

République algérienne démocratique et populaire ministère de
l'enseignement supérieur et de la recherche scientifique

Université Saad Dahlab-Blida 1

Faculté des sciences

Département d'informatique

Domaine Mathématique et informatique



Mémoire de Master en Informatique

Option : Traitement Automatique de la Langue

Présenté par :

Mohamed Nadhir Tsouri Ben Tsouri

&

Baya Meziane

Thème :

**Détection de tumeurs mammaires via les réseaux de
neurones convolutifs pour l'aide à la décision**

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

Année universitaire 2021-2022

ملخص

الغرض من هذه الأطروحة هو تطوير نظام دعم القرار الذي يسمح بالكشف عن أورام الثدي من خلال استغلال صور الموجات فوق الصوتية. يتم تنفيذ هذا الإدراك من خلال الشبكات العصبية التلافيفية عن طريق نقل التعلم لتصنيف صور قاعدة الثدي المصرية إلى ثلاث فئات: "خبيثة" و "حميدة" و "طبيعية". في هذا السياق ، تم اختيار العديد من النماذج ، على وجه الخصوص DenseNet121 و DenseNet201 و VGG16 و VGG19 و ResNet50 وأخيراً ResNet152.

بعد تقسيم قاعدة البيانات وتدريب النماذج المذكورة ، أظهر تقييم النتائج أن أقل معدل تم الحصول عليه لدقة التدريب هو 70٪ لنموذج ResNet152 ، مقابل دقة 95٪ لنموذج VGG16 من أجل التحقق من الصحة ، لا يزال أدنى معدل متبقي لـ ResNet152 ، كما هو الحال بالنسبة للنماذج الأخرى ، فقد أعطوا جميعاً معدلاً أعلى من 90 ٪. لهذا الغرض ، يعد VGG16 هو النموذج الأنسب للبيانات المدروسة.

الكلمات الرئيسية: صور الموجات فوق الصوتية ، والشبكات العصبية التلافيفية ، ونقل التعلم ، و VGG16 و VGG19 ، و Resnet152 و Resnet50 ، و DenseNet121 و DenseNet201 ، ودعم القرار.

Résumé

Le but de ce mémoire est de développer un système d'aide à la décision, permettant de détecter les tumeurs mammaires, en exploitant les clichés échographiques. Cette réalisation est effectuée par les réseaux de neurones convolutifs par apprentissage par transfert pour la classification d'images de la base mammaire égyptienne en trois classes : "maligne", "bénigne", et "normale". Dans ce cadre, plusieurs modèles ont été retenus, notamment : le DenseNet121, le DenseNet201, le VGG16, le VGG19, le ResNet50 et enfin le ResNet152.

Après le partitionnement de la base de données et l'entraînement des modèles cités, l'évaluation des résultats montre que le plus bas taux obtenu pour la précision de l'entraînement, est de 70% pour le modèle ResNet152, contre une précision de 95% pour le modèle VGG16. Pour la validation, le plus bas taux reste toujours pour le ResNet152, quant aux autres modèles, ils ont tous donnés un taux supérieur à 90%. A cet effet, le VGG16 est le modèle le plus approprié aux données étudiées.

Mots clés : Clichés échographiques, Réseaux de Neurones convolutifs, Apprentissage par Transfert, VGG16 & VGG19, Resnet50 & REsnet152, Densenet121& Densenet201, Aide à la décision.

Abstract

The purpose of this thesis is to develop a decision support system, allowing the detection of breast tumors, by exploiting ultrasound images. This realization is carried out by convolutional neural networks by transfer learning for the classification of images of the Egyptian mammary base into three classes: "malignant", "benign", and "normal". In this context, several models have been selected, in particular: the DenseNet121, the DenseNet201, the VGG16, the VGG19, the ResNet50 and finally the ResNet152.

After the partitioning of the database and the training of the cited models, the evaluation of the results shows that the lowest rate obtained for the accuracy of the training is 70% for the model ResNet152, against an accuracy of 95 % for the VGG16 model. For validation, the lowest rate still remains for the ResNet152, as for the other models, they all gave a rate above 90%. For this purpose, the VGG16 is the most appropriate model for the data studied.

Keywords: Ultrasound images, Convolutional Neural Networks, Transfer Learning, VGG16 & VGG19, Resnet50 & Resnet152, Densenet121& Densenet201, Decision support

Dédicaces

Je dédie ce travail à toute ma famille, à mes parents que j'aime plus que tout.

A mon très cher père Brahim, mon idole qui ne m'a jamais laissé tomber depuis ma venue en ce monde.

A ma très chère maman Salima Fatiha, sans laquelle je n'aurais jamais pu réussir et être qui je suis aujourd'hui.

A mes 2 grandes sœurs Aicha et Affaf, qui ont tout fait pour que je réussisse et qui m'ont tout donner.

A ma petite nièce Céline Mélissa notre petite princesse, qui m'a redonné le sourire même dans les moments les plus durs.

A ma tante adoré Leila, qui depuis mon premier jour à l'université, n'a pas arrêté de me soutenir et de me motiver.

Merci à mon grand-père Kaddour qui a été notre source de bonheur à tous, paix à son âme. Sans oublier mon acolyte Meziane Baya avec qui j'ai beaucoup ris et malgré les hauts et les bas, elle a toujours gardé la tête froide et avec laquelle, nous avons pu réaliser ce projet.

Nadhir

Dédicaces

C'est avec une immense joie que je dédie ce modeste travail, à mon très cher père Ahmed pour son soutien et ses encouragements durant tout le long du chemin.

A ma mère adorée Rabia qui m'a toujours donnée la force, la patience et l'amour pour continuer ma route.

A mes 3 frères Brahim Oussama et Bilel qui ont été l'épaule sur laquelle je m'appuyais à tout instant.

A ma nièce et mes neveux Malek, Mohamed et Amir, mes anges adorés que j'aime par-dessus tout.

A ma belle-famille Hamoul et spécialement, à mes beaux-parents.

A mes sœurs Lizia Meriem et mon frère Zoubir et leurs adorables enfants Nazim et Iyad.

Je remercie mon collègue Nadhir pour tous ces efforts déployés et endurent toutes les difficultés afin d'arriver à ce travail.

Je tiens à dédier ce travail spécialement à mes meilleurs amis YaYa, Wiam, Faiza, Kader, Said, Riadh et Walid.

Et enfin nous gardons le meilleur pour la fin, je dédie ce travail à mon mari Yacin, pour la confiance qu'il a mise en moi, pour la stabilité et le courage, qu'il m'a donné même dans les pires moments.

Baya

Remerciements

Arrivant à l'aboutissement du travail, des remerciements s'imposent. Nous tenons tout d'abord, à remercier Allah notre dieu bienveillant qui nous a donné la force le courage et la patience d'arriver là où nous sommes aujourd'hui.

Nous tenons à remercier Dr. Reguieg et Pr. Benblidia nos promotrices, qui ont cru en nous dès le début et qui nous en ont été d'une aide précieuse, dans l'élaboration de ce mémoire grâce à leur encadrement de qualité. Merci à elles de nous avoir donné les moyens de réaliser ce travail.

Nous remercions les honorables membres du jury qui nous font l'honneur de juger ce travail, un jury d'experts dans le domaine de ce mémoire.

Merci à nos professeurs qui nous ont accompagné tout le long du chemin. Un grand merci à Dr. Mezzi qui a été un vrai mentor pour nous, durant les 2 dernières années de Master. Elle a été d'un soutien moral énorme et d'une grande aide. Merci pour tous les bons conseils et vos encouragements.

Merci à nos familles pour leurs soutiens dans chaque étape, nos pères, nos mères, nos sœurs, nos frères, sur qui nous avons toujours pu compter et qui, nous ont suivis dès notre premier pas.

Un très grand merci à tous nos amis qui nous ont aidés dans ce travail, ainsi qu'à chaque personne ayant participé de près ou de loin au bon déroulement de ce mémoire.

Table des matières

I. Contexte Médicale

1.1 Introduction.....	2
1.2 Anatomie du sein.....	2
1.3 Introduction au cancer du sein.....	4
1.3.1 Symptômes.....	5
1.3.2 Causes.....	5
1.4 Diagnostic clinique.....	5
1.4.1 Interrogatoire.....	6
1.4.2 Examen clinique.....	6
1.4.3 Classification clinique et diagnostic.....	7
1.5 Système d'acquisition en imagerie mammaire.....	7
1.5.1 Mammographie.....	8
1.5.2 IRM.....	9
1.5.3 Echographie.....	9
1.5.4 Prélèvement à visée histologique.....	10
1.6 Anomalies mammaires.....	11
1.6.1 Masses visibles sur une mammographie.....	11
1.6.2 Calcifications.....	12
1.7 Classification des clichés mammographiques selon ACR.....	14
1.8 Classification BI-RADS ultrasonore.....	15
1.9 Sémiologie écho-mammographique.....	16
1.10 Limites de l'échographie.....	18
1.11 Conclusion.....	18

II. Etude De L'intelligence Artificielle Via Le Machine Learning Et Le Deep Learning

2.1 Introduction.....	20
2.2 Introduction à l'intelligence artificielle.....	20
2.3 Présentation du Machine Learning.....	21
2.4 Introduction aux réseaux de neurones classiques.....	23
2.4.1 Principe des réseaux de neurones artificiels classiques.....	23
2.4.2 Famille des réseaux à propagation avant ou feed-forward.....	24
2.4.3 Réseaux autoorganisés : les cartes autoorganisées.....	26
2.5 Introduction au Deep Learning.....	26
2.5.1 Historique du Deep Learning.....	27
2.5.2 Différence entre le Deep Learning et le Machine Learning.....	29
2.6 Modèles des réseaux de neurones en apprentissage profond.....	30
2.6.1 Réseaux de neurones récurrents.....	30
2.6.2 Réseaux de neurones convolutifs.....	31
2.7 Fonctions d'activation.....	39
2.7.1 Fonction Relu.....	39
2.7.2 Tangente hyperbolique.....	39
2.7.3 Fonction sigmoïde.....	40
2.7.4 Fonction Softmax.....	40
2.8 Apprentissage par transfert.....	40
2.9 Outils de mesure de performances.....	41
2.9.1 Matrice de confusion.....	41
2.9.2 Critères d'évaluation.....	42
2.10 Conclusion	43

III. Conception D'un Système D'aide Au Diagnostic Du Cancer Du Sein Dans Les Images Echographiques

3.1	Introduction.....	44
3.2	Description de quelques travaux en imageries mammaire.....	44
3.3	Présentation du système conceptuel.....	45
3.3.1	Base de données.....	46
3.3.2	Préparation des données.....	46
3.3.3	Entraînement et validation.....	47
3.3.4	Test et décision.....	47
3.4	Modèle des réseaux de neurones adoptés.....	47
3.4.1	Modèle VGGNet.....	47
3.4.2	Modèle ResNet50.....	51
3.4.3	Avantages de ResNet50.....	54
3.4.4	Modèle de DenseNet.....	54
3.5	Fine tuning.....	59
3.5.1	Fine tuning total.....	60
3.5.2	Fine tuning partiel.....	60
3.5.3	Fonctionnement du fine tuning.....	60
3.6	Conclusion.....	61

IV. Implémentation D'un Système D'aide Au Diagnostic Du Cancer Du Sein Dans Les Images Echographiques

4.1	Introduction.....	62
4.2	Environnement de travail.....	62
4.2.1	Matériel utilisé.....	62
4.2.2	Langage de programmation Python.....	63
4.2.3	Google Colab.....	63
4.3	Implémentation du système 'BreastUltraCNN4'.....	65
4.3.1	Description du Data-set.....	65
4.3.2	Préparation des données.....	65
4.4	Résultats et interprétation.....	67
4.4.1	Résultats de VGG16 et VGG19.....	67
4.4.2	Résultats de ResNet50 et ResNet152.....	72
4.4.3	Résultats de DenseNet201 et DenseNet121.....	76
4.4.4	Etude comparative.....	81
4.5	Comparaison avec les travaux antérieurs.....	81
4.6	Conclusion.....	82

Liste des figures

Figure 1.1 : Anatomie du sein

Figure 1.2 : Ganglions Lymphatiques

Figure 1.3 : Mammographe

Figure 1.4 : Cliché d'une mammographie

Figure 1.5 : IRM du sein

Figure 1.6 : Echographe utilisé pour un dépistage + Cliché correspondant

Figure 1.7 : Mammographie de dépistage avec quatre incidences

Figure 1.8 : Classification des microcalcifications

Figure 1.9 : Formes des calcifications mammaires selon la classification de Le Gal

Figure 2.1 : Intelligence artificielle et ses sous domaines

Figure 2.2 : Structure d'un neurone biologique

Figure 2.3 : Structure d'un neurone artificiel

Figure 2.4 : Architecture du perceptron simple

Figure 2.5 : Architecture du perceptron multicouche

Figure 2.6 : Exemple de l'augmentation du nombre de couches et du nombre de neurones

Figure 2.7 : Deux exemples de cartes autoorganisées

Figure 2.8 : Réseaux de neurones récurrents

Figure 2.9 : Architecture classique d'un réseau de neurones convolutif

Figure 2.10 : Exemple d'une convolution avec un noyau de taille 3*3 et un pas de 2

Figure 2.11 : Opération de pooling avec un noyau MaxPool de taille 2*2 et d'un pas de 2

Figure 2.12 : Exemple applicatif des différents pooling

Figure 2.13 : Exemple de Flatten

Figure 2.14 : Schéma d'une couche entièrement connectée avec 6 neurones

Figure 2.15 : Exemple de Dropout

Figure 2.16 : Exemple de convergence pour différents α

Figure 2.17 : Fonction ReLU

Figure 2.18 : Fonction tangente hyperbolique

Figure 2.19 : Fonction sigmoïde

Figure 2.20 : Matrice de confusion

Figure 3.1 : Schéma conceptuel de détection de tumeurs mammaires pour l'aide à la décision

Figure 3.2: Architecture du réseau neuronal VGG16

Figure 3.3 : Architecture de VGG16

Figure 3.4 : Architecture de VGG19

Figure 3.5 : Modèle de VGG19

Figure 3.6 : Bloc résiduel

Figure 3.7 : Architecture de ResNet50.

Figure 3.8 : Un bloc dense à 5 couches avec un taux de croissance de $k = 4$.

Figure 3.9 : DenseNet profond avec trois blocs denses.

Figure 3.10 : Illustration schématique de l'architecture DenseNet-121

Figure 3.11 : Une illustration schématique de l'architecture DenseNet201

Figure 3.12 : Le Fine Tuning

Figure 4.1: Interaction machine/drive/google colab

Figure 4.2 : Exemples d'images de la base BUSI

Figure 4.3 : Répartition de la base

Figure 4.4 : Modèle VGG16

Figure 4.5 : Modèle VGG19

Figure 4.6 : Entraînement pour vgg16

Figure 4.7 : Entraînement pour vgg19

Figure 4.8 : Graphes des résultats de VGG16

Figure 4.9 : Graphes des résultats deVGG19

Figure 4.10 : Modèle ResNet50

Figure 4.11 : Modèle ResNet152

Figure 4.12 : Entraînement de ResNet50

Figure 4.13 : Entraînement de ResNet152

Figure 4.14 : Graphes des résultats de ResNet50

Figure 4.15 : Graphe des résultats ResNet152

Figure 4.16 : Modèle DenseNet121

Figure 4.17 : Modèle DenseNet201

Figure 4.18 : Entraînement du modèle DenseNet121

Figure 4.19 : Entraînement du modèle DenseNet201

Figure 4.20 : Graphe du modèle DenseNet201

Figure 4.21 : Résultats du modèle DenseNet121

Liste des tableaux

Tableau 1.1 : Classification ACR-BIRADS des anomalies mammographies.

Tableau 1.2 : Classification BI-RADS échographique.

Tableau 1.3 : Caractéristiques mammographiques et échographiques d'une tumeur du sein

Tableau 2.1 : Evolution historique du deep learning

Tableau 2.2 : Différence entre le deep et le machine learning

Tableau 3.1 : Détails de l'architecture ResNet50

Tableau 3.2 : Architecture de Densenet121

Tableau 4.1 : Résumé des caractéristiques de la machine utilisée

Tableau 4.2 : Paramètres utilisés et résultats obtenus pour les modèles VGG16 et VGG19

Tableau 4.3 : Paramètres utilisés et résultats obtenus pour les modèles ResNet50 et ResNet152

Tableau 4.4 : Paramètres utilisés et résultats obtenus pour les modèles DenseNet121 et DenseNet201

Tableau 4.5 : Précision obtenue par tous les modèles

Tableau 4.6 : Etude comparative avec les travaux antérieurs

Introduction Générale

1. Contexte du projet

Depuis notre venue sur terre, nous sommes constamment confrontés à divers dangers et maladies. Ainsi, pendant des milliers d'années, l'Homme n'a cessé de chercher des solutions et des remèdes à ses maux. Au fil du temps, plusieurs techniques et outils ont été développés pour aider à soigner, diagnostiquer et administrer les bons remèdes. Il fallait attendre jusqu'au 20^{ème} siècle, avec l'évolution de la science et le boom technologique, le temps de l'informatique et de l'intelligence artificiel que les chercheurs et les médecins, ont travaillé d'arrachepied pour mettre en œuvre des systèmes aptes à détecter, à prédire et aider au diagnostic de maladies dangereuses.

Aujourd'hui nous vivons dans une ère nouvelle, faites d'intelligence artificielle, de super ordinateurs et de robots. Mais malgré toutes les innovations et les avancées technologiques, le cancer en général reste un mystère pour l'être humain. Tant de scientifiques ont proposé des moyens pour éradiquer ce fléau, mais en vain. Cependant un groupe de chercheurs, de médecins et de développeurs ont commencé à mettre au point des techniques informatiques, visant à diagnostiquer les tumeurs à leurs tout début et ainsi, permettre au patient d'avoir une chance de guérir avant que la maladie, ne soit à un stade trop avancé.

L'objectif de ce projet, est le développement d'un système basé sur les réseaux de neurones convolutifs, afin de détecter les tumeurs mammaires à partir d'images échographiques, pour l'aide à la décision.

Les données considérées dans le cadre de ce mémoire, sont basées sur une base publique 'BUSI', qui présente 3 classes : classe maligne, classe bénigne et classe normale.

2. Problématique

'L'erreur est humaine'. Cette phrase résume parfaitement notre problématique qui se trouve être la détection des tumeurs sur des images échographiques mammaires. En effet, les spécialistes peuvent se tromper lors de la consultation des clichés des patients. Ce qui peut induire en erreur le patient, où même lui donner de faux espoirs. Par ailleurs, la confirmation de la présence d'une tumeur, demande un énorme effort de concentration et de maîtrise, ce qui rend cette tâche très ardue.

C'est ce qui a motivé la communauté informatique ces dernières années, à vouloir trouver une solution à ce problème. Ainsi, c'est tout le principe de cette étude qui vise à aider les médecins dans leurs diagnostics et le suivi de leurs patients, pour des résultats plus fiables et un gain de temps pas moins négligeable.

3. Objectif du mémoire

La résolution de la problématique susmentionnée, consiste à mettre au point un système d'aide à la décision, basé sur les techniques du deep learning, suivant les modèles pré-entraînés : VGG16, VGG19, Resnet50, Resnet152, Densenet121 et Densenet201. Ce système représente la solution informatique pour l'aide au diagnostic du cancer du sein, afin de permettre la détection précise, des tumeurs dans les clichés échographiques et donner le bon diagnostic quant au stade du cancer. Ainsi les erreurs humaines seront diminuées et les anomalies seront détectées à un stade précoce, ce qui favorisera le traitement et le rétablissement du patient.

4. Plan du mémoire

Pour mener à bien le travail demandé, le mémoire est organisé comme suit :

- ✓ Le premier chapitre consiste à présenter l'anatomie du sein, définir le cancer du sein et évoquer également, les machines utilisées actuellement en imagerie médicale, pour détecter les tumeurs présentes.
- ✓ Le second chapitre est consacré aux notions de base sur le 'machine learning', aux techniques du deep learning et principalement aux réseaux de neurones convolutifs.
- ✓ Le troisième chapitre couvrira le côté théorique de notre solution. Nous y mentionnerons le schéma conceptuel de notre travail, qui sera détaillé étape par étape. Ensuite, seront décrits les modèles choisis pour la réalisation du système d'aide à la décision.
- ✓ Le quatrième et dernier chapitre consistera à présenter l'environnement du travail réalisé, les résultats obtenus, leur interprétation ainsi que leur comparaison à des travaux antérieurs, élaborés par l'équipe 'IMAD' des laboratoires LATSI et LRDSI. Ce chapitre sera suivi par une conclusion générale et quelques perspectives.

Chapitre 1 Contexte Médical

1.1 Introduction

Le cancer du sein est le cancer le plus fréquent chez la femme. Il est actuellement la première cause des décès féminins de 35 à 64 ans [1]. En effet, 25 % des cancers féminins, sont des cancers du sein. Par ailleurs une femme porteuse d'un nodule même bénin doit bénéficier d'une surveillance à long terme et d'un traitement du déséquilibre hormonal, qui fait le lit des mastopathies bénignes, ce qui permet de prévenir certains cancers.

Ce chapitre est dédié à l'étude du contexte médical du cancer mammaire, en vue de comprendre le domaine étudié de notre mémoire. Dans ce cadre, sont décrits la structure anatomique du sein, les différentes pathologies mammaires, ainsi que les systèmes d'acquisition, permettant leur détection, tels que ; la mammographie, l'IRM, ou l'échographie, le centre d'intérêt de ce travail.

1.2 Anatomie du sein

La fonction biologique du sein (figure 1.1), est de produire du lait afin de nourrir un nouveau-né. Chaque sein contient une glande mammaire (elle-même composée de quinze à vingt compartiments séparés par du tissu graisseux) et du tissu de soutien qui contient des vaisseaux, des fibres et de la graisse.

Chacun des compartiments de la glande mammaire, est constitué de lobules et de canaux. Le rôle des lobules est de produire le lait en période d'allaitement. Les canaux transportent le lait vers le mamelon [2, 3].

Le sein est constitué de graisse, de tissu conjonctif, de glandes et de canaux. Chaque sein repose sur un large muscle du thorax appelé "muscle grand pectoral". Le sein couvre une région assez grande : en hauteur jusqu'à la clavicule et en largeur, de l'aisselle (creux axillaire) jusqu'au milieu du sternum environ. Il est composé de différentes parties :

- **Les ligaments** qui sont des bandes serrées de tissu conjonctif soutenant les seins. Ils traversent le sein de la peau jusqu'aux muscles où ils se fixent au thorax.

- **Les lobules** qui sont des groupes de glandes qui produisent le lait. Chaque sein comporte de 15 à 25 lobules. Les glandes produisent du lait quand elles sont stimulées par les hormones de la femme durant la grossesse.

- **Les canaux lactifères** qui sont des tubes qui transportent le lait des lobules au mamelon.

- **Le mamelon** désigne la région située au centre de l'aréole et d'où sort le lait à une extrémité. Il est aussi appelé "ostium papillaire" ou "papille". Le mamelon est fait de fibres musculaires. Le mamelon est composé de l'aréole et du téton, avec un muscle sphincter qui permet au téton de se contracter. La peau du mamelon est particulièrement fine, afin de laisser passer le lait maternel au moment de l'allaitement.

- **L'aréole** est la surface ronde, rosée ou brunâtre qui entoure le mamelon. Elle contient de petites glandes qui libèrent, ou sécrètent, une substance huileuse qui agit comme lubrifiant pour le mamelon et l'aréole. De taille variable selon les individus, elle a en général un diamètre de 3 cm, mais peut recouvrir entièrement la surface du sein, ou être à peine visible. L'aréole est recouverte de quelques glandes aréolaires, qui protègent le sein contre les infections et le dessèchement.

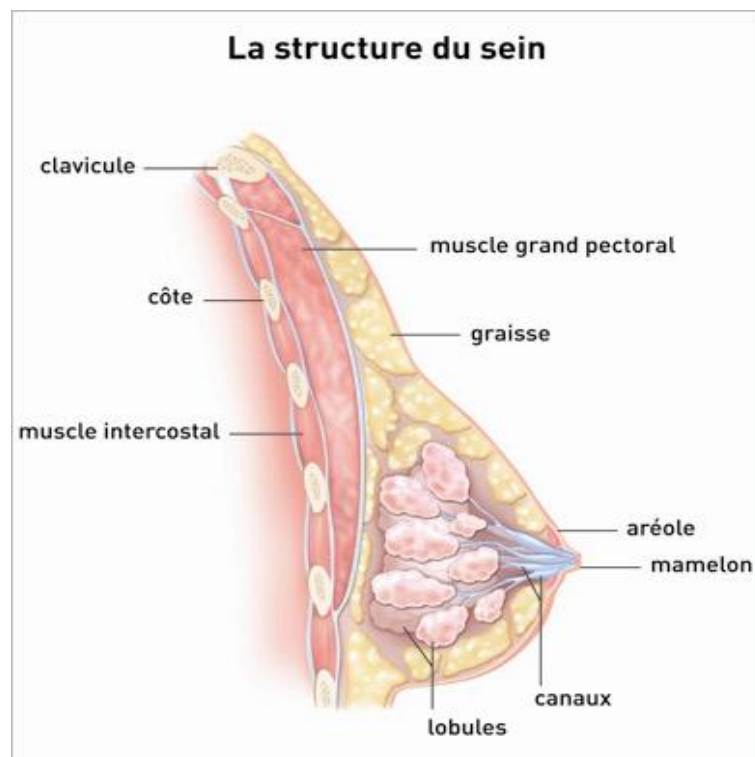


Figure 1.1 : Anatomie du sein [2]

La glande mammaire se développe et fonctionne sous l'influence des hormones, fabriquées par les ovaires [4]. Ces hormones sont de deux types :

- Les œstrogènes, qui permettent notamment le développement des seins au moment de la puberté et jouent un rôle important tout au long de la grossesse (assouplissement des tissus, augmentation du volume sanguin nécessaire à l'alimentation du bébé, etc.) ;
- La progestérone qui joue notamment un rôle dans la différenciation des cellules du sein et sur le cycle menstruel, en préparant par exemple l'utérus à une éventuelle grossesse (densification et développement de la vascularisation la muqueuse de l'utérus).

Le sein est parcouru de vaisseaux sanguins et de vaisseaux lymphatiques (figure 3.2). Les ganglions et les vaisseaux lymphatiques composent le système lymphatique qui aide notamment à combattre les infections. Les ganglions lymphatiques du sein sont principalement situés :

- Au niveau de l'aisselle (ganglions axillaires) ;
- Au-dessus de la clavicule (ganglions sus-claviculaires); sous la clavicule (ganglions sous-claviculaires ou infra-claviculaires) ;
- A l'intérieur du thorax, autour du sternum (ganglions mammaires internes).

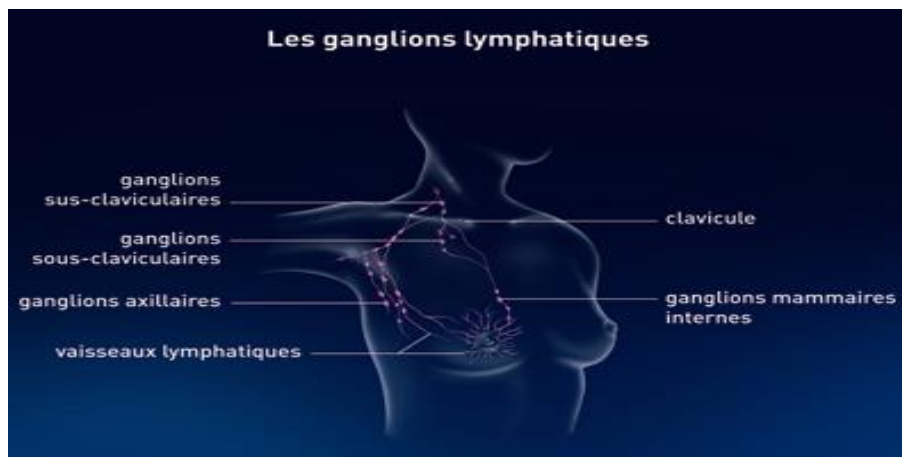


Figure 1.2 : Ganglions Lymphatiques [4]

1.3 Introduction au cancer du sein

Un cancer du sein résulte d'un dérèglement de certaines cellules, qui se multiplient et forment le plus souvent une masse appelée tumeur. Il en existe différents types qui n'évoluent pas de la même manière. Certains sont « agressifs » et évoluent très rapidement, d'autres plus lentement. Les cellules cancéreuses peuvent rester dans le sein. Elles peuvent aussi se propager dans d'autres organes, ce qui est une situation encore plus menaçante. On parle alors de métastases. Dans la majorité des cas, le développement d'un cancer du sein prend plusieurs mois, voire plusieurs années [4, 5].

1.3.1 Symptômes

La majorité des cancers de seins, sont découvertes lors de l'autopalpation. De ce fait, le symptôme principal du cancer du sein, est une grosseur. Les autres changements observés, sont des plaques rouges sur ou autour du mamelon, un changement d'aspect du mamelon (renforcement, etc.) ou une douleur dans le sein, ou l'aisselle.

1.3.2 Causes

On ne sait pas exactement pourquoi les cancers du sein se produisent, mais plusieurs facteurs de risque ont été identifiés, qui rendent plus probable le développement de la maladie. Le risque de cancer du sein, augmente avec l'âge. Avoir dans sa famille une personne qui a eu un cancer du sein, accroît la probabilité d'en avoir un soi-même. Si une femme a eu un cancer du sein, elle a plus de risques, d'en avoir un dans l'autre sein. D'autre part, la présence d'une grosseur bénigne, ou le fait d'avoir des seins denses (forte concentration de cellules), peuvent légèrement augmenter le risque d'avoir un cancer du sein. Parfois, les cellules cancéreuses dans le sein sont stimulées par l'œstrogène. Par conséquent, le risque d'avoir un cancer du sein, est légèrement plus élevé si le cycle menstruel est à un âge précoce, ou si la ménopause est à un âge tardif, parce que l'exposition aux œstrogènes aura duré plus longtemps. Les autres facteurs de risques sont le surpoids, l'obésité, une grande taille, une consommation régulière de fortes quantités d'alcool et la prise d'un traitement hormonal substitutif [5].

