

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب البلدية
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الآلية والالكترونقى
Département d'Automatique et d'Electrotechnique



Mémoire de Master

en Automatique

Spécialité : Automatique et Informatique industrielle

Présenté par

BenM'Hamed Imène

&

Nekkaz Hanane

Analyse de carcinomes mammaires par les techniques du deep learning pour l'aide à l'interprétation

Proposé et dirigé par : Dr. Reguieg F. Zohra & Pr. Benblidia Nadja

Année Universitaire 2021-2022

Dédicaces

Je tiens à dédier ce modeste travail à :

Mes très chers parents, pour tous leurs sacrifices, leurs amours, leurs soutiens et leurs conseils durant toutes mes années d'étude, merci d'être toujours là.

A mes sœurs Amani, Nour EL Houda, Sawssen et Amira et à mon frère Mohamed El Hadi, pour leurs encouragements.

A mon cher fiancé Mohamed, pour ses encouragements, son soutien et ses judicieux conseils.

A mes meilleures amies, Dallel et Samah qui m'ont soutenu, partagé avec moi de durs moments, mais aussi des moments de joie.

Imène

Je dédie ce modeste travail à :

*Mes chers parents, pour tous leurs sacrifices,
leur amour, leur tendresse, leur soutien et leurs
prières tout au long de mes études.*

*Mes grands-parents mes oncles et mes tantes,
pour leur amour et leur soutien moral.*

Mes chers frères pour leurs encouragements.

*Que ce travail soit l'accomplissement de vos
vœux, et le fruit de votre soutien, « MERCI »
d'être toujours là pour moi.*

Hanane

Remerciements

Tout d'abord, on remercie Allah pour la patience, la volonté, la force et la santé qu'il nous a donné afin de réaliser ce travail.

Nos chaleureux remerciements s'adressent à notre promotrice Dr. Reguieg F. Zohra. Nous lui exprimons toute notre reconnaissance pour son encadrement, ses conseils, l'aide et le temps qu'elle a bien voulu nous consacrer et sans qui, ce mémoire n'aurait jamais vu le jour.

Nous tenons aussi à exprimer nos sincères remerciements, à notre co-promotrice Pr. Benblidia Nadjia, pour son soutien et, son aide qui nous a été très utile.

Nous remercions aussi nos parents, pour leur soutien moral, leur disponibilité, leur compréhension et surtout leurs encouragements quotidiens : « Un grand merci ».

Enfin, nous remercions toute personne qui a, à des degrés divers, contribué sur le plan intellectuel, technique, moral à l'achèvement de ce travail.

ملخص: من بين السرطانات الأكثر شيوعاً سرطان الثدي. ولسنوات عديدة ، كان هذا السرطان هو السبب الرئيسي للوفاة بين النساء. تم تطوير العديد من التقنيات والأساليب لاكتشافها وتصنيفها بشكل أفضل. في هذا السياق ، يتيح عملنا تصنيف الدرجات المختلفة لسرطان الثدي باستخدام نظام تصنيف مجهز بالذكاء الاصطناعي. يعتمد هذا DenseNet-121 و VGG-16 وهي CNN النظام على استخدام تقنيات التعلم العميق ، وأربعة نماذج مختلفة من ، للحصول على دقة أفضل أثناء هذا التصنيف. يتم تطبيق النظام على ResNet-50 و DenseNet-201 و BreakHis (تحتويان على صور نسيجية مرضية ، وتستخدم قاعدة بيانات IDC و BreakHis قاعدة بيانات لتصنيف درجات العينات الخبيثة. كشفت دراستنا أن IDC لتصنيف العينات الحميدة والخبيثة ، قاعدة بيانات كان الأنسب لنظامنا. DenseNet-121.

كلمات المفاتيح: سرطان الثدي ، نظام التصنيف ، الذكاء الاصطناعي ، التعلم العميق ، نماذج CNN ، الصور النسيجية المرضية ، العينات الحميدة والخبيثة.

Résumé : Le cancer du sein, représente la première cause de mortalité chez la femme, dans le monde. Plusieurs techniques et méthodes, ont été développées pour le détecter à un stade précoce. Notre travail s'inscrit dans ce cadre, afin d'analyser les différents grades des carcinomes mammaires, à l'aide d'un système de classification doté de l'intelligence artificielle. Ce système se base sur l'utilisation des techniques d'apprentissage profond, selon quatre différents modèles des réseaux convolutifs pré-entraînés, qui sont le VGG-16, le DenseNet-121, le DenseNet-201 et le ResNet-50, pour avoir une meilleure précision, lors de la détection. Le système développé est appliqué sur deux bases de données histopathologiques mammaires (BreakHis et IDC). La base BreakHis sert à catégoriser les échantillons bénins et malins et la base IDC, sert à classifier les grades des échantillons malins pour une aide à l'interprétation. Notre étude a révélé que le DenseNet-121 est le plus adapté à la classification des carcinomes mammaires, avec une précision de 99%.

Mots clés : Images histopathologiques mammaires, Carcinomes mammaires, Intelligence artificielle, Classification, Apprentissage profond, Réseaux convolutifs pré-entraînés, Aide à l'interprétation.

Abstract: Breast cancer is the leading cause of women death. Several techniques and methods have been developed, to detect and classify it at early stage. Our work, is part of this context, making possible to analyze the different grades of breast carcinomas, using a classification system through artificial intelligence. This system is based on the use of deep learning techniques, and four different models of pre-trained convolutional network which are the VGG-16, the DenseNet-121, the DenseNet-201 and the ResNet-50, to have a better precision during this classification. The system is applied on two histopathological images databases (BreakHis and IDC). The BreakHis database is used to categorize benign and malignant samples and the IDC database, is used to classify the grades of malignant samples. Our study revealed that DenseNet-121 is the most suitable for the classification of breast carcinomas with a precision of 99%.

Keywords: Breast histopathological images Breast carcinomas, Artificial intelligence, Classification, Deep learning, Pre-trained convolutional network, Computer aided interpretation.

Listes des acronymes et abréviations

ADAM : Adaptive Moments.

CNN: Convolutional Neural Network

DenseNet: Densely connected convolutional network

DL: Deep Learning

ELU: Exponential Linear Units

FC: Fully connected

FN: False negative

FP: False positive

IA: Intelligence Artificielle.

IDC: Invasive ductal carcinoma

ML: Machine Learning.

MLP: Multi-Layer Perceptron

Numpy: Numerical python

ReLU: Rectified Linear Unit

RMSProp : Root Mean Square Propagation

RNN : Recurrent Neural Network

SGD : Stochastic Gradient Descent

Spyder : Scientific python Development environment

Tanh: Hyperbolic tangent

TN: True negative

TP: True positive

VGG : Visual Geometry Group

Table des matières

Introduction générale	1
Chapitre 1 Contexte médical	
1.1 Introduction	3
1.2 Anatomie du sein	3
1.2.1 Configuration externe	4
1.2.2 Structure	5
1.3 Pathologies du sein	6
1.3.1 Les principales maladies du sein	6
1.3.2 Les Pathologies bénignes du sein	8
1.4 Les système d'acquisitions en imagerie radiographique mammaire	9
1.4.1 Mammographie	9
1.4.2 Echographie	11
1.4.3 Imagerie par résonance magnétique mammaire (IRM)	12
1.5 Etude histologique	13
1.5.1 Lésions bénignes	14
1.5.2 Carcinomes	15
1.5.3 Classification histologique	20
1.6 Intelligence artificielle dans la médecine	23
1.7 Conclusion	24
Chapitre 2 Notions d'apprentissage automatique et profond	
2.1 Introduction	25
2.2 Introduction à L'intelligence artificielle	26
2.2.1 Intelligence artificielle faible	26
2.2.2 Intelligence artificielle forte	26
2.3 Apprentissage automatique	26
2.3.1 Types d'apprentissage automatique	26
2.4 Réseaux de neurones	29
2.4.1 Neurones biologiques	29
2.4.2 Premiers neurones artificiels	30
2.4.3 Principe	31
2.4.4 Fonctions d'activation	32
2.4.5 Perceptron monocouche	37
2.4.6 Perceptron multicouche (MLP)	38
2.5 Apprentissage profond	39
2.5.1 Différence entre l'apprentissage automatique et l'apprentissage profond ...	40
2.5.2 Réseaux de neurones convolutifs	41
2.5.3 Optimiseurs	46
2.5.4 Réseaux neuronaux récurrents (RNN)	47

2.6 Apprentissage par transfert	48
2.7 Présentation de quelques modèles des CNN	49
2.7.1 Modèle Alexnet	49
2.7.2 Modèle VGG	50
2.7.3 Modèle GoogleNet	50
2.8 Conclusion	51

Chapitre 3 Etude conceptuelle du système d'aide à la classification des carcinomes mammaires

3.1 Introduction	52
3.2 Description de quelques travaux en imagerie mammaire	52
3.3 Conception d'un système d'aide à la classification des carcinomes mammaires	53
3.3.1 Architecture du système conçu	53
3.3.2 Exploration de la base de données	55
3.3.3 Prétraitement des images	56
3.3.4 Augmentation des données	56
3.4 Architectures des réseaux de neurones convolutifs	57
3.4.1 ImageNet	58
3.4.2 Modèle VGG-16	59
3.4.3 Modèle ResNet-50	60
3.4.4 Modèle DenseNet	62
3.4.5 Différence entre DenseNet et ResNet	67
3.5 Fine tuning	67
3.5.1 Fine tuning total	68
3.5.2 Fine tuning partiel	68
3.6 Outils de mesures des performances	68
3.6.1 Matrice de confusion	69
3.6.2 Exactitude	69
3.6.3 Précision	70
3.6.4 Spécificité	70
3.6.5 Sensibilité	70
3.6.6 F-score	71
3.6.7 Entropie croisée	71
3.7 Conclusion	72

Chapitre 4 Mise en œuvre du système d'aide à l'interprétation des carcinomes

4.1 Introduction	73
4.2 Environnement de travail	73
4.2.1 Matériel utilisé (Hardware)	73
4.2.2 Langage de programmation utilisé (Software)	73
4.2.3 Bibliothèques utilisés	77
4.3 Bases de données	80
4.3.1 Base de données IDC	80
4.3.2 Base de données BreakHis_V1	81

4.4 Préparation des données	82
4.5 Architectures des modèles du système 'BreastCarciCNN'	83
4.5.1 Modèle VGG-16	84
4.5.2 Modèle ResNet-50	87
4.5.3 Modèles DenseNet-201 et DenseNet-121	90
4.6 Discussion et étude comparative des modèles développés	94
4.7 Conclusion	99
Conclusion générale	100
Bibliographie.....	101

Liste des figures

Figure 1.1. Schéma d'une coupe sagittale du sein	4
Figure 1.2. Anatomie du sein	5
Figure 1.3. Carcinome	6
Figure 1.4 : Cancer du sein	6
Figure1.5. Cancer « in situ » et cancer invasif	7
Figure 1.6. Mammographe	9
Figure 1.7. Utilisation d'un mammographe	10
Figure 1.8. Clichés mammographiques	11
Figure1.9. Appareil échographique	11
Figure 1.10. Imagerie par résonance magnétique	12
Figure1.11. IRM d'une tumeur	13
Figure1.12. Grade histologique du cancer du sein. (a) tumeur différenciée de grade I, (b) tumeur différenciée de grade II, (c) tumeur de grade III avec pléomorphisme nucléaire et cellules en mitose	13
Figure 1.13. Type architectural papillaire et glanduliforme x40	14
Figure 1.14. Prolifération papillaire dans un galactophore x75	14
Figure 1.15. Tubules réguliers séparés par un tissu conjonctif peu abondant x40	15
Figure 1.16. Nettes vacuoles cytoplasmiques et lumière dilatées comme dans le tissu mammaire	15
Figure 1.17. Architecture massive et en comédons x120	16
Figure 1.18. Architecture papillaire x40	16
Figure 1.19. Architecture cribriforme x180	16
Figure 1.20. Carcinome lobulaire in situ x460	16
Figure 1.21. Carcinome lobulaire in situ	16
Figure 1.22. Cellules tumorales canalaire sans signe d'invasion x300	16
Figure 1.23. Cellules tumorales semblables en diamètre x480	17

Figure 1.24. Cellules tumorales Polymorphes autour du tissu conjonctif	17
Figure 1.25. Cellules tumorales avec une quantité modérée de stroma dense x180	17
Figure 1.26. Cellules tumorales infiltrent le tissu conjonctif x75	18
Figure 1.27. Cellules tumorales disposées en massifs irréguliers x300	18
Figure 1.28. Petits ilots de cellules tumorales régulières x100	18
Figure 1.29. Noyaux vésiculeux x240	19
Figure 1.30. Petits tubules réguliers infiltrent x100	19
Figure 1.31. Tubules tumoraux avec un stroma dense x300	19
Figure 1.32. Cellules tumorales Polymorphes	19
Figure 1.33. Cellules tumorales Polymorphes	19
Figure 1.34. Cellules normales et cancéreuses	20
Figure 1.35. Grades histopronostiques	22
Figure 1.36. Echantillon de tumeurs tubulaires et mitotiques (a) Cellules épithéliales et tumorales tubulaires (b) Cellules mitotiques	23
Figure 2.1. Interactions entre l'intelligence artificielle, le machine learning et le Deep Learning	25
Figure 2.2. Types d'apprentissage automatique	27
Figure 2.3. Schéma d'apprentissage supervisé	27
Figure 2.4. Apprentissage automatique non-supervisé	28
Figure 2.5. Exemple d'apprentissage automatique semi-supervisé	28
Figure 2.6. Apprentissage automatique par renforcement	29
Figure 2.7. Premiers neurones artificiels	30
Figure 2.8. Neurone artificiel simple	31
Figure 2.9. Structure des réseaux de neurones	31
Figure 2.10. Structure du neurones formel	32
Figure 2.11. Fonction ReLU	33
Figure 2.12. Fonction tangente hyperbolique	33
Figure 2.13. Fonction softmax	34

Figure 2.14. Fonction sigmoïde binaire	35
Figure 2.15. Fonctions sigmoïde bipolaire	35
Figure 2.16. Fonction échelon	36
Figure 2.17. Fonction seuil	36
Figure 2.18. Perceptron simple	37
Figure 2.19. Schéma détaillé d'un perceptron	37
Figure 2.20. Exemple de séparation des patterns	38
Figure 2.21. Perceptron multiple	38
Figure 2.22. Modèle du perceptron multicouche	39
Figure 2.23. Machine Learning et Deep Learning	41
Figure 2.24. Architecture d'un CNN	42
Figure 2.25. Architecture de couche de convolution	42
Figure 2.26. Exemple de convolution	43
Figure 2.27. Opération de max pooling	44
Figure 2.18. Couche entièrement connectée	45
Figure 2.29. Rétropropagation	47
Figure 2.30. Architecture d'un RNN	48
Figure 2.31. Machine Learning traditionnel et le transfer Learning	49
Figure 2.32. Architecture d'un CNN de Alexnet	49
Figure 2.33. GoogleNet architecture	50
Figure 3.1. Architecture conceptuelle du système d'aide à l'interprétation	54
Figure 3.2. Architecture du modèle VGG-16	60
Figure 3.3. ResNet-50	61
Figure 3.4. Architecture DenseNet-201	65
Figure 3.5. 'DenseBlock'	66
Figure 3.6. Architecture complète de DenseNet	66
Figure 3.7. Architectures de DenseNet et ResNet.....	67

Figure 4.1. Logo d'Anaconda	75
Figure 4.2. Navigateur Anaconda	76
Figure 4.3. Environnement spyder.....	76
Figure 4.4. Logo de keras	77
Figure 4.5. Plateforme Tensorflow	78
Figure 4.6. Un exemple de la validation croisée à 5 folds (plis)	80
Figure 4.7. Images x40 de la base de données IDC	81
Figure 4.8. Images de la base de données BreakHis_v1	82
Figure 4.9. Nombre d'images après la division du dataset	83
Figure 4.10. Nombre d'images après la division du dataset	83
Figure 4.11. Augmentation des données	84
Figure 4.13. Architecture du VGG16	86
Figure 4.14. Exactitude et Perte du modèle VGG16 en utilisant la base (Breakhis_v1)	87
Figure 4.15. Exactitude et Perte du modèle VGG16 en utilisant la base (IDC)	87
Figure 4.16. Architecture ResNet50 adaptée aux bases histologiques	88
Figure 4.17. Architecture de ResNet-50	89
Figure 4.18. Exactitude et perte du modèle ResNet-50 en utilisant la base (BreakHis_V1) ..	89
Figure 4.19. Exactitude et perte du modèle ResNet-50 en utilisant la base (IDC)	90
Figure 4.20. Architecture DenseNet adaptée aux bases histologiques.....	91
Figure 4.21. Architecture DenseNet-201	91
Figure 4.22. Architecture DenseNet-121	91
Figure 4.23. Exactitude et perte du modèle DenseNet-201 en utilisant la base (BreakHis_V1)	92
Figure 4.24. Exactitude et perte du modèle DenseNet-201 en utilisant la base (IDC)	93
Figure 4.25. Exactitude et perte du modèle DenseNet-121 en utilisant la base (BreakHis_V1)	93
Figure 4.26 : Exactitude et perte du modèle DenseNet-121 en utilisant la base (IDC)	94
Figure 4.27. Matrice de confusion pour la base de données (IDC)	95

Figure 4.28. Matrice de confusion pour la base de données (Breakhis_v1)	96
Figure 4.29. Résultats de 'BreastCarcicNN' obtenus pour la 1 ^{ère} base de données (IDC)	96
Figure 4.30. Résultats de 'BreastCarcicNN' pour la 2 ^{ème} base de données (Breakhis_v1)	97

Liste des tableaux

Tableau 1.1. Caractéristiques d'une cellule tumorale	20
Tableau 3.1. Résultats de différentes architectures dans le challenge	59
Tableau 3.2. Architecture du ResNet50	62
Tableau 3.3. Différents modèles de DenseNet	63
Tableau 3.4. Architecture du DenseNet-121	64
Tableau 3.7. Matrice de confusion	69
Tableau 4.1. Résultats du système 'BreastCarcicNN'	95
Tableau 4.2 : Résultats du modèle DenseNet-121 suivant deux optimiseurs (Adam et SGD) ..	98
Tableau 4.3 : Résultats du modèle DenseNet-121 suivant différents batch size (Breakhis_v1) (LR=0.0001 pour l'optimiseur Adam)	98

Introduction générale

Le cancer est un terme général désignant une maladie pour lesquelles certaines cellules du corps humain, se divisent d'une manière incontrôlée. Les nouvelles cellules résultantes peuvent former une tumeur maligne (un néoplasme), ou se propager à travers le corps. Parmi les cancers les plus fréquents chez les femmes, le cancer du sein est en tête avec 2.3 millions de cas confirmés dans le monde [1]. Il arrive aussi, que les hommes soient touchés par cette pathologie, mais rarement.

Une femme sur 8 développe un cancer dans sa vie, surtout à partir de 50 ans. C'est la première cause du décès féminin, mais détecté suffisamment tôt, il se guérit dans 9 cas sur 10 [1]. Dans ce contexte, l'examen clinique des pathologies mammaires, peut être effectué suivant plusieurs systèmes d'acquisition, tels que la mammographie, pour analyser des clichés, ou l'histologie pour déterminer les différents grades de la maladie. Cet examen anatomopathologique, consiste à analyser au microscope des cellules ou des tissus prélevés du sein, pour établir de façon précise le diagnostic du cancer. On parle de preuve histologique. Cependant, il arrive souvent que des erreurs d'interprétation se produisent, à cause de la fatigue du médecin. L'intelligence artificielle est devenue alors, cruciale dans la prise de décisions dans le domaine médical. Elle est capable d'exploiter les données cliniques et permettre des diagnostics, plus fiables et plus rapides [2].

Notre travail se concentre sur la contribution de l'intelligence artificielle dans la prédiction et la classification des carcinomes mammaires, en utilisant quatre architectures des réseaux de neurones convolutifs. L'opération consiste d'extraire automatiquement des informations pertinentes, à partir d'images histologiques mammaires, afin d'établir la classification des 3 grades des carcinomes.

Une comparaison des résultats obtenus par les 4 modèles implémentés, tels que le VGG-16, le DenseNet-201, le DenseNet-121, et le ResNet-50, est réalisée afin de trouver le modèle le plus performant dans cette application.

Pour mener à bien notre projet, le mémoire est organisé en quatre chapitres :

- Le premier chapitre, est consacré au contexte médical. Nous débuterons par décrire l'anatomie du sein, les différentes pathologies mammaires, ainsi que les différents systèmes d'acquisitions.
- Le deuxième chapitre, introduit le machine learning, le deep learning et les différents types de réseaux de neurones profonds.
- Le troisième chapitre, est consacré à l'étude conceptuelle du système d'aide à la classification des carcinomes mammaires, ainsi qu'une présentation des différentes architectures utilisées dans cette étude.
- Le quatrième chapitre est destiné à la mise en œuvre du système développé, la présentation de l'étude expérimentale et des résultats obtenus par divers modèles neuronaux.

1.1 Introduction

En termes de mortalité, le cancer du sein féminin, occupe la 5^{ème} place dans le monde [3]. Avec l'âge, le risque de développer un cancer, augmente fortement, c'est pourquoi le dépistage est recommandé aux femmes de 50 à 70 ans tous les 2 ans, nécessitant une prise en charge pluridisciplinaire.

Dans ce contexte, le dépistage permet de détecter une éventuelle anomalie ou un cancer, avant l'apparition de symptômes. En effet, la détection précoce suivant des systèmes d'acquisition spécifiques à l'imagerie médicale, augmente les chances de guérison. Ainsi, si un cancer est découvert, il sera traité tôt avec des traitements moins lourds.

Dans ce chapitre, nous commencerons par découvrir l'anatomie du sein, les différentes pathologies mammaires, dont les carcinomes et ses différents stades. Nous montrons aussi, le rôle de l'intelligence artificielle dans la médecine.

1.2 Anatomie du sein

Les seins, du latin "sinus" qui signifie "courbure, sinuosité, pli" sont des organes pairs présents chez la femme et sous une forme atrophiée chez l'homme.

Le sein étant l'organe de la lactation, est une glande exocrine, paire et lobulée de morphologie très variable selon la phase de vie. Il est situé sur la paroi thoracique antérieure entre le 3^{ème} et le 7^{ème} cotes, en dehors de sternum et en dedans de la ligne axillaire antérieure (figure 1.1). Limité en bas par le sillon sus mammaire (qui représente sa limite supérieure), ses dimensions sont variables, en fonction de l'activité glandulaire. On note une augmentation de sa taille au cours de la grossesse ainsi que de l'allaitement [4, 5] [6].

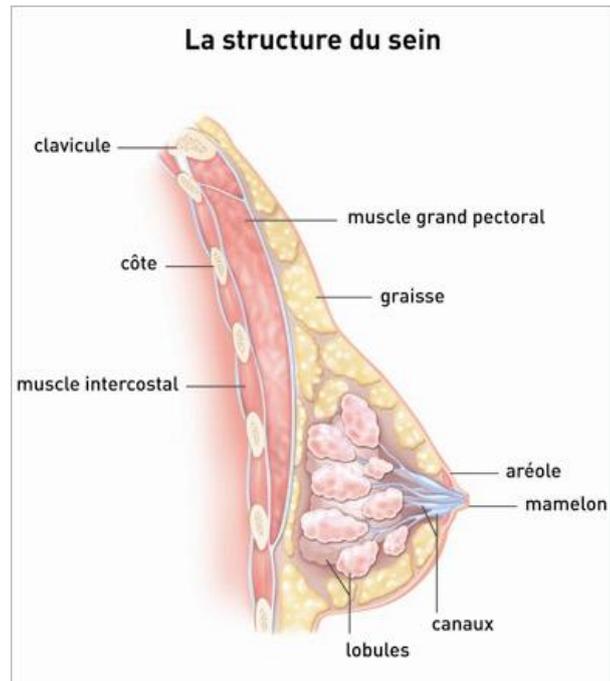


Figure 1.1 : Schéma d'une coupe sagittale du sein [6]

1.2.1 Configuration externe

La configuration externe du sein se distingue par la peau, l'aréole et le mamelon (figure 1.2).

- **Peau**

Plus épaisse à la périphérie de la glande qu'à proximité de l'aréole, la peau adhère intimement à la glande par les ligaments de Cooper. Elle ne glisse pas sur les tissus sous-jacents car dépourvue de tissu adipeux sous-jacent. Elle est séparée de la glande par le muscle mamillaire, fait de fibres circulaires et de fibres radiés.

- **Aréole**

C'est un disque cutané de 3 à 5cm de diamètre, pigmenté, de coloration brunâtre. Sa surface est irrégulière, on y observe de petites saillies (15 à 20) les tubercules de Morgagni, qui sont des glandes sébacées.

- **Mamelon**

Il est de forme cylindrique ou conique, pigmenté, séparé de l'aréole par un sillon. A la surface du mamelon, il y a les orifices d'aboutement (ou les pores) des canaux galactophores. Il est doublé à sa face profonde par le muscle mamillaire.

Le mamelon et l'aréole forment une unité, la **plaque aréolo-mamelonnaire**.

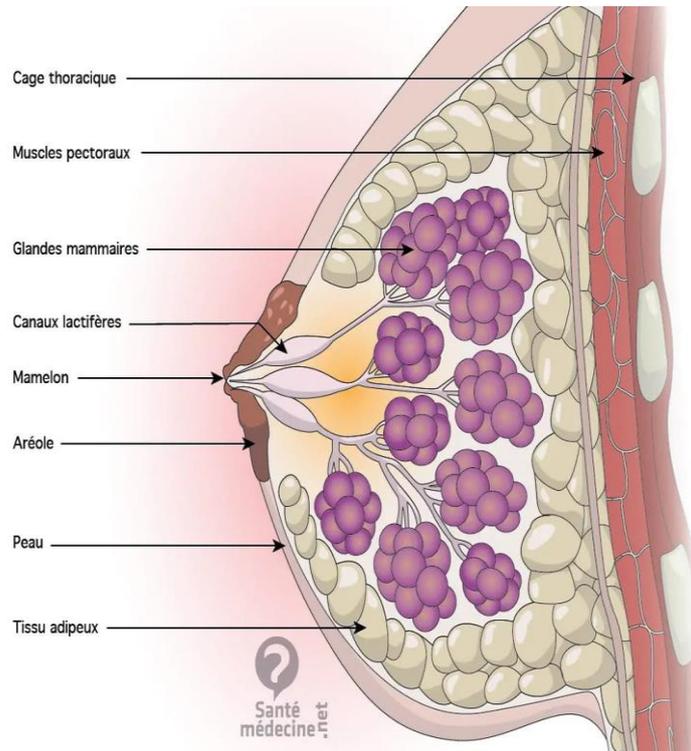


Figure 1.2 : Anatomie du sein [6]

1.2.2 Structure

Le sein est constitué de la glande mammaire d'un tissu adipeux et conjonctif [5, 6].

- **Glande mammaire** : c'est une masse de densité variable, discoïde, aplatie d'avant en arrière, de contour irrégulier. Elle est organisée en une vingtaine de lobes.

Chaque lobe est composé de 20 à 40 lobules, où chacun d'eux contient 10 à 100 alvéoles. L'unité de base est l'acinus ou l'alvéole.

L'alvéole est une cavité en forme de cul de sac, qui constitue la partie sécrétrice de la glande. Chaque acinus se draine par un canal intra-lobulaire (canal de 3^{ème} ordre).

Les acini et les canaux intra-lobulaires forment un lobule, qui se draine par un canal interlobulaire (canal de 2^{ème} ordre). Plusieurs lobules se réunissent pour former un lobe glandulaire, qui se draine par un canal galactophore de premier ordre.

- **Tissu adipeux et conjonctif** : Etroitement liée au tissu glandulaire, on distingue dans ce type, deux couches graisseuses :
 - 1) Une couche antérieure pré-glandulaire est cloisonnée par des travées conjonctives : les ligaments de Cooper qui relient la peau à la glande, en formant les crêtes de Ducret. Elle se trouve en dehors de la plaque aréolo-mamelonnaire.
 - 2) Une couche postérieure rétro-mammaire : limitée par le fascia superficialis, elle est séparée de l'aponévrose du grand pectoral par un tissu conjonctif.
 L'ensemble peau-glande-graisse, glisse sur le grand pectoral.

1.3 Pathologies du sein

Les maladies du sein sont de plusieurs natures [7, 8]. Elles peuvent être des pathologies malignes ou bénignes (figures 1.3 et 1.4).

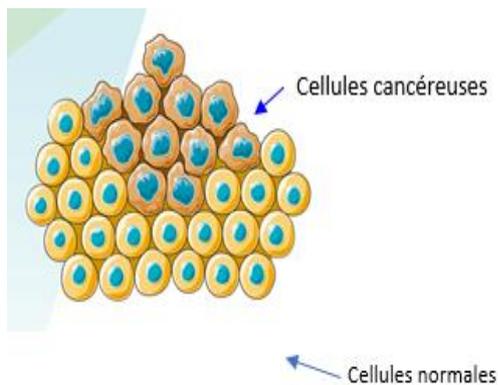


Figure 1.3 : Carcinome [7]

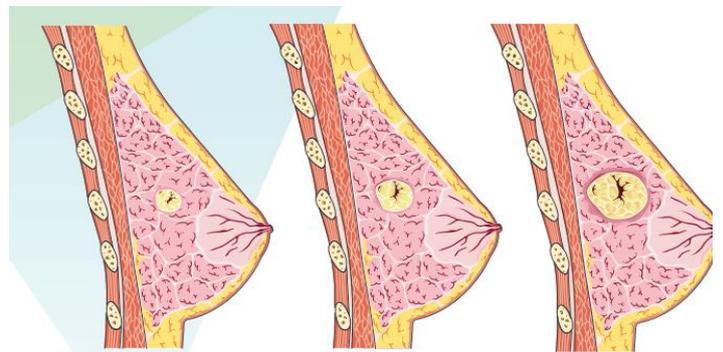


Figure 1.4 : Cancer du sein [7]

1.3.1 Pathologies malignes du sein

Ces pathologies peuvent se regrouper en plusieurs types :

- **Carcinome canalaire** : c'est le plus fréquent qui se forme dans les cellules glandulaires des canaux mammaires (figure 1.5). Il peut être in situ (présence uniquement dans les canaux mammaires) ou infiltrant (extension dans les tissus mammaires avoisinants). Diagnostiqué assez tôt grâce à la mammographie, il reste non invasif et on en guérit dans presque tous les cas avec un traitement adapté.

- En revanche, sans traitement, il peut poursuivre sa croissance et devenir invasif, en se propageant à l'extérieur des canaux de lactation.

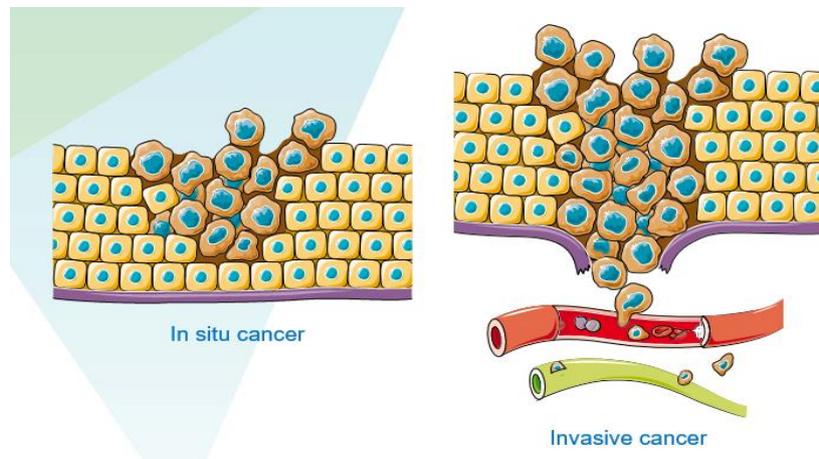


Figure 1.5 : Cancer « in situ » et cancer invasif [8]

- **Carcinome lobulaire** : Il se forme dans les lobules. Les cellules cancéreuses traversent leurs parois et se développent, dans les tissus environnants.

