

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

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

جامعة سعد دحلب البليدة
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique



Mémoire de Projet de Fin d'Études

Présenté par

Cherifi Ilham

pour l'obtention du diplôme de Master en Electronique

Option

Option 'Automatique et informatique industrielle'

Thème

**Apport de l'intelligence artificielle dans la détection
connexionniste du cancer du sein**

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

Année Universitaire 2019-20220

Remerciements

Au préalable, je remercie le bon dieu de m'avoir donné le courage, la force et la patience de poursuivre mes études malgré toutes les difficultés

Je tiens à exprimer toute ma reconnaissance à ma directrice de mémoire, Madame **Reguieg F. Zohra** Je la remercie de m'avoir encadrée, orientée, aidée, me conseiller, m'encourager et d'être toujours là malgré les circonstances que nous traversons durant cette pandémie par ses e-mails, appels et ses Visioconférence.

Je tiens aussi à remercier ma co-promotrice Madame **Benblidia Nadja**, pour ses encouragements, son soutien et ses conseils.

J'adresse mes sincères remerciements à tous les professeurs, intervenants et toutes les personnes qui par leurs paroles, leurs écrits, leurs conseils et leurs critiques ont guidé mes réflexions et ont accepté de me rencontrer et de répondre, à mes questions durant mes études.

Je dédie ce modeste travail en signe de respect,

reconnaissance et de remerciement :

A mes très chers parents qui ont toujours été là pour moi, pour leurs amours, leurs sacrifices, leurs encouragements, leurs conseils leur éducation à eux je suis là où je suis. Je ne pourrai jamais les remercier assez.

A ma sœurs Yousra et mes frères Oussama et Anes

A mes Grands-parents, mes oncles et mes tantes

A mes cousins et cousines notamment Abdrrahim et Haithem qui m'ont tellement aidé.

A mes meilleures amies, mes sœurs : Maroua, Rayen, Soumia, Foufou, Melissa et Hafsa merci d'être toujours là à mes côtés.

A mon amie Ilhem pour ses précieuses conseils.

A ceux avec qui j'ai passé les meilleurs moment à l'université : Soumia , Anya et lilia.

A tous mes amis et camarade de la promotion automatique et informatique industrielle.

ملخص:

يستخدم الذكاء الاصطناعي لتصوير الثدي لجعل تفسير صور الموجات فوق الصوتية أقل تعقيدًا لهذا الهدف من هذه الدراسة هو جلب الذكاء الاصطناعي في الكشف عن أورام الثدي هذا التنبؤ يتم من خلال تهجين الشبكات العصبية التلافيفية ونقل التعلم بشكل ملموس أكثر، نقدم أربعة بنى مختلفة RESNET ، VGG16 ، VGG19 ، XCEPTION، قمنا بمقارنة العديد من فترات التدريب التي تتراوح من 1 إلى 50 حقة ، من قاعدتي بيانات ومحسنتين مختلفتين ADAM و ADADELTA. أفضل النتائج التي تم الحصول عليها هي 100% و 93% دقة.

كلمات المفاتيح: الذكاء الاصطناعي، صور الموجات فوق الصوتية، الشبكات العصبية الالتفافية، هندسيات سي أن أن، المساعدة على الترجمة .

Résumé : L'intelligence artificielle est mise au service de l'imagerie mammaire, afin de rendre l'interprétation des clichés échographiques moins fastidieuse et, plus précise. A cet effet, l'objectif de cette étude est la prédiction du cancer du sein, par les techniques de l'intelligence artificielle. Cette prédiction est faite grâce à l'hybridation réseaux de neurones convolutifs suivant, le transfert d'apprentissage. Plus concrètement nous proposons l'étude de quatre architectures CNN différentes : RESNET, VGG16, VGG19, XCEPTION appliquées sur deux bases de données échographiques, pour la catégorisation des classes malignes, bénignes et normales. Le système réalisé engendre une exactitude de 100% et une exactitude de 93%, pour une aide à l'interprétation.

Mots clés : Intelligence artificielle, Imagerie mammaire, Clichés échographiques, Réseaux de neurones convolutionnels, Architectures CNN, Aide à l'interprétation.

Abstract : The Artificial intelligence is adopted for breast imaging, to make the interpretation of ultrasound images, easier and accurate. For this purpose, the main objective of this research is the prediction of breast cancer, by using the artificial intelligence techniques. This prediction is made by the hybridization of Convolutional Neuron Networks and the transfer learning. More concretely, we propose the study of four different CNN architectures: RESNET, VGG16, VGG19, XCEPTION, applied on two breast ultrasound databases, to categorize the benign, the malign and the normal classes. The produced system generates an accuracy of 100% and an accuracy of 93%, for interpretation assistance.

Keywords : Artificial intelligence, Breast imaging, Breast ultrasound images, Convolutional Neuron, CNN architectures, Interpretation assistance.

Listes des acronymes et abréviations

AA : apprentissage automatique

ADAM : adaptive learning rate optimization

CNN : Convolutional neural network

DL : Deep learning

IA : Intelligence artificielle

FP : False positive

FN : False negative

KNN : K-Nearest Neighbor

ML : Machine learning

OMS : Organisation mondiale de la Santé

RNA : Réseaux de neurones artificiels

RNN : Réseaux de neurones récurrents

SVM : Support vector machines

TN : True négative

TP : True positive

Table des matières

Introduction générale	1
Chapitre 1 Contexte médical	3
1.1 Introduction.....	3
1.2 Structure du sein	3
1.3 Anatomie descriptive mammaire.....	5
1.4 Système lymphatique du sein	6
1.5 Lésions bénignes du sein	7
1.5.1 Adénofibrome	7
1.5.2 Tumeur phyllode bénigne	7
1.5.3 Dystrophie fibrokystique.....	8
1.5.4 Abscess	8
1.5.5 kystes.....	8
1.6 Introduction au Cancer du sein	8
1.6.1 Définition d'un cancer.....	8
1.6.2 Symptômes du cancer du sein	9
1.7 Classification et stadification des cancers du sein	10
1.7.1 Classification	10
1.7.2 Différents stades du cancer du sein.....	11
1.7.3 Différents types tumeurs malignes du sein	11
1.8 Moyens de dépistages et du diagnostic	13
1.8.1 Dépistage.....	13
1.8.2 Age et facteurs de risque	13
1.8.3 Diagnostic.....	15
1.9 Imagerie mammaire	16
1.9.1 Mammographie.....	16
1.9.2 Echographie.....	17
1.9.3 Imagerie par résonance magnétique mammaire.....	19
1.10 Etude échographique.....	19
1.10.1 Aspects pathologiques	20
1.10.2 Aspect échographique du ganglion axillaire envahi.....	21
1.11 Classification BIRADS des pathologies mammaires.....	22
1.12 Traitement médical.....	23

1.13	Application de l'intelligence artificielle à la prédiction et au diagnostic en oncologie.....	23
1.14	Conclusion.....	24
Chapitre 2 Notions fondamentales d'intelligence artificielle.....		25
2.1	Introduction.....	25
2.2	BIG DATA, data science et intelligence Artificielle.....	25
2.2.1	BIG data.....	25
2.2.2	La data science.....	26
2.2.3	Introduction à l'intelligence Artificielle.....	26
2.3	De la naissance de l'intelligence artificielle jusqu' à 2020.....	27
2.4	Apprentissage automatique.....	29
2.4.1	Données d'apprentissage automatique.....	30
2.4.2	Différents types d'apprentissage automatique.....	30
2.5	Réseaux de neurones artificiels.....	32
2.5.1	Histoire des réseaux de neurones.....	33
2.5.2	Principe.....	34
2.5.3	Modélisation.....	34
2.6	Choix de la fonction d'activation et la fonction de perte.....	40
2.7	Les propriétés des réseaux de neurones.....	41
2.8	Différents types de réseaux de neurones.....	41
2.8.1	Les réseaux de neurones mono couches.....	42
2.8.2	Réseau de neurones multicouches.....	42
2.9	Différentes architectures des réseaux de neurones.....	44
2.9.1	Réseaux Feed-Forward.....	44
2.9.2	Réseaux par rétropropagation (back forward).....	44
2.10	Différence entre époque, taille de lot et une itération.....	44
2.11	Apprentissage profond.....	45
2.11.1	Extraction des caractéristiques.....	46
2.12	Différence entre le deep Learning et le machine Learning.....	47
2.13	Réseaux de neurones récurrents RNN.....	48
2.14	Auto-encodeur.....	48
2.15	Outils de mesures des performances.....	48
2.15.1	Matrice de confusion.....	48
2.15.2	Critères d'évaluations.....	49

2.16	Conclusion.....	51
Chapitre 3	Fonctionnalités des CNN dans la classification d'images.....	52
3.1	Introduction.....	52
3.2	Réseaux de neurones convolutif (CNN).....	52
3.2.1	Couches d'un réseau de neurones convolutif.....	53
3.2.2	Opération de convolution.....	53
3.2.3	Couche de convolution (conv).....	54
3.2.4	Couche de pooling.....	55
3.2.5	Couche d'activation.....	55
3.2.6	Couches Fully Connected (Fc).....	55
3.2.7	Couche flatten.....	56
3.2.8	Couche Dropout.....	56
3.3	Diverses architectures des réseaux de neurones convolutif.....	56
3.3.1	Imagenet et la compétition « large scale visual recognition challenge ».....	56
3.3.2	LeNet.....	56
3.3.3	Alex-NET.....	57
3.3.4	Modèles VGG.....	59
3.3.5	RESNET.....	61
3.3.6	Xception.....	62
3.4	Architecture globale de Xception.....	64
3.5	Comparaison entre les différentes architectures.....	65
3.6	Augmentation des données.....	67
3.7	Détection d'objets par les réseaux de neurones convolutifs.....	67
3.8	Apprentissage profond par transfert.....	68
3.9	Optimiseurs.....	69
3.10	Travaux en imagerie échographique.....	70
3.11	Conclusion.....	71
Chapitre 4	Mise en œuvre de l'application et discussions des résultats.....	72
4.1	Introduction.....	72
4.2	Environnement de travail.....	72
4.2.1	Hardware.....	72
4.2.2	Software (Logiciel).....	72
4.3	Présentation du système développé.....	75
4.4	Bases de données.....	75

4.4.1	Dataset BUSI.....	76
4.4.2	US dataset	78
4.5	Augmentation des données	79
4.5.1	Transformations appliquées sur la base BUSI.....	80
4.5.2	Augmentation des données de la base US.....	80
4.6	Préparation des données pour l'apprentissage	81
4.7	Architectures des modèles de 'BreastCNNPredict'	82
4.7.1	Modèle VGG16 appliqué à la multiclassification	83
4.7.2	Architecture VGG16 pour la classification binaire	87
4.7.3	VGG19	88
4.7.4	Modèle ALEXNET adopté dans BreastCNNPredict.....	90
4.7.5	Modèle Xception	93
4.8	Optimiseurs utilisés	94
4.9	Fonction de perte adoptée.....	95
4.10	Résultats de 'BreastCNNPredict'	95
4.10.1	Résultats obtenus par le modèle XCEPTION	95
4.10.2	Résultats obtenus par ALEXNET	98
4.10.3	Résultats obtenus par le VGG16	100
4.10.4	Résultats obtenus par le modèle VGG19	102
4.10.5	Modèle type obtenu par VGG16 pour BUSI dataset.....	105
4.10.6	Modèle type obtenu par VGG 16 pour la base US.....	108
4.10.7	Modèle type obtenu par VGG19.....	111
4.10.8	Modèle type obtenu par l'optimiseur ADAM	114
	Le rapport de classification	117
4.11	Présentation de l'interface 'BreastCNNPredict'	117
4.12	Comparaison entre les différents modèles	120
4.13	Conclusion.....	122
	Conclusion générale.....	123
	Bibliographie	125

Liste des figures

Figure 1.1 : Structure du sein.....	5
Figure 1.2 : Quatre quadrants du sein	6
Figure 1.3 : Ganglions du sein	7
Figure 1.4 : Principaux types de cancers chez les femmes et les hommes en Algérie en 2018	9
Figure 1.5 Symptômes du cancer du sein	10
Figure 1.6 : Cancers canaux in situ et infiltrant	12
Figure 1.7 : Mammographe analogique	17
Figure 1.8 : Chronologie de l'échographie.....	18
Figure 1.9 : Echographe	19
Figure 1.10 : kyste	20
Figure 1.11 : Tumeur du sein d'aspect malin (lacune hypoéchogène, hétérogène, forme irrégulière, sans cône d'ombre postérieure, non parallèle au plan cutané).....	21
Figure 1.12 : Tumeur du sein d'aspect malin (contours irréguliers : indentations, spicules)	21
Tableau 1.2 Classification BI_RADS (breast imaging reporting and data system).....	23
Figure 2.1 : Cycle de travail d'un datascientist	26
Figure 2.2 : IA, ML, DL	27
Figure 2.3 Différents types d'apprentissage automatique.....	31
Figure 2.4 Neurone biologique	33
Figure 2.5 Propagation des données	35
Figure 2.6 Fonction échelon	36
Figure 2.7 Fonction Sigmoïde	37
Figure 2.8 Fonction tanh.....	38
Figure 2.9 Fonction ReLu	38
Tableau 2.1 Choix de la fonction d'activation et la fonction perte	41
Figure 2.10 : perceptron simple	42
Figure 2.11: Réseau de neurones multicouches	43
Figure 2.12 Extraction des caractéristiques par apprentissage automatique et profond	47
Tableau 2.2 Différence entre le deep Learning et le machine Learning	47
Tableau 2.3 Matrice de confusion	49
Figure 3.1 : Exemple d'une convolution	54
Figure 3.2 Architecture LeNet	57
Tableau 3.1 Architecture LeNet	57
Figure 3.3 Architecture LeNet	58
Tableau 3.2 Différentes configurations de VGG	60
Figure 3.4 Architecture de VGG16 et VGG19	61
Figure 3.5 Architecture Resnet	61

Figure 3.6 Modèle résiduel	62
Figure 3.7 Convolution originale séparable en profondeur	63
Figure 3.8 Convolution modifiée séparable en profondeur	64
Figure 3.9 Architecture de Xception	65
Tableau 3.3 Paramètres des différentes architectures	66
Tableau 3.4 top 5 d'erreurs	67
Figure 3.10 Architecture YOLO	68
Figure 4.1 : Synoptique de BreastCNNPredict	75
Figure 4.2 : Images de la base BUSI	76
Figure 4.3 Répartition des classes pour BUSI	76
Figure 4.4 Répartition des pixels d'une image bénigne de BUSI	77
Figure 4.5 Répartition des pixels d'une image maligne de BUSI	77
Figure 4.6 Répartition des pixels d'une image maligne de BUSI	78
Figure 4.7 Clichés échographiques de la base US.....	78
Figure 4.8 Répartition des classes pour US.....	78
Figure 4.9 Répartition des pixels en fonction des niveaux de gris d'un cliché bénin.....	79
Figure 4.10 Répartition des pixels en fonction des niveaux de gris d'un cliché malin...	79
Figure 4.11 Exemple d'augmentation des données de la base BUSI	80
Figure 4.12 Exemple d'augmentation des données de la base BUSI	81
Figure 4.13 Répartition des données.....	81
Figure 4.14 Nombre de clichés dans chaque classe dans BUSI	82
Figure 4.15 Nombre de clichés dans chaque classe de la base US.....	82
Figure 4.16 Modèle de BreastCNNPredict.....	83
Figure 4.17 VGG16 pour la base de données BUSI.....	84
Figure 4.18 : Exemple d'une convolution	85
Figure 4.19 Exemple d'un maxpooling 2*2	85
Figure 4.20 : Exemple de l'opération flatten	86
Figure 4.21 : Exemple de Dropout	86
Figure 4.22 : Dernière couche fully connected.....	86
Figure 4.23 vgg 16 pour US.....	87
Figure 4.24 : Dernière couche fully connected pour US.....	88
Figure 4.25 : VGG19 pour BUSI	89
Figure 4.26 : VGG19 pour US	90
Figure 4.27 Alex net pour la multiclassification.....	91
Figure 4.28 Alexnet pour la classification binaire.....	92
Figure 4.29 : Architecture Xception pour la multiclassification	93
Figure 4.30 : Architecture Xception pour la classification binaire.....	94
Figure 4.31 Fonction loss	95
Tableau 4.1 : Résultats obtenus par le modèle Xception pour Busi dataset.....	96
Figure 4.32 : Exactitude du modèle xception pour la base BUSI.....	96
Tableau 4.2 : Résultats obtenus par le modèle Xception pour la base US.....	97

Figure 4.33 : Exactitude pour le modèle xception pour la base US.....	97
Figure 4.34: Exactitude du modèle ALEXNET pour BUSI	99
Figure 4.35: Exactitude du modèle ALEXNET pour BUSI	100
Figure 4.36: Exactitude du modèle vgg16 pour Busi	101
Figure 4.37: Exactitude du modèle vgg16 pour la base US	102
Figure 4.38: Exactitude du modèle vgg19 pour la base BUSI	104
Figure 4.39: Exactitude du modèle vgg19 pour la base US	104
Figure 4.40 : Evaluation de l'exactitude des ensembles d'entraînement et de validation en fonction du nombre des époques (vgg16 et busi).....	105
Figure 4.41 : Evaluation de la fonction Loss des ensembles d'entraînement et de validation en fonction du nombre des époques.....	106
Figure 4.42 : Matrice de confusion et table de confusion (vgg16 et Busi).....	106
Figure 4.43: Prédiction d'une classe normale	107
Figure 4.44: Evaluation de l'exactitude des ensembles d'entraînement et de validation en fonction du nombre des époques (vgg16 et us).....	108
Figure 4.45 : Evaluation de la fonction Loss des ensembles d'entraînement et de validation en fonction du nombre des époques.....	109
Figure 4.45: la courbe ROC pour le vgg16	109
Figure 4.46 : matrice et table de confusion (vgg16 et us)	110
Figure 4.47 : Prédiction d'une classe maligne	111
Figure 4.48 Evaluation de l'exactitude des ensembles d'entraînement et de validation en fonction du nombre des époques (vgg19 et busi).....	112
Figure 4.49 : Evaluation de la fonction Loss des ensembles d'entraînement et de validation en fonction du nombre des époques.....	112
Figure 4.50 : table et matrice de confusion (busi et vgg19)	113
Figure 4.51 : Prédiction d'une classe maligne	114
Figure 4.52 Evaluation de l'exactitude des ensembles d'entraînement et de validation en fonction du nombre des époques.....	115
Figure 4.53: Evaluation de la fonction Loss des ensembles d'entraînement et de validation en fonction du nombre des époques.....	115
Figure 4.54 courbe ROC (vgg19)	116
Figure 4.55: table et matrice de confusion (vgg16 et us).....	116
Figure 4.56 : Prédiction d'une classe maligne	117
Figure 4.57 : Interface principale de " BreastCNNPredict"	118
Figure 4.58 Choix d'un cliché	118
Figure 4.59 : Aperçu de l'exécution	119
Figure 4.60 Classification d'une tumeur bénigne	119
Figure 4.61 : Classification d'une tumeur maligne	120
Figure 4.62 : Classification d'une tumeur normale.....	120
Figure 4.62 Exactitude des différents modèles de 'BreastCNNPredict'	121

Liste des tableaux

Tableau 1. 1 Classification BI_RADS (breast imaging reporting and data system).....	22
Tableau 2.1 Choix de la fonction d'activation et la fonction perte	40
Tableau 2.2 Différence entre le deep Learning et le machine Learning	47
Tableau 2.3 table de confusion.....	499
Tableau 3.1 Architecture LeNet.....	57
Tableau 3.2 Différentes configurations de VGG	60
Tableau 3.3 Paramètres des différentes architectures	65
Tableau 1 top 5 d'erreurs.....	66
Tableau 4.1 Résultats obtenus par le modèle Xception pour Busi dataset.....	95
Tableau 4.2 Résultats obtenus par le modèle Xception pour US dataset.....	96
Tableau 4.3 Résultats obtenus par le modèle ALEXNET pour Busi dataset.....	97
Tableau 4.4 Résultats obtenus par le modèle ALEXNET pour US dataset.....	98
Tableau 4.5 Résultats obtenus par le modèle VGG16 pour BUSI	99
Tableau 4.6 Résultats obtenus par le modèle VGG16 pour US	101
Tableau 4.7 Résultats obtenus par le modèle VGG19 pour BUSI	102
Tableau 4.8 Résultats obtenus par le modèle VGG19 pour US	103
Tableau 1 : Comparaison entre les différents modèles développés.....	121

Introduction générale

Le cancer du sein, est le cauchemar des femmes qui doit être abattu. En Algérie selon l'organisation mondiale de la santé, il se positionne dans le premier rang suivant son incidence, sa mortalité et sa prévalence [1]. Une détection précoce est nécessaire et, un diagnostic rapide est crucial pour sauver des vies.

L'échographie mammaire, est un examen d'investigation du second ordre non invasif à faible coût, qui assure un affichage en temps réel. Utilisé dans le cadre de dépistage et de diagnostic, cet examen souvent prescrit aux jeunes femmes et à celles, ayant les seins denses, n'entraîne aucun risque, grâce à ses rayonnements non ionisants. Cependant, l'interprétation des clichés échographiques mammaires, reste difficile et fastidieuse pour les médecins et les radiologues ; car elle nécessite d'être complétée par une panoplie d'examen en l'occurrence les examens anatomopathologiques. En effet, ces analyses biologiques, prennent énormément de temps dans l'interprétation, plus le temps s'écoule, plus le danger de mortalité augmente. De nos jours l'intelligence artificielle traite cette problématique, suivant l'apprentissage profond.

Notre travail porte sur l'apport de l'intelligence artificielle dans la prédiction des tumeurs du sein, en utilisant différentes architectures des réseaux de neurones convolutifs, par la technique du transfert d'apprentissage. Dans ce cadre, les réseaux de neurones convolutifs assurent une extraction automatique des informations pertinentes, des clichés échographiques au fur et à mesure des opérations de convolutions. Ce processus sera suivi par une classification, grâce à un réseau de neurones multi couches. La dernière étape de notre étude, consiste à comparer les différents résultats obtenus par quatre modèles pré-entraînés : VGG16, VGG19, ALEXNET, XCEPTION afin de représenter le modèle type dans l'interface destinée, à l'aide au diagnostic clinique.

Pour mener à bien cette initiation à la recherche, nous structurons notre mémoire selon quatre chapitres :

- Le premier chapitre parcourt l'anatomie mammaire et le cancer du sein, ainsi qu'une présentation des différents outils d'imagerie mammaire.
- Le deuxième chapitre est consacré à l'intelligence artificielle, l'apprentissage automatique et, l'apprentissage profond, en décrivant les réseaux de neurones.
- Dans le troisième chapitre, nous mettons en relief les réseaux de neurones convolutifs, ses différentes architectures et la technique d'apprentissage par transfert.
- Le quatrième chapitre est consacré à l'application développée et, aux différents résultats obtenus.

1.1 Introduction

Le cancer du sein, est l'ennemi effrayant de toute femme, notamment les femmes ménopausées. Plus d'1,67 million nouveaux cas sont enregistrés chaque année, dans le monde dont, 12.000 nouveaux cas sont confirmés en Algérie. Les chiffres alarmants, ne cessent d'augmenter. On le considère comme la première cause de mortalité chez le sexe féminin lorsqu'il est détecté à un stade avancé. L'OMS a déclaré environ 3 115 décès en Algérie, en 2017 [1]. Il vaut mieux alors prévenir que guérir. A cet effet, un examen de dépistage est recommandé aux femmes selon leurs facteurs de risque.

On distingue plusieurs moyens de dépistage, les plus pertinents sont les outils d'imagerie mammaires tels que la mammographie et l'échographie, qui ont connu une évolution remarquable ces dernières années. Aujourd'hui un système de décision doté d'une certaine intelligence artificielle, est capable d'interpréter plusieurs clichés dans un temps restreint. Mais est-ce qu'un diagnostic machine est aussi fiable qu'un diagnostic médecin ?

Dans ce chapitre nous présentons l'anatomie du sein, les différentes pathologies mammaires ainsi que, l'imagerie mammaire. A la fin du chapitre, nous révélerons si l'hybridation de l'intelligence artificielle et l'imagerie mammaire, est une réalité ou une fiction.

1.2 Structure du sein

Le sein est un organe globuleux (figure 1.1), qui sous une enveloppe cutanée, comprend du tissu glandulaire, des tissus fibreux et adipeux situés entre les lobes et lobules glandulaires, des vaisseaux sanguins et lymphatiques ainsi que des nerfs [2, 3].

1. Enveloppe cutanée

Elle comprend le mamelon et l'aréole.

- Le mamelon : situé au centre de l'aréole séparé de ce dernier, par un sillon, pigmenté cylindrique composé de fibres musculaires pouvant se contracter et durcir ; sa principale fonction est d'évacuer le lait maternel.
- L'aréole : est un disque cutané de 15 à 30 mm de diamètre plus ou moins pigmentée. Souvent brunâtre ou rosée, elle contient des glandes sébacées appelées glandes aréolaires.

2. Glande mammaire

C'est une masse de densité variable discoïde, aplatie d'avant en arrière. Elle est composée de 10 à 25 lobes, chacun d'eux contient 10 à 100 alvéoles. La principale fonction des lobules est la production de lait maternel pendant le post-partum, véhiculé par des canaux jusqu' au mamelon.

3. Tissu adipeux et conjonctif

Etroitement liée au tissu glandulaire, la quantité de tissu adipeux est en grande partie responsable du volume des seins, lequel n'a aucun effet sur la production et la qualité du lait. On distingue deux couches graisseuses. La couche antérieure pré glandulaire est cloisonnée par des travées conjonctives : les ligaments de Cooper qui relient la peau à la glande en formant les crêtes de Ducret. La couche postérieure est limitée par le fascia superficialis, elle est séparée de l'aponévrose du grand pectoral par du tissu conjonctif. L'ensemble peau-glande-graisse, se trouve sur le grand pectoral.

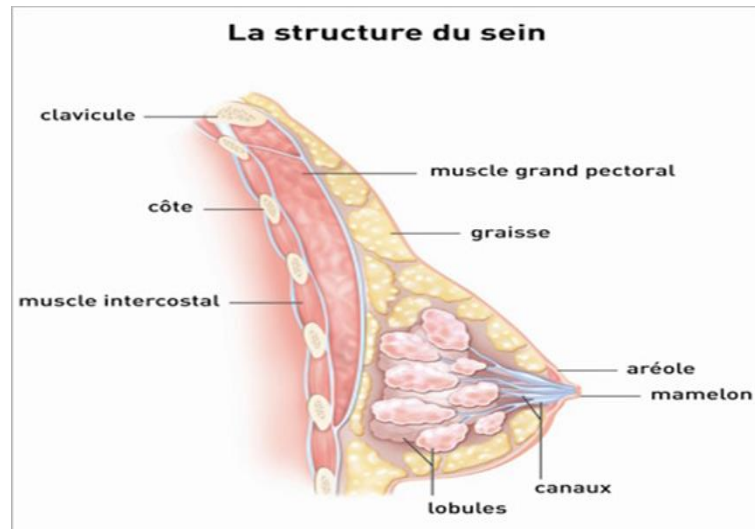


Figure 2.1 : Structure du sein [3]

1.3 Anatomie descriptive mammaire

La glande mammaire est une unité cutanée et glandulaire, qui commence son développement anatomique dès les premières semaines de la vie fœtale (4e semaine) en suivant les étapes hormonales de la vie de la femme (puberté, grossesse allaitement sevrage, ménopause) [2].

Les seins occupent la partie antéro-supérieure du thorax, de part et d'autre du sternum en avant des muscles pectoraux, en regard de l'espace compris entre la 3ème et la 7ème côte, le mamelon se situe en regard de la 9ème vertèbre dorsale. En position debout, sous l'influence de son propre poids, le sein tombe légèrement, ce qui crée le sillon infra-mammaire entre la moitié inférieure du sein et le thorax [2].

Selon la convention de repérage on peut diviser le sein en quatre quadrants, plus une partie axiale. Ceci ne correspond à aucune réalité anatomique, c'est juste pour la réalisation de compte- rendu des observations cliniques et radiologiques (figure 1.2).

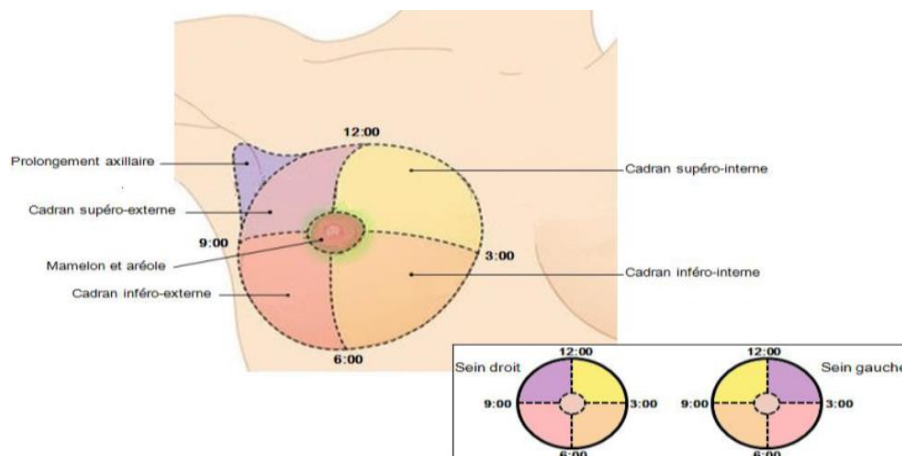


Figure 3.2 : Quatre quadrants du sein [3]

1.4 Système lymphatique du sein

Le sein contient aussi un réseau de vaisseaux lymphatiques dont, l'enjeu principal des drainages lymphatiques, est la transportation de la lymphe des tissus jusqu'aux ganglions lymphatiques, entourant la zone mammaire. Parfois le cancer du sein est infiltrant, il peut atteindre les ganglions lymphatiques afin de développer des métastases [4]. Les groupes ganglionnaires capitaux en matière d'extension des cancers du sein sont :

a Les ganglions axillaires

Situés dans le creux axillaire (aisselle), au nombre moyen de 40, les ganglions (figure 1.3) sont répartis sur 3 niveaux en fonction de leur proximité au muscle grand pectoral. On mesure ainsi la propagation d'un cancer du sein, selon le niveau :

- Niveau I : les ganglions sont situés à la partie la plus externe du grand pectoral, ou région axillaire inférieure ;
- Niveau II : ils sont situés sous le muscle petit pectoral, ou région axillaire centrale ;
- Niveau III : ils sont situés au bord le plus interne du petit pectoral, ou région axillaire supérieure.

b Les ganglions sus-claviculaires : sont situés au-dessus de la clavicule.

c Les ganglions sous-claviculaires : sont situés sous la clavicule.

d **Les ganglions mammaires internes** : sont situés autour du sternum, à l'intérieur du thorax.

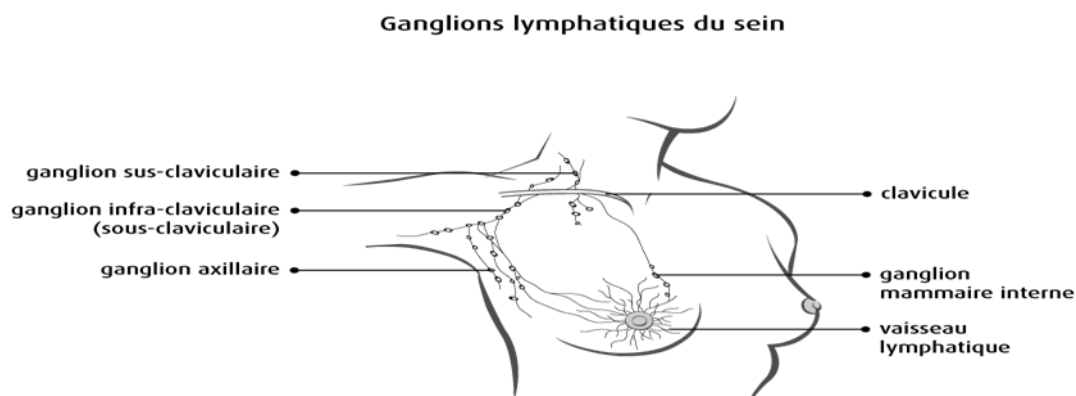


Figure 4.3 : Ganglions du sein [4]

1.5 Lésions bénignes du sein

Les lésions bénignes sont des tumeurs moins néfastes limitées, encapsulées, faciles à repérer avec une croissance lente. Elles n'entraînent pas de métastase. La tumeur du sein la plus fréquente est l'adénofibrome observé chez la femme non ménopausée. La deuxième catégorie est dite maligne, plus néfaste, non encapsulé mal déterminé, des cellules avec une croissance très rapide, elles peuvent entraîner une métastase. La tumeur maligne la plus fréquente est l'adénocarcinome [5].

1.5.1 Adénofibrome

Adénofibrome ou le fibroadénome est la tumeur bénigne la plus fréquente, chez les femmes jeunes. Elle est composée de tissu conjonctif fibreux et de tissu glandulaire. La plupart des fibroadénomes mesurent de 1 à 2 cm, mais elles peuvent atteindre jusqu'à 5 cm. Cette tumeur se présente par un nodule induré, mobile, limité et de consistance élastique.