1.4 Diagnostic clinique

La découverte d'un nodule mammaire, impose une investigation complète comprenant un interrogatoire, un examen clinique rigoureux et des examens complémentaires [6].

La démarche diagnostique doit être la même devant tout nodule. Elle repose sur :

- L'interrogatoire ;
- L'examen clinique ;
- L'imagerie : mammographie, échographie, IRM, thermographie ;
- La cytoponction ou la ponction-biopsie.

Nous présentons dans cette section, les deux premiers éléments du diagnostic clinique.

1.4.1 Interrogatoire

L'interrogatoire précise :

- Les circonstances du diagnostic ;
- Les facteurs de risque des pathologie mammaires personnels et familiaux ;
- Les antécédents des pathologie mammaires, ponctions ou biopsies mammaires, la durée et la régularité des cycles ou bien la date de la ménopause ;
- Les traitements hormonaux (contraception orale, traitement hormonal substitutif) ;
- Les caractéristiques de la tumeur (date de découverte et évolution durant le cycle et au cours des 6 derniers mois, sensibilité, écoulement mamelonnaire associé, modifications de la forme du sein).

1.4.2 Examen clinique

L'examen clinique est bilatéral et comparatif, de face et de profil.

- A l'inspection : le clinicien recherche différentes déformations mammaires, telles que : la forme et la taille du sein, une asymétrie, une voussure, une ride ou rétraction cutanée, un aspect en peau d'orange, un œdème, des signes inflammatoires, une anomalie du mamelon.
- La palpation : se fait sur les quatre quadrants, y compris la région mamelonnaire, la région para-mammaire et le prolongement axillaire. Ainsi, Si un nodule est palpé, ses caractéristiques sont précisées : le sein est ferme ou dur ; régulier ou non ; bien ou mal limité. Le clinicien recherche une adhérence à la peau, au grand pectoral ou au mamelon, et un écoulement mamelonnaire provoqué par la pression du mamelon, au d'un des quadrants du sein. Il mesure dans ce cas, la taille du nodule, dans son plus grand diamètre et, note le siège précis dans le sein, sur un schéma daté.
- L'examen clinique mammaire se termine par la palpation des creux axillaires et sous-claviculaires. En cas de suspicion de cancer ou de signe fonctionnel particulier, un examen abdominal (palpation hépatique), rhumatologique ou neurologique peut être nécessaire.

1.4.3 Classification clinique et diagnostique

En cas de tumeur manifestement maligne, le clinicien effectue la tumeur présente certaines caractéristiques :

- A l'inspection : voussure ; méplat ; fossette ; déviation ou rétraction du mamelon.
- A la palpation :
 - Tumeur plus ou moins arrondie, irrégulière mal limitée, dure non douloureuse, qui s'accompagne avec modification de la peau.
 - L'adhérence au plan profond est rare.
 - On recherche un écoulement spontané ou provoqué.

En cas d'impression clinique rassurant, on ne remarque aucun signe suspicieux.

- A l'inspection : absence totale de signes cutanés.
- A la palpation : nodule bien limité, ferme mais non dur, sphérique, mobile et indépendant s/p.

Les signes décrits ci-dessus, peuvent indiquer certaines tumeurs bénignes.

- Chez la femme jeune : ces éléments évoquent l'adénome ou l'adénofibrome.
- Chez la femme âgée : surtout si la masse est rénitent d'apparition brutale + une mastodynie : on évoque le kyste qu'il soit solitaire ou multiple.
- Autres nodules bénins :
 - Adénose,
 - Mastite chronique,
 - Cystostéatonécrose post-traumatique.

1.5 Systèmes d'acquisition en imagerie mammaire

L'imagerie médicale est un domaine de la médecine, qui regroupe un ensemble de techniques d'acquisition et de restitution d'images, de différents organes du corps humain. Ce domaine a connu d'importants progrès permettant non seulement un meilleur diagnostic, mais offrant aussi de nouvelles possibilités de traitement pour plusieurs maladies. L'imagerie du sein, est devenue primordiale pour le diagnostic et la caractérisation d'anomalies visualisées en examen de dépistage. La mammographie est la radiographie de base du diagnostic d'anomalies du sein, qui est généralement couplée à une échographie mammaire, pour un examen complémentaire.

D'autres examens peuvent être réalisés, tels que ; l'IRM, l'histologie, etc. Nous présentons dans ce qui suit, quelques systèmes d'acquisition d'images mammaires [7, 8, 9].

1.5.1 Mammographie

Une mammographie est une radiographie des seins, réalisée par un appareil spécifique (Figure 1.3). Elle permet d'obtenir des images de l'intérieur du sein à l'aide de rayons X et de détecter ainsi d'éventuelles anomalies. Autrement dit c'est un examen qui permet d'orienter le médecin dans son diagnostic (peut repérer des nodules, des opacités, des microcalcifications (qui peuvent correspondre à des lésions suspectes) [7].

Une mammographie peut être réalisée soit dans le cadre d'un dépistage du cancer du sein (mammographie de dépistage), soit en présence de symptômes (mammographie de diagnostic).

La mammographie est plus efficace chez la patiente âgée et notamment ménopausée que chez la jeune femme, car le sein est moins dense. Cette caractéristique, permet de présenter une meilleure visualisation des anomalies.

La mammographie standard comprend 3 clichés par sein : un cliché de face, un profil, et un oblique (figure 1.4).

La mammographie réalisée dans le cadre des campagnes de dépistage ne comprend qu'un seul cliché par sein, en général un oblique ; il ne s'agit alors que d'un test de dépistage qui sera confirmé par une mammographie normale (à 6 clichés en cas d'anomalie).



Figure 1.3 : Mammographe [7]



Figure 1.4 : Cliché d'une mammographie [7]

1.5.2 IRM

L'imagerie par résonance magnétique (figure 1.5) est un examen sans radiation, réalisé grâce à un appareil qui produit un champ magnétique très puissant, projeté sur le sein permettant de visualiser avec une très grande précision, la structure intérieure du sein. En effet, l'IRM permet une exploration approfondie des pathologies, dans tous les plans de l'espace. Cet examen est généralement effectué après une biopsie positive pour un cancer, où on a besoin de plus d'informations sur l'étendue de la maladie. Cependant, cette technique ne constitue pas un examen systématique du diagnostic du cancer du sein. Il s'agit plutôt d'un outil utilisé, pour compléter l'investigation de l'imagerie standard (mammographie ou échographie).

Des recherches récentes, ont montré que l'IRM permet de localiser certaines petites lésions mammaires, parfois omises par la mammographie [8].

L'IRM permet en effet, de caractériser des nodules solides ou des surveillances des seins opérés et/ou irradiés ainsi que chez les jeunes femmes qui, ont des tissus mammaires denses.



Figure 1.5 : IRM du sein [8]

1.5.3 Echographie

Le fonctionnement de l'échographie (figure 1.6) est basé sur l'utilisation des ultrasons (avec une fréquence supérieure à 20 000 Hz). L'échographie a pour but d'envoyer des ultrasons dans le corps, puis de les réceptionner afin de permettre d'établir un diagnostic.

Pour effectuer l'émission et la réception des ultrasons, le médecin doit se munir d'une sonde pourvue d'un émetteur et d'un récepteur, puis d'un système informatique afin de transformer les signaux reçus en images pour visualiser la zone étudiée [9].

L'échographie des seins est d'autant plus performante pour la jeune la patiente dont les seins sont denses. Elle est donc très complémentaire à la mammographie.

L'échographie est peu utile dans les seins gras, des patientes ménopausées. L'échographie n'est pas un bon examen de dépistage du cancer du sein, car les images sont trop polymorphes et difficiles à interpréter.

L'échographie trouve donc son intérêt dans les seins très denses, chez la femme de moins de 30 ans, enceinte ou allaitante.



Figure 1.6 : Echographe utilisé pour un dépistage + Cliché correspondant [9]

1.5.4 Prélèvement à visée histologique

Un examen histologique par micro biopsie peut être réalisé, afin de préciser le diagnostic. Ainsi, l'analyse histologique des lésions du sein, représente la lecture au microscope des tissus (prélevés par le chirurgien ou le radiologue), par un médecin anatomo-pathologiste spécialisé dans la lecture des prélèvements du sein.

L'examen histologique permet de classer les lésions cancéreuses, en trois catégories : bénignes, suspectes et malignes [10].

1.6 Anomalies mammaires

Les anomalies mammographiques détectées le plus fréquemment, sont les masses et les calcifications [11, 12, 13].

1.6.1 Masses visibles sur une mammographie

Les anomalies mammographiques détectées le plus fréquemment, sont les masses (figure 1.7). La masse se définit comme une lésion occupant un espace, vue sur deux incidences différentes. Dans le contexte du dépistage, afin de rester centré sur l'image suspecte et non sa nature potentielle, on l'appelle aussi «opacité».

Lorsque l'image n'est visible que sur une seule incidence, on l'appelle alors asymétrie de densité. Les masses se caractérisent par leur forme ronde, ovale, lobulée ou irrégulière, leur contour circonscrit, micro-lobulé, indistinct, masqué ou spiculé et leur densité. L'analyse de ces différentes caractéristiques, permet de déterminer s'il s'agit d'une anomalie bénigne ou suspecte. Si l'anomalie est bénigne, la mammographie de dépistage reste négative. Dans les autres cas, le dépistage est positif et des investigations complémentaires, doivent être entreprises [11, 12].

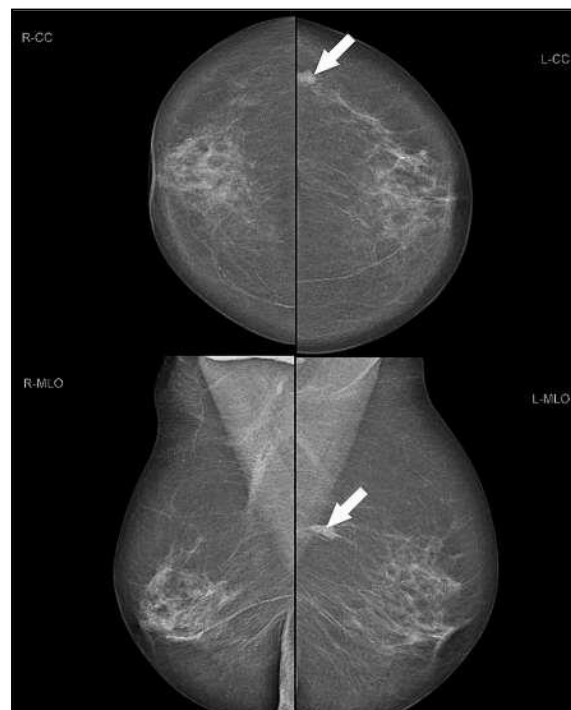


Figure 1.7 : Mammographie de dépistage avec quatre incidences

(RCC = crânio-caudal droit, LCC = crânio-caudal gauche, RMLO = médio-latéral oblique droit, LMLO = médio-latéral oblique gauche). Une opacité bosselée centimétrique est visible dans le quadrant supéro-externe à gauche (flèche) [11]

1. Opacités à contours irréguliers

Les opacités à contours irréguliers, spiculés et notamment les images de type stellaire, sont presque toujours spécifiques du cancer : 95 % d'entre elles, sont malignes [12].

L'opacité stellaire maligne typique, comporte une condensation tissulaire centrale plus ou moins dense et hétérogène, de dimension inférieure à celle de la masse palpée.

Son contour est irrégulier, frangé, spiculé, avec des prolongement opaques linéaires, tentaculaires, rayonnants à distance. La prolifération néoplasique induit une rétraction du tissu péri-tumoral créant un vide apparent péri tumoral (halo clair graisseux) [12].

2. Opacités à contours nets multiples et bilatérales

Certains nodules solides, en apparence bénins, peuvent correspondre à des cancers, de variétés histologiques habituelles ou plus rares (cancers papillaires, mucoïdes ou médullaires) [12].

1.6.2 Calcifications

Les calcifications mammaires sont des dépôts de calcium, qui se forment dans le tissu du sein. Ces calcifications ne sont pas liées à la quantité de calcium qu'une personne absorbe par son alimentation, ou les suppléments qu'elle prend.

Les calcifications mammaires sont assez courantes et on les détecte souvent, lors d'une mammographie de dépistage. Elles apparaissent sous forme de points blancs à la mammographie. Le radiologiste observe la taille, la forme et la disposition des calcifications et note le tout, dans les résultats de la mammographie.

La plupart des calcifications mammaires ne sont pas associées au cancer. Cependant, certaines caractéristiques des calcifications, comme une forme irrégulière ou certains regroupements, peuvent être inquiétantes [12, 13].

Il existe 2 types de calcifications :

1. Macro calcifications

Les macro calcifications sont des dépôts grossiers de calcium dans le sein. Elles sont plutôt fréquentes chez les femmes âgées de plus de 50 ans. On les associe souvent à des changements bénins qui se produisent dans le sein et qui sont causés par : le vieillissement des artères du sein, des lésions aux tissus du sein, l'inflammation des tissus du sein (mastite), des masses non cancéreuses (bénignes) au sein, des dépôts de calcium dans la peau où les vaisseaux sanguins.

Les macro calcifications ont un aspect très distinct à la mammographie. Elles sont habituellement associées, à des affections bénignes. La plupart des macro calcifications ne requièrent pas de biopsie.

2. Microcalcifications

Les différents types de microcalcifications ont été décrits par Le Gal [12], qui propose une classification en 5 stades (figures 1.8 et 1.9).

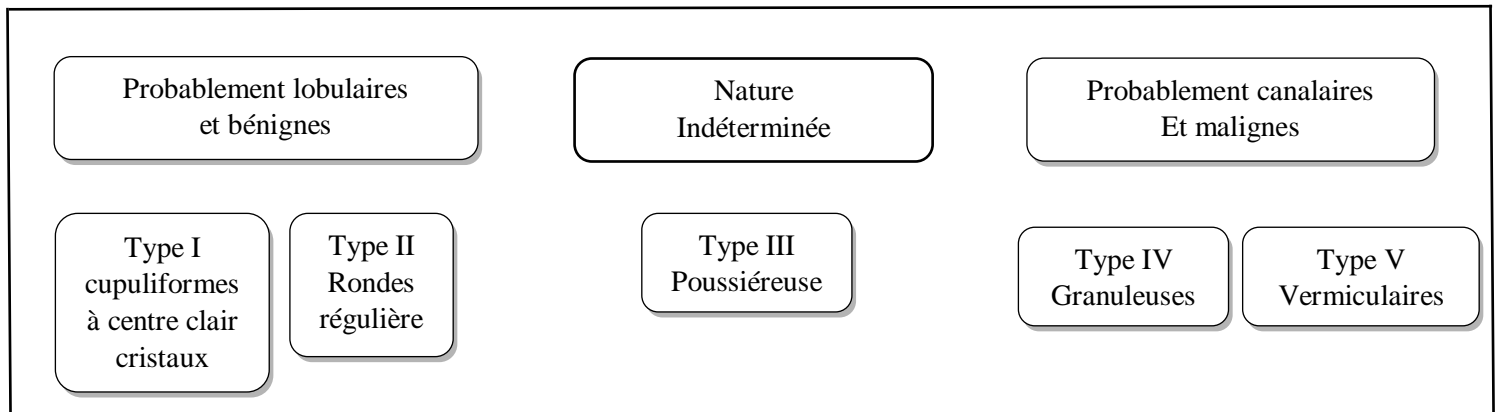


Figure 1.8 : Classification des microcalcifications [12]

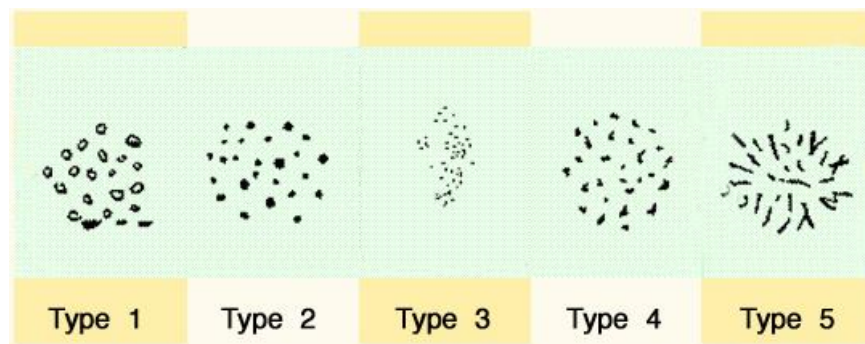


Figure 1.9 : Formes des calcifications mammaires selon la classification de Le Gal [12]

Cette classification permet :

- D'associer à la morphologie des microcalcifications**, un risque croissant de malignité.
Les calcifications suspectes : sont typiquement irrégulières, vermiculaires ou granuleuse, en grain de sel. Elles sont nombreuses, polymorphes, de densité différente de l'une à l'autre et dans une même calcification. Celles qui moulent les embranchements canaliculaires sont très suspectes.
- D'analyser la topographie**. A cet effet, les foyers qui ont une disposition sphérique, sont bénins. Les calcifications canalaire malignes, jalonnent le trajet des galactophores ou occupent un territoire plus ou moins triangulaire ou pyramidal, à

sommet orienté vers le mamelon. La variété histologique qui comporte le plus souvent ces calcifications typiques est le comédocarcinome.

1.7 Classification des clichés mammographiques selon ACR

A l'heure actuelle, les hypothèses diagnostiques des comptes rendus mammographiques standards (tableau 1.1), sont formulées selon le degré de suspicion de la classification ACR (l'American Collège of Radiology) [14].

ACR 0 : Des investigations complémentaires sont nécessaires	Comparaison avec les documents antérieurs, incidences complémentaires, clichés centrés comprimés, agrandissement de microcalcifications, échographie, etc. C'est une classification « d'attente », qui s'utilise en situation de dépistage ou dans l'attente d'un second avis, avant que le bilan d'imagerie soit complété pour une classification définitive.
ACR 1 : Mammographie normale	
ACR 2 : Il existe des anomalies bénignes ne nécessitant ni surveillance ni examen complémentaire	Opacité ronde, avec macrocalcifications, adénofibrome ou kyste, ou ganglion intra mammaire. Opacité(s) ronde(s) correspondant à un (des) kyste(s) typique(s) en échographie Image(s) de densité grasseuse ou mixte (lipome, hamartome, galactocèle, kyste huileux) Cicatrice(s) connue(s) et calcification(s) sur matériel de suture Macro calcification sans opacité (adénofibrome, kyste, adiponécrose, ectasie canalaire sécrétante, calcifications vasculaires, etc.) Microcalcifications annulaires ou arciformes, semi-lunaires, sédimentées, rhomboédriques (calcifications d'aspect carré ou rectangulaire de face, losangiques ou trapézoïdales de profil, à étudier sur des agrandissements) Calcifications cutanées et calcifications punctiformes régulières diffuses.
ACR 3 : Il existe une anomalie probablement bénigne pour laquelle une surveillance à court terme est conseillée	Microcalcifications rondes ou punctiformes régulières ou pulvérulentes, peu nombreuses, en petit amas rond isolé Petit(s) amas rond(s) ou ovale(s) de calcifications amorphes, peu nombreuses, évoquant un début de calcification d'adénofibrome Opacité(s) bien circonscrite(s), ronde(s), ovale(s) ou discrètement polycyclique(s) sans micro lobulation.
ACR 4 : Il existe une anomalie indéterminée ou	Calcification(s), non liquidiennes en échographie Asymétrie focale de densité à limites concaves et/ou

<p>suspecte qui indique une vérification histologique</p>	<p>mélangée à de la graisse. Microcalcifications punctiformes régulières nombreuses et/ou groupées en amas aux contours ni ronds, ni ovales Microcalcifications pulvérulentes groupées et nombreuses Microcalcification irrégulières, polymorphes ou granulaires, peu nombreuses. Image(s) spiculée(s) ou ovale(s) aux contours lobulés, ou masqués, ou ayant augmenté de volume. Distorsion architecturale en dehors d'une cicatrice connue et stable Asymétrie(s) ou surcroît(s) de densité localisé(s) à limites convexes ou évolutif(s)</p>
<p>ACR 5 : Il existe une anomalie évocatrice d'un cancer</p>	<p>Microcalcifications vermiculaires, arborescentes ou microcalcifications irrégulières, polymorphes ou granulaires, nombreuses et groupées. Groupement de microcalcifications quelle que soit leur morphologie, dont la topographie est galactophorique. Microcalcifications associées à une anomalie architecturale ou à une opacité. Microcalcifications groupées ayant augmenté en distribution devenues plus suspectes. Opacité mal circonscrite aux contours flous et irréguliers Opacité spiculée à centre dense.</p>

Tableau 1.1 : Classification ACR-BIRADS des anomalies mammographiques [14]

La classification doit tenir compte du contexte clinique et des facteurs de risque. La comparaison avec des documents anciens, ou le résultat d'investigation complémentaires, peuvent modifier la classification d'une image : une opacité ovale régulière classée ACR 3 mais présente sur des documents anciens, peut être reclassée ACR 2. Quelques calcifications résiduelles après prélèvement percutané contributif bénin d'un amas classé ACR 4, peuvent être reclassées ACR 2 [14].

1.8 Classification Bi-RADS ultrasonore

Les indications de l'échographie, sont à moduler en fonction de la densité des seins et de la catégorie ARC de l'anomalie. Les principales indications de l'échographie sont les anomalies ACR 3 et 4 sur des seins denses [14].

Dans ce cadre, une classification BI-RADS ultrasonore (tableau 1.2) a également été établie pour décrire ces nodules et orienter l'examen, plutôt vers une histologie bénigne ou maligne (tableau 1.2).

BI-RADS 1	Absence de masse	
BI-RADS 2	Bénignité certaine : lipome, ganglion intra mammaire	Masse elliptique / macro lobulée plus large que haute encapsulée
BI-RADS 3	Bénignité probable (2% de risque de cancer) : adénofibrome	
BI-RADS 4	Malignité probable : 4a : de 3 à 49% de risque de cancer 4b : de 50 à 89% de risque de cancer	
BI-RADS 5	Malignité certaine (90% de risque de cancer)	Hypo échogénicité marquée Halo hyperéchogène plus haute que large contours spiculés Micro lobulations Aspect branché Extension intra canalaire Cône d'ombre postérieur Microcalcification

Tableau 1.2 : Classification BI-RADS échographique [14]

L'échographie permet essentiellement, de différencier un kyste d'une structure tissulaire et apporte des arguments supplémentaires de malignité ou de bénignité.

L'échographie ne sert pas à apprécier directement le degré de suspicion, mais à éliminer une lésion radiologiquement latente associée [14, 15].

1.9 Sémiologie écho-mammographique

Il n'existe pas d'aspect échographique parfaitement typique du cancer. Cependant, certaines images sont suspectes comme : une masse hypoéchogène, solide, non compressible, d'aspect lacunaire, à contours irréguliers avec cône d'ombre postérieur ou ayant des dimensions antéro-postérieures supérieures aux dimensions transversales.

Les nodules cancéreux sont en général peu échogènes ; ils ont une écho structure hétérogène. Ils présentent une zone hyperéchogène postérieure (renforcement postérieur).

La mammographie peut montrer un nodule isolé ou des nodules multiples ; avec parfois des calcifications de grande taille, irrégulières, une zone dense inhomogène, parfois réticulée, parsemée de nodules arrondis de taille plus ou moins grosse, parfois associées à des microcalcifications. Elle peut être difficile d'interprétation, lorsque le sein est trop dense.

A l'échographie, les nodules se manifestent par des opacités hétérogènes mais sans atténuation de faisceaux. Ils sont associés à des formations kystiques (masse anéchogène bien limitée, ovalaire ou polylobée, de taille et de topographie variables, avec renforcement postérieur).

Le tableau 2.2, indique l'aspect sémiologique écho-mammographique [14, 15, 16, 17].

	Aspect mammographique	Aspect échographique
Tumeur bénigne	<p>Opacité</p> <ul style="list-style-type: none"> ▪ Régulière ▪ Homogène ▪ Identique à la clinique <p>Signes associés</p> <ul style="list-style-type: none"> ▪ Calcification non suspectes annulaires à centre clair, cupuliformes, punctiformes régulières. ▪ Absence de début de malignité. 	<ul style="list-style-type: none"> ▪ Régulière ▪ Hypoéchogène et homogène ▪ Absence de cône d'ombre +/- renforcement postérieur. ▪ Image horizontale.
Tumeur maligne	<p>Opacité</p> <ul style="list-style-type: none"> ▪ Mal limitée, irrégulière, spiculée ▪ Petite que la masse palpable ▪ Dense et hétérogène <p>Signes associés</p> <ul style="list-style-type: none"> ▪ Calcifications vermiculaires ou calcifications groupées en foyer ▪ Désorganisation architecturale 	<ul style="list-style-type: none"> ▪ Irrégulière ▪ Echogène ▪ Hétérogène ▪ Cône d'ombre ▪ Image verticale.

Tableau 1.3 : Caractéristiques mammographiques et échographiques d'une tumeur du sein [14]

Il existe d'autres signes comme les adénopathies axillaires et, l'épaississement cutané. Ce dernier est dû, soit à l'invasion tumorale directe ou à l'œdème lié aux embolies lymphatiques [13].

1.10 Limites de l'échographie

La capacité de l'échographie à diagnostiquer des lésions malignes et, principalement les cancers infracliniques [15, 16, 17] dépend de plusieurs facteurs, essentiellement :

- ❖ **La taille des lésions** : est un élément déterminant dans la détection des petites masses solides, le seuil de détection pour ces derniers, est supérieur par rapport au kyste.
- ❖ **Le type du sein.** Dans ce cas, le taux des faux négatifs à l'échographie est plus important dans les seins gras, que dans les seins denses, malgré l'amélioration du contraste par les sondes de haute fréquence.
- ❖ **Le type de lésions détectées à l'échographie** (les microcalcifications isolées sont exclues). La capacité de l'échographie est assez satisfaisante dans la détection des cancers, qui se traduisent par des opacités nodulaires et spiculées ou des distorsions architecturales ; alors que dans les asymétries de densité, l'échographie reste limitée.

1.11 Conclusion

Ce chapitre a été consacré principalement à l'anatomie du sein, la présentation des principales pathologies tumorales, en passant par l'aspect sémiologique écho-mammographique, pour comprendre le contexte de ce projet. En effet, la prise en charge d'un nodule du sein, implique de suivre une démarche diagnostique définie, passant par l'interrogatoire, l'examen clinique, des examens d'imagerie, voire des examens à visée histologique.

- La conduite à tenir dépend de l'âge de la patiente et surtout, des données de l'examen clinique.
- L'imagerie de première intention comprend une mammographie et éventuellement une échographie.
- Le cancer du sein est le diagnostic à éliminer, quel que soit l'âge de la patiente.

- Le traitement de ces nodules, dépend de leur nature histologique. En cas de doute diagnostique, l'exérèse chirurgicale est indispensable.

Le cancer du sein est un véritable problème de santé publique mondial par sa fréquence de plus en plus croissante, son impact sur la qualité de vie des patientes et sa gravité. Il s'agit du second cancer le plus fréquent dans le monde et le premier chez la femme.

Chaque année, malgré toutes les précautions prises et les multitudes de traitements contre cette maladie, on constate la perte de milliers de femmes à travers le monde, victimes de ce fléau. Cela a un tel impact sur nos vies, que le monde entier s'est mis d'accord pour instaurer une campagne de sensibilisation, pour le dépistage du cancer du sein : c'est le mois national de la sensibilisation au cancer du sein (*National Breast Cancer Awareness Month*).

Cette campagne, est organisée chaque mois d'Octobre (**Octobre Rose**) dans le but d'accroître la sensibilisation à cette maladie, et récolter des fonds nécessaires à la recherche. On reconnaît facilement son symbole, qui est un ruban rose.

Nous terminons ce chapitre en priant Allah, de protéger toutes les femmes du monde de cette maladie et qu'il les garde, en pleine santé le plus longtemps possible, pour la grande joie de leurs proches.

Le second chapitre est consacré à des notions de base sur l'apprentissage automatique et profond.

Chapitre 2 Introduction au Machine
Et au
Deep Learning

2.1 Introduction

Depuis de nombreuses années, la charge de travail en radiologie, ne cesse de s'accroître en se diversifiant, causant ainsi des erreurs de diagnostic. Devant cette équation menant inévitablement à un problème sanitaire, comme nous l'observons actuellement pendant la crise du Covid, l'IA pourrait être une partie de la solution pour le plus grand bien de tous. Elle permet alors, d'aider les médecins, dans la prise de décision.

L'intelligence artificielle (IA) en imagerie fait partie des grandes évolutions médicales d'aujourd'hui. Elle est utilisée par les radiologues à travers deux axes importants : le **machine learning** et le « **deep learning** », non pas pour les remplacer, comme certains se plaisent à l'imaginer, mais plutôt pour les seconder [18]. Dans le cas de l'apprentissage automatique, c'est l'humain qui apprend à la machine, d'identifier une anomalie. Dans le second cas, on lui fournit des images normales et pathologiques ainsi que des comptes-rendus et c'est la machine, qui apprend seule à reconnaître les images anormales. Le danger du « machine learning » est que l'algorithme ne sait discerner que ce qu'on lui a strictement appris. Il considérera comme une classe normale, une anomalie qu'il n'aura jamais rencontrée.

Dans ce chapitre, nous présentons quelques notions de base, sur l'apprentissage automatique ainsi que les différents modèles de l'apprentissage profond, particulièrement les réseaux de neurones convolutifs.

2.2 Introduction à l'intelligence artificielle

Par intelligence artificielle, on comprend l'acte de faire reproduire par des machines, des tâches qui sont jugées comme complexes par les humains, typiquement : faire raisonner une machine, apprendre à une machine la planification de tâches, ou encore apprendre à une machine à représenter des connaissances d'une manière structurée [19, 20].

L'intelligence artificielle est pluridisciplinaire, car elle croise de nombreux domaines tels que ; l'informatique, la théorie de l'information, le traitement du signal, les statistiques, les neurosciences et la biologie. Il existe plusieurs sous-domaines de l'IA, les plus pertinents sont l'apprentissage automatique et l'apprentissage profond (figure 2.1).