D'autres cancers plus rares ont des caractéristiques plus visibles comme :

- **Cancer inflammatoire du sein** : le sein devient rouge, enflé, chaud et peut avoir un aspect de peau d'orange. Les cellules cancéreuses se forment d'abord dans les canaux mammaires, puis migrent rapidement vers les vaisseaux lymphatiques de la peau. Cette pathologie mammaire est de forme plutôt agressive.
- **Maladie de Paget** : elle est caractérisée par une petite plaie au mamelon qui ne guérit pas. Elle se forme au niveau de la partie superficielle de la glande mammaire. Cette pathologie est fréquemment associée à un autre cancer du sein.
- **Pathologies malignes plus rares** : On distingue dans ce cas, des lymphomes diffus à grandes cellules, des sarcomes des tissus mous, des mélanomes, des tumeurs métaplastiques, des carcinomes adénoïdes kystiques, des tumeurs phyllodes ainsi que des carcinosarcomes, un cancer de type basal.

1.3.2 Pathologies bénignes du sein

Les **pathologies bénignes du sein** sont très fréquentes, survenant à tous les âges. Elles sont dominées par les calcifications, la **pathologie** kystique et par la **pathologie** fibroadénomateuse [9, 10, 11].

- **Calcifications mammaires** : ce sont des dépôts de calcium, qui se forment dans le tissu du sein. Elles sont assez fréquentes et la plupart d'elles, ne sont pas liées au cancer.
- **Kyste mammaire** : c'est un sac rempli de liquide, produit par les glandes mammaires se formant dans les tissus du sein. Il peut se stabiliser ou augmenter de taille rapidement, générant malaise et inconfort. C'est le type le plus fréquent de masses bénignes, chez les femmes âgées de 30 à 50 ans.
- **Adénofibromes** : encore appelé le fibroadénome ou fibrome. C'est une tumeur bénigne fréquente, qui ne comporte aucune gravité. La tuméfaction est indolore, palpable (nodule arrondi ou ovalaire composée de tissus fibreux et glandulaires), qui se développe dans le sein à partir de la glande mammaire.

Ses caractéristiques principales sont les suivantes :

- Une taille comprise entre 1 et 15 cm, le plus souvent entre 2 et 3 cm ;
 - Une localisation possible dans toutes les zones du sein ;
 - Un aspect lisse, ferme et bien délimité ;
 - L'absence d'adhérences à la peau et aux autres tissus du sein ;
 - Un seul ou plusieurs **adénofibromes** susceptibles d'affecter un seul ou les deux seins.
-
- **Tumeur phyllode** : appelée aussi cystosarcome phyllode, est une tumeur non épithéliale et rare qui touche le sein. Elle peut être bénigne ou maligne [10]. Il s'agit d'une tumeur fibroépithéliale, dont la composante du tissu conjonctif prédomine. Elle se présente comme une masse qui peut être bosselée et grossir rapidement. Elle peut même parfois déformer le sein [11].

1.4 Systèmes d'acquisition en imagerie radiographique mammaire

L'imagerie médicale est un domaine de la médecine, qui regroupe différentes techniques permettant d'observer l'intérieur du corps humain. Elle est aussi un élément essentiel à la recherche clinique, l'étude des maladies et la mise au point de nouvelles possibilités de traitement. L'imagerie recouvre une grande variété de technologies développées, grâce à l'exploitation des grandes découvertes de la physique du 20^e siècle. Les outils d'imagerie souvent utilisés pour le dépistage et le diagnostic du cancer du sein, sont généralement : l'échographie (par ultrasons), l'IRM (par résonance magnétique) et la mammographie (par rayon X) [12, 13, 14, 15].

1.4.1 Mammographie

Une mammographie est une radiographie des seins. Elle permet d'avoir une image des tissus intérieurs du sein à l'aide de rayons X et de détecter ainsi, des anomalies qui peuvent être le signe d'un cancer du sein.

L'appareil qui permet de faire la mammographie est le mammographe. Cet appareil se compose d'un générateur de rayons X de faible énergie (entre 20 et 50 keV) et d'un système de compression du sein (figure 1.6).

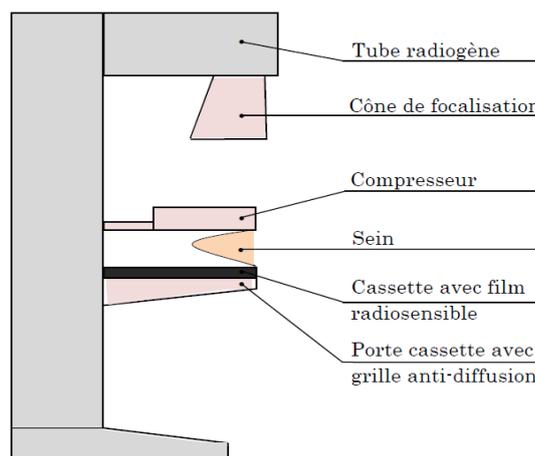


Figure 1.6 : Mammographe [13]

Deux incidences sont toujours faites. Une incidence de face, où le sein est mis à l'horizontale et est comprimé suffisamment, pour qu'il ne bouge pas. Une seconde incidence oblique, permet d'avoir le sein en deux dimensions.

Ce qui est important à savoir, c'est que le sein ne doit absolument pas bouger. Car la mammographie va permettre de rechercher de tout petits signes de cancer (des microcalcifications, qui font en moyenne 0,3 mm). C'est pour cela qu'il faut une certaine compression, pour obtenir une image parfaite (figure 1.7).

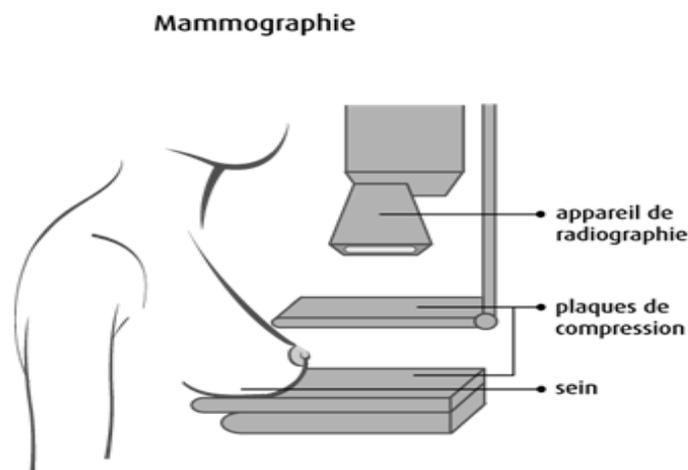


Figure 1.7 : Utilisation d'un mammographe [13]

L'examen mammographique peut être réalisé soit dans le cadre d'un dépistage du cancer du sein (mammographie de dépistage), soit en présence de symptômes (mammographie de diagnostic).

- **Mammographie de dépistage :** est une radiographie des seins (figure 1.8) qui a pour but de détecter la présence des cancers de petite taille, avant l'apparition des symptômes (stade précoce) ou un trouble du sein. C'est la meilleure méthode, pour dépister le cancer du sein [15]. Parfois, le radiologue peut compléter cet examen, par une échographie qui lui permet d'analyser plus finement la lésion.

• La tumeur est une boule.

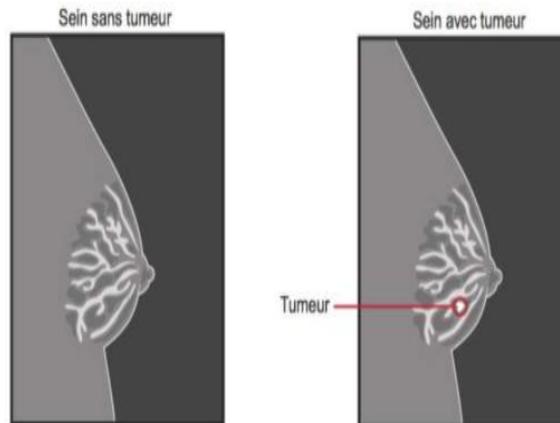


Figure 1.8 : Clichés mammographiques [15]

- **Mammographie de diagnostic :** elle est utilisée lors de la présence des symptômes ou de lésions du sein, comme une masse ou une région anormale. Elle permet d'avoir des images plus détaillées du sein prises sous différents angles. Elle peut se faire pour un sein (mammographie unilatérale) ou pour les 2 seins (mammographie bilatérale), afin que le médecin puisse comparer le tissu de chaque sein. Elle dure un peu plus longtemps, que la mammographie de dépistage.

1.4.2 Echographie

Une échographie du sein est un système d'imagerie médicale, qui utilise des ultrasons (des ondes de haute fréquence) pour visualiser les tissus du sein (figure 1.9). L'image est obtenue à l'aide d'une sonde, qui se déplace sur le sein recouvert d'un gel hypoallergénique [12].



Figure 1.9 : Appareil échographique [12]

Cet examen est considéré comme un examen complémentaire à la mammographie, qui permet d'identifier toutes les structures du sein et de déterminer s'il s'agit simplement de kystes remplis d'eau liquide, de tumeurs cancéreuses ou d'un autre type de pathologie. L'échographie n'est pas invasive et ne présente aucun risque connu pour la patiente.

Une raison pour laquelle un médecin peut demander une échographie mammaire, est lorsqu'une patiente a des seins extrêmement denses, la visualisation du tissu mammaire sur une mammographie, devient difficile.

1.4.3 Imagerie par résonance magnétique mammaire (IRM)

L'IRM mammaire est un examen non-radiant (figure 1.10), qui utilise la résonance magnétique (figure 1.11). Il est souvent pratiqué après une échographie mammaire, car il fournit des résultats plus précis et présente des avantages évidents pour les femmes à haut risque familial du cancer du sein. Il est donc destiné à affiner le diagnostic [14].



Figure 1.10 : Imagerie par résonance magnétique [14]

Une IRM est effectuée :

- Pour faire la différence entre une anomalie bénigne et une anomalie cancéreuse dans le cadre du dépistage des femmes à haut risque de cancer du sein.
- Lorsque l'imagerie standard (mammographie ou échographie) ne permet pas de conclure avec certitude à l'absence de malignité.
- Pour guider un prélèvement par biopsie.
- Dans le cadre du bilan d'extension.

- Pour évaluer la réponse aux thérapies néo adjuvantes (chimiothérapie réalisée avant l'opération).
- A la recherche de récurrence locale après la chirurgie conservatrice.
- Pour vérifier l'état d'implants mammaires, en particulier si on croit qu'il y a eu rupture.



Figure 1.11 : IRM d'une tumeur [14]

1.5 Etude histologique

C'est l'examen anatomopathologique d'un échantillon de la tumeur, qui permet d'évaluer le type exact du cancer et de définir son grade (figure 1.12). En effet, il existe plusieurs degrés de malignité. Ainsi, plus une cellule cancéreuse ressemble aux cellules normales (elle est dite alors bien différenciée), moins elle est agressive. Plus une cellule se modifie par rapport aux cellules normales (elle est alors indifférenciée), plus elle est agressive [16, 17, 18] [19, 20, 21].

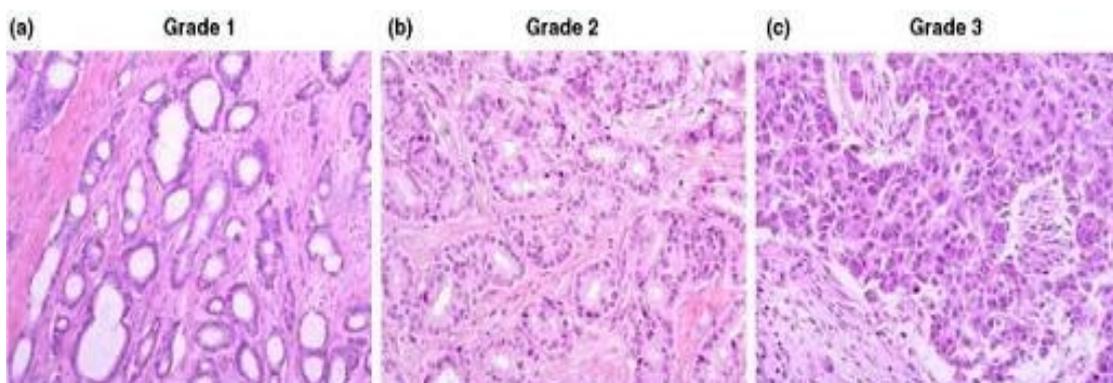


Figure 1.12 : Grades histologiques du cancer du sein
 (a) tumeur différenciée de grade I, (b) tumeur différenciée de grade II, (c) tumeur de grade III avec pléomorphisme nucléaire et cellules en mitose [17]

Le rôle de l'anatomie pathologique, consiste à poser ou confirmer un diagnostic sur la base d'analyses des tissus et/ou des cellules prélevées chez la patiente, en utilisant des techniques principalement fondées sur la morphologie macroscopique et microscopique. Cet examen comprend l'évaluation de facteurs pronostiques et, la mise en évidence de mutations génétiques des cellules tumorales. Ceci permet de prédire l'efficacité des thérapies [18, 21]. Dans ce cadre, le diagnostic du pathologiste apporte des informations déterminantes, qui conditionnent les modalités de prise en charge thérapeutique du cancer et de son suivi.

1.5.1 Lésions bénignes

On distingue plusieurs types de tumeurs bénignes à partir d'images histologiques [17, 21].

- **Papillome intracanalair** : c'est une tumeur papillaire bien individualisée (figure 1.13) se développant dans un canal galactophore. Dans sa forme typique, elle se caractérise par une architecture régulière, souvent glanduliforme avec des axes conjonctifs et vasculaires bien développés.

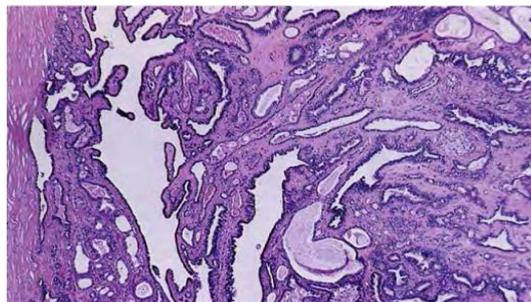


Figure 1.13 : Type architectural papillaire et glanduliforme x40 [17]

- **Adénome du mamelon** : c'est une tumeur épithéliale bénigne (figure 1.14) se développant dans les canaux galactophores rétromamelonnaires et montrant une prolifération intracanalair, souvent associée aux stades évolués à une composante tubuliforme.

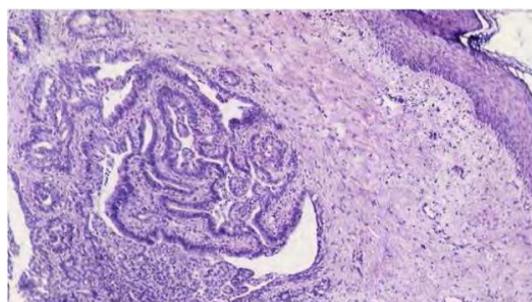


Figure 1.14 : Prolifération papillaire dans un galactophore x75 [17]

- **Adénome** : c'est une tumeur bénigne peu fréquente, bien limitée, faite de formations glanduliformes et d'un stroma peu abondant. Ses deux types sont :
 - **Adénome tubuleux** : c'est un adénome constitué de tubules réguliers (figure 1.15) ressemblant aux canalicules d'un lobule quiescent.

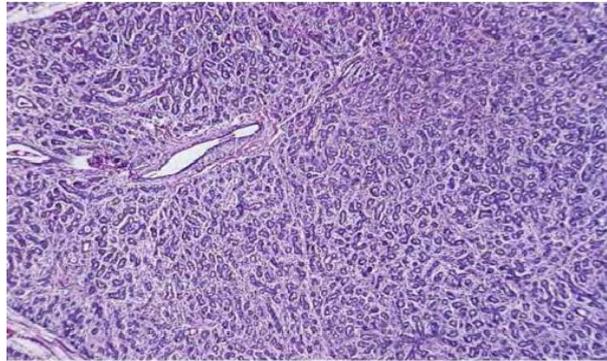


Figure 1.15 : Tubules réguliers séparés par un tissu conjonctif peu abondant x40 [17]

- **Adénome lactant** : c'est un adénome constitué de formations tubulo-acineuses (figure 1.16), comportant d'importants signes sécrétoires observables, pendant la grossesse et la lactation [23].

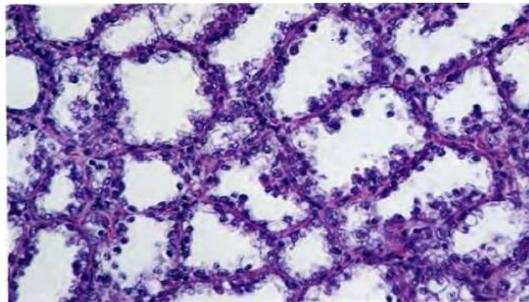


Figure 1.16 : Vacuoles cytoplasmiques nettes et dilatées [17]

1.5.2 Carcinomes

Les tumeurs malignes que nous remarquons lors de l'examen histologique, peuvent se regrouper en des carcinomes infiltrants et non infiltrants.

a Carcinomes non infiltrants

- **Carcinome intracanalair** : c'est un carcinome des canaux galactophores n'infiltrant pas le tissu conjonctif voisin.

- Il est caractérisé par quatre types architecturaux : massif, en comédons, papillaire et cribriforme (figures 1.17, 1.18, 1.19 et 1.20).

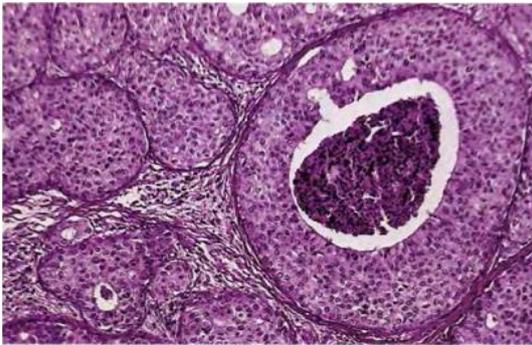


Figure 1.17 : Architecture massive et en comédons x120 [19]

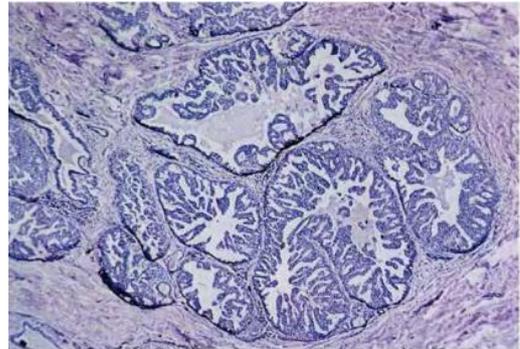


Figure 1.18 : Architecture papillaire x40 [19]

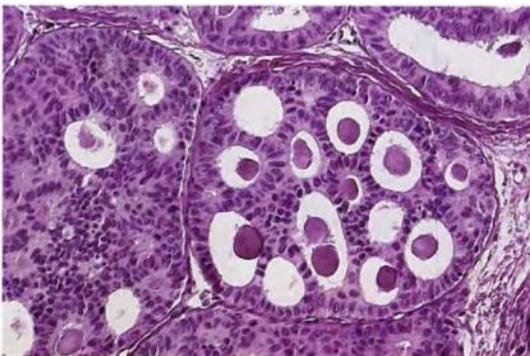


Figure 1.19 : Architecture cribriforme x180 [19]

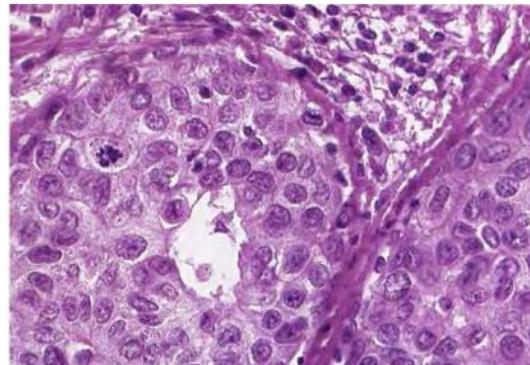


Figure 1.20 : Carcinome lobulaire in situ x460 [19]

- **Carcinome lobulaire in situ** : c'est un carcinome intéressant les canalicules intralobulaires (figures 1.21, 1.22 et 1.23), qui sont comblés et distendus par une prolifération de cellules peu jointives, sans envahissement du tissu conjonctif voisin.

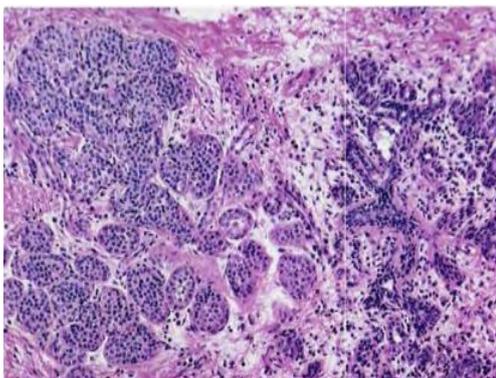


Figure 1.21 : Carcinome lobulaire in situ [19]

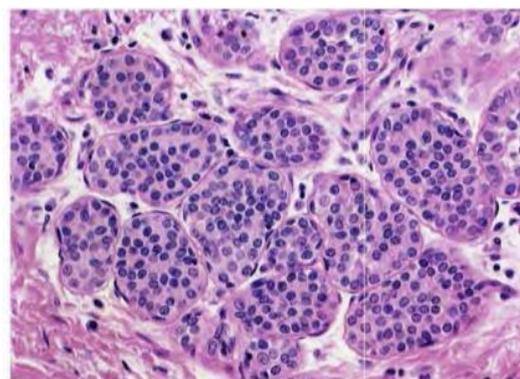


Figure 1.22 : Cellules tumorales canaliculaires sans signe d'invasion x300 [19]

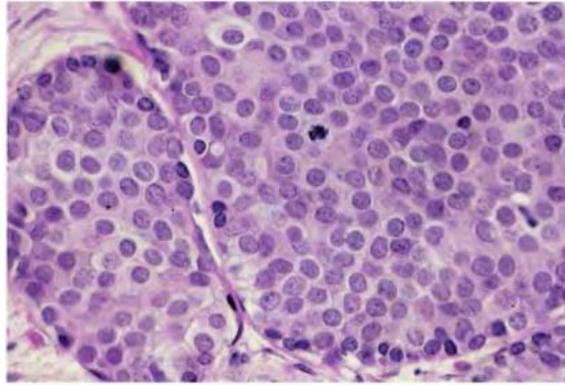


Figure 1.23 : Cellules tumorales semblables en diamètre x480 [19]

***b* Carcinomes Infiltrants**

- **Carcinome canalaire infiltrant** : c'est la forme la plus fréquente des tumeurs malignes du sein. Ce type tumoral (figures 1.24 et 1.25) n'appartient à aucune des autres catégories du carcinome mammaire infiltrant. Les aspects macro- et microscopiques sont extrêmement variables. Les cellules tumorales se disposent généralement en îlots, en travées ou en formations glanduliformes.

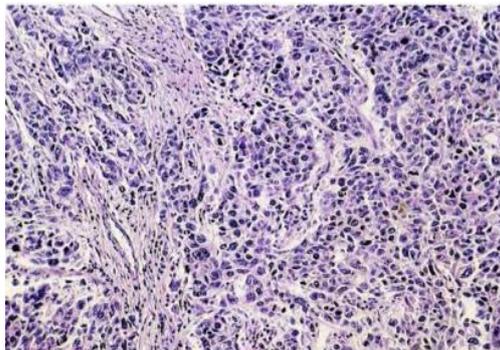


Figure 1.24 : Cellules tumorales polymorphes autour du tissu conjonctif [19]

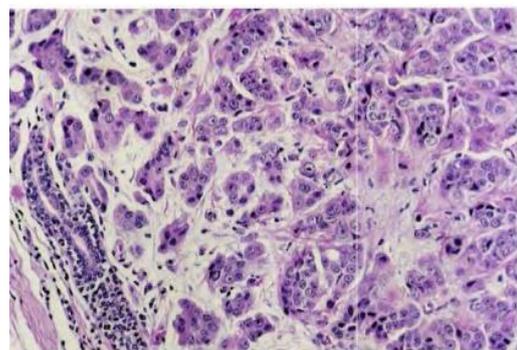


Figure 1.25 : Cellules tumorales avec une quantité modérée de stroma dense x180 [19]

- **Carcinome canalaire infiltrant avec composante intracanaulaire prédominante** : c'est un carcinome essentiellement intracanaulaire (figure 1.26) mais présentant des foyers d'infiltration du tissu conjonctif.

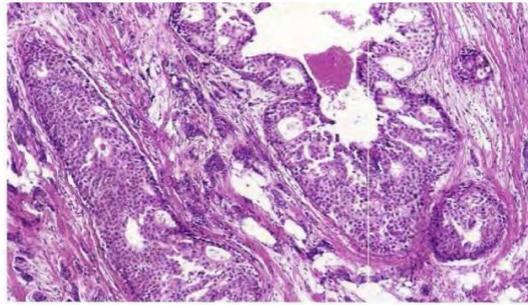


Figure 1.26 : Cellules tumorales infiltrant le tissu conjonctif x75 [19]

- **Carcinome lobulaire infiltrant :** c'est un carcinome infiltrant (figure 1.27) formé de cellules régulières, ressemblant à celles du carcinome lobulaire in situ et ayant en général un faible taux de mitoses.

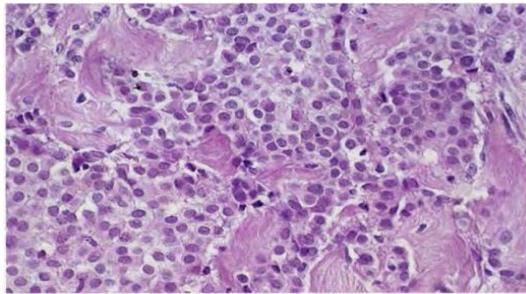


Figure 1.27 : Cellules tumorales disposées en massifs irréguliers x300 [19]

- **Carcinome mucineux :** c'est un carcinome riche en mucus épithélial extracellulaire (figure 1.28). Il peut être visible macroscopiquement et se reconnaître au microscope, comme extra- et intra-cellulaire.

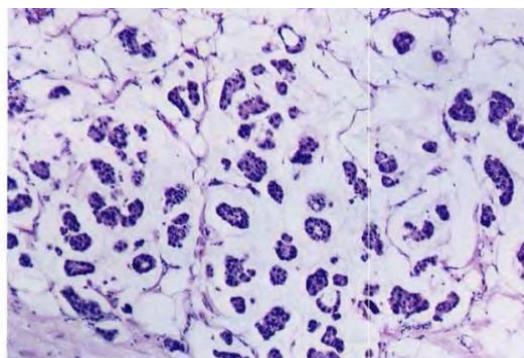


Figure 1.28 : Petits îlots de cellules tumorales régulières x100 [20]

- **Carcinome médullaire** : c'est un carcinome bien limité (figure 1.29) constitué de cellules peu différenciées, dans un stroma peu abondant, avec une intense infiltration lymphoïde.

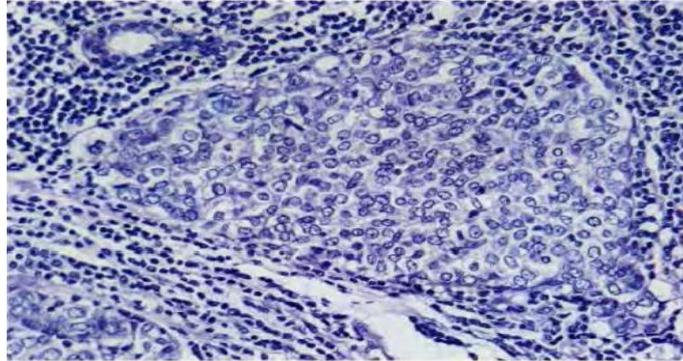


Figure 1.29 : Noyaux vésiculeux x240 [20]

- **Carcinome papillaire** : c'est un carcinome rare dont l'architecture infiltrante (figures 1.30 et 1.31), est surtout faite de structures papillaires.

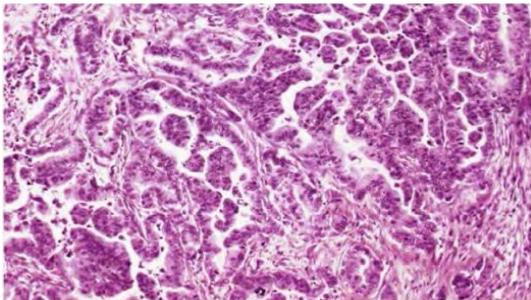


Figure 1.30 : Petits tubules réguliers infiltrant x100 [20]

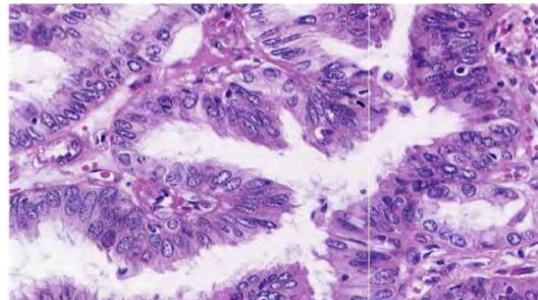


Figure 1.31 : Tubules tumoraux avec un stroma dense x300 [20]

- **Carcinome tubuleux** : c'est un carcinome infiltrant extrêmement différencié (figures 1.32 et 1.33), dont les cellules sont régulières et disposées en tubules bien structurés, typiquement faits d'une seule couche cellulaire et entourées d'un abondant stroma fibreux.

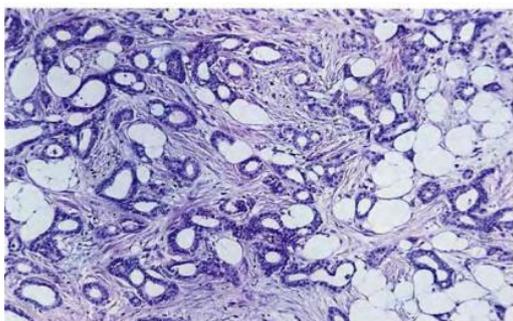


Figure 1.32 : Cellules tumorales Polymorphes [20]

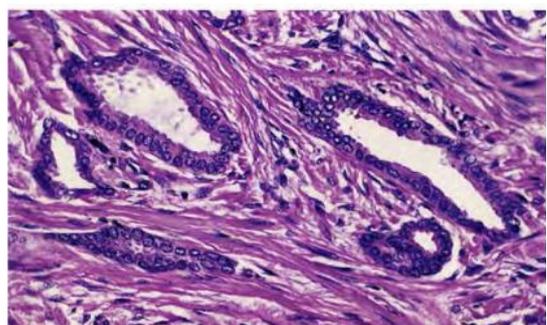


Figure 1.33 : Cellules tumorales Polymorphes [20]

L'aspect microscopique d'une cellule tumorale, est caractérisée suivant 3 critères résumés dans le tableau 1.1.

Critère	Note 1	Note 3
Architecture	La tumeur contient beaucoup de structures bien formées.	La tumeur contient peu ou pas du tout de structures bien formées.
Noyau	Les noyaux de la tumeur sont petits et uniformes.	Les noyaux de la tumeur sont gros. Leur taille et leur forme varient.
Activité mitotique	Les cellules de la tumeur se divisent lentement = faible nombre de mitoses.	Les cellules de la tumeur se divisent rapidement = important nombre de mitoses.

Tableau 1.1 : Caractéristiques d'une cellule tumorale [17]

1.5.3 Classification histologique

a Classification manuelle

Pour déterminer le grade du cancer du sein, les pathologistes doivent examiner un échantillon de tissu prélevé dans la tumeur par un microscope. Une comparaison est ensuite effectuée entre l'apparence des cellules cancéreuses, avec celle des cellules normales (figure 1.34). Dans ce contexte, afin d'assigner un grade au cancer du sein, le pathologiste doit évaluer les caractéristiques des cellules cancéreuses [20, 21].

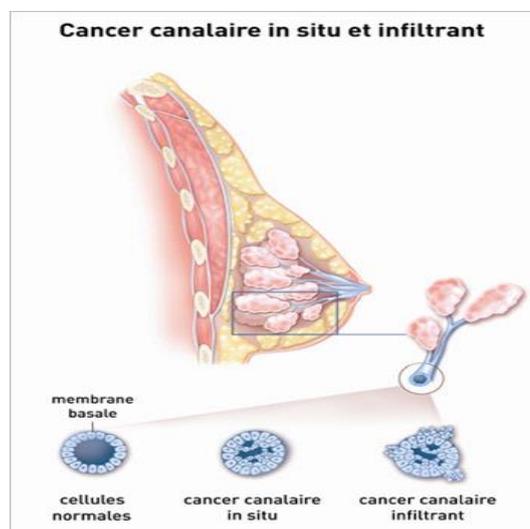


Figure 1.34 : Cellules normales et cancéreuses [21]

Comment les pathologistes déterminent-ils le grade du carcinome canalaire invasif ?

