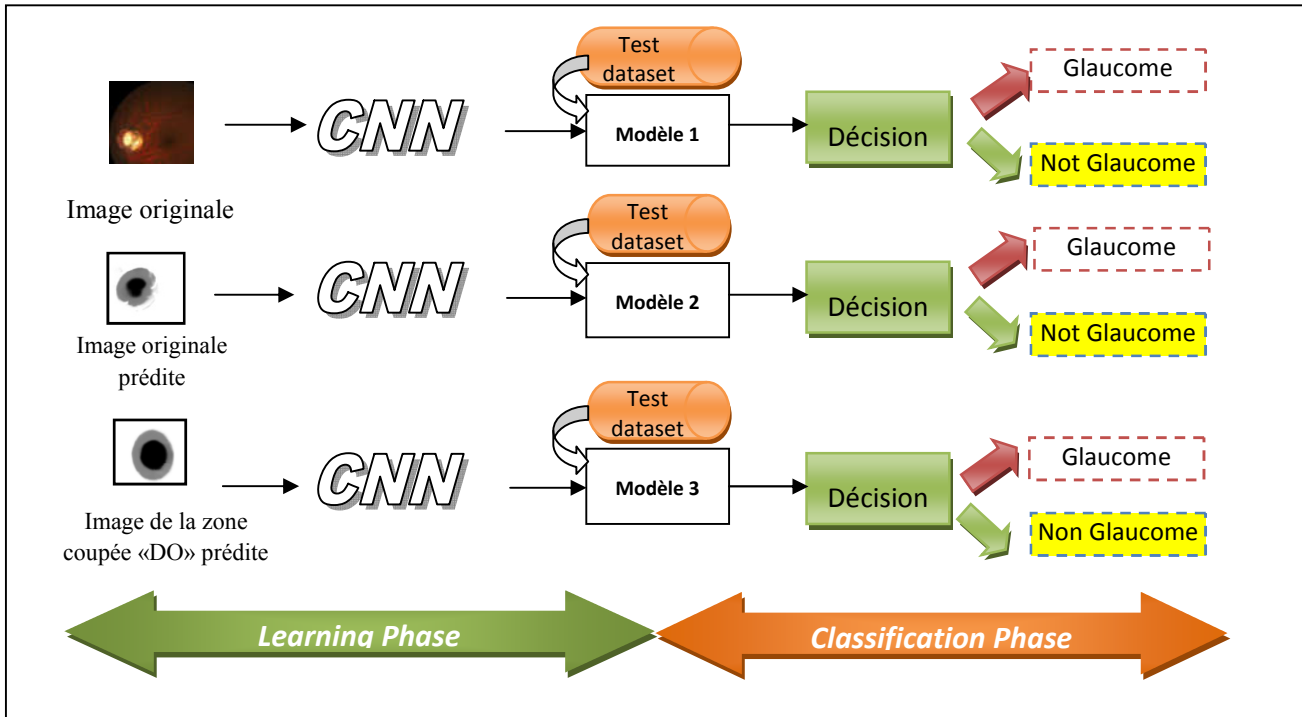


Figure 4.12 : la phase de classification [162]

- **L'augmentation des données:**

Les applications d'apprentissage en profondeur nécessitent toujours un nombre élevé d'images pour assurer la diversité des données. **Pour cela**, nous avons renforcé le système avec une étape d'augmentation des données pour obtenir de meilleurs résultats de performance. Nous avons sélectionné certaines techniques d'augmentation de données et les **avons** appliquées sur l'ensemble d'images originales et les images résultantes de la segmentation (l'annotation prédite) avant d'appliquer l'étape de classification.

Quatre techniques d'augmentation des données sont adoptées:

- Égalisation
- Miroir
- Rotation 45 °
- Rotation 90 °

- Nous avons commencé par 400 échantillons pour chaque image et nous avons obtenu 2000 images pour chacune après l'étape d'augmentation de données.

- Notre ensemble de données est partitionné en 70% pour la phase d'apprentissage, 30% pour le test (pour chaque modalité).

- Voici un exemple d'une image subie les techniques d'augmentation de données utilisées dans notre travail

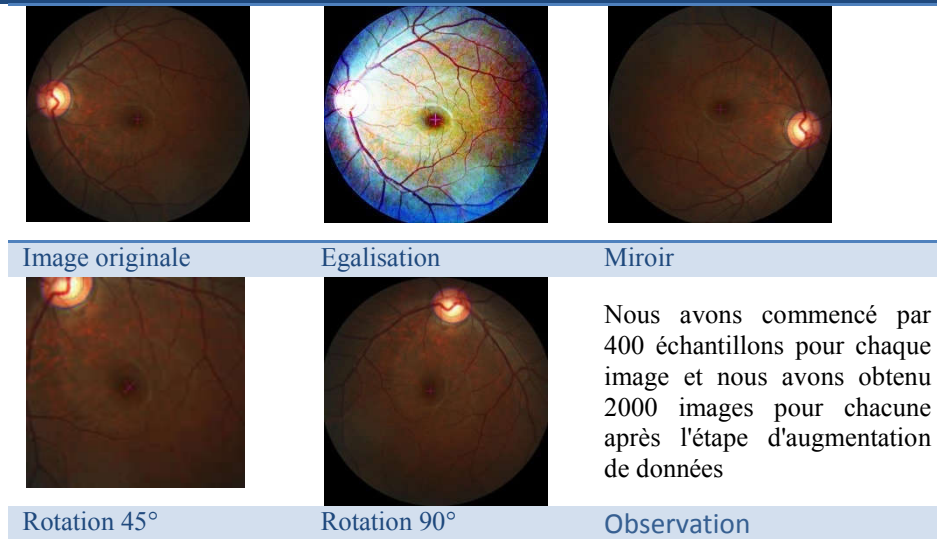


Figure 4.13: Les techniques d'augmentation de données utilisées dans notre approche [162]

V.2.1. Architecture CNN:

Une architecture CNN avec cinq couches de convolution est choisie, où chaque couche de convolution est suivie d'une couche de max pooling. La sortie de la dernière couche de pooling est connectée à la couche entièrement connectée afin de classifier notre jeu de données.

Tableau 4.1: les paramètres choisis de notre CNN

Les couches de convolutions	Paramètres
Conv1	Conv2D(32, kernel_size = (3, 3), activation='relu'
Max_pooling	size=(2,2)
Conv2	Conv2D(64, kernel_size=(3,3), activation='relu'
MaxPooling	size=(2,2)
Conv3	Conv2D(96, kernel_size=(3,3), activation='relu'
Max_pooling	size=(2,2)
Conv4	Conv2D(96, kernel_size=(3,3), activation='relu'
Max_pooling	size=(2,2)
Conv5	Conv2D(64, kernel_size=(3,3), activation='relu'
Max_pooling	size=(2,2)
Couche entièrement connectée	Soft-max

V.3. Description de l'ensemble de données REFFUGE

Reffuge est un grand ensemble de données [42] d'environ 1200 images annotées du fond rétinien utilisé pour analyser les maladies oculaires, divisé en 3 sous-ensembles (400 pour la l'apprentissage, 400 pour le test et 400 pour la validation). Toutes les images ont été acquises avec deux caméras de fond différentes (Zeiss Visucam 500 «2124x2056 pixels» et Canon CR-2 «1634x1634 pixels»). Cet ensemble de données peut être utilisé pour ces trois tâches [158]:

- a. Classification du glaucome clinique
- b. Segmentation du disque optique
- c. Localisation de la fovéa (centre maculaire)

Pour la tâche de segmentation, des annotations manuelles par pixel du disque optique et de la cupule ont été obtenues par des spécialistes du Zhongshan Ophthalmic Center, Université Sun Yat-sen, Chine. Elles sont stockés sous forme d'image BMP avec la même taille que l'image de fond de l'œil correspondante représentant par deux étiquettes Optic Disc (couleur grise) et Optic Cup (couleur noire).

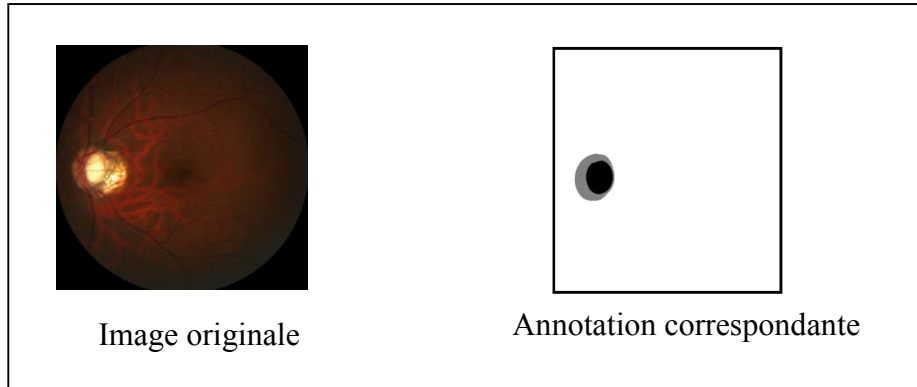


Figure 4.14: image originale et son annotation correspondante à partir de l'ensemble de données Reffuge [158]

Dans notre travail, nous nous sommes intéressés aux deux premières tâches (classification et segmentation) pour évaluer notre approche.