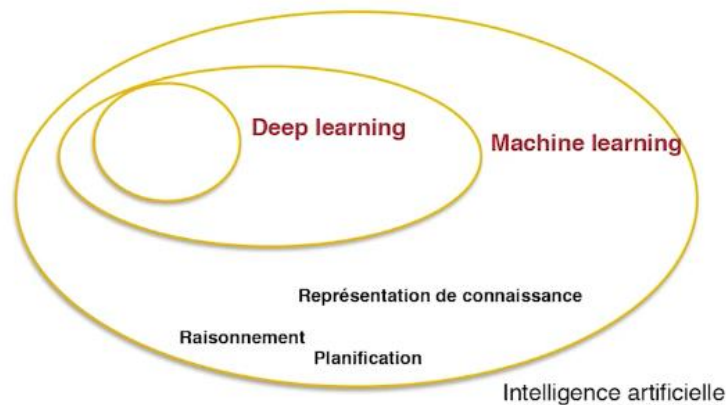


Figure 2.1 : Intelligence artificielle et ses sous domaines [19]

2. 3 Présentation du machine learning

Le machine learning (ML) est le processus qui consiste à utiliser des modèles mathématiques de données, pour aider un ordinateur à apprendre sans instruction directe. Il est considéré comme un sous-ensemble de l'intelligence artificielle. Le Machine Learning utilise des algorithmes pour identifier des patterns dans les données. Ces patterns sont ensuite utilisés, pour créer un modèle de données qui peut effectuer des prédictions. Avec des données et une expérience accrue, les résultats du machine learning sont plus précis, à l'instar de la manière dont les êtres humains s'améliorent avec la pratique [18, 19, 20].

Il existe quatre types d'apprentissage automatique :

1. Machine learning avec supervision

Le machine learning avec supervision est une technologie élémentaire mais stricte. Les opérateurs présentent à l'ordinateur des exemples d'entrées et les sorties souhaitées. L'ordinateur recherche des solutions, pour obtenir ces sorties, en fonction de ces entrées (avec label). Le but recherché est que l'ordinateur, apprenne la règle générale qui mappe les entrées et les sorties [19].

Le machine learning avec supervision peut être utilisé pour faire des prédictions, sur des données indisponibles ou futures (on parle de « modélisation prédictive »).

L'algorithme essaie de développer une fonction qui prédit avec précision, la sortie à partir des variables d'entrée. Par exemple, prédire la valeur d'un bien immobilier (sortie) à partir d'entrées telles que le nombre de pièces, l'année de construction, la surface du terrain, l'emplacement, etc.

Le machine learning avec supervision peut se subdiviser en deux types :

- **Classification** : la variable de sortie est une catégorie.
- **Régression** : la variable de sortie est une valeur spécifique.

Les principaux algorithmes du machine learning avec supervision sont les suivants : forêts aléatoires, arbres décisionnels, méthode du k plus proche voisin (k-NN), régression linéaire, classification naïve bayésienne, machine à vecteurs de support (SVM), régression logistique et boosting des gradients [20, 21].

2. Machine learning sans supervision

Dans le Machine Learning sans supervision, l'algorithme est laissé à lui-même pour déterminer la structure de l'entrée (aucun label n'est communiqué à l'algorithme). Cette approche peut être un but en soi (qui permet de découvrir des structures enfouies dans les données) ou un moyen d'atteindre un certain but. Cette approche est également appelée « apprentissage des caractéristiques » (feature learning) [20].

Il existe deux types de machine learning sans supervision :

➤ **Clustering** : L'objectif consiste à trouver des regroupements dans les données.

➤ **Association** : L'objectif consiste à identifier les règles qui permettront de définir de grands groupes de données.

Les principaux algorithmes du machine learning sans supervision sont les suivants : K-Means, clustering/regroupement hiérarchique et réduction de la dimensionnalité.

3. Machine learning semi supervisé

C'est l'hybridation de l'apprentissage supervisé et non supervisé. Certaines données communiquées à la machine, sont étiquetées et d'autres données ne le sont pas. Cela permet de ne pas laisser de côté et d'utiliser toute l'information [18].

4. Machine learning par renforcement

Dans le Machine learning par renforcement, un programme informatique interagit avec un environnement dynamique, dans lequel il doit atteindre un certain but ; par exemple conduire un véhicule ou affronter un adversaire dans un jeu. Le programme-apprenti reçoit du feedback sous forme de « récompenses » et de « punitions » pendant qu'il navigue dans l'espace du problème et qu'il apprend à identifier le comportement le plus efficace, dans le contexte considéré [20].

Il existe deux types de machine learning par renforcement :

➤ **Monte Carlo** : Le programme reçoit ses récompenses à la fin de l'état « terminal ».

➤ **Machine learning par différence temporelle** : Les récompenses sont évaluées et accordées à chaque étape.

Les principaux algorithmes du machine learning par renforcement sont les suivants : Q-learning, Deep Q Network (DQN) et SARSA (State-Action-Reward-State-Action).

2.4 Introduction aux réseaux de neurones classiques

Les réseaux de neurones, communément appelés des réseaux de neurones artificiels sont des imitations simples des fonctions d'un neurone du cerveau humain (figure 2.3), pour résoudre des problématiques d'apprentissage de la machine.

Le neurone est l'une des principales cellules du système nerveux. En effet, ce sont des cellules excitables, qui peuvent transmettre un message nerveux, autrement dit un signal électrique. Le signal électrique est reçu au niveau des dendrites, activant le neurone [22, 23]. Le signal est ensuite acheminé vers le soma pour se propager à travers l'axone, vers un autre neurone grâce à la synapse.

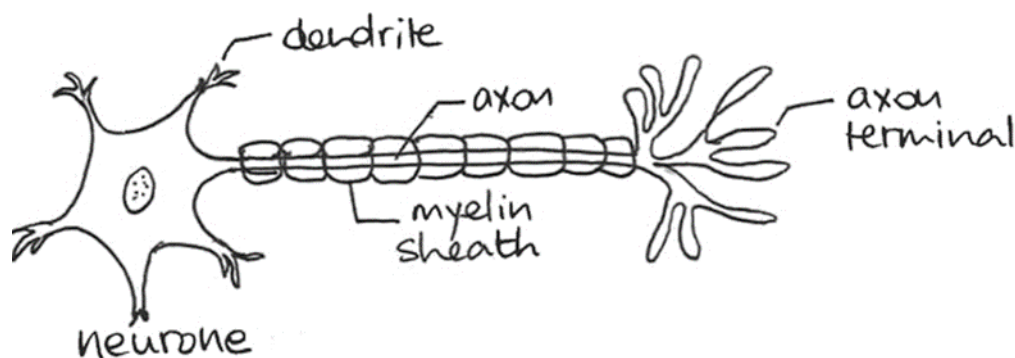


Figure 2.2 : Structure d'un neurone biologique [22]

Les réseaux neuronaux peuvent être appliqués à plusieurs axes, tels que :

- La reconnaissance de formes ;
- La classification de textes ou d'images ;
- l'identification d'objets ;
- la prédiction de données.

2.4.1 Principe des réseaux de neurones artificiels classiques

Plus connus de la communauté machine learning que leur homologues impulsionnels, ces réseaux se composent de neurones artificiels : autrement dit, de neurones à taux de décharges ou à spike "discret" ou explicite. Ces modèles permettent de décrire l'activité (taux de décharge instantané) d'une population de neurones. Le réseau (figure 2.3) reçoit des entrées qu'il module selon plusieurs paramètres (taux d'apprentissage, nombre de neurones), afin de fournir ainsi une réponse [22].

Le réseau de neurones artificiel est un automate qui contient (n) entrées et (m) sorties. Il se présente sous la forme d'un graphe dirigé étiqueté, constitué par un ensemble de nœuds, neurones ou unités, effectuant un calcul simple.

Il se compose aussi d'un ensemble d'arcs associé par des poids variables, reliant deux nœuds. Chaque arc transporte la sortie du nœud en amont vers le nœud en aval, sous la forme d'un signal électrique afin de changer l'état actuel du neurone (activer ou désactiver le neurone).

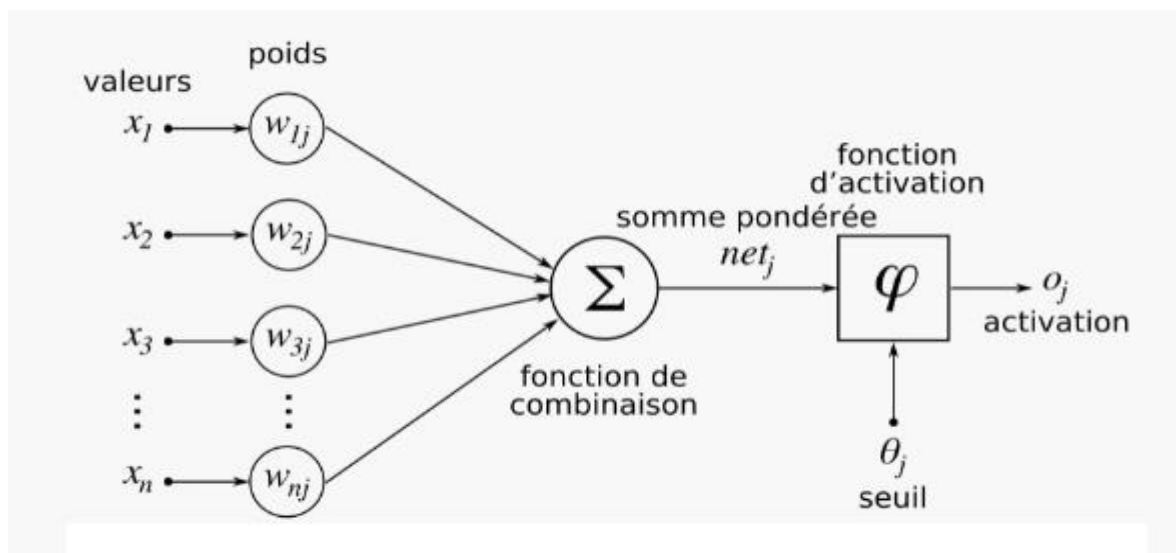


Figure 2.3 : Structure d'un neurone artificiel [23]

2.4.2 Famille des réseaux à propagation avant ou feed-forward

Lorsqu'un réseau propage l'information au travers de ses couches successives, de la couche d'entrée à la couche de sortie, sans avoir un retour de l'information en arrière, alors il s'agit d'un réseau à propagation avant (ou feed-forward). Les exemples les plus connus, sont le perceptron simple et sa version multicouche [23].

1. Perceptron simple

Le perceptron simple est un réseau mono-couche (figure 2.4), acyclique (il ne comporte pas de boucle) et dont la dynamique (l'activité), est déclenchée par la réception en entrée d'information. Ce réseau est dit simple car il ne se compose que de deux couches : une couche d'entrée et une couche de sortie ce qui implique une seule matrice de poids. L'ensemble des unités de la couche d'entrée, sont connectés à celles de la couche de sortie.

Ces structures lui permettent d'être considéré comme un classifieur linéaire : autrement dit il peut classer/séparer les données, selon deux caractéristiques.

- Un neurone possède des entrées ;
- Chaque entrée possède un poids ;

- La sortie est une fonction du poids et des entrées suivant l'équation 2.1.

$$Y = f(W_1 * X_1 + W_2 * X_2) \quad (2.1)$$

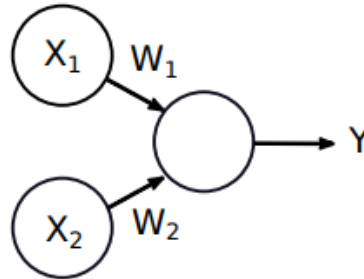


Figure 2.4 : Architecture du perceptron simple [23]

2. Perceptron multicouche

Désigné par le sigle MLP (pour Multi-layer Perceptron), le perceptron multicouche (figure 2.5) se compose d'une couche d'entrée, d'une couche de sortie et d'une ou plusieurs couches cachées.

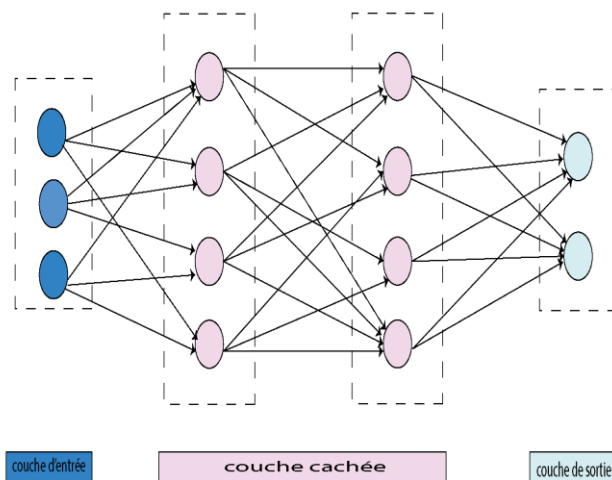


Figure 2.5 : Architecture du perceptron multicouche [23]

Si le réseau possède n couches, alors il possède $n-1$ matrices de poids (une entre chaque suite de couches). Le MLP doté d'une couche cachée, est théoriquement un approximateur universel de fonctions. En théorie, il suffit donc d'ajouter un nombre de neurones suffisant au niveau de la couche cachée (figure 2.6), pour approximer n'importe quelle fonction non linéaire [24]. Dans ce cadre, l'augmentation du nombre de couches et du nombre de neurones, accroît le pouvoir de séparation.

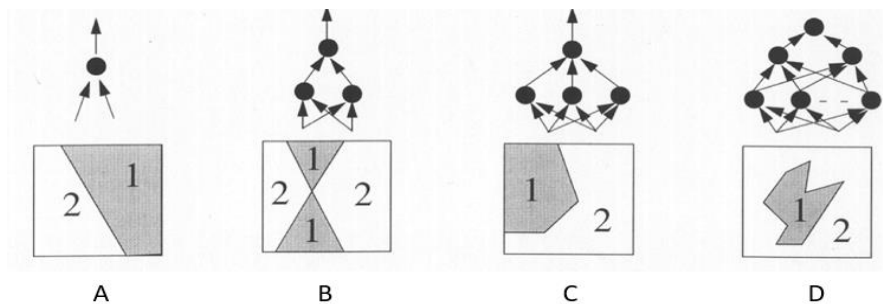


Figure 2.6 : Exemple de l'augmentation du nombre de couches et du nombre de neurones [23]

Les réseaux de neurones à convolution (CNN pour Convolutionnal Neural Network) sont aussi des réseaux feed-forward, qui correspondent, à un empilement de perceptrons multicouches. Ces réseaux sont surtout utilisés pour la reconnaissance d'images, de vidéos ou encore dans le traitement naturel du langage.

2.4.3 Réseaux autoorganisés : les cartes autoorganisées

Inspiré de l'organisation du cortex, les réseaux auto-organisés (figure 2.7) se distinguent par une connectivité locale. Ils sont surtout adaptés pour le traitement d'informations spatiales.

Ces réseaux utilisent des méthodes d'apprentissage non-supervisées. Ils peuvent être utilisés pour cartographier un espace réel, ou encore étudier la répartition de données dans un espace de grandes dimensions comme dans le cas de problème de quantification vectorielle, de clustering ou de classification [23].

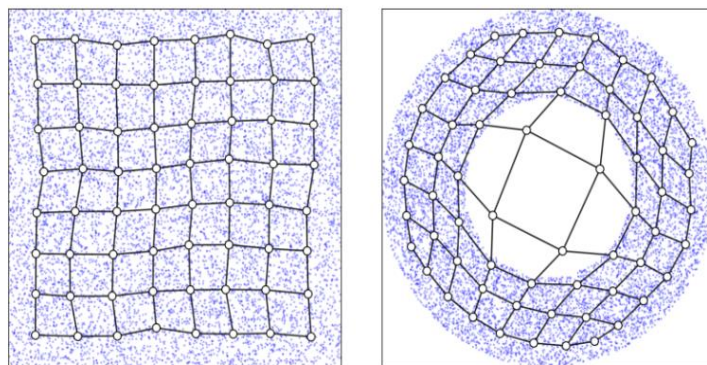


Figure 2.7 : Deux exemples de cartes auto-organisées [23]

2.4 Introduction au deep learning

L'apprentissage profond (deep learning) est basé sur l'idée des réseaux de neurones artificielles, conçu pour traiter de larges quantités de données, en ajoutant des couches au réseau.

Un modèle de ce type d'apprentissage a la capacité d'extraire des caractéristiques, à partir de données brutes grâce aux multiples couches de traitement. Il permet d'apprendre sur ces caractéristiques, à travers chaque couche suivant une intervention humaine, minimale [25, 26].

2.4.3 Historique du deep learning

Le tableau ci-dessous, montre brièvement l'évolution du deep learning [26].

Année	Inventeurs	Invention	Explication
1943	Warren McCulloch Et Walter Pitts	« Invention de premiers neurones artificiels » et plus tard renommée «Threshold logic unit »	Les premiers réseaux de neurones sont inventés par deux mathématiciens et neuroscientifiques qui expliquent comment ils ont pu programmer des neurones artificiels en s'inspirant du fonctionnement des neurones biologique, Ils ont utilisé une combinaison d'algorithmes et de mathématiques qu'ils ont appelées « logique de seuil » pour imiter le processus de pensée.
1950	Frank Rosenblatt	Invention de perceptron	Le psychologue américain Frank inventa le modèle linéaire« perceptron »,le premier algorithme de reconnaissance de formes, basé sur un réseau de neurones informatiques à deux couches. Il utilisa une simple addition et soustraction. Il a également proposé des couches supplémentaires avec des notations mathématiques, mais celles-ci ne seront réalisées qu'en 1975.Grace au perceptron il sera possible de construire des machines

			capable de lire, de parler, de marcher et même d'avoir une conscience.
1986	David Rumelhart	Invention Le perceptron multicouches	Les premiers réseaux de neurones n'étaient pas capables de résoudre des problèmes non linéaires alors David Rumelhart inventa le modèle de perceptron multicouche avec l'utilisation de la technique Back propagation.
1990	Yann LeCun	Réseaux de neurones convolutifs CNN (LeNet)	Le célèbre Yann inventa les premiers réseaux de neurones convolutifs, réseaux qui sont capable reconnaître et de traiter des images, en introduisant au début de ces réseaux des filtres mathématiques qu'on appelle convolution et pooling.
1997	Sepp Hochreiter et Jürgen Schmidhuber	Les réseaux de neurones récurrents (LSTMs)	Les réseaux de neurones récurrents sont une variante du perceptron multicouche qui permet de traiter efficacement les problèmes de séries temporelles comme la lecture de textes ou encore la reconnaissance vocale.
Mid-2000s	Geoffrey Hinton et RuslanSalakhutdinov	Deep learning	Le terme « deep learning » commence à gagner en popularité après qu'un article de Geoffrey Hinton et RuslanSalakhutdinov a montré comment un réseau de neurones à plusieurs couches pouvait être pré-formé une couche à la fois.
2012		L'expérience du chat	AP n'a réellement pris envol qu'en 2012 lors d'une compétition de vision par ordinateur nommée ImageNet, où une équipe de chercheurs menée par Hinton

			développa un réseau de neurones capable de reconnaître n'importe quelle image avec une meilleure performance que tous les autres algorithmes de l'époque.
--	--	--	---

Tableau 2.1 : Evolution historique du deep learning [26]

2.4.4 Différence entre le deep et le machine learning

Le Machine learning est la technologie la plus ancienne et la plus simple. Elle s'appuie sur un algorithme qui adapte lui-même le système à partir des retours faits par l'humain. La mise en place de cette technologie implique l'existence de données organisées. Le système est ensuite alimenté par des données structurées et catégorisées lui permettant de comprendre comment classer de nouvelles données similaires. En fonction de ce classement, le système exécute ensuite les actions programmées. Il sait par exemple identifier si une photo montre un chien ou un chat et classer le document dans le dossier correspondant.

Après une première phase d'utilisation, l'algorithme est optimisé à partir des feedbacks du développeur, qui informent le système des classifications erronées et lui indiquent les bonnes catégories.

L'apprentissage profond n'a pas besoin de données structurées. Le système fonctionne à partir de plusieurs couches de réseaux neuronaux, qui combinent différents algorithmes en s'inspirant du cerveau humain. Ainsi, le système est capable de travailler à partir de données non structurées.

Cette approche est particulièrement adaptée pour les tâches complexes, lorsque tous les aspects des objets à traiter ne peuvent pas être catégorisés en amont. Le système du deep learning identifie lui-même les caractéristiques discriminantes. Dans chaque couche, il recherche un nouveau critère spécifique de l'objet, qui sert de base pour décider de la classification retenue pour l'objet à la fin du processus [27].

Le tableau 2.2 indique les différences entre le 'machine et le deep learning'.

	Machine learning	Deep learning
Organisation des données	Données structurées	Données non-structurées
Base de données	Contrôlable	>1million de données

Entraînement	Entraînement par l'humain nécessaire	Système d'apprentissage autonome
Algorithme	Algorithme modifiable	Réseau de neuronal d'algorithmes
Champ d'application	Actions simple de routine	Tâches complexes
Temps d'entraînement	De quelques minutes à quelques heures	Jusqu'à des semaines

Tableau 2.2 : Différence entre le deep et le machine learning [27]

2.5 Modèles des réseaux de neurones en apprentissage profond

Les réseaux de neurones profonds ont des applications spectaculaires dans des domaines très divers dont la vision par ordinateur, la compréhension de la parole, l'analyse de langages naturels, mais aussi pour la robotique, la prédiction de phénomènes physiques divers et le diagnostic médical [28, 29, 30, 31].

2.6.1 Réseaux de neurones récurrents

Les réseaux de neurones récurrents (ou RNN pour Recurrent Neural Networks), sont des réseaux de neurones dans lesquels, l'information peut se propager dans les deux sens ; de la couche profonde à la première couche.

Ces réseaux (figure 2.8) ont des connexions répétitives dans le sens, où ils gardent des informations en mémoire : ils peuvent considérer à tout moment, un certain nombre d'états passés.

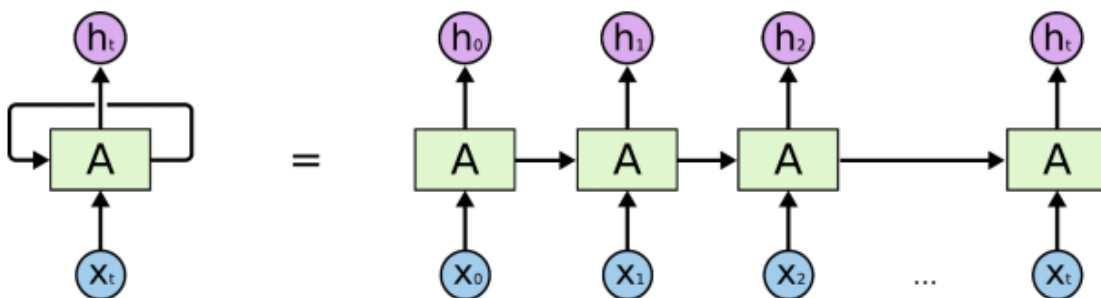


Figure 2.8 : Réseaux de neurones récurrents [29]

2.6.2 Réseaux de neurones convolutifs

La classification d'images a connu une avancée majeure en termes de performance, grâce à l'essor des réseaux de neurones convolutifs (Convolutional Neural Network, CNN). Ces réseaux de neurones artificiels sont capables de catégoriser les informations des plus simples, aux plus complexes. Ils consistent en un empilage multicouche de neurones, des fonctions mathématiques à plusieurs paramètres ajustables, qui pré-treatent de petites quantités d'informations.

On distingue deux parties dans ces réseaux ; la première est la partie convolutive du modèle. Elle permet l'extraction de caractéristiques des images et, la seconde est la classification du modèle constituée de couches entièrement connectées (perceptron multicouches) [30].

L'architecture d'un réseau de neurones convolutifs (figure 2.8) est formée par une succession de blocs de traitement, pour extraire les caractéristiques discriminant la classe d'appartenance de l'image des autres. Un bloc de traitement se compose d'une à plusieurs :

- Couches de convolution (CONV) qui traitent les données d'un champ récepteur.
- Couches de correction (ReLU), souvent appelée par abus « ReLU » en référence à la fonction d'activation (Unité de rectification linéaire) ;
- La couche flatten, la couche qui va transformé la matrice en un vecteur
- Couches de pooling (POOL), qui permet de compresser l'information en réduisant la taille de l'image intermédiaire (souvent par sous-échantillonnage).

Les blocs de traitement s'enchaînent jusqu'aux couches finales du réseau qui réalisent la classification de l'image et le calcul de l'erreur entre la prédiction et la valeur cible :

- Couche « entièrement connectée » (FC), qui est une couche de type perceptron ;
- Couche de perte (LOSS).

La façon dont s'enchaînent les couches de convolution, de correction et de pooling dans les blocs de traitement, ainsi que les blocs de traitement entre eux, font la particularité de l'architecture du réseau [9].

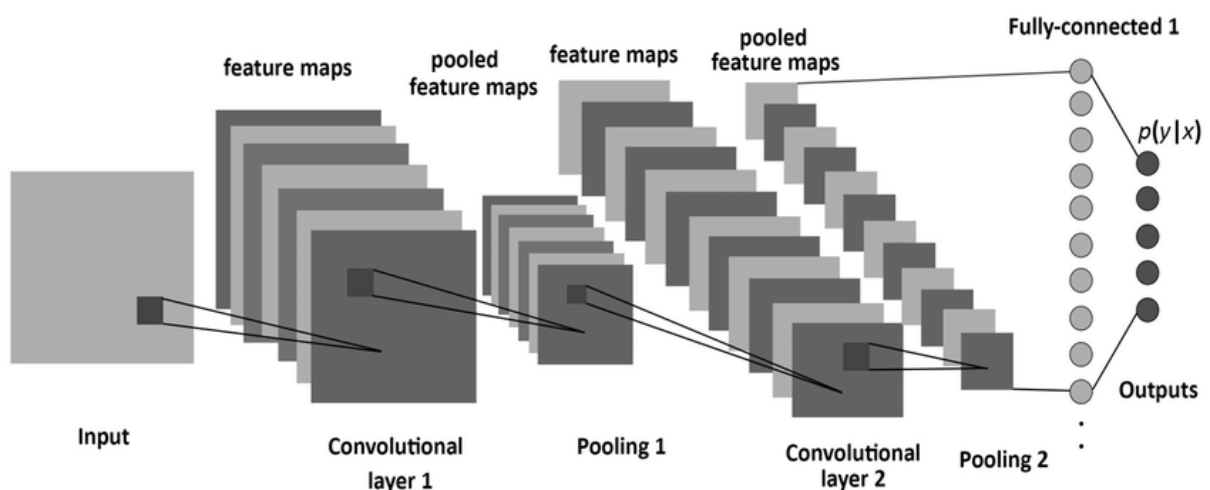


Figure 2.9 : Architecture classique d'un réseau de neurones convolutif [30]

1. Couche de convolution

La couche de convolution (figure 2.9) est l'élément essentiel des réseaux de neurones convolutifs. Elle constitue au moins leur première couche. Son but est de repérer la présence d'un ensemble de caractéristiques, dans les images reçues en entrée.

La convolution (figure 2.9) consiste en un filtrage de l'image suivant un masque (noyau), effectué pour extraire une carte d'activation, qui définit la position des caractéristiques de l'image. On glisse le masque sur l'image et, on calcule le produit de convolution entre le filtre et chaque portion de l'image balayée. Ainsi, plus la valeur est élevée, plus l'endroit correspondant est affecté aux attributs pertinents.

Une couche de convolution est un empilement de convolutions. En effet, l'image est parcourue par plusieurs noyaux de convolution qui donnent lieu à plusieurs cartes de caractéristiques de sorties. Chaque noyau de convolution possède des paramètres spécifiques à l'information, qui est recherchée dans l'image (par exemple : un noyau de convolution de type filtre Sobel a des paramètres permettant de rechercher les contours dans l'image).

Le choix des paramètres du noyau de convolution dépend de la tâche à résoudre. Ainsi, ces paramètres sont automatiquement appris par l'algorithme à partir des données d'entraînement. Notamment, grâce à la technique de rétropropagation du gradient, qui permet l'ajustement des paramètres en fonction de la valeur du gradient de la fonction de perte. Cette fonction calcule l'erreur entre la valeur prédite et la valeur cible.

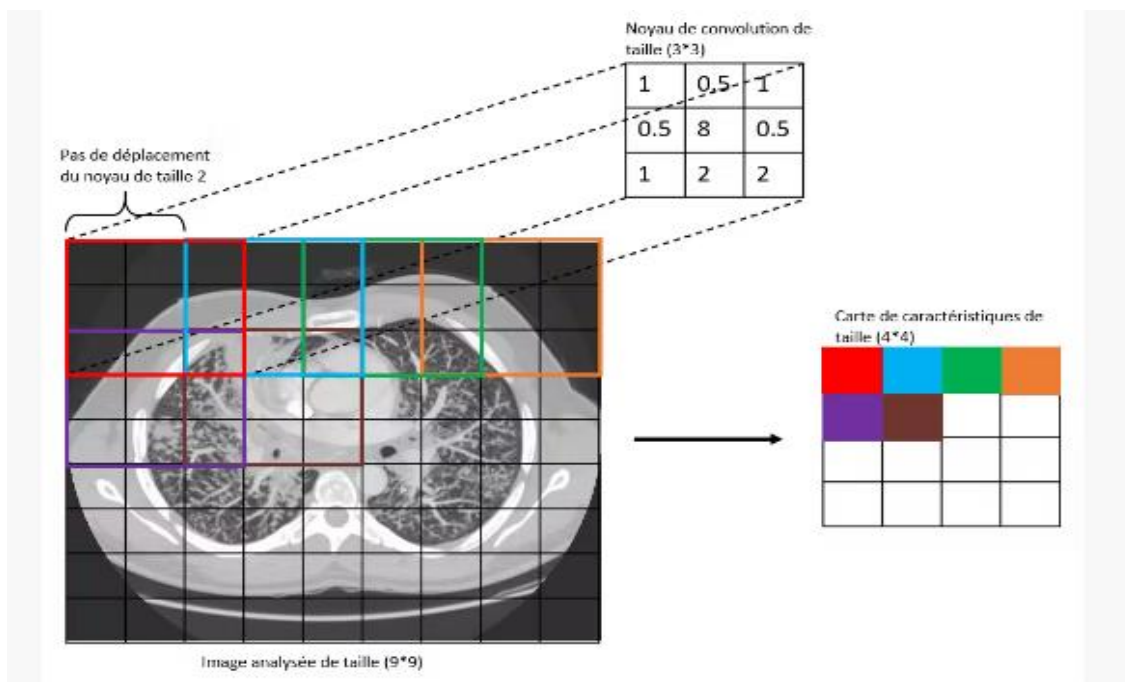


Figure 2.10 : Exemple d'une convolution avec un noyau de taille 3*3 et un pas de 2 [31]

2. Couche de correction (ReLU)

La couche de correction est l'application d'une fonction non-linéaire aux cartes de caractéristiques, en sortie de la couche de convolution.