Le classement du cancer du sein est effectué selon le système de notation de Nottingham, également appelé système de Bloom-Richardson modifié, qui est préféré par la plupart des pathologistes, car il donne une évaluation plus objective que les systèmes précédents [17, 20, 21].

Ce système de notation combine trois critères :

1-Pléomorphisme nucléaire

Le pléomorphisme (ou pléomorphe) est un mot, que les pathologistes utilisent lorsqu'il y a une différence entre les noyaux de cellule tumorales.

La notation pour le pléomorphisme nucléaire est de score 1 à 3 :

- **Grade 1** : Il se manifeste lorsqu'il n'y a pas de nécrose dans l'échantillon. Celui-ci présente une grande ressemblance entre les cellules et où la plupart des cellules cancéreuses, sont de petite taille.
- **Grade 2** : Ce grade présente pratiquement la même observation, que le grade 1. On observe dans ce grade, quelques petites régions nécrosées, dans l'échantillon.
- **Grade 3** : Ce grade se manifeste par des cellules cancéreuses de taille élevée et une nécrose dans l'échantillon.

2- formation tubulaire

Un grade de 1 à 3 est attribué en fonction du pourcentage de cellules cancéreuses, formant des tubules.

Dans le grade 1 (bas), les tumeurs sont composées principalement de tubules.

Dans le grade 3 (haut), les tumeurs sont composées de très peu de glandes.

3- nombre de mitoses

Le taux mitotique est la division et la création de nouvelles cellules (mitose). Le pathologiste compte le nombre de figures mitotiques (les cellules qui se divisent) dans une zone spécifique : appelée champ à haute puissance.

- **Dans le grade 1**, on constate très peu de figures mitotiques.
- **Dans le grade 3**, on constate de nombreuses les figures mitotiques.

Lors de l'examen de la tumeur, les pathologistes recherchent les caractéristiques microscopiques représentées par la figure 1.35.

CARCINOMES INFILTRANTS

Facteurs pronostiques et prédictifs Grade histopronostique

Grade reconnu par l'OMS : SBR modifié par Elston et Ellis

A APPLIQUER A TOUS LES TYPES HISTOLOGIQUES

- Différenciation			
Glandes et tubules > 75% de la tumeur	1	Grade I D+A+M : 3 à 5	
Glandes et tubules 10 à 75% de la tumeur	2		
Glandes et tubules < 10% de la tumeur	3		
- Anisocaryose			
Noyaux petits réguliers et uniformes	1	Grade II D+A+M : 6-7	
Augmentation de taille et variation modérées	2		
Variation marquée	3		
- Mitoses (à adapter au Ø du champ x400 : ex=0,63 mm)			
0 à 11/10 CFG	1	Grade III D+A+M : 8-9	
12 à 22	2		
> 23	3		
<i>Place de l'index de prolifération Ki67?</i>			

Figure 1.35: Grades histopronostiques [21]

Le grade global de la patiente est défini en fonction des scores pour la formation de tubules D, du pléomorphisme nucléaire A et des cellules mitotiques M. Ces différents éléments sont additionnés, pour déterminer le niveau de chaque grade G [38, 39]. Ainsi, le grade global est attribué comme suit :

- Grade I (grade faible) : G = 3, 4, 5
- Catégorie II (catégorie intermédiaire) : G = 6, 7
- Grade III (grade élevé) : G = 8, 9.

b Classification automatique

Les pathologistes doivent traiter environ 100 cas de classement par jour. Chacun d'entre eux, est composé d'environ 2000 images qui affectent la précision du classement. Comme il s'agit d'une tâche qui prend beaucoup de temps, un classement automatique et précis, est effectué par un système informatique pour aider les pathologistes dans le diagnostic, le pronostic et réduire la charge de travail [17, 21].

A Partir de micros images histologiques, le système localise le néoplasme et classe les cellules tubulaires et mitotiques formées sur l'image (figure 1.36).

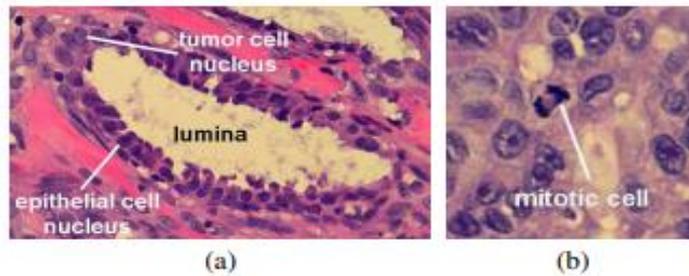


Figure 1.36 : Echantillon de tumeurs tubulaires et mitotiques
 (a) Cellules épithéliales et tumorales tubulaires (b) Cellules mitotiques [21]

Connaître le grade tumoral permet d'aider le clinicien, à établir le pronostic et à prévoir comment le cancer pourrait réagir au traitement. La classification tumorale, indique aussi, la probabilité de propagation du cancer.

1.6 Intelligence artificielle dans la médecine

L'intelligence artificielle (IA) est un processus évolutionniste, né dans les années 1950, qui permet à des machines de simuler une forme d'intelligence réel grâce à l'assemblage de théorie et de techniques vérifier. Cette pratique multidisciplinaire s'applique principalement dans plusieurs domaines d'applications, comme la médecine. Elle intervient dans la recherche fondamentale et clinique, la pratique hospitalière, les examens médicaux, les soins ou encore la logistique. Ce qui contribue à l'affinement des diagnostics. Elle aide les médecins oncologues à prendre des décisions, permettant ainsi, une plus grande précision dans l'imagerie médicale.

Cependant, l'Intelligence Artificielle ne peut pas remplacer les médecins, mais vient compléter sa décision, pour lui permettre de modifier ou confirmer le diagnostic. Le clinicien prendra sa décision, en utilisant ses qualités d'intelligence relationnelle et émotionnelle, aspect de l'intelligence que les machines ne possèdent pas.

L'IA, par les algorithmes, aide donc principalement à l'élaboration de diagnostics. En effet, la machine prescrit le même diagnostic que les médecins dans 99% des cas, et dans 30% des cas, elle propose un traitement plus adapté que celui des spécialistes. Elle réussit à détecter les cancers du sein dans 89% des cas, alors que les spécialistes les détectent dans 73% des cas.

Les machines arrivent également à détecter des crises cardiaques, par téléphone. Ainsi, la robotique étend sa toile dans de nombreux secteurs de la médecine [22, 23].

Les applications sont encore à un stade précoce de la médecine, mais le potentiel est énorme, et ce, dans divers domaines :

- Médecine de précision pour les maladies cardiovasculaires et les cancers.
- Reconnaissance d'images pour le diagnostic des cancers et des maladies rétinienne.
- Conception de médicaments.
- Amélioration de la prestation des soins de santé.

1.7 Conclusion

Ce chapitre, a été consacré à la présentation des différents systèmes d'acquisition utilisés dans la détection du cancer du sein, après avoir décrit l'anatomie du sein, ainsi que les différentes pathologies mammaires. Nous avons introduit les principales formes du cancer du sein et les différents stades de ce dernier.

Après une simple présentation sur l'intelligence artificielle dans le domaine médical, le second chapitre, sera consacré à l'apprentissage automatique et l'apprentissage profond (deep Learning).

2.1 Introduction

Le machine Learning et le deep learning, sont des termes parfois confondus, alors qu'ils ont chacun une signification bien distincte. Si ces termes sont confondus, c'est que même s'ils ne signifient pas exactement la même chose, ils restent très étroitement liés. En réalité, ils ne sont que des différents niveaux d'une même discipline mère, qui est l'intelligence artificielle (figure 2.1). Aujourd'hui le machine learning et le deep learning sont en train de révolutionner le monde entier : comme dans le domaine de l'industrie des transports, évidemment avec les voitures autonomes, la finance, l'industrie agricole, l'industrie aérospatiale, l'industrie bio mécanique, mais également le monde de la santé, parce qu'aujourd'hui le machine learning est capable de diagnostiquer des millions de cancer tous les ans [24, 25].

Dans ce chapitre les divers concepts et les méthodes d'apprentissage automatique et profond, sont définis, surtout les réseaux neuronaux profonds utilisés dans notre étude, comme les réseaux convolutifs.

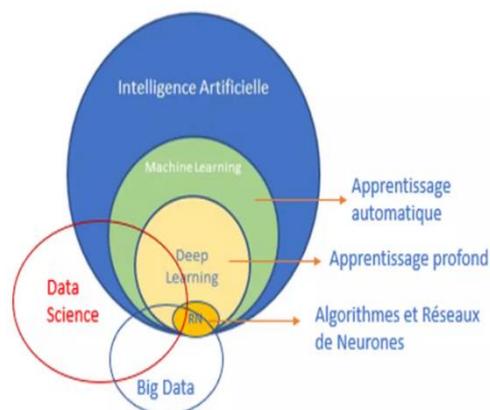


Figure 2.1 : Interactions entre l'intelligence artificielle, le machine learning et le Deep Learning [24]

2.2 Introduction à l'intelligence artificielle

L'intelligence artificielle (IA) est un processus évolutionniste, né dans les années 1950, qui permet à des machines, de simuler une forme d'intelligence réelle en se basant sur le traitement de connaissances et sur le raisonnement.

L'intelligence artificielle peut être divisée en deux grands types : l'**IA faible** et l'**IA forte** [26, 27].

2.2.1 Intelligence artificielle faible

Également appelée intelligence artificielle étroite ou Narrow AI, l'intelligence artificielle faible, est utilisée dans le but de réaliser une tâche simple.

2.2.2 Intelligence artificielle forte

Dans l'optique de ressentir de réels sentiments, l'IA dite forte, consiste à faire comprendre à une machine, ce qui la mener à effectuer certaines actions. De cette manière, le système reproduit de façon identique le fonctionnement du système cognitif humain. Ainsi bien que privés de conscience qui reste quant à elle, un propre organisme vivant, la machine acquiert de l'expérience lui permettant de modifier son propre fonctionnement.

2.3 Apprentissage automatique

Issu de l'intelligence artificielle, l'apprentissage automatique (machine learning) est un concept qui tend à rendre une machine, capable d'apprendre de ses expériences basées sur l'analyse de patterns, dans le but de faire de la prédiction [25, 28, 29].

2.3.1 Types d'apprentissage automatique

Il existe différents types d'apprentissage automatique : le supervisé, le semi-supervisé, le non-supervisé et celui par renforcement (figure 2.2).

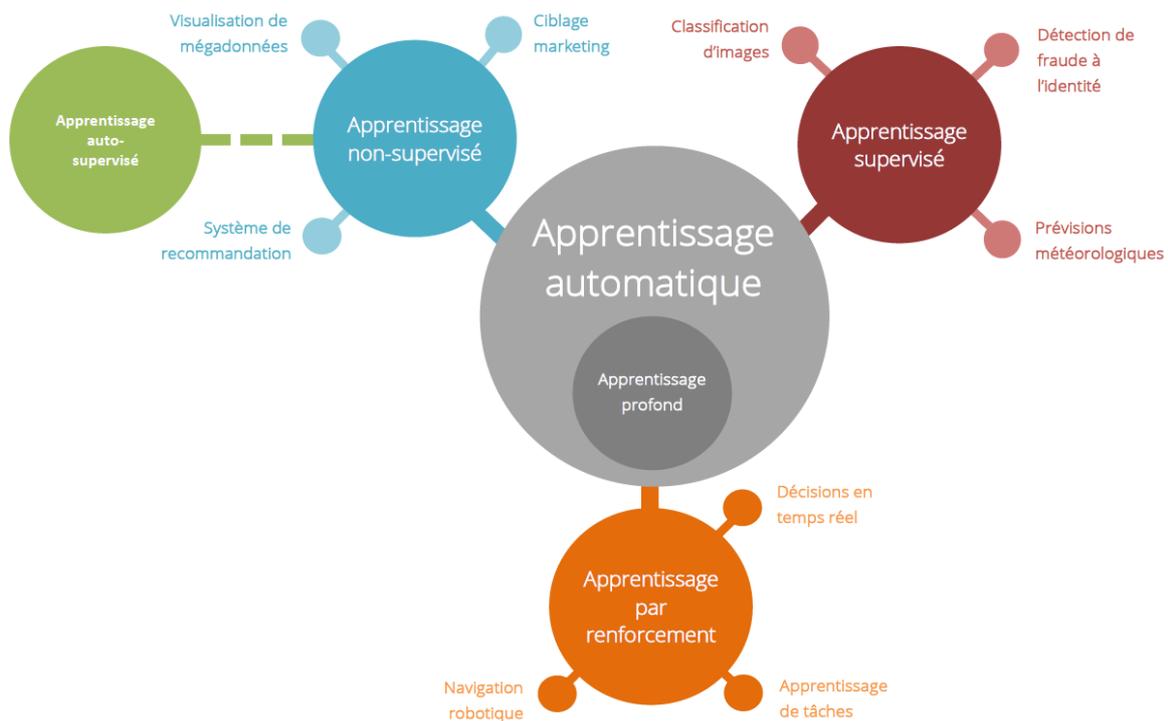


Figure 2.2 : Types d'apprentissage automatique [28]

a Apprentissage supervisé

Cet apprentissage utilise des images étiquetées, à partir desquelles l'algorithme doit générer une fonction qui délivre le résultat désiré (figure 2.3).

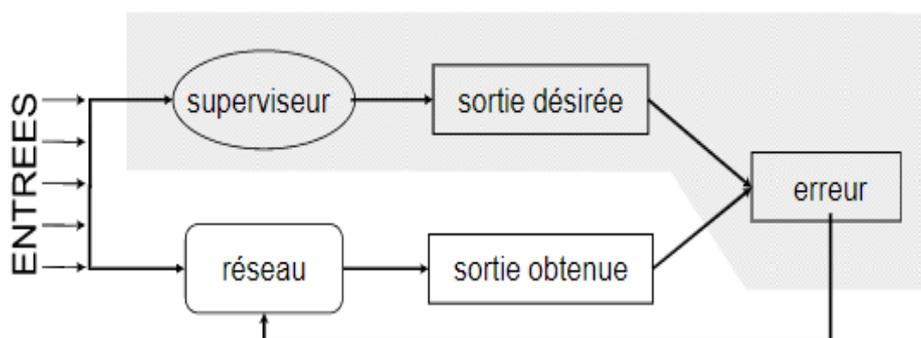


Figure 2.3 : Schéma d'apprentissage supervisé [28]

b Apprentissage semi supervisé

Cette catégorie utilise les données non-annotées, pour compléter l'apprentissage supervisé. Il consiste à fournir de nombreux exemples au système, ainsi que la bonne réponse pour certains d'entre eux (figures 2.4 et 2.5).

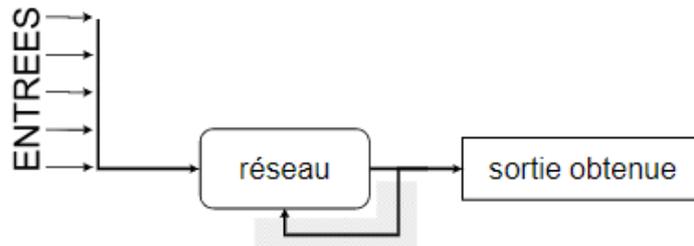


Figure 2.4 : Apprentissage automatique non-supervisé [28]

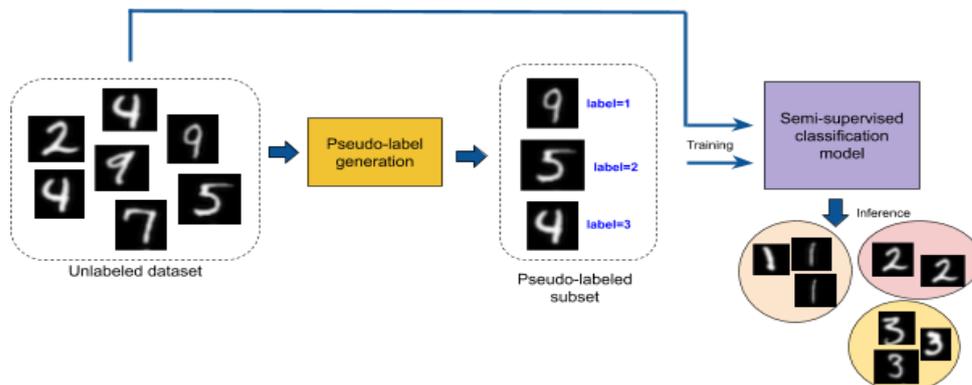


Figure 2.5 : Exemple d'apprentissage automatique semi-supervisé [29]

c Apprentissage non supervisé

Cet apprentissage utilise une approche plus indépendante, dans laquelle un ordinateur apprend à identifier des processus et des schémas complexes, sans un quelconque guidage humain constant et rigoureux. L'apprentissage non-supervisé implique une formation basée sur des données sans étiquette, ni résultat spécifique défini.

d Apprentissage par renforcement

Le jeu de données d'apprentissage par renforcement (figure 2.6) est similaire aux jeux de données d'apprentissage supervisé. Cependant, l'idée de l'apprentissage par renforcement est d'apprendre à un agent, comment se comporter dans un environnement, en utilisant un système de récompenses pour un bon choix ou de pénalité pour une mauvaise action.

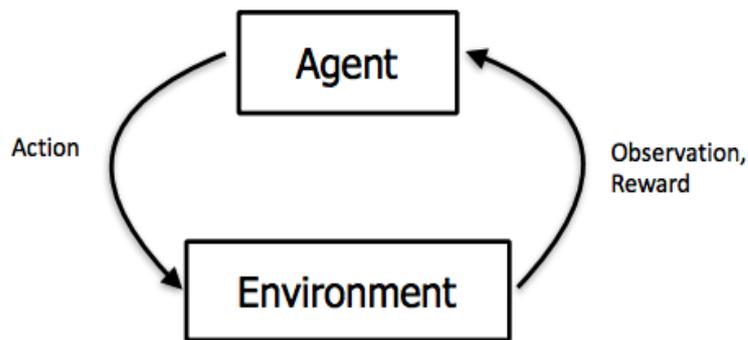


Figure 2.6 : Apprentissage automatique par renforcement [28]

L’algorithme apprend quelle stratégie (ou choix d’actions), maximise le cumul de récompenses.

Ce type d’apprentissage est souvent utilisé dans le cadre de la robotique, de la théorie des jeux et des véhicules autonomes [30, 31].

L’apprentissage supervisé, peut être réalisé par plusieurs méthodes, telles que les arbres de décision, les machines à vecteurs de support, ou le perceptron multicouche, [31].

On retrouve dans l’apprentissage non supervisé, les méthodes du clustering. Dans notre cas, ce sont les réseaux de neurones qui nous intéressent.

2.4 Réseaux de neurones

2.4.1 Neurones biologiques

Dans l’Antiquité, il n’était pas évident que le cerveau constitue l’organe de la pensée. Ainsi pour le philosophe Ariston de Chios : « le cœur est au centre des processus sensitifs, le cerveau ayant alors un simple rôle de réfrigération de ce dernier. » [31].

A la fin du 19^{ème} siècle, le neuroscientifique espagnol Santiago Ramon y Cajal [32], met en évidence que les cellules du cerveau, sont des entités fonctionnelles autonomes connectées entre elles.

En 1891, l’anatomiste allemand Heinrich Wilhelm Waldeyer [33], confirme la thèse de Cajal et invente le mot « neurone », ainsi que le mot « chromosome ».

Dans ce contexte, le médecin et scientifique britannique Charles Scott Sherrington propose le terme « synapse » en 1987 ; qui correspond à un point de contact entre un neurone et un autre neurone [33]. Ainsi, un neurone (figure 2.7) peut comporter jusqu'à 7000 dendrites. Le nombre de neurones estimés dans le cerveau humain, est entre 90 et 100 milliards.

Le neurone reçoit des signaux via les dendrites et, émet un signal via l'axone. La communication entre les neurones, se fait via les synapses. Cela ressemble à une fonction mathématique à plusieurs variables $f(x_1, x_2, \dots, x_n)$.

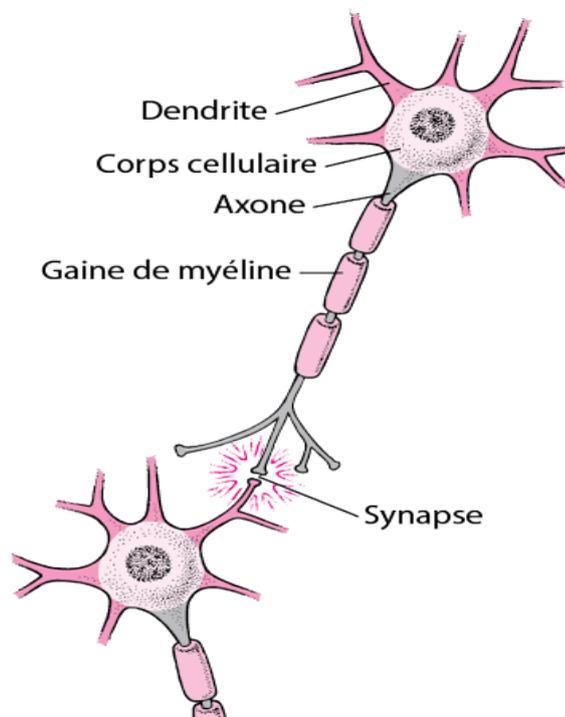


Figure 2.7 : Structure du neurone [34]

2.4.2 Premiers neurones artificiels

L'américain et chercheur neurologue Warren McCulloch, est le créateur du premier modèle de neurone formel avec le mathématicien Walter Harry pitts [34, 35].

En 1943, c'est l'année de la création du premier neurone artificiel (figure 2.8). C'est un modèle mathématique simplifié d'un neurone biologique. Il comporte un certain nombre d'entrées e_i et, des poids w_i appelés coefficients synaptiques. Il effectue une sommation y pondérée des entrées via les poids suivant l'équation 2.1.

$$y = w_0 e_0 + w_1 e_1 + w_2 e_2$$

2.1

Il déclenche un signal de sortie en fonction d'un certain seuil (fonction seuil).

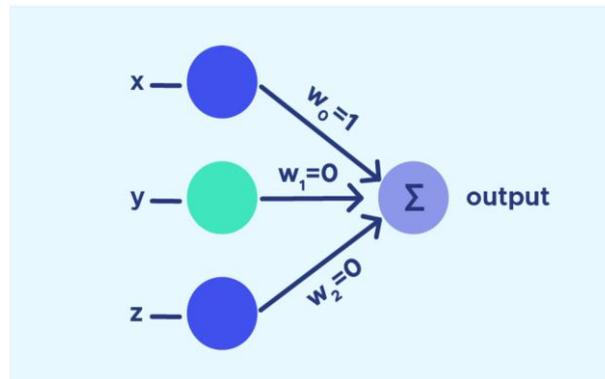


Figure 2.8 : Neurone artificiel simple [35]

2.4.3 Principe

Ces réseaux (figure 2.9) sont inspirés du fonctionnement des neurones biologiques. Ce système algorithmique est de conception schématique (constitués de plusieurs neurones artificiels connectés entre eux) qui se rapproche des méthodes statistiques [31, 35, 36].

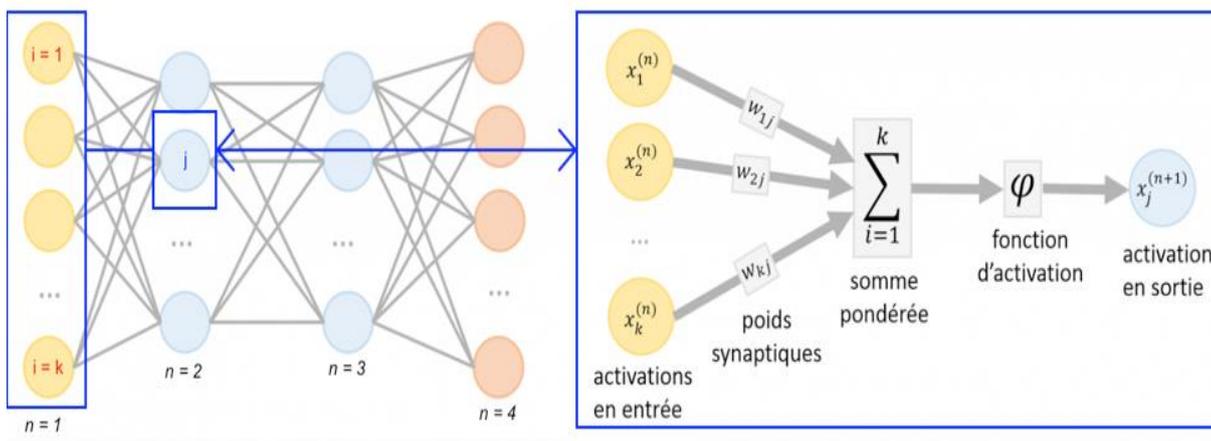


Figure 2.9 : Structure des réseaux de neurones [31]

Le neurone dit formel est conçu comme un automate doté d'une fonction de transfert, qui transforme ses entrées en sortie selon des règles logiques, arithmétiques et symboliques précises.

Assemblées en réseau, ces neurones formels ont la capacité d'opérer rapidement des classifications et d'apprendre progressivement à les améliorer.

Le neurone calcule la somme de ses entrées x pondérées par les poids w , avant d'appliquer la fonction d'activation ϕ .

$$O_j = \phi(b_j + \sum_{i=1}^n w_{ij} x_i) \quad 2.2$$

Où b_j : est le biais.

Le neurone est caractérisé par :

- Son état de sortie **O_j** (binaire, discret, continu).
- Le poids des connections en entrée **w**.
- Le niveau d'activation reçu en entrée (continu).

2.4.4 Fonctions d'activation

Le neurone formel, l'unité élémentaire d'un réseau, se compose de deux parties comme le montre le schéma ci-dessous :

- La fonction E permet l'évaluation du signal reçu.
- La fonction f responsable de l'activation, définit l'état interne.

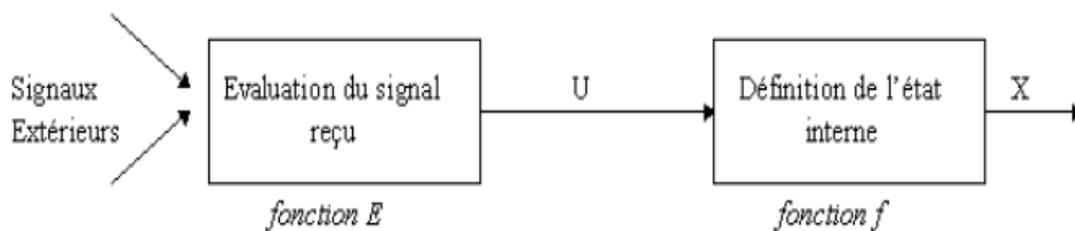


Figure 2.10 : Structure du neurone formel [37]

La **fonction d'activation** (aussi appelée seuil d'activation), est une fonction mathématique appliquée à un signal en sortie d'un neurone artificiel. Le terme de "fonction d'activation" vient de l'équivalent biologique "potentiel d'activation", un seuil de stimulation qui, une fois atteint entraîne une réponse du neurone. La fonction d'activation est souvent une fonction non linéaire [37].

Un neurone ne va faire qu'appliquer la fonction suivante :

$$Y = \sum (\text{entrée} * \text{poids}) + \text{biais} \quad 2.3$$

C'est sur cette sortie, que la fonction d'activation va s'appliquer. On distingue différents types de fonctions d'activation [38, 39, 40].

a Fonction ReLU

La fonction Relu (Unité linéaire rectifiée) est la fonction d'activation la plus couramment utilisée, dans les réseaux de neurones convolutifs. Elle est de la forme :

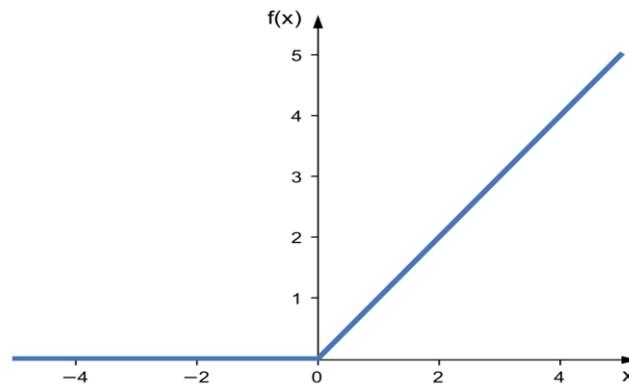


Figure 2.11 : Fonction ReLU [38]

L'expression mathématique est la suivante :

$$y_{out} = f(y_{sum}) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad 2.4$$

Cela signifie que f(x) est égale à zéro lorsque x est inférieur à zéro et f(x), est égale à x lorsque x est supérieur ou égal à zéro. Cette fonction est dérivable, sauf en un seul point x = 0. En ce sens, la dérivée d'une ReLU est en fait une sous-dérivée.

b Tangente hyperbolique (Tanh)

C'est une fonction d'activation largement adoptée (figure 2.11), pour un type spécial de réseau de neurones, appelé réseau de rétropropagation. Elle est de nature bipolaire [38].

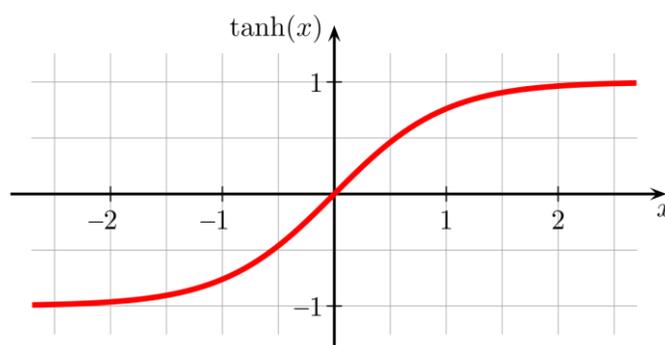


Figure 2.12 : Fonction tangente hyperbolique [38]

L'expression mathématique est la suivante :

$$y_{out} = f(x) \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad 2.5$$

Elle est similaire à la fonction sigmoïde bipolaire.

c Softmax

La fonction Softmax permet de transformer un vecteur réel, en un vecteur de probabilité [59]. On l'utilise souvent, dans la couche finale d'un modèle de classification, notamment pour les problèmes multiclasse.

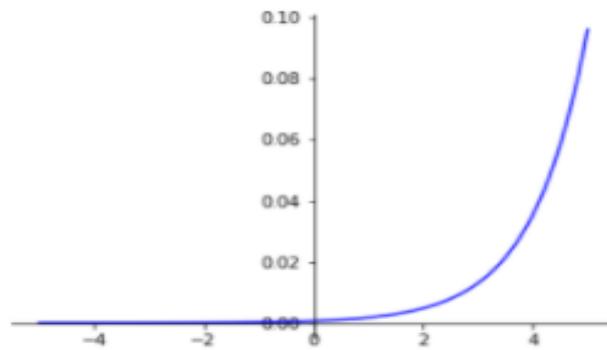


Figure 2.13 : Fonction softmax [38]

Dans la fonction Softmax (2.6), chaque vecteur est traité indépendamment.

$$\text{Softmax}(z_i) = \frac{e^{z_i}}{\sum e^{z_i}} \quad 2.6$$

Où z_i c'est le vecteur d'entrée :

d Logistique(sigmoïde)

C'est la fonction d'activation la plus couramment utilisée, dans les réseaux de neurones. Le besoin d'une fonction sigmoïde provient du fait que de nombreux algorithmes d'apprentissage, nécessitent que la fonction d'activation soit dérivable et donc continue.

Il existe deux types de fonction sigmoïde :

➤ **Fonction sigmoïde binaire**

Sa forme est représentée par la figure 2.14. Elle est indiquée par l'équation 2.7.

$$y_{out} = f(x) = \frac{1}{1+e^{-kx}} \quad 2.7$$

Où k est le paramètre de pente. Ainsi, en faisant varier la valeur de k, une fonction sigmoïde avec différentes pentes peut être obtenue. La pente d'origine est k/4. Lorsque la valeur de k devient très grande, la fonction sigmoïde devient une fonction de seuil [39].