V.4. Résultats et discussion:

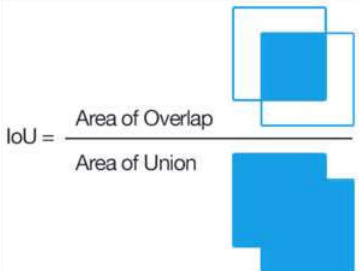
Cette section analyse les résultats des deux étapes: segmentation et classification

V.4.1. Mesures de performances:

Pour évaluer la performance de notre approche, nous avons résumé toutes les métriques que nous avons utilisées comme suit :

La première métrique utilisée dans ce travail est l'Intersection over Union qui est une métrique d'évaluation utilisée pour mesurer la précision d'un détecteur d'objet sur un ensemble de données particulier

- **Intersection over Union (IoU):** C'est essentiellement une méthode pour quantifier le pourcentage de chevauchement entre le masque cible et notre sortie de prédiction

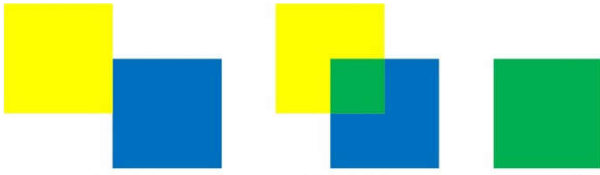


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (4.1)$$

$x \cap y$: Area overlap

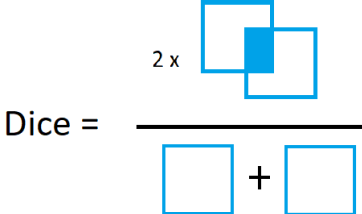
$x \cup y$: Area of union

Si la prédiction est tout à fait correcte, $\text{IoU} = 1$. Plus l'IoU est faible, plus le résultat de la prédiction est mauvais. En générale si ($\text{IoU} \geq 0,50$) est un bon résultat



$$\text{IoU}=0 \quad \text{IoU}=1/7 \quad \text{IoU}=1 \quad (4.2)$$

- **Dice similarity coefficient or Sørensen–Dice index:** Nous avons opté pour le coefficient de Dice pour évaluer les performances de la segmentation du disque et de la cupule car il est l'outil le plus utilisé dans la validation de la segmentation d'images. C'est un outil statistique qui mesure le chevauchement entre le masque réel (cible) et la sortie prédite. La valeur de **Dice** est : **2 * la zone de chevauchement divisée par le nombre total de pixels dans les deux images.**



$$\text{Dice} = \frac{2 \times \text{Area of Overlap}}{\text{Area of Rectangle 1} + \text{Area of Rectangle 2}} \quad (4.3)$$

V.4.2. Implémentation

Nous avons implémenté notre approche avec Python version 3.6 basée sur Keras [159], en utilisant la bibliothèque de logiciels open source TensorFlow [160]. La majorité des opérations de traitement d'images a été réalisées avec la bibliothèque graphique libre, OpenCV .

La plate-forme d'apprentissage et de test était sur un système Windows 7 avec un processeur I5 sans GPU et 4 Go.

Pendant l'apprentissage, nous utilisons l'estimation adaptative du moment (**ADAM**) pour optimiser le modèle profond. Le rôle de l'optimiseur ADAM est d'améliorer les *prédictions* au fil de l'apprentissage : l'erreur commise sur chaque exemple doit être de plus en plus faible (c'est-à-dire diminuer le taux d'erreur).

V.4.3. Résultats de l'étape de segmentation:

Afin de générer un bon système de segmentation sémantique du disque optique (DO) et de la cupule optique (CO) basé sur l'architecture U-Net, nous avons effectué plusieurs analyses empiriques dans le but de choisir les meilleurs paramètres s'adaptant à l'ensemble de données REFFUGE. Nous faisons plusieurs tests pour analyser l'impact de nombre de batch (voir tableau 4.2 et 4.3).

Nous avons testé la segmentation de disque optique par U-Net en utilisant deux types d'images; les images originales du fond de l'œil prises directement à partir du jeu de données REFFUGE et les images obtenues après la pré-segmentation qui sont les images de la zone (DO) coupée pour se focaliser sur la zone d'intérêt (ROI).

Pour cette segmentation automatisée, le modèle a été formé pour 400 images originales et 400 images avec DO coupés avec leurs annotations correspondantes. Une fois le modèle formé, nous avons testé un ensemble des images non étiquetées pour obtenir plus de prédictions dont nous avons besoin dans l'étape suivante (classification) figure 4.12.

Le tableau suivant présente les résultats obtenus dans la segmentation des images originales avec un nombre variable des batch:

Tableau 4.2:les performances de segmentation basées sur les images originales

Modèle utilisé	Batch	IoU	Dice
U-Net	2	0.71	0.80
U-Net	8	0.68	0.76
U-Net	16	0.63	0.73

Le tableau suivant présente les résultats obtenus dans la segmentation des images résultantes de la phase de pré-segmentation:

Tableau 4.3: les performances de segmentation basées sur les images pré-segmentées (zones DO coupées)

Modèle utilisé	Batch	IoU	Dice
U-Net	2	0.80	0.90
U-Net	8	0.77	0.84
U-Net	16	0.74	0.80

- D'après les résultats obtenus lors de plusieurs tests empiriques, nous remarquons que lorsque le nombre de batch augmente, les performances de notre modèle **diminuent**. Pour cette raison, un nombre de batch = 2 **sera le plus approprié**.
- A partir des tableaux 4.2 et 4.3, on peut dire que l'impact de la phase pré-segmentation joue un rôle important pour produire de meilleurs résultats, Dans notre cas, les images avec un DO coupé ont abouties à des résultats très satisfaisantes dans l'étape de segmentation en les comparant avec les résultats obtenus par les images originales
- Cependant, l'étape de segmentation est considérée comme un succès car toutes les valeurs d'IoU et de Dice sont supérieures à 50%.

V.4.4. Résultats de l'étape de classification:

Dans cette étape, nous avons validé le système d'aide au diagnostic de la maladie de glaucome basé sur deux systèmes, le premier est passé par deux étapes de segmentation et le second est passé directement par le segmenteur U-Net. La classification de deux modalités d'images issues de l'étape de segmentation et les images originales **est appliquée sur** :

- Images originales
- Images originales prédites (après segmentation par U-net)
- Images de la zone coupée «DO» prédites (après deux phase de segmentation)

Après l'étape d'augmentation de données, nous avons obtenu 2000 images pour chaque modalité de test de notre ensemble de données, où nous avons partitionné 70% des images pour la phase d'apprentissage et 30% pour le test.

- Après plusieurs tests empiriques, nous avons fixé le nombre d'époque à 20
- Une fois le modèle généré pendant la phase d'apprentissage, notre système sera capable de classifier n'importe quelle image **de test ou de validation** .

Les résultats obtenus sont résumés selon les trois modalités d'images comme suit :

a) images originales:

Le tableau ci-dessous **illustre** les résultats des performances du premier CNN appliqué sur les images originales ; on peut trouver ici toutes les métriques de performance nécessaires : la matrice de confusion, le diagramme de précision et la courbe de ROC