La couche de correction ReLU remplace donc, toutes les valeurs négatives reçues en entrée, par des zéros. Souvent, la correction Relu est préférable, car il en résulte la formation d'un réseau neuronal plusieurs fois plus rapide, sans faire une différence significative à la généralisation de précision [32]. Elle est basée sur la fonction Relu (équation 2.2).

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x & \text{si } x \geq 0 \end{cases} \quad 2.2$$

3. Couche de pooling

Le pooling est une technique de sous-échantillonnage. Elle est généralement insérée régulièrement entre les couches de correction et de convolution. En réduisant la taille des cartes de caractéristiques, donc le nombre de paramètres du réseau, cela accélère le temps de calcul et diminue le risque de sur-apprentissage.

L'opération de pooling la plus courante est celle du maximum : MaxPool(2*2, 2). Elle est plus efficace que la moyenne, car elle maximise le poids des activations fortes. Elle est appliquée à la sortie de la couche précédente, comme un filtre de convolution de taille (2*2) et se déplace avec un pas de 2 (figure 2.10). En sortie de la couche de pooling est obtenue une carte de caractéristiques compressée par un facteur de 4 [30].

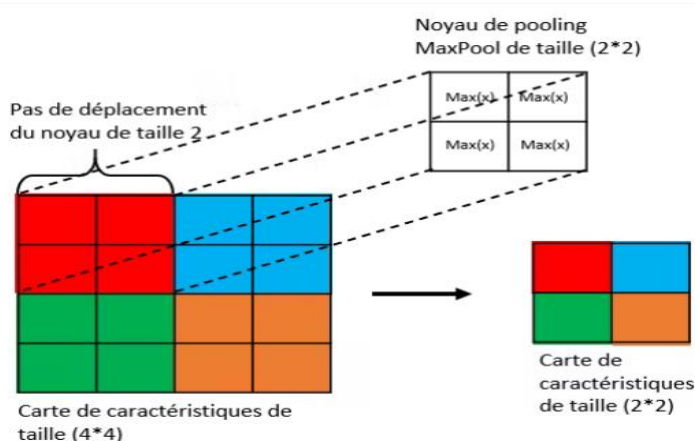


Figure 2.11 : Opération de pooling avec un noyau MaxPool de taille 2*2 et d'un pas de 2 [30]

Le pooling (mise en commun) est un algorithme visant à réduire la dimension d'une image tout en gardant des sous-parties importantes. La fenêtre de Pooling est une matrice $m \times m$, qui parcourt l'image afin de produire un seul pixel de sortie parmi $m \times m$ pixels voisins.

Plusieurs méthodes de production de pixels sont mises en place en fonction de la pertinence de l'information voulue. Parmi ces méthodes, on compte l'average pooling, le max et le min Pooling [32].

➤ **Average-Pooling**

Le principe de l'average-Pooling est de prendre la moyenne des $m \times m$ pixels voisins. Dans le traitement d'image, on parle de "lissage". La fenêtre de Pooling la plus utilisée est de 2×2 .

➤ **Max-Pooling**

Le Max-Pooling permet d'extraire uniquement la plus grande valeur dans un ensemble de pixels. Ce type de Pooling est avantageux pour analyser une image dont le fond est noir. Le but est ainsi d'ignorer tous les pixels de faible valeur (en noir) et d'en extraire l'avant-plan (plus clair).

➤ **Min-Pooling**

Le Min-Pooling permet d'extraire le pixel le plus faible parmi ses $m \times m$ pixels voisins. Cette méthode est adaptée dans les cas où le fond de l'image est clair et l'avant plan sombre.

La figure 2.11 illustre les différents résultats de pooling et leurs effets réducteurs.

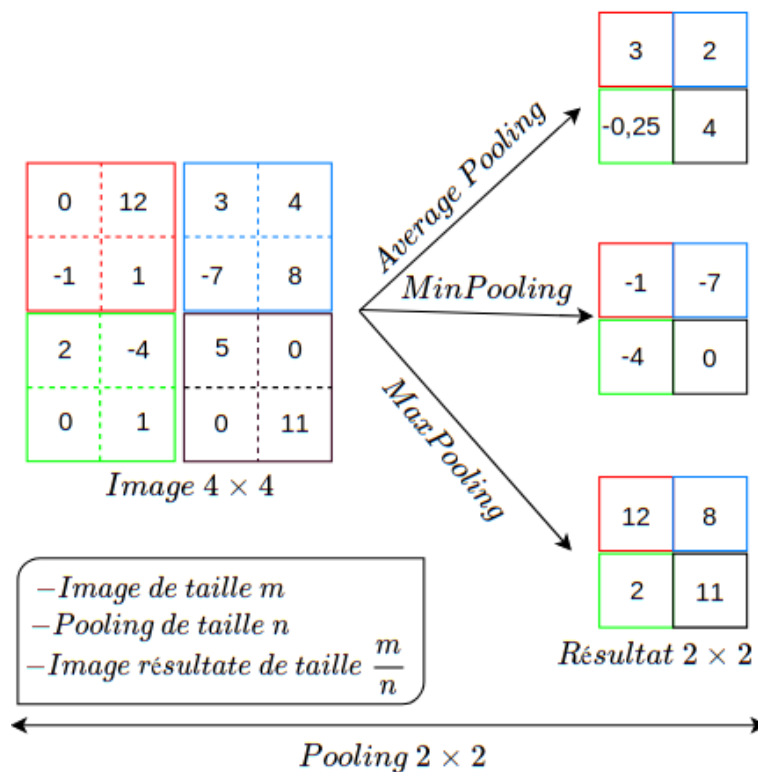


Figure 2.12 : Exemple applicatif des différents pooling [32]

4. Couche Flatten

Elle représente l'aplatissement (figure 2.12) qui est une opération qui transforme la matrice, en un vecteur $N \cdot 1$ (N est le nombre total d'éléments dans l'entrée) [32, 33].

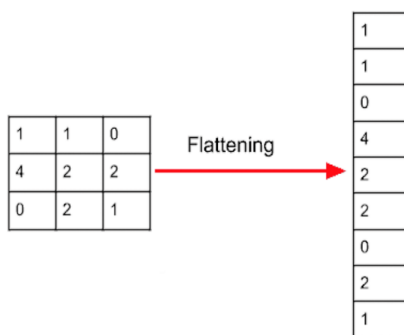


Figure 2.13 : Exemple de Flatten [33]

5. Couche entièrement connectée

Cette couche entièrement connectée (FC) est à la fin du réseau (figure 2.13). Elle permet la classification de l'image, à partir des caractéristiques extraites par la succession de blocs de traitement. Elle est entièrement connectée, car toutes les entrées de la couche sont connectées aux neurones de sorties de celle-ci.

Contrairement à la phase d'extraction des caractéristiques où les neurones, sont indépendants entre eux et ont uniquement accès à l'information du champ réceptif qu'ils traitent ; les neurones de la couche FC, ont accès à la totalité des informations d'entrée. Chaque neurone attribue à l'image, une valeur de probabilité d'appartenance à la classe i parmi les C classes possibles [32, 33].

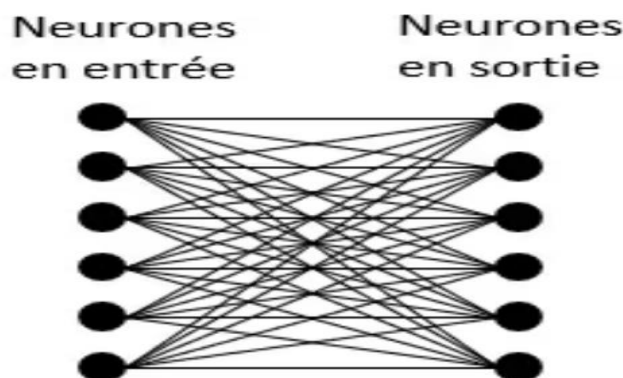


Figure 2.14 : Schéma d'une couche entièrement connectée avec 6 neurones [33]

6. Couche de perte

La couche de perte (LOSS) est la dernière couche du réseau. Elle calcule l'erreur entre la prédiction du réseau et la valeur réelle. Lors d'une tâche de classification, la variable aléatoire est discrète, car elle peut prendre uniquement les valeurs 0 ou 1, représentant l'appartenance (1) ou non (0) à une classe. C'est pourquoi la fonction de perte la plus courante et la plus adaptée est la fonction d'entropie croisée (cross-entropy) [32].

Celle-ci mesure la différence globale entre deux distributions de probabilités (celle de la prédiction du modèle et, celle du cas réel) pour une variable aléatoire ou un ensemble d'évènements [15]. Formellement, elle s'écrit :

$$\text{LOSS}(x, \text{class}) = - \sum_{\text{class}=1}^c y_{x, \text{class}} \log(p_{x, \text{class}}) \quad 2.3$$

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

7. Couche drop-out

Le drop-out consiste à invalider des neurones de manière aléatoire (figure 2.14). Chaque neurone désactivé garde son poids, mais ne contribue pas à la comptabilisation de la couche suivante [32].

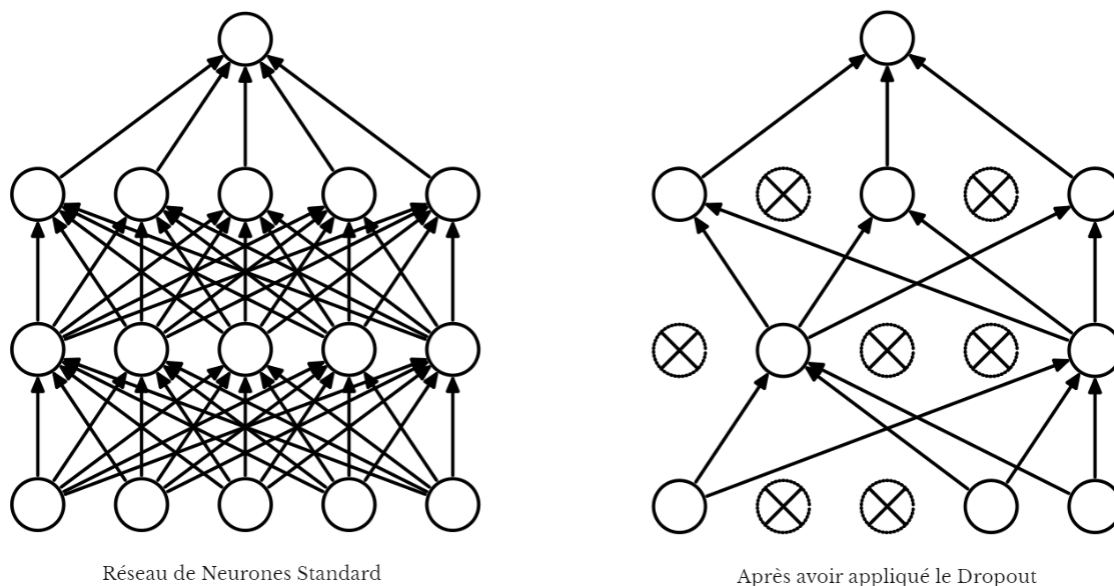


Figure 2.15 : Exemple de Dropout [32]

À chaque itération, on applique cette désactivation aléatoire. C'est-à-dire qu'à chaque 'forward propagation', le modèle apprendra avec une configuration de neurones différentes, les neurones s'activant et se désactivant aléatoirement. On désactive ainsi, temporairement certains neurones dans le réseau, ainsi que toutes ses connexions entrantes et sortantes.

L'objectif du dropout est une technique permettant de réduire le sur-apprentissage (l'overfitting) lors de l'entraînement du modèle.

8. Hyperparamètres

Les hyperparamètres permettent de contrôler l'apprentissage automatique en grande dimension [30, 32]. Ils sont répartis en deux catégories :

- **Les hyperparamètres de modèles :** Ils permettent de définir la taille du réseau (largeur, profondeur) et son type (ex : auto-encodeur). Ces paramètres ne sont modifiés, que dans le cas d'un fine tuning.
- **Les hyperparamètres d'algorithme :** Ils n'influencent pas la performance du modèle, mais le processus d'apprentissage (rapidité, qualité).

Le modèle est réalisé suivant les paramètres ci-dessous :

- La taille du 'batch size' en entrée du réseau.
- Le nombre d'itérations, qui définit le nombre de cycles d'apprentissage (training-validation),
- La fonction de perte qui calcule la valeur de l'erreur existante, entre l'estimation et l'observation.
- L'optimiseur qui est la fonction d'optimisation utilisée, pour la descente du gradient.
- Le taux d'apprentissage, qui représente le pas, dans l'algorithme de la descente du gradient.

8. Fonctions de l'optimisation

L'optimisation consiste à trouver les entrées d'une 'fonction objectif' qui aboutissent à la sortie minimale ou maximale de la fonction. 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 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 [32, 34].

a. Descente de gradient stochastique

La descente de gradient stochastique est une solution aux inconvénients de la méthode classique. Celle-ci consiste à rechercher un poids optimal (équations 2.4 et 2.5) en mettant à jour un paramètre, à partir d'un seul exemple d'entraînement.

$$W = w - \alpha \frac{dJ(w,b)}{dw} \quad 2.4$$

$$B = w - \alpha \frac{dJ(w,b)}{db} \quad 2.5$$

Où :

W : le nouveau poids, w : l'ancien poids.

B : le nouveau biais, b : l'ancien biais.

α : taux d'apprentissage.

b. Fonction Adam

L'optimisation Adam 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.

c. Taux d'apprentissage α et optimisation

L'optimisation consiste à chercher les paramètres qui minimisent la fonction d'erreur. L'optimisation est également une recherche du taux d'apprentissage α optimal, afin de converger de manière rapide et efficace vers un poids w optimal.

Le choix de la valeur de α peut être assez fluctuant. Il est considéré comme un hyperparamètre à optimiser car si la valeur est trop petite, la réduction de l'erreur sera très lente, tandis que si elle est trop grande, des oscillations divergentes peuvent en résulter. La figure 2.15 illustre un exemple de convergence pour différentes valeurs du taux d'apprentissage [18, 29]. Généralement, le taux d'apprentissage est choisi dans l'intervalle suivant : $\alpha \in [10^{-6}, 1[$.

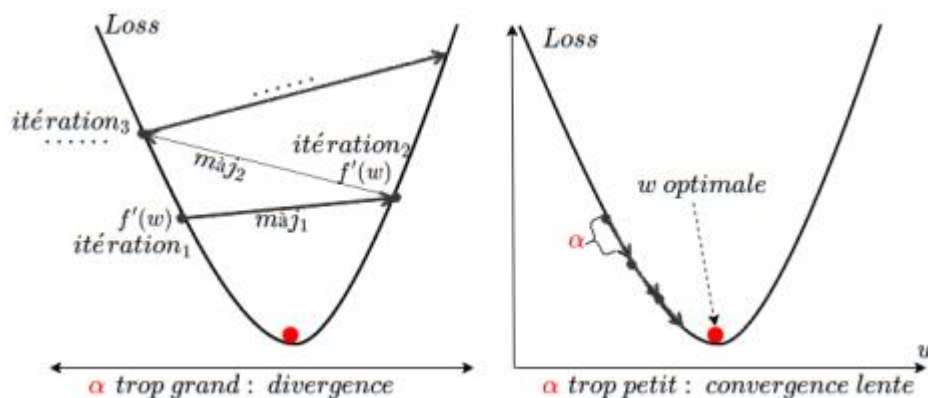


Figure 2.16 : Exemple de convergence pour différents α [32]

2.7 Fonctions d'activation

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 [35, 36].

Il existe plusieurs types de fonctions d'activation données ci-dessous :

2.7.1 Fonction Relu

La fonction Relu (Unité linéaire rectifiée) est la fonction d'activation la plus couramment utilisée (figure 2.16), dans les réseaux de neurones convolutifs.

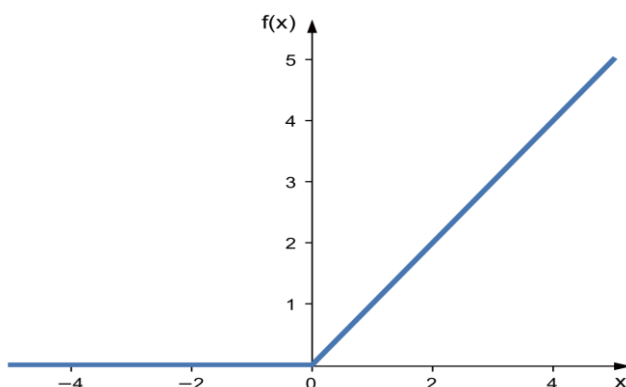


Figure 2.17 : Fonction ReLU [36]

2.7.2 Tangente hyperbolique

C'est une fonction d'activation largement adoptée (figure 2.16), pour un type spécial de réseau de neurones, appelé réseau de rétropropagation [36].

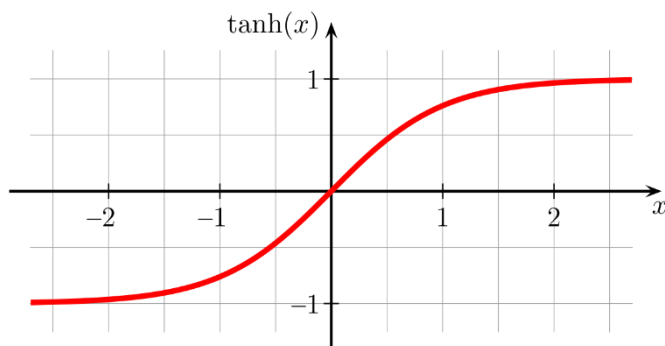


Figure 2.18 : Fonction tangente hyperbolique [36]

Elle est donnée par l'expression mathématique 2.6.

$$\text{Tanh}(x) = f(x) \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad 2.6$$

2.7.3 Fonction sigmoïde

Appelée aussi la fonction logistique, la sigmoïde est une fonction mathématique classique, la plus ancienne et la plus populaire (figure 2.17). C'est une fonction non-linéaire, dérivable, continue et symétrique par rapport à l'axe des y, elle produit une courbe sigmoïde S (équation 2.7) [20].

$$S = \frac{1}{1+e^{-x}} \quad 2.7$$

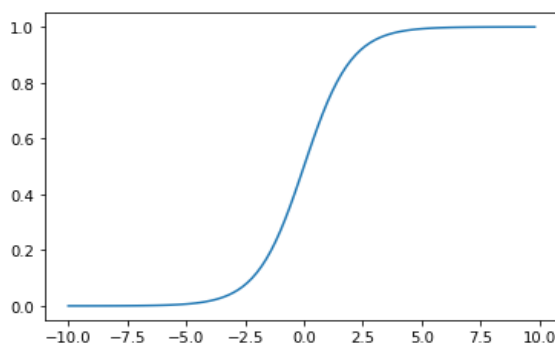


Figure 2.19 : Fonction sigmoïde [36]

2.7.4 Fonction softmax

La fonction Softmax (équation 2.8) ou régression logistique multinomiale, est souvent utilisée pour résoudre les problèmes de multiclassification. Elle utilise la distribution de probabilité.

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum e^{x_i}} \quad 2.8$$

Où : x_i est le vecteur d'entrée.

2.8 Apprentissage par transfert

L'intelligence artificielle (IA), et plus spécifiquement le deep learning, nécessite un grand nombre de données pour faire fonctionner à leur meilleur niveau, les réseaux de neurones artificiels. Pour améliorer la qualité de l'apprentissage d'un logiciel, il est possible de mutualiser des savoirs d'un modèle à un autre par le 'transfer learning'.

Le Transfer Learning, ou apprentissage par transfert, désigne les techniques permettant de transférer les connaissances de plusieurs sources vers d'autres problèmes à régler.

Il s'agit des compétences apprises à partir de tâches antérieures, servant à améliorer de nouvelles tâches similaires. Ce transfert de savoir global est très efficace dans le cas du deep learning.

Il représente une optimisation qui permet une progression rapide ou une amélioration des performances, lors de la modélisation de la deuxième tâche.

Il existe plusieurs modèles basés sur l'apprentissage par transfert, tels que le VGG16, le Resnet50, ou le densenet que nous décrirons dans le prochain chapitre.

Le *transfer learning* repose sur la ré-exploitation des connaissances acquises dans d'autres configurations (sources), pour la résolution d'un problème particulier (cible). Dans ce contexte, on peut distinguer plusieurs approches selon ce que l'on souhaite transférer, quand et comment réaliser le transfert. Le transfer learning peut être de différents types, tels que : l'apprentissage par transfert inductif, transductif et non-supervisé [37, 38].

1. Apprentissage par transfert inductif

Dans cette configuration, les domaines source et cible sont les mêmes (même données), mais les tâches source et cible sont différentes, mais proches. L'idée consiste alors à utiliser les modèles existants, pour réduire de manière avantageuse le champ d'application des modèles possibles (biais de modèle).

2. Apprentissage par transfert non supervisé

Les sources et cibles sont similaires, et les tâches sont différentes. Par contre, les données ne sont pas labellisées.

3. Apprentissage par transfert transductif

Les tâches et sources sont les mêmes, mais les domaines sont différents.

2.9 Outils de mesures des performances

On utilise les outils de mesure de performance, pour évaluer les résultats d'un algorithme d'apprentissage du 'machine et du deep learning' [39].

2.9.1 Matrice de confusion

Il s'agit d'un outil de mesure de performances, dédié aux algorithmes de classification binaire, ou à la multiclassification. Cette matrice (figure 2.18) utilise la technique de comparaison des classes réelles (positive ou négative) et, des classes prédites.

Ceci permet de mesurer les erreurs commises par le modèle d'apprentissage et de détecter les types de ces erreurs. La matrice de confusion permet aussi de visualiser rapidement, les données et, les statistiques.

		classe réelle	
		-	+
classe prédite	-	vrai négatifs	faux négatifs
	+	faux positifs	vrai positifs

Figure 2.20 : Matrice de confusion [39]

- Vrai positif (true positive TP), ce paramètre est observé quand le résultat du modèle (la prédiction) est positif et la valeur réelle est vraiment positive.
- Vrai négatif (true negative TN) ce paramètre est, observé quand la prédiction est négative et, la valeur réelle est négative.
- Faux positif (false positif FP) : il s'agit d'une erreur observée quand la prédiction est positive et, la valeur réelle est négative.
- Faux négatif (false negative FN) : il s'agit d'une erreur observée quand la prédiction est négative et la valeur réelle est positive.

2.9.2 Critères d'évaluation

Les critères d'évaluations d'un système, peuvent être donnés par les facteurs suivants :

- **Le nombre total d'erreur =FP+FN**
- **La spécificité** (équation 2.13), est la capacité d'un système d'identifier la proportion du nombre d'échantillons négatifs, que le modèle a correctement identifié.

$$\text{spécificité} = \frac{TN}{FP+TN} \quad 2.9$$

- **Précision (accuracy)** : c'est le nombre d'échantillons positifs ou négatifs, qui ont correctement été prédits.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad 2.10$$

- **Rappel (recall)** : appelé aussi sensibilité (sensitivity) est le taux des vrais positifs. Le rappel (équation 2.12) a pour le but d'identifier la proportion du nombre d'échantillons positifs, que le modèle a correctement identifiée.

$$\text{Rappel} = \frac{TP}{TP + FN} \quad 2.11$$

- **F-score** : est la moyenne du rappel et de la précision (équation 2.12).

$$F - \text{score} = 2 * \frac{\text{precision} * \text{rappel}}{\text{precision} + \text{rappel}} \quad 2.12$$

- **Courbe ROC**

La courbe ROC (Receiver Operating Characteristic), est un outil de visualisation de performance d'un modèle de classification binaire, en représentant la sensibilité (rappel) en fonction de 1-spécificité pour différents seuils de la classification.

2.10 Conclusion

Dans ce chapitre, nous avons présenté quelques notions de bases sur le 'machine learning' et le 'deep learning' ainsi que les réseaux de neurones convolutifs (CNN), l'un des algorithmes de la classification d'images.

Le plus important à retenir dans un neurone formel, est la mise à jour des poids, la fonction de perte et, la descente de gradient pour son optimisation, car toutes ces notions participent à la synthèse de la solution finale.

Dans le chapitre suivant, nous présenterons les différents modèles appliqués sur notre base de données échographiques. Chaque solution sera suivie d'une discussion, qui aboutira pas à pas à un modèle final.

Chapitre 3 Conception D'un Système D'aide à la Décision pour la Détection du Cancer du Sein dans Les Images Echographiques

3.1 Introduction

Dans ce chapitre, nous abordons comment la conception de notre système d'aide à la décision du cancer du sein en imagerie échographique est effectuée ; en passant en revue la description des modèles de classification utilisés, ainsi que de la base de données.

Il existe plusieurs modèles de classification par les réseaux de neurones, tous ont donné des résultats instructifs dans la recherche [40].

Dans le cadre de ce travail, nous optons pour 5 différents modèles, que nous allons comparer pendant les expérimentations décrites, dans le dernier chapitre. Ainsi, nous serons en mesure de décider lequel de ces derniers, est le mieux adapté pour notre problème. Les modèles qui ont retenu notre attention sont le 'VGGNet', le 'ResNet', et le 'DenseNet'. Ils seront décrits suivant leur différentes fonctions.

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.

L'étude de Al-Dhabyani et al [41] a été consacrée à la classification des masses, en échographie mammaire. La classification des tumeurs mammaires, est réalisée par les réseaux CNN pré-entraînés, avant et après l'augmentation des données par le modèle NASNET, suivant une précision de 99%.

Dans l'article de Yap et al [42], une détection automatique des lésions échographiques mammaires suivant plusieurs architectures CNN (Lenet, UNET et Alexnet), a été réalisée sur deux types de bases d'images échographiques. Le meilleur résultat est obtenu par l'hybridation des deux ensembles de données.

Chia-Yen Lee et al [43] rapportent que l'approche proposée, permet la détection de tumeurs mammaires sur des images échographiques, en se basant sur plusieurs modèles des réseaux convolutifs. Ils ont comparé 4 modèles d'apprentissages par transfert (AlexNet, Inception v3, ResNet et, DenseNet), pour déduire que le réseau Densenet, a obtenu la meilleure précision d'identification. A cet effet, la précision obtenue est de 83%.

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 [44, 45, 46], la détection des masses, suivant VGG19, VGG16 et Inception V3 s'est faite avec une précision de 94,64%.

La catégorisation des carcinomes mammaires [47], par les réseaux VGG16, VGG19, Densenet, Resnet, ainsi qu'un CNN adapté, a permis une précision de 94% dans un premier travail [66] et une précision de 98% dans une autre étude [48].

3.3 Présentation du système conceptuel

Le schéma qui suit (figure 3.1) représente la démarche entreprise, pour la détection de tumeurs mammaires en imagerie échographique pour l'aide à la décision, suivant 6 modèles des réseaux convolutifs : VGG16, VGG19, Resnet50, Resnet152, Densenet121 et Densenet201.

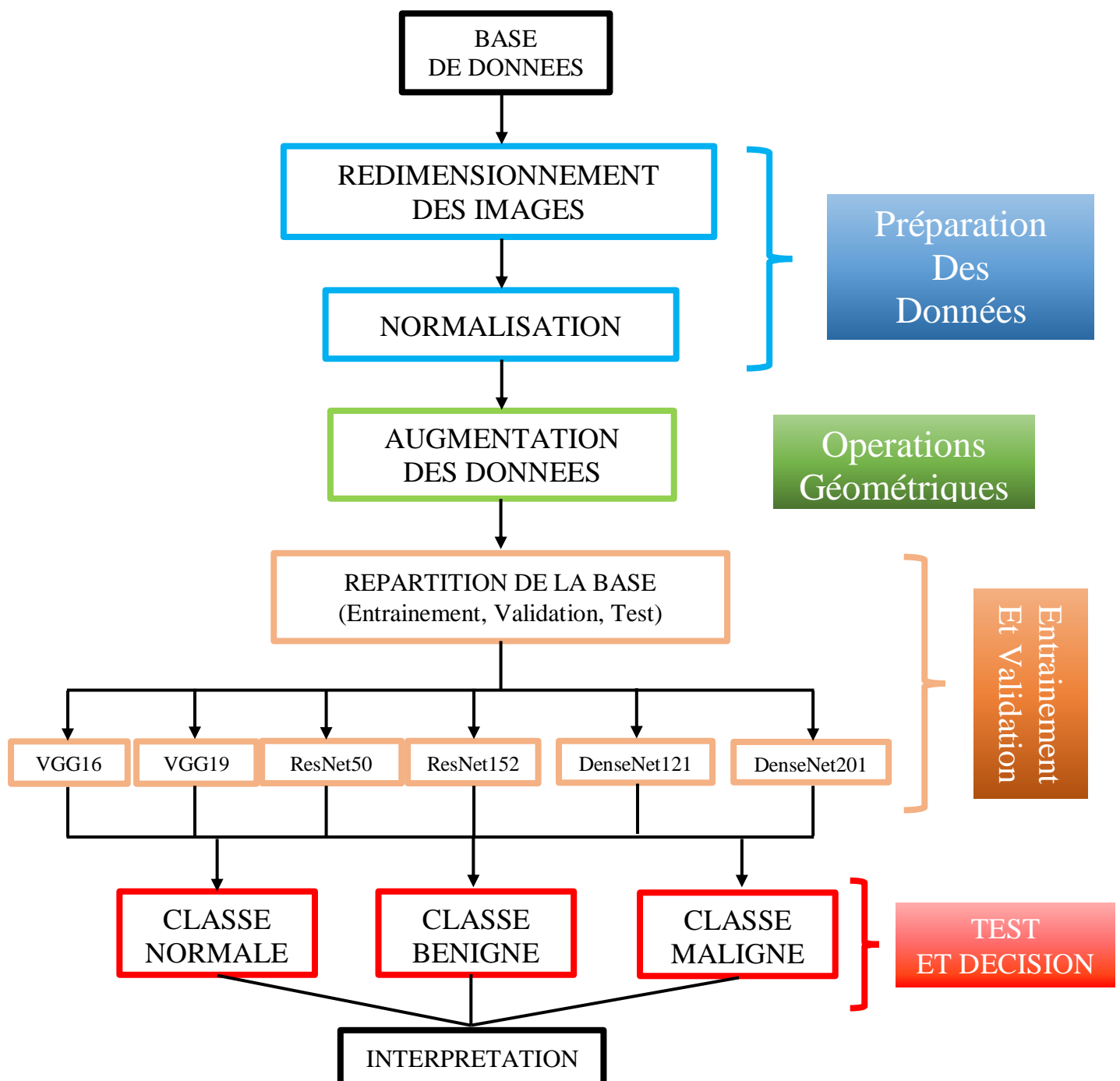


Figure 3.1 : Schéma conceptuel de détection de tumeurs mammaires pour l'aide à la décision

3.3.1 Base de données

La base de données explorée dans le cadre de ce projet, représentant des images échographiques mammaires, est d'origine égyptienne « DatasetBusi » (BreastUltrasound Images Dataset) [49]. Cette base de données est divisée en trois ensembles, où chacun d'eux, est utilisé pour une phase de création du modèle.