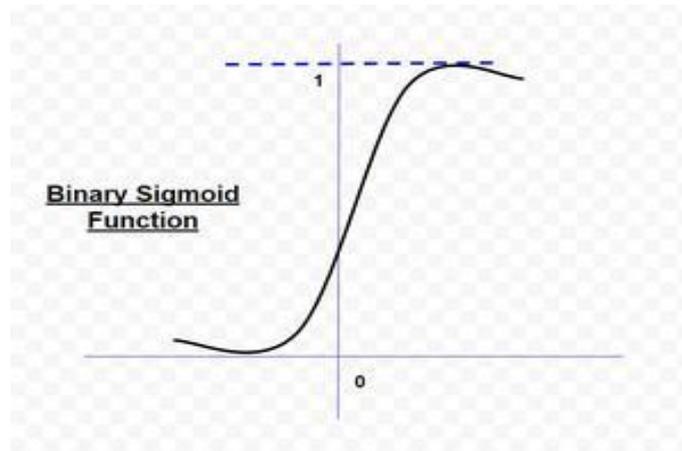


Figure 2.14 : Fonction sigmoïde binaire [38]

➤ **Fonction sigmoïde bipolaire**

Elle est représentée par la courbe de la figure 2.15. La fonction sigmoïde bipolaire est donnée par l'équation 2.8.

$$y_{out} = \frac{1-e^{-kx}}{1+e^{-kx}} \quad 2.8$$

La plage de valeurs des fonctions sigmoïdes, peut varier en fonction de l'application. Cependant, la plage de (-1, +1) est la plus couramment adoptée [40].

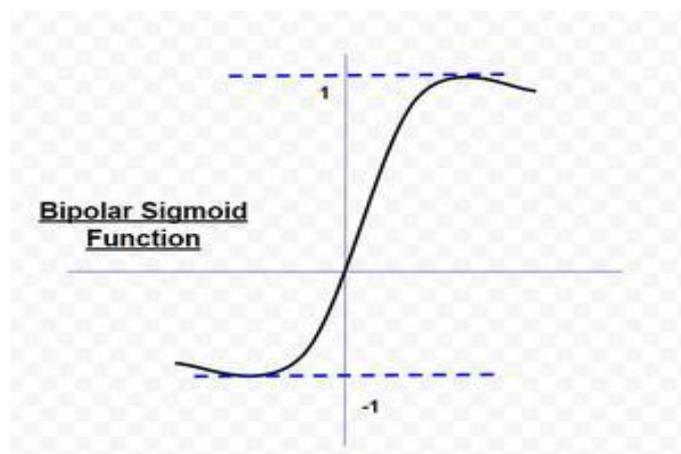


Figure 2.15 : Fonction sigmoïde bipolaire [38]

e Fonction d'identité

C'est une fonction linéaire ayant la forme suivante :

$$y_{out} = f(x) = x, \forall x \quad 2.9$$

La sortie reste la même que l'entrée. Elle est utilisée comme fonction d'activation pour la couche d'entrée.

f Fonction échelon

C'est la fonction la plus utilisée (figure 2.16). Elle est égale à 1 quand l'entrée est soit 0, soit positive comme le montre la figure 2.16. Si l'entrée est négative, cela donne 0 en sortie [60].

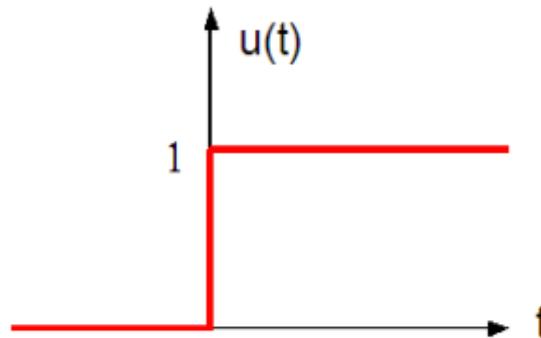


Figure 2.16 : Fonction échelon [38]

L'expression mathématique est la suivante :

$$y_{out} = f(y_{sum}) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad 2.10$$

g Fonction seuil

La fonction de seuil (figure 2.17) ressemble presque à la fonction échelon. La seule différence est que le paramètre θ est utilisé, comme valeur de seuil au lieu de 0 [39].

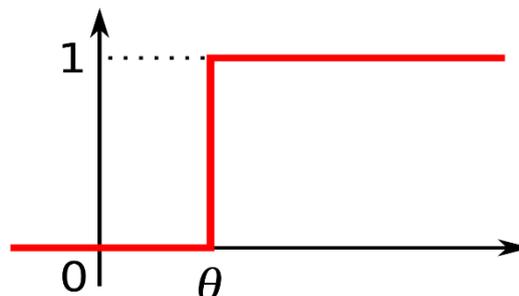


Figure 2.17 : Fonction seuil [38]

L'expression mathématique est la suivante :

$$y_{out} = f(y_{sum}) = \begin{cases} 1, x \geq \theta \\ 0, x < \theta \end{cases} \quad 2.11$$

2.4.5 Perceptron monocouche

Il s'agit d'un **modèle de classification binaire**, capable de séparer linéairement 2 classes. Le perceptron est le premier réseau de neurones artificiels évolutif, c'est-à-dire capable d'apprentissage (figure 2.18). En 1957, le psychologue américain Frank Rosenblatt un des pionniers de l'IA, réalise le perceptron, une machine capable d'apprentissage [31, 41, 42].

Le perceptron avait pour but de reconnaître des lettres de l'alphabet, avec des cellules photo-électriques comme capteur.

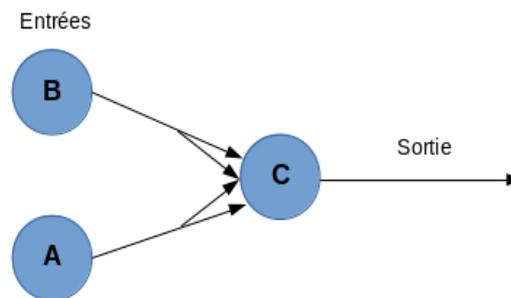


Figure 2.18 : Perceptron simple [42]

Comme on peut voir dans le schéma ci-dessous, le perceptron est un réseau à propagation avant, composé de neurones à seuil seulement, avec deux couches (entrée et sortie) entièrement interconnectées.

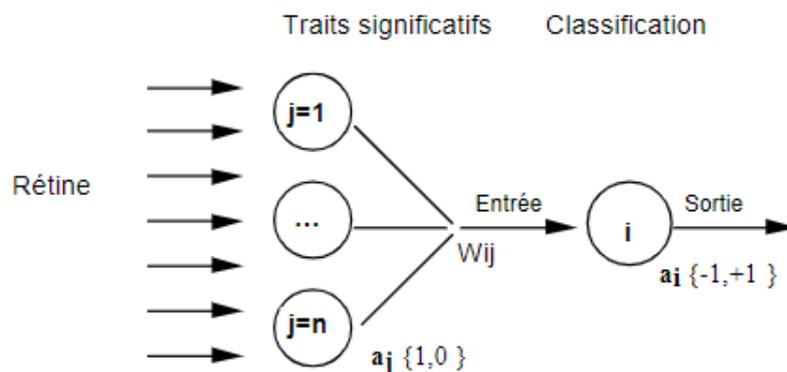


Figure 2.19 : Schéma détaillé d'un perceptron [42]

Les interconnexions ne sont autorisées qu'entre deux couches voisines. Le réseau est de type feed forward, c'est-à-dire que les signaux de traitement se propagent de l'entrée vers la sortie [42].

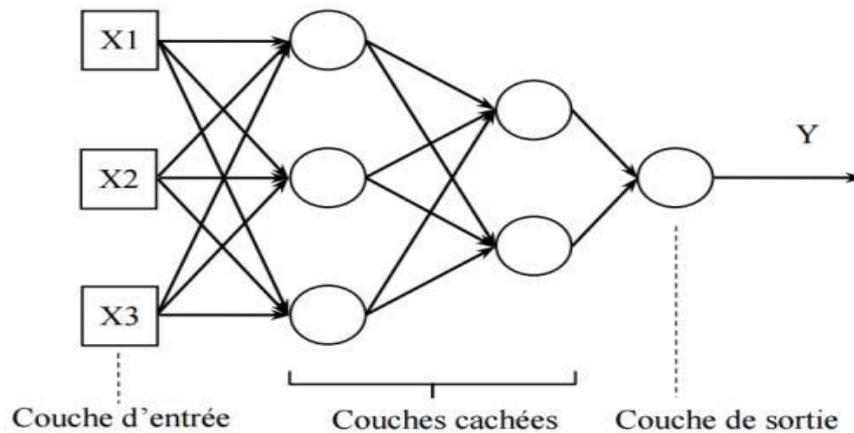


Figure 2.22 : Modèle du perceptron multicouche [42]

2.5 Apprentissage profond

L'apprentissage profond (deep learning) est une technologie d'intelligence artificielle, qui correspond à une assimilation automatique d'informations, par une machine en temps réel. Les données utiles sont identifiées et intégrées dans un processus d'apprentissage profond, sans aucune intervention humaine. Il s'inspire du cerveau humain et fonctionne avec un réseau de neurones artificiels assemblé en plusieurs couches. A travers ce système, le processus permet ainsi de reconnaître des lettres dans un texte ou visualiser un objet sur une image. La technique est basée sur des méthodes spécifiques, où un nombre de données initiales est à établir, pour que le système accumule les expériences et que grâce à ce point de départ, il devient performant. Cette technologie [43] est aujourd'hui utilisée dans un certain nombre de domaines comme :

- La reconnaissance de formes ;
- L'assistance vocale ;
- Les services de traduction automatique ;
- La rédaction de texte ;
- La cybersécurité ;
- Les transactions financières ;
- La conduite autonome ;
- Les diagnostics médicaux.

La notion d'apprentissage profond est une traduction directe du terme anglais « Deep Learning », que certains préfèrent traduire par la notion d'apprentissage statistique. C'est un nouveau concept qui a émergé depuis les années 2000 et formalisé en 2007 à partir des nouvelles architectures de réseaux de neurones sur lequel se base le Deep Learning. En 1943, Warren McCulloch et Walter Pitts publient leur premier modèle mathématique et informatique du neurone biologique, appelé le neurone formel. Ils proposent le premier neurone artificiel, qui est un modèle très simplifié du neurone biologique. Ce neurone artificiel possède une ou plusieurs entrées et une sortie binaire. Son fonctionnement est simple, le neurone active sa sortie (sortie active = 1) selon ses entrées, si elles dépassent un certain seuil.

À partir de ce neurone, il est possible de construire n'importe quel réseau de neurone artificiel et même, des réseaux contenant des milliards de neurones, capables de résoudre des opérations logiques, ainsi que des problèmes extrêmement complexes. La révolution du Deep Learning est liée à la puissance des ordinateurs, qui ne cesse d'augmenter permettant la création de réseaux neuronaux, avec des dizaines de couches cachées [44].

2.5.1 Différence entre l'apprentissage automatique et l'apprentissage profond

Dans l'apprentissage machine (automatique) un programmeur doit effectuer une extraction de caractéristiques numériques comme, la surface, le diamètre, le volume, etc. avant la classification. Dans l'apprentissage profond il suffit juste d'avoir des données gigantesques, pour apprendre des modèles (figure 2.23).

Les approches basées sur cet apprentissage, permettent l'extraction automatique des caractéristiques, dans les différentes couches du réseau, pour être apprises par le modèle de lui-même [45].

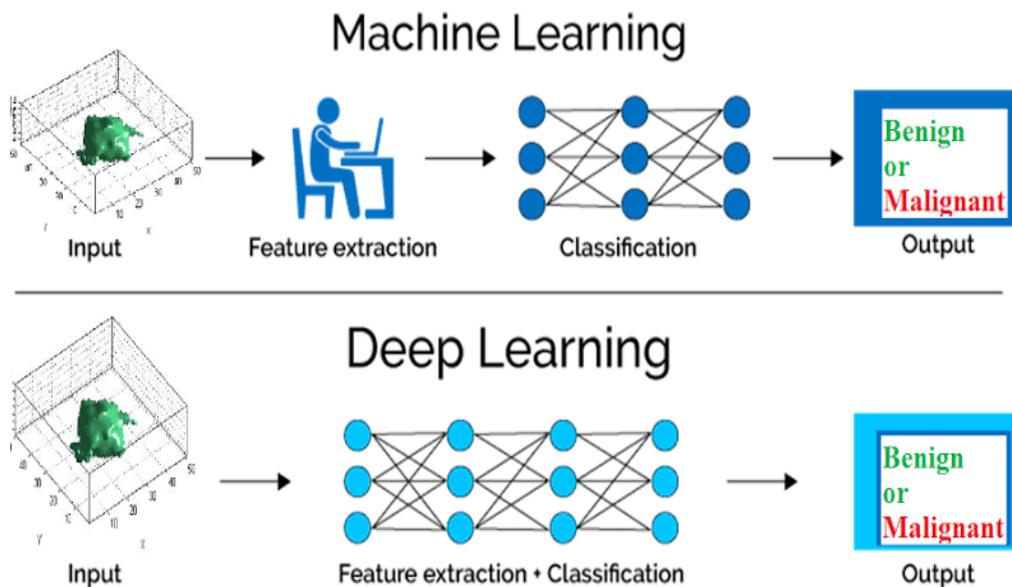


Figure 2.23: Machine Learning et Deep Learning [45]

2.5.2 Réseaux de neurones convolutifs

Les réseaux de neurones convolutifs profonds (CNN convolutional neural networks) sont devenus ces dernières années, l'une des innovations les plus importantes dans le domaine de l'apprentissage automatique et l'apprentissage profond. Ils peuvent extraire directement des caractéristiques représentatives des images, à partir de couches de convolution et les classifier [46, 47, 48, 49].

Un réseau neuronal convolutif se compose de trois couches principales :

- 1- Les couches convolutionnelles.
- 2- Les couches de mise en commun.
- 3- Les couches entièrement connectées.

L'architecture suivante montre comment les entrées, sont transmises aux trois couches du réseau neuronal convolutif.

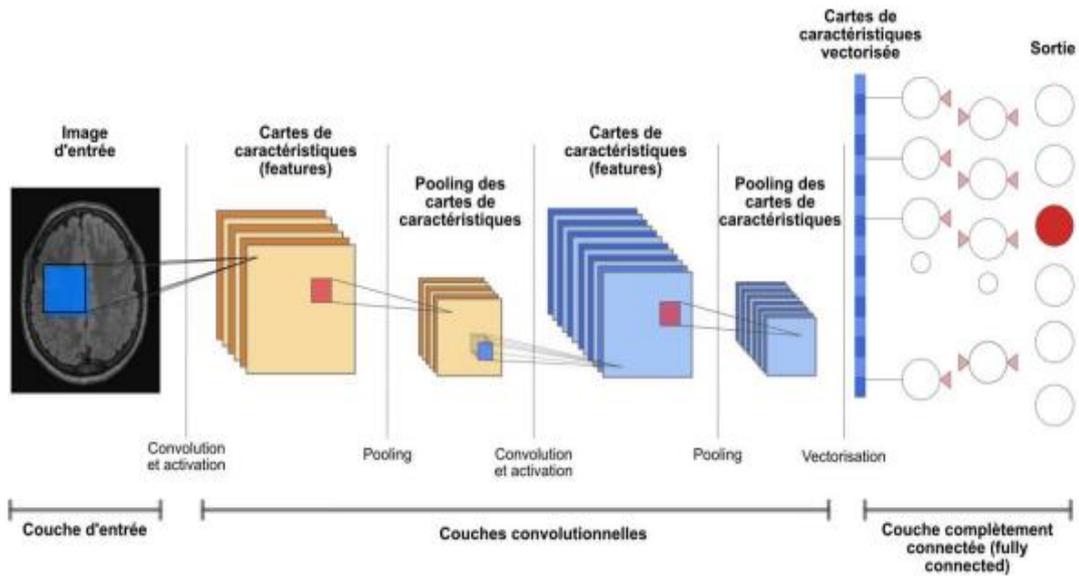


Figure 2.24 : Architecture d'un CNN [49]

a Couche convolutionnelle

La couche convolutionnelle est la première couche d'un CNN. Cette couche prend une image en entrée sur laquelle une série d'opérations de convolution, est effectuée. L'objectif des couches convolutionnelles est d'extraire des motifs et des informations d'une image [47, 49, 50].

Les couches d'un CNN sont rangées en 3 dimensions (largeur, hauteur, profondeur), où la profondeur est liée aux types de couleurs de l'image, en entrée comme le montre la figure 2.25.

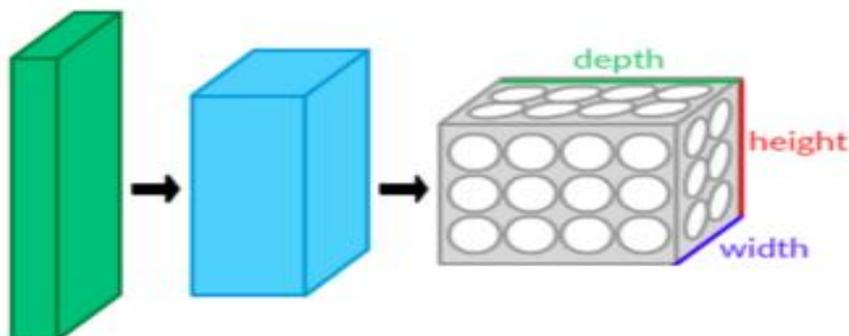


Figure 2.25 : Architecture de couche de convolution [47]

Ce type de couche est souvent placé entre deux couches de convolution : elle reçoit en entrée plusieurs 'feature maps' et applique à chacune d'entre elles, l'opération de pooling [49, 50]. L'opération de pooling la plus utilisée, est celle du maxpooling.

- **Max pooling**

L'opération de max pooling (figure 2.27) prend en entrée les cartes des caractéristiques (feature-maps) pour en extraire la valeur max. Elle maximise le poids des activations et garde seulement les informations importantes.

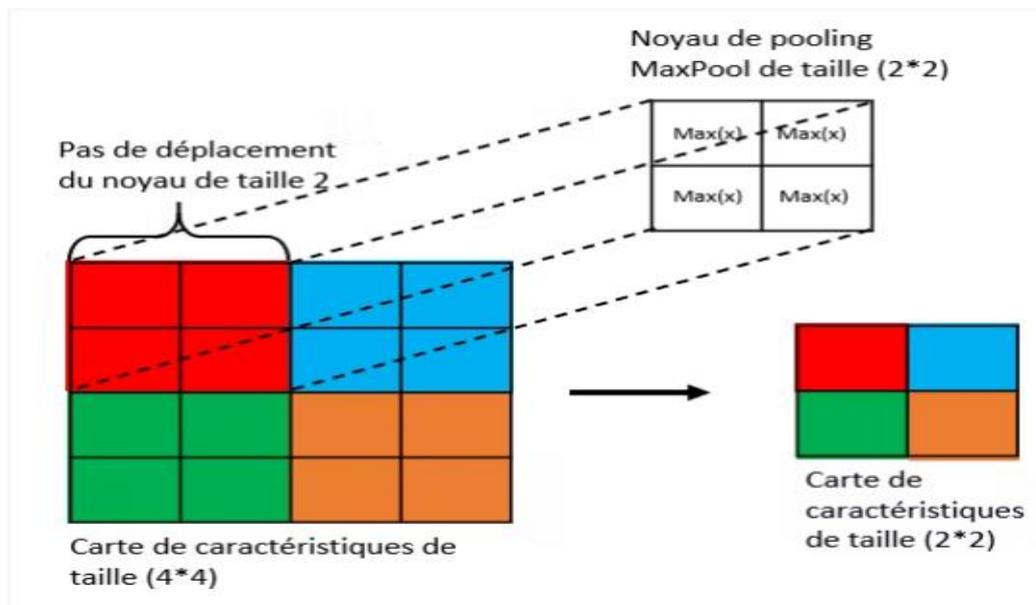


Figure 2.27 : Opération de max pooling [49]

- **Average pooling**

Le « mean pooling » (ou average pooling), représente la moyenne des pixels de la sélection : on calcule la somme de toutes les valeurs et on divise par le nombre de valeurs. On obtient ainsi une valeur intermédiaire pour représenter un lot de pixels.

- **Sum pooling**

Le « sum pooling », représente la somme des pixels de la sélection.

La couche du pooling a deux paramètres :

- Un noyau de taille (2, 2) ;

- Un pas de déplacement pris généralement à 2.

La couche de *pooling* permet de réduire le nombre de paramètres et le coût de calculs dans le réseau. On améliore ainsi l'efficacité du réseau et on évite le sur-apprentissage [49].

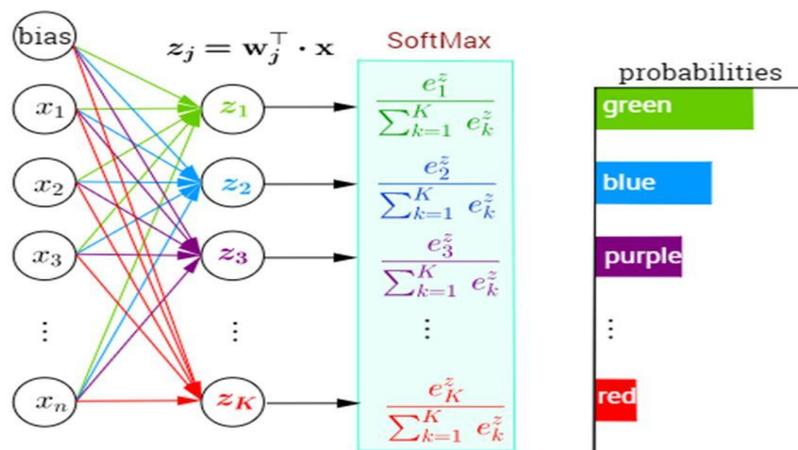
c Couche Flatten

C'est la couche d'aplatissement qui sert à transformer une matrice de dimensions, $n \times n$ en un vecteur de $n \times 1$.

d Couche entièrement connectées (fully connected)

La couche entièrement connectée est responsable de l'entrée d'un vecteur d'image aplati (tenseur d'image 1-D) et de la détermination d'un score de probabilité, pour chaque étiquette dans l'ensemble de données d'apprentissage.

La dernière couche entièrement connectée (figure 2.28), est responsable de la classification de l'image en fonction de la tâche à accomplir, appelée aussi couche de classification.



44

Figure 2.28 : Couche entièrement connectée [49]

e Dropout

Le dropout est une méthode de régularisation, qui permet d'approximer la formation d'un grand nombre de réseaux neuronaux avec différentes architectures en parallèle [50].

2.5.3 Optimiseurs

Les optimiseurs sont utilisés pour minimiser l'erreur par l'ajustement du poids. Dans ce cadre, la descente du gradient, est l'un des algorithmes les plus importants de tout le machine et le deep Learning. Il permet de trouver le minimum de n'importe quelle fonction convexe. Il est utilisé dans les problèmes d'apprentissage supervisé, pour trouver le minimum de la fonction coût. Il existe beaucoup de variantes de cet algorithme : Stochastic Gradient Descent, ADAM, RMSProp [49, 50, 51].

a Descente du gradient stochastique

C'est une méthode de descente du gradient (stochastic gradient descent) itérative. C'est l'algorithme d'optimisation le plus simple, qui permet de trouver des paramètres minimisant la fonction coût donnée.

b ADAM

Cette fonction est utilisée pour la formation de modèles d'apprentissage profond. Il s'agit d'une extension de la descente du gradient stochastique. Dans cet algorithme d'optimisation, les moyennes courantes des gradients et des seconds moments des gradients sont utilisées. La fonction permet de calculer les taux d'apprentissage adaptatifs, pour chaque paramètre [50].

c RMSProp

RMSprop est un algorithme d'optimisation qui accélère la descente du gradient. Il est utilisé pour entraîner des modèles à base de réseaux de neurones profonds.

d Rétropropagation

La rétropropagation du gradient de l'erreur (ou backpropagation) est un algorithme d'optimisation, permettant d'ajuster les paramètres d'un réseau de neurones multicouches, pour mettre en correspondance des entrées et des sorties référencées dans une base d'apprentissage.

La rétropropagation permet de calculer le gradient de l'erreur pour chaque neurone, de la dernière couche vers la première (figure 2.29). Cela permet de corriger les erreurs, selon l'importance des éléments qui ont justement participé à la réalisation de ces erreurs. Ainsi, les poids synaptiques qui contribuent à engendrer une erreur importante, se verront modifiés de manière plus significative, que les poids qui ont engendré une erreur marginale.

Les procédures d'optimisation finissent par aboutir à une configuration stable, généralement un extremum local, au sein du réseau de neurones [51].

Retropropagation

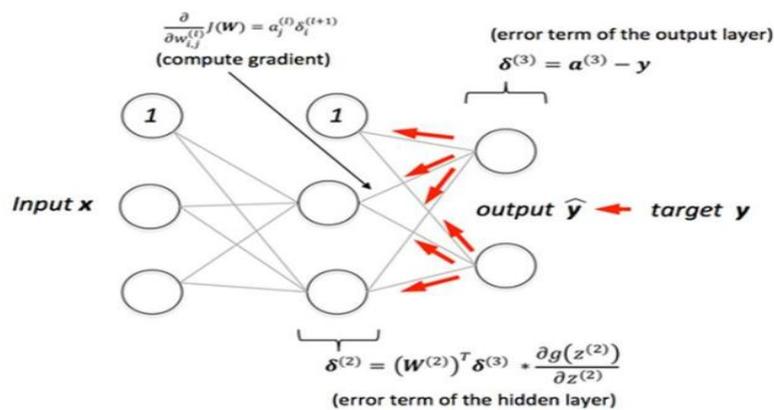


Figure 2.29 : Rétropropagation [51]

2.5.4 Réseaux neuronaux récurrents (RNN)

Le réseau neuronal récurrent (RNN) est une autre classe des réseaux neuronaux artificiels, qui utilisent une alimentation séquentielle des données. Les RNN ont été développés pour résoudre le problème des séries temporelles de données d'entrée séquentielles. L'entrée du RNN (figure 2.30) se compose de l'entrée actuelle et des échantillons précédents. Par conséquent, les connexions entre les nœuds forment un graphe, dirigé le long d'une séquence temporelle. En outre, chaque neurone d'un RNN possède une mémoire interne qui conserve les informations du calcul des échantillons précédents [49, 52].

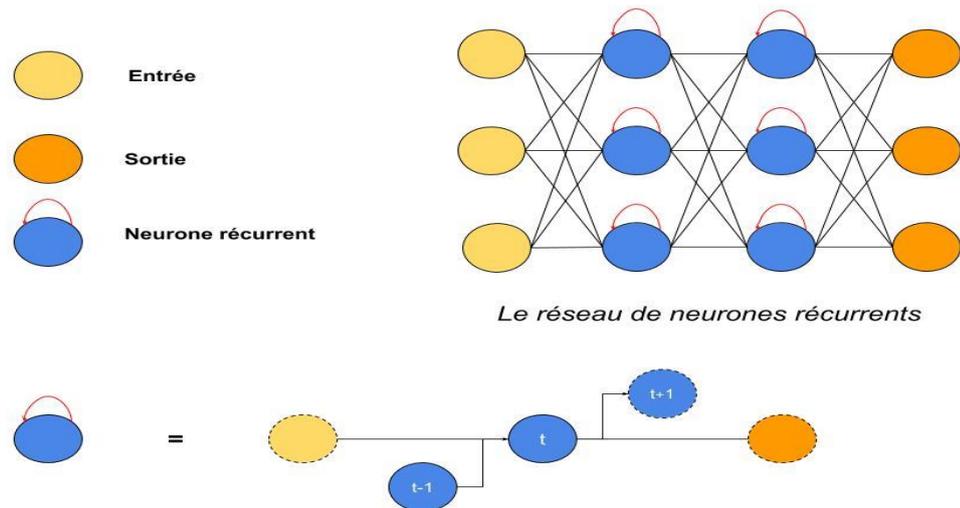


Figure 2.30 : Architecture d'un RNN [52]

2.6 Apprentissage par transfert

L'apprentissage par transfert consiste à mutualiser des données et savoirs, afin d'améliorer la qualité des réseaux de neurones artificiels. Le deep Learning nécessite en effet, un nombre important de modèles et d'algorithmes pour l'apprentissage automatique (figure 2.31).

Il permet d'élaborer des techniques permettant de transférer des connaissances provenant d'une ou de plusieurs tâches, afin d'améliorer l'apprentissage dans l'exécution d'une tâche cible similaire. L'objectif de ce transfert de stratégies d'apprentissage, est de faire évoluer l'apprentissage automatique afin qu'il soit aussi efficace que l'apprentissage humain [53, 54, 55].

L'apprentissage par transfert permet alors, de réexploiter des connaissances acquises dans d'autres configurations (sources) pour la résolution d'un problème particulier (cible).

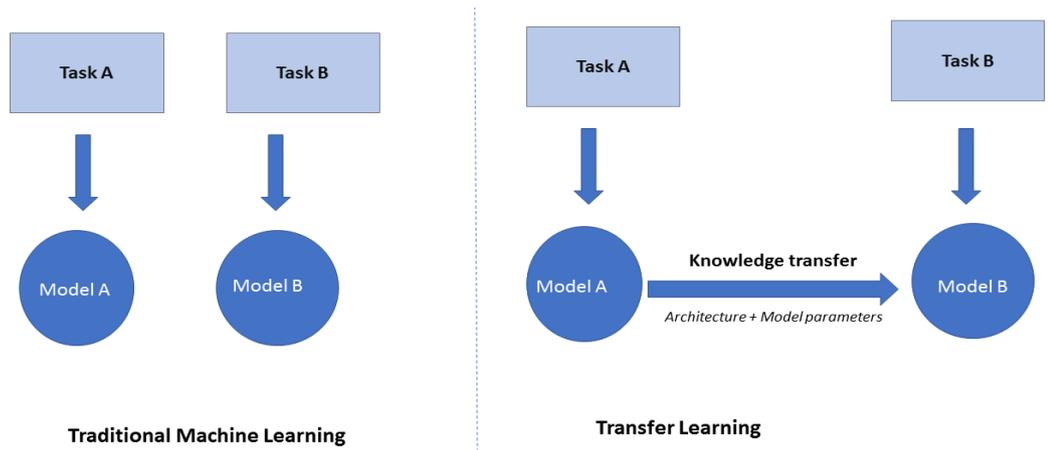


Figure 2.31 : Machine Learning traditionnel et le transfer Learning [55]

2.7 Présentation de quelques modèles des CNN

Les modèles d'apprentissage automatique CNN sont très populaires pour les tâches de vision par ordinateur. Ils peuvent capturer la représentation de haut niveau des données d'entrée, en appliquant divers filtres convolutifs [56, 57]. Dans cette section, sont présentés brièvement quelques modèles.

2.7.1 Modèle Alexnet

Le modèle Alexnet est pour la classification d'images, le premier réseau neuronal CNN à avoir remporté le défi ImageNet en 2012.

AlexNet (figure 2.32) se compose de cinq couches de convolution, de deux couches entièrement connectées suivies d'une couche softmax pour les prédictions. Ainsi, AlexNet a besoin de 61 millions de poids et de 724 millions d'opérations d'addition et de multiplication (OAM), pour classifier l'image d'une taille de 227×227 [56].

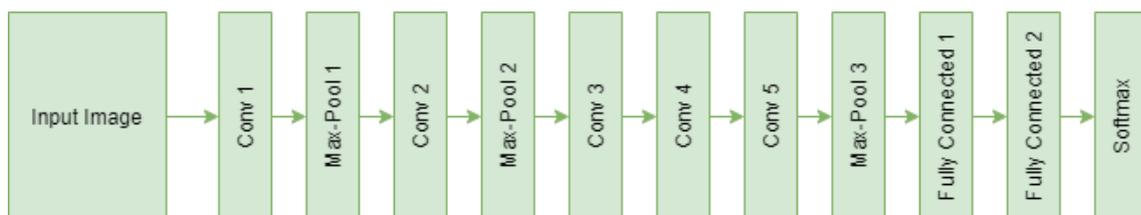


Figure 2.32 : Architecture CNN de Alexnet [56]

2.7.2 Modèle VGG

Le VGG est une abréviation du « Visual Geometry Group ». C'est une architecture réseau de neurones convolutifs classique proposés par K. Simonyan et A. Zisserman de l'université d'Oxford. Le modèle a atteint une précision de 92.7% sur ImageNet, ce qui est un des meilleurs scores obtenus. Il a marqué une progression par rapport aux modèles précédents tels que Alexnet, en proposant, dans les couches de convolution, des noyaux de convolution de plus petites dimensions (3x3), que ce qui avait été fait jusque-là. Le modèle a été entraîné sur des semaines, en utilisant des cartes graphiques de pointes [57].