1.5.2 Tumeur phyllode bénigne

Ce sont des tumeurs rares mixtes fibroépithéliales, particulières par le déséquilibre des deux composants en faveur du tissu conjonctif.

1.5.3 Dystrophie fibrokystique

C'est une affection bénigne qui touche plus de la moitié des femmes, âgées de 30 à 50 ans. Elle se caractérise par la formation de nodules et de kystes à l'intérieur du sein, à la suite de la stimulation des hormones féminines. Elle peut causer un gonflement, de la douleur ou de la sensibilité aux deux seins, surtout dans la partie supéro-externe. Parfois il peut y avoir un écoulement du sein de type ectasie.

1.5.4 Abscesses

Ils se présentent sous la forme d'un ensemble des bosses rouges, sensibles, qui contiennent du pus. On les diagnostique beaucoup plus, chez la femme enceinte ou allaitante.

1.5.5 Kystes

Il s'agit d'une cavité remplie de liquide ou semi liquide, produit par la glande mammaire. Si le liquide s'accumule, la patiente peut sentir une masse ovale ou ronde de 1 cm ou 2cm de diamètre.

1.6 Introduction au Cancer du sein

1.6.1 Définition d'un cancer

Provoqué par un dysfonctionnement, le cancer est une anomalie qui atteint certaines cellules du corps [6], qui finissent par former une masse. Parfois les cellules cancéreuses migrent à travers les vaisseaux sanguins et les vaisseaux lymphatiques afin de former une nouvelle tumeur détachée d'une tumeur primitive dans une autre partie du corps (métastase). Il existent plusieurs types de cancer, en particulier en Algérie (figure 1.4).

Number of new cases in 2018, females, all ages

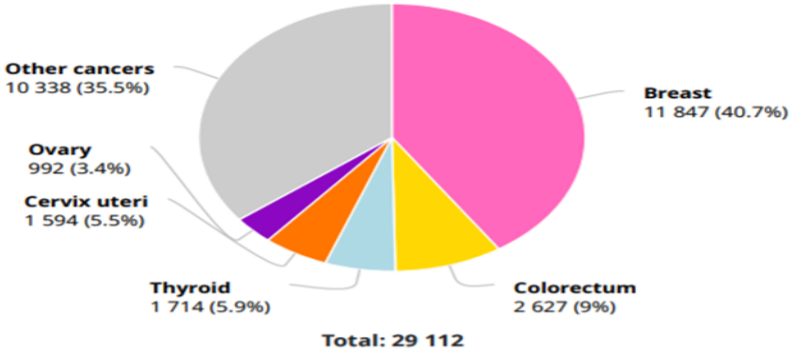


Figure 5.4 : Principaux types de cancers chez les femmes et les hommes en Algérie en 2018[7]

Le cancer du sein se définit comme une prolifération anarchique et incontrôlée, des cellules épithéliales du sein. Qu’il s’agisse des cellules des canaux galactophores « carcinome canalaire » ou de celle des lobules « carcinome lobulaire », on parle « d’adénocarcinome » c’est-à-dire un cancer du tissu glandulaire. Le carcinome peut être in situ ou infiltrant, selon qu’il y ait ou non, effraction de la membrane basale, il possède ou non, un potentiel métastatique. Rarement la tumeur se développe dans le tissu conjonctif du sein, on parle alors du cancer non glandulaire ou « tumeur phyllode du sein » [8, 9, 10].

Plusieurs symptômes jalonnent cette maladie, les tumeurs du sein ne causent aucun signe ni symptôme au début plus ils se développent, plus ils se manifestent.

1.6.2 Symptômes du cancer du sein

Le carcinome canalaire ou lobulaire se caractérise par l’apparition d’une grosseur (boule) au niveau du sein ou de l’aisselle (figure 1.5). Une déformation du sein est une modification de la texture, de la pigmentation de la peau du sein ou de l’aréole ou, l’apparition d’écoulements anormaux au niveau du mamelon [8, 11].

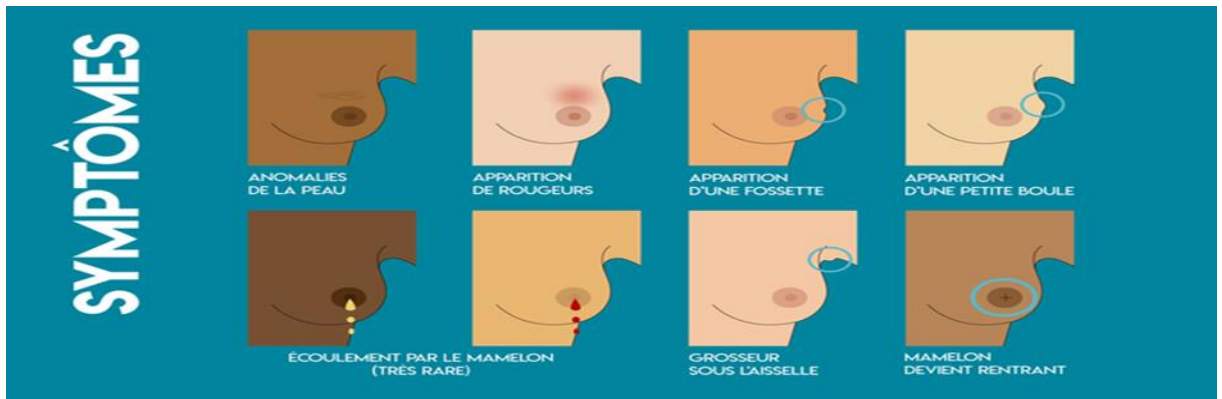


Figure 6.5 Symptômes du cancer du sein [8]

1.7 Classification et stadification des cancers du sein

Le système de stadification est basé sur la classification TNM ainsi que, les résultats du diagnostic initial, afin de planifier un traitement efficace [10, 11].

1.7.1 Classification

Il est extrêmement nécessaire de classer les tumeurs du sein, afin de créer des groupes homogènes sur un plan pronostique et un plan thérapeutique.

La classification TNM (tumeur, nodule, métastase) est fondée sur l'extension anatomique de la tumeur, déterminée par la clinique et l'histopathologie. C'est un ensemble de lettres T N M, combiné à un ensemble de chiffres en romain I, II, III, IV qui augmentent en fonction de la gravité du cancer. C'est-à-dire une relation de corrélation directe dans le but de décrire la taille de la tumeur et le degré de la propagation dans le corps.

- La cotation T (tumeur) : dépend du volume tumoral et son degré d'extension, sur une échelle de T I à T IV.
- La cotation N (nodule, ganglion) : désigne le degré d'atteinte des ganglions lymphatiques situés à proximité de la tumeur, sur une échelle de N I à N III
- La cotation M (métastase) : signifie s'il y a des tumeurs détachées d'une tumeur primitive, dans une autre partie du corps. M I signifie la présence d'une métastase et M 0 représente l'absence d'une métastase.

1.7.2 Différents stades du cancer du sein

Il existe cinq stades du cancer du sein présentés par un ensemble de chiffres romain allant de 0 jusque à IV. Plus le chiffre augmente plus la tumeur, devient néfaste et envahissante.

- Stade 0 : Les cellules anormales restent confinées à l'intérieur du canal où elles sont apparues initialement.
- Stade I : La tumeur mesure moins de 2 cm et de petits amas de cellules cancéreuses sont retrouvés dans les ganglions lymphatiques. Le cancer de stade I est divisé en stades IA et IB.
- Stade II : soit la tumeur mesure moins de 2 cm qui s'est propagée aux ganglions lymphatiques de l'aisselle, soit la tumeur mesure entre 2 cm et 5 cm de diamètre sans s'être propagée vers les ganglions lymphatiques de l'aisselle. Le cancer de stade II est divisé en stades, IIA et IIB.
- Stade III : la tumeur peut avoir n'importe quelle taille, mais elle s'est propagée à la paroi thoracique et/ou à la peau du sein ainsi qu'aux différents ganglions lymphatiques. Le cancer de stade III est divisé en stades IIIA, IIIB et IIIC
- Stade IV : ou Le cancer métastatique. Le cancer s'est propagé à d'autres organes du corps.

1.7.3 Différents types tumeurs malignes du sein

On peut distinguer plusieurs types du cancer du sein selon leur localisation, extension et évolution [6, 9,11]. Les trois formes de cancer les plus connues sont les suivants :

- Le carcinome in situ : il s'agit du cancer non invasif ou le carcinome in situ (CCIS), le plus fréquent. C'est une tumeur maligne non métastatique ; les cellules cancéreuses prolifèrent à l'intérieur des canaux galactophores. Ce type de pathologie mammaire est souvent diagnostiqué, grâce à la mammographie. Un bon pronostic avec un suivi global, peut abaisser le taux de récurrence invasive.

- Les carcinomes lobulaires in situ (CLIS) : le cancer néoplasique lobulaire in situ ou le cancer lobulaire in situ, est le cancer du sein in situ le plus rare. Il est considéré comme un facteur de risque, qui peut développer un cancer du sein.
- Cancer infiltrant : ou le cancer invasif. On parle de ce type de cancer lorsque les cellules cancéreuses, envahissent le tissu qui entoure les canaux et les lobules. Il se propage ensuite vers les ganglions axillaires, puis vers d'autres parties du corps. Il existe deux types du cancer infiltrant (figure 1.6) :
 - Carcinome canalaire infiltrant : un cancer qui surgit dans les canaux, se propageant aux tissus mammaires voisins.
 - Carcinome lobulaire infiltrant : un cancer qui apparait dans les lobules, envahissant les tissus mammaire environnants.

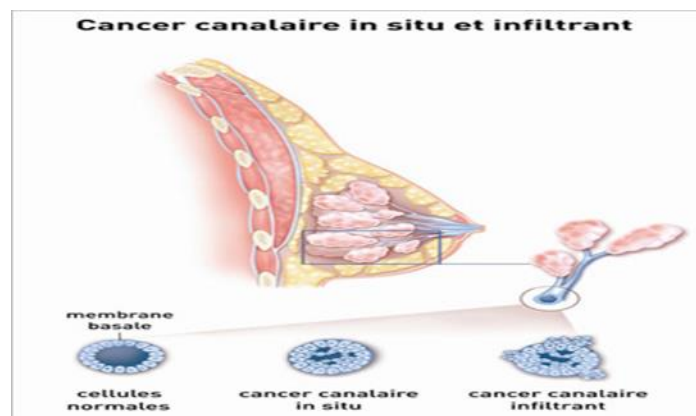


Figure 7.6 : Cancers canaux in situ et infiltrant [11]

- Cancer métastatique : il est connu aussi par le cancer du sein avancé, ou de stade IV. Les cellules néoplasiques des carcinomes invasifs, envahissent le tissu mammaire environnant et peuvent disséminer à partir de la tumeur primaire, le plus souvent par le réseau lymphatique, mais aussi par les vaisseaux sanguins.
- Cancer inflammatoire : c'est un cancer rare et agressif, car il se développe et se propage rapidement. Il affecte le plus souvent, les femmes jeunes et les femmes d'origine africaine.

- Le cancer inflammatoire se présente comme une mastite. Il apparaît lorsque les cellules cancéreuses, bloquent les vaisseaux lymphatiques de la peau du sein. La rougeur, le gonflement du sein et parfois la douleur, sont les symptômes majeurs associés au cancer inflammatoire.
- La maladie de paget : c'est un cancer de la peau du mamelon ou l'aréole, qui touche les femmes à partir de 50 ans, notamment les femmes qui ont déjà été atteintes par un carcinome canalaire infiltrant ou un carcinome canalaire in situ (CCIS). La maladie de paget correspond à la présence de cellules de paget, au niveaux de la peau mamelonnaire .

1.8 Moyens de dépistages et du diagnostic

Personne n'est malheureusement à l'abri du cancer du sein. Cela nécessite une grande vigilance, qui exige un dépistage, un diagnostic et un traitement [12, 13].

1.8.1 Dépistage

Le dépistage est un examen de prévention, qui vérifie la présence d'une maladie avant l'apparition de tous les symptômes. Il est parfois judicieux de se présenter à un test de dépistage. Car lorsqu'on détecte un cancer sein à un stade précoce et qu'on le traite, les chances de guérison sont élevées. En plus de l'examen de dépistage, l'examen de l'auto palpation joue un rôle majeur dans la détection précoce d'une pathologie mammaire.

L'examen de dépistage contient trois phases cruciales ; un interrogatoire suivi par un examen clinique, un examen radiologique complété éventuellement par un examen biologique.

1.8.2 Age et facteurs de risque

Les femmes devraient être au courant de leur vulnérabilité au cancer du sein, selon leurs âges et leurs facteurs de risque.

a L'âge : l'âge joue un rôle majeur dans la prévention. La maladie est rare chez les femmes de moins de 30 ans, les femmes de 50 ans ou plus doivent passer une mammographie tous les 2 ans, à l'âge de 75 ans une double lecture des clichés par deux radiologues différents est nécessaire.

b Facteurs de risque

- Les antécédents familiaux : le cancer du sein (jeune âge, bilatéral multifocal...), ou d'autres cancers (ovaire ...), sont des arguments en faveur d'une mutation génétique.
- Des antécédents en pathologie mammaire, en particulier mastose histologiquement à risque.
- La densité mammographique : pour les femmes ayant des seins denses en mammographie, le risque est multiplié de deux à six fois.
- Le risque hormonal :
 - Principalement l'exposition aux œstrogènes : âge précoce de 1ere menstruation, la ménopause tardive, exogène (prise d'une chimiothérapie substituée prolongée (plus de 5ans)).
 - Les facteurs de la vie reproductive : première grossesse tardive, l'absence de l'allaitement maternel prolongé.
- Examen Radiation ionisante. Un suivi intensif de plusieurs groupes de population, a montré que le sein est l'un des organes les plus sensibles, aux effets des radiations.
- Les facteurs de risques liés au mode de vie tels que l'obésité, l'inactivité physique, la consommation de l'alcool, le tabac et la mauvaise alimentation, peuvent favoriser l'apparition d'un cancer du sein.
- Le sexe : les femmes sont beaucoup plus touchées par cette maladie, a contrario aux hommes.
- Les prédispositions génétiques.

1.8.3 Diagnostic

Le diagnostic désigne le processus qui permet d'identifier la cause, d'un problème de santé. Le processus diagnostique du cancer du sein, commence dès la présence d'un nodule (une masse) au niveau des seins ou bien lorsqu'une mammographie de dépistage révèle une anomalie.

La démarche conventionnelle du processeur diagnostique liée à la tumeur du sein est caractérisée par trois phases cruciales : l'examen clinique, l'imagerie mammaire et anatomopathologique.

a L'examen clinique

C'est un examen réalisé avant et après, l'imagerie mammaire. Il débute par un interrogatoire suivi d'un examen physique.

- L'interrogatoire : c'est un ensemble de questions posé par le médecin, qui lui permet de révéler le contexte de la découverte et les signes d'accompagnement, les facteurs de risques associés, les événements et les troubles médicaux.
- L'examen physique, appelé aussi l'examen complet repose sur le triplet (observation, palpation et écoute).

b L'imagerie mammaire

Il existe trois examens complémentaires classiques :

- La mammographie ;
- L'échographie mammaire ;
- L'imagerie par résonance magnétique

c Anatomopathologie

L'examen anatomopathologique par abréviation <<examen anapath>> permet d'analyser des cellules ou des tissus prélevés au niveau d'une anomalie du sein, réalisé avant et après la chirurgie. Il est effectué par un médecin, un chirurgien ou un radiologue.

L'examen se fait d'abord à l'œil nu, puis il est complété par la ponction cytologique et la biopsie.

1.9 Imagerie mammaire

L'imagerie médicale notamment l'imagerie mammaire a connu un énorme progrès ces vingt dernières années [20]. Ces outils donnent le privilège d'étudier la physiologie ou le métabolisme du corps humain sans interagir directement avec celui-ci, dans le but de la détection des anomalies radiographiques.

L'imagerie médicale est fondée sur plusieurs phénomènes physiques tels que la résonance magnétique, la radioactivité, l'absorption et l'atténuation des rayons X, la réflexion d'ondes d'ultrasons et l'effet photoélectrique, afin d'assurer une meilleure acquisition et restitution des images.

Il existe plusieurs de modalités médicales, la mammographie à rayon X, l'échographie, et l'imagerie par résonance magnétique [14, 15, 16, 17, 18, 19].

1.9.1 Mammographie

La mammographie (figure 1.7) nommée aussi mastographie ou la radiographie des seins est une technique d'imagerie par transmission et par émission, qui utilise l'énergie des rayons X pour visualiser les structures de l'intérieur du sein et, de détecter d'éventuelles anomalies mammaires. Il faut savoir que les rayons X délivrés lors du déroulement d'examen sont très faibles, pour éviter quelques risques. Néanmoins chez certaines femmes les risques sont toujours présents, c'est le cas chez les femmes enceintes ou allaitantes. La mammographie peut être réalisée soit dans le cadre du dépistage soit dans le cadre de diagnostic [15].

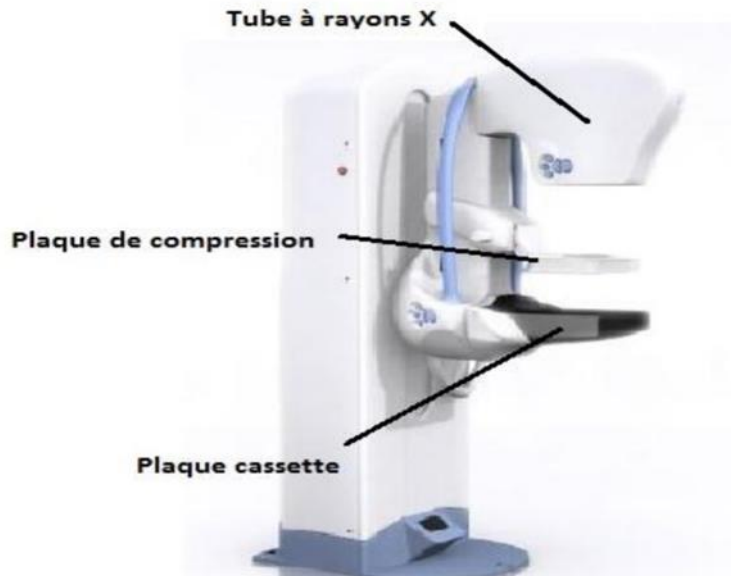


Figure 8 : Mammographe analogique [15]

1.9.2 Echographie

Créée en 1990, l'échographie a connu une grande évolution, jusqu'à l'acquisition d'images 4D (figure 1.8). L'échographie mammaire est un examen d'investigation, de seconde intention et relativement peu coûteux, qui se base sur l'utilisation des ultrasons pour produire des images interne du sein. Ces dernières assurent une étude structurale, de la glande mammaire. Cet examen est prescrit en complément de la mammographie dans le but d'orienter le diagnostic vers la bénignité ou la malignité d'anomalies. L'échographie, permet de guider les examens de la biopsie. Elle contribue aussi, à la surveillance des seins opérés ou traités par la chimio ou radiothérapie, après un cancer. Il sert aussi à mesurer la taille des anomalies. Parfois il arrive que l'échographie soit réalisée avant la mammographie, chez les femmes enceintes, allaitantes, les femmes ayant des seins denses notamment les femmes jeunes (avant 40 ans) et les femmes ayant des prothèses [16, 17].

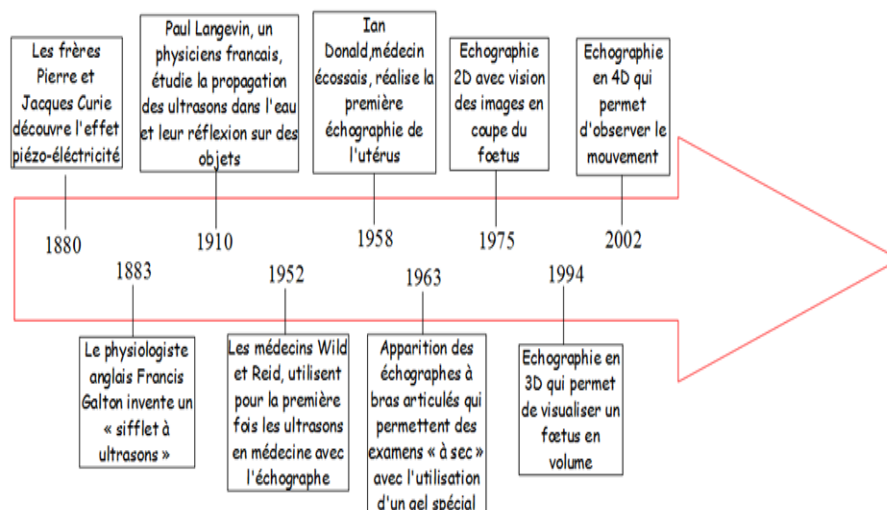


Figure 9 : Chronologie de l'échographie [17]

a Principe et appareillage

L'échographie est réalisée à l'aide d'un appareillage (figure 1.9) en temps réel à balayage manuel, qui ressemble à un ordinateur dit échographe. Cet appareil est constitué d'une sonde, d'un gel hypoallergénique, d'un moniteur, d'une console et d'un système informatique. L'échographie mammaire est réalisée à l'aide des sondes à hautes fréquences. Ces fréquences sont comprises entre 7 et 13 MHz, selon le type des lésions et la morphologie des seins. L'élément de base de l'échographie est la céramique située dans la sonde, qui génère un effet appelé effet piézo-électrique. Elle fonctionne alternativement en émettrice et en réceptrice.



Figure 10 : Echographe [17]

b Limites de l'échographie

Comme toutes les modalités d'imagerie, l'échographie présente aussi quelques limites. L'échographie ne peut pas acquérir une image d'un sein complet, en une seule fois. Elle n'est pas capable de détecter les anomalies les plus profondes, dans le tissu mammaire et elle ne permet pas de classer les microcalcifications [17].

1.9.3 Imagerie par résonance magnétique mammaire

L'IRM mammaire ou l'IRM des seins, est un examen indolore qui ne remplace pas la mammographie, ni l'échographie mammaire. Ceci est parfois recommandé par un médecin pour renforcer l'examen de diagnostic, ou réaliser un bilan d'extension, pour guider un prélèvement par biopsie [18, 19].

1.10 Etude échographique

Les structures échographiques étudient les différents aspects du tissu mammaire [20].

1.10.1 Aspects pathologiques

Les lésions bénignes se caractérisent par une structure régulière, homogène et une déformation minimale des structures adjacentes.

- **Kystes** : à l'échographie, on observe une absence d'échos internes. Ils présentent un contour lisse, de forme ronde ou ovale (figure 1.10).

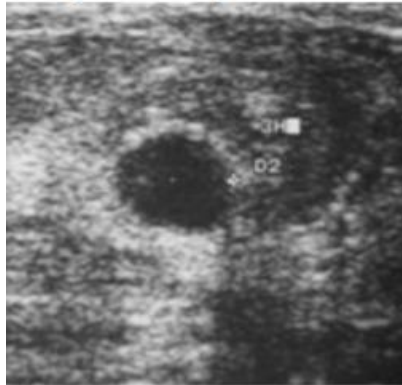


Figure 11 : kyste [20]

- **Fibroadénome** : il présente une image lacunaire et, pauvre en échos internes. Les contours sont étroits et réguliers, homogènes, long ou oval avec un diamètre parallèle à la zone.
- **Carcinome** : il est de forme irrégulière (figures 1.11 et 1.12) avec un axe vertical. Les contours sont larges et irréguliers, avec une structure échographique interne hétérogène et pauvre en échos, non compressible, de mauvaise mobilité, on observe une déformation de l'architecture environnante.



Figure 12 : Tumeur du sein d'aspect malin (lacune hypoéchogène, hétérogène, forme irrégulière, sans cône d'ombre postérieure, non parallèle au plan cutané) [20]



Figure 13 : Tumeur du sein d'aspect malin (contours irréguliers : indentations, spicules) [20]

1.10.2 Aspect échographique du ganglion axillaire envahi

la sensibilité de l'échographie dans la détection des métastases axillaires, est de l'ordre de 60 à 70%. Les ganglions visibles sont pratiquement hypoéchogènes et souvent suspects. Les ganglions sont détectés à partir de 5 à 6 mm. Les ganglions envahis peuvent apparaître arrondis ou irréguliers, s'ils sont volumineux, avec une perte de l'échostructure normale. Ils sont nettement hypoéchogènes. L'échographie de la région axillaire, présente surtout un intérêt dans l'évaluation du stade des cancers avancés ne pouvant pas bénéficier de la chirurgie, des tumeurs traitées par radiothérapie seule, des tumeurs opérées lorsque le curage axillaire n'a pas été complet ou de la surveillance volumique des ganglions sous chimiothérapie [20].

1.11 Classification BIRADS des pathologies mammaires

Les radiologues utilisent la classification BIRADS (breast imaging reporting and data system) l'ACR (l'American College of Radiology) pour l'interprétation des clichés mammographiques. Cette classification établit six catégories (tableau 1.1), en cas d'une investigation radiologique [21].

Catégorie	Image	Risque de cancer	Conduite
0	Investigation incomplète		Investigation à compléter ou comparer avec films antérieurs
1	Normal	0 %	Retour au dépistage
2	Anomalie bénigne	0 %	Retour au dépistage
3	Anomalie probablement bénigne	> 0 % mais ≤ 2%	Suivi 6 mois
4 4A 4B 4C	Anomalie demandant une biopsie faiblement suspecte modérément suspecte très suspecte	> 2 % mais < 95 % > 2 % à ≤ 10 % > 10 % à ≤ 50 % > 50 % à < 95 %	Biopsie

5	Anomalie fortement suspecte d'un cancer	≥ 95 %	Biopsie
6	Cancer prouvé à la biopsie	100 %	Chirurgie

Tableau 1.1 Classification BI_RADS (breast imaging reporting and data system) [21]

1.12 Traitement médical

Il existe plusieurs types de traitement [22] selon le type de la tumeur, le stade du cancer, les préférences personnelles et l'âge. Un suivi clinique est indispensable tous les trois mois, pendant un an puis tous les six mois pendant cinq ans. Une mammographie est exigée tous les six mois puis tous les ans pendant cinq ans.

Les différents types de traitements se résument suivant :

- Un traitement local : est souvent utilisé lorsque la tumeur est présente sur une région bien spécifique par exemple (la radiothérapie et la chirurgie).
- Un traitement systémique : se fait par l'injection des molécules chimiques, absorbées par les cellules et les tissus, afin de détruire les cellules cancéreuses.
- Un traitement ciblé : consiste à utiliser des médicaments < ciblés >. pour attaquer les cellules cancéreuses, afin d'arrêter la propagation de ces dernières.

1.13 Application de l'intelligence artificielle à la prédiction et au diagnostic en cancérologie

L'intelligence artificielle est devenue un outil précieux dans la lutte contre le cancer. Cette technologie est mise au service de l'imagerie mammaire et de la santé de la femme. En effet, les outils d'imagerie mammaire jouent un rôle important dans le diagnostic et dans le dépistage du cancer du sein. Néanmoins l'interprétation directe d'un très grand nombre des clichés est peut-être fastidieuse. Pendant les débuts des années 1980, des systèmes de diagnostic assisté par ordinateur (CAD) ont été développés. Dans le début des années 1970 aux années 1990, l'analyse des images est réalisée à l'aide d'un traitement de bas niveaux ; le débruitage, le rehaussement du

contraste, la détection de bords et de lignes, la segmentation des régions en se basant sur les différents outils mathématiques [22].

A l'heure actuelle de nouveaux systèmes d'intelligences artificielles inspirées du cerveau humain, sont à la disposition des médecins.

L'intelligence artificielle a fait preuve d'une meilleure précision, pour identifier un cancer du sein à un stade précoce et, réduire le taux de faux positifs et de faux négatifs.

Les outils d'intelligence artificielle peuvent analyser un nombre important d'images médicales, en un temps restreint. Avec l'adoption des techniques adéquates, l'intelligence artificielle est aussi capable de prédire un cancer du sein, avant qu'il ne soit présent.

1.14 Conclusion

Ce chapitre nous a permis d'introduire quelques notions de base sur l'anatomie du sein, les différents types de pathologies mammaires, le dépistage, le diagnostic, le traitement et de mettre en évidence l'impact majeur des outils d'imagerie mammaire en l'occurrence l'échographie. Malgré la compétence des médecins et le développement incessant des outils d'imagerie médicale, le taux de mortalité due au cancer du sein, augmente jour après jour.

A cet effet l'idée de notre projet est née, qui consiste en la prédiction de l'apparition d'un cancer du sein quelques années avant qu'il ne soit présent. Dans ce cadre, des connaissances approfondies sur l'intelligence artificielle, sont indispensables, pour la réalisation de notre travail. C'est l'objectif du second chapitre.

2.1 Introduction

À présent, l'intelligence artificielle se glisse jusqu'au plus intime, de nous-même. Elle reproduit nos pensées par l'intervention à chaque recherche sur Google, une publication sur un des réseaux sociaux ou à chaque utilisation de notre ami inséparable, le smartphone. Elle pénètre au cœur de nos cellules, afin de décrypter notre code génétique en quelques minutes. AI est capable aussi de sauver des vies, grâce à la prédiction des cancers, notamment le cancer du sein. Alors demain à ses côtés qu'elle sera notre place ? L'intelligence artificielle se nourrit grâce aux tonnes de données, dites BIG DATA.

Dans ce qui suit nous introduisons l'intelligence artificielle et ses dérivés, l'apprentissage automatique et l'apprentissage profond, afin de construire un système de prédiction du cancer du sein, en échographie. Nous mettrons par la suite, l'accent sur le concept choisi dans le cadre de la réalisation de notre modèle de prédiction, du cancer du sein.

2.2 BIG DATA, data science et intelligence Artificielle

Ces trois termes sont souvent utilisés d'une manière interchangeable, mais il existe une différence subtile entre ces derniers [23, 24] [25, 26].

2.2.1 BIG data

Le terme BIG data ou données massives représente un ensemble massif de données numériques structurées ou non structurées, collectées à travers les réseaux sociaux les applications mobiles, les moteurs de recherche..., puis traitées en temps réel.

Ce traitement n'est pas compatible avec les outils classiques, ce qui donne naissance à l'intelligence artificielle et au data science [24, 25].

2.2.2 La data science

La data science ou la science des données, est un domaine pluridisciplinaire qui utilise des algorithmes, des théories issues des mathématiques appliquées et de la statistique, pour exploiter les données (les connaissances). Le cycle de travail d'un Data Scientist se résume par la figure 2.1.

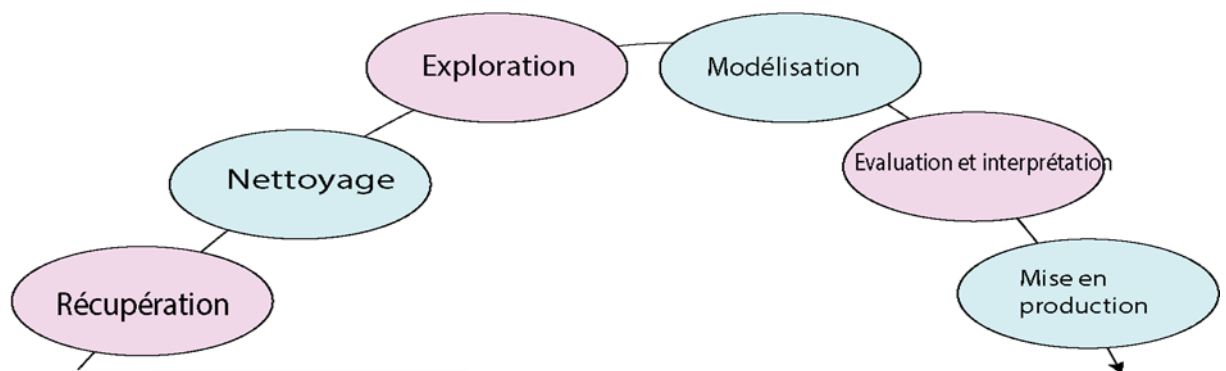


Figure 14 : Cycle de travail d'un datascientist [24]

2.2.3 Introduction à l'intelligence Artificielle

L'intelligence artificielle ou IA est une discipline scientifique, qui tente de reproduire les processus cognitifs humains dans le but de réaliser des actions intelligentes, en se basant sur le traitement de connaissances et, sur le raisonnement.

Il existe 3 grands niveaux de l'intelligence artificielle.

a Intelligence artificielle étroite (IA faible)

IA faible est la capacité d'une machine, à accomplir une seule tâche spécifique mieux qu'un être humain. On ne peut pas parler d'AI étroite, sans mentionner le programme développé par Deep Mind ALPHA Go, qui a abattu Lee Sedol le meilleur joueur de GO en 2017 [26].

b Intelligence artificielle générale (IA générale)

C'est une IA capable d'effectuer toute tâche intellectuelle, avec le même niveau de précision qu'un être humain moyen [26].

c Intelligence artificielle active (IA active)

IA active est capable de battre les humains dans de nombreuses tâches, y compris la créativité scientifique, la clairvoyance et les aptitudes sociales [26].

L'intelligence artificielle (figure 2.2) est pluridisciplinaire, car elle croise de nombreuses disciplines telles que l'informatique, la théorie de l'information, le traitement du signal, la théorie d'optimisation, les statistiques, les neurosciences et la biologie.