- ✓ La base entraînement ou « training set », représente 80% de l'ensemble de données d'origine). Elle est utilisée pour entraîner le modèle choisi.
- ✓ La base de validation, représente 10% de l'ensemble des données originelles.
- ✓ La base de test représente 10% de l'ensemble de données.

Les données comprennent des images échographiques mammaires chez les femmes âgées de 25 à 75 ans. Ces données égyptiennes ont été recueillies en 2018 sur 600 patientes. L'ensemble de données se compose de 780 images avec une taille d'image moyenne de 500*500 pixels. Les images sont au format PNG. Les images de vérité terrain sont présentées avec des images originales. Les images sont classées en trois classes, normales, bénignes et malignes.

3.3.2 Préparation des données

Nous définirons dans cette section les différentes notions relatives à la préparation des données.

1. Redimensionnement des images

Comme cité précédemment, les images explorées, sont de la taille 500*500 pixels. Or, ces dimensions ne conviennent pas aux modèles utilisés. Il est alors obligatoire de redimensionner les images à la taille de 224*224, suivant le réseau utilisé [50].

2. Normalisation

La normalisation est un processus qui modifie la plage des valeurs d'intensité des pixels, de [0 255] à [0 1].

3. Augmentation des données

Augmenter son dataset signifie tout simplement construire de nouvelles images, à partir des images dont on dispose, suivant des opérations géométriques. Comment faire ? Plusieurs méthodes et techniques existent pour faire de l'augmentation de données. On en dénombre plus d'une dizaine de méthodes. Les plus utilisés sont : le retournement c'est-à-dire qu'on retourne notre image horizontalement et verticalement, ou simplement les faire pivoter à 180°.

On peut également utiliser le recadrement, le zoom, la rotation, l'élimination du bruit gaussien..., et tant d'autres méthodes toutes aussi efficaces, les unes que les autres. Dans notre cas, nous utilisons la rotation de l'image, le zoom, le décalage horizontal et vertical [51].

Dans notre étude, la base de données initial contenait 780 images, et après applications de l'augmentation de données, la base de données contiens maintenant 1572 images.

3.3.3 Entraînement et validation

Cette partie contient le travail le plus important, puisqu'il s'agit du traitement des données suivant les 5 modèles pré-entraînés sélectionnés, pour la classification des tumeurs mammaires. Nous modifions les architectures choisies, en remplaçant leurs dernières couches par un nouveau classifieur adapté, aux trois classes de la base de données étudiée, pour l'entraînement et la validation des résultats. Ces modèles sont affinés selon les stratégies du fine tuning [52].

3.3.4 Test et décision

A la fin du déroulement du système, les résultats obtenus seront évalués suivant la précision et la perte. C'est grâce à ces mesures, que la comparaison sera faite, pour savoir lequel de ces modèles, est le meilleur.

3.4 Modèles des réseaux de neurones adoptés

Dans cette section, sont présentés les modèles pré-entraînés adoptés, dans le cadre de ce mémoire : le VGG16, le VGG19, le ResNet50, le DenseNet 121 et le DenseNet 201 [53]. Les numéros après les noms des modèles indiquent le nombre de couche convolutionnelles, dont se compose le modèle, par exemple le VGG19 se compose de 19 couches convolutionnelles.

3.4.1 Modèle VGGNet

Basé sur une architecture du réseau neuronal à convolution profonde (CNN), le modèle VGG (Visual Geometry Group) a été développé pour augmenter les performances du modèle [54].

Le "profond" fait référence au nombre de couches avec VGG-16 ou VGG-19, composés respectivement de 16 et 19 couches convolutionnelles.

L'architecture VGG (décrite ci-dessous) est à la base de modèles de reconnaissance d'objets révolutionnaires. Développé en tant que réseau neuronal profond, le VGGNet dépasse

également les lignes de base, sur de nombreuses tâches et ensembles de données au-delà d'ImageNet.

- **Entrée** : Le VGGNet prend une taille d'entrée d'image de 224×224 .
- **Couches convolutives** : Les couches convolutives de VGG tirent parti d'un champ de réception minimal, c'est-à-dire 3×3 , la plus petite taille possible qui capture toujours le haut/bas et la gauche/droite de l'image. De plus, il existe également des filtres de convolution 1×1 agissant comme une transformation linéaire de l'entrée. Vient ensuite une fonction d'activation d'unité linéaire rectifiée Relu, qui est une énorme innovation d'AlexNet qui réduit le temps de formation. La foulée de convolution est fixée à 1 pixel pour conserver la résolution spatiale préservée après la convolution (la foulée est le nombre de décalages de pixels sur la matrice d'entrée).
- **Couches cachées** : Toutes les couches cachées du réseau VGG, utilisent la fonction ReLU. VGG n'utilise généralement pas la normalisation de la réponse locale (LRN), car elle augmente la consommation de mémoire et le temps de formation. De plus, il n'apporte aucune amélioration à la précision globale.
- **Couches entièrement connectées** : Le VGGNet a trois couches entièrement connectées. Sur les trois couches, les deux premières ont 4096 canaux chacune et, la troisième a 1000 canaux, 1 pour chaque classe.

1. Modèle VGG16

Le modèle VGG16, qui prend en charge 16 couches, est un modèle du réseau neuronal convolutionnel proposé par A. Zisserman et K. Simonyan de l'Université d'Oxford [55]. Ces chercheurs ont publié leur modèle, dans le document de recherche intitulé " Réseaux convolutifs très profonds pour la reconnaissance d'images à grande échelle ".

Le modèle VGG16 a atteint une précision de 92,7 % dans le top 5 des tests dans ImageNet.

ImageNet est un jeu de données composé de plus de 14 millions d'images, appartenant à 1000 classes. De plus, c'était l'un des modèles les plus populaires soumis à ILSVRC-2014 (ImageNet Large Scale Visual Recognition Challenge) [56].

Le modèle VGG16 remplace les filtres de grande dimension, par plusieurs filtres de la taille d'un noyau de 3×3 , apportant ainsi des améliorations significatives par rapport à AlexNet. Le modèle VGG16 a été formé à l'aide de GPU Nvidia Titan Black pendant plusieurs semaines.

Chapitre 3 Conception d'un système d'aide à la décision pour la détection du cancer du sein dans les images échographiques

Comme mentionné ci-dessus, le VGGNet-16 (figures 3.2 et 3.3) prend en charge 16 couches et peut classer les images en 1000 catégories d'objets, y compris le clavier, les animaux, le crayon, la souris, etc.

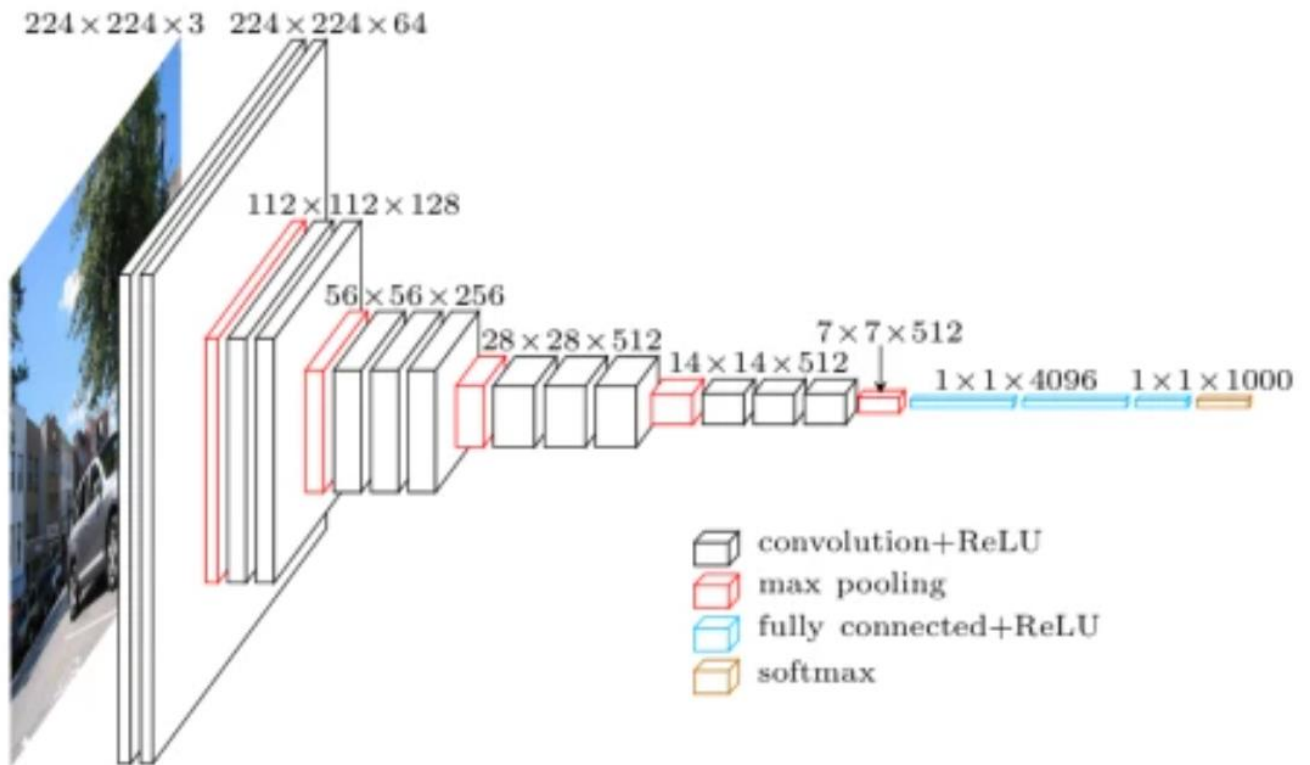


Figure 3.2: Architecture du réseau neuronal VGG16 [54]

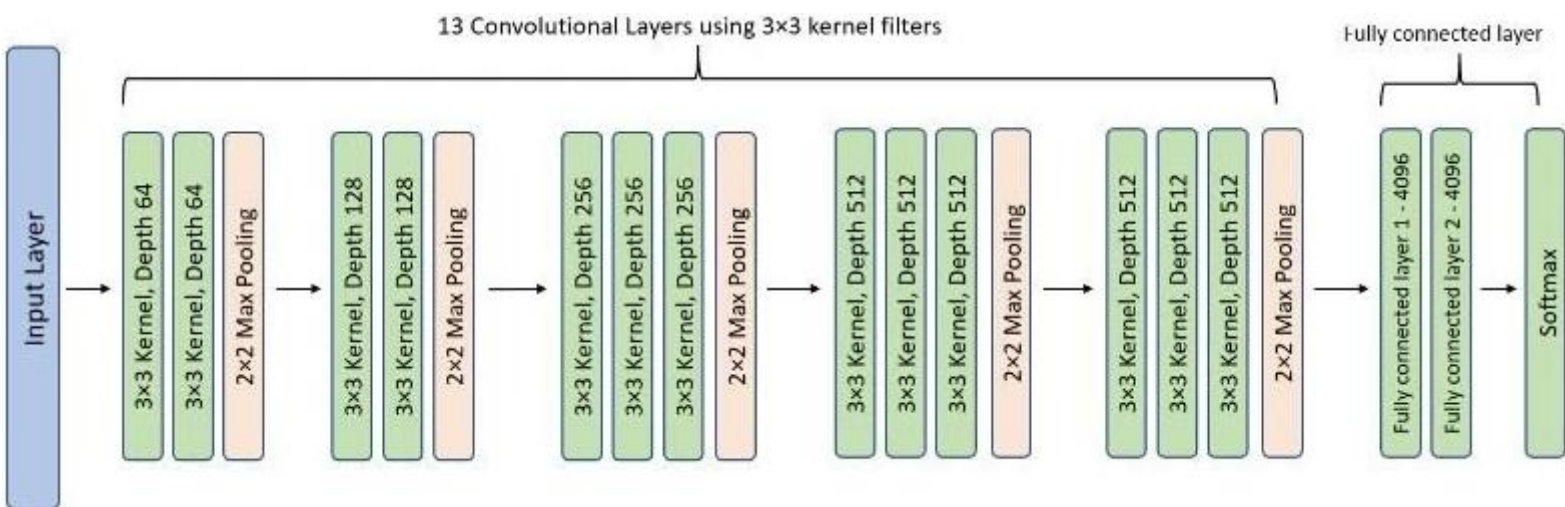


Figure 3.3 : Architecture de VGG16 [54]

Le chiffre 16 dans le nom VGG fait référence au fait qu'il s'agit d'un réseau neuronal profond à 16 couches (VGGnet). Cela signifie que VGG16 est un réseau assez étendu et compte au total environ 138 millions de paramètres. Même selon les normes modernes, c'est un énorme réseau. Cependant, la simplicité de l'architecture VGGNet16 est ce qui rend le réseau plus attrayant. Rien qu'en regardant son architecture, on peut dire qu'elle est assez uniforme [54].

Nous constatons sur les deux architectures de VGG16, qu'il se compose de :

- 13 couches convolutives (convolutional layers), représentées par conv3-XXX.
- 3 couches entièrement connectées (fully connected layers), représentées par FC-XXXX.
- 5 couches de mise en commun (couches de pooling), représentées par Maxpool de taille 2x2. Le nombre total de couches, est donc de $13 + 3 = 16$, soit 16 dans le VGG16.

Il y a quelques couches de convolution, suivies d'une couche de regroupement qui réduit la hauteur et la largeur. Si nous regardons le nombre de filtres que nous pouvons utiliser, environ 64 filtres sont disponibles, que nous pouvons doubler à environ 128 puis à 256 filtres. Dans les dernières couches, nous pouvons utiliser 512 filtres.

2. Modèle VGG19

Le concept du modèle VGG19 (figures 3.4 et 3.5) est le même que celui du VGG16, sauf qu'il prend en charge 19 couches [57, 58]. Le « 16 » et le « 19 » représentent le nombre de couches de poids dans le modèle (couches convolutives). Cela signifie que VGG19 a trois couches convolutionnelles de plus que le VGG16.

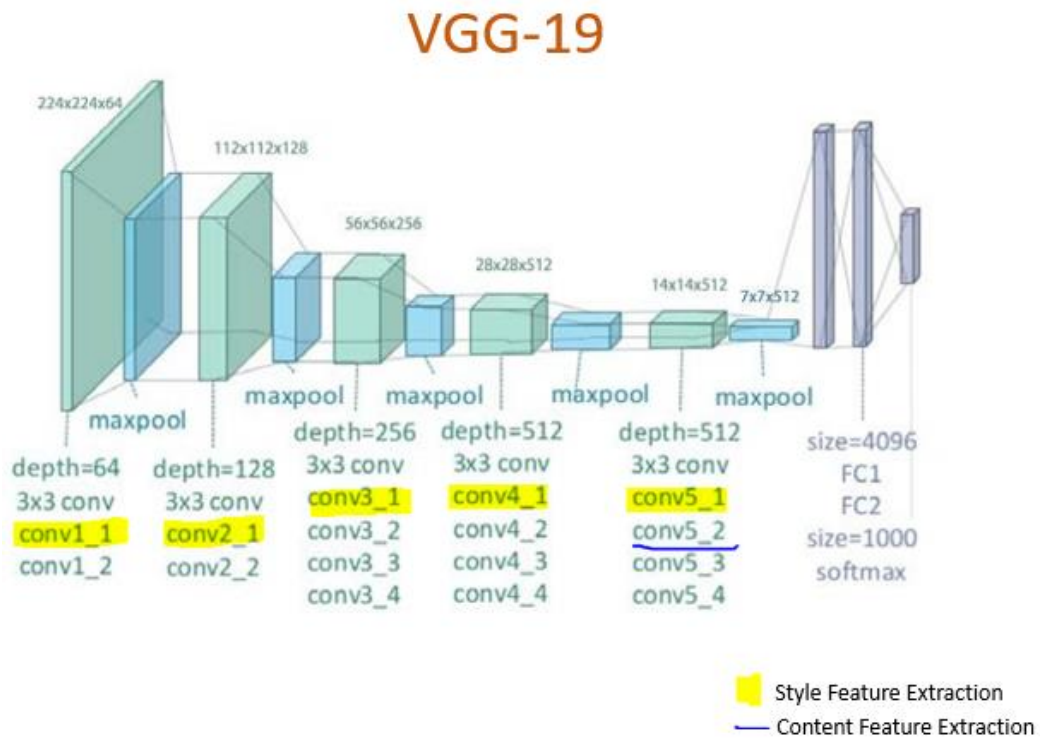


Figure 3.4 : Architecture de VGG19 [57]

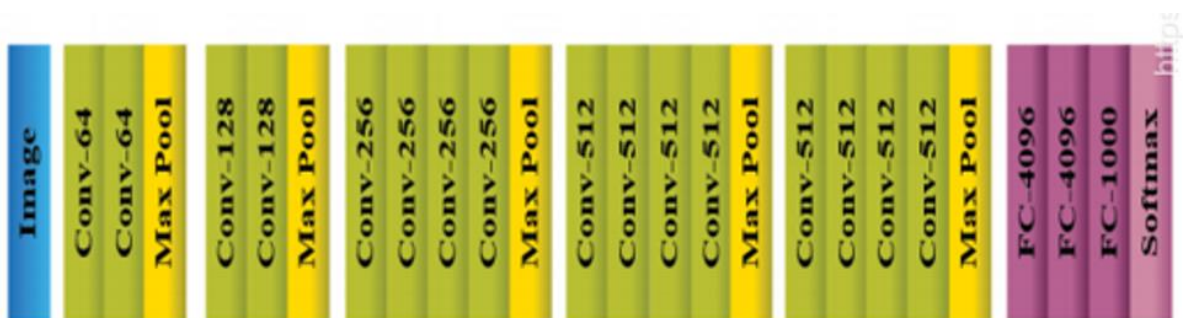


Figure 3.5 : Modèle de VGG19 [58]

3. Performances du modèle VGG16

VGG16 surpasse largement les versions précédentes des modèles des compétitions ILSVRC-2012 et ILSVRC-2013. De plus, le résultat VGG16 est en compétition pour le vainqueur de la tâche de classification GoogLeNet avec une erreur de 6,7 %. En termes de performances à réseau unique, le modèle VGG16 obtient le meilleur résultat avec environ 7,0 % d'erreur de test, dépassant ainsi, GoogLeNet d'environ 0,9 % [54].

3.4.2 Modèle ResNet 50

ResNet signifie Residual Network [59, 60]. Il s'agit d'un réseau de neurones innovant, qui a été introduit pour la première fois par Kaiming He, Xiangyu Zhang, Shaoqing Ren et Jian Sun dans leur article de recherche sur la vision par ordinateur de 2015 intitulé "Apprentissage résiduel profond pour la reconnaissance d'images" [61].

Ce modèle a connu un immense succès, comme en témoigne le fait que son ensemble a remporté la première place au concours de ILSVRC 2015, avec une erreur de seulement 3,57 %. De plus, il est également arrivé premier, dans la détection ImageNet, la localisation ImageNet, la détection COCO et la segmentation COCO dans les compétitions ILSVRC & COCO de 2015 [62].

Les réseaux de neurones résiduels, utilisent des connexions sautées ou des raccourcis pour sauter par-dessus certaines couches (figure 3.6). Les avantages de ces réseaux, résident en leur capacité à réduire la fuite du gradient.

ResNet a de nombreuses variantes qui fonctionnent sur le même concept, mais ont des nombres de couches différents.

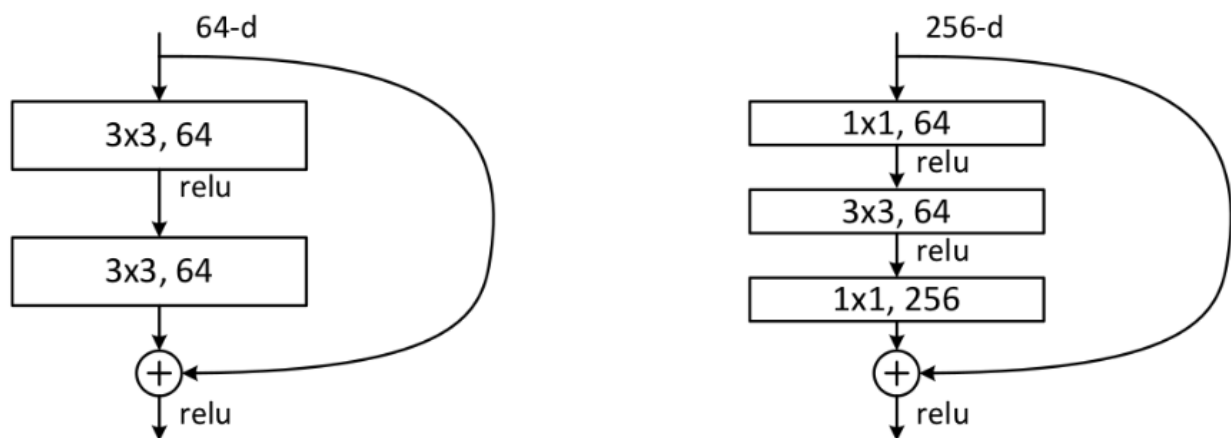


Figure 3.6 : Bloc résiduel [60]

Resnet50 (figure 3.7) est utilisé pour désigner la variante, qui peut fonctionner avec 50 couches des réseaux de neurones. C'est cette architecture qui est employée, dans le cadre de ce travail.

Chapitre 3 Conception d'un système d'aide à la décision pour la détection du cancer du sein dans les images échographiques

Le modèle ResNet-50 se compose de 5 étapes, chacune d'elles, avec une convolution et un bloc d'identité (figure 3.25). Chaque bloc de convolution possède 3 couches de convolution, tout comme le bloc d'identité, qui se compose également de 3 couches de convolution. Le ResNet-50 possède plus de 23 millions de paramètres, pouvant être entraînés [59, 63].

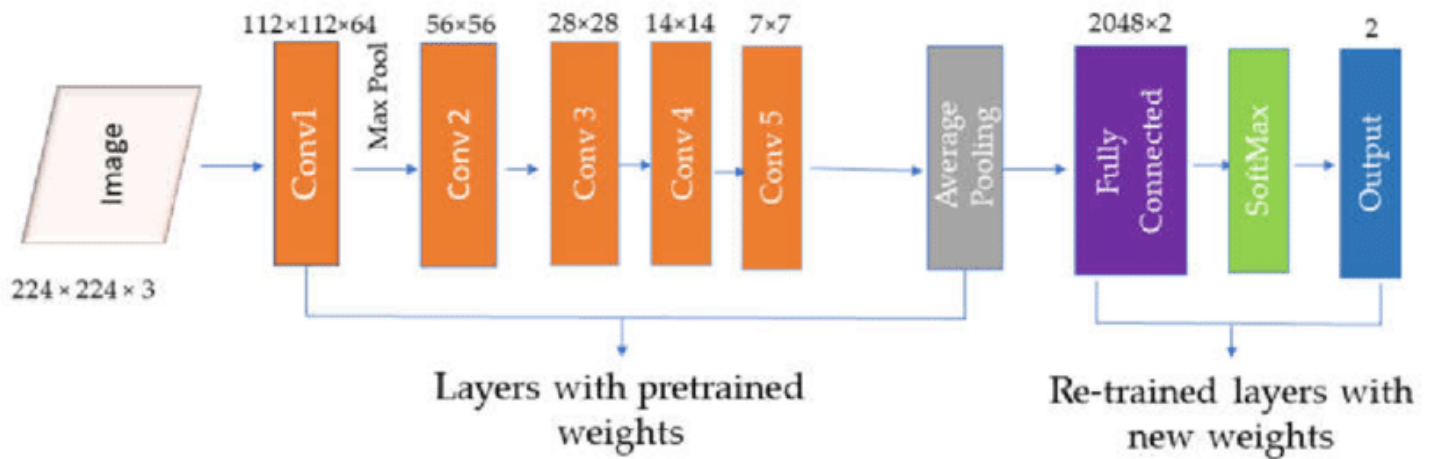


Figure 3.7 : Architecture de ResNet50 [63]

Le tableau 3.1 indique les détails de l'architecture de Resnet50.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
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 3.1 : Détails de l'architecture ResNet50 [60]

Comme on peut le voir dans le tableau 3.1, l'architecture Resnet 50 contient les éléments suivants :

- Une convolution avec une taille de noyau de $7 * 7$ et 64 noyaux différents tous avec une foulée de taille 2 nous donnant 1 couche.
- Une mise en commun maximale avec une taille de foulée de 2.
- Dans la convolution suivante, il y a 64 filtres d'un noyau $1 * 1$, suivis de 64 filtres d'un noyau $3 * 3$, et enfin de 256 filtres d'un noyau $1 * 1$. Ces trois couches sont répétées au total 3 fois, ce qui nous donne 9 couches dans cette étape.
- Ensuite, nous voyons 128 filtres d'un noyau de $1 * 1$, 128 filtres d'un noyau de $3 * 3$ et enfin 512 filtres d'un noyau de $1 * 1$. Cette étape est répétée 4 fois, nous donnant ainsi 12 couches.
- Après cela, on remarque 256 filtres d'un noyau de $1 * 1$, 256 filtres suivant un noyau de $3 * 3$ et 1024 filtres de taille $1 * 1$. Ceci est répété 6 fois, nous donnant un total de 18 couches.
- 512 filtres d'un noyau $1 * 1$, 512 filtres avec un noyau de $3 * 3$ et 2048 filtres suivant un noyau de $1 * 1$. Cela a été répété 3 fois, nous donnant un total de 9 couches.
- Un pooling moyen suivi d'une couche entièrement connectée contenant 1000 nœuds. A la fin une fonction softmax, qui donne 1 couche.

Donc en totalisant ces couches, on remarque un réseau convolutif profond, de $1 + 9 + 12 + 18 + 9 + 1 = 50$ couches.

3.4.3 Avantages de ResNet 50

À mesure que le nombre de couches augmente dans un CNN, la capacité du modèle à s'adapter à des fonctions plus complexes, augmente également. Par conséquent, plus de couches promettent de meilleures performances. Cela ne doit pas être confondu avec un réseau de neurones artificiels (ANN), où une augmentation du nombre de couches n'entraîne pas nécessairement de meilleures performances.

Un réseau neuronal résiduel utilise l'insertion de connexions de raccourcis, pour transformer un réseau simple en son homologue de réseau résiduel. Comparés aux VGGNets, les ResNets sont moins complexes, car ils ont moins de filtres.

Le réseau résiduel, ne permet pas au problème de gradient de fuite, de se produire. Les liaisons de saut agissent comme des autoroutes en pente, ce qui permet à la pente de s'écouler sans être perturbée. C'est également l'une des raisons les plus importantes, pour lesquelles ResNet est disponible dans des versions telles que ResNet50, ResNet101 et ResNet152 [59].

3.4.4 Modèle de DenseNet

Dans un réseau de neurones convolutif (CNN) traditionnel, chaque couche convolutive, à l'exception de la première (l'entrée), reçoit la sortie de la couche convolutive précédente et produit une carte de caractéristiques de sortie, qui est ensuite transmise à la suivante.

Cependant, à mesure que le nombre de couches dans le CNN augmente, c'est-à-dire à mesure qu'elles s'approfondissent, le problème du « gradient de fuite » se pose. Cela signifie qu'au fur et à mesure que le chemin des informations entre les couches d'entrée et les couches de sortie augmente, certaines informations peuvent « disparaître » ou se perdre, ce qui réduit la capacité du réseau à s'entraîner efficacement.

Les DenseNets résolvent ce problème en modifiant l'architecture CNN standard, avec la simplification de la connectivité entre les couches. Dans une architecture DenseNet, chaque couche est connectée directement à toutes les autres couches, d'où le nom de réseau convolutif Densely Connected [59, 64, 65].

La figure 3.8 représente un bloc dense avec 5 couches.