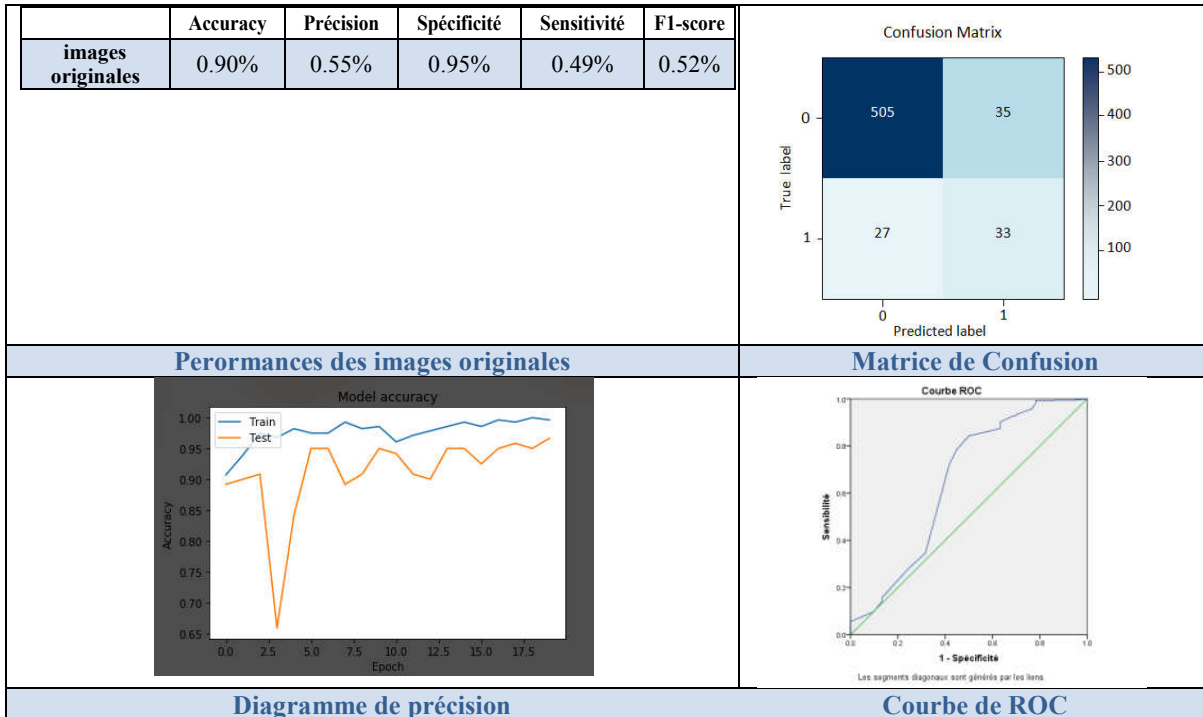


Figure 4.15: les résultats de performance de la modalité Image originale

b) Images originales prédites (après la segmentation par U-Net)

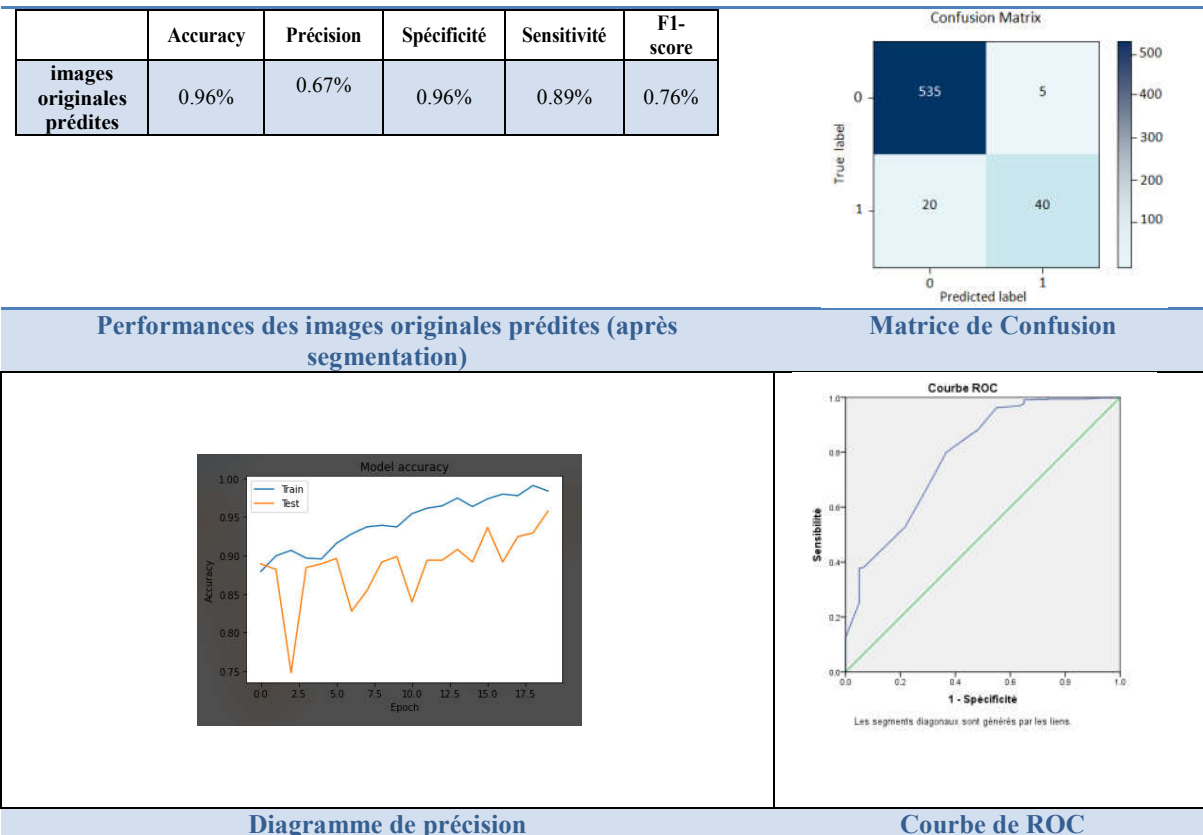


Figure 4.16: les résultats de performance des mages originaux prédites

c) Images de la zone coupée «DO» prédites (après les deux phases de segmentation)

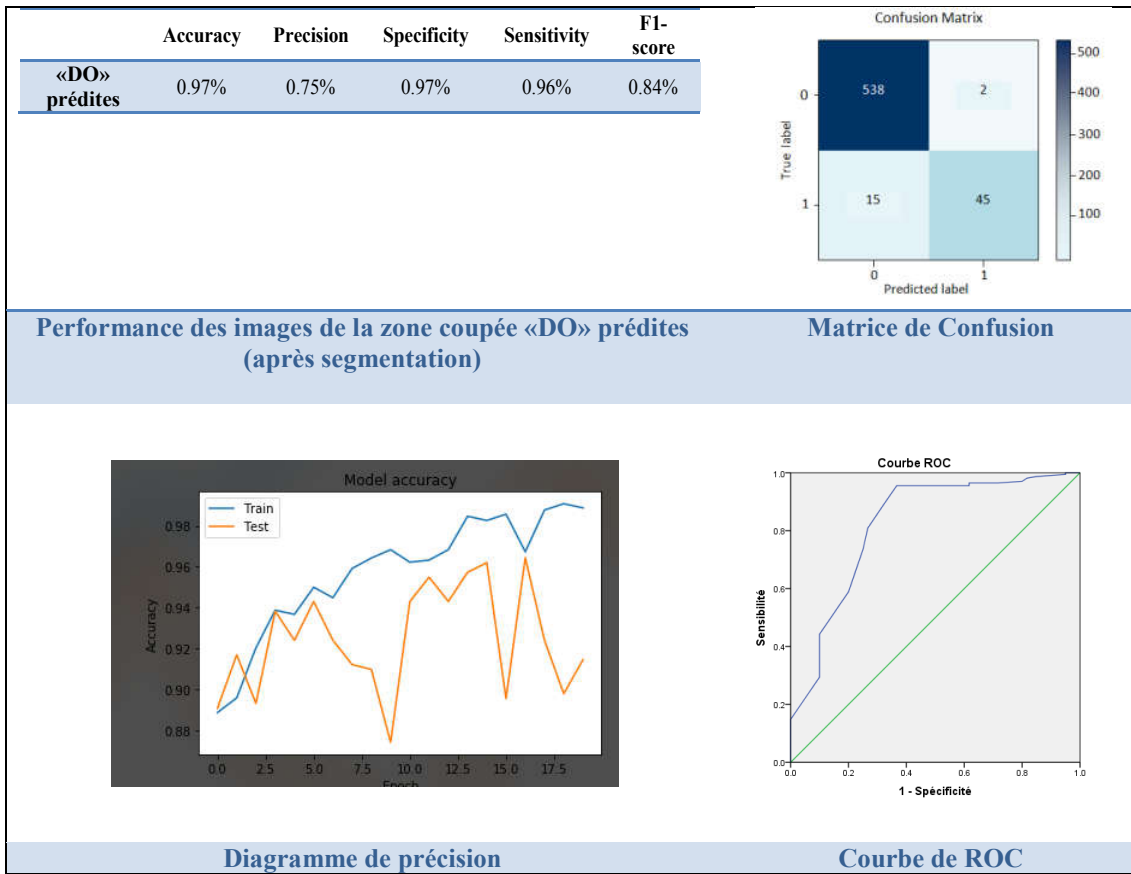


Figure 4. 17: les résultats de performance des images de la zone coupée «DO» prédites (après segmentation)

Discussion:

- Nous remarquons d’après l’étape de classification que les images prédites de la zone coupée «DO» (après segmentation) donnent des résultats très satisfaisants comparés aux images originales et aux images originales prédites avec une précision d’environ 0,97%.
- Nous observons à partir de la classification des images résultantes avant et après la pré-segmentation (figures 4.15 et 4.17) que les performances de classification augmentent plus nous nous concentrons sur la zone de disque optique (ROI) ce qui permet au CNN de classifie ces images d’une manière robuste. Cela confirme que, plus que la segmentation est efficace plus on obtient une bonne précision de classification.
- Nous avons également utilisé les caractéristiques de fonctionnement du récepteur (courbes ROC) pour évaluer notre approche proposée, d’où les trois courbes des trois modalités citées

dans (figure 4.15, 4.16 et 4.17) sont très efficaces selon le tracé des courbes où tous les points de chaque courbe sont situés sur la partie supérieure de l'espace ROC.