Il existe plusieurs sous-ensembles d'IA les plus pertinents, sont l'apprentissage automatique et l'apprentissage profond.

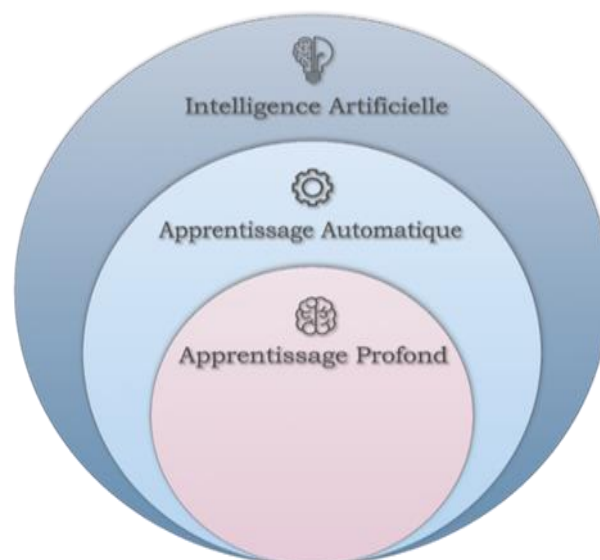


Figure 15 : IA, ML, DL [27]

2.3 De la naissance de l'intelligence artificielle jusqu' à 2020

IA est apparue la première fois en 1950 dans l'article d'Alan Turing intitulé « Computing Machinery and Intelligence » qu'Alan Turing publie en 1950 >> dans lequel le problème de la conscience d'une machine a été exploré [26, 28, 29]. Dans cet article, il a décrit comment une machine s'approche d'une IA. Il appelait cette méthode le jeu de l'imitation, ce qu'on appelle aujourd'hui le test de Turing.

Pendant l'été de 1956 aux États-Unis, une conférence a été organisée par des scientifiques américains, qui s'est tenue au Dartmouth Collège et c'est là où le terme intelligence artificielle a été utilisé pour la première fois.

À partir de là, c'est l'essor de l'IA. En 1974, l'IA a connu une période difficile dite l'hiver de l'IA. Dans cette période les gouvernements cessent le financement des projets, à cause de l'échec récurrent des expériences des experts. Dans les années 80, la recherche sur l'intelligence artificielle fut relancée grâce aux succès des systèmes des experts.

Entre 1990 et 2000, le développement des performances informatiques permet d'exploiter l'IA dans de nouveaux domaines. Dans les années 2000-2010 : l'IA représente un enjeu de société. Au début des années 2000, l'intelligence artificielle s'intègre en grande partie dans l'industrie du cinéma pour des contenus plus spectaculaires. Le film Matrix reste l'exemple le plus marquant du nouveau millénaire.

La société vit un véritable big-bang informatique entre 2000 et 2010, avec des ordinateurs plus performants, le déploiement d'Internet et le développement des smartphones qui lancent l'ère de l'Homo Numericus.

À partir de 2010, c'est l'IA sans limites. En 2011, l'IA brille encore grâce aux prouesses des Watson d'IBM marquant un tournant, dans la médiatisation des recherches.

La loi de Moore continue de guider les progrès de l'intelligence artificielle, donnant vie à de nouveaux procédés pour le Machine learning puis le Deep Learning. Ainsi, de nombreux autres projets suivent cette voie. En 2012, la firme de Mountain View n'avait que quelques projets d'usage, contre 2700 trois ans plus tard. Suivi par Facebook en 2013 qui ouvre Facebook Artificial Intelligence Research (FAIR).

En 2020, l'IA est utilisée dans la lutte contre le coronavirus. AliBaba [29] met à la disposition des médecins une IA capable de détecter le SRAS-CoV-2 en 20 secondes, par scanner pour accélérer la détection du virus.

La gestion des données permettra de mettre l'IA, en application pour surpasser l'Homme dans plusieurs domaines ce qui soulève des questions sur le rôle professionnel de ce dernier dans les années à venir !

2.4 Apprentissage automatique

En premier lieu, il est judicieux de bien cerner la signification du mot apprentissage. L'apprentissage est basé sur la construction d'un modèle général, en utilisant des données particulières du monde réel, dans le but de prédire un comportement ou approximer une fonction.

L'apprentissage est divisé en deux catégories ; la première catégorie est dite apprentissage symbolique issu de l'intelligence artificielle, la deuxième catégorie est dite apprentissage numérique issu des statistiques [30, 31, 32].

La meilleure méthode pour apprendre à un enfant de faire la distinction entre une orange et une pomme, est de lui montrer plusieurs exemples de ces deux fruits.

Après un certain nombre d'expériences, l'enfant sera capable de distinguer les oranges et les pommes sans l'intervention de ses parents. Grâce à la conception du développement de l'enfant, l'apprentissage automatique est apparu.

L'apprentissage automatique nommé aussi l'apprentissage artificiel (AA) en anglais Machine Learning (ML) est un sous-ensemble de l'intelligence artificielle. Son inventeur ARTHUR SAMUEL le définit comme étant un domaine d'étude, qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmé. Dans ce cadre, plusieurs définitions se retrouvent dans la littérature, nous présentons l'une d'entre elles.

➤ Définition de l'AA

L'apprentissage artificiel consiste à extraire, analyser et comprendre les patterns existant dans un ensemble de données. L'objectif principal de cette discipline est d'offrir la possibilité à un ordinateur, d'apprendre seul et d'améliorer ses performances (adaptation) à partir de ces patterns pour reconnaître des exemples jamais vus auparavant.

Le champ d'application de l'apprentissage artificiel est très étendu. Il a envahi quasiment tous les domaines, du diagnostic médical, à la prédiction des prix, ou la détection des fraudes, jusqu'à la recommandation des produits.

2.4.1 Données d'apprentissage automatique

Les données sont divisées en trois catégories. La première catégorie correspond aux variables prédictives. C'est un ensemble d'attributs utilisé pour générer le modèle d'apprentissage. La deuxième catégorie est dite l'ensemble de validation. La dernière catégorie désigne l'ensemble de tests destiné à tester l'algorithme.

Il n'y a pas de règle générale pour la proportion des données, mais on utilise souvent 70 % pour les variables d'entraînement, 15 % pour la validation et 15 % pour l'ensemble du test.

2.4.2 Différents types d'apprentissage automatique

L'apprentissage artificiel se divise en quatre grandes familles, l'apprentissage supervisé, non supervisé, semi-supervisé et par renforcement.

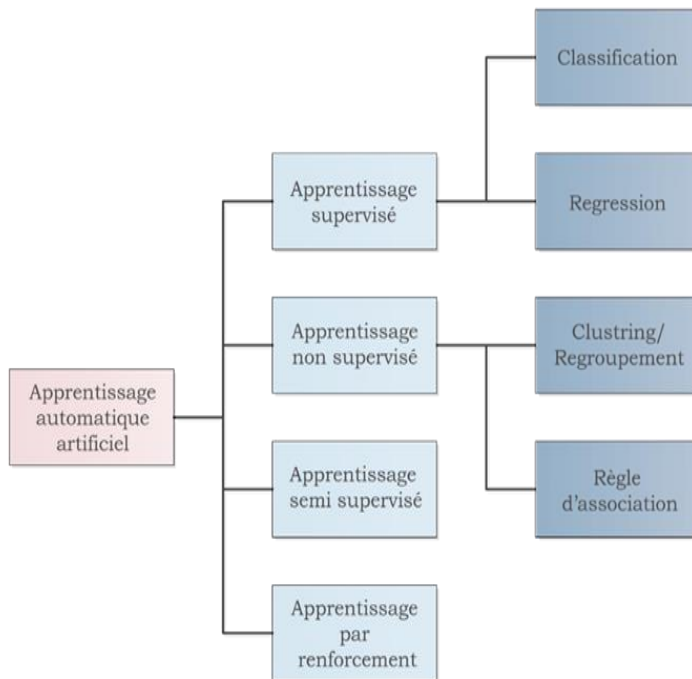


Figure 16 Différents types d'apprentissage automatique [30]

a Apprentissage supervisé

L'apprentissage supervisé est une approche qui a pour objectif, la conception des modèles à partir des données annotées. Ainsi, chaque entrée est déjà associée à une label ou classe cible. Une fois les modèles entraînés, Ils seront capables d'être utilisés dans différentes applications telles que, la prédiction ou la classification des données non annotées.

b Apprentissage non supervisé

Dans ce cas, l'apprentissage se fait d'une manière autonome. Les données livrées à la machine ne sont pas annotées.

Dans ce cas, c'est à l'algorithme de trouver les similarités et les distinctions au sein de ces données, afin de regrouper celles qui ont des caractéristiques communes dans la même classe.

c Apprentissage semi- supervisé

C'est l'hybridation d'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.

d Apprentissage par renforcement

Le jeu de données d'apprentissage par renforcement est similaire aux jeux de données d'apprentissage supervisé. Cependant, l'idée de l'apprentissage par renforcement est d'apprendre à un agent, comment se comporter dans un environnement, en utilisant un système de récompenses ou de pénalité. Pour mieux comprendre une analogie est souvent citée, c'est celle du cycle de la dopamine : une bonne expérience augmente la dopamine d'un agent, ce qui lui donne envie de refaire cette expérience. C'est le cas aussi, de véhicules autonomes.

Parmi les algorithmes supervisés, nous citons le perceptron multicouche [30, 31]. On retrouve dans l'apprentissage non supervisé, les méthodes du clustering. Dans notre cas, ce sont les réseaux de neurones qui nous intéressent.

2.5 Réseaux de neurones artificiels

Les réseaux de neurones artificiels (RNA) sont inspirés par le cerveau humain, qui contient environ 10 milliards de neurones chacun de ces neurones est connecté à environ 10 000 d'autres. A cet effet, il serait astucieux de comprendre le fonctionnement d'un neurone biologique (figure 2.4). 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 [31, 32, 33].

Le signal est ensuite acheminé vers le soma pour se propager à travers l'axone, vers un autre neurone grâce à la synapse.

Les réseaux de neurones biologiques effectuent facilement certaines applications, telles que la reconnaissance des formes, la mémorisation ou la compréhension d'un langage humain tel qu'il est parlé, néanmoins cette simulation reste loin du cerveau humain.

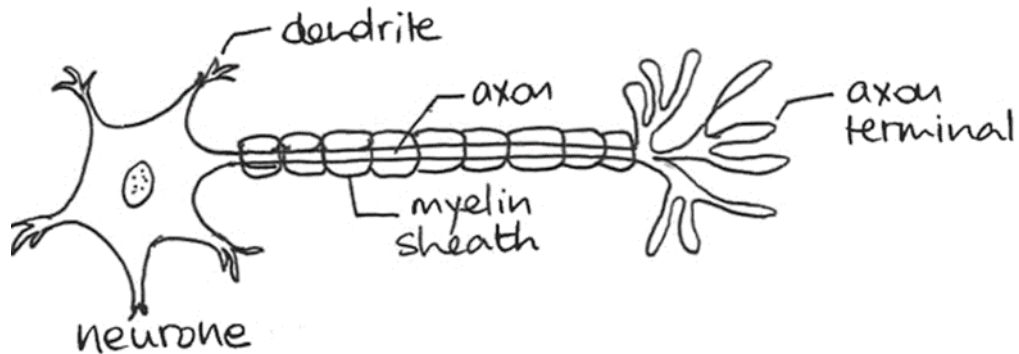


Figure 17 Neurone biologique [31]

2.5.1 Histoire des réseaux de neurones

Tout a commencé en 1943 dans un article de MC CLOCH et PITTS [31,32]. Cet article a révélé pour la première fois, une modélisation d'un neurone physiologique, en neurone formel. Le but majeur de cette étude, était de démontrer que le cerveau humain est analogue à une machine de Turing.

En 1949, une fameuse règle d'apprentissage appelé plasticité synaptique a été présentée par D.O.Hebb dans son ouvrage <<The organization of Behavior>> [31].

Cette règle d'apprentissage inspire plusieurs modèles de réseaux neurones artificielles aujourd'hui. En 1958 ROSENBLATT fait apparaître le premier modèle opérationnel des réseaux de neurones <LE PERCEPTRON>.

En 1960, B. WIDROW présentait le modèle ADALINE (Adaptive Linear Element) inspiré du perceptron. En 1969, M. Minsky et S. Papert ont critiqué les propriétés du perceptron en présentant ses limites, mais cette recherche a été abandonnée. En 1972 T. Kohonen a présenté des travaux, sur les mémoires associatives et proposa des applications à la reconnaissance des formes.

En 1986, Rumelhart, Hinton et Williams, ont réalisé un algorithme de rétropropagation d'erreur, qui permet d'optimiser les paramètres d'un réseau de neurones multicouches. A partir de ce moment-là, plusieurs chercheurs s'intéressent à ce domaine.

De nos jours les réseaux de neurones envahissent plusieurs domaines tels que, le traitement du signal, le diagnostic et le suivi médical.

2.5.2 Principe

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).

2.5.3 Modélisation

Le neurone reçoit une ou plusieurs entrées x à travers un neurone source ou plus. Ces entrées sont des valeurs numériques binaires (0,1), bipolaires (-1,1) ou réelles (représenté en virgule flottante), qui représentent l'état du neurone émetteur. Chaque entrée est multipliée par un certain poids w .

Le nœud fait une somme pondérée de ses entrées. On appelle cette opération, la somme (**Z**). Pour faire apparaître la sortie (**a**) du nœud, une fonction d'activation est appliquée à la somme (**Z**), afin d'activer ou désactiver le neurone (figure 2.5).

La valeur obtenue par ce calcul, représente l'entrée du neurone de la couche suivante. Cette démarche se répète selon le nombre de couches du réseau de neurones, jusqu'à la couche sortie. A ce niveau-là, une sortie prédictive (\hat{Y}) (sortie globale) est obtenue. Cette sortie est ajustée au fur et à mesure, de l'utilisation de la fonction coût et la descente du gradient, de façon qu'elle soit similaire à la sortie réelle (**Y**).

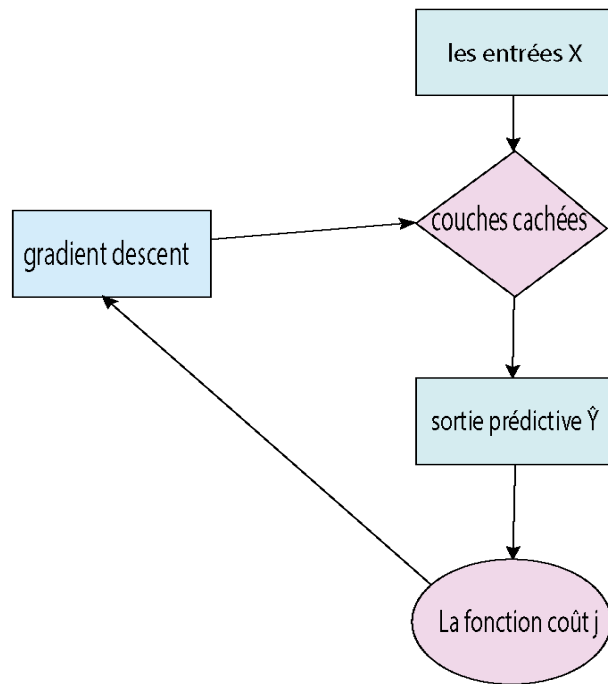


Figure 18 Propagation des données [32]

a Somme pondérée

Une somme pondérée (z) ou une moyenne pondérée, est la moyenne d'un certain nombre de valeurs affectées par des coefficients.

Un neurone artificiel calcule la somme de ses entrées X, multipliées par le poids associé W.

La formule de la somme pondérée Z est représentée par l'équation 2.1.

$$Z_k^i = \sum_{n=0}^n (W_n * X_n) + b \quad 2.1$$

i : est le numéro de la couche suivante.

K : est le numéro de neurone de la couche actuelle.

W : le poids et b : le biais

- Le poids w : le poids d'une valeur aléatoire négative ou positive, ajustable multipliée à l'entrée du neurone.
- Le biais : est un neurone d'entrée qui produit souvent une valeur de 1 en sortie ce qui donne le privilège de déplacer le seuil de classification, pour que le réseau de neurones effectue l'apprentissage.

b Fonction d'activation

Nommée aussi la fonction de transfert (g), elle représente une fonction mathématique linéaire ou non-linéaire, appliquée à la somme pondérée précédente Z , avant de la faire sortir complètement du neurone. Cette fonction est équivalente au potentiel d'action biologique. Si le résultat de la fonction d'activation(g) dépasse un certain seuil, le neurone sera activé. Dans ce cas, la valeur obtenue est le résultat de l'activation (la fonction a .) sinon le neurone reste désactivé (il n'y a pas de transmission.). Il existe plusieurs fonctions d'activations. Leur choix dépend de l'application. A cet effet, les fonctions d'activation les plus exploitées sont les fonctions linéaires.

➤ **La fonction échelon** (figure 2.6) est définie par la relation 2.2.

$$g(Z)=\begin{cases} 1 & \text{si } z > 0 \\ 0 & \text{si } z < 0 \end{cases} \quad 2.2$$

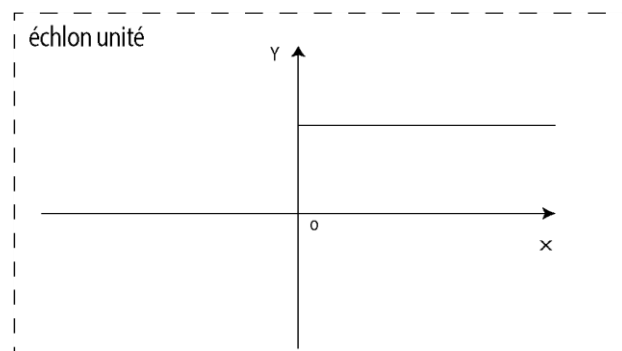


Figure 19 Fonction échelon [33]

La sortie ne prend que deux valeurs : 0 pour un signal négatif et 1 pour un signal positif. Cette fonction n'est pas différentiable, ce qui peut poser des soucis lors de l'application du gradient.

➤ **La fonction sigmoïde** : appelée aussi la fonction logistique, la sigmoïde est une fonction mathématique classique, la plus ancienne et la plus populaire (figure 2.7). C'est une fonction non-linéaire, dérivable, continue et symétrique par rapport à l'axe des y , elle produit une courbe sigmoïde S est définie comme suit :

$$S = \frac{1}{1+e^{-t}}$$

2.3

Avec e : une constante exponentielle

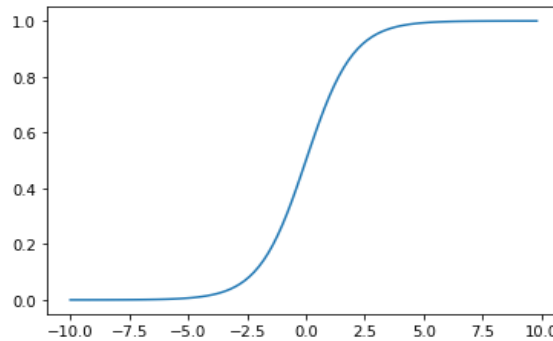


Figure 20 Fonction Sigmoïde [33]

L'utilisation de cette fonction assure que les valeurs restent dans un intervalle étroit [0,1].

Pour les valeurs négatives de Z, la sortie se rapproche de 0, le neurone reste désactivé.

Pour les valeurs positives de Z, la sortie se rapproche de 1, le neurone s'activera.

Pour les valeurs de Z nulles, la sortie est égale à $\frac{1}{2}$.

➤ **La fonction tangente hyperbolique**

C'est une fonction trigonométrique dont la forme de sa courbe (figure 2.8), lui donne le privilège d'être choisie parmi les fonctions d'activations. La seule différence entre cette fonction et la fonction sigmoïde est que Tan(h) est centrée sur le zéro. Elle produit des sorties qui sont comprises dans la plage [-1,1], lui permettant de donner une efficacité en plus aux résultats de calcul.

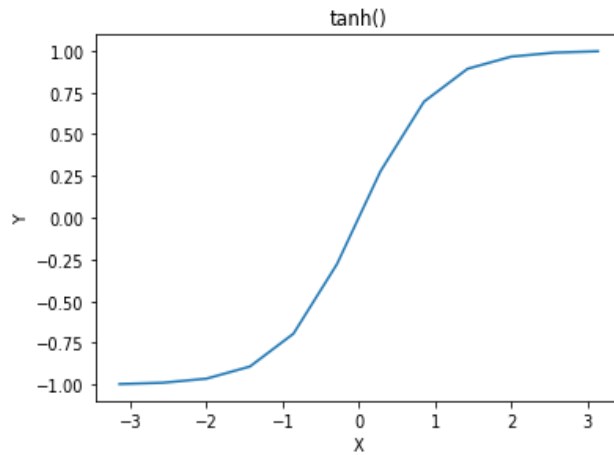


Figure 21 Fonction than [33]

L'équation 2.4 représente la fonction tangente hyperbolique.

$$\text{than}(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad 2.4$$

- **La fonction d'activation rectified linear unit (Relu)** : est une fonction présentée par Tech & Hinton en 2000 qui est devenue à l'heure actuelle la fonction la plus utilisée notamment par les réseaux de neurones profonds, pour résoudre le problème de saturation des fonctions précédentes et l'augmentation de la rapidité d'apprentissage. Relu est définie comme :

$$r(z) = \max(0, z) \quad 2.5$$

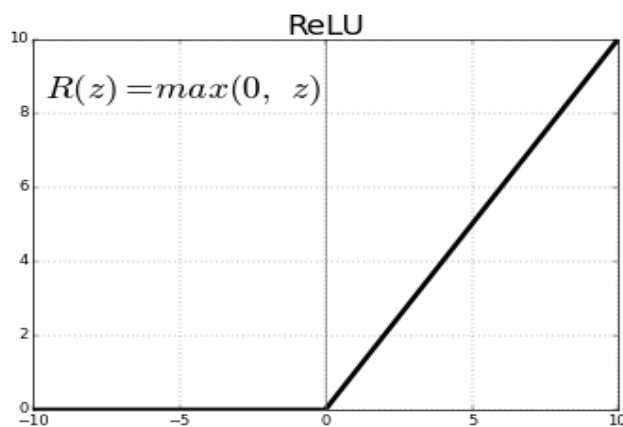


Figure 22 Fonction ReLu [33]

En sortie, cette fonction n'épargne que les valeurs positives de la somme pondérée Z.

➤ **La fonction softmax**

La fonction Softmax (équation 2.6) ou régression logistique multinomiale, est souvent utilisée pour résoudre les problèmes de multiclassification. Elle utilise la distribution de probabilité.

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

c Sortie a

C'est la sortie de neurone de la couche actuelle et, l'entrée de neurone de la couche suivante (équation 2.7).

$$a_k^i = g(z) \quad 2.7$$

Avec :

i : le numéro de la couche suivante.

K : le numéro de neurone de la couche actuelle.

d Fonction prédictive h (x) ou $\hat{Y}(x)$

C'est la sortie finale. Autrement dit, c'est la sortie d'un des neurones de la couche sortie, qui est par la suite comparée à la sortie réelle **Y** du problème proposé, grâce à la fonction coût.

e Fonction coût j

La fonction cout ou la fonction perte, utilisée afin d'estimer la performance de la phase d'apprentissage. En comparant la sortie prédictive à la sorte réelle pour déduire si la sortie prédictive se rapproche de la sortie réelle [20].

Cette comparaison est réalisée à l'aide de plusieurs métriques (2.8, 2.9 et 2.10).

- Erreur quadratique moyenne

$$J = 1/n \sum_{i=1}^n (Y_i - \hat{Y})^2 \quad 2.8$$

-Entropie croisée

$$J = -\sum_{i=1}^n Y_i \log(\hat{Y}) \quad 2.9$$

-Entropie croisée binaire

$$J = -\sum_i^n (y \log(\hat{Y}) + (1 - y) \log(1 - \hat{Y})) \quad 2.10$$

f Descente du gradient

C'est un algorithme d'optimisation utilisé pour minimiser les erreurs du modèle entraînement, en dérivant la fonction de perte pour trouver les valeurs des poids et les valeurs des biais ajustés. Le but de cet algorithme est de trouver le minimum global ($J=0$), après avoir effectué plusieurs itérations.

$$W = w - \alpha \frac{dJ(w,b)}{dw} \quad 2.11$$

$$B = b - \alpha \frac{dJ(w,b)}{db} \quad 2.12$$

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

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

α : Taux d'apprentissage.

2.6 Choix de la fonction d'activation et la fonction de perte

Le tableau ci-dessous résume le choix de la fonction d'activation et, de la fonction coût utilisées pour la classification des données [34].

type de problème	type de sortie	la dernière fonction d'activation	La fonction coût
regression	valeurs numériques	linéaire	L'erreur quadratique moyenne
classification	sortie binaire	sgmoide	Entropie croisée binaire
classification	une seule étiquette multi_classes	softmax	Entropie croisée
classification	multi-étiquettes multi_classes	sgmoid	Entropie croisée binaire

Tableau 231 Choix de la fonction d'activation et la fonction perte [34]

2.7 Les propriétés des réseaux de neurones

Les réseaux de neurones [32] sont fondés sur six propriétés :

- **Parallélisme** : le parallélisme désigne l'ensemble des entrées qui travaillent en même temps.
- **La tolérance aux pannes** : grâce au travail collectif, simultané et à la structure des RNN, le réseau de neurones fait preuve de résistance au bruit et aux pannes.
- **La capacité d'adaptation** : c'est l'aptitude de tenir compte, face à des nouveaux problèmes ou des nouvelles données arrivant de l'extérieur.
- **La capacité d'auto-organisation** : celle-ci permet aux réseaux de neurones de se stabiliser dans le cas des systèmes dynamiques.
- **La généralisation** : c'est la capacité d'un réseau de neurones à reconnaître des nouveaux exemples qui ne sont pas déjà vus.
- **Structure de connexion** : la combinaison de ces différentes connexions, décrit une topologie.

2.8 Différents types de réseaux de neurones

On distingue deux grandes familles des réseaux de neurones, les réseaux monocouches et les réseaux de neurones multicouches [32, 33, 34].

2.8.1 Les réseaux de neurones mono couches

Ce sont des réseaux de neurones composés par deux couches : la couche d'entrée et la couche de sortie (figure 2.10). Le rôle principal de la couche d'entrée, est de fournir les données qui vont être traitées dans la couche de sortie (la sommation et l'activation).

Le perceptron qui est le premier réseau de neurones évolutif supervisé, est inspiré par le système visuel humain. Ce réseau de neurones a pour objectif la reconnaissance des lettres d'alphabet en utilisant un capteur et, en effectuant des calculs logiques simples (AND ou OR).

Le perceptron interagit avec les valeurs binaires (0,1) seulement que ce soit en entrée ou en sortie, avec un sens unique de la propagation des données (Feed forward network), suivant l'échelon unité, comme une fonction d'activation.

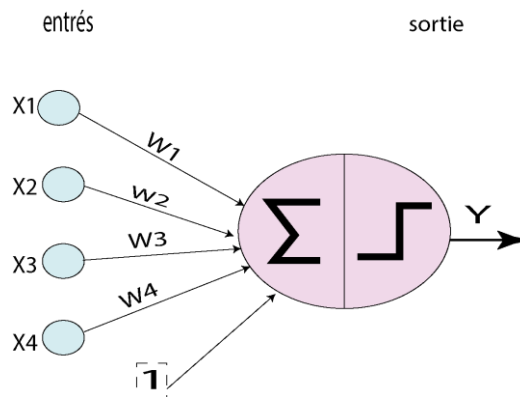


Figure 24 : perceptron simple [33]

2.8.2 Réseau de neurones multicouches

Le réseau de neurones multicouches est une association de plusieurs réseaux de neurones simples, pour former un réseau de neurones de (n) couches, où chaque couche contient (m) nœuds (figure 2.11). Les données peuvent se propager dans un seul sens (feed forward) ou dans les deux sens.

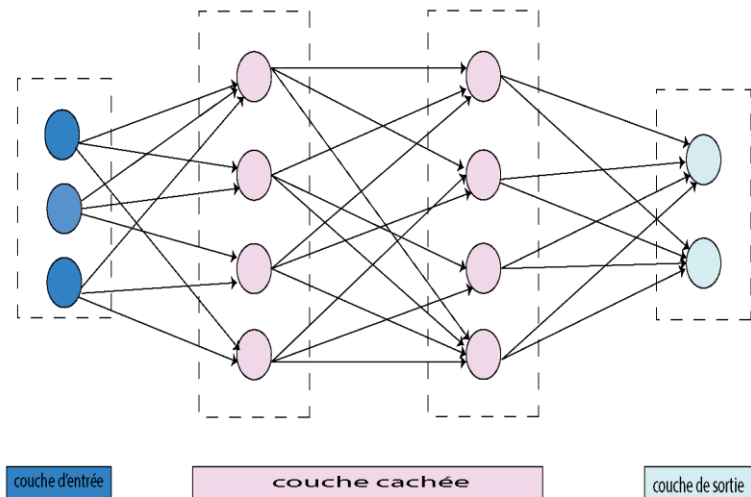


Figure 25: Réseau de neurones multicouches [33]

La couche 0 : est la couche d'entrée. Son rôle principal est d'alimenter le réseau de neurones, en fournissant les données du monde réel.

Les couches cachées ou couches d'apprentissage : représentent le pont qui relie les couches d'entrées et les couches de sorties. Leurs principales fonctionnalités sont de préparer les données, pour la couche de sortie en calculant les sommes pondérées, grâce à la fonction d'activation.

Le nombre de ces couches ne doit pas dépasser les deux couches.

La couche de sortie : est la dernière couche du réseau de neurones qui sert à récolter toutes les données pour présenter les résultats finaux.

Avant de construire n'importe quel réseau de neurones, il faut bien identifier les hyperparamètres :

le nombre de couches cachées ;

le nombre de nœuds dans chaque couche cachée ;

les fonctions d'activations utilisées dans les couches cachées et dans la couche de sortie.

2.9 Différentes architectures des réseaux de neurones

2.9.1 Réseaux Feed-Forward

C'est la structure de connexion la plus simple et la moins utilisée, destinée à résoudre les problèmes de classification basique. En effet, la propagation des données se fait dans un sens unique aller sans retour, en commençant par les nœuds d'entrée pour passer par les couches cachées et atteindre au final les couches des sorties [32, 33, 34].

Ces réseaux sont caractérisés par :

- La simplicité d'apprentissage, de réglage de paramètres et de calcul.
- La tolérance à l'erreur.

2.9.2 Réseaux par rétropropagation (back forward)

Ce type d'architecture est le plus utilisé, le plus fiable, car il est basé sur la minimisation des erreurs, pour que la sortie prédictive soit presque similaire à la sortie réelle, au fur et à mesure de l'entraînement [32, 34].

La propagation des données passe par deux étapes cruciales ; un aller (feed forward) et un retour (back forward) dans le sens inverse. Le sens de la couche sortie vers les couches entrées passant par les couches cachées en calculant les erreurs partielles (delta) entre les nœuds et l'erreur globale en appliquant le gradient descend sur ces derniers afin de les minimiser pour ajuster les poids.

Le nombre des allées et le nombre des retours, doivent être équitables.

2.10 Différence entre époque, taille de lot et une itération

Ces trois paramètres sont utilisés dans le cas où, la base de données est gigantesque. Car il est impossible de transmettre toutes les données, à l'ordinateur d'un seul coup. Pour cela une division du dataset en tailles plus petites, est cruciale [33].

- Une époque est le passage en avant et en arrière, d'un ensemble de données, dans le réseau de neurones.

- La taille de lots (batch size) est le nombre d'échantillons d'apprentissages utilisés dans une propagation aller-retour (feed-forward et back-forward).
- Une itération est le nombre nécessaire de lots, pour compléter une époque.

2.11 Apprentissage profond

En réalité, le terme <<apprentissage profond>> existait depuis les années 40. Néanmoins il a été utilisé officiellement pour la première fois en 2010, suite à son développement, dans divers domaines [23, 24, 33].

La puissance des ordinateurs qui ne cesse d'augmenter, l'accumulation des tonnes de données et une meilleure compréhension des fonctions d'initialisation des poids ainsi que la fonction d'activation, sont les causes majeures de l'apparition de l'apprentissage profond. C'est dans ce cadre, que seules les approches basées sur cet apprentissage, permettent l'extraction automatique des caractéristiques, dans les différentes couches du réseau.

Le deep learning est aussi un logiciel informatique puissant, qui utilise diverses techniques pour entraîner des réseaux de neurones complexes (profonds). Ces réseaux composés de plusieurs couches cachés à partir d'un ensemble de données large, servent à prédire un résultat exploitable, dont sa structure imite celle du cerveau humain.

Dès qu'un réseau de neurones profond reçoit une donnée ou une information, celle-ci excite l'ensemble des nœuds de la première couche, en changeant leurs états actuels afin de faire circuler l'information, jusqu'à la dernière couche et vice-versa. Ceci permet au système d'apprendre d'une manière autonome, en améliorant ses décisions et ses prévisions.

La profondeur du modèle d'apprentissage, est mesurée par le nombre de couches qui le constituent. Un réseau de neurones contenant plus de 10 couches, est considéré comme un réseau de neurones très profond (very deep neural networks).