Ce travail a montré une amélioration importante par rapport aux architectures antérieures, en augmentant la profondeur à 16 pour le VGG16 (étudié dans le troisième chapitre) et 19 couches pour le VGG19.

L'impact de cette technologie est vaste : dans le domaine de la robotique, la vision par ordinateur pouvant ainsi, être utilisée dans n'importe quelle application où les images et les vidéos doivent être analysées avec plus de précision, que les humains [57].

2.7.3 Modèle GoogleNet

Pour améliorer la précision tout en réduisant le calcul de l'inférence du réseau de neurones profond, GoogleNet (figure 2.34) introduit un module de départ composé de filtres de différentes tailles. En conséquence, GoogleNet obtient une meilleure précision que VGG-16, tout en ne nécessitant que sept millions de poids et 1,43 G d'opérations, pour traiter une image de même taille [58].

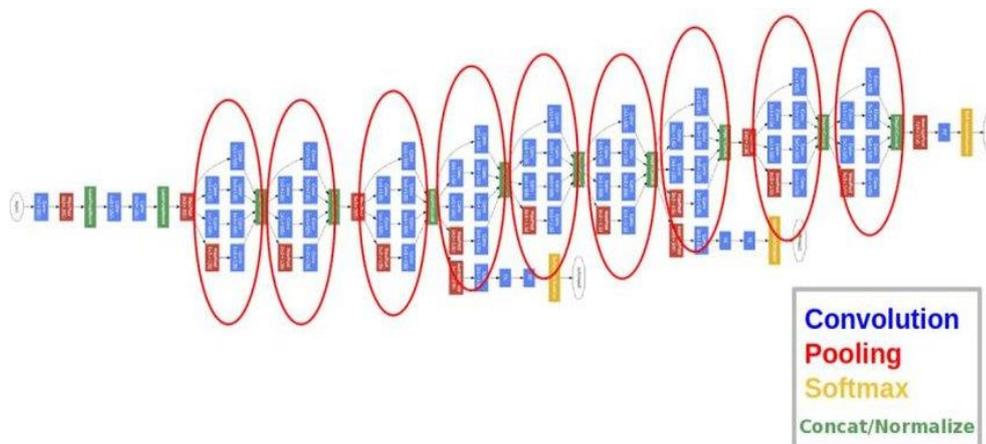


Figure 2.33 : GoogleNet architecture [58]

2.8 Conclusion

Dans ce chapitre, nous avons introduit les concepts de base du machine Learning ainsi que ceux du deep Learning. A partir de ces connaissances, nous remarquons que le deep learning est préférable au machine learning, en raison de ses hautes performances et de sa capacité cognitive, basée sur l'IA. Cet apprentissage permet seulement d'améliorer la précision des traitements, mais également de réduire le coût et la durée du processus. Le chapitre suivant, est consacré à l'étude conceptuelle d'un système d'aide à la détection des carcinomes mammaires, par les modèles des réseaux convolutifs.

d'un système d'aide à la classification des carcinomes mammaires

3.1 Introduction

L'intelligence artificielle est aujourd'hui essentielle dans le diagnostic des pathologies, avec une plus grande précision qu'un radiologue, grâce à des approches d'apprentissage automatique et profond.

Ce chapitre a pour objectif de construire un système d'aide à la classification des carcinomes mammaires selon 3 grades, où les cancers du sein de grade I sont les moins agressifs, le grade II correspond aux tumeurs intermédiaires et enfin les tumeurs de grade III qui sont les plus agressifs. L'étude est consacrée aux modèles utilisés dans cette conception.

3.2 Description de quelques travaux en imagerie mammaire

De nombreux travaux permettent la reconnaissance du cancer du sein, suivant les approches du deep learning. Nous présentons quelques-uns de ces travaux. Bayramoglu et al [59] ont réalisé, une classification des tumeurs mammaires sur l'ensemble de données BreakHis suivant un CNN, où des noyaux de convolution de taille différente (7×7, 5×5 et 3×3) ont été utilisés. Le taux de reconnaissance obtenu, est de 83,25%.

La classification des images histologiques, provenant de la base IDC [60] a été accomplie suivant plusieurs modèles des réseaux de neurones convolutifs.

Les images ont été regroupées selon quatre classes différentes : tissu normal, lésion bénigne, carcinome in situ et carcinome invasif et catégorisées en classes binaires (tissus normaux et carcinomes). L'approche basée sur les CNN, a atteint une précision de 77,8% lors de l'expérience à quatre classes et 83,3%, pour les classes binaires. Dans l'étude de Han et al [61], une précision moyenne de 93,2 % a été atteinte pour la classification des carcinomes suivant un CNN adapté. Les chercheurs Barsha et al [62] ont présenté une étude comparative approfondie, sur la classification automatisée des carcinomes, à l'aide d'un ensemble de modèles pré-entraînés d'apprentissage profond, tels que les DenseNet 121, 169, les ResNets 101v2 et 50. Dans cette étude, le modèle DenseNet-121 a engendré les meilleurs résultats, avec une précision de 92,70 % et un score F1 de 95,70%, surpassant les résultats précédents de l'état de l'art.

Dans les travaux des laboratoires LATSI et LRDSI, plusieurs modèles ont été traités dans le cadre de la détection du cancer du sein. Sur les images échographiques [63, 64, 65], la détection des masses, suivant VGG19, VGG16 et Inception V3 s'est faite avec une précision de 94,64% et une AUC de 0,98. La catégorisation des carcinomes mammaires [66], par les réseaux VGG16, VGG19 ainsi qu'un CNN adapté, a permis une précision de 94%.

3.3 Conception d'un système d'aide à la classification des carcinomes mammaires

Notre étude consiste à développer un système d'aide à la décision, afin de classer le cancer du sein selon 3 grades, à partir d'images histologiques. Le concept est d'utiliser plusieurs architectures de CNNs, afin de choisir la plus adaptée.

3.3.1 Architecture du système conçu

L'architecture du système conçu est présentée dans la figure 3.1. Cette conception s'appuie sur quatre modèles pré-entraînés : VGG16, Resnet50, Densenet201 et Densenet121.

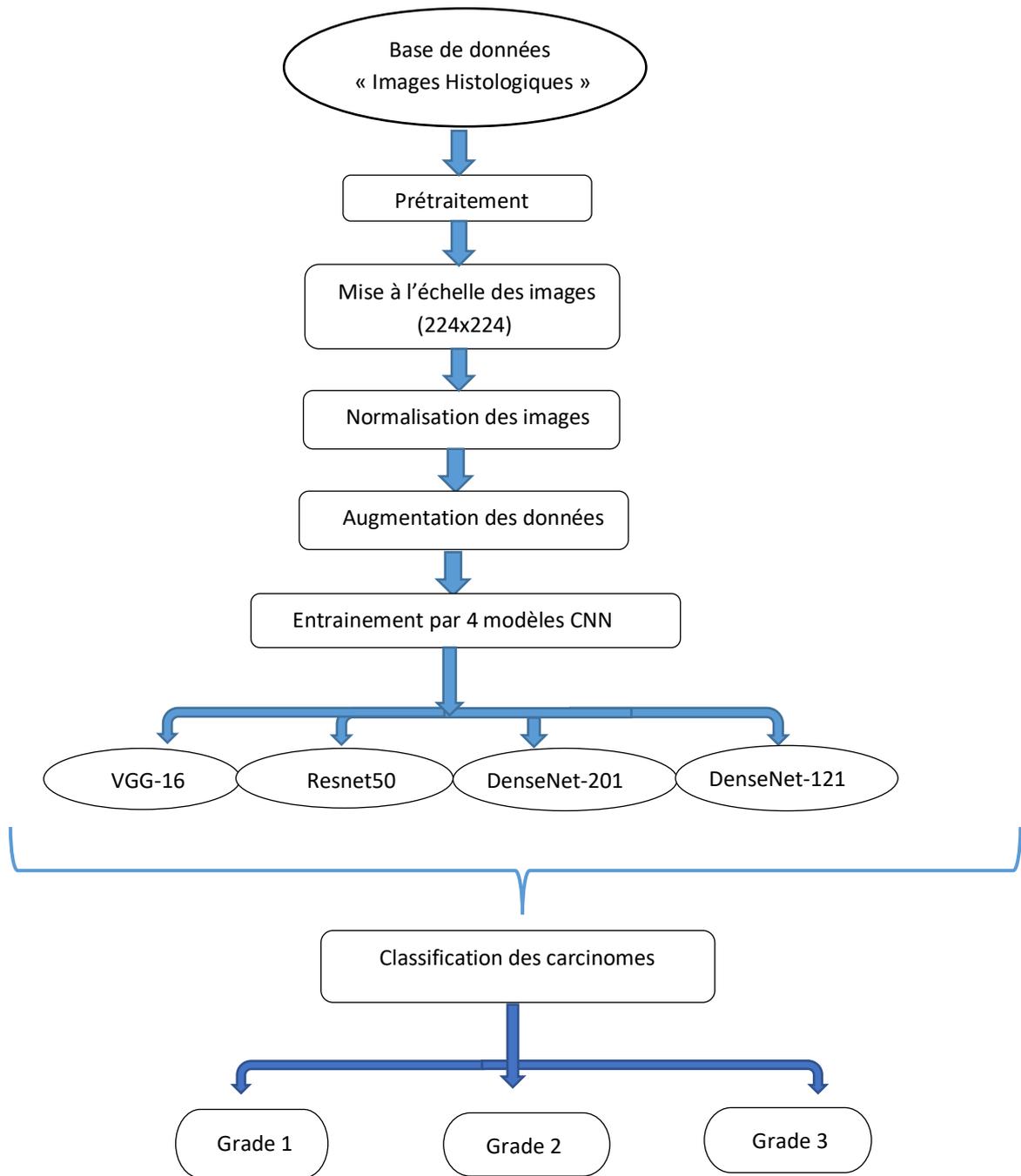


Figure 3.1 : Architecture conceptuelle du système d'aide à l'interprétation

Après avoir introduit les images de la base de données, le système lance le processus d'extraction des caractéristiques, convertissant ainsi, les données à des configurations plus réduites, afin d'obtenir les informations pertinentes, qui seront utilisées lors de la classification. Cette étape réduit les dimensions de l'image d'entrée à une dimension inférieure ((224x224), (299x299), ...etc), pour permettre l'apprentissage en mini-batch et pour respecter les limitations de calcul, ainsi que pour l'amélioration des performances des modèles pré-entraînés utilisés.

L'étape suivante est la normalisation des images, c'est-à-dire mettre toutes les images à la même échelle (de 0.255 à 0.1).

Pendant l'apprentissage, l'ensemble de données d'apprentissage est fortement augmenté et les images sont normalisées entre 0 et 1. Pour l'augmentation, 50% des images sont retournées horizontalement et 20% des images sont retournées verticalement, recadrées, affinées, dilatées, assombries et éclaircies en multipliant tous les pixels par des valeurs aléatoires (entre 0.5 et 1.0). Des valeurs entre 10 et -10 sont ajoutées aux pixels de 50% des images. 1% et -10% des pixels de l'image, sont transformés en pixels noirs.

En dehors de ces augmentations, des transformations affines sont également effectuées où les images sont tournées (entre - 10° et + 10°) et translattées (entre - 25° et + 25°).

Nous modifions les architectures choisies, en remplaçant leurs dernières couches par un nouveau classifieur adapté, aux trois classes de la base de données étudiée. Ces modèles sont affinés selon les stratégies du fine tuning.

Le principal but, est d'atteindre les meilleures performances, lors de la classification des carcinomes mammaires.

3.3.2 Exploration de la base de données

Dans le cadre de ce projet, nous utilisons une base d'images histopathologiques mammaires importée depuis la base de données nommée « A histopathological image dataset for grading breast invasive ductal carcinomas » [67].

Cette base comprend 922 images histopathologiques en RGB, au format JPEG, extraites de 124 patientes diagnostiqués avec un IDC entre 2014 et 2019 au centre de recherche Poursina Hakim de l'université des sciences médicales d'Isfahan en Iran, et publiée en 2020.

Pour chaque spécimen, quatre niveaux de grossissement (4x, 10x, 20x et 40x) sont fournis. Toutes les images de l'ensemble de données sont étiquetées, en fonction de leur grade diagnostiqué, qui peut être utilisé pour l'entraînement des algorithmes d'apprentissage automatique pour la détection du grade IDC.

L'ensemble des données sera réparti en trois ensembles, l'un pour l'entraînement, le second pour la validation et le troisième pour le test.

3.3.3 Prétraitement des images

Cette étape permet d'uniformiser les images utilisées dans le traitement, pour l'utilisation des modèles pré-entraînés ; car en réalité, les images sont sur différentes formes, résolutions et en niveaux de gris. Il est alors nécessaire de les redimensionner à une taille de 224x224 pixels suivant le modèle utilisé.

3.3.4 Augmentation des données

L'augmentation des données peut être utilisée, pour augmenter artificiellement la taille d'un ensemble de données de formation, en créant des versions modifiées des images de cet ensemble. Elle est souvent employée pour améliorer les performances et réduire l'erreur de généralisation, lors de la formation de modèles de réseaux neuronaux dans la vision par ordinateur.

Cette technique souvent utilisée dans l'apprentissage profond, consiste à générer artificiellement de nouvelles données, en déformant les données primaires, à travers des transformations qui préservent l'information pertinente. Les opérations de cette technique, permettent d'améliorer la diversité des données d'entraînement d'un classifieur, dans le but de parfaire ses performances [65, 68, 69]. On cite quelques opérations dans l'augmentation des données.

a Retournement

Un retournement (flip) d'image signifie l'inversion des lignes ou des colonnes de pixels respectivement, dans le cas d'un flip vertical ou horizontal. Certaines bibliothèques ne fournissent pas de fonction pour les flips verticaux, ce qui n'est pas problématique. En effet, un flip vertical peut être obtenu, grâce à une rotation d'une image de 180 degrés.

b Translation

La translation (shift) consiste à déplacer l'image, sur l'axe des X et/ou Y. Le shift horizontal translate l'image vers la gauche ou la droite, tandis que le shift vertical translate l'image vers le haut ou le bas.

c Rotation

Une chose importante à noter à propos de cette opération, est que les dimensions de l'image ne peuvent pas être conservées après la rotation. Si l'image est un carré, la rotation de 90° préservera la taille de l'image. Si c'est un rectangle, la rotation de 180 degrés préserverait la taille. La rotation de l'image selon des angles plus fins, modifie également la taille finale de l'image.

d Mise à l'échelle

L'image peut être mise à l'échelle (Scale) vers l'extérieur (outward), ou vers l'intérieur (inward). Lors de la mise à l'échelle vers l'extérieur, la taille finale de l'image sera plus grande que la taille d'origine. La plupart des cadres d'image découpent une section de la nouvelle image, avec une taille égale à l'image d'origine. Le scaling inward réduit la taille de l'image.

e Découpage d'une section

Contrairement à la mise à l'échelle, nous échantillons aléatoirement une section de l'image d'origine. Nous redimensionnons ensuite cette section à la taille de l'image d'origine.

3.4 Architectures des réseaux de neurones convolutifs

Depuis quelques années, les modèles des réseaux de neurones convolutifs, connaissent plusieurs challenges, dans la classification des données, à partir d'une gigantesque base d'images Imagenet. L'étude conceptuelle de ce projet, est élaborée sur quatre modèles issus des challenges effectués.

3.4.1 ImageNet

C'est une base de données d'images composée de 14,197,122 images avec 1000 classes [70]. Il s'agit d'une initiative visant à aider les chercheurs, dans le domaine de la vision.

ImageNet organise également des concours, dont l'un était ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), qui a mis au défi les chercheurs du monde entier pour proposer des solutions qui produisent les taux d'erreur les plus bas (le taux d'erreur des 5 premiers serait le pourcentage d'images, où l'étiquette correcte n'est pas l'une des cinq étiquettes les plus probables du modèle). Le concours donne un ensemble de test de 150 000 images.

En 2014, l'architecture du VGG a surpassé d'autres modèles de pointe et est toujours préférée, pour de nombreux problèmes difficiles [57].

Le défi ImageNet a été crucial pour démontrer l'efficacité des CNN profonds. Le problème consiste à reconnaître des catégories d'objets, dans une image typique que l'on peut trouver sur Internet. La tâche de classification de 2012 'ImageNet Large Scale Visual Recognition Challenge (ILSVRC)', consiste à classer des images obtenues à partir de Flickr et d'autres moteurs de recherche, parmi 1000 catégories d'objets possibles. Cette tâche sert de référence standard pour l'apprentissage profond. Les images ont été étiquetées manuellement en fonction de la présence ou de l'absence d'un objet, appartenant à ces catégories. L'ensemble d'entraînement compte 1,2 million d'images, avec 732 à 1300 images d'entraînement disponibles par classe. Un sous-ensemble aléatoire de 50 000 images a été utilisé comme ensemble de validation et 100 000 images ont été utilisées pour l'ensemble de test, avec respectivement, 50 et 100 images par classe (tableau 3.1).

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.713	0.901	138,357,544	23
VGG19	549 MB	0.713	0.900	143,667,240	26
ResNet50	98 MB	0.749	0.921	25,636,712	-
ResNet101	171 MB	0.764	0.928	44,707,176	-
ResNet152	232 MB	0.766	0.931	60,419,944	-
ResNet50V2	98 MB	0.760	0.930	25,613,800	-
ResNet101V2	171 MB	0.772	0.938	44,675,560	-
ResNet152V2	232 MB	0.780	0.942	60,380,648	-
InceptionV3	92 MB	0.779	0.937	23,851,784	159
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
MobileNet	16 MB	0.704	0.895	4,253,864	88
MobileNetV2	14 MB	0.713	0.901	3,538,984	88
DenseNet121	33 MB	0.750	0.923	8,062,504	121
DenseNet169	57 MB	0.762	0.932	14,307,880	169
DenseNet201	80 MB	0.773	0.936	20,242,984	201
NASNetMobile	23 MB	0.744	0.919	5,326,716	-
NASNetLarge	343 MB	0.825	0.960	88,949,818	-

Tableau 3.1 : Résultats de différentes architectures dans le challenge [71]

3.4.2 Modèle VGG-16

VGG-16 est un modèle de réseau neuronal convolutif, proposé par K. Simonyan et A. Zisserman de l'université d'oxford dans l'article « Very Deep Convolutional Networks for Large-Scale Image Recognition » [57, 72]. Le modèle a atteint une précision dans le test de 92.7% dans le top 5 dans ImageNet. C'était l'un des célèbres modèles, soumis à ILSVRC-2014. Il a apporté une amélioration par rapport à AlexNet, en remplaçant les filtres des deux premières couches (11*11 et 5*5), par plusieurs filtres de taille 3x3 l'un [128]. VGG-16 a été formé pendant des semaines avec un GPU NVIDIA Titan Black.

Les entrées dans le modèle VGG16 (figure 3.2), sont des images RVB de taille 224x224. L'image passe à travers un empilement de couches convolutives, où les filtres sont utilisés avec une taille de 3*3. Le pas de la convolution est fixé à 1 pixel, tel que la résolution spatiale soit préservée après convolution. La mise en commun spatiale est effectuée par cinq couches de max pooling, qui suivent certaines des couches de convolution (toutes les couches de convolution, ne sont pas suivies par un max-pooling). La mise en commun maximale, est effectuée sur une fenêtre de 2x2 pixels, avec un pas de 2.

Trois couches entièrement connectées (FC) suivent une pile de couches convolutives (qui ont une profondeur différente dans différentes architectures) : les deux premières ont 4096 canaux chacune, la troisième effectue une classification ILSVRC à 1000 voies et contient donc 1000 canaux (un pour chaque classe). La couche finale est la couche soft-max.

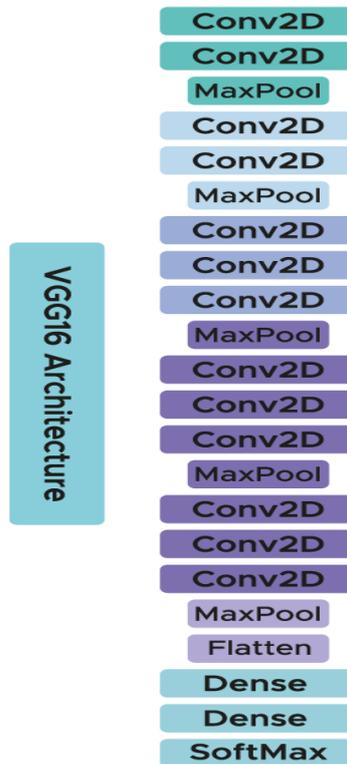


Figure 3.2 : Architecture du modèle VGG-16 [72]

La configuration des couches entièrement connectées, est la même dans tous les réseaux. Toutes les couches cachées, sont équipées de la non-linéarité de rectification (ReLU).

3.4.3 Modèle ResNet-50

ResNet-50 (figure 3.3) est un réseau neuronal convolutif, variant du modèle ResNet avec 50 couches de profondeur [10]. Il comporte 48 couches de convolution avec 1 couche MaxPool et 1 couche Average Pool. Il réalise $3,8 \times 10^9$ opérations en virgule flottante. C'est un modèle ResNet largement utilisé [73].

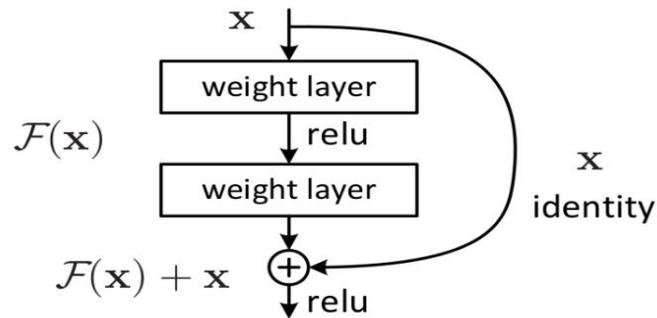


Figure 3.3 : ResNet-50 [73]

L'architecture ResNet (tableau 3.2) contient les éléments suivants :

- Une convolution avec une taille de noyau de 7*7 et 64 noyaux différents tous avec un pas (stride) de taille 2, nous donnant 1 couche.
- Une mise en commun maximale avec également un pas de taille 2.
- Dans la convolution suivante, il y a un noyau de 1*1.64 suivi d'un noyau de 3*3.64 et enfin d'un noyau de 1*1.256. Ces trois couches sont répétées au total 3 fois, ce qui nous donne 9 couches dans cette étape.
- Dans l'étape suivante, nous remarquons un noyau de 1*1.128, un noyau de 3*3.128 et enfin un noyau de 1*1.512. Cette étape est répétée 4 fois, nous donnant ainsi 12 couches dans cette étape.
- Dans les couches suivantes, il y a un noyau de 1*1.256 ainsi que deux autres noyaux avec 3*3.256 et 1*1.1024. Ces opérations sont répétées 6 fois, nous donnant un total de 18 couches.
- Dans les couches suivantes, nous remarquons, un noyau de 1*1.512 avec deux autres de 3*3.512 et 1*1.2048. Ces opérations sont répétées 3 fois, nous donnant un total de 9 couches.
- Après ces différentes opérations, un moyen pooling est effectué suivi de la couche entièrement connectée, contenant 1000 nœuds. A la fin, une fonction softmax, indique 1 couche.

En totalisant toutes ces couches, le modèle Resnet50, représente un réseau convolutif profond de $1+9+12+18+9+1=50$ couches [73].

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Tableau 3.2 : Architecture du ResNet50 [73]

3.4.4 Modèle DenseNet

Dans le DenseNet, chaque couche obtient des entrées supplémentaires de toutes les couches précédentes et transmet ses propres cartes de caractéristiques, à toutes les couches suivantes. A cet effet, la concaténation est utilisée. Chaque couche reçoit une ‘connaissance collective’ de toutes les couches précédentes.

Etant donné que chaque couche reçoit des cartes d’entités de toutes les couches précédentes, le réseau peut être plus fin et compact, c’est-à-dire que le nombre de canaux peut être inférieur. Le taux de croissance k , est le nombre supplémentaire de canaux pour chaque couche. Ainsi, le réseau Densenet a une efficacité de calcul et une efficacité de mémoire plus élevées [74].

a Avantages du DenseNet

DenseNet a été spécialement développé pour améliorer la précision causée par le gradient de fuite, dans les réseaux de neurones de haut niveau. Ceci est réalisé à cause de la longue distance entre les couches d’entrée, de sortie et des informations qui disparaissent avant d’atteindre leur destination.

Le tableau 3.3 montre les erreurs Top-1 et Top-5 pour différentes tailles de DenseNet sur ImageNet.

Model	top-1	top-5
DenseNet-121	25.02 / 23.61	7.71 / 6.66
DenseNet-169	23.80 / 22.08	6.85 / 5.92
DenseNet-201	22.58 / 21.46	6.34 / 5.54
DenseNet-264	22.15 / 20.80	6.12 / 5.29

Tableau 3.3 : Différents modèles de DenseNet [74]

Le modèle DenseNet existe en différentes versions, comme le DenseNet-121, le le DenseNet-160, le DenseNet-201, etc. Les chiffres indiquent le nombre de couches dans le réseau neuronal. Le nombre 121 est calculé comme suit :

$$\text{DenseNet-121:-} \\ 5 + (6 + 12 + 24 + 16) * 2 = 121$$

- 5 – Convolution and Pooling Layer
- 3 – Transition layers (6,12,24)
- 1 – Classification Layer (16)
- 2 – DenseBlock (1x1 and 3x3 conv)

b Architecture du modèle DenseNet-121

Chaque bloc dense a un nombre variable de couches (répétitions) comportant chacune deux convolutions, un noyau de taille 1x1 comme couche de goulot d'étranglement et un noyau 3x3 pour effectuer l'opération de convolution [74].

Chaque couche de transitions à une couche convolutive 1x1 et une couche de mise en commun moyenne 2x2 avec une foulée de 2.

Les couches présentes sont les suivantes :

1. Couche conv avec 64 filtres de taille 7x7 et stride of 2.
2. Couche pooling de 3x3 max pooling et stride of 2.
3. Block Dense 1 avec 2 convolutions repeated 6 fois.
4. Transition layer 1 (1 Conv + 1 AvgPool).
5. Block Dense 2 avec 2 convolutions répéter 12 fois.
6. Couche de Transition 2 (1 Conv + 1 AvgPool).

7. Block Dense 3 avec 2 convolutions répéter 24 fois.
8. Couche de Transition 3 (1 Conv + 1 AvgPool).
9. Block Dense 4 avec 2 convolutions répéter 16 fois.
10. Couche "Global Average pooling" -accepts all the feature maps of the network to perform classification.
11. Couche d'Output.

Par conséquent, DenseNet-121 a les couches suivantes :

- 1 7x7 Convolution
- 58 3x3 Convolution
- 61 1x1 Convolution
- 4 AvgPool
- 1 couche Fully Connected

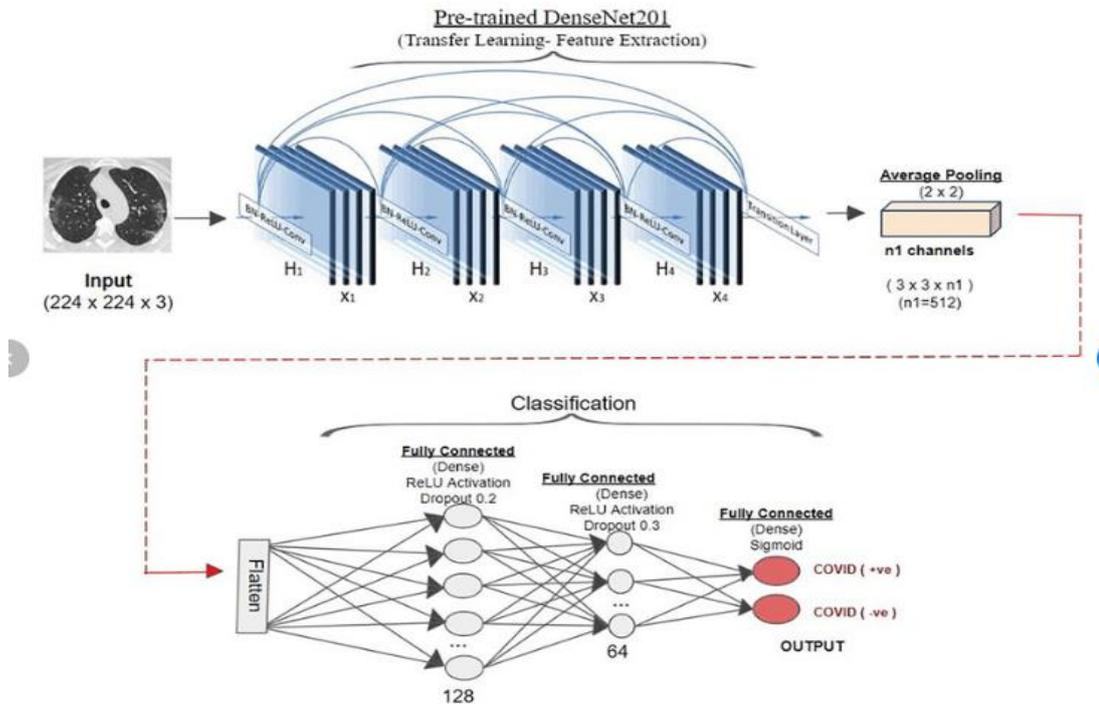
En conclusion, le DenseNet-121 à 120 convolutions et 4 pooling moyens [74].

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

Tableau 3.4 : Architecture du DenseNet-121 [74]

c Architecture du modèle DenseNet-201

DenseNet-201 (figure 3.4) est un réseau neuronal convolutif de 201 couches de profondeur.



Architecture of proposed transferred DenseNet201 for feature extraction with CNN for classification.

Figure 3.4 : Architecture DenseNet-201 [74]

d Blocs denses et couches

Qu'il s'agisse d'une addition ou d'une concaténation, le regroupement des couches, n'est possible que si les dimensions des cartes de caractéristiques sont identiques.

Le DenseNet est divisé en 'DenseBlocks' où un certain nombre de filtres sont différents, mais où les dimensions à l'intérieur du bloc sont les mêmes. La couche de transition applique la normalisation par lots en utilisant le sous-échantillonnage, une étape essentielle du CNN [62, 74].

Voyons ce qu'il y a à l'intérieur du DenseBlock et de la couche de transition selon la figure 3.5.

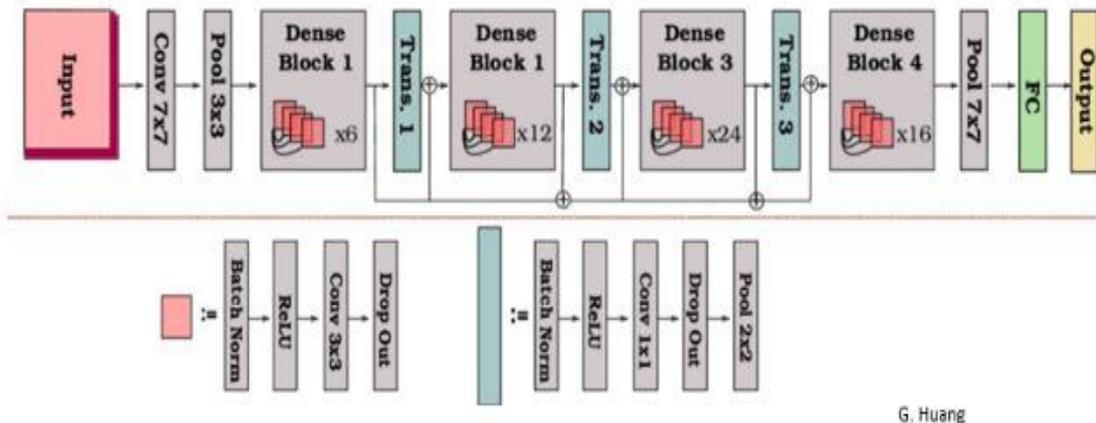


Figure 2.5 : 'DenseBlock' [74]

Le nombre de filtres change entre les 'DenseBlocks', ce qui augmente les dimensions du canal. Le taux de croissance (k) aide à généraliser la l-ième couche. Il contrôle la quantité d'informations à ajouter à chaque couche (équation 3.1).

$$k^{[l]} = (k^{[0]} + k(l - 1)) \tag{3.1}$$

L'architecture complète de densenet, est donnée par la figure 3.6.