- Cette étude nous amène à confirmer que la segmentation de la zone d'intérêt (ROI) avant classification donne toujours de meilleurs résultats par rapport aux images originales passées directement pour classification.

VI. Conclusion

L'utilisation d'un système de diagnostic assisté par ordinateur est très efficace dans le dépistage précoce et le diagnostique du glaucome qui peut aider les cliniciens à alléger considérablement leur charge de travail. Dans cette contribution, une nouvelle approche locale a été proposée se focalisant sur la zone porteuse de la maladie au lieu d'aborder la classification de toute l'image. Pour assurer cet objectif, une segmentation incrémentale à deux phases est proposée.

Nous avons voulu se rapprocher plus de la zone d'intérêt qui est dans notre étude le disque optique par une segmentation sémantique basée sur une architecture incrémentale à deux phases : une pré-segmentation de la zone d'intérêt suivi d'une segmentation sémantique par un modèle encodeur-décodeur de type (U-Net). Une fois extraite, elle sera classifié par un modèle neuronal de type réseau convolutif.

Nous avons validé notre système sur la base des images rétiniennes de la base « REFFUGE » qui contient deux catégories d'images représentant un seul échantillon : chaque image de fond d'œil est présentée avec son annotation correspondante. Toutes ces images subissent une phase de prétraitement et une phase d'augmentation de données, afin d'avoir des meilleures performances du système.

La première phase de détection de la zone est assurée par une pré-segmentation basée sur la localisation et la coupure de la zone de disque optique de l'image du fond de la rétine. La segmentation sémantique est ensuite abordée par une architecture très élégante d'apprentissage profond qui est le U-Net pour segmenter le disque optique et cupule optique.

Le système basé sur l'apprentissage en profondeur a été conçu afin de souligner l'importance de l'étape de segmentation dans l'augmentation des performances de classification.

Cette approche a permis un diagnostic très efficace des images rétiniennes en générant des performances élevées, plus particulièrement après l'intégrations des deux modules de segmentation.

Conclusion Générale

L'objectif de cette thèse est l'interprétation des d'images médicales, plus particulièrement les images de la rétine par les algorithmes d'apprentissage profond. Le choix de ces méthodes a été motivé par leurs avantages d'extraction automatiques des caractéristiques grâce aux différentes couches de convolutions suivies de pooling intégrées.

Les réseaux de neurones convolutifs (CNNs) et l'architecture U-Net sont les principales architectures investiguées dans ce travail: cela est justifié par leurs capacités d'adaptation à la classification et à la possibilité de réaliser une segmentation sémantique des images 2D ou 3D.

En 1998, l'utilisation des architectures CNNs pour la classification supervisée était un peu bornée à cause de faiblesse de la puissance des ordinateurs et la pauvreté au niveau du nombre d'images disponibles. Dans le même temps, les méthodes d'apprentissage automatique ont connu un grand intérêt en raison des exigences minimales en puissance de matériel de traitement et en volume de données. En 2012, l'architecture AlexNet implémentée sur la base ImageNet présentait un meilleur taux d'erreur avec une grande réduction du temps de calcul grâce à la capacité du GPU ; cela a encouragé la communauté du traitement d'images à réutiliser et à optimiser ce type de modèles de réseaux pour les applications de vision par ordinateur.

Dans ce contexte, ce type de modèles basés sur l'apprentissage profond est investi dans cette thèse pour l'interprétation d'images médicales, en particulier les images rétinienne. L'objectif principal est la conception d'un système robuste d'aide à la décision de la maladie du Glaucome avec des meilleures performances.

Le diagnostic des images rétinienne reste jusqu'à ce jour en stade de recherche vu la difficulté de la tâche, car l'examen manuel de ces images pose plusieurs problèmes, qui sont liés à la subjectivité des décisions des ophtalmologistes.

Cette thèse se situe à l'intersection de l'informatique et de la médecine, dont le but est d'aider les ophtalmologistes à prendre une décision fiable, évitant ainsi les décisions subjectives. L'objectif principal de nos approches est d'améliorer les performances de la classification des images rétinienne par les méthodes DL en introduisant plusieurs stratégies telles que l'intégration de la multimodalité par fusion multimodale de plusieurs sources de données, l'augmentation de données pour assurer la diversité d'informations renforçant le processus d'apprentissage ...etc.

Afin de comprendre la structure et la technologie de traitement des images rétiniennes, nous avons mené une étude préliminaire dans le domaine médicale concernant l'anatomie de l'œil humain ainsi que les moyennes physiques utilisés par les spécialistes pour le diagnostic de ces images. D'autre part, nous avons analysé l'architecture d'un modèle de classification basé sur les réseaux à convolution afin d'analyser les techniques d'ajustement des hyper-paramètres pour concevoir une architecture optimale s'adaptant aux données d'entrée et générant des performance de décision meilleure.

La fusion multimodale est proposée dans cette thèse comme moyen essentiel permettant de générer plusieurs sources d'informations à partir d'une base d'origine ; également d'assurer la complémentarité des différents modèles de classification par la fusion au niveau des classifieurs. Ce choix a permis de renforcer un nouvel ensemble de caractéristiques provenant de diverses sources et de bénéficier des diversités de différentes modalités.

Nous avons proposé à travers la présente thèse deux contributions ;

Le but de la première contribution est de présenter une approche globale multimodale pour la classification de la maladie de glaucome en se basant sur une architecture CNN bien choisie, dans le cadre d'analyser l'impact de la réalisation d'une approche multimodale appliquée à plusieurs sources d'informations.

Pour mieux renforcer la première contribution en se focalisant plus particulièrement sur la zone d'intérêt, nous avons proposé une approche locale incrémentale à deux phases de segmentation suivie par un module de classification. L'architecture encodeur-décodeur (l'architecteur U-net) est adopté dans cette deuxième contribution afin d'analyser l'impact de la focalisation sur la zone d'intérêt ROI (disque optique) et son rendement sur la phase de classification.

- Dans un premier lieu, nous avons proposé une approche globale multimodale nommée β plus-CNN basée CNN pour la classification des images de fond d'œil en deux classes: "Glaucome" et "Non-Glaucome" à partir de plusieurs modalités de l'image en entrée (couleurs, binaires, niveaux de gris et les images passées par le filtrage de Gabor). Afin d'analyser l'impact d'utiliser la représentation du texture (le filtre de Gabor) comme une source complémentaire de notre système d'aide au diagnostic médical et montrer le rôle de fusion des classifieurs issus de plusieurs sources d'informations.
- À l'issue de l'analyse des résultats obtenus dans le cadre de cette étude, nous avons remarqué que le modèle couleur et le modèle de Gabor ont pu avoir lesdes meilleurs résultats de classification.

- D'une manière générale, le système β plus-CNN a réussi à apprendre et à classifier les images rétiniennes avec des performances très élevées, principalement parce que la fusion de quatre classifieurs a obtenu des résultats de prédiction de modèle toujours supérieure à celles obtenus par les modèles pris individuellement.
- Ces expériences confirment la supériorité de la qualité de notre modèle proposé par rapport aux autres méthodes de pointe dans le même domaine.
- Nous avons proposé également une approche locale du système de diagnostic assisté par ordinateur nommée DSC-U-Net basée sur une architecture incrémentale à deux phases de segmentation bénéficiant de la particularité de brillance du disque optique de l'image rétinienne et également le modèle U-Net . Notre approche a été validée sur l'ensemble de données de la base «REFFUGE».

Perspectives

Comme le travail proposé dans cette thèse aborde l'aide au diagnostic médical par les techniques avancées d'apprentissage profond en touchant les deux phases de classification et de détection de zones, des perspectives futures peuvent être envisageables afin de mieux suivre :

- Intégrer les nouvelles techniques intelligentes d'augmentation de données basées deep learning telles que le GAN(Generative Adversarial Networks) pour une meilleure augmentation des données d'apprentissage
- Généraliser nos contributions sur d'autres domaines médicaux et d'autres types d'images médicales en déployant les adaptations nécessaires
- Analyser d'autres architectures de type Deep Learning comme le Res-Net et le VGG16 dans le contexte multimodal.
- Investir d'autres techniques de segmentation dans le contexte multimodal.