De nos jours l'apprentissage profond a envahi divers domaines tels que :

- les assistants vocaux :Siri, Alexa, Google ;
- la rédaction des textes ;
- la cyber sécurité ;
- la finance ;
- la médecine ;
- le marketing et la vente ;
- l'agriculture ;
- l'exploration spatiale.

2.11.1 Extraction des caractéristiques

Un jeu de données est composé par un ensemble infini de caractéristiques, qui ne sont pas toutes utiles pour l'apprentissage d'une machine. A cet effet, l'extraction des caractéristiques pertinentes est primordiale [24, 33]. Pour ce faire (figure 2.12), on distingue deux méthodes :

- La méthode classique en utilisant l'apprentissage automatique. Dans ce cas, l'extraction des caractéristiques est faite manuellement, en appliquant les algorithmes de réduction de dimensions.
- Les méthodes par apprentissage profond. Dans ce cas, l'architecture des réseaux de neurones profonds notamment les réseaux de neurones convolutifs, assurent une extraction automatique des caractéristiques. Cette extraction est réalisée dans les premières couches à l'aide des techniques de filtrage.

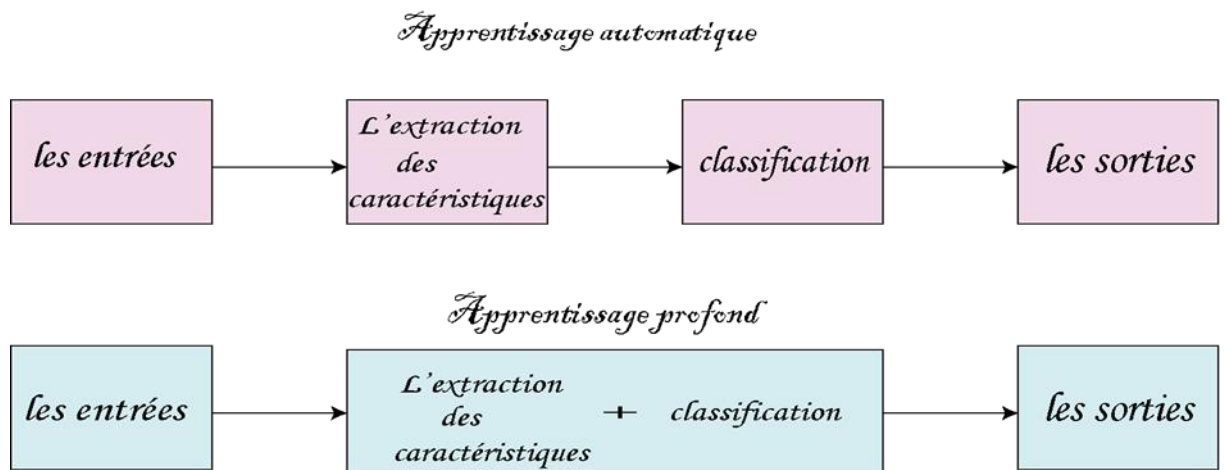


Figure 26 Extraction des caractéristiques par apprentissage automatique et profond

2.12 Différence entre le deep Learning et le machine Learning

La différence entre le deep Learning et le machine Learning [28] se résume dans le tableau 2.2.

	l'apprentissage automatique	apprentissage profond
la taille des datasets	excellente performance sur un petit/moyen ensemble de données	Excellente performance sur un ensemble de données massives
Matériel	peut être implémenté sur une machine bas de gamme.	exige une machine puissante avec GPU car le DL effectue une quantité importante de multiplication matricielle
l'interprétation des algorithmes	Certains algorithmes sont faciles à interpréter (logistique, arbre de décision), certains sont presque impossible (SVM, XGBoost)	Difficile à impossible
les caractéristiques (Features)	Besoin de comprendre les caractéristiques que représentent les données	Pas besoin de comprendre la meilleure caractéristique que Représentent les données
durée d'entraînement du modèle	De quelques minutes à des heures.	Jusqu'à quelques semaines, le réseau neuronal doit calculer un nombre important de poids

Tableau 27 Différence entre le deep Learning et le machine Learning [28]

2.13 Réseaux de neurones récurrents RNN

Il s'agit d'un réseau multicouches dynamique, dont le flux de données se propage dans deux sens : feed forward et backforward. Ces réseaux de neurones sont destinés au traitement des signaux, de taille variable. La notion de temps est extrêmement importante, car le fonctionnement du réseau en représente une évolution dépendante [32, 33].

Les connexions entre les nœuds contiennent des bouclages partiels ou totaux, qui leurs permettent de stocker, d'apprendre et de reproduire les séquences.

Ces réseaux sont souvent utilisés dans le traitement de texte et audio.

2.14 Auto-encodeur

L'auto-encodeur [33, 35] est un algorithme d'apprentissage non supervisé, qui est fondé sur l'utilisation des réseaux de neurones profonds. Leur architecture est combinée de deux parties, l'encodeur et le décodeur :

L'encodeur traite les données, pour construire une nouvelle représentation dite encodée.

Le décodeur : reçoit les présentations précédentes pour reconstruire les données de départ.

L'erreur est mesurée par la différence entre les données initiales et, les données reconstruites.

2.15 Outils de mesures des performances

On utilise les outils de mesure de performance, pour évaluer les résultats d'un algorithme d'apprentissage du deep learning, du machine learning, ou des tests médicaux [36].

2.15.1 Matrice de confusion

Appelée aussi la matrice d'erreur ou tableaux de contingence (tableau 2.4), il s'agit d'un outil de mesure de performances, dédié aux algorithmes de classification binaire, ou à la multiclassification. Cette matrice utilise la technique de comparaison des classes réelles (positive ou négative) et, des classes prédites. Ceci permet non seulement, de mesurer les erreurs commises par le modèle d'apprentissage mais, de révéler 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

Tableau 28 table de confusion

- 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 de type 1, observée quand la prédiction est positive et, la valeur réelle est négative.
- Faux négatif (false negative FN) : il s'agit d'une erreur de type 2, observée quand la prédiction est négative et la valeur réelle est positive.

2.15.2 Critères d'évaluations

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.

$$\textit{spécificité} = \frac{TN}{FP+TN} \quad 2.13$$

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

$$\textit{exactitude} = \frac{TP + TN}{TP + FP + FN + TN} \quad 2.14$$

- **Précision** : elle est le rapport entre les observations positives correctement prédites et le total des observations positives prédites (équation 2.15).

$$\textit{précision} = \frac{Tp}{TP + FP} \quad 2.15$$

- **Rappel (recall)** : appelé aussi sensibilité (sensitivity) est le taux des vrais positifs. Le rappel (équation 2.16) a pour le but d'identifier la proportion du nombre d'échantillons positifs, que le modèle a correctement identifiée.

$$\textit{rappel} = \frac{TP}{TP + FN} \quad 2.16$$

- **F-score** : est la moyenne du rappel et de la précision.

$$F - \textit{score} = 2 * \frac{\textit{precision} * \textit{rappel}}{\textit{precision} + \textit{rappel}} \quad 2.17$$

- **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.

- **Courbe AUC**

Elle représente l'aire sous la courbe Roc, l'augmentation de l'aire désigne l'augmentation de la performance.

2.16 Conclusion

Dans ce chapitre, nous avons passé en revue les principales fonctionnalités de l'intelligence artificielle, en se focalisant sur le fonctionnement, de certains réseaux connexionnistes.

Nous avons mis en avant l'importance des réseaux de neurones profonds dans la l'analyse des signaux et des images. L'un des types dédiés à l'analyse des données non structurées notamment les clichés d'imagerie mammaire, est le réseau de **neurones convolutif** qui est l'objectif du chapitre suivant.

Chapitre 3

Fonctionnalités des CNN dans la classification d'images

« L'âme ne pense jamais sans image » Aristote

3.1 Introduction

L'interprétation médicale des clichés échographiques faite par un radiologue, est basée sur son système visuel et d'autres examens cliniques. Ce fonctionnement peut être émulé par des approches d'intelligence artificielle, grâce aux réseaux de neurones convolutifs. Ces réseaux assurent alors, une extraction automatique des informations pertinentes. Celles-ci sont ensuite classifiées selon des classes bien précises.

Ce chapitre introduit les réseaux de neurones convolutifs, ses différentes couches ses architectures, en se focalisant sur les modèles utilisés, pour la conception de notre système de prédiction du cancer du sein.

3.2 Réseaux de neurones convolutif (CNN)

Un réseau de neurones convolutif est un réseau de neurones composé par une succession de couches, dont au moins une seule, est une « couche de convolution ». Ce type d'architecture est conçu pour extraire automatiquement des caractéristiques

(features) à partir d'un ensemble de données non structurées, afin de reconnaître des objets. Lors de l'apprentissage des réseaux CNN, la hauteur et la largeur du volume d'entrée diminuent grâce à l'application d'une fonction d'activation, sur les résultats des couches de convolution, ou aux résultats des couches de convolution et des couches de Pooling, jusqu'à atteindre « la couche fully-connected » [34, 35, 37, 38, 39].

3.2.1 Couches d'un réseau de neurones convolutif

On distingue plusieurs types de couches utilisées [38, 39], pour construire un CNN mais les plus pertinentes sont :

- ✓ La couche de convolution.
- ✓ La couche de Pooling.
- ✓ La couche d'activation.
- ✓ La couche flatten.
- ✓ La couche Fully Connected.
- ✓ Dropout.

Dans les sections qui suivent, sont décrites les principales opérations des CNN.

3.2.2 Opération de convolution

La convolution (équation 3.1) est une transformation linéaire (figure 3.1), qui permet de remplacer la valeur de chaque pixel en entrée par une combinaison linéaire, des intensités de ses voisins avec un noyau (kernel).

✓ Définition d'un noyau

Le noyau (Kernel (H)), est un masque ou noyau de convolution, représenté par une fenêtre carrée de taille fixe impaire $2*k+1$.

$$Y=H*X \tag{3.1}$$

Avec,

* : le produit de convolution
Y : l'image filtrée

H : le masque
 X : l'image originale
 La taille de l'image filtrée est $y = k - H + 1$

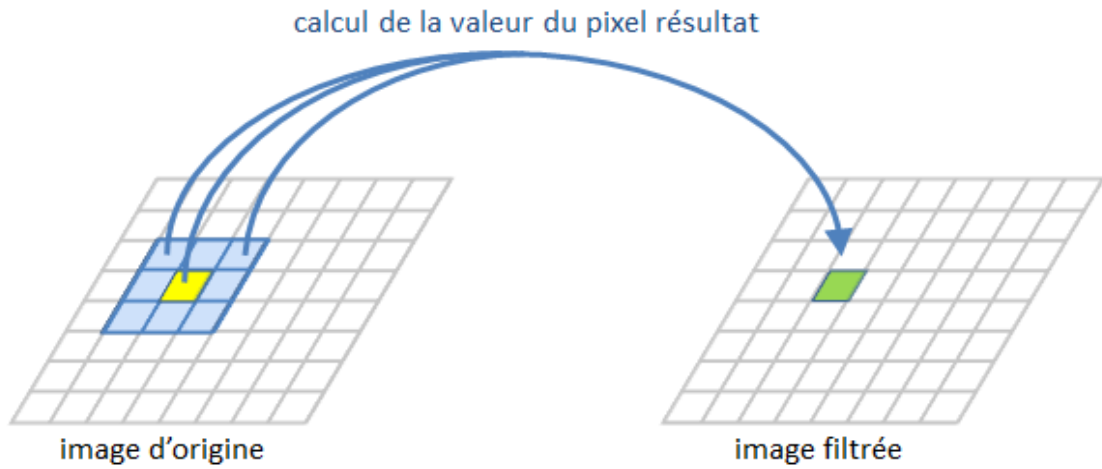


Figure 29 : Exemple d'une convolution [35]

3.2.3 Couche de convolution (conv)

La couche conv est la base d'un réseau de neurones convolutif [38, 39]. Dans cette couche, les données se propagent dans un seul sens (Feed Forward). Son but est de repérer toutes les caractéristiques (features), en appliquant des filtrages par convolution.

A cet effet, l'image d'entrée est de taille $W * H * D$.

Avec,

W : la largeur en pixel
 H : la hauteur en pixel
 D : le nombre des canaux

Après avoir effectué l'opération de convolution, la nouvelle taille d'image est $W_c * H_c * D_c$.

Avec,

$$W_c = \frac{w - F + 2P}{s} + 1 \tag{3.2}$$

$$H_c = \frac{H - F + 2p}{s} + 1 \tag{3.3}$$

K : nombre de filtres

F : taille du filtre $F \times F \times D$

Les paramètres assurant les opérations des couches de convolution sont indiqués par :

➤ **La profondeur**

C'est le nombre de noyaux à appliquer pendant la convolution.

➤ **Stride (s)**

C'est le pas de saut de pixels entre deux zones, dans une image. Par exemple si le stride est égale à 1 ($S=1$), le filtre se déplace d'un pixel ou d'une unité à la fois.

➤ **Zero padding (p)**

Cette technique permet d'ajouter des lignes et des colonnes de pixels nuls, aux bordures des images originales, afin de conserver les tailles de celles-ci, lors de l'application de la convolution. Cette opération est effectuée pour résoudre le problème de la diminution, de la taille de l'image d'entrée.

3.2.4 Couche de pooling

C'est une couche facultative. Il s'agit de la deuxième méthode utilisée pour réduire la taille d'une carte des caractéristiques (features map). Dans le cas du maxpooling, le principe est de découper l'image en plusieurs carrés, pour retenir le maximum correspondant à chacun de ces carrés et, réduire la complexité du calcul des poids

3.2.5 Couche d'activation

Dans cette couche une fonction d'activation non linéaire « Relu », est appliquée aux résultats de chacune des couches de convolution. L'objectif est d'obtenir les valeurs $W_c \times H_c \times D_c$ seuils, à la sortie de cette couche en gardant les mêmes dimensions d'avant.

3.2.6 Couches Fully Connected (Fc)

Les couches Fully Connected doivent être placées à la fin d'un réseau de neurones convolutif, après avoir effectué toutes les opérations de convolution, de pooling et

d'activation. Les sorties des couches d'activation sont ainsi, connectées à la première couche de Fully connected.

L'utilisation de plusieurs couches FC, permet d'augmenter la puissance de calcul.

Les FC fonctionnent exactement comme les réseaux de neurones multicouches, avec un grand nombre de nœuds et d'arcs. De plus, dans la dernière couche, les données se propagent en Feed forward et Back forward, afin d'ajuster les poids, pour activer les sorties grâce à la fonction softmax ou sigmoïde.

3.2.7 Couche flatten

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

3.2.8 Couche Dropout

Le **décrochage**, ou l'**abandon** est une couche de régularisation utilisée, afin de réduire le surajustement en diminuant les connexions, entre les neurones de deux couches du 'fully connected' avec une probabilité aléatoire.

3.3 Diverses architectures des réseaux de neurones convolutif

3.3.1 Imagenet et la compétition « large scale visual recognition challenge »

Imagenet est une base de données open source qui contient environ 15 millions images avec 1000 différentes classes, dédiée aux travaux de vision par ordinateur [40].

Imagenet organise une compétition annuelle nommée <large scale visual recognition challenge>, depuis 2010, afin de choisir le meilleur modèle de classification d'images et de détection d'objets. En effet, cette compétition fait apparaître plusieurs architectures des CNN, qu'on appelle aujourd'hui les modèles pré-entraînés.

3.3.2 LeNet

LeNet est la première architecture qui est apparue en 1998 dans les travaux de Yann LeCun et al [41]. L'objectif principal de cette étude, était de construire un modèle de reconnaissance optique des caractères.

LeNet est une architecture simple (figure 2.2), destinée aux images noir et blanc de taille 28*28. La fonction d'activation utilisée est la fonction tangente hyperbolique.

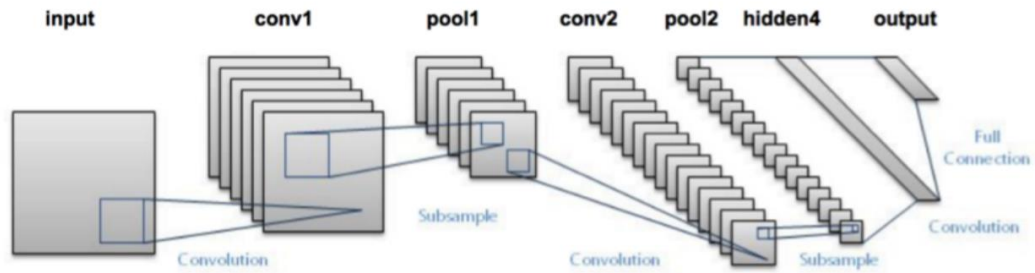


Figure 30 Architecture LeNet [41]

Le tableau 3.1 résume l'architecture « LeNet ».

Les différents types de couches	la taille de la sortie	la taille de filtre
image d'entrée	28*28*1	
la couche convolution	28*28*20	5*5,k=20
La couche d'activation	28*28*20	
La couche pooling	14*14*20	2*2
la couche convolution	14*14*50	5*5,k=50
La couche d'activation	14*14*50	
la couche pooling	7*7*50	2*2
Fully Connected	500	
La couche d'activation	500	
Fully Connected	10	
SoftMax	10	

Tableau 31 Architecture LeNet [41]

3.3.3 Alex-NET

Alex-NET en hommage à Alex Krizhevsky, [41, 42], le premier auteur du document de la classification révolutionnaire d'ImageNet. Ce réseau (figure 3.3) de neurones a été le socle de 'Scale Visual Recognition Challenge 2012'.

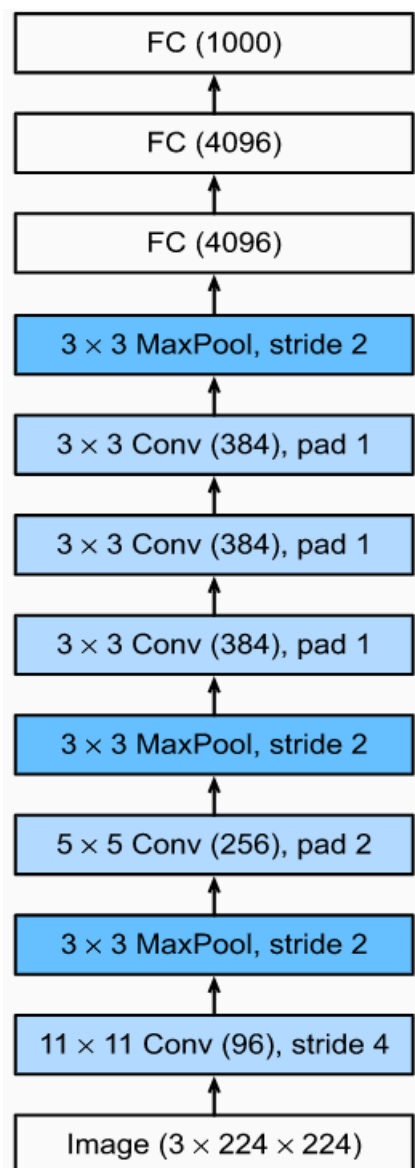


Figure 32 Architecture LeNet [42]

Le modèle d'Alex-net est composé de 8 couches de convolution. Les cinq premières couches utilisent une opération de maxpooling, par une fenêtre de taille 3*3 avec un pas de 2.

Alex-net comprend deux couches fully connected, de taille 4096. Chaque couche est suivie par la fonction d'activation Relu. La troisième couche, qui est celle de la sortie, est activée par la fonction softmax. Elle génère un vecteur de 1000 classes.

Les tailles des filtres utilisés diffèrent d'une couche à une autre. Ainsi, la taille du filtre de la première couche est de 11*11, celle de la deuxième couche est de 5*5 et celle des sept couches restantes est de 3*3.

3.3.4 Modèles VGG

Il existe plusieurs types d'architectures de VGG [41, 43] de tailles différentes, dont le nombre de couches est compris entre 13 et 19 couches avec une architecture spéciale, qui réduit la taille du filtre et augmente sa profondeur. Le VGG a été introduit par « Karen Simonyan » et « Andrew Zisserman » du groupe « Visual Geometry » du laboratoire de l'université d'Oxford, en 2014 lors de la compétition « l'ISLVRRC ».

On distingue plusieurs configurations du VGG (tableau 3.2). Les plus utilisées sont VGG19 et VGG16 (figure 3.4). A cet effet, le VGG16 est composé de 13 couches convolutives, alors que le VGG19, est composé de 16 couches convolutives. Dans ce cas, les couches de convolution sont reliées par la fonction d'activation ReLU, qui assure une linéarisation.

- ✓ Les images d'entrée doivent être de dimensions de (244*244*3).
- ✓ Le filtre utilisé dans les couches convolutives, est de taille 3*3 avec un pas d'un pixel.
- ✓ L'opération du pooling est réalisée par une fenêtre de taille 2*2.

Les VGG16 et 19 sont composés de trois couches Fully Connected. La première couche représente un vecteur de taille 4096, suivie par la fonction d'activation ReLU. La deuxième couche est similaire à la première. La dernière couche désigne un vecteur de classes.

Le nombre maximal de classes est 1000 classes, à chacune d'elles est attribuée une probabilité. La fonction d'activation softmax assure la distribution des probabilités.

les divers configurations					
A	A-LRN	B	B	D	E
11 COUCHES	11 COUCHES	13 COUCHES	16 COUCHES	16 COUCHES	19 COUCHES

Tableau 33 Différentes configurations de VGG [43]

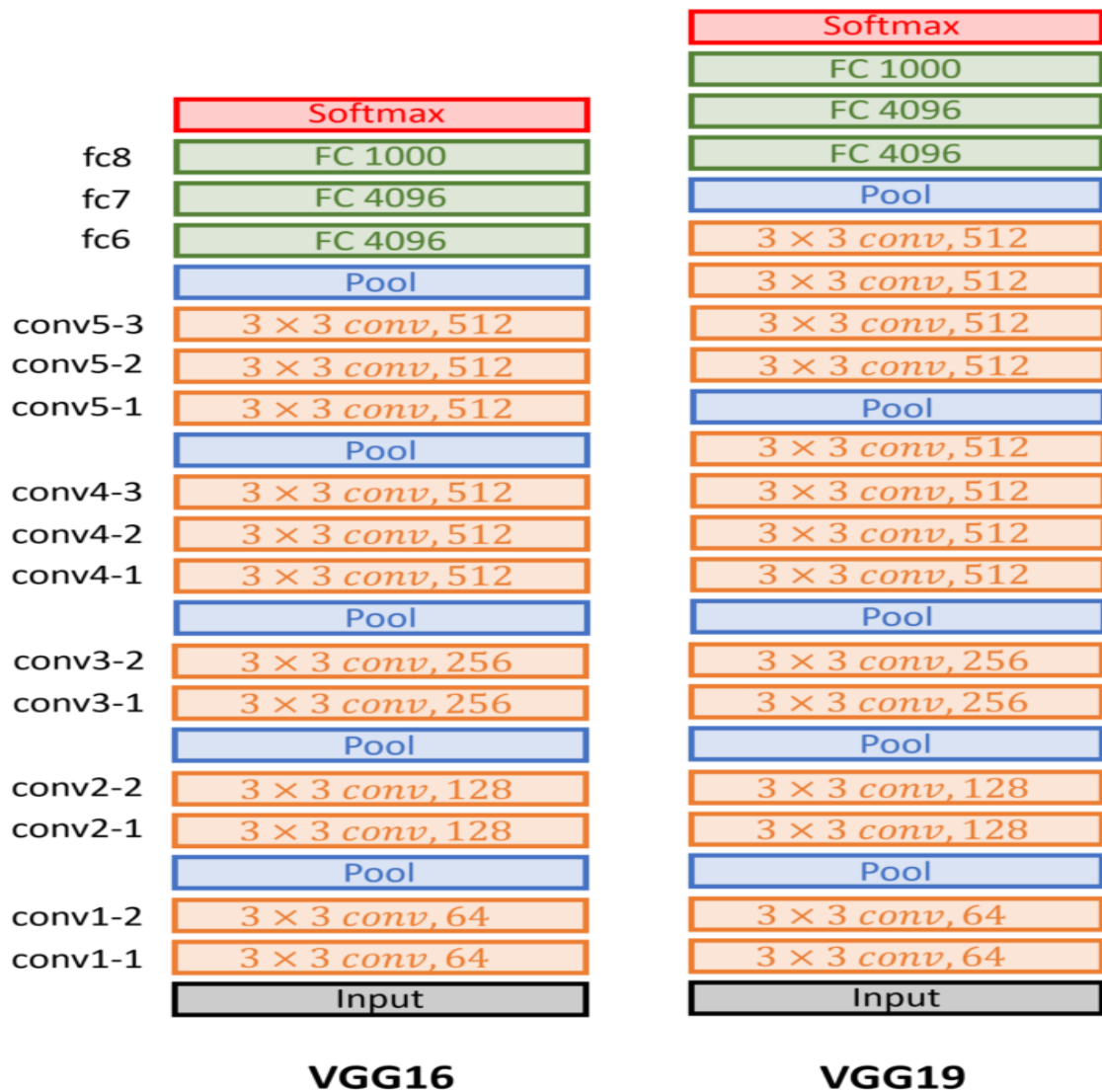


Figure 34 Architecture de VGG16 et VGG19 [41]

3.3.5 RESNET

Le modèle Resnet (residual neural network) (figure 3.5) est le vainqueur de la compétition « ILSVRC 2015 » avec un taux d’erreur de 3.6% [41]. Ceci lui confère un niveau de performance, similaire à celui d’un être humain dans la classification des images.

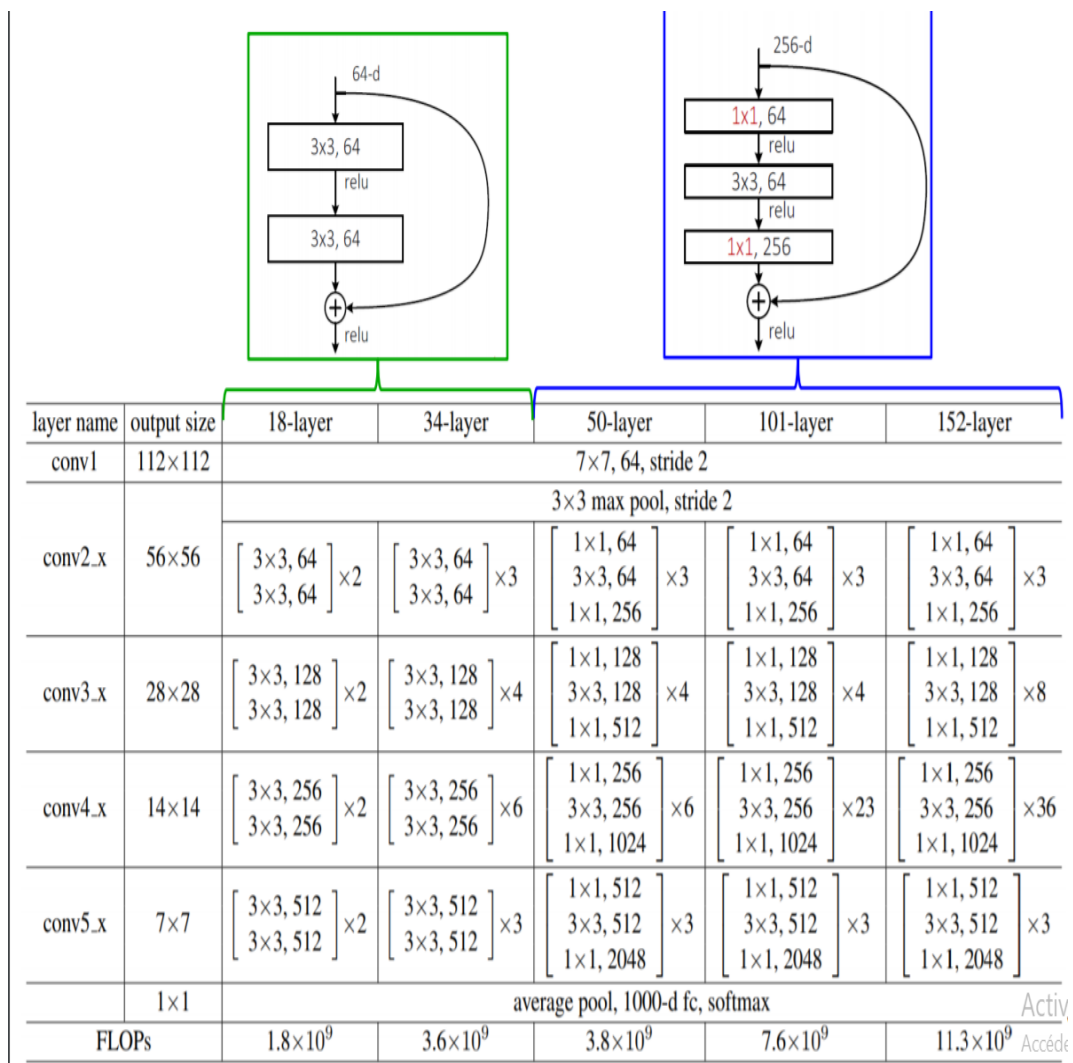


Figure 35 Architecture Resnet [41]

Ce réseau est composé de 50 jusqu’à 152 couches avec une taille des filtres adoptés de 3*.3. Cependant, l’entraînement de ce réseau de neurones avec un très grand nombre

de couches, génère certains problèmes ; tels que celui du flux et de la disparition du gradient. A cet effet, une nouvelle technique appelée 'ignorer les connexions' (skip connections), est introduite dans les réseaux Resnet pour réduire ces problèmes et réaliser un apprentissage rapide.

➤ **Principe de skip connections**

Cette technique consiste à sauter la connexion résiduelle. Son principe est d'ajouter à l'entrée de la couche en amont (x) la somme pondérée de la couche en aval, avant que la fonction d'activation ne soit appliquée $f(x)-x$ (figure 3.6).

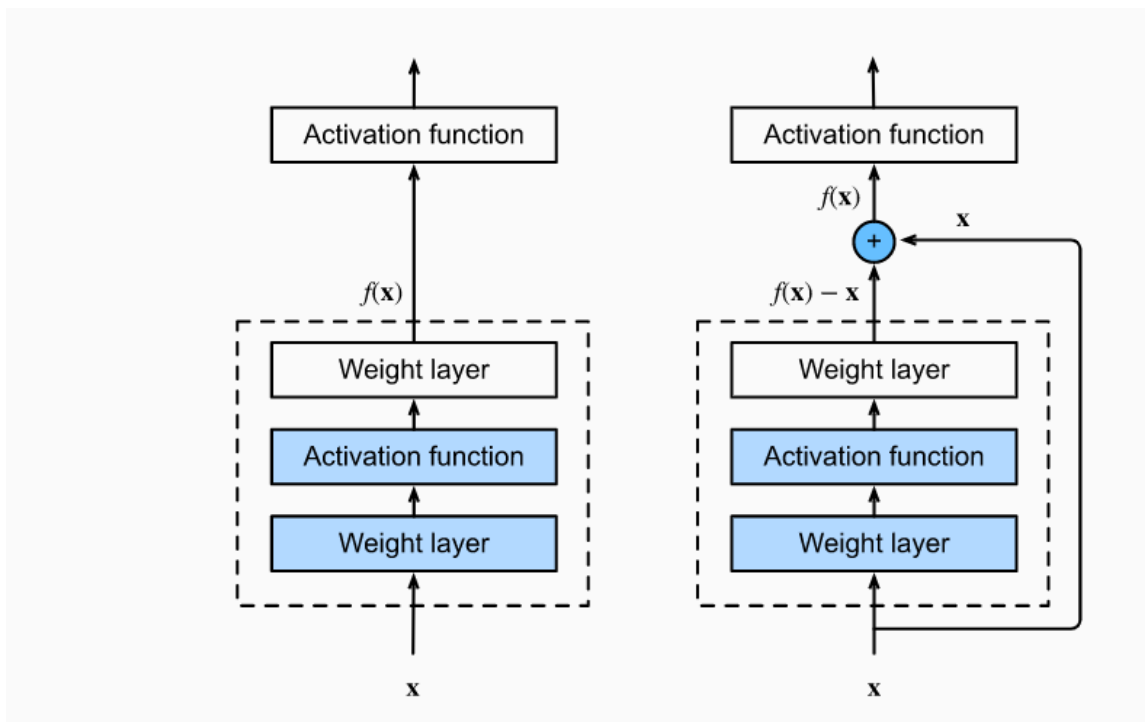


Figure 36 Modèle résiduel [41]

3.3.6 Xception

Xception est une architecture inspirée par Inception V3, Resnet et VGG16. Elle a été développée par les chercheurs de Google [41, 44]. Il implique des convolutions séparables, en profondeur modifiée.

➤ **Convolution originale séparable en profondeur**

La convolution séparable en profondeur originale (figure 3.7), est la convolution en profondeur suivie d'une convolution ponctuelle.

- La convolution en profondeur est représentée par des convolutions spatiales alternatives aux convolutions normales. Elle est plus rapide et adopte une taille des nombres de canaux de $n \times n$.
- La convolution ponctuelle est une convolution classique de taille $1 \times 1 \times N$ appliquée sur une image, pour obtenir une nouvelle dimension de $K \times K \times N$.