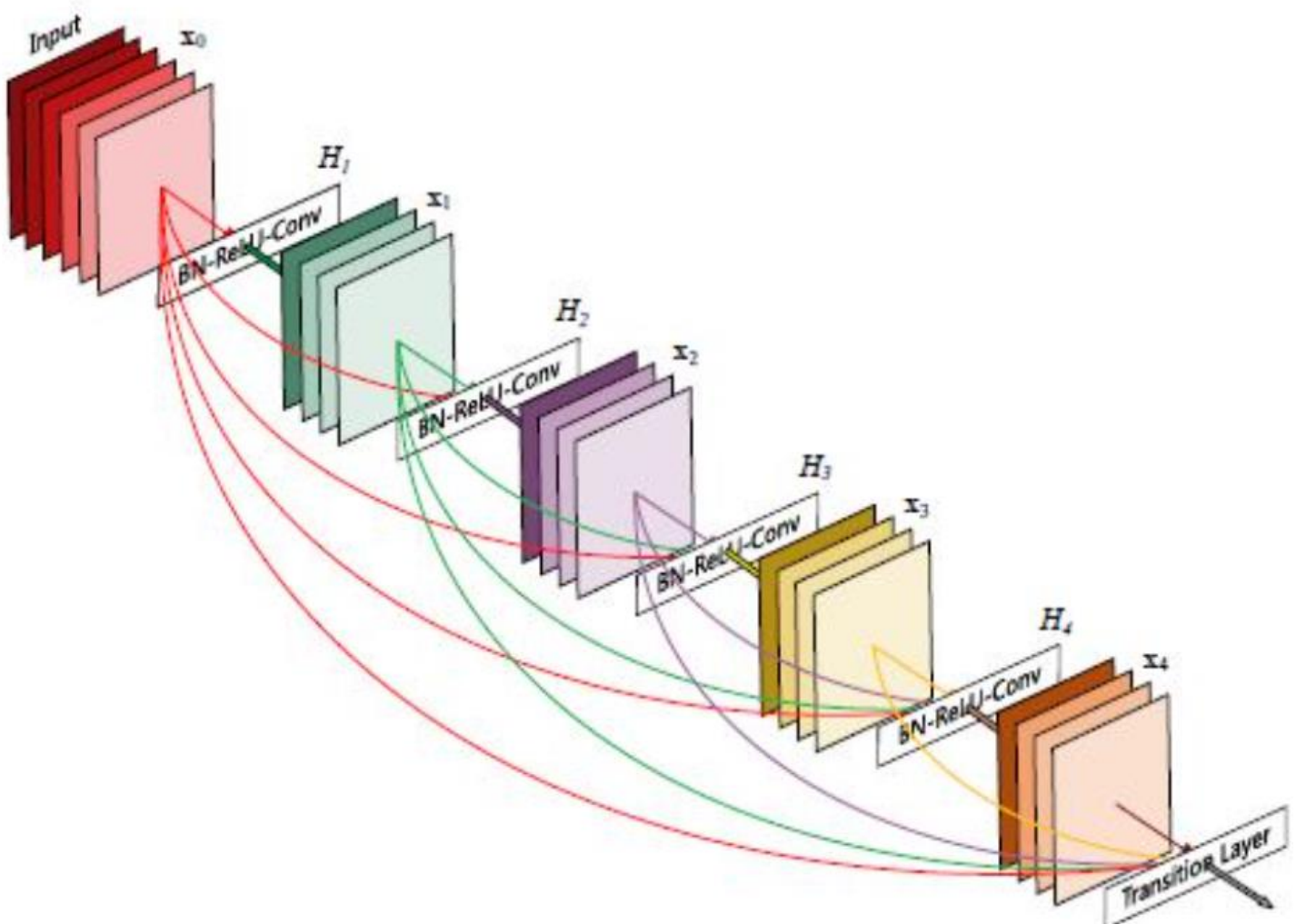


Figure 3.8 : Bloc dense à 5 couches avec un taux de croissance de $k = 4$ [64]

1. Composants du modèle DenseNet

Les composants de DenseNet [64, 65] incluent :

- Une Connectivité ;
- Des Blocs denses ;
- Un Taux de croissance ;
- Des Couches de goulot d'étranglement.

a. Connectivité

Dans chaque couche, les cartes d'entités de toutes les couches précédentes ne sont pas additionnées, mais concaténées et utilisées comme entrées. Par conséquent, les DenseNets nécessitent moins de paramètres qu'un CNN traditionnel équivalent.

Ce qui permet la réutilisation des fonctionnalités, car les cartes de fonctionnalités redondantes sont supprimées. Ainsi, la $l^{\text{ième}}$ couche reçoit les feature-maps de toutes les couches précédentes, x_0, \dots, x_{l-1} , en entrée, l'équation 3.1 correspond à cette connectivité.

$$X_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad 3.1$$

Où $[x_0, x_1, \dots, x_{l-1}]$ est la concaténation des feature-maps, c'est-à-dire la sortie produite dans toutes les couches précédentes $l(0, \dots, l-1)$. Les multiples entrées de H_l sont concaténées en un seul tenseur, pour faciliter la mise en œuvre.

b. Blocs denses

L'utilisation de l'opération de la concaténation n'est pas possible lorsque la taille des cartes d'entités change. Cependant, une partie essentielle des CNN, est le sous-échantillonnage des couches qui réduit la taille des cartes de caractéristiques grâce à la réduction de la dimensionnalité, pour gagner des vitesses de calcul plus élevées.

Pour permettre cela, les DenseNets sont divisés en DenseBlocks, où les dimensions des cartes d'entités restent constantes dans un bloc, mais le nombre de filtres entre eux, est modifié. Les couches entre les blocs sont appelées couches de transition, qui réduisent le nombre de canaux à la moitié de celui des canaux existants.

Pour chaque couche, à partir de l'équation 3.1, H_l est défini comme une fonction composite qui applique trois opérations consécutives : une normalisation par lots (BN, batch normalisation), une unité linéaire rectifiée (ReLU) et une convolution (Conv).

Dans la figure 3.9, un DenseNet profond avec trois blocs denses est affiché. Les couches entre deux blocs adjacents, sont les couches de transition qui effectuent un sous-échantillonnage (c'est-à-dire modifient la taille des cartes de caractéristiques), via des opérations de convolution

et de regroupement, tandis que dans le bloc dense, la taille des cartes de caractéristiques, est la même pour permettre la concaténation des caractéristiques.

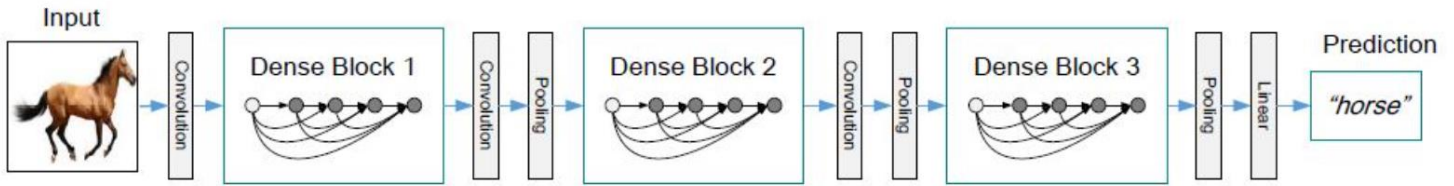


Figure 3.9 : DenseNet profond avec trois blocs denses [65]

c. Taux de croissance

On peut considérer les fonctionnalités comme un état global du réseau. La taille de la carte d'entités augmente après un passage à travers chaque couche dense, chaque couche ajoutant des entités "K" au-dessus de l'état global (entités existantes). Ce paramètre « K » est appelé le taux de croissance du réseau, qui régule la quantité d'informations ajoutées dans chaque couche du réseau. Si chaque fonction H_1 produit k cartes d'entités, alors la $l^{ème}$ couche a K_j cartes de caractéristiques d'entrée.

$$K_j = K_0 + K * (l - 1) \quad 3.2$$

Où k_0 est le nombre de canaux dans la couche d'entrée. Contrairement aux architectures de réseau existantes, les DenseNets peuvent avoir des couches très étroites.

d. Couches de goulot d'étranglement

Bien que chaque couche ne produise que k cartes d'entités en sortie, le nombre d'entrées peut être assez élevé, en particulier pour les couches supplémentaires. Ainsi, une couche de convolution 1x1 peut être introduite comme couche de goulot d'étranglement avant chaque convolution 3x3 pour améliorer l'efficacité et la vitesse des calculs.

2. Architecture DenseNet-121

L'architecture du réseau Densenet121 [65, 66], est illustrée par la figure 3.10 et détaillée dans le tableau 3.1.

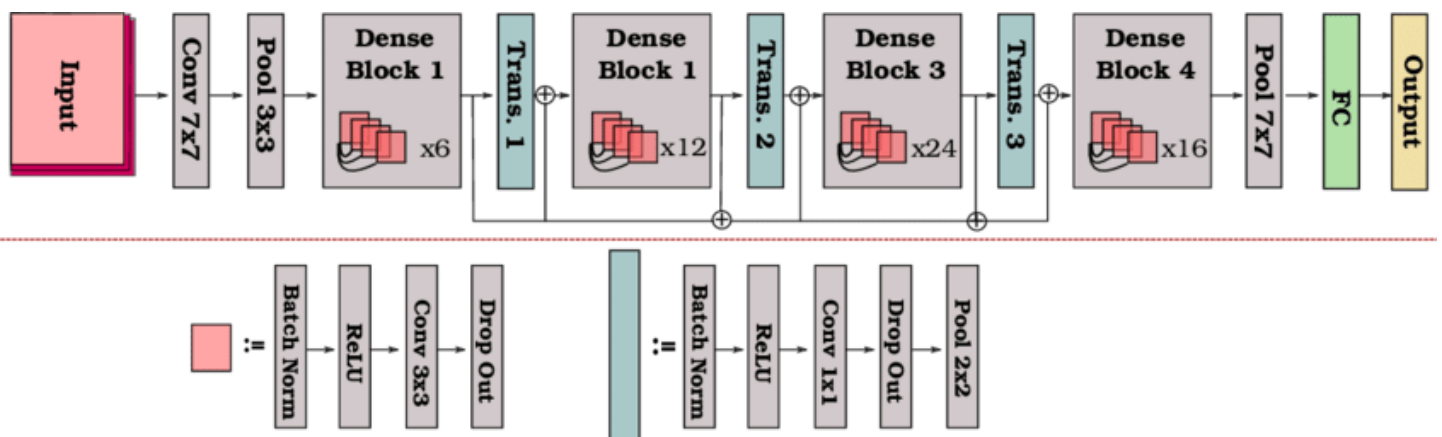


Figure 3.10 : Illustration schématique de l'architecture DenseNet-121 [66]

Chapitre 3 Conception d'un système d'aide à la décision pour la détection du cancer du sein dans les images échographiques

Dans l'architecture DenseNet-121, nous pouvons voir que 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.

De plus, chaque couche de transition a une couche convolutive 1x1 et une couche de regroupement moyenne 2x2, avec une foulée de 2.

Ainsi, les couches présentes dans le modèle Densenet121 sont les suivantes :

- a. Couche de convolution de base avec 64 filtres de taille 7X7 et une foulée de 2;
- b. Couche de mise en commun de base avec une mise en commun maximale de 3x3 et une foulée de 2;
- c. 'Dense Block 1' avec 2 circonvolutions répétées 6 fois;
- d. Couche de transition 1 (1 Conv + 1 AvgPool);
- e. 'Dense Block 2' avec 2 circonvolutions répétées 12 fois;
- f. Couche de transition 2 (1 Conv + 1 AvgPool);
- g. 'Dense Block 3' avec 2 circonvolutions répétées 24 fois;
- h. Couche de transition 3 (1 Conv + 1 AvgPool);
- i. 'Dense Block 4' avec 2 circonvolutions répétées 16 fois;
- j. Couche 'Global Average Pooling', accepte toutes les cartes d'entités du réseau pour effectuer la classification.
- k. Couche de sortie

Par conséquent, DenseNet-121 a les couches suivantes :

- 1 convolution de 7x7;
- 58 Convolutions de 3x3;
- 61 Convolutions de 1x1;
- 4 MoyPool;
- 1 couche entièrement connectée.

En bref, DenseNet-121 à 120 convolutions et 4 AvgPool.

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 3.2 : Architecture de Densenet121 [65]

Toutes les couches, c'est-à-dire celles qui se trouvent dans le même bloc dense et les couches de transition, répartissent leurs poids sur plusieurs entrées, ce qui permet aux couches plus profondes, d'utiliser les fonctionnalités extraites dès le début.

Étant donné que les couches de transition produisent de nombreuses caractéristiques redondantes, les couches des deuxième et troisième blocs denses, attribuent le moins de poids à la sortie des couches de transition. De plus, même si les poids de l'ensemble du bloc dense sont utilisés par les couches finales, il peut encore y avoir plus de caractéristiques de haut niveau générées plus profondément dans le modèle, car il peut y avoir une concentration plus élevée, vers les cartes de caractéristiques finales dans les expériences.

3. Architecture DenseNet201

L'architecture du modèle Densenet201 est illustrée par la figure 3.11. Ce réseau est représenté par 201 couches [64].

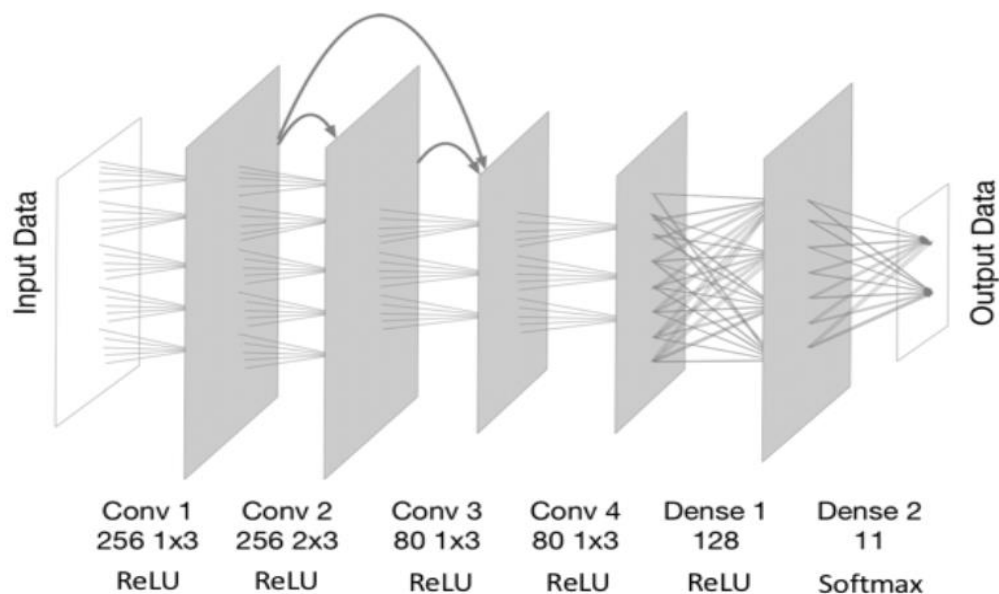


Figure 3.11 : Illustration schématique de l'architecture DenseNet201 [64]

3.5 Fine tuning

Le fine tuning permet des réglages spécifiques dans les modèles des réseaux de neurones de l'apprentissage par transfert (transfer learning), suivant des ajustements progressifs et des gels de certains poids du modèle pré-entraîné [52, 67].

L'application d'un fine tuning nous permet d'utiliser des réseaux préformés, pour reconnaître les classes sur lesquelles ils n'ont pas été formés à l'origine.

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.

3.5.3 Fonctionnement du Fine Tuning

Le Fine Tuning (figure 3.12) comprend les étapes suivantes :

- Considérer un modèle de réseau neuronal pré-entraîné, c'est-à-dire le modèle source, sur un ensemble de données source ImageNet.
- Créer un nouveau modèle de réseau de neurones, c'est-à-dire le modèle cible. Cela copie toutes les conceptions du modèle et leurs paramètres sur le modèle source, à l'exception de la couche de sortie.
- Ajouter une couche de sortie au modèle cible, dont le nombre de sorties correspond au nombre de catégories dans le jeu de données cible. Initialiser ensuite aléatoirement les paramètres du modèle de cette couche.
- Entraîner le modèle cible sur le jeu de données cible. La couche de sortie sera formée à partir de zéro, tandis que les paramètres de toutes les autres couches sont affinés en fonction des paramètres du modèle source.

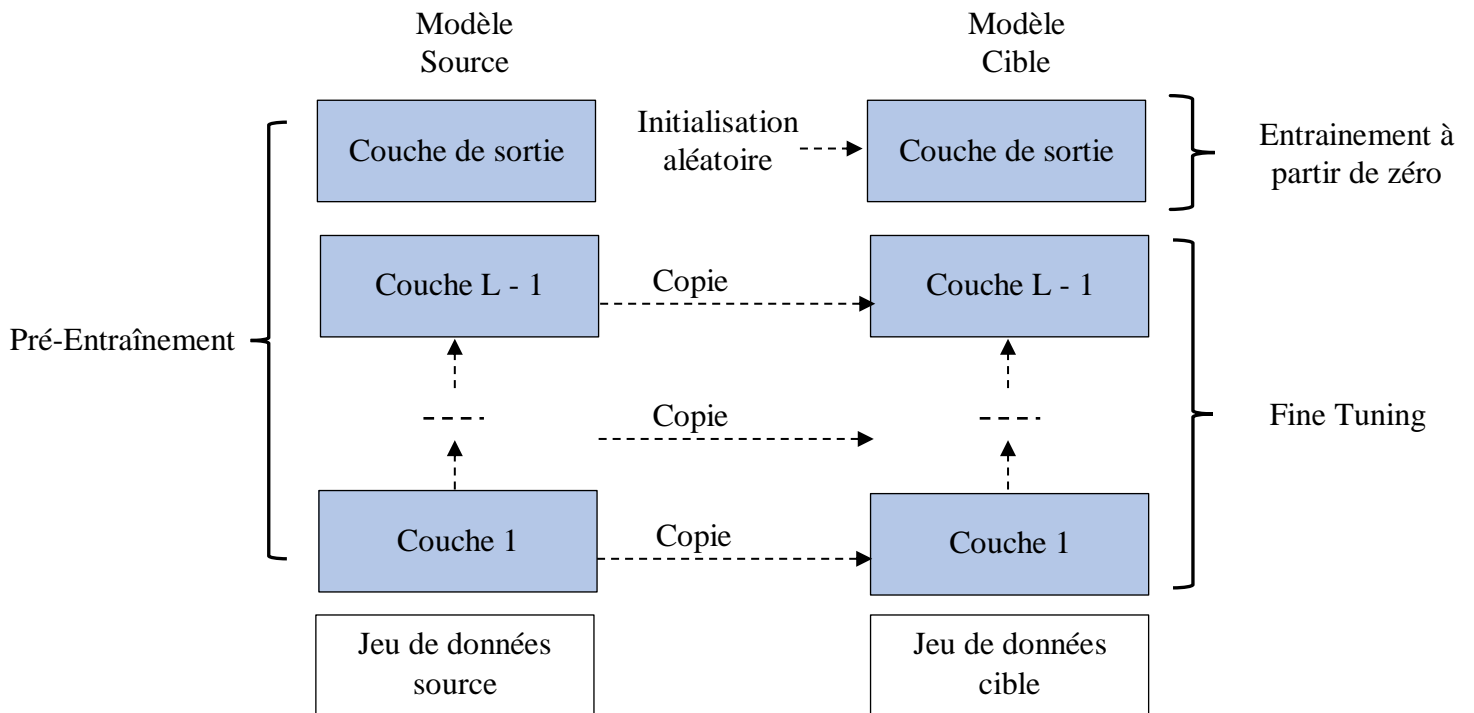


Figure 3.12 : Fine Tuning [67]

3.6 Conclusion

Ce chapitre a été consacré à l'étude conceptuelle d'un système de détection de tumeurs mammaires suivant 5 modèles des réseaux de neurones pré-entraînés.

Le dernier chapitre, présente l'environnement de travail, la mise en œuvre du système d'aide à la décision en imagerie échographique, les résultats obtenus, ainsi qu'une comparaison entre les modèles.

Chapitre 4 Implémentation du Système de la Classification de Tumeurs Mammaires dans les Images Echographiques

4.1 Introduction

Ce chapitre, est consacré à la réalisation du système de classification de tumeurs mammaires dans les images échographiques, suivant six modèles des réseaux convolutifs pré-entraînés. Ces modèles, ont été formés suivant plusieurs hyperparamètres donnés, tels que la fonction d'optimisation.

Les résultats les plus importants seront présentés, ainsi que les obstacles rencontrés et les solutions appliquées afin d'aboutir à l'objectif visé à savoir un taux de précision satisfaisant, pour un domaine aussi exigeant que la classification de tumeurs mammaires.

Les résultats obtenus par les six réseaux adoptés par le VGG16, le VGG19, le Resnet50, le Resnet152, le Densenet121 et le Densenet201, sont comparés pour déterminer le meilleur.

4.2 Environnement de travail

Le système de la classification d'images mammaires échographiques 'BreastUltraCNN4', a été développé suivant le langage de programmation python, sous environnement Google Colab.

4.2.1 Matériel Utilisé

Le matériel utilisé dans le cadre de ce mémoire, est un laptop de la marque ASUS ROG (Republic Of Gamers), de son nom de série 'ASUS ROG GL753VE', dont les caractéristiques sont présentées dans le tableau ci-dessous :

Processeur	Intel Core i7-7700HQ Quad-Core 2.8 GHz / 3.8 GHz Turbo - Cache 6 Mo
Mémoire vive RAM	16 Go DDR4
Ecran	17.3" Full HD 1920 x 1080 pixels
Carte Graphique	NVIDIA GeForce GTX 1050 TI 4 Go de mémoire dédiée GDDR5
Disque dur	SSD M.2 SATA 128 Go 7200 RPM + HDD 1 To 5400 RPM
Système d'exploitation	Windows 10 Famille 64 bits

Tableau 4.1 : Résumé des caractéristiques de la machine utilisée

4.2.2 Langage de programmation Python

Python est le langage de programmation le plus utilisé en deep learning (DL) depuis ces dernières années, avec le langage R [68, 69].

Nous avons choisi python pour les raisons suivantes :

- Disponibilité d'un grand nombre de bibliothèques dédiées pour le DL.
- Simplicité et fluidité de la syntaxe.

Remarque

Les bibliothèques de python sont innombrables, c'est pourquoi il n'est pas toujours facile de les nommer et citer leurs fonctionnalités. Nous citerons donc les plus importantes dans notre travail : matplotlib, Keras, numpy, tensorflow.

4.2.3 Google Colab

Google Colaboratory ou Colab (figure 4.1), est un outil de Google simple et gratuit initié au Deep Learning. Il permet de d'écrire et d'exécuter un code python dans un navigateur. Bien qu'il soit gratuit, la fonctionnalité qui le rend aussi populaire, est qu'il offre l'accès à un processeur graphique GPU totalement gratuit [70].

Remarques

- Ces ressources sont extensibles en fonction des besoins et du nombre d'utilisateurs, qui sont connectés à la fois.
- Les données peuvent être récupérées à partir de l'espace de stockage de la machine locale, de Google Drive ou de la mémoire temporaire de Google Colab.



Figure 4.1: Interaction machine/drive/google colab [70]

4.2.4 Bibliothèques utilisées

Pour le bon déroulement de ce projet, un certain nombre de bibliothèques ont été utilisés. Elles sont définies dans cette section.

1. Numpy

NumPy [71] est le package fondamental pour le calcul scientifique en Python. Elle étend les capacités de Python, pour travailler sur des tableaux et des matrices à n dimensions, de façon bien plus optimisée. Elle offre ainsi, des fonctions mathématiques de haut niveau, telles que la transformée de Fourier discrète, les opérations statistiques de base, la simulation aléatoire et bien plus encore...

Elle est majoritairement écrite en C (pour les parties clés à optimiser) et Python.

2. Matplotlib

Matplotlib [72] est la bibliothèque de traçage pour Python qui fournit une API orientée objet, pour intégrer des tracés. Grâce à Matplotlib, on peut tracer des histogrammes, des graphes de fonctions ...etc.

Dans notre cas matplotlib nous a permis de tracer les courbes d'apprentissage et de la fonction *Loss*, pour visualiser l'avancement des modèles.

3. Tensorflow 2.3.0

TensorFlow [73] 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. 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.

4. Keras

Keras est une API de réseau de neurones écrite en langage Python. Il s'agit d'une bibliothèque Open Source, exécutée par-dessus des frameworks tels que Theano et TensorFlow.

Conçue pour être modulaire, rapide et simple d'utilisation, Keras a été créée par l'ingénieur François Chollet de Google [74]. Elle offre une façon simple et intuitive, de créer des modèles du deep learning.

Aujourd'hui, Keras est l'une des APIs les plus utilisées, pour le développement des réseaux de neurones.

4.3 Implémentation du Système ‘BreastUltraCNN4’

Le système ‘BreastUltraCNN4’ (figure 3.1) a été implémenté avec le langage python, pour la classification de tumeurs mammaires échographiques, suivant 5 réseaux de neurones pré-entraînés.

4.3.1 Description du Data-set

L'ensemble de données d'échographie mammaire BUSI [49], est regroupé en trois classes : images normales, bénignes et malignes.

La figure ci-dessous représente, un exemple de chaque catégorie.

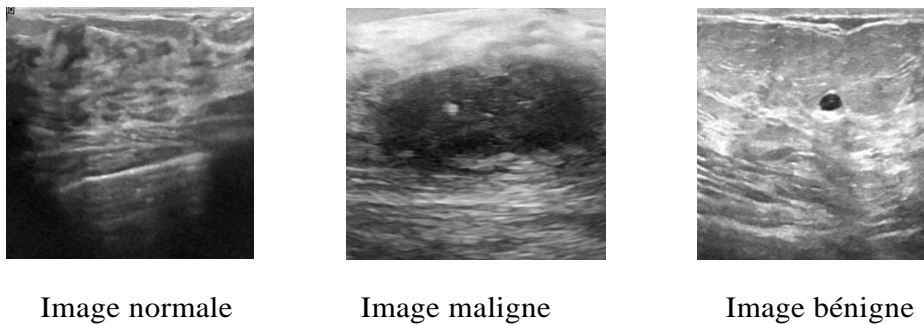


Figure 4.2 : Exemples d'images de la base BUSI [49]

4.3.2 Préparation des données

La préparation des données passe par plusieurs étapes, décrites dans les différents paragraphes.

1. Répartition de la base

Avant d'entamer le traitement de la base de données, il est primordial de la diviser en trois ensembles de données par l'entraînement, la validation et le test (figure 4.3).

80% pour l'entraînement, 10% pour la validation, et 10% pour le test.

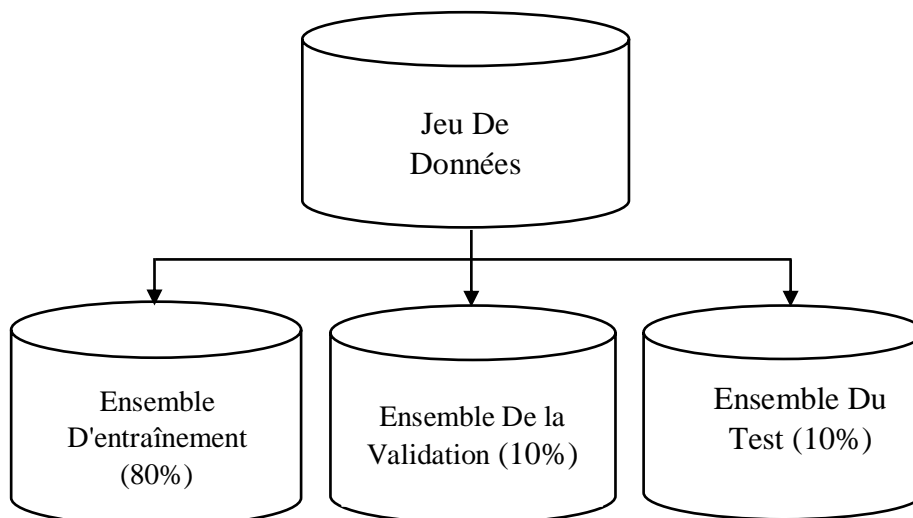


Figure 4.3 : Répartition de la base

2. Augmentation de données

L'augmentation des données est une technique qui peut être utilisée, pour augmenter artificiellement la taille d'un ensemble de données, en créant des versions modifiées des images dans la base initiale, pour maximiser la précision.

Pour effectuer cette tâche, nous avons eu recours à 'Keras ImageDataGenerator', qui permet de faire de l'augmentation de données par plusieurs fonctions telles que : le Zoom, la rotation, la translation, etc...

3. Paramètres de prétraitement utilisés

Les paramètres utilisés pour la réalisation du système 'BreastUltraCNN4', sont donnés ci-dessous :

➤ Taille d'image

Il est recommandé d'utiliser une taille d'image de 224x224, car les entrées des CNN sont des images de 224x224 pixels.

➤ Taille du batch

C'est la taille du lot en français, c'est le nombre d'exemples d'apprentissage en une seule étape. Plus la taille du lot est élevée, plus nous aurons besoin d'espace mémoire.

En termes pratiques, pour déterminer la taille du lot optimale, il est recommandé d'essayer d'abord des tailles de lot petites (généralement 32 ou 64), en gardant également à l'esprit que les petites tailles de lot, nécessitent de petits taux d'apprentissage. Le nombre de tailles de lots doit être une puissance de 2, pour tirer pleinement parti du traitement des GPU. Dans notre cas nous choisissons 32.

➤ Nombre d'époques

Dans la terminologie des réseaux de neurones, une époque est définie comme le moment où un ensemble de données, est transmis en avant et en arrière, via le réseau de neurones une seule fois.

Une partie des données d'apprentissage est dédiée à la validation du modèle, pour vérifier les performances du modèle après chaque époque d'apprentissage. La perte et la précision sur l'ensemble d'apprentissage ainsi que sur l'ensemble de validation, sont surveillées pour examiner le numéro d'époque, après lequel le modèle commence le sur ajustement. Le nombre choisi est =30.

➤ **Taux d'apprentissage**

Le taux d'apprentissage (**LEARNING_RATE, lr**) est un hyperparamètre qui contrôle le degré de modification, du modèle en réponse à l'erreur estimée, chaque fois que les poids du modèle sont mis à jour.

Le choix du taux d'apprentissage est difficile, car une valeur trop petite peut entraîner un long processus d'entraînement qui pourrait rester bloqué ; tandis qu'une valeur trop grande peut entraîner l'apprentissage trop rapide, d'un ensemble de poids sous-optimal ou un processus d'entraînement instable.

Après avoir commencé avec des valeurs basses comme $lr = 0.1$ et 0.01 , la valeur la mieux appropriée est 0.0001 (**LEARNING_RATE = 0.0001**).

➤ **Dropout**

Le dropout peut être utilisé après les couches convolutionnelles et après les couches de pooling. Souvent, le dropout n'est utilisé qu'après les couches de pooling, mais il ne s'agit que d'une heuristique grossière. Dans ce cas, le dropout est appliqué à chaque élément ou cellule des cartes de caractéristiques.

Après avoir essayé les valeurs 0.5 , 0.4 , 0.3 , et 0.2 , les meilleurs taux de précision et de perte ont été obtenus avec 0.2 .

4.4 Résultats et Interprétation

Après initialisation de tous les paramètres vus ci-dessus, l'importation des modèles et de leurs poids, l'entraînement des modèles s'est fait. La durée de l'entraînement se compte en heures. Voici les résultats obtenus pour chaque modèle choisi pour la réalisation de BreastUltraCNN4.