Productions scientifiques

1. Publication dans des revues internationales

- Touahri, R., Azizi, N., Hammami, N. , Al Dwairi, M.,Benzebouchi.N.,Moumen,O., “Multi source Retinal Fundus Image Classification Using Convolution Neural Networks Fusion and Gabor-Based Texture Representation”, Int. J. of Computational Vision and Robotics, Vol 11, No 4, 2020. Pp. 401-428.
- R.Touahri, N.Azizi, N.E.Hammami, F.Benaida, N.Zemmal: “A two-stage optic disc segmentation based on U-Net architecture for glaucoma diagnosis”, (submitted)

2. Publication dans des conférences internationales

- Touahri.R, Azizi.N , Hammami .N.E , Benzebouchi .N.E, Moumene.O: “A Comparative Study of Convolutional Neural Network and Twin SVM for Automatic Glaucoma Diagnosis”, 2018 International Conference on Signal, Image, Vision and their Applications (SIVA), IEEE, Guelma, 2018.
- Touahri .R, Azizi.N , Hammami .N.E: “ Convolution Neural Network Features Based SVM For Automated Breast Cancer Classification ”, First international workshop of deep learning 1st IW DLMA, Bejia, 2018.
- Touahri .R, Azizi.N , Hammami .N.E , Aldwairi.M , Benaida. F: “Automated Breast Tumor Diagnosis Using Local Binary Patterns (LBP) Based on Deep Learning Classification”, 2019 International Conference on Computer and Information Sciences (ICCIS), el jouf , Saudia Arabia KSA, 2019.

Bibliographie

- [1] M.H. Alsun, « indexation guidée par les connaissances en imagerie médicale », l'Université européenne de Bretagne, 2012.
- [2] R.A. Miller, «Computer-assisted diagnostic decision support: history, challenges, and possible paths forward», *Advances in Health Sciences Education*, vol.14, p.106, 2009.
- [3] J.C. Sournia, « Histoire du diagnostic en médecine », Santé, 1995.
- [4] K. Kawamoto, C. A. Houlihan, E. A. Balas, et D. F. Lobach, « Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success », *BMJ*, vol. 330, no. 7494, p. 765, 2005.
- [5] G.Kong, D.L. Xu, & J.B. Yang, "Clinical decision support systems: a review on knowledge representation and inference under uncertainties", *International Journal of Computational Intelligence Systems*, 1(2), 159-167.2008.
- [6] L. A. Stojkowska & S.Loskovska, "Clinical decision support systems: medical knowledge acquisition and representation methods". *The International Conference on Electro/Information Technology*, Normal, IL, USA, 1-6, 2010.
- [7] L.Burhenne, S.Wood, C.S.A.Feig, D.B.Kopans, K. F .O'Shaughnessy, E. A.Sickles, , L. Tabar, C. J.Vyborny, & R. A. Castellino, "Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology*", 215 (2), 554-562. 2000.
- [8] N.Zemmal, thèse doctorat « Techniques d'apprentissage pour la sélection d'attributs : Application à la reconnaissance des formes », université Annaba, Algérie, 2018
- [9] L. Benmimoune, thèse doctorat « Une approche pour la conception de systèmes d'aide à la décision médicale basés sur un raisonnement mixte à base de connaissance », Université de Technologie de Belfort-Montbéliard, 2019
- [10] N.Mekhalidi, « Analyse des images médicales par des techniques hybrides: application a la détection des kystes rénaux », thèse de magister, Université d' ORAN, 2014.
- [11] C.M BENTAOUZA, Thèse de magister , « classification des images médicales par les séparateurs à vaste marge », Université des Sciences et de la Technologie d'Oran USTOMB, 2007 .
- [12] S. W .Shin, & K. J. Kim, "The Role of Computer Aided Diagnosis (CAD) in Medical Imaging", *Central Asian Journal of Medical Science*, vol.1, 1-15, (2015).
- [13] A.MEZIANE «: Systèmes d'information en Imagerie médicale : Moyen de diagnostic, de thérapie et d'économie », *RIST*, vol. 11, N° 02 , 2001.
- [14] A. CAPRI « Caractérisation des objets dans une image en vue d'une aide à l'interprétation et d'une compression adaptée au contenu : application aux images échographique », Université d'Orléans. Faculté des Sciences et Technologies Industrielles. Thèse de doctorat, 2007
- [15] https://www.doctissimo.fr/html/sante/imagerie/imagerie_sommaire.htm
- [16] <https://www.biron.com/fr/imagerie-medicale/mammographie/radiographie-des-seins/>
- [17] P. Maurel, 'cours Imagerie par magnéto-et électro-encéphalographie (MEG/EEG)', France 2016
- [18] <http://cleophtalmo.com/explorations/rnm/> dernière vue 19/02/2021
- [19] <https://www.provisu.ch/fr/dossiers/oeil-et-vision.html>
- [20] <https://www.frm.org/recherches-autres-maladies/maladies-de-la-retine/focus-maladies-retine#types-maladies-retine>
- [21] G. K.Matsopoulos, P. A. Asvestas, K. K.Delibasis, N. A. Mouravliansky, T. G.Zeyen, "Detection of glaucomatous change based on vessel shape analysis." *Computerized Medical Imaging and Graphics* 32, no. 3 183-192, 2008.
- [22] K.P.Noronha et al.: "Automated classification of glaucoma stages using higher order cumulant features" *J. Biomed. Signal Process.* vol 10, pp.174–183, 2014.
- [23] N.M.Tan et al.: "Robust multi-scale superpixel classification for optic cup localization", *Comput.Med. Imaging Graphics*, vol.40, pp.182–193, 2015.
- [24] A.Sing et al.: "Image processing based automatic diagnosis of glaucoma using wavelet features of segmented optic disc from fundus image", *Comput. Methods Progr. Biomed*, vol. 124, pp.108–120, 2015.

- [25] X.Chen, Y.Xu, D. Wing Kee Wong and J.Liu , “Glaucoma Detection based on Deep Convolutional Neural Network”, Institute for Infocomm Research, Agency for Science Technology and Research, IEEE conference, Singapore , pp.1-4, 2015.
- [26] S.Maheshwari, R. B. Pachori, V.Kanhangad, S. V. Bhandary, U. R. Acharya: “Iterative variational mode decomposition based automated detection of glaucoma using fundus images”, Computers in Biology and medicine, vol.88, pp. 142-149, 2017.
- [27] J.Zilly, J.M.Buhmann, D.Mahapatra : “ Glaucoma Detection Using Entropy Sampling And Ensemble Learning For Automatic Optic Cup And Disc Segmentation”,Computerized Medical Imaging and Graphics, ,pp.1-32, 2016.
- [28] N.E.Benzebouchi, N.Azizi, S.E.Bouziane, “Glaucoma diagnosis using cooperative convolutional neural network”, International Journal of Advances in Electronics and Computer Science, Volume-5, Issue-1, Jan.-2018.
- [29] B.Al-Bander, W. Al-Nuaimy, M.A.Al-Tae, Y.Zheng,: “Automated Glaucoma Diagnosis using Deep Learning Approach”, 14th International Multi-Conference on Systems, Signals & Devices (SSD), pp.1-4, 2017.
- [30] U.R. Acharya, S.Bat, J.E.Koh, S.V.Bhandary, H. Adeli,: “A novel algorithm to detect glaucoma risk using texton and local configuration pattern features extracted from fundus images”,Computers in Biology and Medicine,vol. 88 ,pp. 72-83, 2017.
- [31] U.Raghavendra, H.Fujita,S.V. Bhandary, A.Gudigar, J.HongTan,U. R. Acharya, “ Deep Convolution Neural Network for Accurate Diagnosis of Glaucoma Using Digital Fundus Images”,Information Sciences, vol:441, pp. 41-49, 2018.
- [32] Y. Chai, H.Liua ,J. Xub : “Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models”, Knowledge-Based Systems, Vol.161, pp. 147-156, 2018.
- [33] R. Touahri, N. Azizi , .N.E. Hammami, N.E. Benzebouchi, O. Moumene: “A Comparative Study of Convolutional Neural Network and Twin SVM for Automatic Glaucoma Diagnosis”, International Conference on Signal, Image, Vision and their Applications (SIVA), pp.1-5,2018.
- [34] R. Touahri,N. Azizi , N.E. Hammami, M.Al Dwairi, N.E.Benzebouchi, O.Moumen, “ Multi source Retinal Fundus Image Classification Using Convolution Neural Networks Fusion and Gabor-Based Texture Representation”, Int. J. of Computational Vision and Robotics, 2020.
- [35] C. Shorten and T. M. Khoshgoftaar: “A survey on Image Data Augmentation for Deep Learning”, Journal of Big Data ,2019.
- [36] <https://towardsdatascience.com/exploring-image-data-augmentation-with-keras-and-tensorflow> a8162d89b844 Last access: 23/06/2019.
- [37] JM-B. Francisco, S. Fiammetta, MJ. Jose, U. Daniel, F .Leonardo: “Forward noise adjustment scheme for data augmentation”, IEEE Symposium Series on Computational Intelligence (SSCI), 2019.
- [38] J.McCarthy, M.Minsky, N.Rochester, C. E.Shannon : “A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence”,1955.
- [39] G.Idiou, thèse de magistère, :« régression et modélisation par les réseaux de neurones », université Mentouri – constantine, 2009.
- [40] G. Rejala, A. Ravi, S.Churiwala : “An Introduction to Machine Learning”, Springer Nature Switzerland AG, 2019.
- [41] N.Guerroumi : thèse « Segmentation automatique par apprentissage profond de la colonne vertébrale scoliotique à partir d’images de résonance magnétique », école de technologie supérieure université du Québec, Montréal, 2019.
- [42] A. L. Samuel: “Some studies in machine learning using the game of checkers.IBM Journal of research and development”, 44(1.2): 206–226, 2000.
- [43] R. S. Michalski, J. G.Carbonell and T. M. Mitchell; “An artificial intelligence approach”, Springer Science & Business Media, 2013.
- [4 4] J. Brownlee, “Deep Learning for Computer Vision”, 2019.
- [45] Y. Lecun, Y. Bengio,G. Hinton: “Deep Learning”, Nature , vol. 521 , 2015
- [46] B.Jia, W.Feng, M.Zhu,: “Obstacle detection in single images with deep neural networks”. Signal Image Video Process, Vol.10, pp.1–8 , 2015.