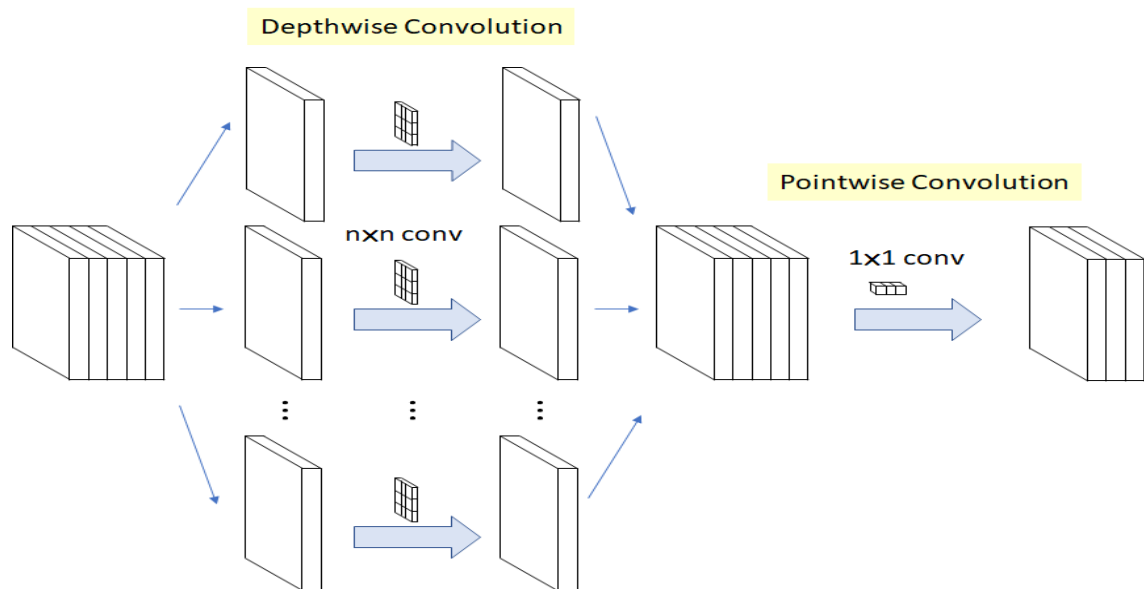


Figure 37 Convolution originale séparable en profondeur [41]

➤ **Convolution modifiée séparable en profondeur**

Il s'agit d'une convolution ponctuelle suivie d'une convolution en profondeur (figure 3.8).

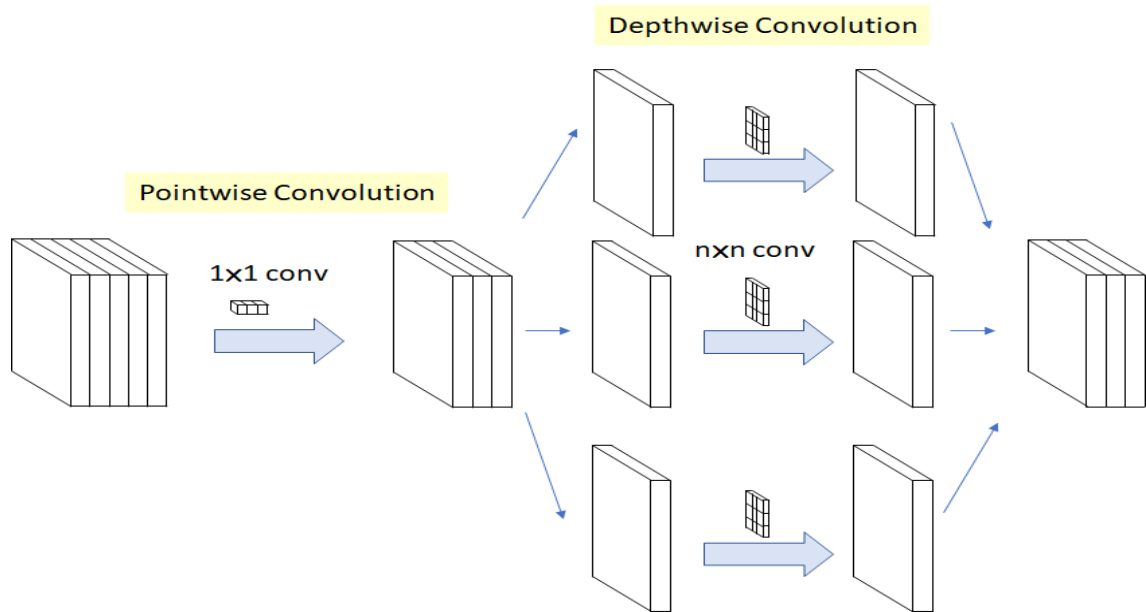


Figure 38 Convolution modifiée séparable en profondeur [41]

3.4 Architecture globale de Xception

Dans le cas de cette architecture (figure 3.9), l'ensemble des données passe préalablement par le flux d'entrée, puis par le flux du milieu qui est répété huit fois et, enfin par le flux de sortie. Toutes les couches de convolution classique et de convolution séparable, sont suivies d'une normalisation par lots [41, 44].

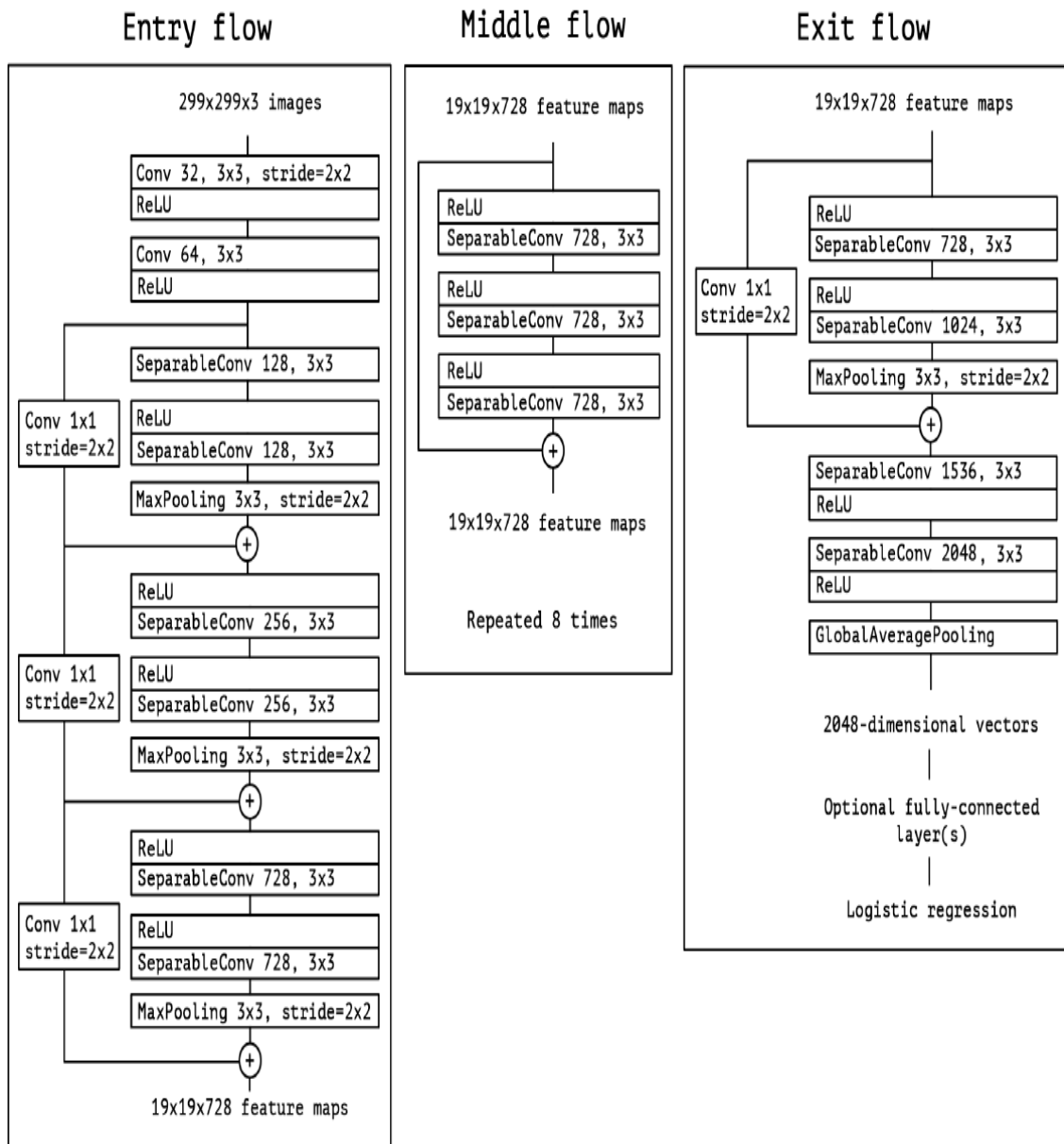


Figure 39 Architecture de Xception [44]

3.5 Comparaison entre les différentes architectures

Le tableau suivant révèle une comparaison entre les différentes architectures Cnn, pré-entraînés sur Keras [45]. Le top 5 d'erreurs générés par certains modèles, est indiqué dans le tableau 3.4.

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

Tableau 40 Paramètres des différentes architectures [45]

de l'architecture	Année	Nombre de couches	Top 5 d'erreurs
-	Avant 2012	≤ 5	25
AlexNet	2012	8	15.4%
ZfNet/Clarifai	2013	8/>8	14.8% / 11.1%
VGG	2014	19	7.3%
GoogLeNet	2014	22	6.7%
ResNet	2015	152	3.6%

Tableau 41 top 5 d'erreurs [45]

3.6 Augmentation des données

L'augmentation des données (data augmentation) [35] est une technique utilisée, pour générer de nouvelles données à partir des données d'origine, sans changer les étiquettes des images originales. Le but de cette opération, est d'augmenter la performance du modèle en utilisant plusieurs transformations géométriques sur les images d'origine telles que :

- La rotation
- Le changement d'échelle
- La translation
- Le zoom

3.7 Détection d'objets par les réseaux de neurones convolutifs

Les applications des CNN ne se limitent pas à la classification des images. En effet, les réseaux convolutifs ont aussi, un rôle majeur dans la détection des objets et leur localisation au sein des images. On distingue plusieurs algorithmes de détection d'objets dont, l'architecture la plus sophistiquée est YOLO [46].

Yolo est l'abréviation de **You Only Look Once**. La signification de cette nomination en français est, 'vous ne regardez qu'une seule fois', c'est-à-dire que le modèle parcourt l'image une seule fois. Cet algorithme a été développé par Joseph Redmond et al [46] de l'Université de Washington en 2015.

YOLO est un algorithme très rapide (figure 3.10), qui traite les images en temps réel, en passant l'image une seule fois dans le réseau de neurones, pour obtenir la localisation de l'objet.

Sur un **Pascal Titan X**, Yolo traite les images à 30 FPS (image par seconde) avec une précision moyenne de 57.9%.

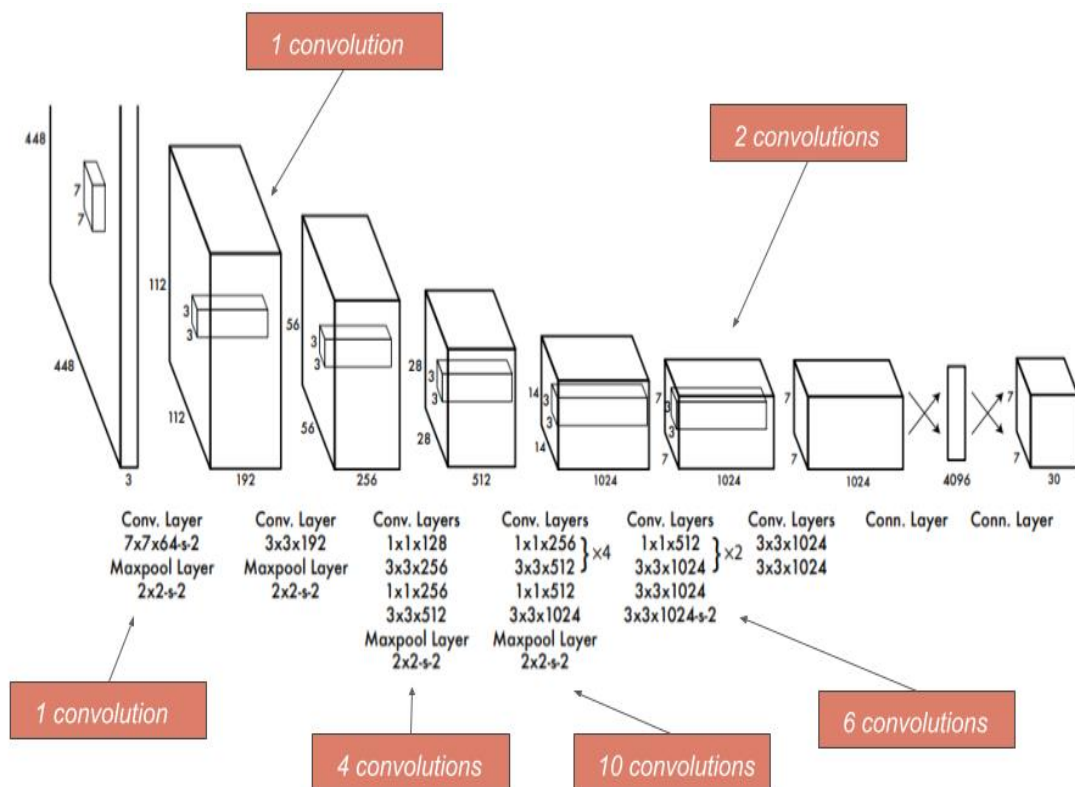


Figure 42 Architecture YOLO [46]

3.8 Apprentissage profond par transfert

L'apprentissage profond par transfert (deep transfer learning) est une technique, qui permet d'entraîner un réseau de neurones dans un temps restreint, sans avoir besoin d'un processeur graphique (GPU) très sophistiqué [47].

L'apprentissage profond par transfert, par permet d'éviter le sur-apprentissage. Cette technique est basée sur des réseaux de neurones pré- entraînés, dont l'architecture est fixe et les poids calculés. La technique d'apprentissage par transfert utilisé dans le cadre de la réalisation de notre projet, est basée sur le '**fine-tuning**'. Le concept de cette technique est d'utiliser des architectures telles que VGG, RESNET déjà entraînés sur Image net en gardant les même couches convolutives, tout en remplaçant la dernière couche, par un nouveau classifieur adapté à notre problème.

3.9 Optimiseurs

Considérés comme des fonctions d'optimisation, les optimiseurs sont utilisés pour minimiser l'erreur par l'ajustement du poids. On distingue plusieurs algorithmes tels que ADAM.

➤ ADAM

Ce paramètre est basé sur l'estimation du moment adaptatif, pour l'optimisation du taux d'apprentissage (adaptive learning rate optimization algorithm) [33]. Il utilise les gradients carrés, pour une mise à l'échelle du taux d'apprentissage. Les différentes étapes de ce calcul sont indiquées ci-dessous :

$$m_{ij} = \beta_1 m_{ij} + (1 - \beta_1) \frac{\partial u}{\partial w_{ij}} \quad 3.5$$

$$g_{ij} = \beta_2 g_{ij} + (1 - \beta_2) \left(\frac{\partial u}{\partial w_{ij}} \right)^2 \quad 3.6$$

Où,

m_{ij} : est le taux de décroissance

g_{ij} : est le gradient pour chaque poids w_{ij}

β_1 et β_2 sont des constantes < 1

Avec,

m : est la moyenne normalisé des gradients

v : est la moyenne des gradients carrées

$$\hat{m}_{ij} = \frac{m}{1 - \beta_1} \quad 3.7$$

$$\hat{v}_{ij} = \frac{v}{1 - \beta_2} \quad 3.8$$

La règle finale pour chaque poids w_{ij} est :

$$W_{ij} = w_{ij} - \frac{\eta}{\sqrt{\hat{v}_{ij} + \epsilon}} \hat{m}_{ij} \quad 3.10$$

Où η est le taux d'apprentissage

3.10 Travaux en imagerie échographique

A l'heure actuelle, l'échographie mammaire est devenue parmi les examens les plus utilisés dans le dépistage et le diagnostic du cancer du sein grâce à son faible coût, son affichage en temps réel et surtout, sa sécurité. En effet, elle n'entraîne aucun risque sur la santé de la patiente. Néanmoins la détection des tumeurs mammaires dans les clichés échographiques de nature bruitée est fastidieuse, ce qui fait appel à l'apprentissage profond, pour ses performances. Plusieurs travaux dans la classification et la détection des pathologies mammaires, ont été réalisés dans ce cadre.

L'étude de Al-Dhabyani et al [48] a été consacrée à la classification des masses, en échographie mammaire. Afin d'agrandir la base des images, les auteurs, ont utilisé l'application de deux méthodes d'augmentation des données : la méthode classique et la méthode engendrée par les réseaux génératifs (generative adversarial network (GAN)), sur la base BUSI [49]. La classification des pathologies échographiques, est réalisée par les réseaux CNN pré-entraînés, avant et après l'augmentation des données. La meilleure précision de 99% est obtenue par le modèle NASNET, grâce à la combinaison des deux méthodes d'augmentation.

La recherche de Cao et al [50] réalisée en 2019, permet la détection et la classification des tumeurs mammaires. Le principe adopté est divisé en deux étapes : la localisation des régions d'intérêt (ROI) et, la classification de ces lésions en classes bénignes ou malignes. Plusieurs types de réseaux de neurones convolutifs (CNN) ont été testés, dans ce cadre. Pour la tâche de détection des lésions, un détecteur multiboîtes avec une taille d'entrée de 300×300 (SSD300) réalise les meilleures performances en termes de taux de précision moyen (APR), de taux de rappel moyen (ARR) et de F1 score. Pour la tâche de classification, DenseNet est plus adapté aux cas étudiés.

Dans l'article de Yap et al [51], une détection automatique des liaisons échographiques mammaires suivants plusieurs architecture CNN (Lenet, UNET et Alexnet) a été réalisée sur deux types de bases d'images. Le meilleur résultat est obtenu par l'hybridation des deux ensembles de données.

Le travail de Nayeem et al [52] est consacré la détection automatique des informations pertinentes suivant les SVM, avec une combinaison optimale de 25 caractéristiques différentes pour différencier entre les tumeurs malignes et bénignes. Les performances obtenues ont généré une AUC égale à 93,1%.

3.11 Conclusion

Ce chapitre nous a permis de comprendre les réseaux de neurones convolutifs, le fonctionnement de ses diverses couches et leur impact majeur, dans le domaine de la vision par ordinateur et la classification des images.

Le chapitre qui suit est consacré à la mise en œuvre des modèles convolutifs choisis, pour la classification de données échographiques du sein.

4.1 Introduction

Ce chapitre est dédié à la réalisation de notre approche, qui permet la classification des lésions échographiques mammaires, sur deux types de bases de données. Le concept de cette étude est d'utiliser différentes architectures, des réseaux de neurones convolutifs (VGG16, VGG19, ALEXNET, XCEPTION) avec apprentissage par transfert.

Une étude comparative est effectuée entre les modèles implémentés, afin de choisir le plus adapté pour la prédiction du cancer du sein et, de l'intégrer dans une interface nommée 'BreastCNNPredict'.

4.2 Environnement de travail

L'hybridation hardware software nous permet la réalisation de notre interface « BreastCNNPredict », avec le langage python sous environnement anaconda.

4.2.1 Hardware

Ce travail a été réalisé avec un ordinateur personnel avec les caractéristiques suivantes :

- Marque du PC : DELL inspiron 5000
- Processeur : Intel® core™ CPU @ 1.60 GHZ 1.80 GHZ
- Ram : 8.00 G
- Carte graphique : Radeon
- Disque dur : 1000 GB

4.2.2 Software (Logiciel)

L'approche 'BreastCNNPredict' réalisée sous le système d'exploitation « Windows 10 », est développée avec le langage Python [53] sous l'environnement Anaconda, à l'aide d'un ensemble de bibliothèques open source.

a Anaconda

Anaconda est un environnement virtuel de développement et d'accès aux packages, permettant l'installation de plusieurs logiciels et bibliothèques [54].

b Spyder (Scientific Python Development Environment)

C'est un environnement de développement pour python, composé d'un éditeur pour écrire des codes, une console pour saisir les données et voir les résultats [55].

c Python

C'est un langage de programmation open source, interprété, très puissant créé en 1991 par Guido Van Rossum [53]. De nos jours, python est le langage le plus utilisé dans la science des données, l'apprentissage profond et l'apprentissage automatique.

d Différentes bibliothèques utilisées

- TensorFlow

TensorFlow est la bibliothèque du deep learning. Développée par Google Brain Team pour accélérer l'apprentissage des algorithmes de l'apprentissage.

Le terme 'TensorFlow' est la combinaison de deux mots ; le mot Tensor (un tableau multidimensionnel) et le mot Flow (un graphe). Le concept de cette bibliothèque est donc, de faire subir un ensemble de tenseurs à des opérations relatives à des calculs numériques à haute performance [56].

- Keras

C'est une bibliothèque intégrée par Tensor Flow, CNTK ou Theano. Keras est un API de réseaux de neurones de haut niveau, car cette bibliothèque assure une implémentation facile, des réseaux de neurones convolutifs et récurrents ainsi que, des optimiseurs et des fonctions d'activations [45].

Keras offre aussi la possibilité d'enregistrement des modèles, pour effectuer les tests.

- Numpy

C'est une bibliothèque dédiée aux calculs scientifiques, mathématiques, à l'ingénierie et aux sciences des données. Numpy traite les données sous forme d'un tableau multidimensionnel, ou sous la forme d'une matrice [58].

- Scikit-learn

C'est une bibliothèque destinée à l'apprentissage automatique, comme les KNN, ou les forêts aléatoires. Scikit-learn est efficace pour l'exploration et l'analyse des données, ainsi que l'évaluation des performances des modèles entraînés [59].

- Splitfolders

Cette bibliothèque sert à diviser automatiquement, la base de données [60].

- Matplotlib pyplot

C'est une bibliothèque graphique de python, qui permet de visualiser et tracer les données sous forme de graphes [60].

- OpenCV (open for computer vision)

Il s'agit d'une bibliothèque libre développée par Intel dédié au traitement d'image en temps réel [60].

- Tkinter (Tool kit interface)

Tkinter est une bibliothèque graphique de création des interfaces [60].

4.3 Présentation du système développé

Le système de prédiction des lésions échographiques mammaires 'BreastCNNPredict' (figure 4.1), est réalisé en adoptant des réseaux de neurones convolutifs pré-entraînés, suivant quatre modèles différents (VGG16, VGG19, ALEXNET, XCEPTION).

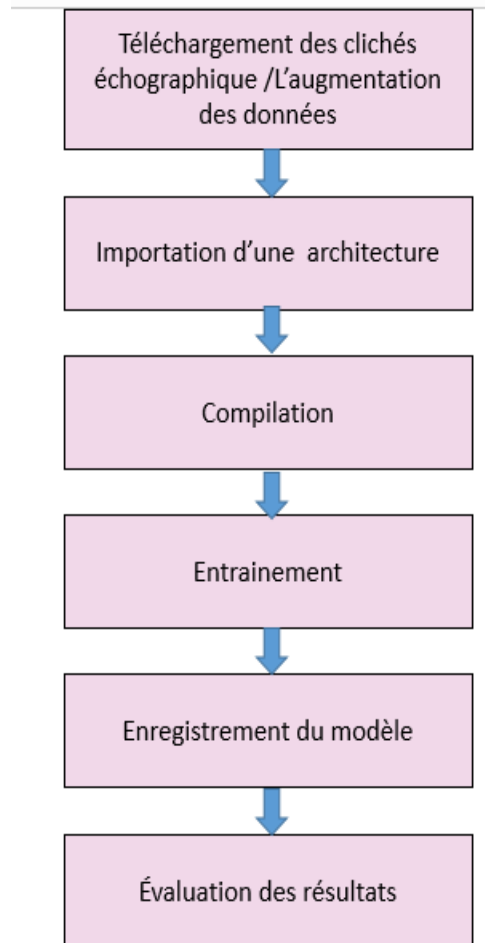


Figure 43 : Synoptique de BreastCNNPredict

4.4 Bases de données

L'approche développée, est appliquée sur deux bases de données échographiques mammaires.

4.4.1 Dataset BUSI

Cette base de données datant de 2018, a été collectée à l'hôpital de Baheya [49] en Égypte. Elle comprend des clichés mammaires de 600 patientes, âgées entre 25 ans et 75 ans. BUSI est constituée de 780 clichés, de taille moyenne de 500 *500 pixels, segmentés et fournis avec des masques. Les images échographiques de BUSI, sont réparties en 133 classes normales, 210 classes bénignes et 437 classes malignes (figures 4.2 et 4.3).

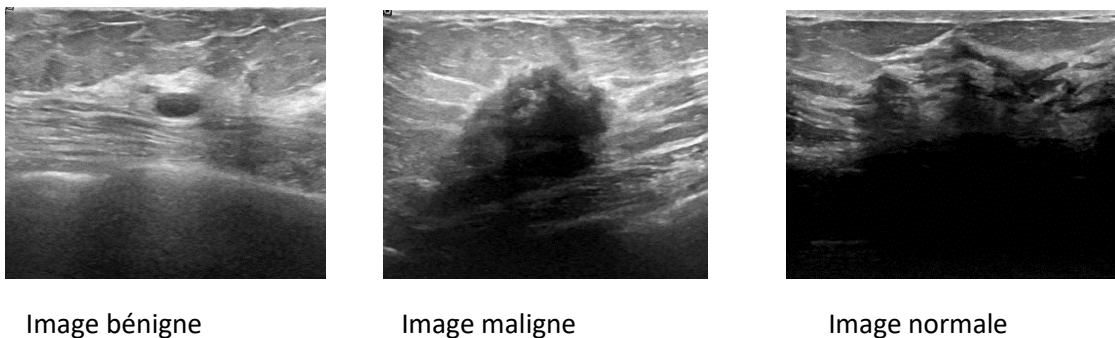


Figure 44 : Images de la base BUSI [2]

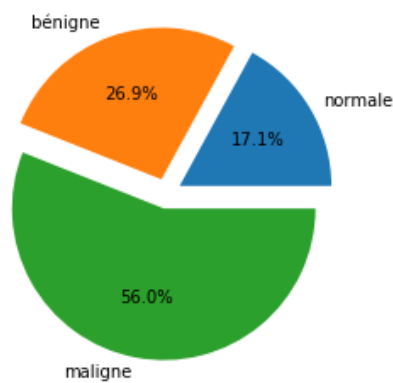


Figure 45 Répartition des classes pour BUSI

➤ Analyse statistique de BUSI Dataset

Les images échographiques de la base BUSI, présentent une certaine répartition des pixels, en fonction des niveaux de gris. Nous illustrons cette répartition par les graphes 4.4, 4.5 et 4.6.

✓ **Cas d'une image bénigne**

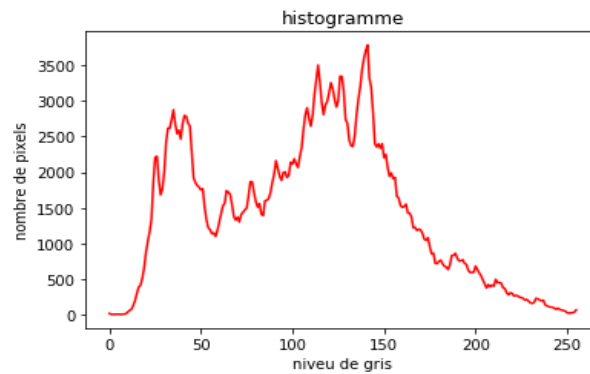


Figure 46 Répartition des pixels d'une image bénigne de BUSI

Nous remarquons que l'image de taille 585*617 pixels, est fortement contrastée.

✓ **Cas d'une image maligne**

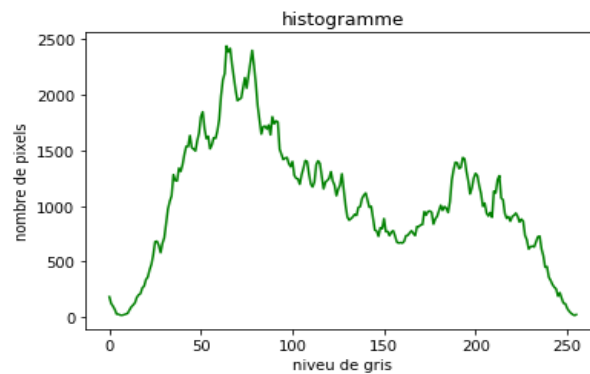


Figure 47 Répartition des pixels d'une image maligne de BUSI

Nous remarquons que l'image de taille 598*449 pixels, est fortement contrastée.

✓ **Cas d'une image normale**

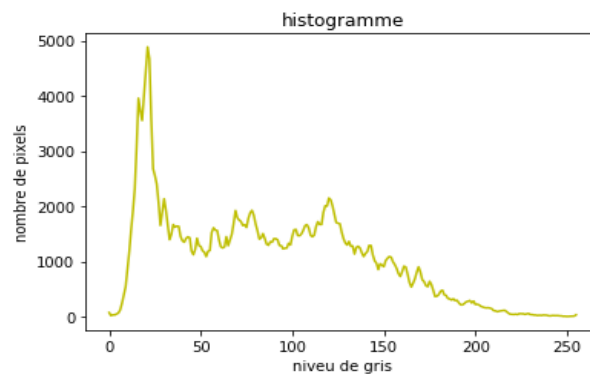


Figure 48 Répartition des pixels d'une image maligne de BUSI

Le cliché échographique normal de taille de 470*568 pixels, présente sur le graphe ci-dessus, un fort contraste. Cette qualité d'image, est présente dans toute la base BUSI.

4.4.2 US dataset

Cette base de données est fournie par le centre de diagnostic UDIAT de la société Parc Tauli, à Sabadell (Espagne) en 2010 [61]. Le nombre global des clichés échographiques est de 163 clichés, dont 53 sont des classes bénignes et 110 désignent un cancer du sein (figures 4.7 et 4.8).

La taille moyenne d'une image de la base US, est de 760*570 pixels.

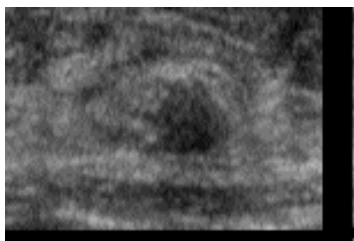


Image bénigne



Image maligne

Figure 49 Clichés échographiques de la base US

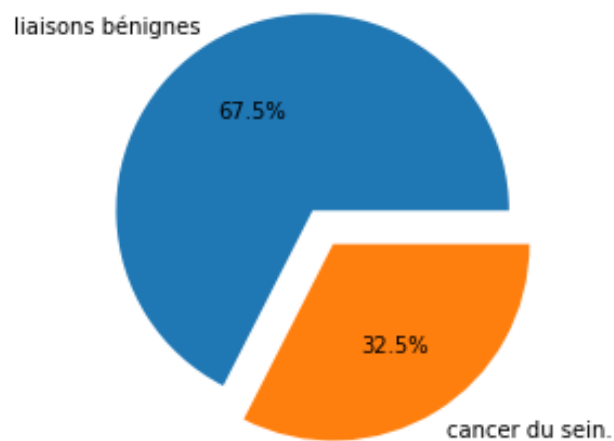


Figure 50 Répartition des classes pour US

➤ **Analyse statistique de US Dataset**

✓ **Cas d'une image bénigne**

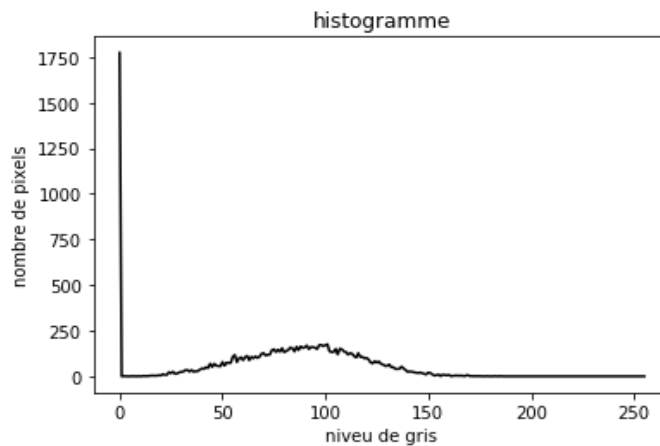


Figure 51 Répartition des pixels en fonction des niveaux de gris d'un cliché bénin
L'image considérée dans ce cas, de taille 87*153pixels, est faiblement contrastée.

✓ **Cas d'une image maligne**

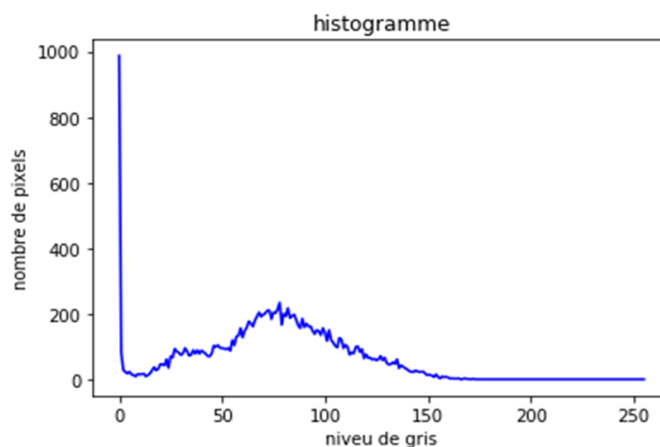


Figure 52 Répartition des pixels en fonction des niveaux de gris d'un cliché malin
L'image de taille 113*137 pixels, présente un faible contraste. Toutes les images de la base US, sont faiblement contrastées.