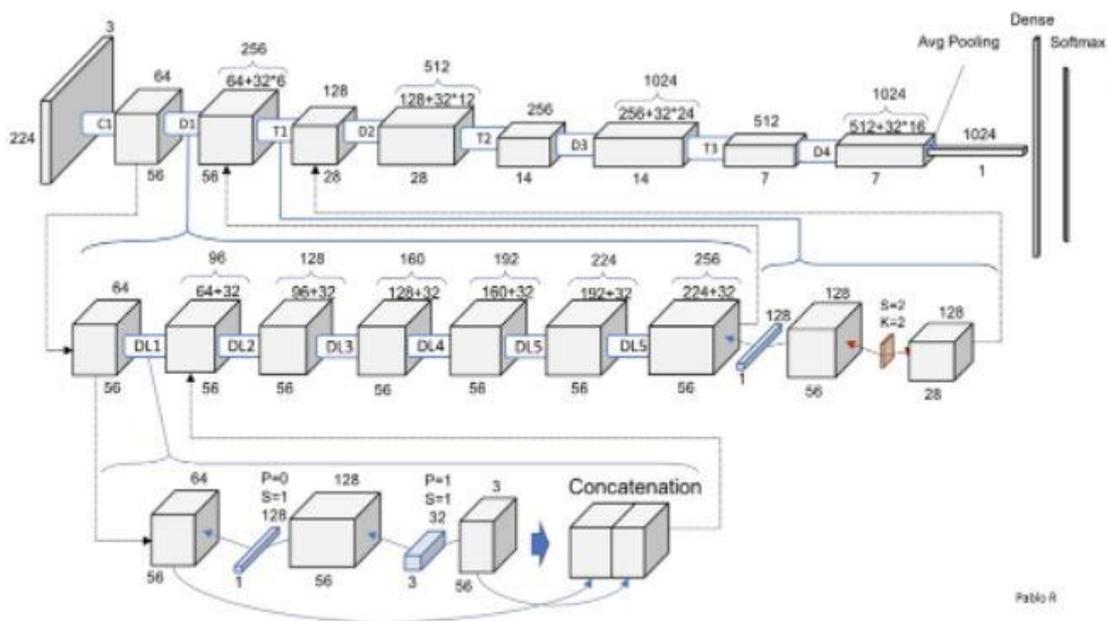


Figure 3.6 : Architecture complète de DenseNet [74]

3.4.5 Différence entre DenseNet et ResNet

DenseNet est l'une des nouvelles découvertes en matière de réseaux neuronaux, pour la reconnaissance visuelle des objets. DenseNet est assez similaire à ResNet, avec quelques différences fondamentales. ResNet utilise une méthode additive, qui fusionne la couche précédente (identité) avec la couche future, alors que DenseNet concatène la sortie de la couche précédente, avec la couche future [75].

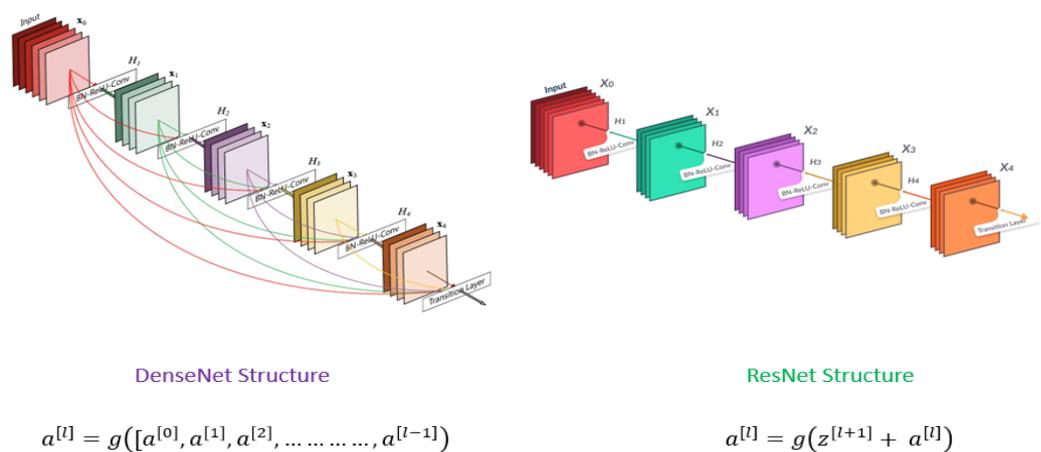


Figure 3.7 : Architectures de DenseNet et ResNet [75]

Une sortie de la couche précédente agit comme une entrée de la deuxième couche, en utilisant une opération de fonction composite. Cette opération composite, se compose de la couche de convolution, de la couche de mise en commun, de la normalisation par lots et de la couche d'activation non linéaire.

Ces connexions signifient que le réseau possède $L(L+1) / 2$ connexions directes. L est le nombre de couches dans l'architecture.

3.5 Fine tuning

Le réglage fin « fine tuning » est une approche de l'apprentissage par transfert, qui consiste à modifier la sortie du modèle, pour l'adapter à la nouvelle tâche et à former uniquement le modèle de sortie [76].

Dans le réglage fin, on optimise à la fois les poids des nouvelles couches de classification ajoutées, ainsi que certaines ou toutes les couches du modèle.

L'ajustement 'fine tuning' permet de réduire le temps d'apprentissage et, le risque de surapprentissage du modèle inhérent à un jeu de données, de taille inférieure au nombre de paramètres du modèle.

Il existe deux stratégies.

3.5.1 Fine tuning total

La couche fully-connected du réseau pré-entraîné, est remplacée par un classifieur adapté au nouveau problème et initialisé aléatoirement. Toutes les couches sont par la suite, entraînées sur les nouvelles images. Cette stratégie est utilisée lorsque la nouvelle base de données est grande.

3.5.2 Fine tuning partiel

La dernière couche fully-connected, est remplacée par le nouveau classifieur initialisé aléatoirement et, les paramètres de certaines couches, du réseau pré-entraîné sont fixés. Ainsi, en plus du classifieur, les couches non-fixées, seront entraînées sur les nouvelles données, qui correspondent généralement aux plus hautes du réseau. Cette stratégie est utilisée lorsque la nouvelle base des données est petite, mais très différente de celle du pré-entraînement.

3.6 Outils de mesures des performances

La prédiction des classes pour les données dans un problème de classification, est basée sur la recherche de la limite optimale entre les classes. En se basant sur les valeurs de certaines métriques, on peut trouver la limite optimale.

La performance du modèle dans un problème de classification, est évaluée à l'aide de plusieurs métriques, telles que la matrice de confusion, la précision, l'exactitude, la spécificité, la sensibilité, le rappel, ainsi que d'autres outils de mesure [77, 78, 79, 80].

3.6.1 Matrice de confusion

La matrice de confusion est en quelque sorte un résumé des résultats de la prédiction, pour un problème particulier de classification. Elle compare les données réelles pour une variable cible, à celles prédites par un modèle. Les prédictions justes et fausses sont révélées et réparties par classe, ce qui permet de les comparer avec des valeurs définies [77].

Les éléments de la matrice de confusion (tableau 3.5) sont utilisés pour trouver trois paramètres importants de la précision, la sensibilité et la spécificité.

Confusion matrix		Reality	
		Negative : 0	Positive : 1
Prediction	Negative : 0	True Negative : TN	False Negative : FN
	Positive : 1	False Positive : FP	True Positive : TP

Tableau 3.5 : Matrice de confusion [77]

Également connue sous le nom de tableau de contingence, la matrice de confusion permet d'évaluer les performances d'un modèle de classification. Elle montre donc à quel point un certain modèle peut être confus, lorsqu'il fait des prédictions. Dans sa forme la plus simple, il s'agit d'une matrice 2X2. Pour des problèmes de classification plus complexes, il est toujours possible d'ajouter des lignes et des colonnes à la forme de base.

3.6.2 Exactitude

L'exactitude (accuracy) est une métrique (équation 3.2), permettant d'évaluer la performance des modèles de classification à 2 classes ou plus [146, 148].

$$Accuracy \% = \frac{TP+TN}{TP+FP+FN+TN} \times 100 \quad 3.2$$

TP = True Positives (vrais positifs), TN = True Negatives (vrais négatifs), FP = False Positives (faux positifs), et FN = False Negatives (faux négatifs).

Où :

- **Vrai positif** ; représente le cas d'une prédiction positive et elle est réellement positive.
- **Vrai négatif** : représente le cas d'une prédiction négative et elle est réellement négative.
- **Faux positif** : représente le cas d'une prédiction positive alors qu'elle est réellement négative.
- **Faux négatif** : représente le cas d'une prédiction négative alors qu'elle est réellement positive.

Avec ces 4 cas, on identifie bien la symétrie des individus négatifs et des individus positifs dans la définition de l'exactitude.

3.6.3 Précision

La précision est la fraction des instances correctement classées. Elle est représentée par l'équation 3.3.

$$\text{Précision \%} = \frac{TP}{TP+FP} \times 100 \quad 3.3$$

3.6.4 Spécificité

La spécificité correspond aux négatifs réels correctement classés. Elle est déterminée par l'équation 3.4.

$$\text{Spécificité \%} = \frac{TN}{TN+FP} \times 100 \quad 3.4$$

3.6.5 Sensibilité

Le rappel ou la sensibilité (équation 3.5), également appelé sensibilité ou taux de vrais positifs, indique correctement la proportion de vrais positifs classés.

$$Rappel \% = \frac{TP}{TP+FN} \times 100 \quad 3.5$$

3.6.6 F-score

La mesure F ou le F- score, est une mesure de la précision d'un test. Il est calculé à partir de la précision et du rappel du test. La plus haute valeur possible d'un score F est 1 ce qui indique une précision et un rappel parfaits (équation 3.6).

$$F1 \% = \frac{2 \times précision \times rappel}{précision + rappel} \times 100 \quad 3.6$$

3.6.7 Entropie croisée

L'entropie croisée (cross entropy) est l'une des fonctions de perte [77, 80], les plus couramment utilisées pour l'entraînement des réseaux de neurones.

La perte d'entropie croisée (Cross entropy loss) est une matrice utilisée, pour mesurer les performances d'un modèle de classification, dans l'apprentissage automatique.

La perte (ou l'erreur) est obtenue pour un nombre compris entre 0 et 1, où le 0 est un modèle parfait. Le but est généralement d'amener le modèle le plus près possible de 0.

La perte d'entropie croisée (équation 3.7), mesure la différence entre la distribution de probabilité découverte d'un modèle de classification d'apprentissage automatique et, la distribution prédite.

$$loss(x, c) = - \sum_{c=1}^C y_{x,c} \log(p_{x,c}) \quad 3.7$$

Où :

y est la probabilité estimée d'appartenance de x à la classe i.

p est la probabilité réelle d'appartenance de x à la classe i, sachant qu'il y a C classes.

3.7 Conclusion

Ce chapitre était consacré à l'étude conceptuelle de notre système d'aide à l'interprétation de carcinomes mammaires par les techniques des réseaux de neurones convolutifs. Les différentes architectures utilisées dans cette conception, ont été décrites, suivant leurs paramètres.

Le chapitre qui suit, introduit l'environnement de travail, la mise en œuvre du système d'aide à l'interprétation des carcinomes mammaires suivant les quatre modèles présentés, ainsi que les résultats obtenus et leur discussion.

d'aide à l'interprétation des carcinomes

4.1 Introduction

L'objectif de ce projet, est dédié à la conception et la réalisation d'un système d'aide à l'interprétation de carcinomes mammaires à partir d'images histologiques ; suivant les techniques de classification par les réseaux de neurones convolutifs pré-entraînés.

Une étude comparative sera effectuée sur les quatre modèles développés, afin de choisir le modèle type.

4.2 Environnement de travail

L'application baptisée 'BreastcarciCNN', a été réalisée sous environnement anaconda et spyder où les codes sources sont développés en python.

4.2.1 Matériel utilisé (Hardware)

Ce travail a été réalisé grâce à un PC portable, avec les caractéristiques suivantes :

- PC : DELL
- RAM : 8,00 Go
- Disque dur : 128 SSD +355 Go HDD
- CPU : Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz 2.40 GHz
- Carte Graphique : AMD Radeon (TM) R5 M335 et Intel(R) HD Graphics 520

4.2.2 Langage de programmation utilisé (Software)

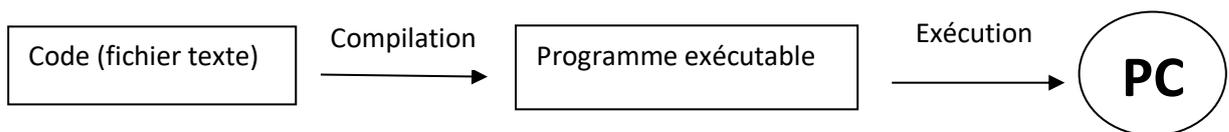
La réalisation de notre application, a été effectuée sous le système d'exploitation « Windows 10 », en utilisant Spyder qui se programme avec le langage python et, à l'aide d'un ensemble de bibliothèques open source.

a Python

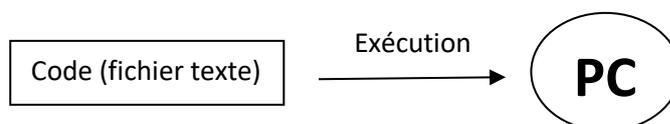
À la fin des années 80, un programmeur du nom de Guido van Rossum profite d'une semaine de vacances pour développer un langage de programmation. Fans de la célèbre série britannique Monty Python's Flying Circus, il décide de l'appeler python [81]. Rapidement adopté par ses collègues, python est en constante évolution encore aujourd'hui. C'est un langage de programmation open source.

Comme il dispose de nombreuses bibliothèques de fonction, il est très utilisé dans le milieu scientifique. Il permet par exemple d'automatiser rapidement des tâches fastidieuses ou encore, réaliser des prototypes d'application. C'est le langage qui est principalement utilisé pour le big data ou encore le machine learning.

Certains langages sont des langages compilés, c'est le cas du C++ ou du C# par exemple. Ainsi, avec un langage compilé, lors de l'écriture d'un code, il faut d'abord le compiler pour pouvoir l'exécuter.



L'inconvénient est que pour un grand programme, la compilation peut prendre un certain temps. L'avantage est que le programme généré est rapide à s'exécuter. D'autres langages sont des langages interprétés. C'est le cas du perl, SQL et python.



Dans ce cas lors de l'écriture du code, le programme généré est lu par un interpréteur, un logiciel qui lit le script pour l'exécuter. L'avantage est qu'il n'y a pas de temps perdu à la compilation. L'inconvénient, est que c'est un peu plus lent à l'exécution.

Certains langages, comme java par exemple sont entre les deux, car ils sont compilés non pas pour créer un exécutable, mais pour être lancés dans une machine virtuelle.

Python permet aux développeurs, de se concentrer sur ce qu'il fait, plutôt que comment il le fait.

C'est un langage très concis, qui permet d'effectuer beaucoup de choses avec peu de code. L'un des grands avantages d'un langage interprété comme python, c'est la possibilité de fonctionner sur de nombreux systèmes d'exploitation. Par contre, dans un langage compilé, il faut créer un exécutable pour chaque forme cible.

Déroutant au début, cela s'avère puissant pour écrire un code lisible par le plus grand nombre.

b Anaconda

Anaconda (figure 4.1) est une distribution libre des langages de programmation Python et R. Cette plateforme est utilisée pour le calcul scientifique, l'analyse prédictive et le traitement des données à grande échelle. Cette dernière est destinée aux scientifiques manipulant des données, aux développeurs, aux analystes travaillant dans des entreprises ou des organisations...etc. Elle comprend également une interface graphique, **Anaconda Navigator**, qui est une alternative graphique à l'interface de ligne de commande (CLI) [82, 83].



Figure 4.1 : Logo d'Anaconda [82]

c Anaconda Navigator

Le navigateur Anaconda (figure 4.2) est une interface graphique (**GUI**), incluse dans la distribution Anaconda, permettant aux utilisateurs de lancer des applications, mais aussi de gérer les librairies conda.

Le navigateur peut également accéder à des bibliothèques présentes sur le Cloud Anaconda ou dans un Repository Anaconda local, afin de les installer dans un environnement, les exécuter et les mettre à jour. Il est disponible pour Windows, macOS et Linux [82].

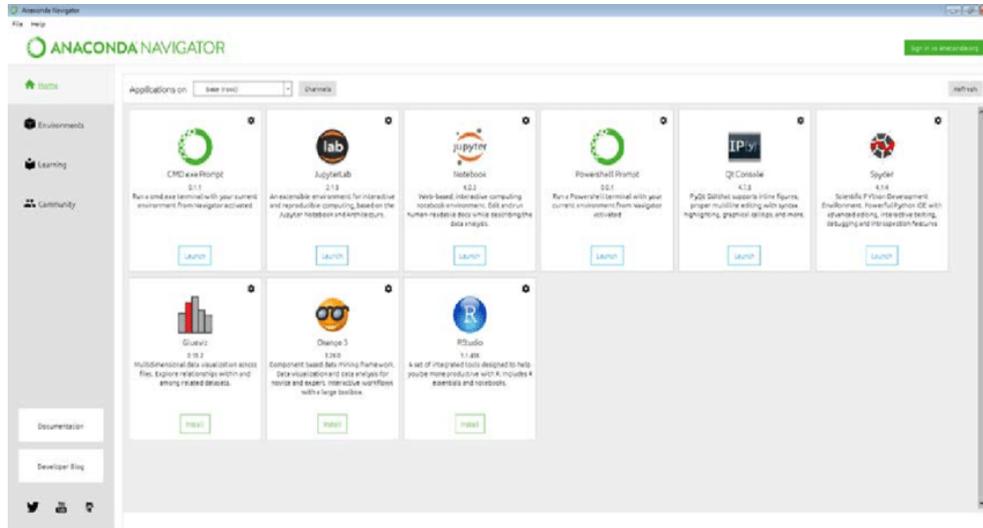


Figure 4.2 : Navigateur Anaconda [83]

d Spyder

Spyder est un environnement de développement pour Python, libre (Licence MIT) et multiplateforme (Windows, Mac OS, GNU/Linux). Conçu par et pour les scientifiques, les ingénieurs et les analystes de données. Il intègre de nombreuses bibliothèques d'usage scientifique : Matplotlib, NumPy, SciPy et IPython [84].

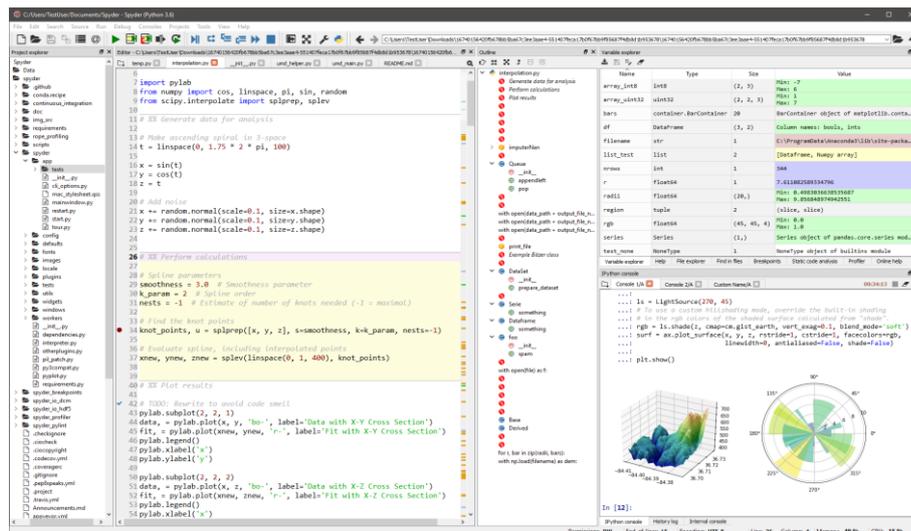


Figure 4.3 : Environnement spyder [84]

4.2.3 Bibliothèques utilisées

Il existe différentes bibliothèques dédiées au calcul matriciel, aux graphes, à l'apprentissage, à la création de réseaux de neurones, etc.

a Keras

Keras est une bibliothèque open source, écrite en Python (sous licence MIT) basée principalement, sur les travaux du développeur de Google François Chollet, dans le cadre du projet ONEIROS (**O**pen-ended **N**euro-**E**lectronic **I**ntelligent **R**obot **O**perating **S**ystem).

Une première version du logiciel multiplateforme, a été publiée le 28 mars 2015. Le but de cette bibliothèque, est de permettre la constitution rapide des réseaux neuronaux. Dans ce cadre, Keras (figure 4.4) ne fonctionne pas comme un framework propre, mais comme une interface de programmation applicative (API), pour l'accès et la programmation de différents frameworks d'apprentissage automatique. Theano, Microsoft Cognitive Toolkit (anciennement CNTK) et TensorFlow, font notamment partie des frameworks pris en charge par Keras [85].



Figure 4.4. Logo de Keras [85]

b TensorFlow

TensorFlow [86] est l'une des bibliothèques open-source les plus utilisées. Développée à l'origine par Google, elle effectue des calculs numériques en utilisant des graphiques de flux de données. À l'ère de l'intelligence artificielle, TensorFlow offre un support solide pour l'apprentissage automatique et profond. Basé sur Python, cette bibliothèque peut exécuter des réseaux neuronaux profonds, pour la reconnaissance d'images, l'intégration de mots, la classification de chiffres manuscrits et la création de divers modèles de séquence.

Son architecture flexible permet de déployer facilement les calculs sur diverses plateformes telles que ; les CPU, les GPU (Graphics Processing Unit) et les grappes de serveurs.

TensorFlow (figure 4.5) peut être utilisé pour créer des algorithmes, permettant de visualiser des objets et d'entraîner une machine à les reconnaître. Elle peut également déployer des modèles d'analyse des sentiments.

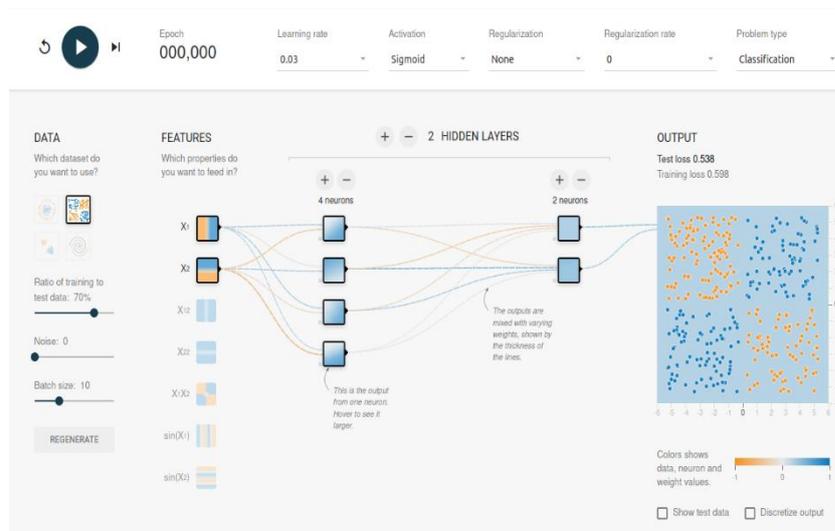


Figure 4.5 : Plateforme Tensorflow [86]

c NumPy

NumPy (Numerical Python) est le paquetage fondamental, pour le calcul scientifique en Python. Il s'agit d'une bibliothèque Python, qui fournit un tableau multidimensionnel, (tels que les tableaux masqués et les matrices), ainsi qu'un assortiment de routines permettant d'effectuer des opérations rapides sur les tableaux, notamment des opérations mathématiques, logiques, de manipulation de formes, de tri, de sélection, d'E/S, de transformées de Fourier discrètes, d'algèbre linéaire de base, d'opérations statistiques de base, de simulation aléatoire et bien plus encore [87].

d Scikit-learn

Scikit-learn est une bibliothèque libre programmée en Python destinée à l'apprentissage automatique. Elle propose dans son framework, de nombreuses bibliothèques d'algorithmes à implémenter, clef en main. Ces bibliothèques sont utiles, notamment pour les data scientists.

Elle comprend des fonctions pour estimer des régressions logistiques et, des algorithmes de classification. Elle est conçue pour s'harmoniser avec d'autres bibliothèques libres, notamment NumPy et SciPy [88].

e Splitfolders

Splitfolders est une bibliothèque en langage python, utilisée pour diviser automatiquement et aléatoirement la base de données en sous-ensembles de formation, de validation et de test [89].

f Kfolders

La procédure de validation croisée k-fold est une méthode standard d'estimation de la performance d'un algorithme d'apprentissage automatique, sur un ensemble de données.

Une approche (figure 4.6) consiste à explorer l'effet de différentes valeurs de k, sur l'estimation de la performance du modèle et à la comparer avec une condition de test idéale. Cela peut aider à choisir une valeur appropriée de k. Une fois que la valeur de k est choisie, elle peut être utilisée pour évaluer une série de différents algorithmes, sur l'ensemble de données. La distribution des résultats peut être comparée à une évaluation des mêmes algorithmes, en utilisant une condition de test idéale, pour voir s'ils sont fortement corrélés ou non. En cas de corrélation, la configuration choisie est une approximation robuste de la condition de test idéale [90].

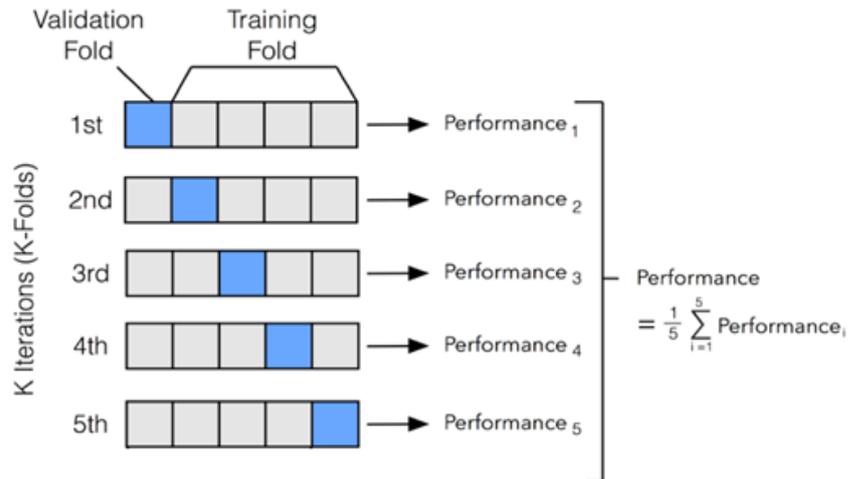


Figure 4.6 : Un exemple de la validation croisée à 5 folds (plis) [90]

L'avantage principal de cette technique est que, grâce à elle, chaque échantillon de données est testé exactement une fois, et est utilisé pour la formation k-1 fois, ce qui réduit le biais. Un autre avantage est qu'elle fonctionne bien sur les petits et grands ensembles de données, car elle garantit que chaque observation de l'ensemble de données original, a une chance d'apparaître dans les ensembles de formation et de test.

g Matplotlib

Matplotlib est une bibliothèque graphique. Elle permet de visualiser et tracer les données sous forme de graphes [91].

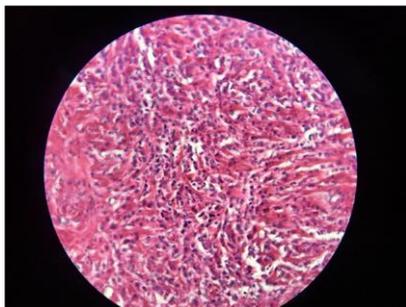
4.3 Bases de données

Dans cette étude, nous avons utilisé deux bases de données histopathologiques mammaires, suivant des tumeurs bénignes et malignes ainsi que des carcinomes à trois grades.

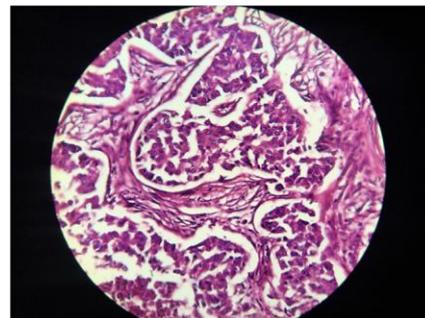
4.3.1 Base de données IDC

IDC est une base de données histopathologiques nommée « A histopathological image dataset for grading breast invasive ductal carcinomas ».

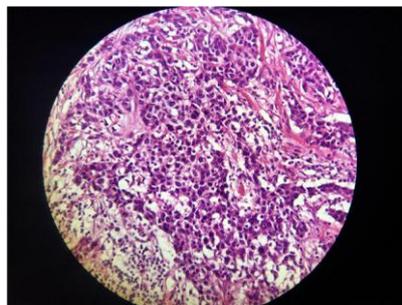
Elle comprend 922 images histopathologiques en RGB (figure 4.7), au format JPEG, extraites de 124 patientes diagnostiqués avec un IDC entre 2014 et 2019 au centre de recherche Poursina Hakim de l'université des sciences médicales d'Isfahan en Iran et publiée en 2020. Pour chaque spécimen, quatre niveaux de grossissement (4x, 10x, 20x et 40x) sont fournis. Toutes les images de l'ensemble de données, sont étiquetées en fonction de leur grade diagnostiqué, qui peut être utilisé pour l'entraînement des algorithmes d'apprentissage automatique pour la détection du grade IDC [67].



Grade 1



Grade 2



Grade 3

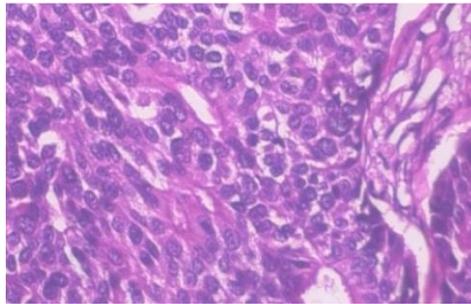
Figure 4.7 : Images x40 de la base de données IDC [67]

4.3.2 Base de données BreakHis_V1

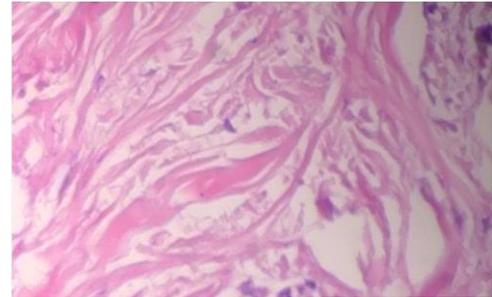
BreakHis [92] est une base de données histopathologiques mammaires. Elle a été construite en collaboration avec le Laboratoire P&D - Anatomie pathologique et cytopathologie, Parana en Brésil, extraites de 82 patientes.

Le nombre global des images microscopiques de tissus de tumeurs du sein, est de 9 109 images, dont 2 480 images sont des échantillons bénins et 5 429 sont des échantillons malins.

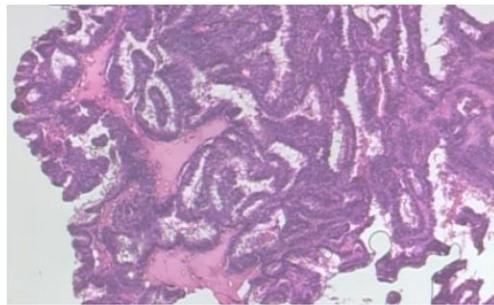
Les images (figure 4.8) de cette base sont en RGB, au format PNG, avec quatre niveaux de grossissement (x40, x100, x200, x400). La taille d'une image est de 700x460 pixels.



Ductal carcinoma



Lobular carcinoma



Papillary carcinoma

Figure 4.8 : Images de la base de données BreakHis_v1 [92]

4.4 Préparation des données

La répartition de la base de données est une étape essentielle, pour l'apprentissage.

Dans ce cadre, elle est divisée en trois ensembles :

- 60% de l'ensemble de données sont utilisés pour l'entraînement (train), permettant de déterminer les paramètres des réseaux.
- 20% sont utilisés pour la validation (val), permettant d'identifier les hyperparamètres.
- 20% pour le test permettant d'évaluer le modèle.

La division est effectuée à l'aide de la bibliothèque Split-folders, où chaque ensemble est utilisé pour une phase de création du modèle.