- [47] A. Alotaibi, A.Mahmood, “Deep face liveness detection based on nonlinear diffusion using convolution neural network”, Springer-Verlag London, pp.1-8, 2016.
- [48] R. Bock, J. Meier, L. G. Nyúl, J. Hornegger, G. Michelson: “Glaucoma risk index:Automated glaucoma detection from color fundus images”, *Medical Image Analysis*, vol. 14 ,pp. 471–481, 2010
- [49] H.Pratta, F.Coenenb, D. M Broadbentc, S.P Hardinga and Y.Zhenga , “Convolutional Neural Networks for Diabetic Retinopathy, ”, Elsevier PCS (MIUA),United Kingdom,vol.90,pp.200-205, 2016.
- [50] D.Doshi, A.Shenoy,D. Sidhpura and P. Gharpure, “ Diabetic Retinopathy Detection using Deep Convolutional Neural Networks,”*IEEE CAST.India*,pp.261-266,2016.
- [51] K.Xu , D.Feng, and H.Mi, “Deep Convolutional Neural Network-Based Early Automated Detection of Diabetic Retinopathy Using Fundus Image ”, *Molecules MDPI .J. China*,vol.22 ,pp.1 7,2017
- [52] R. Touahri, N. Azizi, .N.E.Hammami , M .Aldwairi , F.Benaida: “Automated Breast Tumor Diagnosis Using Local Binary Patterns (LBP) Based on Deep Learning Classification”, 2019 International Conference on Computer and Information Sciences (ICIS),pp.1-5, 2019.
- [53] Y.LeCun, L.Bottou,Y.Bengio,P.Haffner: “Gradient-based learning applied to document recognition”, *Proc. IEEE* , vol.86, pp.2278– 2324,1998.
- [54] Y.Le Cun, , B.Boser, J.S.Denker, D.Henderson, R.E .Howard, W.Hubbard , L.D. Jackel, “Handwritten digit recognition with a back propagation network. In: *Advances in Neural Information Processing Systems*”, Morgan Kaufmann Publishers Inc. San Francisco, pp. 396–404, 1990.
- [55] D.C Ciresan, U.Meier, L. M. Gambardella et J. Schmidhuber : “*Convolutional neural network committees for handwritten character classification*”. In *International Conference on Document Analysis and Recognition*, pp.1135– 1139. IEEE, 2011.
- [56] A. Karpathy, G.Toderici, S. Shetty, T. Leung, R.Sukthankar et L. Fei-Fei: “*Large-scale video classification with convolutional neural networks*”. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 1725–1732, 2014.
- [57] N. Tajbakhsh, J. Y.Shin, S. R. Gurudu, R. Todd Hurst, C. B Kendall, M.B Gotway et J. Liang: “*Convolutional neural networks for medical image analysis :Full training or fine tuning* “, *IEEE transactions on medical imaging*,vol. 35, no. 5, pp. 1299–1312, 2016.
- [58] W.Hu, Y. Huang, L.Weï, F.Zhang et H. Li : “*Deep convolutional neural networks for hyperspectral image classification*”. *Journal of Sensors*, 2015.
- [59] F. A. Spanhol, L. S Oliveira, C. Petitjean et L. Heutte: “*Breast cancer histopathological image classification using convolutional neural networks*”. *international joint conference on neural networks (IJCNN)*, pp. 2560–2567. IEEE, 2016.
- [60] S.H Lee, C. Seng Chan, P. Wilkin et P. Remagnino: “*Deep-plant : Plant identification with convolutional neural networks*”. In 2015 IEEE international conference on image processing (ICIP), pages 452–456. IEEE, 2015.
- [61] P. Lakhani et B.Sundaram: “*Deep learning at chest radiography : automated classification of pulmonary tuberculosis by using convolutional neural networks*”. *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.
- [62] F.Suard : thèse de doctorat : « *Methodes a noyaux pour la detection de pietons* », LITIS - INSA de Rouen, 2006.
- [63] P. Viola, and M. Jones, “Robust real-time face detection”. *International Journal of Computer Vision*, (2004).
- [64] A.Rad, , R.Safabakhsh, , N.Qaragozlou, , and M.Zaheri: “Fast iris and pupil localization and eyelid removal using gradient vector pairs and certainty factors”. In *The Irish Machine Vision and Image Processing Conf*, pp. 82–91, 2004.
- [65] P.Negri, X.Clady, M.Milgram and U.Lisif-parc : « Perceptionvisuelle du geste de prehension : application a la robotique manipulatrice ». *18ème Journée des JJCR, Douai, France*, 2004.
- [66] V.Dumoulin, and F.Visin, :“A guide to convolution arithmetic for deep, learning”. preprint arXiv:1603.07285. 2016.
- [67] J.Wu ;: “Introduction to Convolutional Neural Networks”, LAMDA Group National Key Lab for Novel Software Technology, China,vol.1, pp. 1-31. 2017
- [68] J. Heaton : “Deep Learning and Neural Networks”, vol:3, 2015
- [69] I. Goodfellow, Y. Bengio, A. Courville: “deep learning”, 2016
- [70] M.I. Randrianarivony, « Détection de concepts et annotation automatique d’images médicales par apprentissage profond », université d’ANTANANARIVO, 2018.