4.5 Augmentation des données

L'augmentation des données, est faite sous Keras en utilisant la classe « datagenerator ». Cette classe permet d'effectuer plusieurs transformations sur les images, afin d'augmenter le nombre de données sans perdre les caractéristiques.

Dans le cas de notre approche, nous avons utilisé les opérations suivantes :

- `brightness_range` : cette fonction permet de modifier la luminosité des images.
- `rotation_range` : la fonction 'rotation range' permet de réaliser une rotation de 0 à 180 degrés sur une image.
- `width_shift_range` : il s'agit d'une translation verticale des images.
- `height_shift_range` : c'est une translation horizontale des images.
- `fill_mode` : est une technique de remplissage, utilisée pour les nouveaux pixels apparus après une translation ou une rotation.
- `Rescale` : il s'agit du changement d'échelle.
- `horizontal_flip` : cette commande permet d'effectuer une rotation horizontale d'une image.

4.5.1 Transformations appliquées sur la base BUSI

Les différentes transformations appliquées sur la base BUSI, pour l'augmentation des données sont représentées par la figure 4.11.

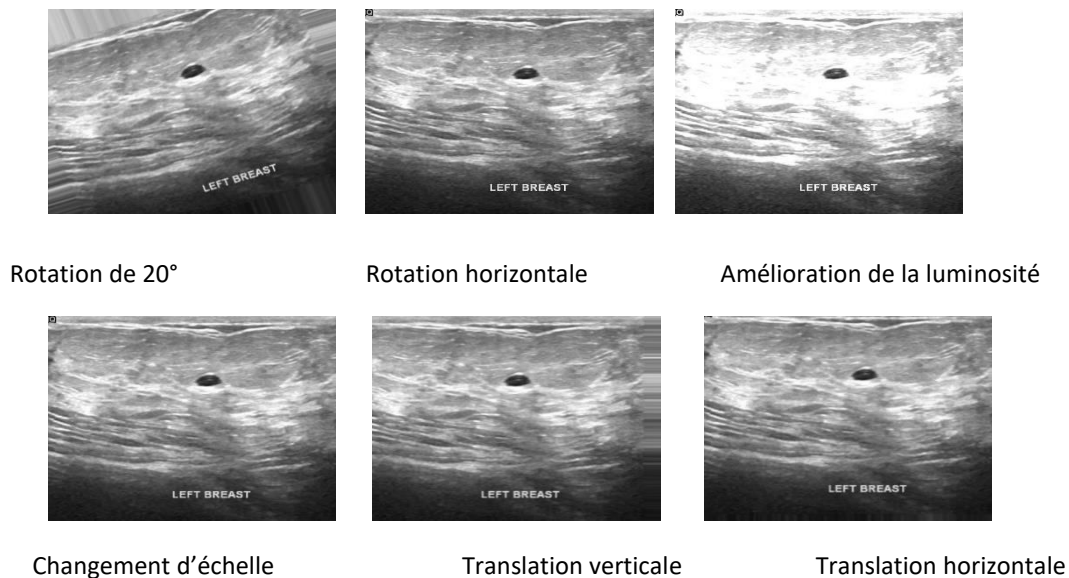


Figure 53 Exemple d'augmentation des données de la base BUSI

4.5.2 Augmentation des données de la base US

Les opérations réalisées pour l'augmentation des données de la base US, sont illustrées par la figure 4.12.

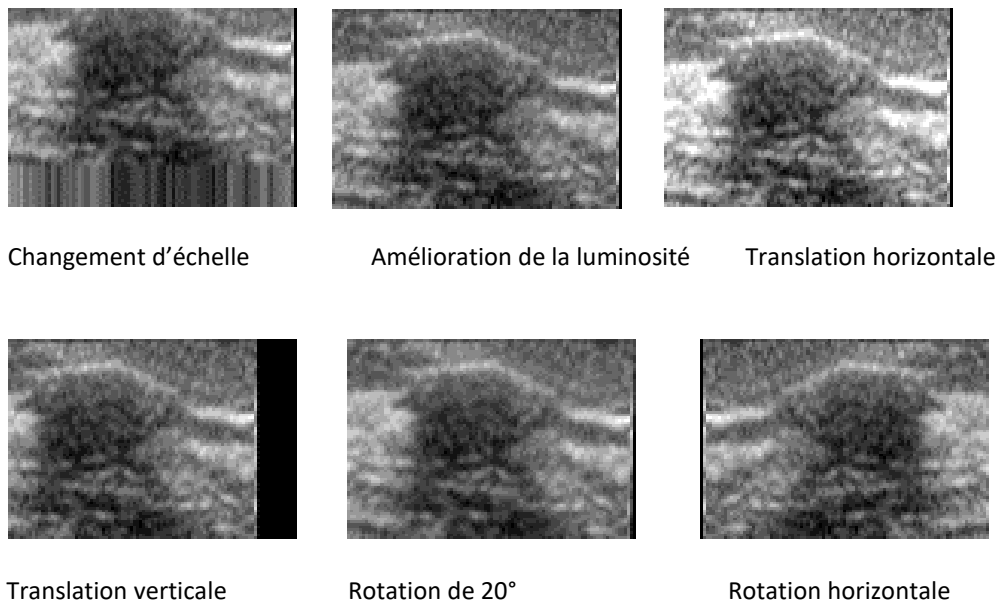


Figure 54 Exemple d'augmentation des données de la base BUSI

4.6 Préparation des données pour l'apprentissage

La répartition de la base des données est cruciale dans l'entraînement des modèles de l'apprentissage profond (figure 4.13). Ainsi,

- 60% sont utilisés pour l'entraînement (train) permettant à cet effet de déterminer les paramètres des réseaux.
- 20% pour la validation (valid) : l'ensemble de validation est destiné à identifier les hyperparamètres.
- 20% pour le test (test) : cet ensemble est destiné à l'évaluation du modèle.

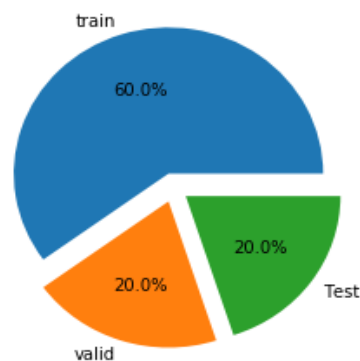


Figure 55 Répartition des données

Grâce à la fonction 'ratio' de la bibliothèque 'split_folders', la division de l'ensemble des données, est automatique et rapide. Dans le cas de la base BUSI, chaque dossier contient les trois classes correspondant à la base. A chaque classe, est attribué un code, pour

effectuer les traitements. Ainsi, la classe bénigne est codée par 0, 1 est affecté à la classe maligne et 2 est attribué à la classe normale.

Le nombre de clichés échographiques appartenant à chaque classe est de 945 pour l'entraînement (train), 315 clichés pour la validation (val) et de 420 clichés pour le test (figure 4.14). Il s'agit alors, de réaliser une multiclassification.

```
Found 945 images belonging to 3 classes.  
Found 315 images belonging to 3 classes.  
Found 420 images belonging to 3 classes.  
{0: 'benign', 1: 'malignant', 2: 'normal'}
```

Figure 56 Nombre de clichés dans chaque classe dans BUSI

Dans le cas de la deuxième base de données, la classification est binaire. Dans ce cas, 0 est attribué à la classe bénigne et 1 à la classe maligne.

Le nombre de clichés échographiques correspondant à chaque classe est de 150 pour l'entraînement (train), 50 pour la validation (val) et 50 pour le test (figure 4.15).

```
Found 150 images belonging to 2 classes.  
Found 50 images belonging to 2 classes.  
Found 50 images belonging to 2 classes.  
{0: 'benign', 1: 'malignant'}
```

Figure 57 Nombre de clichés dans chaque classe de la base US

4.7 Architectures des modèles de 'BreastCNNPredict'

Le concept de notre approche (figure 4.16), est d'adopter les modèles « Alex Net, VGG16, VGG19, XCEPTION » pré-entraînés sur keras, en gelant les couches convolutives pour entrainer juste, les couches fully connected : c'est le «fine tuning ».

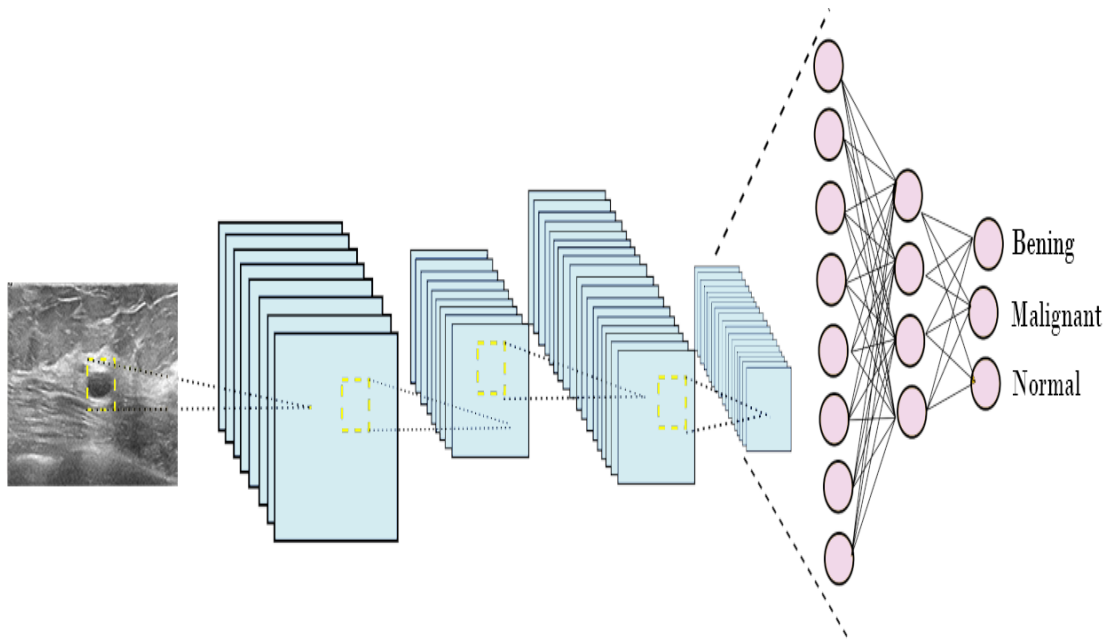


Figure 58 Modèle de BreastCNNPredict

Les couches convolutives permettent l'extraction automatique des informations pertinentes. Les couches 'fully connected', sont chargées de la classification de ces informations pertinentes.

4.7.1 Modèle VGG16 appliqué à la multiclassification

L'architecture du VGG16 a été modifiée, pour être adaptée à la classification des données de la base BUSI (figure 4.17) où toutes les opérations de convolution sont effectuées par un Kernel de taille 3*3.

- **Partie 1** : réalisation de deux produits de convolution avec 64 filtres, suivis par une couche de pooling.
- **Partie 2** : réalisation de deux produits de convolution avec 128 filtres suivis par une couche de pooling.
- **Partie 3** : réalisation de deux produits de convolution avec 526 filtres suivis par une couche de pooling.
- **Partie 4** : réalisation de trois produits de convolution avec 512 filtres suivis par une couche de pooling.
- **Partie 5** : réalisation de trois produits de convolution avec 512 filtres suivis par une couche de pooling.
- **Partie 8** : réalisation des couches fully connected.

- La première couche 'fully connected' contient 256 nœuds (ce nombre de neurones est obtenu grâce à plusieurs testes) suivis par un droup out de 20%, en utilisant la fonction d'activation Relu.
- La deuxième couche fully connected contient 256 nœuds suivis par un droup out de 30% .en utilisant la fonction d'activation Relu.

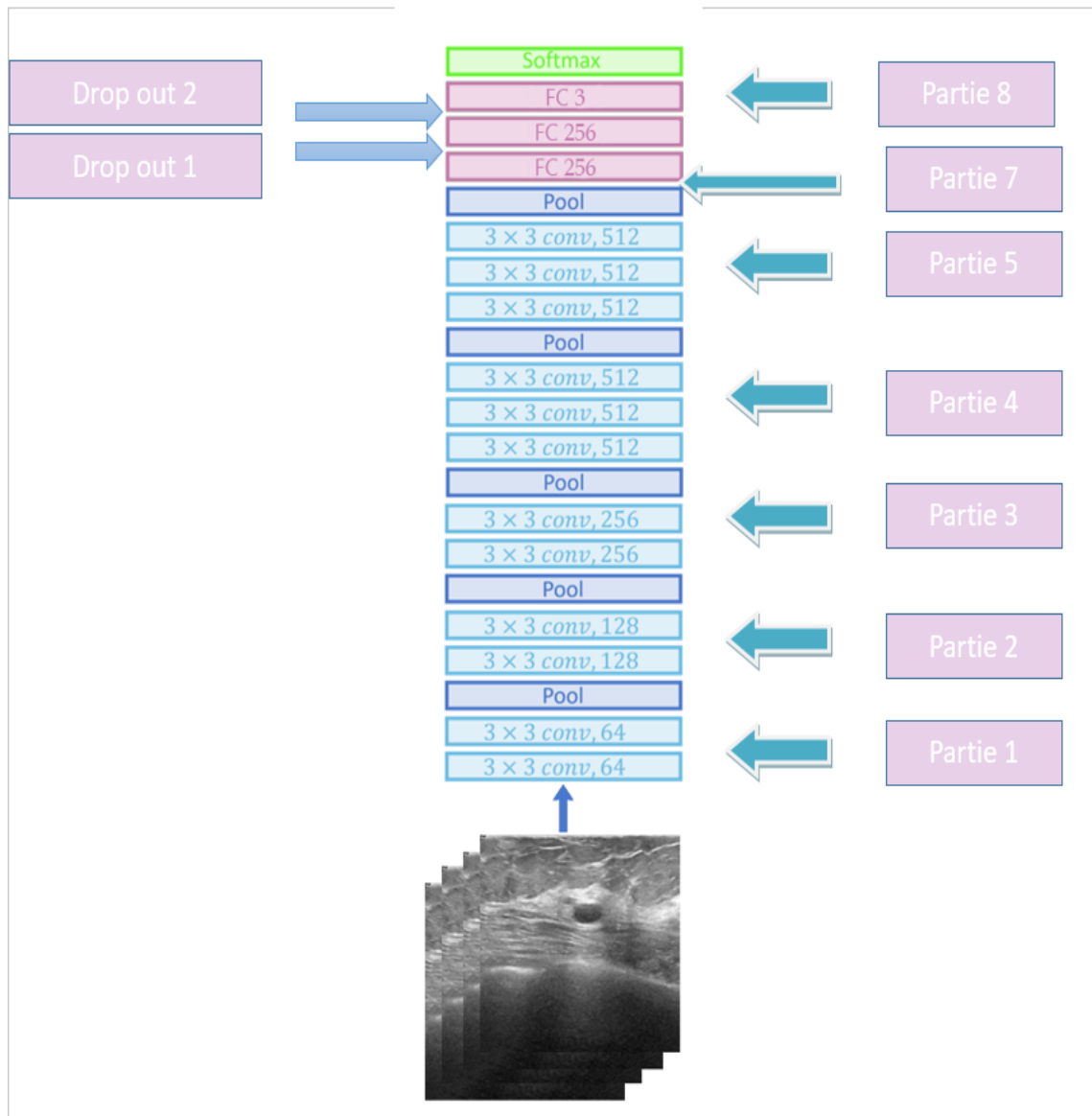


Figure 59 VGG16 pour la base de données BUSI

- Exemple d'un produit de convolution simple d'une image sans padding avec un stride =1 (figure 4.18).

0	30	1	0
1	0	25	12
7	25	40	20
8	9	1	10

(a) image d'entrée 4*4

-1	0	1
1	0	1
1	0	-1

(b) Kernel de 3*3

-6	-13
78	48

(c) Image de sortie 2*2

Figure 60 : Exemple d'une convolution

- Exemple d'un maxpooling 2*2 (figure 4.19)

-2	7	100	12
1	5	23	120
70	15	-90	4
19	50	50	30

7	120
70	50

Figure 61 Exemple d'un maxpooling 2*2

- **Partie 6** : réalisation de l'opération d'aplatissement (flatten). Il s'agit alors de transformer une matrice de dimension n*n en un vecteur de n*1 (figure 4.20).

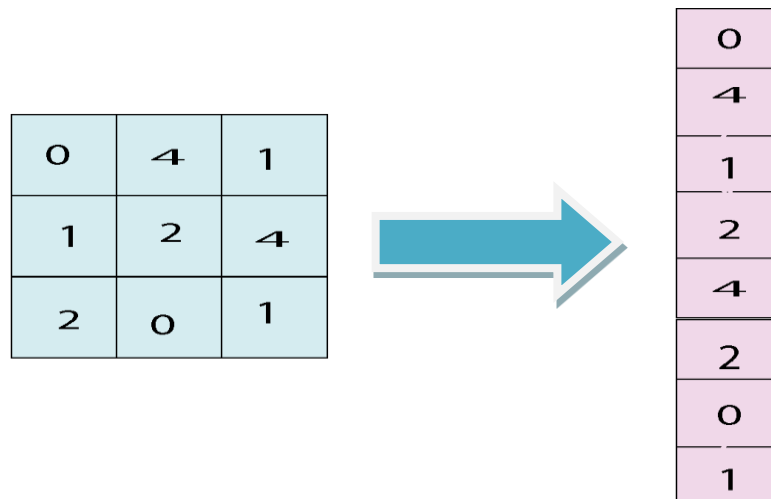


Figure 62 : Exemple de l'opération flatten

➤ Exemple sur Dropout

Dropout ou abandon. Cette opération (figure 4.21) sert à diminuer les connexions entre les couches, afin d'éviter le surapprentissage.

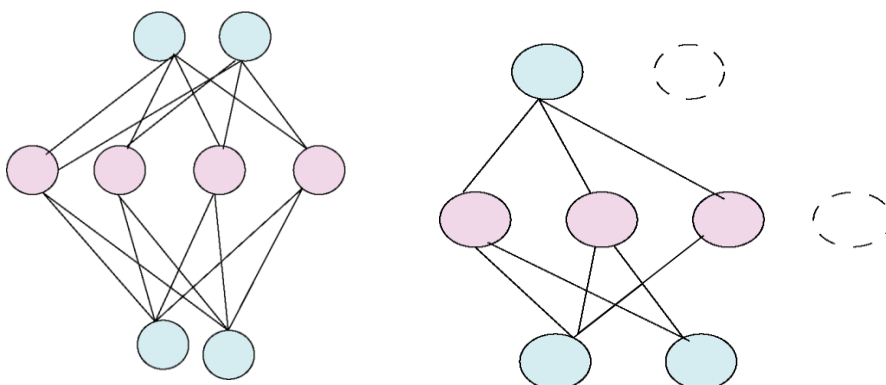


Figure 63 : Exemple de Dropout

La dernière couche (figure 4.21) contient 3 nœuds avec une fonction d'activation softmax, qui permet de distribuer les probabilités pour chaque classe.



Figure 64 : Dernière couche fully connected

4.7.2 Architecture VGG16 pour la classification binaire

Le modèle VGG16 adopté pour la classification binaire des images échographiques, correspond à la figure 4.23.

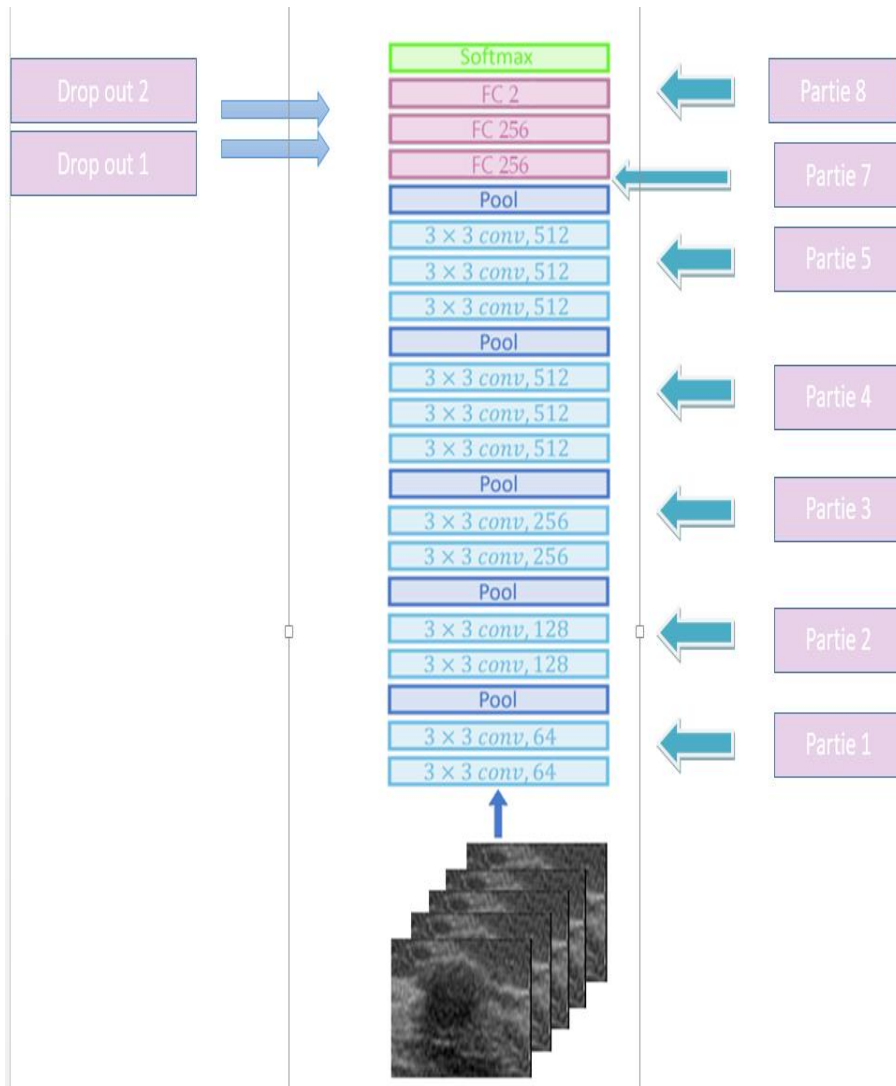


Figure 65 vgg 16 pour US

La seule différence entre le modèle multiclassés, se révèle dans le nombre de nœuds, de la dernière couche 'fully connected' (partie 8). A cet effet, la dernière couche contient 2 nœuds (figure 4.24) avec une fonction d'activation softmax, qui permet de distribuer les probabilités pour chaque classe.

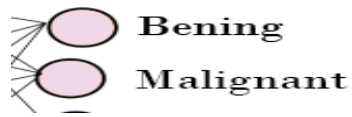


Figure 66 : Dernière couche fully connected pour US

4.7.3 VGG19

La différence entre l'architecture VGG16 et le VGG19 est présentée dans les parties 4 et 5 du modèle de la figure 4.17, où une opération de convolution est ajoutée en plus.

- **Partie 4** : réalisation de quatre opérations de convolution avec 512 filtres suivies par une couche de pooling
- **Partie 5** : réalisation de quatre opérations de convolution avec 512 filtres suivies par une couche de pooling.

Les figures 4.25 et 4.26 représentent respectivement, les modèles VGG19, adoptés pour la multiclassification et la classification binaire des clichés échographiques mammaires.

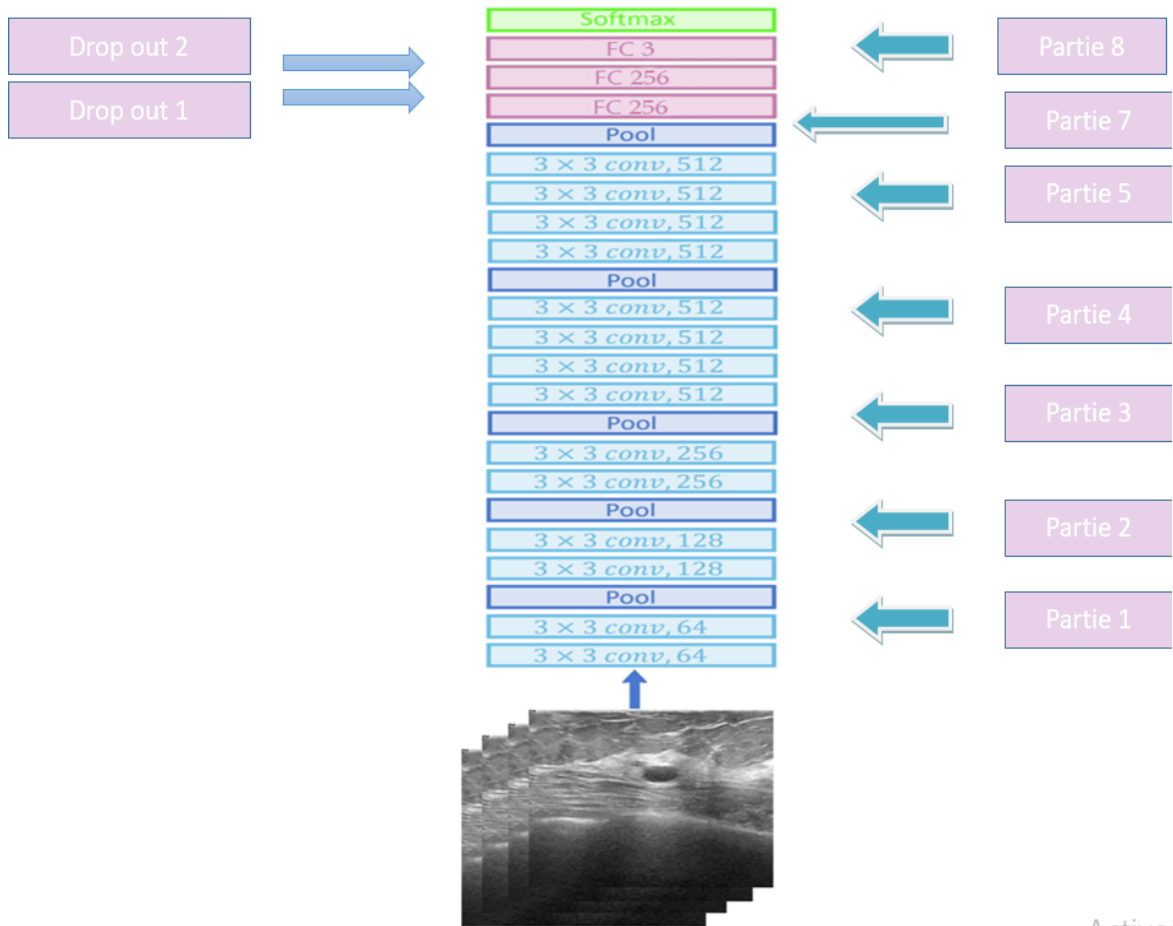


Figure 67 : VGG19 pour BUSI

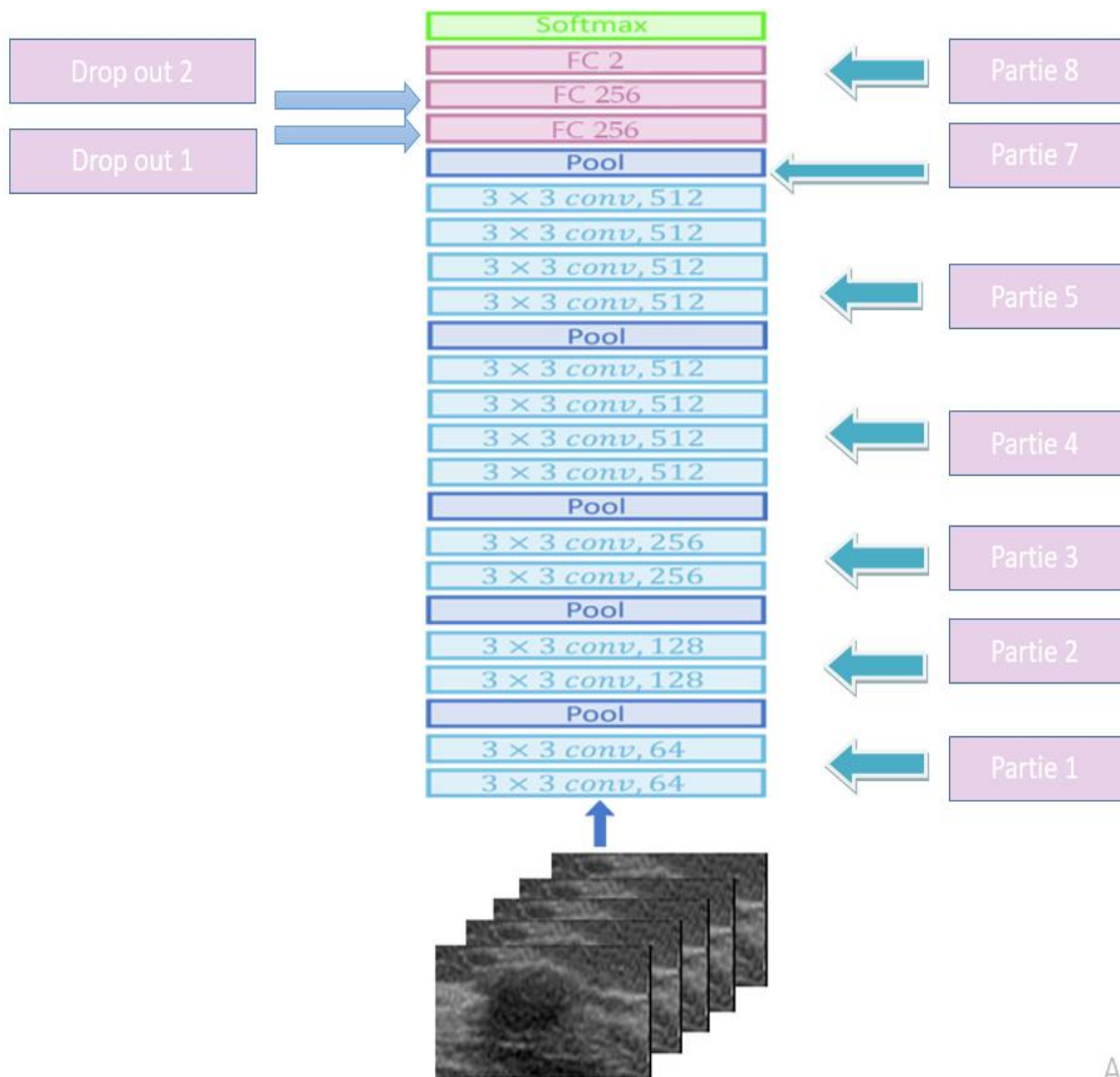


Figure 68 : VGG19 pour US

4.7.4 Modèle ALEXNET adopté dans BreastCNNPredict

✚ Cas de la multiclassification

Le modèle Alexnet adopté dans le cas de la multiclassification est configuré selon la figure 4.27.

- **Partie 1** : réalisation d'une convolution avec 96 filtres de taille 11×11 , avec un $\text{stride} = 4$, suivie par une couche de pooling avec une fenêtre de pooling 3×3 selon un $\text{stride} = 2$.
- **Partie 2** : réalisation d'une convolution avec 256 filtres de taille 5×5 avec un pas $= 2$ suivie par une couche de pooling avec une fenêtre de pooling 3×3 et un $\text{stride} = 2$
- **Partie 3** : réalisation de trois convolutions avec 384 filtres de taille 3×3 , selon un pas $= 2$ suivies par une couche de pooling avec une fenêtre de pooling 3×3 et un $\text{stride} = 2$.

- **Partie 4** : réalisation de la couche 'flatten'.
- **Partie 5** : la première couche fully connected contient 4096 nœuds suivis par un droup out de 20%, en utilisant la fonction d'activation Relu. La deuxième couche fully connected contient 1000 nœuds suivis par un droup out de 30%, en utilisant la fonction d'activation Relu. La dernière couche contient 3 nœuds et, une fonction d'activation Softmax.