Après avoir effectué la division, dans le cas de la base IDC, le nombre d'images histopathologiques appartenant à chaque classe, est de 552 pour l'entraînement (train), 183 pour la validation (val) et 187 pour le test (figure 4.9).

```
Found 552 images belonging to 3 classes.  
Found 183 images belonging to 3 classes.  
Found 187 images belonging to 3 classes.
```

Figure 4.9 : Nombre d'images après la division du dataset

Dans le cas de la 2^{ème} base de données (BreakHis), le nombre d'images histologiques correspondant à chaque classe, est de 365 pour l'entraînement (train), 121 pour la validation (val), et 123 pour le test (figure 4.10).

```
Found 365 images belonging to 3 classes.  
Found 121 images belonging to 3 classes.  
Found 123 images belonging to 3 classes.
```

Figure 4.10 : Nombre d'images après la division du dataset

4.5 Architectures des modèles du système 'BreastCarcicNN'

L'application 'BreastCarcicNN', est développée suivant 4 modèles pré-entraînés, le VGG16, le Densenet201, le Densenet121 et le resnet50. Les modèles ont subi un fine tuning, pour éviter le sur-apprentissage.

Avant de faire une augmentation des données, il est important d'uniformiser les images, en appliquant un redimensionnement à l'ensemble de données grâce à l'instruction «image_resize».

Après avoir uniformisé les clichés, la classe «datagenerator» utilisée sous Keras, permet d'effectuer plusieurs transformations sur les images, afin d'augmenter le nombre de données.

Pour notre approche (figure 4.11), nous avons utilisé les opérations suivantes :

- Une rotation de 90° ;
- Un Changement d'échelle ;
- Une translation ;
- Un décalage ;
- Une amélioration de la luminosité de [0.5,1.5].

```
train_datagen = ImageDataGenerator(  
    rescale=1./255,  
    width_shift_range=0.2,  
    height_shift_range=0.2,  
    horizontal_flip=True,  
    fill_mode='nearest',  
    rotation_range=90,  
    brightness_range=(0.5,1.5),  
)
```

Figure 4.11. Augmentation des données

4.5.1 Modèle VGG-16

Le modèle VGG16 présenté dans le troisième chapitre, est réalisé pour la classification des carcinomes mammaires sur les deux bases présentées, avec un fine tuning partiel ainsi qu'un drop out de 20% et de 70% (figures 4.12 et 4.13).

VGG16

Input 224x224x3

2x Conv 224x224x64 (stride 1)(padding 1)+ReLU

Max Pooling 2x2 stride 2 padding 0

2 Conv 112x112x128 (stride 1)(padding 1)+ReLU

Max Pooling 2x2 stride 2 padding 0

3 Conv 56x56x256 (stride 1)(padding 1)+ReLU

Max Pooling 2x2 stride 2 padding 0

3 Conv 28x28x512 (stride 1)(padding 1)+ReLU

Max Pooling 2x2 stride 2 padding 0

3 Conv 14x14x512 (stride 1)(padding 1)+ReLU

Max Pooling 2x2 stride 2 padding 0

FC (out_features 512)+ ReLu

Dropout 0.5

FC (out_features 512)+ ReLu

Dropout 0.5

FC (out_features 3)

Output Softmax

Figure 4.12 : Architecture de VGG16 adaptée aux images histologiques mammaires

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 1, 1, 512)	14714688
global_max_pooling2d_1 (GlobalMaxPooling2D)	(None, 512)	0
dense_5 (Dense)	(None, 512)	262656
dropout_2 (Dropout)	(None, 512)	0
dense_6 (Dense)	(None, 512)	262656
dropout_3 (Dropout)	(None, 512)	0
dense_7 (Dense)	(None, 3)	1539

Figure 4.13 : Architecture du VGG16

Les couches ajoutées au modèle VGG16 sont :

- Une couche de Max pooling 2D
- Une couche « fully connected » qui contient 512 nœuds en utilisant la fonction d'activation ReLU.
- Une couche « Dropout » de 20%.
- Une 2^{ème} couche « fully connected » qui contient 512 nœuds en utilisant la fonction d'activation ReLU.
- Une 2^{ème} couche « Dropout » de 70%.
- Une couche « fully connected » qui contient 3 nœuds (parce qu'on a 3 classes), avec une fonction d'activation softmax.

Les résultats obtenus par le modèle VGG-16, sont donnés par l'exactitude et la perte dans les graphes 4.14 et 4.15.

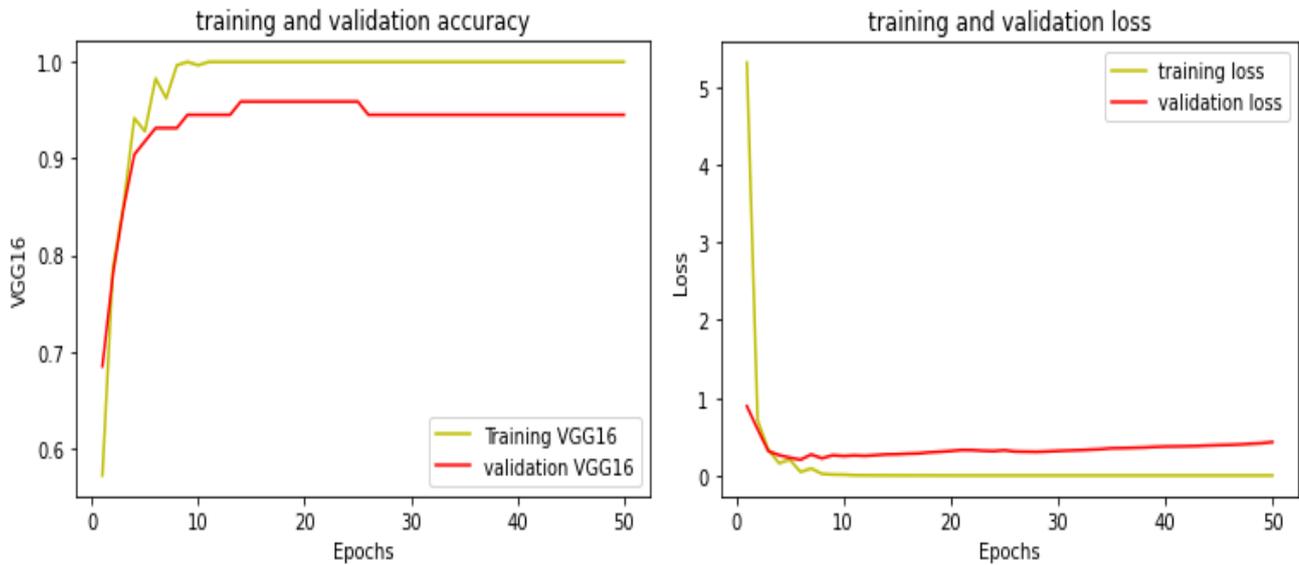


Figure 4.14 : Exactitude et Perte du modèle VGG16 en utilisant la base (Breakhis_v1)

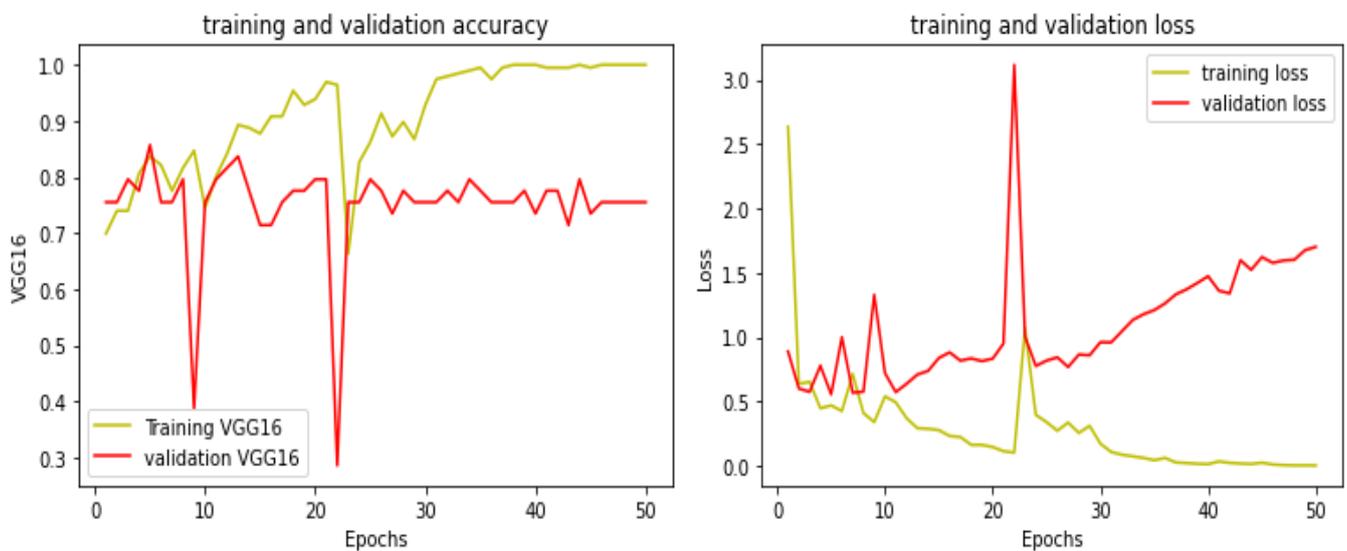


Figure 4.15 : Exactitude et perte du modèle VGG16 en utilisant la base (IDC)

4.5.3 Modèle ResNet-50

Le modèle ResNet a une architecture différente aux autres modèles. Il est constitué de blocs résiduels. Avec un fine tuning, nous avons ajouté les couches suivantes (figures 4.16 et 4.17), au modèle de base :

- Une couche « flatten »
- Une couche « Dropout » de 70%.
- Une couche « fully connected » qui contient 3 nœuds (parce qu'on a 3 classes), avec une fonction d'activation softmax.

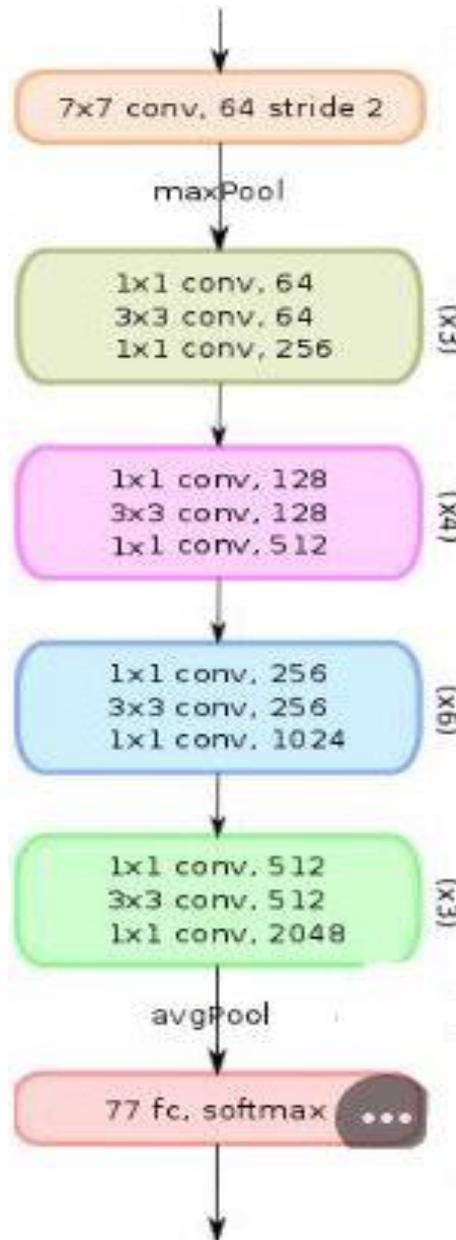


Figure 4.16 : Architecture Resnet50 adaptée aux bases histologiques

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 2, 2, 2048)	23587712
flatten (Flatten)	(None, 8192)	0
dropout (Dropout)	(None, 8192)	0
dense (Dense)	(None, 3)	24579

=====
 Total params: 23,612,291
 Trainable params: 23,559,171
 Non-trainable params: 53,120

Figure 4.17 : Architecture de ResNet-50

Les résultats obtenus par le modèle Resnet50, sont donnés par l'exactitude et la perte dans les graphes 4.18 et 4.19.

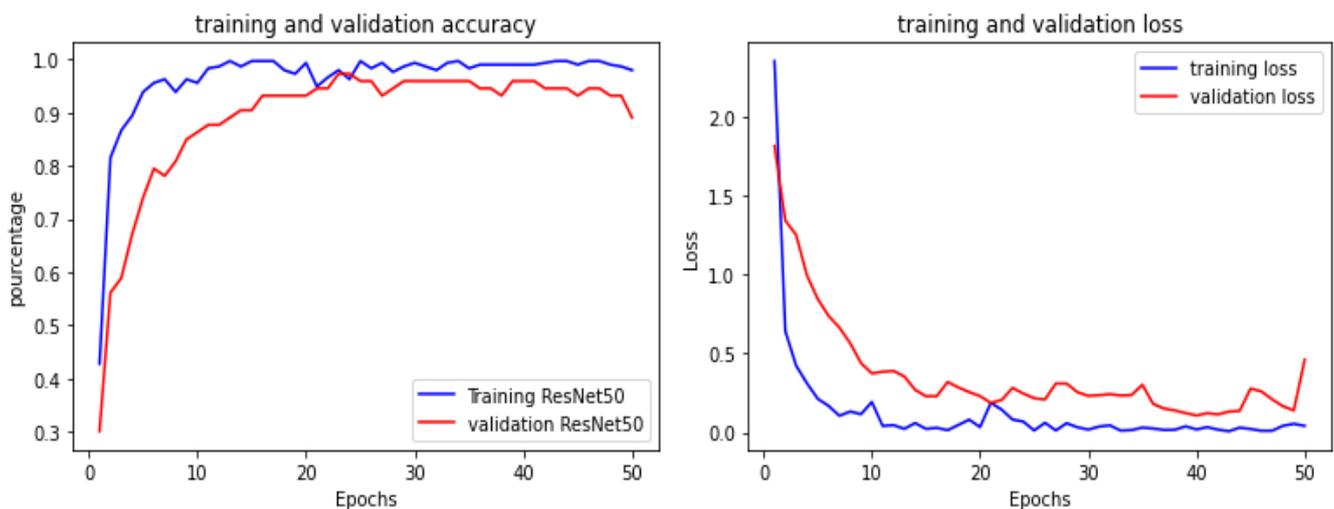


Figure 4.18 : Exactitude et perte du modèle ResNet-50 en utilisant la base (Breakhis_V1)

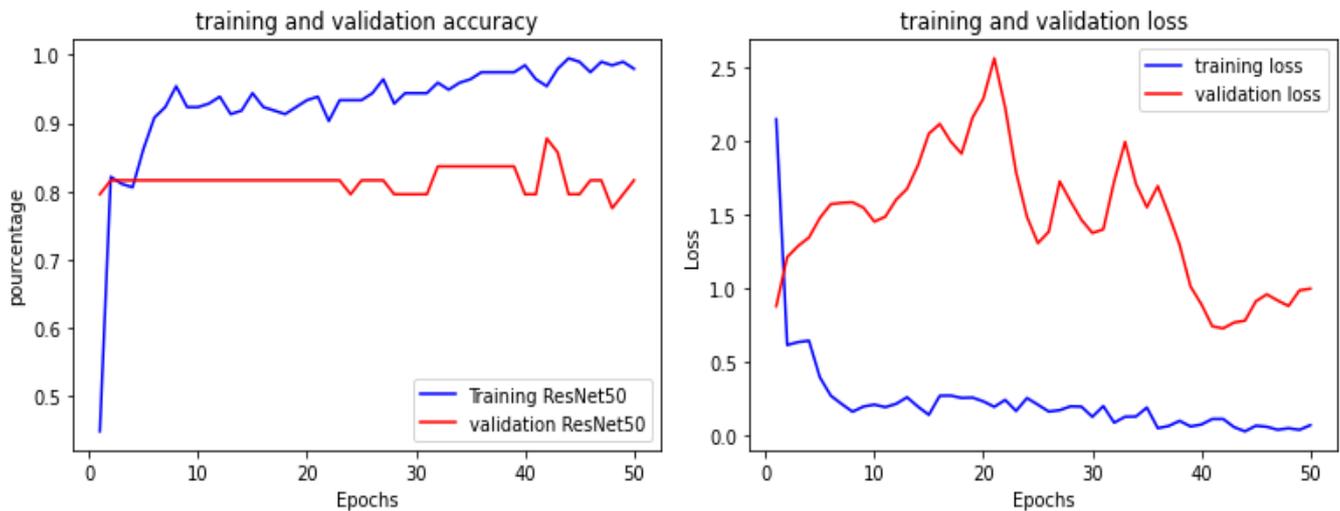


Figure 4.19 : Exactitude et perte du modèle ResNet-50 en utilisant la base (IDC)

4.5.4 Modèles DenseNet-201 et DenseNet-121

L'architecture des deux modèles DenseNet (figures 4.20, 4.21 et 4.22), est composée de la même façon ; la seule différence c'est au nombre de couches convolutives. Ainsi, pour DenseNet-201 comprend les couches suivantes :

- Une couche convolutive (7x7) ;
- 58 couches convolutives (3x3) ;
- 61 couches convolutives(1x1) ;
- Couches AvgPool et max pooling ;
- Une couche entièrement connectée « Fully Connected » .

Les couches ajoutées suivant un fine tuning sont les suivantes :

- Une couche « flatten »
- Une couche « Dropout » de 70%.
- Une couche « fully connected » qui contient 3 nœuds (parce qu'on a 3 classes), avec une fonction d'activation softmax.

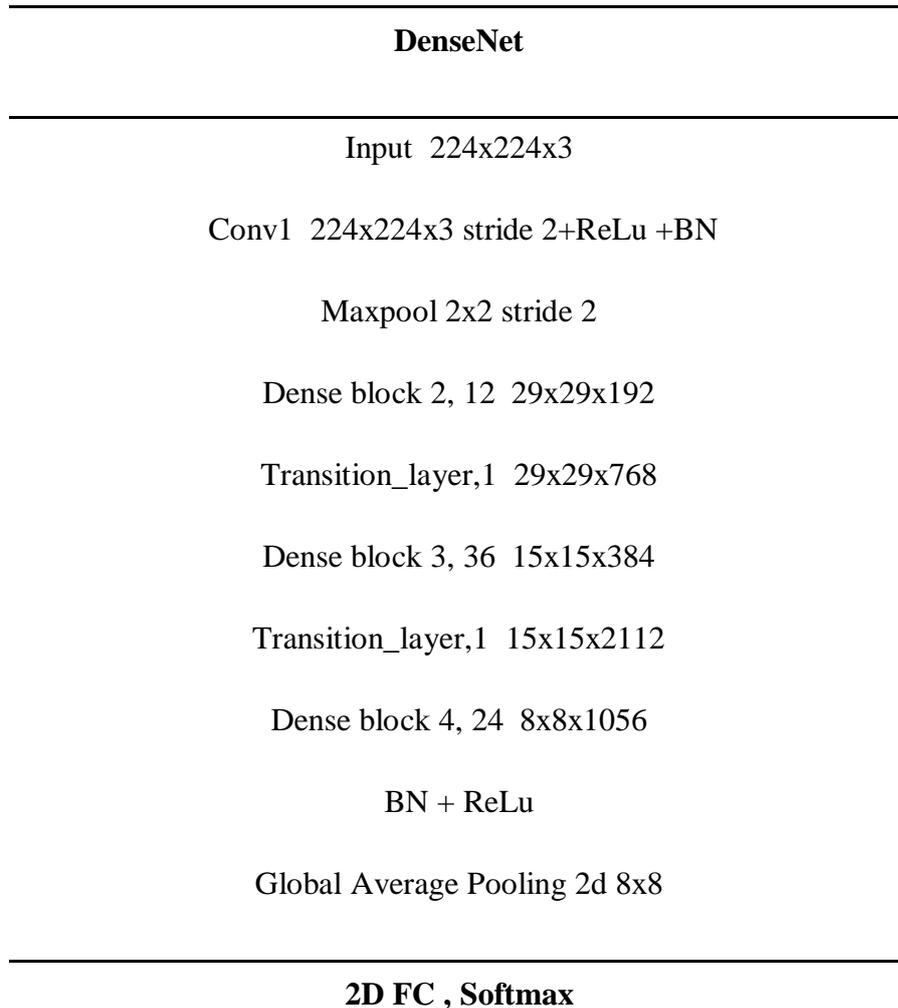


Figure 4.20 : Architecture DenseNet adaptée aux bases histologiques

Layer (type)	Output Shape	Param #
densenet201 (Functional)	(None, 1, 1, 1920)	18321984
flatten_2 (Flatten)	(None, 1920)	0
dropout_2 (Dropout)	(None, 1920)	0
dense_3 (Dense)	(None, 3)	5763

=====
 Total params: 18,327,747
 Trainable params: 18,098,691
 Non-trainable params: 229,056
 =====

Figure 4.21 : Architecture DenseNet-201

Layer (type)	Output Shape	Param #
densenet121 (Functional)	(None, 1, 1, 1024)	7037504
flatten (Flatten)	(None, 1024)	0
dropout (Dropout)	(None, 1024)	0
dense (Dense)	(None, 3)	3075

=====
 Total params: 7,040,579
 Trainable params: 6,956,931
 Non-trainable params: 83,648
 =====

Figure 4.22 : Architecture DenseNet-121

L'exactitude et la perte obtenus par les modèles DenseNet201 et Densenet121, sont représentés respectivement par les figures 4.23, 4.24, 4.25 et 4.26.

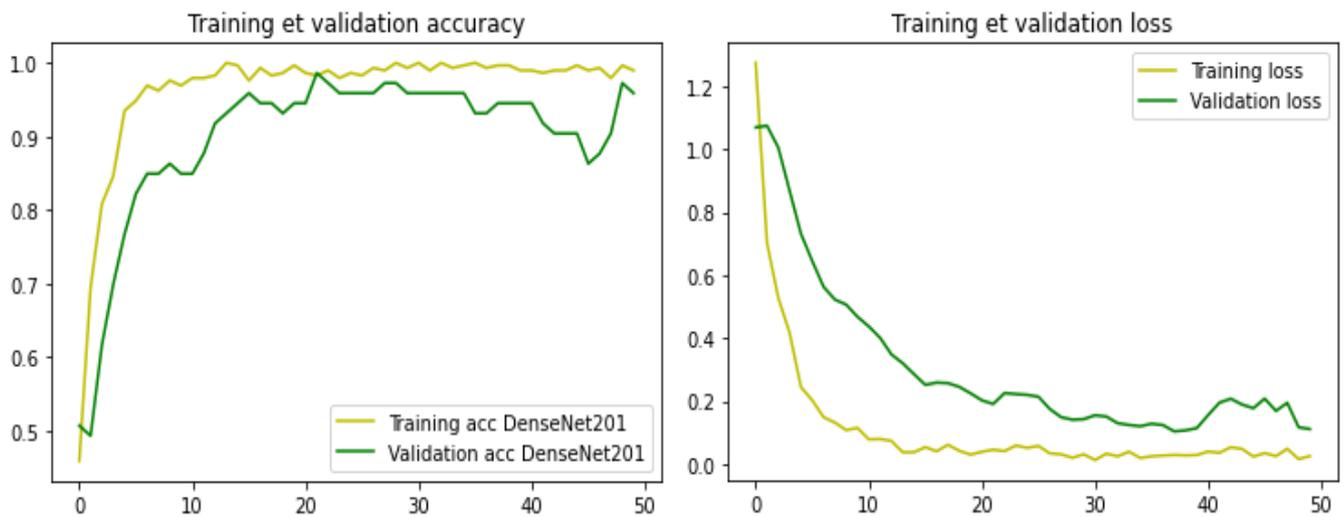


Figure 4.23 : Exactitude et perte du modèle DenseNet-201 en utilisant la base (Breakhis_V1)

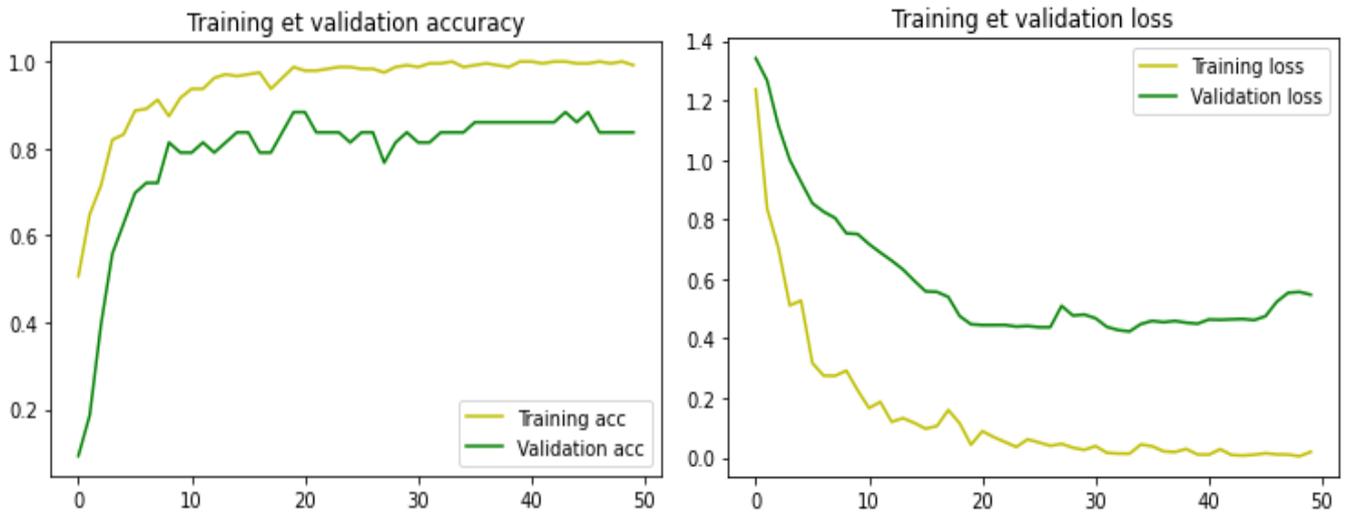


Figure 4.24 : Exactitude et perte du modèle DenseNet-201 en utilisant la base (IDC)

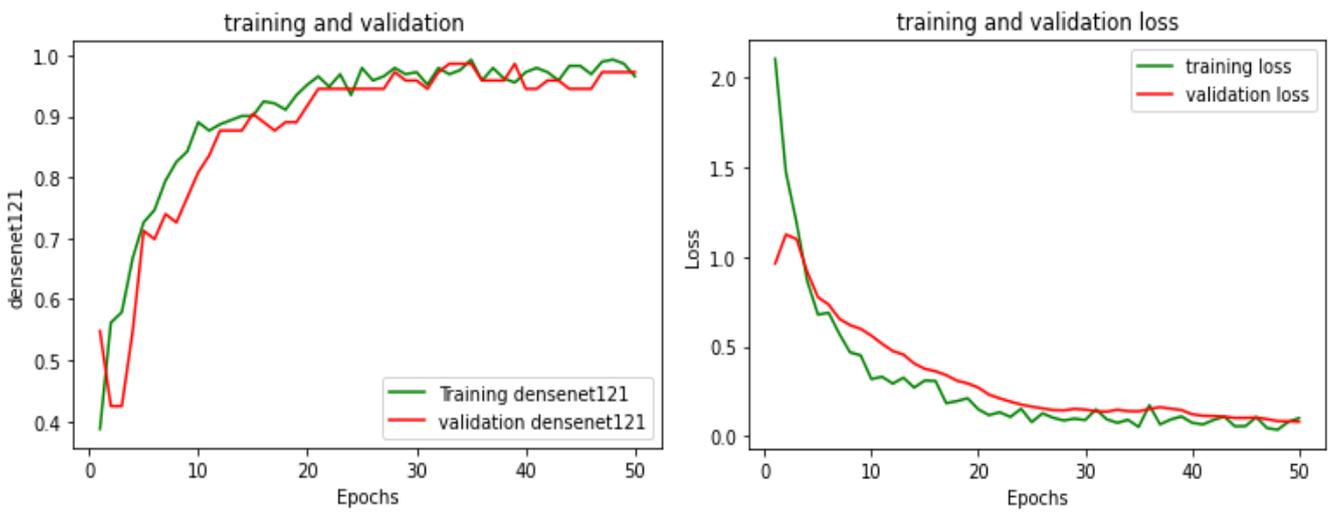


Figure 4.25 : Exactitude et perte du modèle DenseNet-121 en utilisant la base (Breakhis_V1)

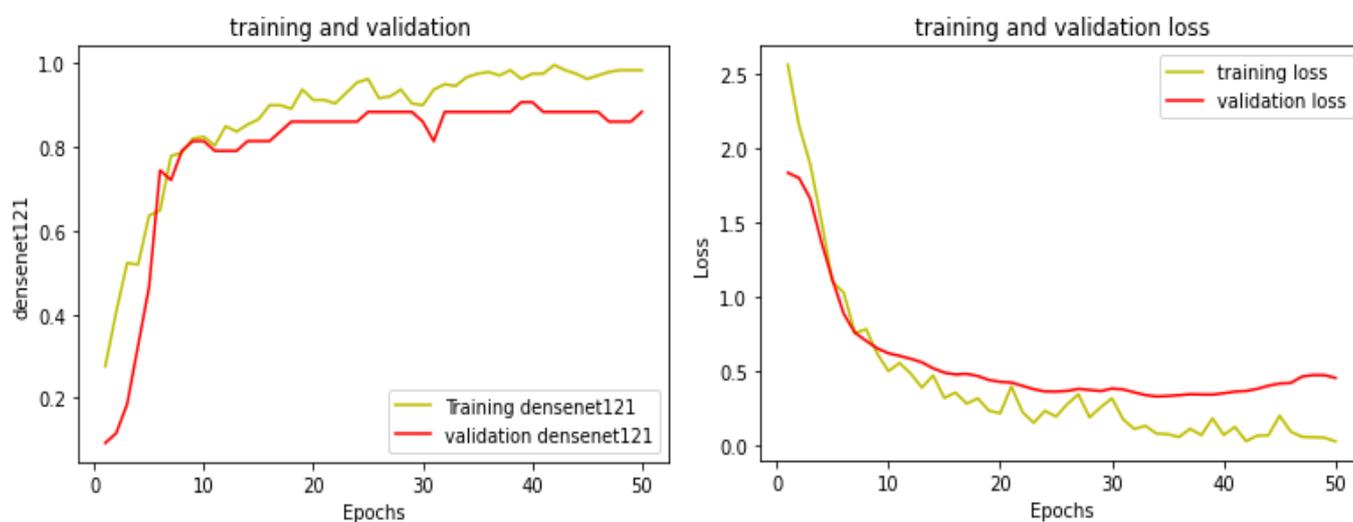


Figure 4.26 : Exactitude et perte du modèle DenseNet-121 en utilisant la base (IDC)

4.6 Discussion et étude comparative des modèles développés

Les différents modèles développés, ont été adaptés aux bases d'images histologiques, suivant un fine tuning. Les paramètres donnés, par le taux d'apprentissage LR, la taille du lot (batch size), les itérations ou les couches de régularisation (dropout), ainsi que les fonctions d'optimisation (Adam et SGBD), ont permis d'éviter le sur-apprentissage et d'atteindre des résultats satisfaisants.

L'apprentissage doit commencer par un taux d'apprentissage relativement important. Car, au début, les poids aléatoires sont loin d'être optimaux, puis le taux d'apprentissage peut diminuer au cours de l'apprentissage, pour permettre des mises à jour de poids. LR=0.0001 montre une performance relativement bonne pour commencer. En effet, le taux d'apprentissage (learning rate) contrôle dans quelle mesure, le modèle doit être modifié en réponse à l'erreur estimée ; chaque fois que les poids du modèle sont mis à jour. Le choix du taux d'apprentissage est un défi, car une valeur trop petite peut entraîner un long processus d'apprentissage qui peut être bloqué, tandis qu'une valeur trop grande peut entraîner l'apprentissage trop rapide d'un ensemble sous-optimal de poids ou un processus d'apprentissage instable.

Le choix de la taille du lot « batch size » est aussi difficile. Plus la taille du lot est élevée, plus le système aura besoin d'espace mémoire.

Plusieurs expériences ont été réalisées lors de l'élaboration du système d'aide à l'interprétation 'BreastCarcicNN', en variant le lot de taille, le nombre d'itérations ainsi que le taux d'apprentissage. Les résultats obtenus (tableau 4.1, figures 4.27, 4.28, 4.29 et 4.30) sur les deux bases histologiques, sont satisfaisants.