- [71] R. O. Duda, P. E. Hart and D. G. Stork: "Pattern classification". Wiley-interscience, 2000.
- [72] C. Desir, « Classification Automatique d'Images, Application à l'Imagerie du Poumon Profond », université de Rouen, 2013.
- [73] M.M. Jadoon, Q. Zhang, I. U. Haq, S.I Butt, A. Jadoon, "Three-Class Mammogram Classification Based on Descriptive CNN Features", BioMed Research International, pp.11, 2017.
- [74] M. Rodriguez-Damian, E. Cernadas, A. Formella, and P. Otero: "Pollen classification using brightness-based and shape-based descriptors". In *ICPR (2)*, pp. 212–215, 2004.
- [75] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Backpropagation applied to handwritten zip code recognition", *Neural computation*, vol.4, pp.541–551, 1989.
- [76] J. Alkhateeb, thèse de doctorat, « Word Based Off-line Handwritten Arabic Classification and Recognition », School of Computing, Informatics and Media, University of Bradford, 2010.
- [77] A. Krizhevsky, I. Sutskever et G. E. Hinton : "Imagenet classification with deep convolutional neural networks", In Advances in neural information processing systems, pp. 1097–1105, 2012.
- [78] M. D. Zeiler et R. Fergus : "Visualizing and understanding convolutional networks", In European conference on computer vision Springer, pp. 818–833., 2014.
- [79] K. Simonyan et A. Zisserman : "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv, ICLR, vol.6, pp.1409.1556, 2015.
- [80] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke et A. Rabinovich. : "Going deeper with convolutions", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9, 2015.
- [81] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens et Z. Wojna.: "Rethinking the inception architecture for computer vision", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2818–2826, 2016.
- [82] K. He, X. Zhang, S. Ren et J. Sun, "Deep residual learning for image recognition", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.
- [83] G. Huang, Z. Liu, L. Van Der Maaten et K. Q. Weinberger. "Densely connected convolutional networks", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708, 2017.
- [84] A. Rajkomar, S. Lingam, A. G Taylor, M. Blum et J. Mongan.,; "High-throughput classification of radiographs using deep convolutional neural networks", Journal of digital imaging, vol. 30, no. 1, pp. 95–101, 2017.
- [85] R. Li, W. Zhang, H.-Il Suk, L. Wang, J. Li, D. Shen et S. Ji: "Deep learning based imaging data completion for improved brain disease diagnosis". In International Conference on Medical Image Computing and Computer-Assisted Intervention Springer, pp. 305–312, 2014.
- [86] J. Xu, X. Luo, G. Wang, H. Gilmore et A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images", Neurocomputing, vol. 191, pp. 214–223, 2016.
- [87] N. benzebouchi, n. azizi, k. ayad: "A Computer-Aided Diagnosis System for Breast Cancer Using Deep Convolutional Neural Networks", Computational Intelligence in Data Mining, 2018.
- [88] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya et al: "Chexnet : Radiologist-level pneumonia detection on chest x-rays with deep learning, 2017.
- [89] R. Girshick, R. Girshick, J. Donahue, T. Darrell et J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation". In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587, 2014
- [90] R. Girshick, "Fast r-cnn. In Proceedings of the IEEE international conference on computer vision", pp.1440–1448, 2015.
- [91] J. Redmon, S. Divvala, R. Girshick et A. Farhadi. "You only look once : Unified, real-time object detection", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788, 2016.
- [92] D. Ribli, A. Horváth, Z. Unger, P. Pollner et I. Csabai, "Detecting and classifying lesions in mammograms with deep learning", Scientific reports, vol. 8, no. 1, pp. 1–7, 2018.

- [93] W. Fan, H. Jiang, L. Ma, J. Gao et H. Yang, "A modified faster R-CNN method to improve the performance of the pulmonary nodule detection". In Tenth International Conference on Digital Image Processing (ICDIP 2018), vol. 10806, pp. 1469–1476, 2018
- [94] D. Yang, S. Zhang, Z. Yan, C.Tan, K. Li et D. Metaxas, "Automated anatomical landmark detection on distal femur surface using convolutional neural network", In 2015 IEEE 12th international symposium on biomedical imaging (ISBI), pp. 17–21, 2015
- [95] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V. CT Mok, L. Shi et P.Heng, "Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks", IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1182–1195, 2016.
- [96] K. Sirinukunwattana, S. E Ahmed Raza, Y-W Tsang, D. RJ Snead, I. A Cree et N. M Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images", IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1196–1206, 2016
- [97] M. Everingham, S. M. A. Eslami, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes Challenge : A Retrospective". *International Journal of Computer Vision*, vol.111(1), pp.98–136, 2014.
- [98] F. Rottensteiner, G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez, and U. Breitkopf. "The ISPRS benchmark on urban object classification and 3d building reconstruction". *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol:1 :3, 2012.
- [99] V. Ulman, M. Maska, K. E. G. Magnusson, O. Ronneberger, C. Haubold, N. Harder, P. Matula, P. Matula, D. Svoboda, M. Radojevic, I. Smal, K. Rohr, J. Jaldén, H. M. Blau, O. Dzyubachyk, B. Lelieveldt, P. Xiao, Y. Li, S.-Y. Cho, A. C. Dufour, J.-C. Olivo-Marin, C. C. Reyes-Aldasoro, J. A. Solis-Lemus, R. Bensch, T. Brox, J. Stegmaier, R. Mikut, S. Wolf, F. A. Hamprecht, T. Esteves, P. Quelhas, . Demirel, L. Malmström, F. Jug, P. Tomancak, E. Meijering, A. Muñoz-Barrutia, M. Kozubek, and C. Ortiz-de Solorzano. "An objective comparison of cell-tracking algorithms", *Nature Methods*, **vol.14**, pp. 1141–1152 advance online publication, 2017.
- [100] J. Long, E. Shelhamer et T. Darrell. « *Fully convolutional networks for semantic segmentation* ». In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440, 2015.
- [101] O. Ronneberger, P. Fischer, T. Brox: "U-Net: Convolutional Networks for Biomedical Image Segmentation ", *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, vol.9351, pp.234-241, 2015
- [102] J. Ni, J. Wu, J. Tong, Z.Chen, J. Zhao, ' GC-Net: Global context network for medical image segmentation', *Computer Methods and Programs in Biomedicine*, pp.1-10,2019.
- [103] M.Dash, N.D. Londhe, S. Ghosh, A.Semwal, R. S. Sonawane, ' PsLSNet: Automated psoriasis skin lesion segmentation using modified U-Net-based fully convolutional network', *Biomedical Signal Processing and Control*, vol.52, pp. 226–237, 2019.
- [104] S. Alqazzaz, X. Sun, X. Yang, L. Noke: "Automated brain tumor segmentation on multi-modal MR image using SegNet", *Computational Visual Media*, Vol. 5, pp. 209–219, 2019.
- [105] P. Kumar, P.Nagar, C. Arora, A. Gupta: "U-segnet: fully convolutional neural network based automated brain tissue segmentation tool", pp.1-5, 2018.
- [106] M.A. Al-masni, M.A. Al-antari, M.Choi, S.Han, T.Kim: "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks", *Computer Methods and Programs in Biomedicine*, vol.162, pp.221–231, 2018.
- [107] J.Bullock, C. Cuesta-L'azaro, A. Quera-Bofarull, "XNet: A convolutional neural network (CNN) implementation for medical X-Ray image segmentation suitable for small datasets", *SPIE Medical Imaging*, vol. 10953, pp.1-11, 2018.
- [108] V. Badrinarayanan, A. Kendall, and R. Cipolla. SegNet : "A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.39(12), pp.2481–2495, 2017
- [109] S. Pereira, A.Pinto, V.Alves et C. A Silva: "Brain tumor segmentation using convolutional neural networks in MRI images", IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1240–1251, 2016
- [110] B. Peng, L. Chen, M. Shang et J. Xu.: "Fully Convolutional Neural Networks for Tissue Histopathology Image Classification and Segmentation", IEEE International Conference on Image Processing (ICIP), pp. 1403–1407,2018.