4.4.1 Résultats de VGG16 et VGG19

Les figures 4.4 et 4.5 représentent respectivement, les architectures de VGG16 et de VGG19, sur lesquelles on peut voir le nombre de paramètres entraînable et non-entraînable. Les paramètres entraînaux sont ceux, dont la valeur est ajustée/modifiée pendant l'entraînement en fonction de leur gradient. Les paramètres non entraînaux par contre, sont ceux dont la valeur n'est pas optimisée lors de l'entraînement selon leur gradient.

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 7, 7, 512)	14714688
flatten_14 (Flatten)	(None, 25088)	0
dense_70 (Dense)	(None, 512)	12845568
dropout_56 (Dropout)	(None, 512)	0
dense_71 (Dense)	(None, 256)	131328
dropout_57 (Dropout)	(None, 256)	0
dense_72 (Dense)	(None, 128)	32896
dropout_58 (Dropout)	(None, 128)	0
dense_73 (Dense)	(None, 64)	8256
dropout_59 (Dropout)	(None, 64)	0
dense_74 (Dense)	(None, 3)	195

=====
Total params: 27,732,931
Trainable params: 20,097,667
Non-trainable params: 7,635,264

Figure 4.4 : Modèle VGG16

Nous remarquons que le nombre de paramètres entraînable est de 20 097 667 et le nombre de paramètres non-entraînable, est de 7 635 264, ce qui fait un total de 27 732 931 paramètres pour l'ensemble.

Layer (type)	Output Shape	Param #
vgg19 (Functional)	(None, 7, 7, 512)	20024384
flatten_1 (Flatten)	(None, 25088)	0
dense_5 (Dense)	(None, 512)	12845568
dropout_4 (Dropout)	(None, 512)	0
dense_6 (Dense)	(None, 256)	131328
dropout_5 (Dropout)	(None, 256)	0
dense_7 (Dense)	(None, 128)	32896
dropout_6 (Dropout)	(None, 128)	0
dense_8 (Dense)	(None, 64)	8256
dropout_7 (Dropout)	(None, 64)	0
dense_9 (Dense)	(None, 3)	195

=====
Total params: 33,042,627
Trainable params: 13,018,243
Non-trainable params: 20,024,384

Figure 4.5 : Modèle VGG19

Nous remarquons ici que le nombre de paramètres entraînable, est de 13 018 243 et le nombre de paramètres non-entraînable est de 20 024 384, ce qui fait un total de 33 042 627 paramètres pour l'ensemble.

1. Entraînement

Les figures ci-dessous montrent le comportement des modèles VGG16 et VGG19 lors de l'entraînement, tout au long des 30 époques. Pour un gain d'espace, nous n'avons affiché que les 5 premières et 5 dernières époques de l'entraînement. La précision qu'on peut voir 'acc : ' et les autres résultats comme la validation et la perte 'loss', sont affichés à chaque fois.

```
Epoch 1/30
39/39 [=====] - 901s 23s/step - loss: 0.9161 - acc: 0.5662 - val_loss: 0.3040 - val_acc: 0.8984
Epoch 2/30
39/39 [=====] - 885s 23s/step - loss: 0.5885 - acc: 0.7410 - val_loss: 0.2640 - val_acc: 0.9141
Epoch 3/30
39/39 [=====] - 880s 23s/step - loss: 0.5212 - acc: 0.7868 - val_loss: 0.3130 - val_acc: 0.8281
Epoch 4/30
39/39 [=====] - 903s 23s/step - loss: 0.4217 - acc: 0.8173 - val_loss: 0.2369 - val_acc: 0.9219
Epoch 5/30
39/39 [=====] - 884s 23s/step - loss: 0.3978 - acc: 0.8284 - val_loss: 0.3021 - val_acc: 0.8828
- . . . . .
Epoch 25/30
39/39 [=====] - 897s 23s/step - loss: 0.1553 - acc: 0.9469 - val_loss: 0.3490 - val_acc: 0.9141
Epoch 26/30
39/39 [=====] - 895s 23s/step - loss: 0.1674 - acc: 0.9428 - val_loss: 0.2771 - val_acc: 0.9062
Epoch 27/30
39/39 [=====] - 898s 23s/step - loss: 0.1806 - acc: 0.9412 - val_loss: 0.3395 - val_acc: 0.8750
Epoch 28/30
39/39 [=====] - 935s 24s/step - loss: 0.1787 - acc: 0.9493 - val_loss: 0.2429 - val_acc: 0.8984
Epoch 29/30
39/39 [=====] - 930s 24s/step - loss: 0.1543 - acc: 0.9542 - val_loss: 0.2030 - val_acc: 0.9297
Epoch 30/30
39/39 [=====] - 926s 24s/step - loss: 0.1857 - acc: 0.9395 - val_loss: 0.2510 - val_acc: 0.9062
```

Figure 4.6 : Entraînement pour vgg16

```
Epoch 1/30
39/39 [=====] - 901s 23s/step - loss: 0.9869 - acc: 0.5351 - val_loss: 0.7401 - val_acc: 0.7969
Epoch 2/30
39/39 [=====] - 891s 23s/step - loss: 0.8690 - acc: 0.5989 - val_loss: 0.5025 - val_acc: 0.8359
Epoch 3/30
39/39 [=====] - 892s 23s/step - loss: 0.7763 - acc: 0.6528 - val_loss: 0.4133 - val_acc: 0.8125
Epoch 4/30
39/39 [=====] - 893s 23s/step - loss: 0.6796 - acc: 0.7092 - val_loss: 0.3701 - val_acc: 0.8906
Epoch 5/30
39/39 [=====] - 888s 23s/step - loss: 0.6322 - acc: 0.7361 - val_loss: 0.3176 - val_acc: 0.8594
- . . . . .
Epoch 25/30
39/39 [=====] - 942s 24s/step - loss: 0.4728 - acc: 0.7949 - val_loss: 0.3086 - val_acc: 0.8594
Epoch 26/30
39/39 [=====] - 943s 24s/step - loss: 0.4511 - acc: 0.8031 - val_loss: 0.2737 - val_acc: 0.8750
Epoch 27/30
39/39 [=====] - 944s 24s/step - loss: 0.4403 - acc: 0.8154 - val_loss: 0.2710 - val_acc: 0.8984
Epoch 28/30
39/39 [=====] - 942s 24s/step - loss: 0.4437 - acc: 0.8227 - val_loss: 0.3658 - val_acc: 0.8281
Epoch 29/30
39/39 [=====] - 971s 25s/step - loss: 0.4203 - acc: 0.8162 - val_loss: 0.2362 - val_acc: 0.9141
Epoch 30/30
39/39 [=====] - 954s 25s/step - loss: 0.4033 - acc: 0.8391 - val_loss: 0.3020 - val_acc: 0.8750
```

Figure 4.7 : Entraînement pour vgg19

2. Détails des architectures et résultats

Le tableau ci-dessous résume tous les paramètres utilisés dans les deux modèles VGG16 et VGG19, ainsi que les résultats obtenus. Les figures 4.8 et 4.9 illustrent ces résultats.

	Vgg16	Vgg19
Nombre de couche	16	19
Nombre d'itérations	30	30
Répartition du Data-Set	80% Train, 10% Validation, 10% Test	80% Train, 10% Validation, 10% Test
Taux d'apprentissage	LR (0,0001)	LR (0,0001)
Taille du batch	32	32
Algorithme d'optimisation	Adam	Adam
DropOut	0.2	0.2
Taille de l'image	(224 x 224 x 3)	(224 x 224 x 3)
Précision (Accuracy)	Accuracy = 95.42 % Validation = 95.31 % Test = 90.78 % Loss = 22.15 %	Accuracy = 83.91 % Validation = 92.97 % Test = 83.55 % Loss = 39.14 %

Tableau 4.2 : Paramètres utilisés et résultats obtenus pour les modèles VGG16 et VGG19

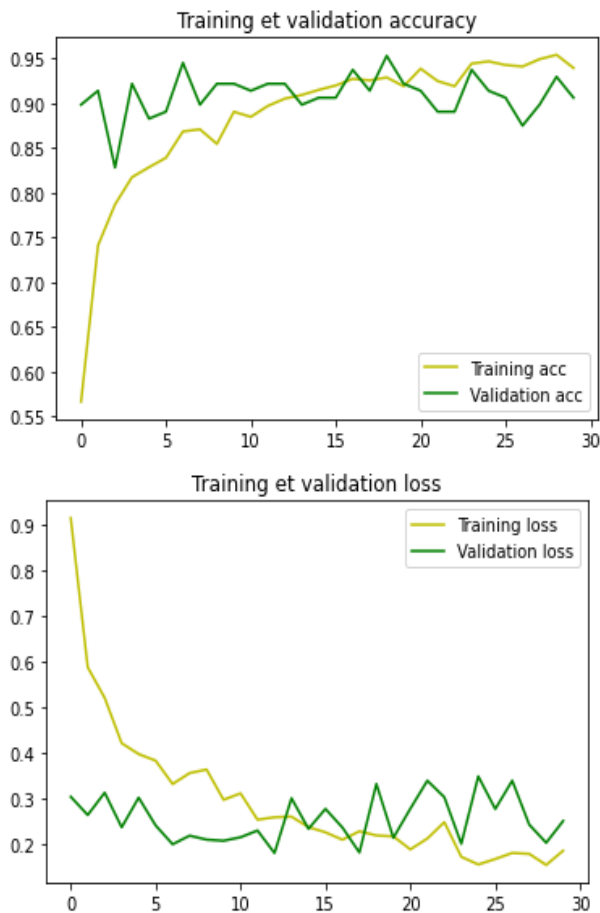


Figure 4.8 : Graphes des résultats de VGG16

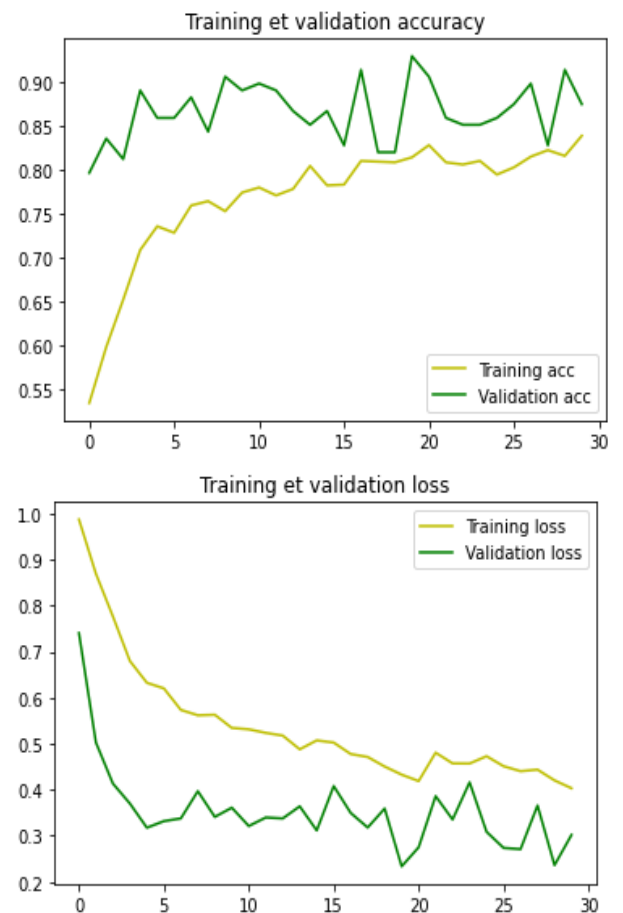


Figure 4.9 : Graphes des résultats de VGG19

A gauche, nous pouvons voir les graphes relatifs au modèle VGG16, qui représentent les graphes de précision (pour le haut), et de la perte (pour le bas). De même pour les graphes de droite, qui sont relatifs au modèle VGG19.

3. Analyse et critique de la méthode

Les deux modèles présentés ci-dessus ont délivré de bons résultats, bien qu'on puisse distinguer une légère différence, dans la précision en faveur de VGG16 avec un taux avoisinant les 95% contre un taux un peu plus bas, estimé à presque 84% pour le VGG19. Nous pouvons donc facilement dire que VGG16 l'emporte haut la main. Quant aux résultats la perte, les deux modèles ont générés des valeurs très basses, ils sont donc à égalité sur ce point.

4.4.2 Résultats de RESNET50 et RESNET152

Les figures 4.10 et 4.11, montrent les résultats des paramètres obtenus respectivement, pour les modèles ResNet50 et ResNet152.

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 7, 7, 2048)	23587712
flatten (Flatten)	(None, 100352)	0
batch_normalization (Batch Normalization)	(None, 100352)	401408
dense (Dense)	(None, 512)	51380736
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
dropout_2 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 64)	8256
dropout_3 (Dropout)	(None, 64)	0
dense_4 (Dense)	(None, 3)	195

=====
 Total params: 75,542,531
 Trainable params: 51,754,115
 Non-trainable params: 23,788,416

Figure 4.10 : Modèle ResNet50

Layer (type)	Output Shape	Param #
resnet152 (Functional)	(None, 1000)	60419944
flatten_1 (Flatten)	(None, 1000)	0
batch_normalization_1 (Batch Normalization)	(None, 1000)	4000
dense_5 (Dense)	(None, 512)	512512
dropout_4 (Dropout)	(None, 512)	0
dense_6 (Dense)	(None, 256)	131328
dropout_5 (Dropout)	(None, 256)	0
dense_7 (Dense)	(None, 128)	32896
dropout_6 (Dropout)	(None, 128)	0
dense_8 (Dense)	(None, 64)	8256
dropout_7 (Dropout)	(None, 64)	0
dense_9 (Dense)	(None, 3)	195

=====
 Total params: 61,109,131
 Trainable params: 687,187
 Non-trainable params: 60,421,944

Figure 4.11 : Modèle ResNet152

Chapitre 4 Implémentation du Système de la classification de tumeurs mammaires dans les Images Echographiques

Le nombre de paramètres de Resnet50 est nettement plus élevé, que pour les modèles VGGNets. Cela est dû au plus grand nombre de couches que possède le modèle ResNet50, par rapport aux modèles vus précédemment. Il y a un total de 75 542 531 de paramètres, dont 51 754 115 sont entraînaibles et les 23 788 416, sont non-entraînables.

Tout comme pour le ResNet50, on retrouve pour le ResNet152, un nombre assez important de paramètres qui s'élève à 61 109 131 au total. Mais seulement 687 187 sont entraînaibles contre 60 421 944 non entraînaibles. Cette dernière valeur est la plus haute, que nous avons vu jusqu'à maintenant pour des paramètres non entraînaibles.

1. Entraînement des modèles

Les figures 4.12 et 4.13 indiquent les détails de l'entraînement des modèles ResNet50 et Resnet152, ainsi que les résultats de la précision et de la perte.

```
Epoch 1/30
39/39 [=====] - 272s 7s/step - loss: 0.9435 - acc: 0.5637 - val_loss: 0.5959 - val_acc: 0.7812
Epoch 2/30
39/39 [=====] - 273s 7s/step - loss: 0.8268 - acc: 0.6348 - val_loss: 0.6092 - val_acc: 0.7969
Epoch 3/30
39/39 [=====] - 256s 7s/step - loss: 0.7677 - acc: 0.6577 - val_loss: 0.4447 - val_acc: 0.8203
Epoch 4/30
39/39 [=====] - 260s 7s/step - loss: 0.7379 - acc: 0.6787 - val_loss: 0.4477 - val_acc: 0.8281
Epoch 5/30
39/39 [=====] - 260s 7s/step - loss: 0.7593 - acc: 0.6740 - val_loss: 0.4534 - val_acc: 0.8281

.....

Epoch 25/30
39/39 [=====] - 260s 7s/step - loss: 0.6273 - acc: 0.7312 - val_loss: 0.3999 - val_acc: 0.8906
Epoch 26/30
39/39 [=====] - 260s 7s/step - loss: 0.6025 - acc: 0.7402 - val_loss: 0.4120 - val_acc: 0.8984
Epoch 27/30
39/39 [=====] - 259s 7s/step - loss: 0.6185 - acc: 0.7508 - val_loss: 0.4379 - val_acc: 0.8828
Epoch 28/30
39/39 [=====] - 259s 7s/step - loss: 0.5654 - acc: 0.7614 - val_loss: 0.3885 - val_acc: 0.9062
Epoch 29/30
39/39 [=====] - 259s 7s/step - loss: 0.6192 - acc: 0.7426 - val_loss: 0.4098 - val_acc: 0.8906
Epoch 30/30
39/39 [=====] - 258s 7s/step - loss: 0.6148 - acc: 0.7271 - val_loss: 0.4694 - val_acc: 0.8906
```

Figure 4.12 : Entraînement de ResNet50

```
Epoch 1/30
39/39 [=====] - 590s 15s/step - loss: 1.0695 - acc: 0.5645 - val_loss: 1.0043 - val_acc: 0.6953
Epoch 2/30
39/39 [=====] - 580s 15s/step - loss: 0.9875 - acc: 0.5662 - val_loss: 0.7890 - val_acc: 0.6953
Epoch 3/30
39/39 [=====] - 580s 15s/step - loss: 0.9465 - acc: 0.5645 - val_loss: 0.7779 - val_acc: 0.6953
Epoch 4/30
39/39 [=====] - 580s 15s/step - loss: 0.9273 - acc: 0.5686 - val_loss: 0.7546 - val_acc: 0.6953
Epoch 5/30
39/39 [=====] - 580s 15s/step - loss: 0.8898 - acc: 0.5817 - val_loss: 0.7547 - val_acc: 0.6953

.....

Epoch 25/30
39/39 [=====] - 585s 15s/step - loss: 0.7040 - acc: 0.6985 - val_loss: 0.4347 - val_acc: 0.8281
Epoch 26/30
39/39 [=====] - 586s 15s/step - loss: 0.7303 - acc: 0.6830 - val_loss: 0.4368 - val_acc: 0.8203
Epoch 27/30
39/39 [=====] - 585s 15s/step - loss: 0.7004 - acc: 0.7026 - val_loss: 0.4167 - val_acc: 0.8281
Epoch 28/30
39/39 [=====] - 587s 15s/step - loss: 0.7085 - acc: 0.6814 - val_loss: 0.4221 - val_acc: 0.8359
Epoch 29/30
39/39 [=====] - 586s 15s/step - loss: 0.7056 - acc: 0.6912 - val_loss: 0.4066 - val_acc: 0.8281
Epoch 30/30
39/39 [=====] - 585s 15s/step - loss: 0.7005 - acc: 0.6846 - val_loss: 0.4036 - val_acc: 0.8281
```

Figure 4.13 : Entraînement de ResNet152

Même si le taux de précision (80%) est presque similaire, il reste relativement bas. Nous distinguons une grande valeur pour la perte. Les résultats au final, restent assez bons dans l'ensemble.

2. Détails des architectures et résultats

Les mêmes paramètres utilisés pour les modèles VGGNets, sont utilisés ici pour les modèles ResNets. Le tableau ci-dessous résume tous les paramètres utilisés dans les deux modèles Resnet50 et Resnet152, ainsi que les résultats obtenus sous formes de graphes. Les figures 4.14 et 4.15 illustrent ces résultats.

	ResNet50	ResNet152
Nombre de couche	50	152
Nombre d'itérations	30	30
Répartition du Data-Set	80% Train, 10% Validation, 10% Test	80% Train, 10% Validation, 10% Test
Taux d'apprentissage	LR (0,0001)	LR (0,0001)
Taille du batch	32	32
Algorithme d'optimisation	Adam	Adam
DropOut	0.2	0.2
Taille de l'image	(224 x 224 x 3)	(224 x 224 x 3)
Précision (accuracy)	Accuracy = 76.14 % Validation = 90.62 % Test = 80.92 % Loss = 47.79 %	Accuracy = 70.26 % Validation = 83.59 % Test = 73.02 % Loss = 55.39 %

Tableau 4.3 : Paramètres utilisés et résultats obtenus pour les modèles ResNet50 et ResNet152

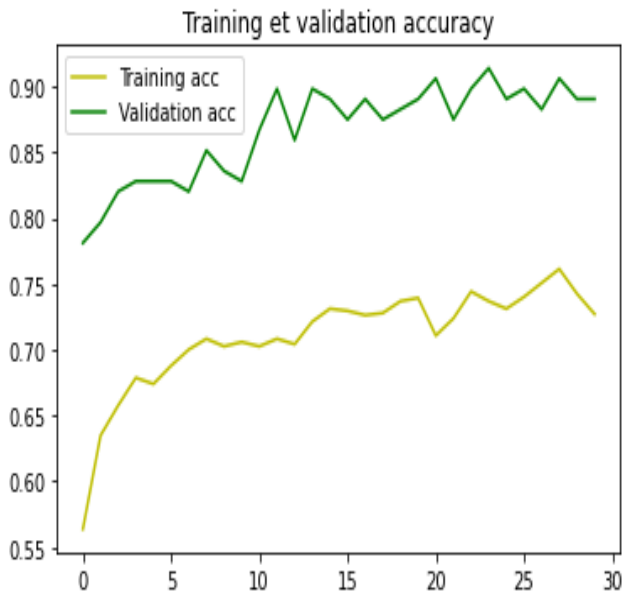


Figure 4.14 : Graphes des résultats de ResNet50

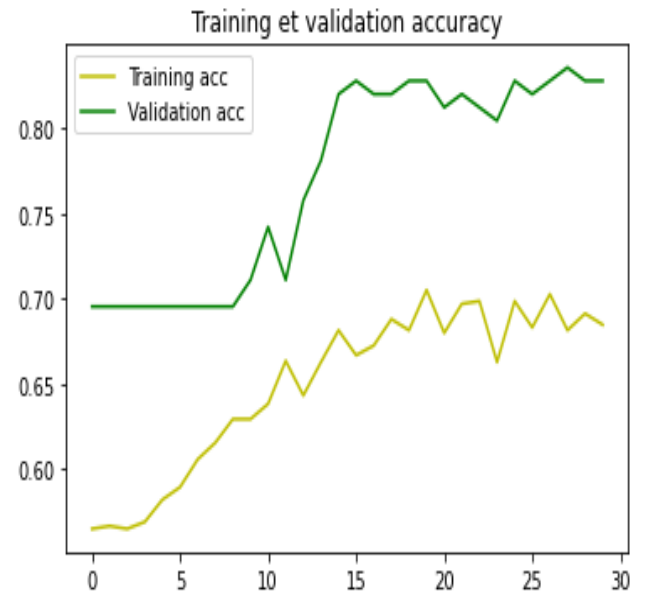
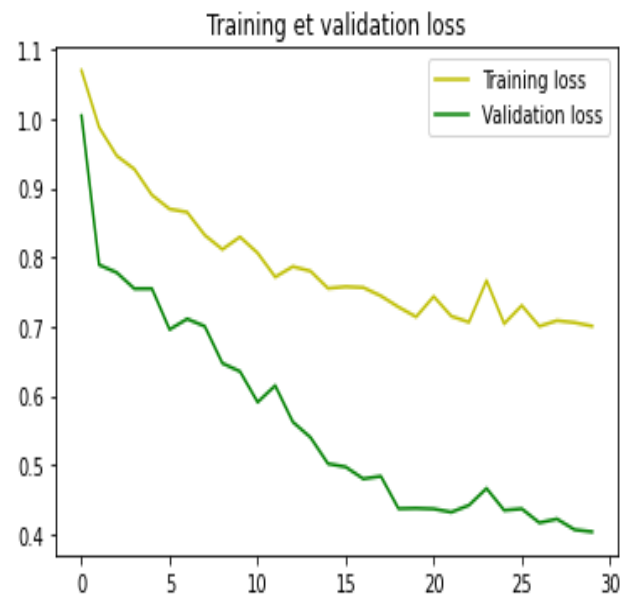
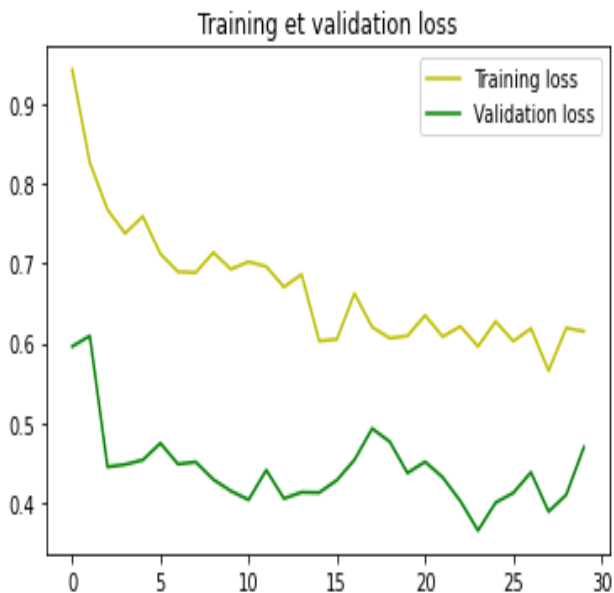


Figure 4.15 : Graphe des résultats ResNet152



Les résultats des deux modèles sont quasiment identiques. Même leurs graphes se ressemblent. Comme nous l'avons dit auparavant, les résultats sont un peu bas, mais restent assez bons dans l'ensemble.

3. Analyse et critique de la méthode

Les deux modèles ont généré des résultats assez bons, avec une précision avoisinant les 70% pour ResNet152, et 76% pour ResNet50. Les chiffres parlent d'eux même, ResNet50 est en tête en termes de précision. Quant aux valeurs de la perte, là encore nous avons des valeurs quasi similaires.

4.4.3 Résultats de DENSENET201 et DENSENET121

La figure 4.16 représente le détail des paramètres pour le modèle DenseNet121. Le nombre total de paramètres dans ce modèle, s'élève à 19 916 099 paramètres, dont 12 878 595 entraînaibles, et les 7 037 504 non-entraînables.

Layer (type)	Output Shape	Param #
densenet121 (Functional)	(None, 7, 7, 1024)	7037504
flatten_1 (Flatten)	(None, 50176)	0
dense_5 (Dense)	(None, 256)	12845312
dropout_4 (Dropout)	(None, 256)	0
dense_6 (Dense)	(None, 128)	32896
dropout_5 (Dropout)	(None, 128)	0
dense_7 (Dense)	(None, 3)	387

=====
Total params: 19,916,099
Trainable params: 12,878,595
Non-trainable params: 7,037,504

Figure 4.16 : Modèle DenseNet121

De même que pour la figure 4.17 qui relate quant à elle, les paramètres du 2^{ème} modèle DenseNet201, nous y remarquons une très grande différence dans le nombre de paramètres ; un total de paramètre de 67 040 451, pour 48 530 307 entraînaibles et 18 510 144 non-entraînables. Cette différence peut s'expliquer par la présence d'une couche supplémentaire dans DenseNet201, qui n'est pas présente dans DenseNet121. Car lors des premiers essais de ce dernier, il n'y a pas eu de différence dans les résultats à la fin, par contre le temps était plus long lors de l'ajout de cette couche. Pour DenseNet201, la suppression de cette couche a fait chuter le taux de précision et augmenter le taux de la perte. C'est pour cette raison, qu'elle est présente dans DenseNet201 et absente dans DenseNet121.

Layer (type)	Output Shape	Param #
densenet201 (Functional)	(None, 7, 7, 1920)	18321984
flatten (Flatten)	(None, 94080)	0
batch_normalization (Batch Normalization)	(None, 94080)	376320
dense (Dense)	(None, 512)	48169472
dropout (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 256)	131328
dropout_1 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 128)	32896
dropout_2 (Dropout)	(None, 128)	0
dense_3 (Dense)	(None, 64)	8256
dropout_3 (Dropout)	(None, 64)	0
dense_4 (Dense)	(None, 3)	195
=====		
Total params: 67,040,451		
Trainable params: 48,530,307		
Non-trainable params: 18,510,144		

Figure 4.17 : Modèle DenseNet201

2. Entraînement des deux modèles

L'entraînement du modèle DenseNet121 a duré aux alentours de 2 heures et 15 minutes. Un taux de précision avoisinant les 90% pour l'entraînement et un taux de précision, pour la validation, dépassant les 93% qui est un très bon résultat. Nous pouvons voir les résultats sur la figure 4.18, qui présente 10 des 30 étapes d'entraînement du modèle, dont les 5 premières et les 5 dernières.

```

Epoch 1/30
39/39 [=====] - 267s 7s/step - loss: 0.6956 - acc: 0.7059 - val_loss: 0.3800 - val_acc: 0.8304
Epoch 2/30
39/39 [=====] - 257s 7s/step - loss: 0.5498 - acc: 0.7582 - val_loss: 0.3518 - val_acc: 0.8237
Epoch 3/30
39/39 [=====] - 256s 7s/step - loss: 0.5135 - acc: 0.7786 - val_loss: 0.3218 - val_acc: 0.8750
Epoch 4/30
39/39 [=====] - 259s 7s/step - loss: 0.4636 - acc: 0.8117 - val_loss: 0.3014 - val_acc: 0.8795
Epoch 5/30
39/39 [=====] - 255s 7s/step - loss: 0.4386 - acc: 0.8056 - val_loss: 0.3296 - val_acc: 0.8482
.
.....

Epoch 25/30
39/39 [=====] - 261s 7s/step - loss: 0.2848 - acc: 0.8856 - val_loss: 0.1943 - val_acc: 0.9219
Epoch 26/30
39/39 [=====] - 261s 7s/step - loss: 0.3023 - acc: 0.8856 - val_loss: 0.1986 - val_acc: 0.9263
Epoch 27/30
39/39 [=====] - 260s 7s/step - loss: 0.2992 - acc: 0.8758 - val_loss: 0.2068 - val_acc: 0.9152
Epoch 28/30
39/39 [=====] - 261s 7s/step - loss: 0.3108 - acc: 0.8791 - val_loss: 0.2122 - val_acc: 0.9085
Epoch 29/30
39/39 [=====] - 261s 7s/step - loss: 0.2894 - acc: 0.8946 - val_loss: 0.2086 - val_acc: 0.9085
Epoch 30/30
39/39 [=====] - 261s 7s/step - loss: 0.2970 - acc: 0.8864 - val_loss: 0.1969 - val_acc: 0.9085

```

Figure 4.18 : Entraînement du modèle DenseNet121

Pour le modèle DenseNet201, la durée de son entraînement a été d'environ 3 heures (figure 4.19). Bien qu'il y ait plus de paramètres dans ce modèle, que dans l'autre et hormis le temps en plus, le taux de précision reste moins élevé que celui du modèle DenseNet121 ; avec une différence de 3% dans la précision, puisqu'il atteint les 87%. Quant à la validation, la précision est de 94% donc, ils sont quasiment pareils sur ce point.

```

Epoch 1/30
39/39 [=====] - 310s 8s/step - loss: 0.8136 - acc: 0.6593 - val_loss: 0.3360 - val_acc: 0.8906
Epoch 2/30
39/39 [=====] - 296s 8s/step - loss: 0.6312 - acc: 0.7255 - val_loss: 0.2777 - val_acc: 0.8906
Epoch 3/30
39/39 [=====] - 301s 8s/step - loss: 0.6446 - acc: 0.7443 - val_loss: 0.2849 - val_acc: 0.8750
Epoch 4/30
39/39 [=====] - 316s 8s/step - loss: 0.5614 - acc: 0.7684 - val_loss: 0.2433 - val_acc: 0.9062
Epoch 5/30
39/39 [=====] - 295s 8s/step - loss: 0.5727 - acc: 0.7819 - val_loss: 0.2512 - val_acc: 0.8828
.
.....

Epoch 25/30
39/39 [=====] - 301s 8s/step - loss: 0.3933 - acc: 0.8399 - val_loss: 0.2155 - val_acc: 0.9453
Epoch 26/30
39/39 [=====] - 302s 8s/step - loss: 0.3634 - acc: 0.8472 - val_loss: 0.2143 - val_acc: 0.9375
Epoch 27/30
39/39 [=====] - 301s 8s/step - loss: 0.3485 - acc: 0.8578 - val_loss: 0.2329 - val_acc: 0.9297
Epoch 28/30
39/39 [=====] - 301s 8s/step - loss: 0.3801 - acc: 0.8693 - val_loss: 0.2138 - val_acc: 0.8984
Epoch 29/30
39/39 [=====] - 301s 8s/step - loss: 0.3770 - acc: 0.8538 - val_loss: 0.2288 - val_acc: 0.9141
Epoch 30/30
39/39 [=====] - 301s 8s/step - loss: 0.3345 - acc: 0.8627 - val_loss: 0.2246 - val_acc: 0.9219

```

Figure 4.19 : Entraînement du modèle DenseNet201

2. Détail des architectures et résultats

Le tableau 4.3 représente les détails des paramètres utilisés dans les deux modèles Densenet121 et Densenet201, ainsi que les résultats obtenus sous formes de graphes. Les figures 4.20 et 4.21 illustrent ces résultats.