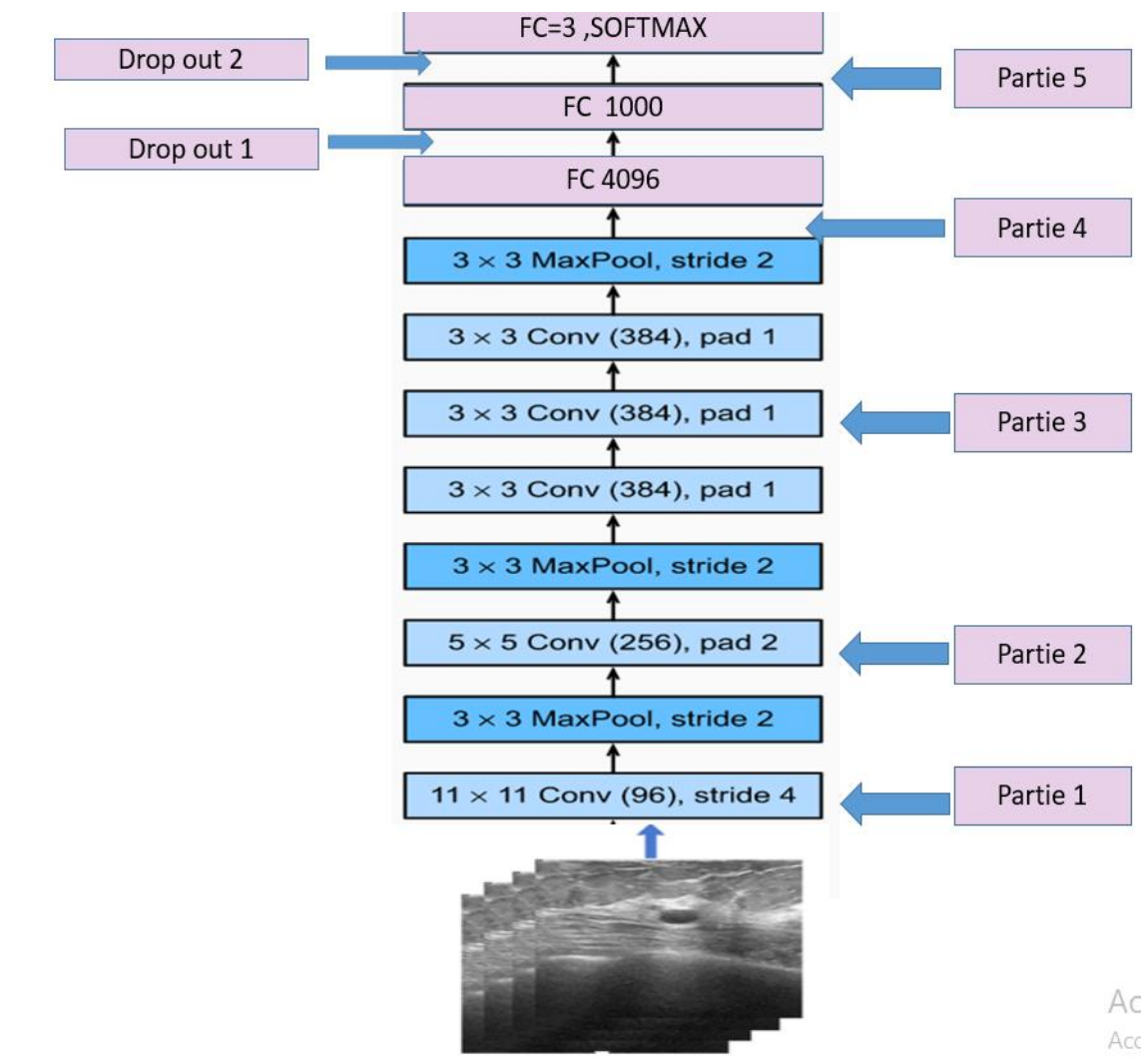


Figure 69 Alex net pour la multiclassification

✚ Architecture d'AlexNet pour la classification binaire

La seule différence entre le modèle de la multiclassification et la classification binaire (figure 4.28), se révèle aussi dans la dernière couche fully connected (partie 5), en utilisant 2 couches.

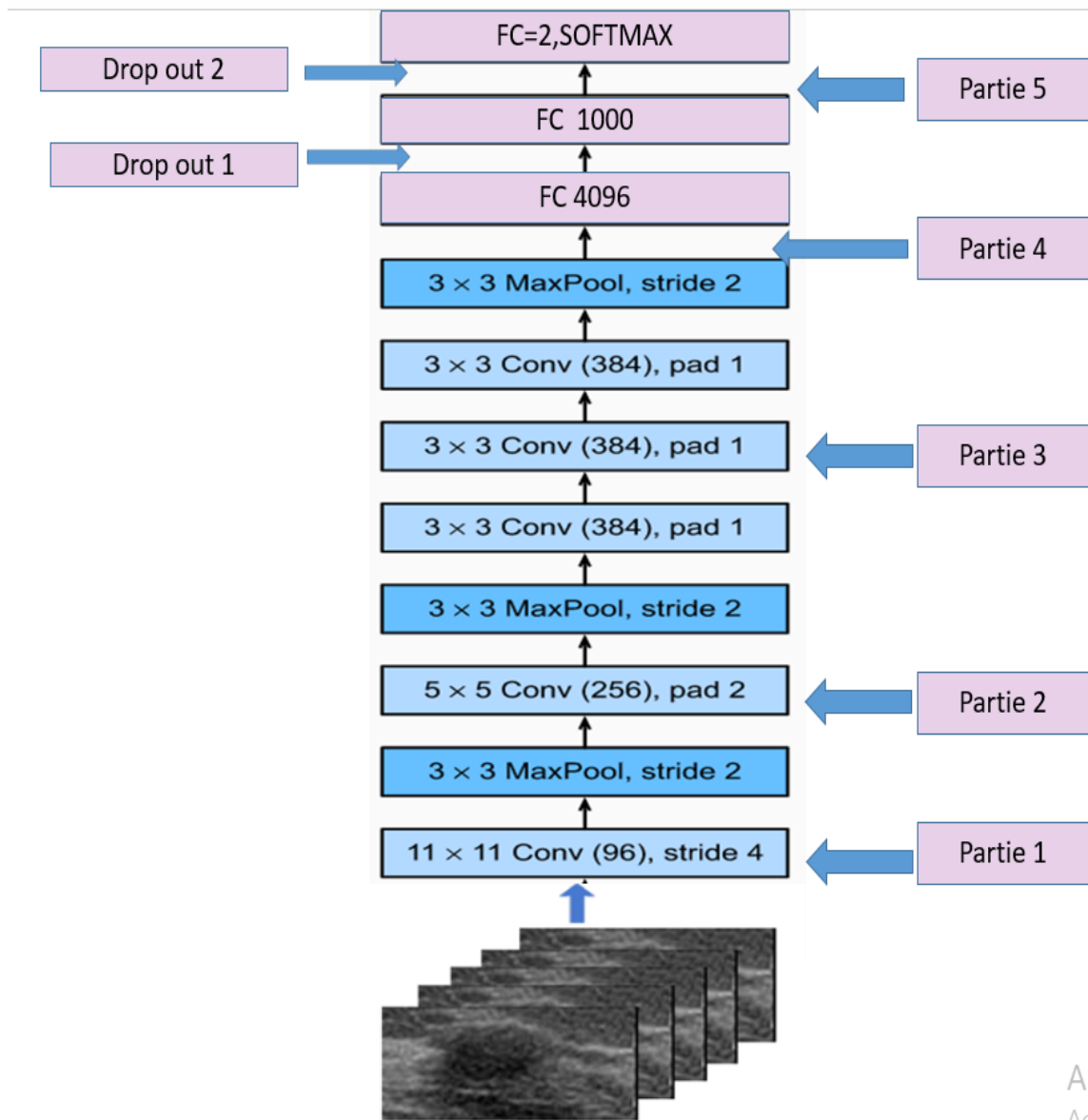


Figure 70 Alexnet pour la classification binaire

4.7.5 Modèle Xception

✚ Xception pour la multiclassification

Le modèle Xception adopté dans le cadre de la multiclassification échographique mammaire, correspond à la figure 4.29.

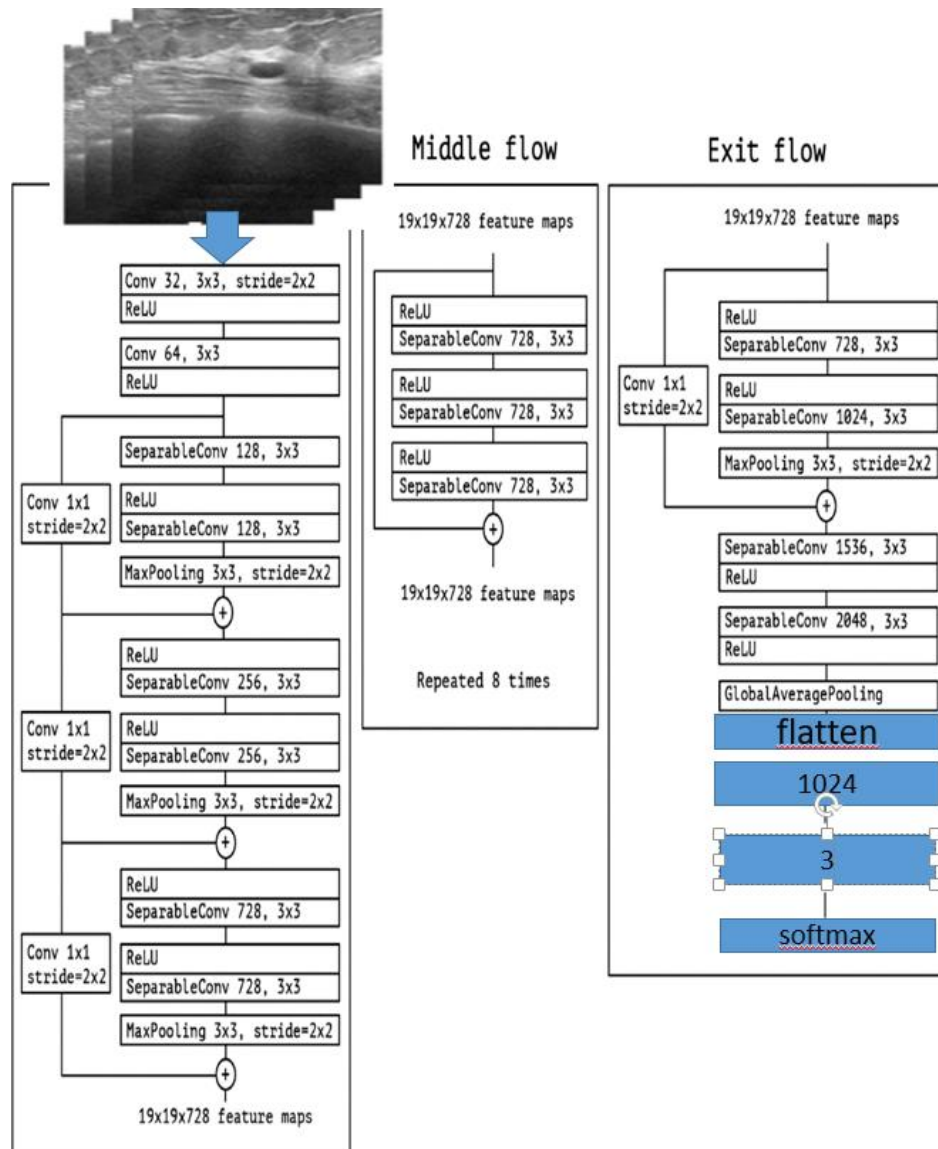


Figure 71 : Architecture Xception pour la multiclassification

✚ Xception pour la classification binaire

Le modèle Xception adopté dans le cadre de la classification échographique mammaire, correspond à la figure 4.30.

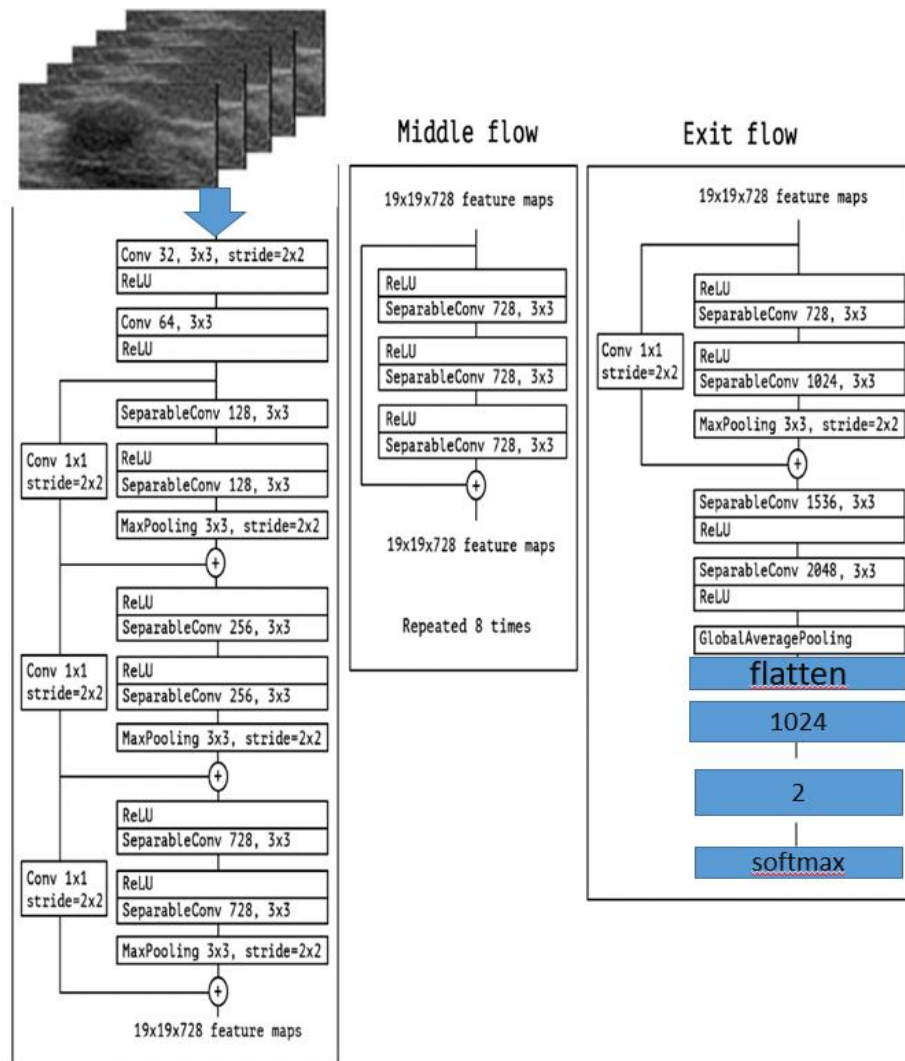


Figure 72 : Architecture Xception pour la classification binaire

4.8 Optimiseurs utilisés

Dans le cadre de nos expériences, nous avons considéré deux optimiseurs ; ADAM et ADADELTA, suivant un taux d'apprentissage (learning rate) variant de 0.1 jusqu'à 0.0001.

4.9 Fonction de perte adoptée

Dans ce travail, nous avons utilisé la fonction de perte, d'entropie croisée catégorique (Categorical Cross-Entropy) appelée Softmax loss [62]. Il s'agit d'une activation Softmax plus une perte par entropie croisée (figure 4.31).

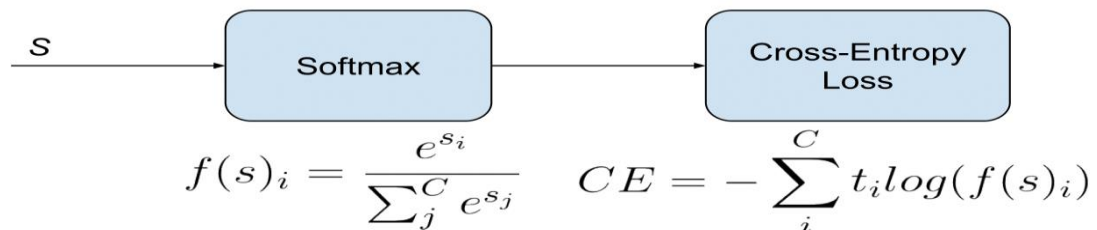


Figure 73 Fonction loss [62]

4.10 Résultats de 'BreastCNNPredict'

Le concept de cette étude est de tester les architectures précédentes, sur les deux bases de données US et BUSI, en utilisant deux optimiseurs ADADELTA et ADAM, suivant différents taux d'apprentissage [0.1, 0.01, 0.001, 0.0001]. Le nombre des époques est varié de 10 jusqu'à 50, afin de déterminer le modèle type et la compatibilité des optimiseurs, suivant la nature des données.

4.10.1 Résultats obtenus par le modèle XCEPTION

XCEPTION et BUSI dataset

Les résultats obtenus pour la classification des données échographiques en trois catégories (malignes, bénignes et normales), en fonction des paramètres de l'apprentissage, sont regroupés dans le tableau 4.1.

L'exactitude du modèle (figure 4.32) est exprimée en fonction des optimiseurs, des taux d'apprentissage du nombre des époques.

	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA
LR Epoques	0,1		0,01		0,001		0,0001	
	10	43%	56%	53%	60%	56%	51%	61%
20	43%	55%	53%	58%	56%	51%	56%	56%
30	43%	55%	53%	59%	56%	51%	56%	56%
40	43%	56%	53%	60%	56%	51%	60%	56%
50	43%	56%	53%	59%	56%	51%	56%	56%

Tableau 74 : Résultats obtenus par le modèle Xception pour Busi dataset

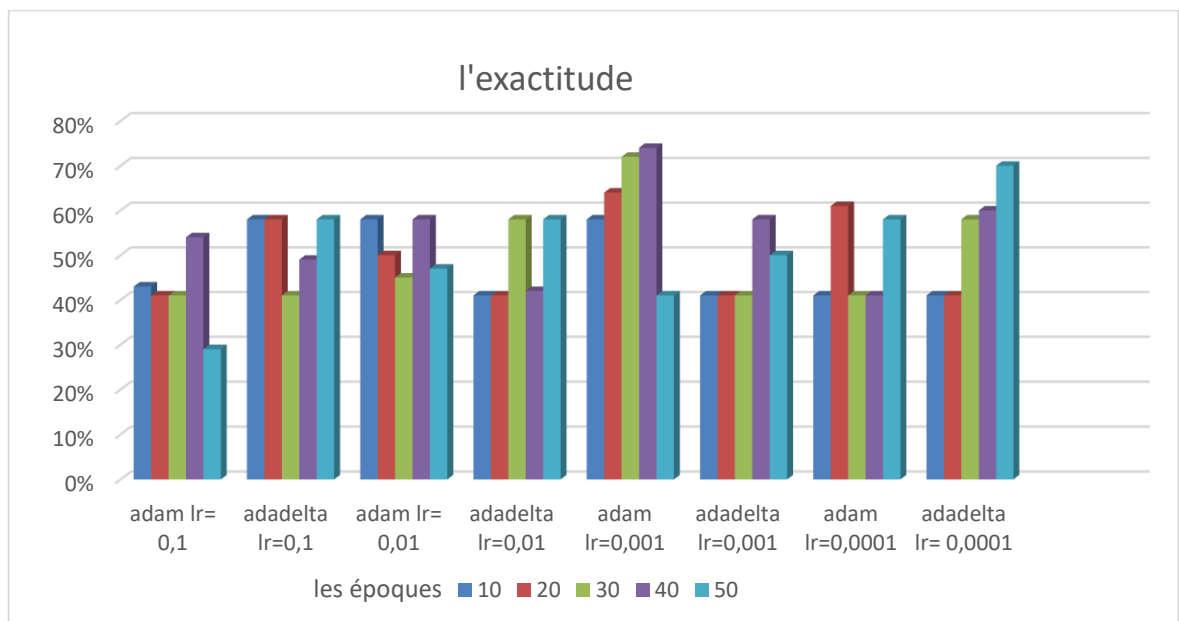


Figure 75 : Exactitude du modèle xception pour la base BUSI

✚ XCEPTION appliqué à la base US

Les résultats obtenus pour cette classification, se résument dans le tableau 4.2 et la figure 4.33.

	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA
LR Epoques	0,1		0,01		0,001		0,0001	
10	54%	43%	56%	43%	43%	56%	47%	54%
20	54%	41%	43%	58%	56%	45%	58%	70%
30	45%	47%	62%	43%	58%	43%	50%	45%
40	41%	60%	52%	60%	58%	50%	60%	41%
50	58%	60%	56%	58%	60%	43%	64%	58%

Tableau 76 : Résultats obtenus par le modèle Xception pour la base US dataset

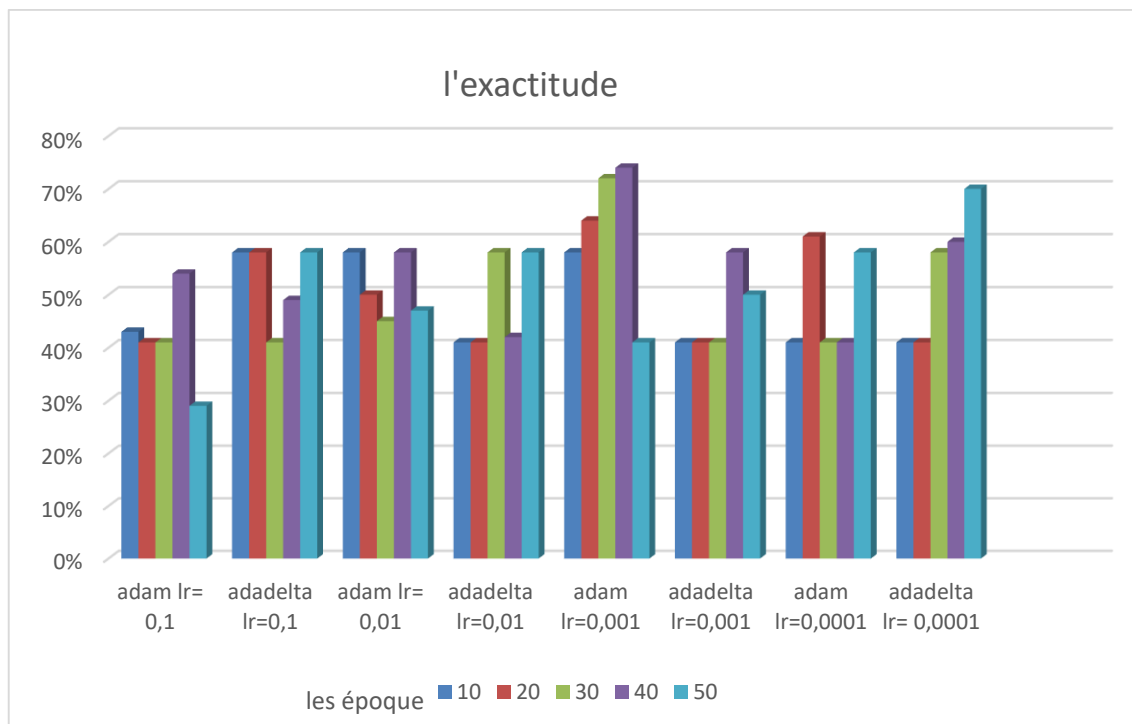


Figure 77 : Exactitude pour le modèle xception pour la base US

Vu l'exactitude obtenue, le modèle Xception n'est pas adapté à notre problème.

4.10.2 Résultats obtenus par ALEXNET

ALEXNET appliqué à la base BUSI

Les résultats de ce modèle, sont indiqués dans le tableau 4.3. L'exactitude d'Alexnet est illustrée par la figure 4.34.

	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA
Epoques \ LR	0,1		0,01		0,001		0,0001	
	10	29%	31%	43%	29%	64%	47%	50%
20	60%	42%	39%	53%	48%	42%	53%	23%
30	44%	43%	51%	67%	61%	53%	65%	32%
40	68%	27%	49%	69%	49%	54%	64%	50%
50	55%	20%	28%	70%	59%	65%	75%	51%
60	71%	31%	77%	56%	65%	61%	68%	43%
70	84%	43%	58%	62%	64%	66%	43%	68%

Tableau 2: Résultats obtenus par le modèle ALEXNET pour BUSI dataset

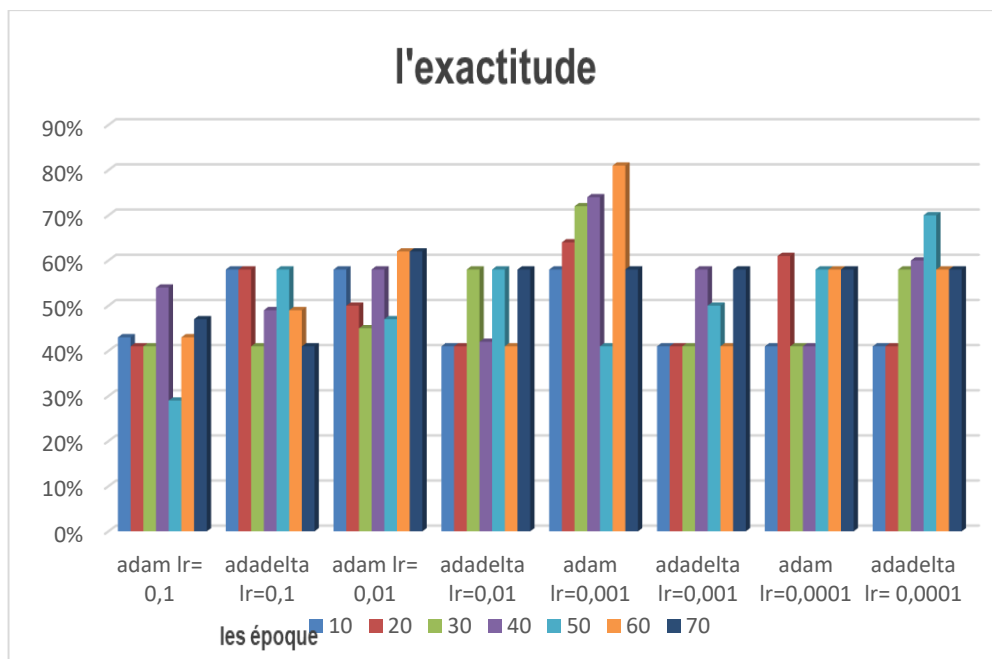


Figure 78: Exactitude du modèle ALEXNET pour BUSI

ALEXNET appliqué sur US

Les résultats réalisés dans ce cas, sont considérés dans le tableau 4.4.

	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA
LR	0,1		0,01		0,001		0,0001	
Epoques	0,1		0,01		0,001		0,0001	
10	43%	58%	58%	41%	58%	41%	41%	41%
20	41%	58%	50%	41%	64%	41%	61%	41%
30	41%	41%	45%	58%	72%	41%	41%	58%
40	54%	49%	58%	42%	74%	58%	41%	60%
50	29%	58%	47%	58%	41%	50%	58%	70%
60	43%	49%	62%	41%	81%	41%	58%	58%
70	47%	41%	62%	58%	58%	58%	58%	58%

Tableau 3 : Résultats obtenus par le modèle ALEXNET pour la base US dataset

La figure 4.35 représente l'exactitude du modèle en fonction des époques et des optimiseurs.

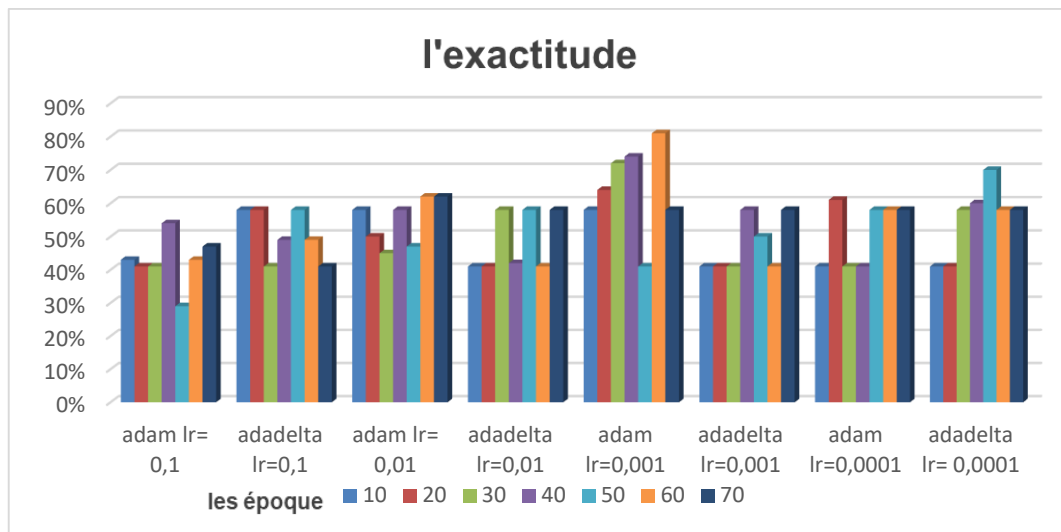


Figure 79: Exactitude du modèle ALEXNET pour US

4.10.3 Résultats obtenus par le VGG16

✚ VGG16 appliqué à la base BUSI

Les résultats de ce modèle, sont regroupés dans le tableau 4.5.

	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA
LR \ Epoques	0,1		0,01		0,001		0,0001	
10	43%	88%	43%	81%	52%	43%	89%	40%
20	43%	87%	43%	86%	72%	43%	92%	43%
30	43%	91%	43%	87%	44%	43%	93%	43%
40	43%	91%	43%	86%	90%	43%	93%	43%
50	43%	92%	43%	91%	88%	43%	93%	43%

Tableau 4: Résultats obtenus par le modèle VGG16 pour BUSI

L'exactitude du modèle VGG16, est représentée par la figure 4.36.

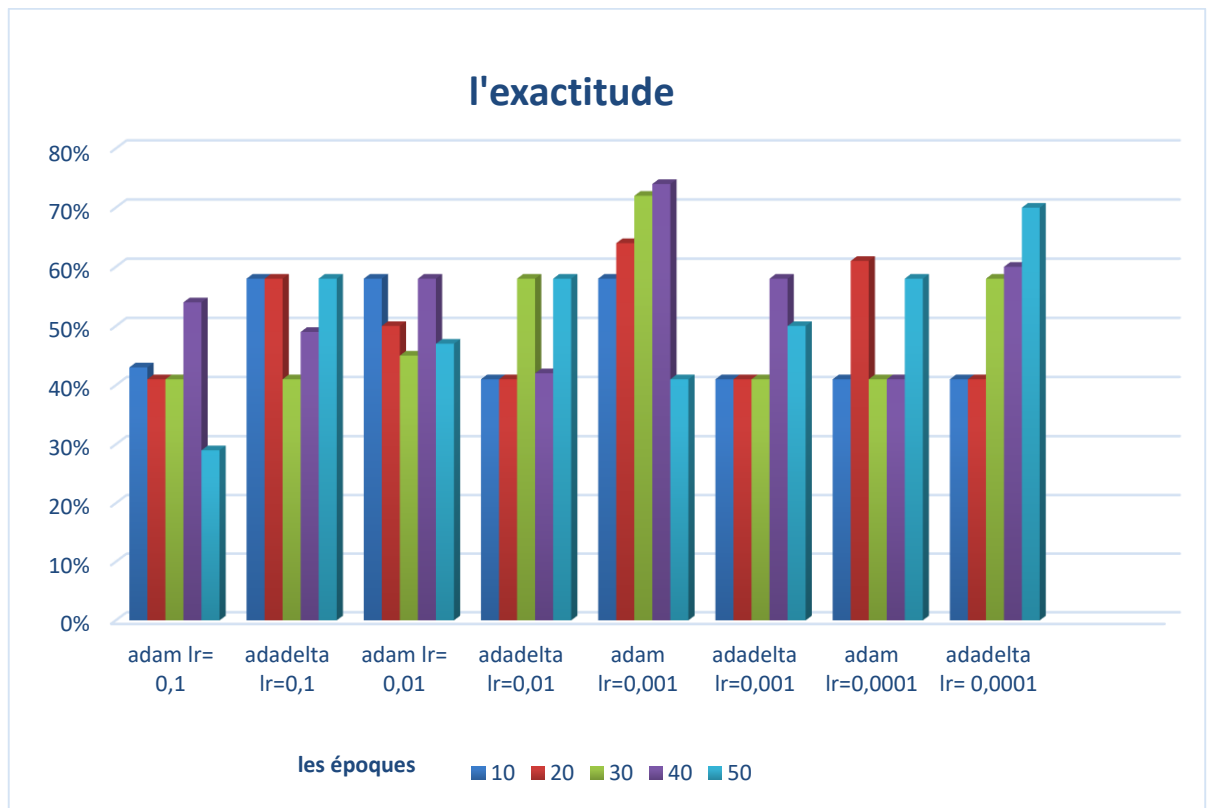


Figure 80: Exactitude du modèle vgg16 pour Busi

L'optimiseur ADAM est compatible avec les taux d'apprentissage 0.001 et 0.0001, contrairement à l'optimiseur Adadelta, qui est compatible avec les taux d'apprentissage 0.1 et 0.01.

Le nombre idéal des époques correspond à 50, dans le cas de nos expériences.

Modèle VGG16 appliqué pour la base US

Les résultats obtenus se résument dans le tableau 4.6 et la figure 4.37.