- [111] Z.Zhou, N. Tajbakhsh, M. R. Siddiquee, J. Liang: "UNet++: A Nested U-Net Architecture for Medical Image Segmentation", Springer Nature Switzerland, pp.1-10 , 2018.
- [112] D. S. Mercadier, B. Besbinar et P.Frossard, "*Automatic Segmentation of Nuclei in Histopathology Images Using Encoding-decoding Convolutional Neural Networks.*" In ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1020–1024. 2019.
- [113] The National Institute of Health, "glaucoma", <https://nei.nih.gov/health/glaucoma>, accessed January 28, 2019.
- [114] H. A. Quigley, and A. T. Broman,: "The number of people with glaucoma worldwide in 2010 and 2020", British journal of ophthalmology, vol. 90(3),pp. 262-267,2006.
- [115] K.Boyd, "What is glaucoma", Available:<https://www.aao.org/eye-health/diseases/what-is-glaucoma>, accessed January 30, 2019.
- [116] A.Tugashetti, T.C.Manjunath, G.Pavithra: "Design of algorithms for diagnosis of primary glaucoma through estimation of CDR in different types of Fundus Images using IP techniques", International Journal of Innovative Research in Information Security (IJIRIS), Vol. 04 , 2017
- [117] J.Y .Choi, T. K. Yoo, J.G. Seo, J. Kwak, T. T. Um and T. H. Rim ;: "Multi-categorical deep learning neural network to classify retinal images: A pilot study employing small database, ",Plos one,vol.12, pp.1-6, 2017.
- [118] J. J. GÓMEZ-VALVERDE., A .ANTÓN, G .FATTI, B. LIEFERS, A .HERRANZ, A.SANTOS, C.I., SÁNCHEZ, M. J LEDESMA-CARBAYO: "Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning", BIOMEDICAL OPTICS EXPRESS 892, Vol. 10, pp.1-22 , 2019.
- [119] C. Mahiba, A .Jayachandran.: "Severity analysis of diabetic retinopathy in retinal images using hybrid structure descriptor and modified CNNs", Measurement, vol.135, pp.762-767,2019.
- [120] S. Molaei , M.E. Shiri, A. Abadi.: "Maintaining filter structure: A Gabor-based Convolutional Neural Network for image Analysis", Applied Soft Computing Journal, pp.1-35, 2019.
- [121] C. Liu, W. Ding, X.Wang, B. Zhang: "Hybrid Gabor Convolutional Networks", Pattern Recognition Letters, pp.11, 2018.
- [122] B.Saçak, M. N.Yildiz, M. Fazal, M. Ahn, R.Feiren, , & Ozdemir, S. (Eds.): "Media Literacy in a Digital Age: Multimodal Social Semiotics and Reading Media. In Handbook of Research on Media Literacy Research and Applications Across Disciplines", IGI Global, pp. 13-26, 2019.
- [123] D.Lahat, T.Adali & C. Jutten: "Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects", Proceedings of the IEEE, vol.103(9), pp.1449–1477, 2015.
- [124] J.Kludas, E.Bruno, S.Marchand-Maillet "Information Fusion in Multimedia Information Retrieval. In: Adaptive Multimedia Retrieval: Retrieval, User, and Semantics., vol. 4918, pp.147-159,2008
- [125] M.H.YANG, J.H.TAO,: "Data fusion methods in multimodal human computer dialog. Virtual Reality & Intelligent Hardware", vol.1(1), pp.21–38, 2019
- [126] L.Xu , A.Krzyzak, , & C. Y. Suen,: "Methods of combining multiple classifiers and their applications to handwriting recognition". IEEE Transactions on Systems, Man, and Cybernetics, 22(3), p.418–435, 1992
- [127] N.E benzebouchi, thèse de doctorat : "Paradigmes Avancés de l'Apprentissage Automatique pour l'Analyse et la Classification des Données Multimodales", Université Annaba, 2020 ;
- [128] O.S.Al-Kadi,: "a gabor filter texture analysis approach for histopathological brain tumour subtype discrimination",ISESCO-JST-journal,vol.12(22),pp.1-8.2017,
- [129] S.Prabhu,, G.Hegde, T. Mungle,: "Retinal pattern matching using the method of gabor filter", Proceedings of RETCOMP, , Vol 1,pp.1-8,2013.
- [130] P.Bandzi, M.Oravec, J.Pavlovičová,: " New statistics for texture classification based on gabor filters", radioengineering, vol. 16, PP. 1-5,2007
- [131] N.Otsu: "A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems", Man and Cybernetics, p.62-66, 1979
- [132] <https://github.com/nismakoto/Gabor-filter>
- [133] F .Fumero , S .Alayon , J.L. Sanchez, J. Sigut, M. Gonzalez-Hernandez.: " RIM-ONE: An Open Retinal Image Database for Optic Nerve Evaluation", Proceedings of the IEEE Symposium on Computer-Based Medical Systems,pp.1-7, 2011.
- [134]: <https://en.wikipedia.org/wiki/DeepLearning4j>

- [135] N.Dey, A.B. Roy, M.Pal, and A .Das. : “ FCM based bloodvessel segmentation method for retinal images”, arXiv:1209.1181, 2012.
- [136] X.Zhang, X. Li, Y. Feng, Z.Liu: “The use of ROC and AUC in the validation of objective fusion evaluation metrics”, *Signal Processing*, , vol. 115, pp. 38-48, 2015.
- [137] J .Sherin, D .Devi KalaRathinam, A .Santhiya Grace : “ analysis on glaucoma detection”, *International Journal of Trend in Scientific Research and Development (IJTSRD)*,vol.2,pp.1-3,2018
- [138] U.R.Acharya, H.Fujita, S.L.Oh,U.Raghavendra, J.HongTan, M.Adam, A.Gertych, Y.Hagiwara: “Automated identification of shockable and non-shockable lifethreatening ventricular arrhythmias using convolutional neural network”, *Future Generation Computer Systems*, .pp.22, 2017
- [139] <https://github.com/clumsy/gaborfilter/commit/91d14225b6ad657fb253aa9d039a2c334e1bad>
- [140] N. E.Benzebouchi, N.Azizi, A. S. Ashour, N.Dey, R. S.Sherratt: “Multi-modal classifier fusion with feature cooperation for glaucoma diagnosis”. *Journal of Experimental & Theoretical Artificial Intelligence*, 31(6),841-874,2019.
- [141] R.Priyanka, S.J.G.Shoba, A.BTherese,: “Segmentation of Optic Disc in Fundus Images using Convolutional Neural Networks for Detection of Glaucoma”, *International Journal of Advanced Engineering Research and Science (IJAERS)* ,Vol.4, pp.1-10,2017.
- [142] N. Thakur, M. Juneja : “ Survey on segmentation and classification approaches of optic cup and optic disc for diagnosis of glaucoma”, *Biomedical Signal Processing and Control* ,vol.42, pp.162–189, 2018
- [143] S.Yu, D. Xiao, S. Frost, Y. Kanagasingam, “Robust optic disc and cup segmentation with deep learning for glaucoma detection”, *Computerized Medical Imaging and Graphics* , vol.74 , pp.61–71, 2019.
- [144] A. Sevastopolsky,: “Optic Disc and Cup Segmentation Methods for Glaucoma Detection with Modification of U-Net Convolutional Neural Network”, *Pattern Recognition and Image Analysis* ,Vol.27,pp.618-624,2017
- [145] H.Fu, J.Cheng, Y.Xu, D. W. K.Wong, J.Liu, X.Cao , “Joint Optic Disc and Cup Segmentation Based on Multi-label Deep Network and Polar Transformation”, *IEEE Transactions on Medical Imaging* ,Vol. 37 , pp.1-9 ,2018.
- [146] J. I. Orlando , Huazhu Fu , J. B. Breda, K. van Keer , D. R. Bathula , A. Diaz-Pinto , R. Fang , P-Ann Heng , J. Kim , J. Lee, J. Lee , X. Li , P. Liu , S. Lu , B. Murugesan , V. Naranjo , S.S. R. Phaye , S.M. Shankaranarayana , A. Sikka , J. Son , A. van den Hengel , S. Wang , J. Wu , Z. Wu , G. Xu , Y. Xu , P. Yin , F. Li , X. Zhang , Y. Xu , H. Bogunovic: “REFUGE Challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs”,*Medical Image Analysis* , vol.59, pp. 1-21,2020 .
- [147] G. Mittal , “Landmark Detection in Retinal Images”, *International Institute of Information Technology Hyderabad, INDIA*, 2018
- [148] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan and I. J. Constable, “Retinal image analysis: Concepts, applications and potential.”, *Progress in Retinal and Eye Research*, vol. 25, pp. 99- 127, 2006
- [149] A.Giachetti, L.Ballerini, E.Truccho, and P.J. Wilson, “The Use of Radial Symmetry to Localize Retinal Landmarks”. *Computerized Medical Imaging and Graphics*, vol. 37(5), pp. 369-376, 2013
- [150] A.Tewari, D.Gupta, and J. Sivaswamy, : “Bilateral Symmetry based Approach for Joint Detection of Landmarks in Retinal Images”, *International Conference on Signal Processing and Communications (SPCOM)*, pp. 1-6, 2014.
- [151] A.Aquino,: “Establishing the Macular Grading Grid by Means of Fovea Centre Detection using Anatomical-based and Visual-based Features”.: *Computers in Biology and Medicine*, vol.55, pp.61-73, 2014.
- [152] P.Burlina, et al. “Detection of age-related macular degeneration via deep learning”, *Biomedical Imaging (ISBI), IEEE 13th International Symposium on. IEEE*, 2016.
- [153] K.K. Maninis, J. Pont-Tuset, P. Arbeláez, et L. Van Gool,: “Deep retinal image understanding”, dans *Medical Image Computing and Computer-Assisted Intervention –MICCAI 6*, pp. 140–148, 2016
- [154] H. Fu, Y. Xu, S. Lin, D. W. Kee Wong, et J. Liu :“Deepvessel : Retinal vessel segmentation via deep learning and conditional random field”, *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pp. 132–139,2016.

- [155] A. Lahiri, K. Ayush, P. K. Biswas, et P. Mitra, “Generative adversarial learning for reducing manual annotation in semantic segmentation on large scale microscopy images”, *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 794–800, 2017
- [156] T. Birgui-Sekou, M.Hidane, O. Julien, H. Cardot.: “Réseaux de neurones à convolution pour la segmentation de vaisseaux sanguins rétiniens – Des patchs aux images de taille réelle » . *Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP)*, Marne-la-Vallée, France, 2018.
- [157] G.P.González : “Détection et segmentation des disques optiques » , université coruna, 2018.
- [158] Reffuge Data set; <https://ai.baidu.com/broad/introduction>, Last access 26/06/2020
- [159] F. Chollet, et al., Keras, 2020, online available (<https://keras.io>)
- [160] M. Abadi, et al., “TensorFlow: a system for large-scale machine learning”, in: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pp. 265–283, 2016
- [162] R. Touahri, N. Azizi, N. E. Hammami, F. Benaïda, N. Zemaï: “A two-stage optic disc segmentation based on U-Net architecture for glaucoma diagnosis”, (submitted)