➤ **Epoques=50, Batch size=32, LR=0.0001**

Modèles	1 ^{ère} Base de données (IDC grades classification)				2 ^{ème} base de données (breakhis_v1)			
	<i>accuracy</i>	<i>Val-acc</i>	<i>Loss</i>	<i>Val-loss</i>	<i>acc</i>	<i>Val-acc</i>	<i>loss</i>	<i>Val-loss</i>
VGG-16	97%	75%	0.018	0.60	100%	94%	0.01	0.43
DenseNet-201	99%	83%	0.01	0.54	98%	95%	0.025	0.11
DenseNet-121	98%	88%	0.08	0.37	99%	98%	0.07	0.08
ResNet-50	97%	81.6%	0.06%	0.99%	97%	90%	0.04	0.45

Tableau 4.1. Résultats du système 'BreastCarcicNN'

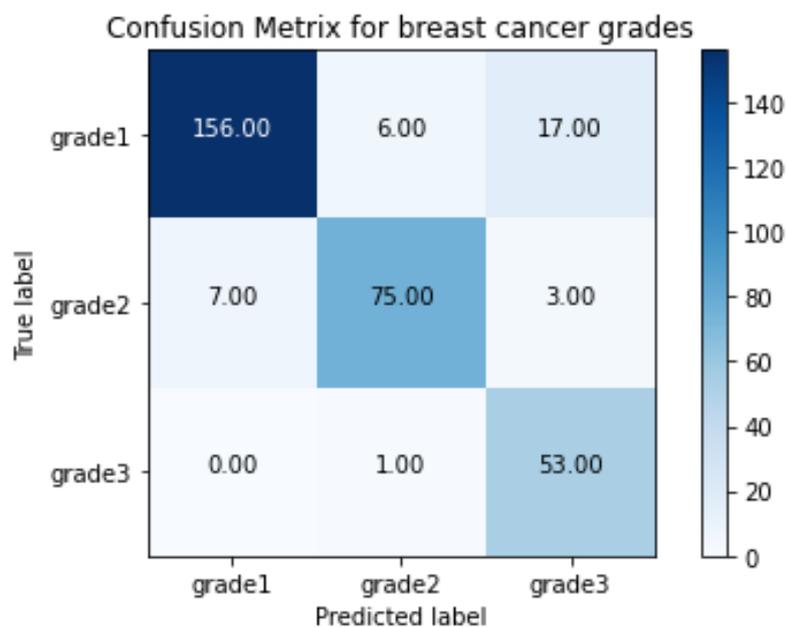


Figure 4.27. Matrice de confusion pour la base de données (IDC)

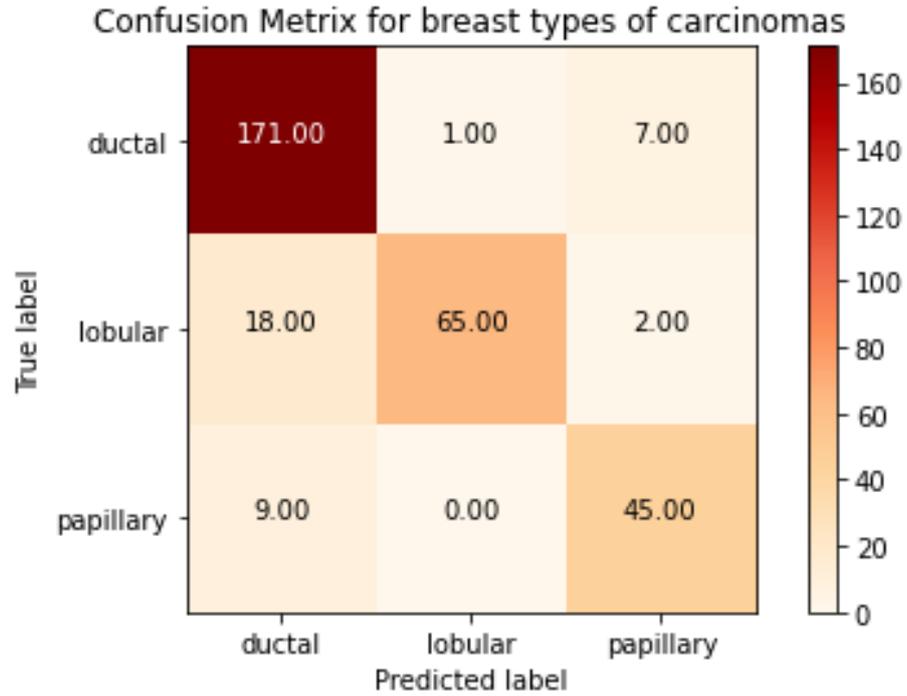


Figure 4.28. Matrice de confusion pour la base de données (Breakhis_v1)

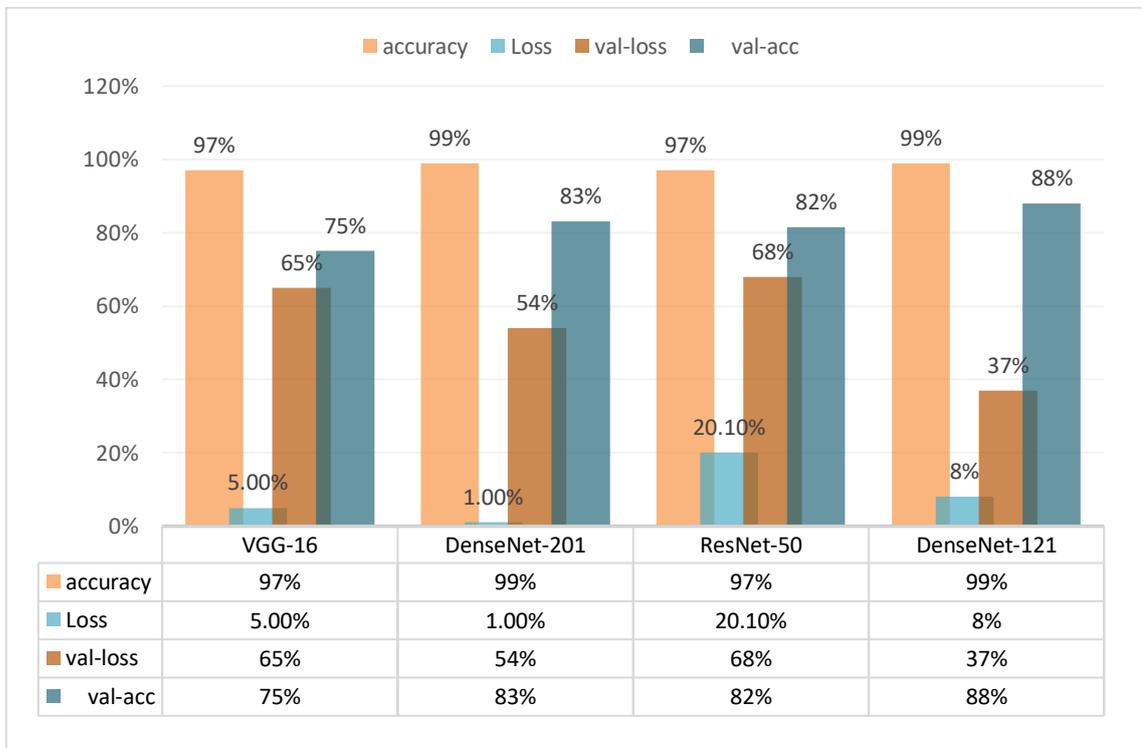


Figure 4.29. Résultats de 'BreastCarcicNN' obtenus pour la 1^{ère} base de données (IDC)



Figure 4.30. Résultats de 'BreastCarcinCNN' pour la 2^{ème} base de données (Breakhis_v1)

Nous remarquons suivant les résultats obtenus, que le modèle le plus adapté à l'analyse des trois grades des carcinomes mammaires (base IDC), est le Densenet 121.

Les échantillons malins et bénins de la base Breakhis_V1, ont été classifiés avec une précision de 100% suivant le modèle VGG-16.

D'autres expériences ont été menées sur le modèle Densenet 121, en changeant la fonction d'optimisation ainsi que le taux d'apprentissage (tableau 4.2) d'une part et en modifiant la taille du lot d'autre part (tableau 4.3), mais en gardant les mêmes paramètres que précédemment.

LR	Epoques	ADAM		SGD	
		<i>accuracy</i>	<i>Val-accuracy</i>	<i>accuracy</i>	<i>Val-accuracy</i>
0.0001	50	98%	88%	99%	75%
	70	99%	90%	99%	79%
0.001	50	94%	78%	97%	90%
	70	99%	78%	99%	83%
0.01	50	75%	60%	60%	81%
	70	78%	58%	75%	79%

Tableau 4.2. Résultats du modèle DenseNet-121 suivant deux optimiseurs (Adam et SGD)

<i>Batch size</i>	<i>Itérations</i>	<i>Accuracy</i>	<i>Val accuracy</i>	<i>loss</i>	<i>Val loss</i>
16	20	98%	91%	0.09	0.4
	40	97%	94%	0.04	0.13
	50	97%	93%	0.05	0.38
	70	97%	82%	0.05	0.41
32	20	96.5%	84.9%	0.10	0.52
	40	97%	93%	0.06	0.12
	50	99%	98%	0.07	0.08
	70	98%	93%	0.05	0.42
64	20	98%	89%	0.05	0.66
	40	98%	94%	0.06	0.13
	50	97%	97%	0.05	0.18
	70	98%	93%	0.06	0.32

Tableau 4.3. Résultats du modèle DenseNet-121 suivant différents batch size (Breakhis_v1)
(LR=0.0001 pour l'optimiseur Adam)

Nous remarquons dans le tableau 4.2, que **l'optimiseur ADAM** converge plus rapidement pour un faible taux d'apprentissage, contrairement à l'optimiseur **SGD**.

Dans le cas des résultats du tableau 4.3, nous remarquons que la taille du lot (batch size) de 32, a donné des résultats parfaits par rapport aux autres tailles.

Dans le cas des travaux de Barsha [114], la classification des différents grades des carcinomes de la base IDC, a été réalisée suivant les modèles Densenet121 et Densenet169. La précision obtenue est de 92.47% pour la première architecture et de 91.43 pour le second réseau. Ceci montre la robustesse du système 'BreastCarciCNN'.

4.7 Conclusion

Dans ce chapitre, nous avons établi avec succès, la classification des images histopathologiques en 3 grades sur deux bases de données, en utilisant divers modèles pré-entraînés de CNN. Ainsi, le DenseNet-121 a donné les meilleurs résultats sur les deux bases ; suivant une exactitude (accuracy) de 98% et une perte de 0.08 sur la 1^{ère} base (IDC), ainsi qu'une exactitude de 99% et une perte de 0.07 sur la 2^{ème} (BreakHis).

Conclusion générale

L'initiation à la recherche réalisée dans le cadre de ce mémoire, s'inscrit dans les objectifs de l'équipe, "IMAD", du département d'électronique du laboratoire LATSI, pour la conception de systèmes d'aide au diagnostic en imagerie médicale, plus particulièrement le cancer du sein. Cette pathologie est un problème de santé, pour les femmes du monde entier. Sa détection précoce, peut augmenter le taux de survie des patients. Par conséquent, les systèmes basés sur le diagnostic assisté par ordinateur peuvent aider les pathologistes à détecter et classier la maladie à différents grades.

Dans ce cadre, l'étude effectuée se rapporte à l'élaboration d'un système d'aide à la classification automatique de carcinomes mammaires 'BreastCarcicNN', via quatre modèles des réseaux de neurones convolutifs pré-entraînés : VGG16, Densenet121, Densenet201 et Resnet50.

Le système 'BreastCarcicNN' a été appliqué sur deux bases histologiques différentes. L'étude comparative approfondie des résultats obtenus par les quatre modèles, a montré un système d'aide à l'interprétation encourageant, puisque la précision varie entre 97% et 100%. En effet, l'utilisation des architectures ResNet-50, VGG-16, DenseNet-121 et DenseNet-201 a apporté de résultats satisfaisants, par rapport aux travaux de la littérature scientifique. Néanmoins, le modèle le plus adapté aux bases étudiées, est le DenseNet-121, qui a permis l'obtention d'une précision de 99% ainsi qu'une validation de 98%, en utilisant l'optimiseur ADAM, avec un taux d'apprentissage de 0.0001.

Les travaux futurs possibles, pourraient éventuellement explorer d'autres variantes des modèles Densenet et Resnet, sur la fusion d'images histologiques et cytologiques à différents grades, ou d'autres modalités du cancer du sein.

Bibliographie

- [1] OMS, Cancer, mars 2022, <https://www.who.int/fr/news-room/fact-sheets/detail/cancer> , consulté en Avril 2022.
- [2] Webhelp medica, « L'usage de l'intelligence artificielle dans la santé », 2021, <https://webhelpmedica.com/intelligence-artificielle-dans-la-sante/> , consulté en janvier 2021.
- [3] Nations Unies, « Le cancer du sein est désormais le plus fréquent au monde », <https://news.un.org/fr/story/2021/02/1088502#:~:text=En%20termes%20de%20mortalit%C3%A9%2C%20le,onze%20meurent%20de%20cette%20maladie> , consulté en janvier 2022.
- [4] Institut national du cancer. Anatomie du sein, <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Anatomie-du-sein>, Consulté en janvier 2022.
- [5] L. Dasinieres, « Sein : anatomie, examens et maladies », le journal des femmes, <https://sante.journaldesfemmes.fr/fiches-anatomie-et-examens/2571039-sein-anatomie-examens-et-maladies/>, 2022, consulté en janvier 2022.
- [6] M.D. El Amrani. Anatomie du sein, cours de 2^{ème} année médecine, Faculté de médecine et de pharmacie, Université C. Ayyad de Marakkech, <http://anatomie-fmpm.uca.ma>, consulté en janvier 2022.
- [7] Institut national du cancer. Les maladies du sein - Cancer du sein, <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Les-maladies-du-sein#:~:text=La%20tumeur%20b%C3%A9nigne%20qui%20se,%C3%A9coulement%20du%20mamelon%2C%20la%20gyn%C3%A9comastie> , consulté en janvier 2022.
- [8] Société canadienne du cancer, « carcinome canalaire », [https://cancer.ca/fr/cancer-information/cancer-types/breast/what-is-breast-cancer/cancerous-tumours/ductal-carcinoma#:~:text=Carcinome%20canalaire%20in%20situ%20\(CCIS\)&text=C'est%20le%20type%20le%20plus%20courant%20de%20cancer%20du,ou%20%C3%A0%20d'autres%20organes](https://cancer.ca/fr/cancer-information/cancer-types/breast/what-is-breast-cancer/cancerous-tumours/ductal-carcinoma#:~:text=Carcinome%20canalaire%20in%20situ%20(CCIS)&text=C'est%20le%20type%20le%20plus%20courant%20de%20cancer%20du,ou%20%C3%A0%20d'autres%20organes), consulté en janvier 2022.
- [9] Institut européen de chimie et biologie, « Pathologie du cancer du sein », <https://www.cellbiol.net/ste/alphERCEPTIN1.php>, consulté en janvier 2022.

[10] Estelle.B, « Adénofibrome du sein », Santé sur le Net, <https://www.sante-sur-le-net.com/sante-femme/adenofibrome-sein/>, consulté en janvier 2022.

[11] Programme québécois de dépistage du cancer du sein, « Tumeur phyllode bénigne », <http://www.depistagesein.ca/tumeur-phyllode-benigne/#.Yj3qNefMLIU>, 2020, consulté en janvier 2022.

[12] cea, « L'imagerie médicale », <https://www.cea.fr/comprendre/Pages/sante-sciences-du-vivant/essentiel-sur-imagerie-medicale.aspx>, 2022, consulté en janvier 2022.

[13] société canadienne du cancer, « Mammographie », <https://cancer.ca/fr/treatments/tests-and-procedures/mammography>, consulté en

[14] Programme québécois de dépistage du cancer du sein, « Imagerie par résonance magnétique », 2020, http://www.depistagesein.ca/imagerie-par-resonance-magnetique/#.Yj2bA_bMLIU, consulté en

[15] Québec. Dépistage du cancer du sein , <https://www.quebec.ca/sante/conseils-et-prevention/depistage-et-offre-de-tests-de-porteur/depistage-du-cancer-du-sein/mammographie-de-depistage#:~:text=La%20mammographie%20est%20une%20radiographie,attribuables%20au%20cancer%20du%20sein>, 2020, consulté en janvier 2022.

[16] Institut national du cancer. Les grades du cancer, <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Les-grades-du-cancer>, consulté en février 2022.

[17] C. Montemagno. Développement de radiotraceurs pour l'imagerie phénotypique des cancers du sein métastatiques, thèse de doctorat en signal et imagerie pour la médecine, université de Grenoble Alpes, 2016.

[18] unf3s, « Moyens et objectifs de l'anatomie pathologique en médecine », http://campus.cerimes.fr/anatomie-pathologique/enseignement/anapath_1/site/html/2.html, 2012, consulté en janvier 2022.

[19] Société canadienne du cancer. Classification histologique du cancer du sein, <https://cancer.ca/fr/cancer-information/cancer-types/breast/grading>, consulté en janvier 2022.

[20] Jason Wasserman. Carcinome canalaire invasive du sein, <https://www.mypathologyreport.ca/fr/breast-invasive-ductal-carcinoma/>, 2022, consulté en janvier 2022.

- [21] O. Kerdraon. Pathologie mammaire, histoire naturelle du cancer du sein, classification histologique des lésions mammaires, CHU de Lille, <https://slideplayer.fr/slide/502071/>, consulté en janvier 2022.
- [22] Berexia. L'intelligence artificielle et la médecine, <http://www.berexia.com/intelligence-artificielle/lintelligence-artificielle-et-la-medecine/>, consulté en février 2022.
- [23] Université de Montréal, Faculté de médecine. Les applications de l'intelligence artificielle en médecine (IAM), <https://medecine.umontreal.ca/recherche/priorites-en-recherche/intelligence-artificielle-en-medecine/>, consulté en février 2022.
- [24] M. J. Iqbal, Z. Javed, H. Sadia, I. A. Qureshi, A. Irshad, R. Ahmed et al. Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future, *Cancer Cell Int* 21, 270 2021, <https://doi.org/10.1186/s12935-021-01981-1>.
- [25] jigsaw. Machine learning model vs deep learning-case study approach, <https://www.jigsawacademy.com/machine-learning-model-vs-deep-learning-case-study-approach/>, consulté en Janvier 2022.
- [26] B. Braunschweig. Intelligence artificielle : Les défis actuels et l'action d'Inria- livre blanc, rapport d'activités, pages 1–21, <https://www.inria.fr/actualite/actualites-inria/livre-blanc-sur-l-intelligence-artificielle>, 2016, consulté en février 2022.
- [27] Le blog des octos. Apprentissage par renforcement, de la théorie à la pratique. <https://blog.octo.com/apprentissage-par-renforcement-de-la-theorie-a-la-pratique>, consulté en février 2022.
- [28] C. Beaulac. Intelligence artificielle avec apprentissage automatique pour l'estimation de la position d'un agent mobile en utilisant les modèles de markov cachés, Mémoire de Maîtrise en mathématiques. Université du Québec, Montréal, 2015.
- [29] ICHI.PRO, « une introduction à l'apprentissage pseudo-semi-supervisé pour le clustering non supervisé », 2020, <https://ichi.pro/fr/une-introduction-a-l-apprentissage-pseudo-semi-supervise-pour-le-clustering-non-supervise-276442106059043>, consulté en
- [30] L. Bastien. Le big data. Analyse de données, <https://www.lebigdata.fr/analyse-de-donnees-algorithmes#:~:text=La%20régression%20linéaire%20est%20l,%20ou%20%20variable%20in dépendante%20%20%20,> 2018, consulté en février 2022.
- [31] C. Touzet. Les réseaux de neurones artificiels, introduction au connexionnisme : cours, exercices et travaux pratiques, ouvrage, Université d'Aix Marseille, 1992.

- [32] Techno-science.net. Santiago Ramón y Cajal, <https://www.techno-science.net/glossaire-definition/Santiago-Ramon-y-Cajal.html>, consulté en février 2022.
- [33] Knowledge one. 5 jalons dans l'histoire des neurosciences, <https://knowledgeone.ca/5-jalons-dans-lhistoire-des-neurosciences/?lang=fr>, consulté en février 2022.
- [34] Le manuel merck, « structure classique d'une cellule nerveuse », <https://www.merckmanuals.com/fr-ca/accueil/multimedia/figure/structure-classique-dune-cellule-nerveuse>, consulté en février 2022.
- [35] Margot P, « Perception : qu'est-ce que c'est et à quoi ça sert, DataScientest, <https://datascientest.com/perceptron>, 2021, consulté en février 2022.
- [36] F. Sur. Introduction à l'apprentissage automatique, cours, École des Mines de Nancy, <https://members.loria.fr/FSur/>, 2022, consulté en février 2022.
- [37] B. Maurice. Fonction d'activation, Deeply learning, <https://deeplylearning.fr/cours-theoriques-deep-learning/fonction-dactivation/>, 2018, 26 consulté en février 2022.
- [38] Acervo Lima. Types de fonction d'activation dans ANN, <https://fr.acervolima.com/types-de-fonction-d-activation-dans-ann/>, 2020, consulté en février 2022.
- [39] T. keldenich. Fonction d'activation, comment ça marche ?-une explication simple, Inside Machine Learning, <https://inside-machinelearning.com/fonction-dactivation-comment-ca-marche-une-explication-simple/#:~:text=Smith%20on%20Unsplash,Qu'est%2Dce%20qu'une%20fonction%20d'activation,mani%C3%A8re%20de%20voir%20une%20donn%C3%A9e.>, 2021, consulté en février 2022.
- [40] Jason Brownlee. How to choose an activation function for deep learning, Machine learning mastery, <https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/>, 2021, consulté en février 2022.
- [41] G. Boesch. Deep neural network: the 3 popular types (MLP CNN, and RNN), viso.ai, <https://viso.ai/deep-learning/deep-neural-network-three-popular-types/>, consulté en mars 2022.
- [42] R. Lambert. Le Réseau de Neurones Artificiels ou Perceptron Multicouche, 2018, <https://penseeartificielle.fr/focus-reseau-neurones-artificiels-perceptron-multicouche/>, consulté en mars 2022.
- [43] Natural solutions. Histoire du deep learning, <https://www.natural-solutions.eu/blog/histoire-du-deep-learning>, 2018, consulté en mars 2022.

- [44] E. Charniak. Introduction au Deep Learning, Dunod, 2021.
- [45] T.K. Senthil kumar. Machine learning model vs deep learning-case study approach, jigsaw, <https://www.jigsawacademy.com/machine-learning-model-vs-deep-learning-case-study-approach/>, 2020, consulté en mars 2022.
- [46] Pensée artificielle. Calcul de la convolution, <https://penseeartificielle.fr/focus-reseau-neurones-convolutifs/convolution-11/>, consulté en mars 2022.
- [47] T. keldenish, « CNN et Couche de Convolution, qu'est-ce que c'est ? », Page Web Inside Machine Learning, 2021, <https://inside-machinelearning.com/cnn-couche-de-convolution/>, consulté en mars 2022.
- [48] W. Lee, A. T.M. Lenferink, C. Otto, H. L. Offerhaus. Classifying Raman spectra of extracellular vesicles based on convolutional neural networks for prostate cancer detection, journal of Raman Spectroscopy, pages 1-8, 2019.
- [49] M. Chane. Classification des images médicales : comprendre le réseau de neurones convolutifs (CNN), IMAIOS, <https://www.imaios.com/fr/Societe/blog/Classification-des-images-medicales-comprendre-le-reseau-de-neurones-convolutifs-CNN>, 2021, consulté en mars 2022.
- [50] Openclassrooms. Découvrez les différentes couches d'un CNN, <https://openclassrooms.com/fr/courses/4470531-classez-et-segmentez-des-donnees-visuelles/5083336-decouvrez-les-differentes-couches-dun-cnn>, consulté en mars 2020.
- [51] Elhani djenaihi. Deep Learning 2, Plan réseaux de neurones artificiels Définition Fonction d'activation Fonction de cout Propagation et rétropropagation Algorithme d'optimisation, Slide player, <https://slideplayer.fr/slide/13975436/>, consulté en mars 2022.
- [52] Pensée artificielle. Architecture RNN, <https://penseeartificielle.fr/tout-pour-bien-debuter-en-deep-learning-5/architecture-rnn-2/>, consulté en mars 2022.
- [53] ICHI.PRO. Un aperçu de l'apprentissage par transfert, <https://ichi.pro/fr/transfer-and-ye-shall-learn-un-apercu-de-l-apprentissage-par-transfert-177126688571069>, consulté en mars 2022.
- [54] DataScientest. Transfer learning : qu'est-ce que c'est ? <https://datascientest.com/transfer-learning>, consulté en mars 2022.
- [55] Data science team. Apprentissage automatique, apprentissage par transfert, Data science, <https://datascience.eu/fr/apprentissage-automatique/apprentissage-par-transfert/>, consulté en mars 2022.

- [56] A. Pedraza, J. Gallego, S. Lopez, L. Gonzalez, A. Laurinavicius, G. Bueno. Glomerulus classification with Convolutional Neural Networks, Annual Conference on Medical Image Understanding and Analysis, Communications in Computer and Information Science, vol 723. Springer, Cham. https://doi.org/10.1007/978-3-319-60964-5_73, 2017.
- [57] Datascientest. VGG : en quoi consiste ce modèle ? <https://datascientest.com/quest-ce-que-le-modele-vgg>, consulté en mars 2022.
- [58] N. A. B. M. Kasim, N. H. B. A. Rahman, Z. Ibrahim, N. N. A. Mangshor. Celebrity Face Recognition using Deep Learning, Indonesian Journal of Electrical Engineering and Computer Science, Vol. 12, No. 2, pages 476-481, 2018.
- [59] N. Bayramoglu, J. Kannala, and J. Heikkila, "Deep learning for magnification independent breast cancer histopathology image classification," *Proc. - Int. Conf. Pattern Recognit.*, vol. 0, pp. 2440–2445, 2016, doi: 10.1109/ICPR.2016.7900002.
- [60] Y. Li, J. Wu, and Q. Wu, "Classification of Breast Cancer Histology Images Using Multi-Size and Discriminative Patches Based on Deep Learning," *IEEE Access*, vol. 7, pp. 21400–21408, 2019, doi: 10.1109/ACCESS.2019.2898044.
- [61] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, 2017, doi: 10.1038/s41598-017-04075-z.
- [62] N. A. Barsha, A. Rahman, M.R.C. Mahdy. Automated detection and grading of Invasive Ductal Carcinoma breast cancer using ensemble of deep learning models, *Computers in Biology and Medicine* Vol 139, pages 1-10, 2021.
- [63] I. Cherifi. Apport de l'intelligence artificielle dans la détection du cancer du sein, mémoire de master en électronique, université de Blida 1, 2020.
- [64] N. Salem. Analyse profonde d'images ultrasonores pour la détection du cancer du sein, mémoire de master en électronique, université de Blida 1, 2020.
- [65] L. Seman, I. Raki. Détection de masses mammaires par apprentissage profond d'images échographiques pour l'aide à la décision, mémoire de master en automatique, université de Blida 1, 2021.
- [66] K. Manseur, S. Mebarki. Détection de pathologies mammaires pour l'aide à l'interprétation, mémoire de master en informatique, université de Blida 1, 2020.
- [67] H. Bolhasani, E. Amjadi, M. Tabatabaeian, S. J. Jassbi, « A histopathological image dataset for grading breast invasive ductal carcinomas, <https://data.mendeley.com/datasets/w7jcx7gj6/1>, consulté en mars 2022.

- [68] INVIVOO beyond tech. Augmentation des images pour améliorer les modèles Machine Learning-Partie 1, <https://www.invivoo.com/augmentation-images-modeles-machine-learning-partie-1/#:~:text=L'augmentation%20des%20donn%C3%A9es%20d,des%20images%20d'apprentissage%20disponibles,2021,Consulté en mars 2022.>
- [69] J. Brownlee, «How to Configure Image Data Augmentation in keras», Page Web Machine Learning Mastery, <https://machinelearningmastery.com/how-to-configure-image-data-augmentation-when-training-deep-learning-neural-networks/>, 2019, consulté en mars 2022.
- [70] ImageNet, <https://www.image-net.org>, consulté en mars 2022.
- [71] L. Akhtyamova, A. Ignatov, J. Cardiff. A large-scale CNN ensemble for medication safety analysis, <https://doi.org/10.48550/arXiv.1706.05549>, 2017.
- [72] O. More. VGG-16 : convolutional neural network», 2020, <https://medium.com/@ommore524/vgg-16-convolution-neural-network-bae747a7494a>, consulté en mars 2022.
- [73] A. stiedemann. ResNet : the working behind it, <https://morioh.com/p/18a0e3c6e6c2>, consulté en mars 2022.
- [74] S. H. Tsang. Review : DenseNet-Dense Convolutional Network (Image Classification), <https://towardsdatascience.com/review-densenet-image-classification-b6631a8ef803>, consulté en mars 2022.
- [75] K. Patel. Architecture comparison of AlexNet, VGGNet, ResNet, Inception, DenseNet, <https://towardsdatascience.com/architecture-comparison-of-alexnet-vggnet-resnet-inception-densenet-beb8b116866d>, 2020, consulté en mars 2022.
- [76] M. Chane. Deep learning en imagerie médicale : stratégies d'apprentissage pour les modèles, Page Web IMaios, <https://www.imaios.com/fr/Societe/blog/IA-pour-l-imagerie-medecale-strategies-d-apprentissage>, 2021, consulté en mars 2022.
- [77] M. Emilion. Matrice de confusion : comment la lire et l'interpréter ? », Page Web Jedha, <https://www.jedha.co/formation-ia/matrice-confusion#:~:text=La%20matrice%20de%20confusion%20est,celles%20pr%C3%A9dites%20par%20un%20mod%C3%A8le>, consulté en mars 2022.
- [78] C. Tremnlay, C. Côme. Partie 7: Accuracy : définition, calcul et limites, Kobia Devançons l'avenir, 2021, <https://kobia.fr/classification-metrics-accuracy/> consulté en mars 2022.

- [79] Machine Learning Crash Course. Classification : Accuracy, <https://developers.google.com/machine-learning/crash-course/classification/accuracy?hl=fr>, consulté en mars 2022.
- [80] S. Maheshkar. What Is Cross Entropy Loss? A Tutorial with Code, Weights & Biases, <https://wandb.ai/sauravmaheshkar/cross-entropy/reports/What-Is-Cross-Entropy-Loss-A-Tutorial-With-Code--VmlldzoxMDA5NTMx>, 2022 consulté en mai 2022.
- [81] P. Fuchs, P. Poulain. Programmation en Python pour les sciences de la vie, Edition Dunod, 2020.
- [82] Académie de Normandie. Distribution Anaconda pour Python, <https://numerique-sciences-informatiques.discip.ac-caen.fr/distribution-anaconda-pour-python>, consulté en janvier 2022.
- [83] Anaconda. Your data science toolkit, <https://www.anaconda.com/products/individual>, consulté en mars 2022.
- [84] Spyder. Spyder : The Scientific Python Development Environment-Documentation, <https://docs.spyder-ide.org/3/index.html>, consulté en mars 2022.
- [85] Digital Guide IONOS. Keras : une bibliothèque open source pour la constitution de réseaux neuronaux, <https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/quest-ce-que-keras/>, consulté en mars 2022.
- [86] R. Vickery. Beginners Guide to Deep Learning with TensorFlow, <https://towardsdatascience.com/beginners-guide-to-deep-learning-with-tensorflow-ca85969b2f2>, consulté en mars 2022.
- [87] S. Bancal. Introduction à Numpy, <https://enacit.epfl.ch/cours/python/scientifique/numpy.html#numpy>, consulté en mars 2022.
- [88] V. Thibault, Scikit-Learn : Découvrez la librairie Python dédié au Machine Learning, Blog Formation Data Science DataScientest.com, <https://datascientest.com/tout-savoir-sur-scikit-learn>, 2021 ; consulté en mars 2022.
- [89] Pythonlang.dev. Splitfolders, <https://pythonlang.dev/repo/jfilter-split-folders/>, consulté en mars 2022.
- [90] J. Brownlee. How to Configure K-Fold Cross-Validation, Page Web Machine Learning Mastery, 2020, <https://machinelearningmastery.com/how-to-configure-k-fold-cross-validation/> consulté en mars 2022.
- [91] Matplotlib, https://matplotlib.org/stable/api/_as_gen/matplotlib.axes.Axes.plot.html, consulté en mars 2022.

[92] Laboratorio Visao Robotica e Imagem. Breast Cancer Histopathological Database (BreakHis), <https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/>, consulté en juin 2022.