	DenseNet 121	DenseNet 201
Nombre de couche	50	152
Nombre d'itérations	30	30
Répartition du Data-Set	80% Train, 10% Validation, 10% Test	80% Train, 10% Validation, 10% Test
Taux d'apprentissage	LR (0,0001)	LR (0,0001)
Taille du batch	32	32
Algorithme d'optimisation	Adam	Adam
DropOut	0.2	0.2
Taille de l'image	(224 x 224 x 3)	(224 x 224 x 3)
Précision (accuracy)	Accuracy = 89.46 % Validation = 93.30 % Test = 87.82 % Loss = 28.91 %	Accuracy = 86.93 % Validation = 94.53 % Test = 84.21 % Loss = 42.30 %

Tableau 4.4 : Paramètres utilisés et résultats obtenus pour les modèles DenseNet121 et DenseNet201

Comme décrit précédemment, les mêmes paramètres ont été utilisés pour les deux modèles. Suite au tableau ci-dessus, nous constatons que leurs comportements est quasi similaire, avec une légère différence dans les résultats, sauf pour la perte qui est assez élevée pour le DenseNet201.

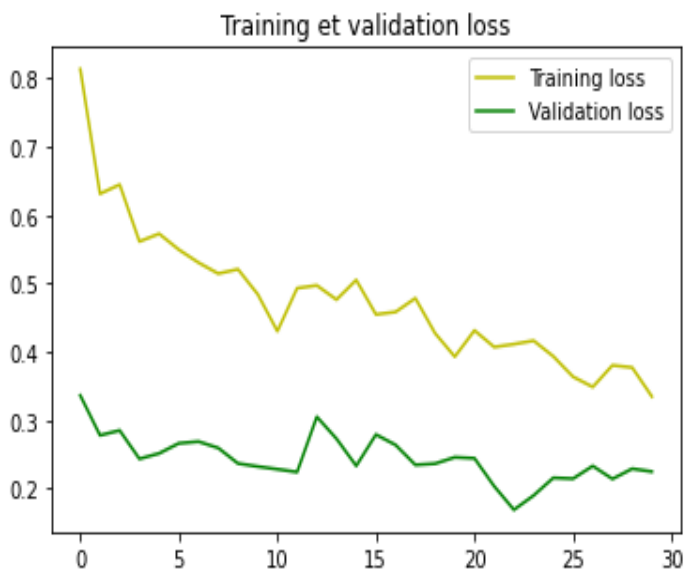
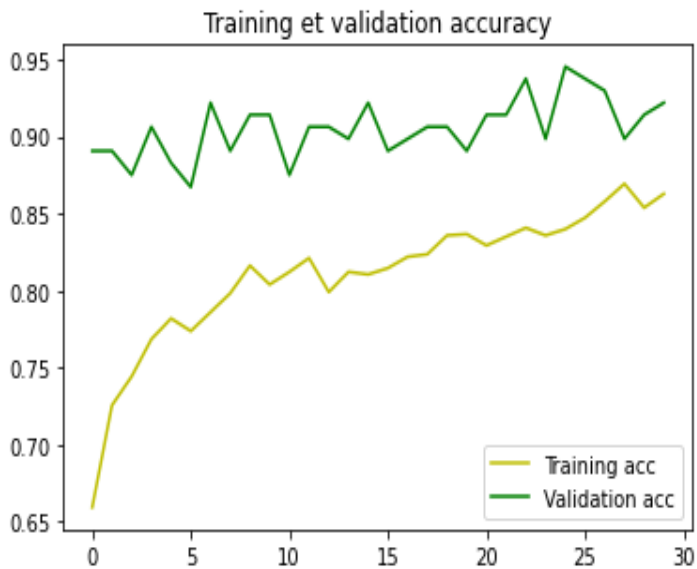


Figure 4.20 : Graphe du modèle DenseNet201

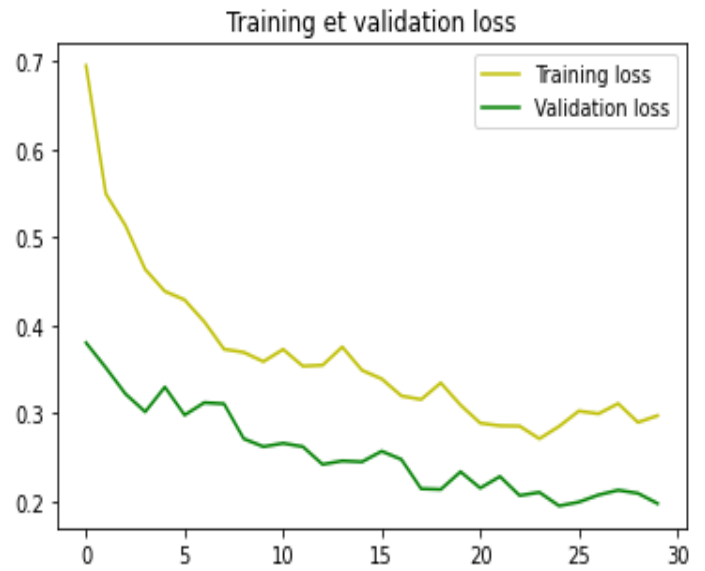
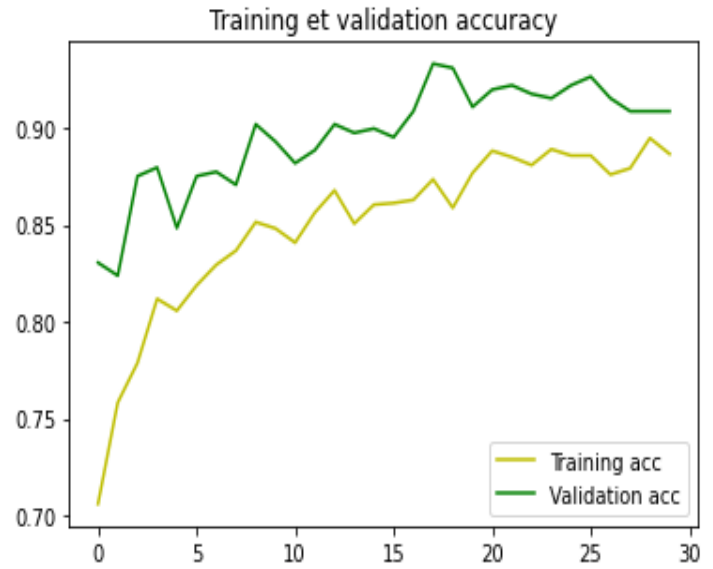


Figure 4.21 : Résultats du modèle DenseNet121

3. Analyse et critique de la méthode

Les deux modèles DenseNet121 et DenseNet201, ont délivré de résultats satisfaisants, avec une précision de 87% pour le DenseNet201 et un taux de 90% pour DenseNet121. Nous pouvons dire qu'ils sont en quasi égalité, en termes de précision. Pour la perte, le DenseNet201 se comporte un peu moins bien avec son taux élevé de plus de 40%.

4.4.4 Etude comparative

Les résultats obtenus pour le système ‘BreastUltraCNN4’, sont représentés dans le tableau 4.4.

	Vgg16	Vgg19	Densnet201	Densnet121	Resnet50	Resnet152
Entraînement	95 %	84 %	87 %	90 %	76 %	70 %
Validation	95 %	93 %	95 %	93 %	91 %	83 %
Test	90 %	83 %	84 %	84 %	81 %	73 %

Tableau 4.5 : Précision obtenue par tous les modèles

Nous distinguons aisément que VGG16 et DenseNet121 avec leurs précisions respectives de 95% et 90%, sont en tête, suivis de près par DenseNet201 et VGG19. ResNet50 et ResNet152 sont en bas de file avec leurs bas taux, lors de l’entraînement.

4.5 Comparaison avec les travaux antérieurs

Le tableau 4.5 relate les principaux résultats obtenus en imagerie échographique, dans les travaux antérieurs de la littérature scientifique et de l’équipe IMAD, des laboratoires LATSI et LRDSI de l’université de Blida 1.

Les travaux de Kalaf et al [75], utilisent deux bases de données échographiques pour la classification de tumeurs mammaires, mais avec les mêmes modèles VGG16 et Attention-VGG16. Les résultats obtenus avec la base BUSI (la même base utilisée dans cette étude), sont satisfaisants et concluants puisque les taux de précision, ont atteint les 93% pour la méthode Attention-VGG16 et un taux de 89% pour VGG16.

Al-Dhabyani et al [41], ont fait la détection de masses malignes, bénignes et normales échographiques, suivant plusieurs CNN pré-entraînés. Les résultats obtenus par VGG16, Resnet et Nasnet, ont atteint respectivement une précision de 88%, 93% et 94%.

Dans les travaux du laboratoire LATSI, au niveau du projet de Salem [45], Nous avons étudié les réseaux de neurones profonds pré-entraînés en l’occurrence VGG19, VGG16 et Inception V3 ; qui exploitent l’enchaînement classique des couches de convolutions, de pooling et les blocs Inception. Les résultats démontrent que l’inception V3, surpasse le VGG16 et le vgg19, avec une précision de 94,64%. Dans le même cadre, au niveau du mémoire de Cherifi [44], quatre architectures CNN différentes : RESNET, VGG16, VGG19, XCEPTION ont été appliquées sur deux bases de données échographiques, pour la catégorisation des classes malignes, bénignes et normales.

Le système réalisé produit une précision de 100%, pour une aide à l'interprétation.

Dans le mémoire de Seman et Raki [46], trois modèles ont été étudiés suivant d'autres stratégies, où VGG16 a dépassé les modèles Resnet50 et InceptionV3, avec une précision de 94%.

En comparant seulement la précision obtenue avec chacun des travaux présentés, notre travail génère des résultats satisfaisants, avec un taux maximal de 95% obtenu avec le modèle VGG16.

Référence	Dataset	Méthode	Précision
Kalaf et al [75]	BUSI	VGG16	89%
		Attention VGG16	93%
Al-Dhabyani et al [41]	BUSI	VGG16	88%
		Resnet	93%
		Nasnet	94%
Salem [45]	BUSI	VGG16	94%
		VGG19	93%
		Inception	95%
Cherifi [44]	BUSI	Resnet50	95%
		VGG16	93%
		VGG19	93%
		Xception	100%
Seman [46]	BUSI	VGG16	94%
		InceptionV3	91%
		Resnet50	90%

Tableau 4.6 : Etude comparative avec les travaux antérieurs

4.6 Conclusion

Le système BreastUltraCNN4 a montré des résultats encourageants, pour la classification des différentes pathologies mammaires à travers des images échographiques, basée sur six architectures des réseaux de neurones convolutifs pré-entraînés, le VGG16, le VGG19, le Resnet50, le Resnet152, le Densenet121 et le Densenet201. Après plusieurs variations des paramètres de base, tels que, le taux d'apprentissage, ou le taux du dropout, l'étude comparative a montré que les modèles les plus appropriés, sont le VGG16 et le Densenet121.

Conclusion générale

L'initiation à la recherche réalisée dans le cadre de ce mémoire, s'inscrit dans les objectifs de l'équipe, 'IMAD', du laboratoire LATSI du département d'électronique et du laboratoire LRDSI du département d'informatique de l'institution universitaire de Blida 1, pour la conception de systèmes d'aide au diagnostic en imagerie médicale.

Les systèmes d'aide au diagnostic sont très utiles et sont en mesures d'apporter des solutions efficaces, afin de faciliter et fiabiliser le diagnostic dans le domaine médical, notamment dans le cas des tumeurs liées au cancer du sein. Les objectifs fixés par ce mémoire, se rapportent à l'élaboration d'un système d'aide à la décision, pour la détection de tumeurs mammaires à partir d'images échographiques, suivant les réseaux de neurones convolutifs.

Le système 'BreastUltraCNNV4' a été réalisé suivant six modèles pré-entraînés, donnés par le VGG16, le VGG19, le Resnet50, le Resnet152, le Densenet121 et le Densenet201, en utilisant un fine tuning et des paramètres appropriés pour classifier les images échographiques, de la base de données égyptienne publique, en classe maligne, bénigne et normale.

Les six modèles de réseaux convolutifs utilisés et mis à l'épreuve, diffèrent entre eux par leurs architectures (nombre de couches de convolution, de pooling, ...) et leur capacité d'apprentissage. En effet, les résultats obtenus montrent, que lors de l'apprentissage de la plus faible précision (70%) est atteinte par ResNet152 et la plus grande (95%) pour le VGG16. Nous pouvons conclure que le VGG16 avec sa précision de 95%, est le modèle le plus approprié à cette étude, pour une aide à la décision.

Le système 'BreastUltraCNNV4' est limité aux clichés mammaires à ultrasons. Il serait intéressant de détecter les tumeurs mammaires, dans le futur à partir d'autres types d'images, telles que : les données de l'IRM ou les mammographies et explorer d'autres modèles du deep learning.

Bibliographie



[1] Global health organization. Global Cancer Observatory, <https://gco.iarc.fr>, date de consultation, avril 2022.

[2] Société canadienne du cancer, Les seins, <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/breast/breast-cancer/the-breasts/?region=on>, date de consultation, février 2022.

[3] E. Sebban. Anatomie du sein : Les différents types de tumeurs bénignes et malignes, mars 2019, <https://www.docteur-eric-sebban.fr/cancer-du-sein/diagnostic-cancer-sein/anatomie-et-pathologies-du-sein/> date de consultation, février 2022.

[4] J. Stinès, Imagerie du sein, Elsevier Masson, Octobre 2012, ISBN 978-2-294-72960-7 https://books.google.dz/books?id=RVQD_vxFR8IC&printsec=frontcover&dq=IMAGERIE+DU+SEIN&hl=fr&sa=X&ved=2ahUKEwiindjJwsHwAhWDTThUIHVVC5oQ6AEwAHoECAIQAg#v=onepage&q=IMAGERIE%20DU%20SEIN&f=false date de consultation, février 2022.

[5] M. Morrow S. Wong L. Venta. The evaluation of breast masses in women younger than forty years of age. *Surgery*; **124** : 634-41, 1998.

[6] R. Givaudan, J. Serres, A. Dupuy, E. Thuane. Le cancer et les cellules cancéreuses, <http://lecancerdusein-tpe.blogspot.com>, date de consultation, février 2022.

[7] J. DÜRRLEMAN, Valeur diagnostique de la mammographie synthétique sans et avec tomosynthèse, Thèse de doctorat en médecine, Université Paris Descartes, 2015.

[8] Blog IRM Québec, IRM du sein, <https://irmquebec.com/tag/irm-du-sein>, date de consultation, février 2022.

[9] M. Dautat. Les bases de l'échographie, <https://ultrasonographie-vasculaire.edu.umontpellier.fr/files/2013/09/MD-Bases-Echographie-Abr-2013A.pdf>, CHU de Nîmes, 2013, date de consultation, février 2022.

[10] Palmer ML, Tsangaris TN. Breast biopsy in women 30 years old or less. *Am J Surg* ; **165**:708-12, 1993.

- [11] Sègbédji Goubalan. Contributions à l'analyse d'images médicales pour la reconnaissance du cancer du sein, thèse de doctorat en informatique, Université de Paris-Saclay, Décembre 2016.
- [12] I. Cheikhrouhou. Description et classification des masses mammaires pour le diagnostic du cancer du sein, thèse de doctorat en informatique, Université d'Evry Val d'Essonne, juin 2012.
- [13] L. Levy, J. Michelin, G. Teman, B. Martin, A. Dana, A. Lacan, D. Meyer, Techniques d'exploration radiologique du sein (mammographie, échographie, imagerie par résonance magnétique), Editions Scientifiques et Médicales Elsevier, Radiodiagnostic - Principes et techniques d'imagerie, 34-800-A-10, 2001, p20 https://fr.slideshare.net/imma-dr/techniques-dexploration-radiologique-du-sein-mammographie?from_action=save, date de consultation, février 2022.
- [14] D. de Korvin, Sémiologie Echographique Lexique Birads Classification ACR Masse et Microcalcifications, Institut Rennais du Sein, CLCC Rennes, 2018. https://moodlesante.univ-brest.fr/moodle/pluginfile.php/53180/mod_folder/content/0/3%20echo%20BdeKorvin%20.pdf?forcedownload=1 date de consultation, janvier 2022.
- [15] A. Zaki. Apport du couple écho-mammographie dans le diagnostic des cancers du sein A propos de 100 cas, thèse de doctorat en médecine, Université Cadi Ayyad de Marrakech, Maroc, 2015.
- [16] I. Leconte. Quelle est la place de l'échographie dans le dépistage du cancer du sein ? Imagerie de la Femme Vol. 16, N°2, pages 95-100, 2006.
- [17] P. David. Valeur primordiale de l'échographie en aval de la mammographie de dépistage du cancer du sein, Imagerie de la Femme, Vol. 14, Supp. 1, Pages 2-18, 2004.
- [18] Institut du sein, H. Harmattan. L'intelligence artificielle va-t-elle aider les radiologues à mieux dépister les cancers du sein ? <https://ishh.fr/cancer-du-sein/intelligence-artificielle-aide-au-depistage-du-cancer-du-sein/>, date de consultation, février 2022.
- [19] Azure. Présentation du machine learning, <https://azure.microsoft.com/fr-fr/overview/what-is-machine-learning-platform/>, date de consultation, février 2022.
- [20] Talend. Tout savoir sur le machine learning, <https://www.talend.com/fr/resources/what-is-machine-learning>, date de consultation, février 2022.
- [21] D. Belhaouci. Démystifier le machine learning, <https://www.juripredis.com/fr/blog/id-19-demystifier-le-machine-learning-partie-2-les-reseaux-de-neurones-artificiels>, date de consultation, février 2022.

[22] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, Volume 79 page 2554-2558, <https://www.pnas.org/doi/epdf/10.1073/pnas.79.8.2554>, 1982, date de consultation février 2022.

[23] I. Chraïbi. Kaadoud & A. Garenne. Architecture des réseaux de neurones : Réseaux de neurones artificiels classiques, <https://scilogs.fr/intelligence-mecanique/architecture-des-reseaux-de-neurones-reseaux-de-neurones-artificiels-classiques-2-3/>, 2018, date de consultation février 2022.

[24] J. Feng, S. Lu. Performance Analysis of Various Activation Functions in Artificial Neural Networks. Journal of Physics: Conference Series, Volume 1237, ICSP 2019.

[25] L. Bastien, Deep Learning ou apprentissage profond : définition, concept, blog LeBigData.fr, 2018, <https://www.lebigdata.fr/deep-learning-definition> , date de consultation mars 2022.

[26] Natural solutions, Histoire du deep learning, blog, <https://www.natural-solutions.eu/blog/histoire-du-deep-learning/>, date de consultation mars 2022.

[27] Digital guide Ionos. Quelles sont les différences entre le Deep learning et le Machine learning ? <https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/deep-learning-vs-machine-learning/>, date de consultation, mars 2022.

[28] Moncoachdata, Comprendre les réseaux de neurones, <https://moncoachdata.com/blog/comprendre-les-reseaux-de-neurones/>, 2019, date de consultation, mars 2022.

[29] Data value. Deep learning, les réseaux de neurones récurrents, <https://datavalue-consulting.com/deep-learning-reseaux-neurones-recurrents-rnn/>, date de consultation, mars 2022.

[30] Actu IA. A travers les réseaux de neurones à convolution en deep learning, <https://www.actuia.com/contribution/cyrille-kone/a-travers-les-reseaux-de-neurones-a-convolution-en-deep-learning/>, date de consultation mars 2022.

[31] M. Chane, Deep learning : les réseaux de neurones convolutifs pour la classification d'images médicales, Page Web IMAIOS, <https://www.imaios.com/fr/Societe/blog/Classification-des-images-medicales-comprendre-le-reseau-de-neurones-convolutifs-CNN>, 2021, date de consultation mars 2022.

[32] K. Vancappel. Tutoriel, deep learning : le réseau neuronal convolutif (CNN), <https://fr.blog.businessdecision.com/tutoriel-deep-learning-le-reseau-neuronal-convolutif-cnn/> date de consultation, mars 2022.

- [33] MI Tensor. Flattening and Full Connection Layers (Neural Networks), <https://www.mltensor.com/post/flattening-and-full-connection-layers-neural-networks>, date de consultation, mars 2022.
- [34] Moualek Djaloul Youcef. Deep Learning pour la classification des images, mémoire de master en informatique, université de Tlemcen, 2017.
- [35] Y. Bengio, P. Simard, P. Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* Volume 5, pages 157–166, 1994.
- [36] Y. Le Cunn. Fonctions d'activation et de perte, <https://atcold.github.io/pytorch-Deep-Learning/fr/week11/11-1/>, date de consultation, mars 2022.
- [37] Thomas B, Lina F, Gary B, Transfer Learning : Qu'est-ce que c'est ? Blog Formation Data Science | DataScientest.com, 2020, <https://datascientest.com/transfer-learning> date de consultation, mars 2022.
- [38] Data Science Test. Transfer Learning : Qu'est-ce que c'est ? <https://datascientest.com/transfer-learning>, date de consultation, mars 2022.
- [39] Geekflare. Qu'est-ce qu'une matrice de confusion dans l'apprentissage automatique ? <https://geekflare.com/fr/confusion-matrix-in-machine-learning/>, date de consultation, mars 2022.
- [40] Mieux comprendre le deep learning appliqué à la reconnaissance d'images. (2022, 27 septembre). *Devoteam France*. Consulté en mars 2022, à l'adresse <https://france.devoteam.com/paroles-dexperts/mieux-comprendre-le-deep-learning-applique-a-la-reconnaissance-dimages/>
- [41] W. Al-Dhabyani, A. Fahmy, M. Gomaa, H. Khaled. Deep Learning Approaches for Data Augmentation and Classification of Breast Masses using Ultrasound Images, *International Journal of Advanced Computer Science and Applications*, Volume 10, n°5, p618-627, 2019.
- [42] Moi Hoon Yap, Gerard Pons, Joan Marti, Sergi Ganau, Melcior Sentis, Reyer Zwiggelaar, Adrian K. Davison, Robert Marti, Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks, *IEEE Journal of Biomedical and Health Informatics*, Vol 22, N°04, pages 1218-1226, 2018.
- [43] C. Y. Lee, G. L. Chen, Z. X. Zhang, Y. H. Chou, C. C. Hsu, Is Intensity Inhomogeneity Correction Useful for Classification of Breast Cancer in Sonograms Using Deep Neural Network? *Journal of Healthcare Engineering*, Volume 2018, page 2040-2309, 2018.

- [44] I. Cherifi. Apport de l'intelligence artificielle dans la détection du cancer du sein, mémoire de master en électronique, université de Blida 1, 2020.
- [45] N. Salem. Analyse profonde d'images ultrasonores pour la détection du cancer du sein, mémoire de master en électronique, université de Blida 1, 2020.
- [46] 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.
- [47] K. Manseur, S. Mebarki. Détection de pathologies mammaires pour l'aide à l'interprétation, mémoire de master en informatique, université de Blida 1, 2020.
- [48] I. BenMh'amed, H. Nekkaz. Analyse de carcinomes mammaires par les techniques du deep learning pour l'aide à l'interprétation, mémoire de master en automatique, université de Blida 1, juin 2022.
- [49] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy. Dataset of breast ultrasound images, Data in brief, volume 28, 2020.
- [50] Esri. Entraîner un modèle de Deep Learning, <https://pro.arcgis.com/fr/pro-app/latest/help/data/las-dataset/train-a-point-cloud-model-with-deep-learning.htm>, date de consultation, mars 2022.
- [51] Neovision. Data augmentation : des solutions face au manque de données, <https://neovision.fr/data-augmentation-solutions-manque-donnees/>, date de consultation, mars 2022.
- [52] Z. El Hamraoui. Fine-tuning in Deep Learning, <https://ai.plainenglish.io/fine-tuning-in-deep-learning-909666d4c151>, date de consultation, mars 2022.
- [53] Visio.ai. Deep Neural Network, <https://viso.ai/deep-learning/deep-neural-network-three-popular-types/>, date de consultation, mars 2022.
- [54] Visio.ai. VGG Very Deep Convolutional Networks (VGGNet), <https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/>, date de consultation, mars 2022.
- [55] K. Simonyan, A. Zisserman. Very Deep Convolutional Networks for Large-Scale Visual Recognition, https://www.robots.ox.ac.uk/~vgg/research/very_deep/, date de consultation, mars 2022.
- [56] Imagenet. ImageNet Large Scale Visual Recognition Challenge 2014 (ILSVRC2014), <https://image-net.org/challenges/LSVRC/2014/>, date de consultation, mars 2022.

- [57] Y. Zhenga, C. Yangb, A. Merkul. Breast Cancer Screening Using Convolutional Neural Network and Follow-up Digital Mammography, Conference: Computational Imaging III, DOI: 10.1117/12.2304564, 2018.
- [58] K. Patel. Architecture comparison of AlexNet, VGGNet, ResNet, Inception, DenseNet, <https://towardsdatascience.com/architecture-comparison-of-alexnet-vggnet-resnet-inception-densenet-beb8b116866d>, date de consultation, mars 2022.
- [59] M. Chane. Classification en imagerie médicale : les architectures CNN les plus performantes en 2021, <https://www.imaios.com/fr/Societe/blog/Classification-en-imagerie-medecale-les-architectures-CNN-les-plus-performantes-en-2021>, date de consultation, mars 2022.
- [60] J. Sparrow. Réseaux résiduels (ResNet) – Deep Learning, <https://fr.acervolima.com/reseaux-residuels-resnet-deep-learning/>, date de consultation, mars 2022.
- [61] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [62] Imagenet. ImageNet and MS COCO Visual Recognition Challenges Joint Workshop, <https://image-net.org/challenges/ilsvrc+mscoco2015>, date de consultation, mars 2022.
- [63] Md. M. Islam, N. Tasnim, J. H. Baek. Human Gender Classification Using Transfer Learning via Pareto Frontier CNN Networks, Inventions 2020, Volume 5, N°2, doi:10.3390/inventions5020016, 2020.
- [64] P. Ruiz. Understanding and visualizing DenseNets, <https://towardsdatascience.com/understanding-and-visualizing-densenets-7f688092391a>, date de consultation, mars 2022.
- [65] DenseNet Architecture & Components, <https://iq.opengenus.org/architecture-of-densenet121/>, date de consultation, mars 2022.
- [66] S. S. Rao. Exploring DenseNets and a comparison with other Deep Architectures, <https://medium.com/analytics-vidhya/exploring-densenets-and-a-comparison-with-other-deep-architectures-85f02597400a>, date de consultation, mars 2022.
- [67] Dive into deep learning. Fine tuning, https://d2l.ai/chapter_computer-vision/fine-tuning.html, date de consultation, mars 2022.
- [68] Jedha. Apprendre python ou R, <https://www.jedha.co/blog/apprendre-python-ou-r>, date de consultation, avril 2022.

[69] Futura Tech, Python : qu'est-ce que c'est ? Page Web, <https://www.futura-sciences.com/tech/definitions/informatique-python-19349/> date de consultation, avril 2022.

[70] Colaboratory. Bienvenue dans colab, <https://colab.research.google.com/?hl=fr>, date de consultation, avril 2022.

[71] Numpy. Numpy documentation, <https://numpy.org/doc/stable/>, date de consultation, avril 2022.

[72] Matplotlib. Matplotlib documentation, <https://matplotlib.org/stable/index.html>, date de consultation, avril 2022.

[73] L. Bastien. TensorFlow : tout savoir sur la bibliothèque Machine Learning open source, Blog LeBigData.fr,2018, <https://www.lebigdata.fr/tensorflow-definition-tout-savoir> date de consultation, avril 2022.

[74] Team Keras. Keras documentation, <https://keras.io/about/> date de consultation, avril 2022.

[75] E. Y. Kalaf, A. Jodeiri, S. Kamaledin Setarehdan. Classification of Breast Cancer Lesions in Ultrasound Images by using Attention Layer and loss Ensembles in Deep Convolutional Neural Networks, Diagnostics (Basel), pages 1-14, doi: 10.3390/diagnostics111101859, 2021.