	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA
LR								
Epoques	0,1		0,01		0,001		0,0001	
10	58%	87%	58%	93%	58%	58%	89%	41%
20	58%	93%	58%	80%	58%	58%	91%	52%
30	58%	95%	58%	89%	58%	75%	93%	58%
40	58%	93%	58%	95%	58%	75%	93%	58%
50	58%	89%	58%	97%	58%	87%	93%	58%

Tableau 5: Résultats obtenus par le modèle VGG16 pour US

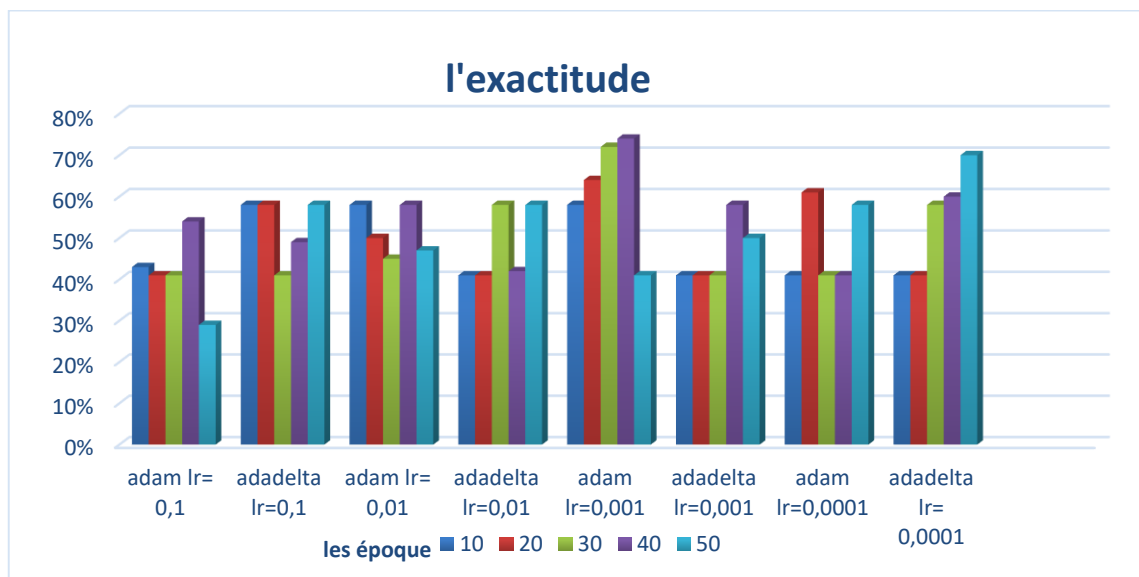


Figure 81: Exactitude du modèle vgg16 pour la base US

4.10.4 Résultats obtenus par le modèle VGG19

+ Cas de la multiclassification

Les résultats obtenus par ce modèle, sont montrés dans le tableau 4.7 et la figure 4.38.

	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA
LR								
Epoques	0,1		0,01		0,001		0,0001	
10	43%	86%	43%	80%	43%	47%	89%	28%
20	43%	87%	43%	85%	82%	57%	88%	44%
30	43%	87%	43%	83%	85%	58%	91%	43%
40	43%	87%	43%	85%	43%	64%	89%	43%
50	43%	87%	43%	88%	83%	73%	91%	43%

Tableau 6: Résultats obtenus par le modèle VGG19 pour BUSI

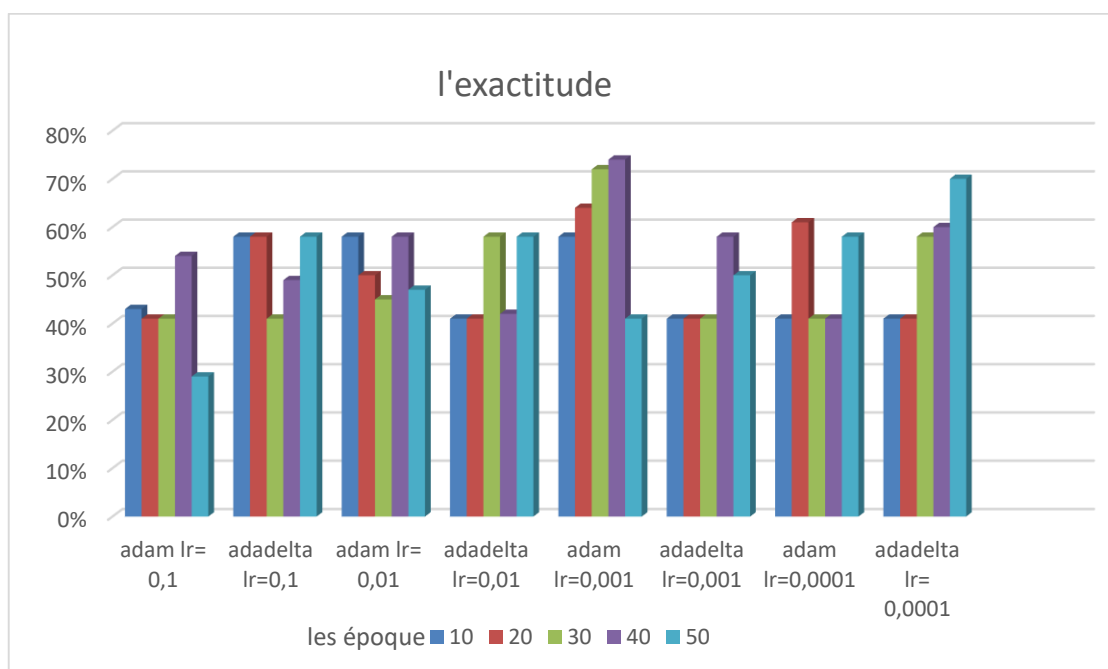


Figure 82: Exactitude du modèle vgg19 pour la base BUSI

Cas de la classe binaire

Les résultats de la seconde base de données, se regroupent dans le tableau 4.8 et la figure 4.39.

	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA	ADAM	ADADELTA
LR \ Epoques	0,1		0,01		0,001		0,0001	
10	58%	47%	58%	85%	58%	58%	89%	58%
20	58%	88%	58%	83%	58%	58%	95%	58%
30	58%	89%	58%	87%	58%	58%	93%	58%
40	58%	90%	58%	89%	58%	58%	97%	58%
50	58%	87%	58%	91%	58%	58%	100%	58%

Tableau 7 Résultats obtenus par le modèle VGG19 pour US

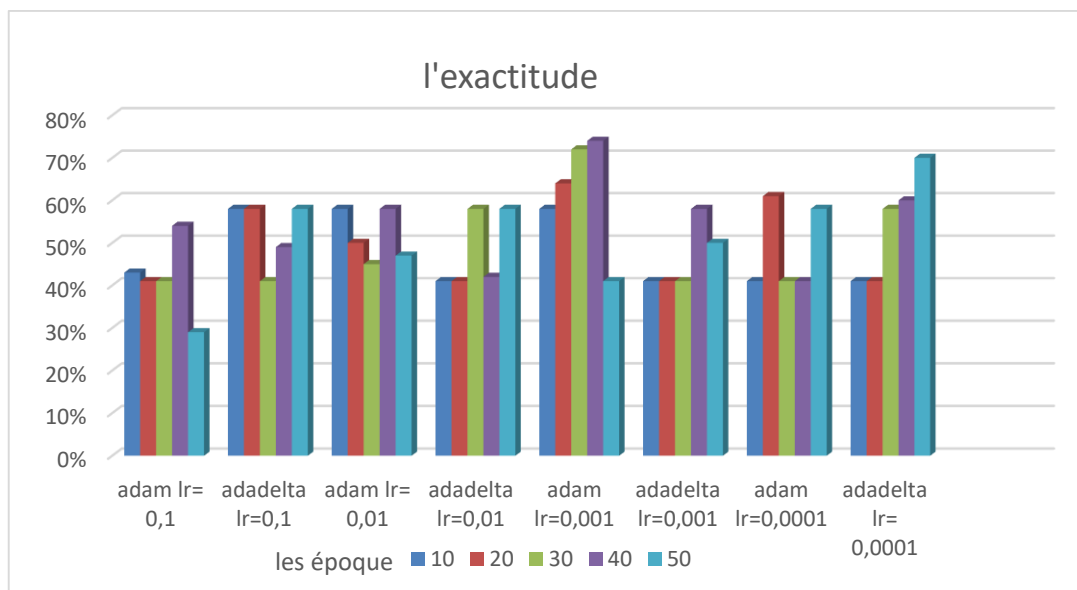


Figure 83: Exactitude du modèle vgg19 pour la base US

4.10.5 Modèle type obtenu par VGG16 pour BUSI dataset

Nous avons obtenu le modèle type avec l'exactitude la plus élevée (94%), à la 50 -ème époque, en utilisant l'optimiseur ADAM avec un taux d'apprentissage égal à 0.0001.

```
Accuracy: 0.935096  
Loss: 0.043329
```

L'analyse des courbes d'apprentissage et de validation, a engendré les figures 4.40 et 4.41.

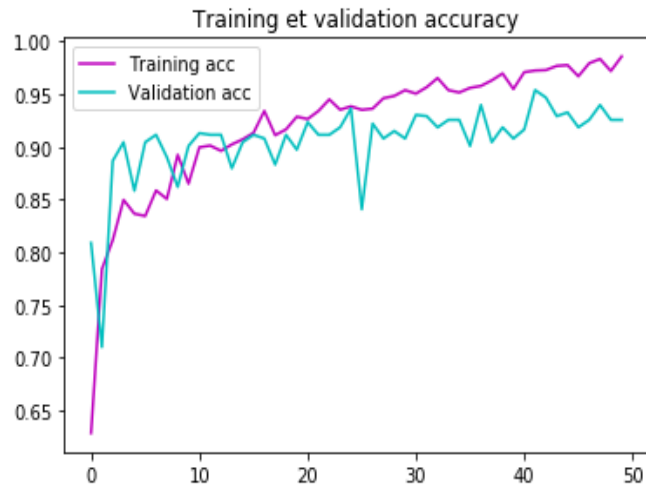


Figure 84 : Evaluation de l'exactitude des ensembles d'entraînement et de validation en fonction du nombre des époques (vgg16 et busi)

A chaque augmentation du nombre des époques, l'exactitude augmente, pour les deux ensembles. Ce qui montre une relation de corrélation directe entre l'entraînement et la validation.

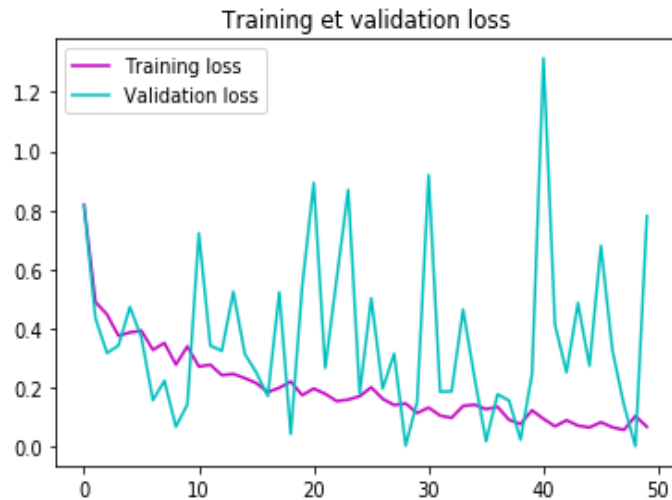


Figure 85 : Evaluation de la fonction Loss des ensembles d’entraînement et de validation en fonction du nombre des époques

Nous remarquons que l’erreur de validation ne tend pas vers zéro et reste instable, entre 0.8 et 1.2 après 50 époques ; au contraire de l’erreur d’entraînement qui diminue avec les époques.

✚ **Matrice de confusion et table de confusion**

Le système obtenu est évalué par la matrice de confusion ci-dessous :

```
[[175  2  2]
 [ 11 74  0]
 [ 11  1 144]]
```

(a) matrice de confusion

		les classes prédites		
		bénigne	maligne	normal
les classes réelles	bénigne	170	2	2
	maligne	11	74	0
	normal	11	1	144

(b) table de confusion

Figure 86 : Matrice de confusion et table de confusion (vgg16, Busi)

Les vrais positifs : il s’agit d’une classe maligne présentée comme maligne. Le nombre des vrais positifs est 74.

Les vrais négatifs : il s’agit d’une classe normale présentée comme normale. Le nombre des vrais négatifs est 144.

Les faux positifs : il s'agit d'une classe bénigne présentée comme maligne. Le nombre de faux positifs est de 2.

Les faux négatifs : il s'agit d'une classe maligne présentée comme normale. Le nombre de faux positifs est de 0.

Le rapport de classification ci-dessous, indique les critères d'évaluation du modèle étudié.

	precision	recall	f1-score	support
benign	0.89	0.98	0.93	179
malignant	0.96	0.87	0.91	85
normal	0.99	0.92	0.95	156
accuracy			0.94	420
macro avg	0.95	0.92	0.93	420
weighted avg	0.94	0.94	0.94	420

Une prédiction d'un cliché échographique aléatoirement choisi, indique la classe appropriée par ce modèle (figure 4.43).

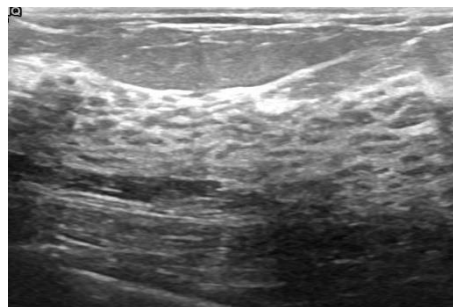


Figure 87: Prédiction d'une classe normale

La classe réelle de cette image de l'ensemble du test, est normale. Les probabilités des classes prédites sont :

```
les probabilités sont: [[0. 0. 1.]]  
benign: 0.0  
malignant: 0.0  
normal: 100.0
```

Tableau 8: Probabilités des classes obtenues par le modèle VGG16 pour la base BUSI

4.10.6 Modèle type obtenu par VGG 16 pour la base US

Nous avons obtenu le modèle type avec la précision la plus élevée (97%) à la 50 -ème époque, en utilisant l'optimiseur ADAelta avec un taux d'apprentissage égale à 0.001.

```
Accuracy: 0.979167  
Loss: 0.001149
```

Les graphes suivants représentent respectivement l'exactitude et la perte, sur les ensembles de test et d'entraînement.

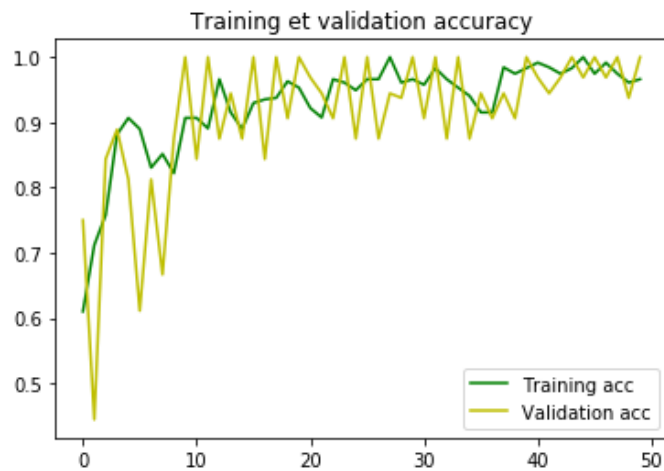


Figure 88: Evaluation de l'exactitude des ensembles d'entraînement et de validation en fonction du nombre des époques (vgg16 et us)

Nous remarquons la même évolution que pour le graphe précédent,

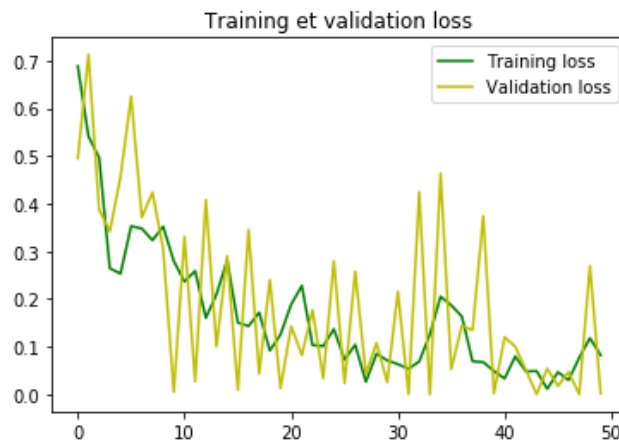


Figure 89: Evaluation de la fonction Loss des ensembles d’entraînement et de validation en fonction du nombre des époques

Nous remarquons que l’erreur de validation ne tend pas vers zéro et reste instable entre 0.3 et 0.7 après 50 époques, au contraire de l’erreur d’entraînement qui diminue avec les époques.

Ce modèle est évalué par la matrice de confusion et la courbe ROC (figures 4.45 et 4.46).

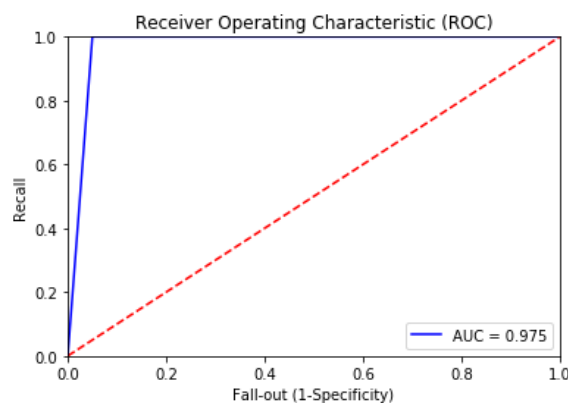


Figure 90: la courbe ROC pour le vgg16

La courbe ROC représente le taux de vrais positifs en fonction de taux de vrais négatifs. Nous avons obtenu une valeur de AUC (l’aire sous la courbe ROC) égale à 97%. Ce qui signifie que la classification est efficace.

✚ Matrice et table de confusions

La matrice de confusion est extraite, pour évaluer la classification (figure 4.46).

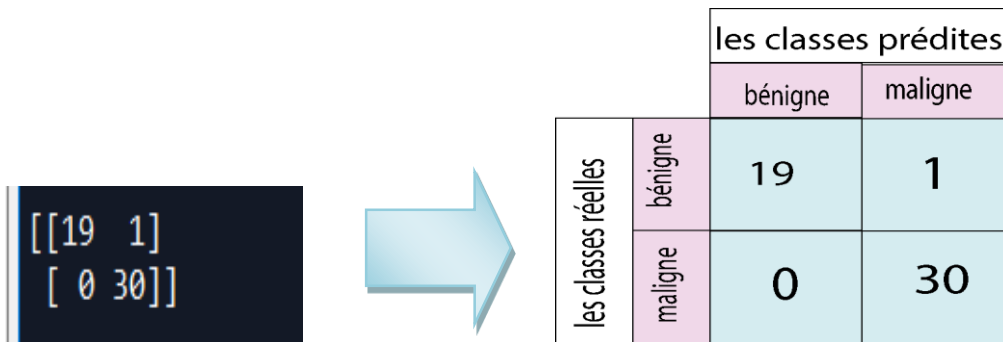


Figure 91 : matrice et table de confusion (vgg16 , us)

Les vrais positifs : il s'agit d'une classe maligne présentée comme maligne. Le nombre des vrais positifs est de 30.

Les vrais négatifs : il s'agit d'une classe bénigne présentée comme bénigne. Le nombre des vrais négatifs est de 17.

Les faux positifs : il s'agit d'une classe bénigne présentée comme maligne. Le nombre de faux positifs est de 1.

Faux négatifs : il s'agit d'une classe maligne présentée comme bénigne. Le nombre de faux positifs est de 0.

Le rapport de classification suivant, montre les critères d'évaluation obtenus pour le modèle VGG16.

	precision	recall	f1-score	support
benign	1.00	0.95	0.97	20
malignant	0.97	1.00	0.98	30
accuracy			0.98	50
macro avg	0.98	0.97	0.98	50
weighted avg	0.98	0.98	0.98	50

Une prédiction d'un cliché échographique aléatoirement choisi, montre une classification correcte.



Figure 92 : Prédiction d'une classe maligne

La classe réelle de cette image de l'ensemble de test, est maligne.

Les probabilités des classes prédites sont :

```
les probabilités sont: [[1.1868584e-04 9.9988127e-01]]  
benign: 0.011868584260810167  
malignant: 99.98812675476074
```

4.10.7 Modèle type obtenu par VGG19

Nous avons obtenu le modèle type avec la précision la plus élevée à la 50 -ème époque, en utilisant l'optimiseur ADAM avec un taux d'apprentissage égale à 0.0001.

La valeur de la précision est de 91% dans la phase d'apprentissage. La valeur de la fonction perte est 0.58.

```
Accuracy: 0.915865  
Loss: 0.582448
```

Les résultats de l'entraînement et de la validation sont illustrés par les graphes des figures 4.48 et 4.49.

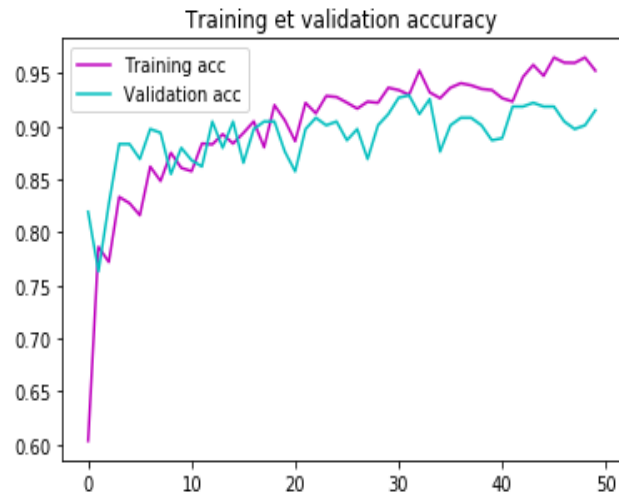


Figure 93 Evaluation de l'exactitude des ensembles d'entraînement et de validation en fonction du nombre des époques (vgg19 et busi)

Nous remarquons dans ce cas, que l'exactitude augmente, au fur et à mesure que le nombre d'époques augmente.

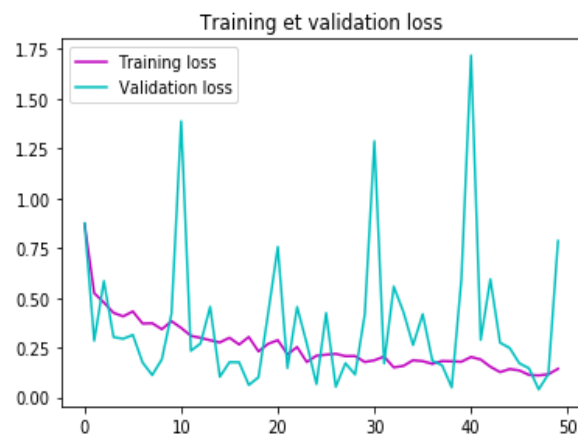
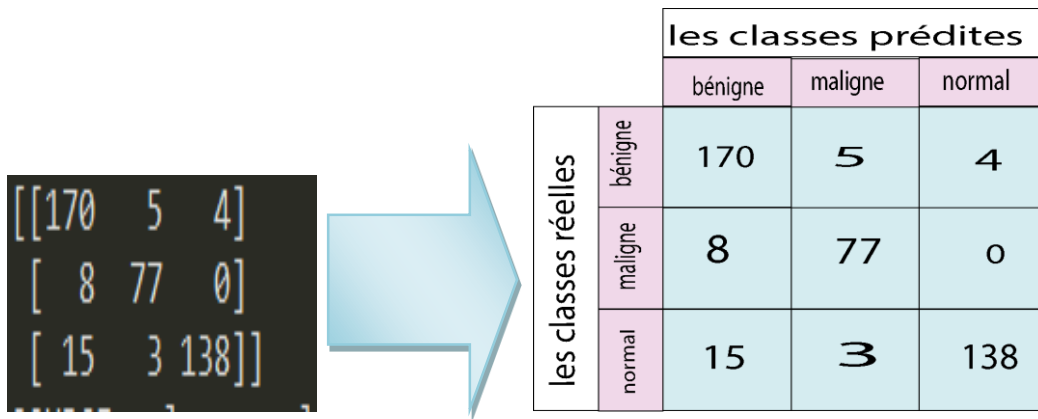


Figure 94: Evaluation de la fonction Loss des ensembles d'entraînement et de validation en fonction du nombre des époques

Nous remarquons que l'erreur de validation ne tend pas vers zéro et reste instable, entre 1 et 1.75 après 50 époques ; au contraire à de l'erreur d'entraînement qui diminue avec les époques.

✚ Matrice et table de confusion

La matrice de confusion est représentée par la figure 4.



(a) : matrice de confusion

(b) : table de confusion

Figure 95 : table et matrice de confusion (busi ,vgg19)

Les vrais positifs : il s’agit d’une classe maligne présentée comme maligne. Le nombre des vrais positifs est de 77.

Les vrais négatifs : il s’agit d’une classe normale présentée comme normale. Le nombre des vrais négatifs est de 170.

Faux positifs : il s’agit d’une classe bénigne présentée comme maligne. Le nombre de faux positifs est de 5.

Faux négatifs : il s’agit d’une classe maligne présentée comme normale. Le nombre de faux positifs est de 0.

Les performances du modèle sont indiquées ci-dessous :

	precision	recall	f1-score	support
benign	0.88	0.95	0.91	179
malignant	0.91	0.91	0.91	85
normal	0.97	0.88	0.93	156
accuracy			0.92	420
macro avg	0.92	0.91	0.92	420
weighted avg	0.92	0.92	0.92	420

Une prédiction d’un cliché échographique aléatoirement choisi, correspond à la classe maligne.

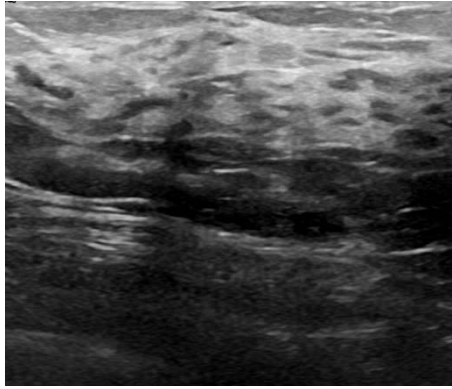


Figure 96 : Prédiction d'une classe maligne

La classe réelle de cette image de l'ensemble de test, est normale. Ceci donne alors les probabilités des classes prédites :

```
les probabilités sont: [[0. 0. 1.]]  
benign: 0.0  
malignant: 0.0  
normal: 100.0
```

4.10.8 Modèle type obtenu par l'optimiseur ADAM

Nous avons obtenu le modèle type avec la précision la plus élevée à la 50^{ème} époque en utilisant l'optimiseur ADAM avec un taux d'apprentissage égal à 0.0001.

```
Accuracy: 1.000000  
Loss: 0.000427
```

La valeur de l'exactitude est de 100% dans la phase d'entraînement. La valeur de la fonction perte est 0.000427 (figures 4.52 et 4.53).

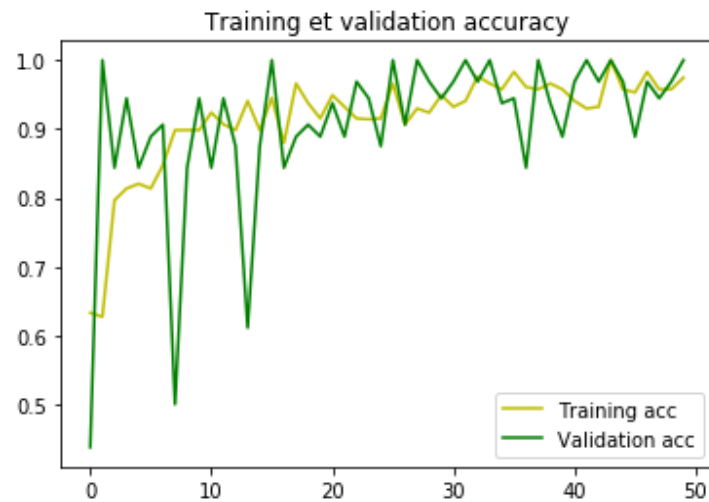


Figure 97 Evaluation de l'exactitude des ensembles d'entraînement et de validation en fonction du nombre des époques

L'exactitude se comporte de la même manière que pour les autres modèles.

Pour l'ensemble de validation, la précision reste instable entre 0 et 1.

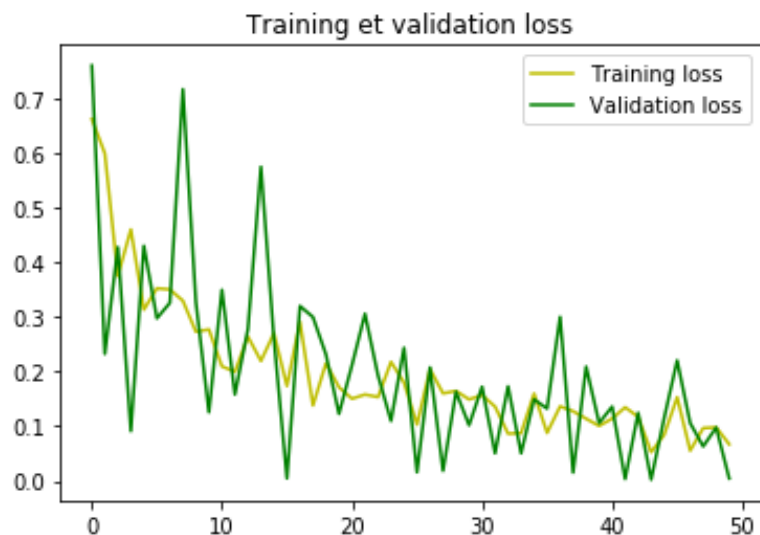


Figure 98: Evaluation de la fonction Loss des ensembles d'entraînement et de validation en fonction du nombre des époques

Nous remarquons que l'erreur de validation ne tend pas vers zéro et reste instable, entre 0.3 et 0.7 après 50 époques au contraire de l'erreur d'entraînement, qui diminue avec les époques.

La valeur de AUC (l'aire sous la courbe ROC) obtenue, pour le meilleur modèle est égale à 1 ce qui signifie que le classifieur catégorise efficacement les lésions (figure 4.54).

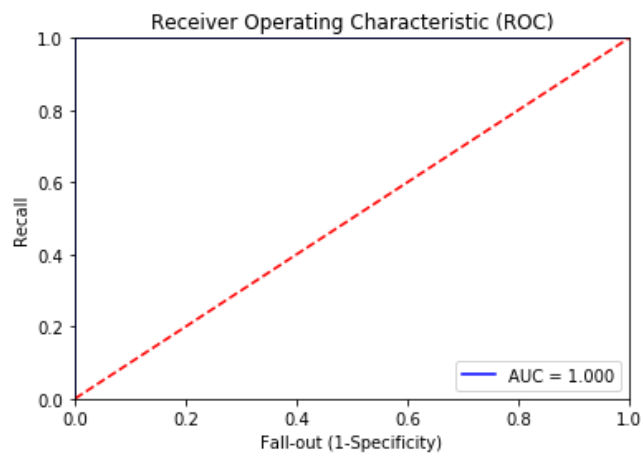


Figure 99 courbe ROC (vgg19)

✚ La matrice de confusion

La matrice de confusion à deux classes, indique la prédiction des différentes classes de la base US (figure 4.55).

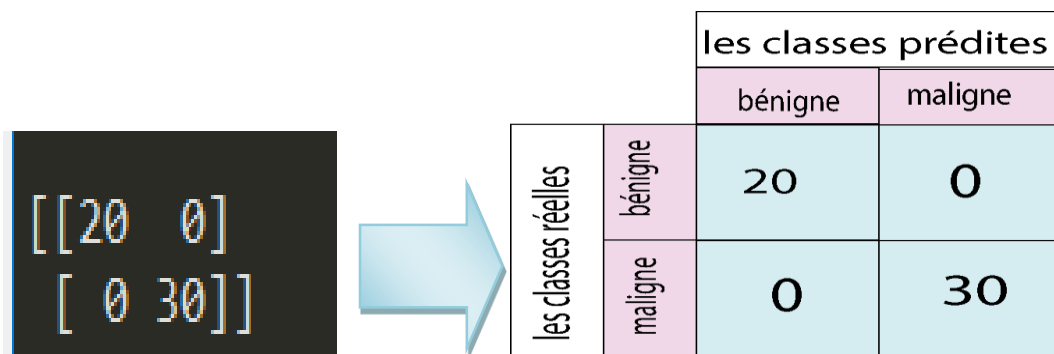


Figure 100: table et matrice de confusion (vgg16 et us)

Les vrais positifs : Le nombre des vrais positifs est de 30.

Les vrais négatifs : le nombre des vrais négatifs est 20.

Les faux positifs : le nombre de faux positifs est de 0.

Les faux négatifs : le nombre de faux positifs est de 0.

Le rapport de classification indique les paramètres ci-dessous :

```
precision    recall  f1-score   support

   benign     1.00     1.00     1.00     20
  malignant     1.00     1.00     1.00     30

 accuracy                   1.00     50
  macro avg     1.00     1.00     1.00     50
  weighted avg     1.00     1.00     1.00     50
```

Une prédiction d'un cliché échographique aléatoirement choisi, détermine la véritable classe.

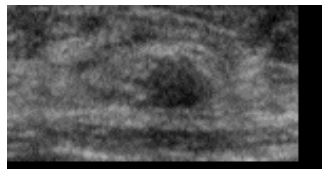


Figure 101 : Prédiction d'une classe maligne

La classe réelle de cette image de l'ensemble de test est bénigne. Les probabilités des classes prédites :

```
les probabilités sont: [[1. 0.]]
benign: 100.0
malignant: 0.0
```

4.11 Présentation de l'interface 'BreastCNNPredict'

La bibliothèque TKinter intégrée à la bibliothèque standard du python, nous donne le privilège de créer l'interface « BreastCNNPredict», à partir des approches implémentées. L'interface développée (figure 4.57), permet d'interpréter un cliché échographique mammographique en deux cliques seulement.

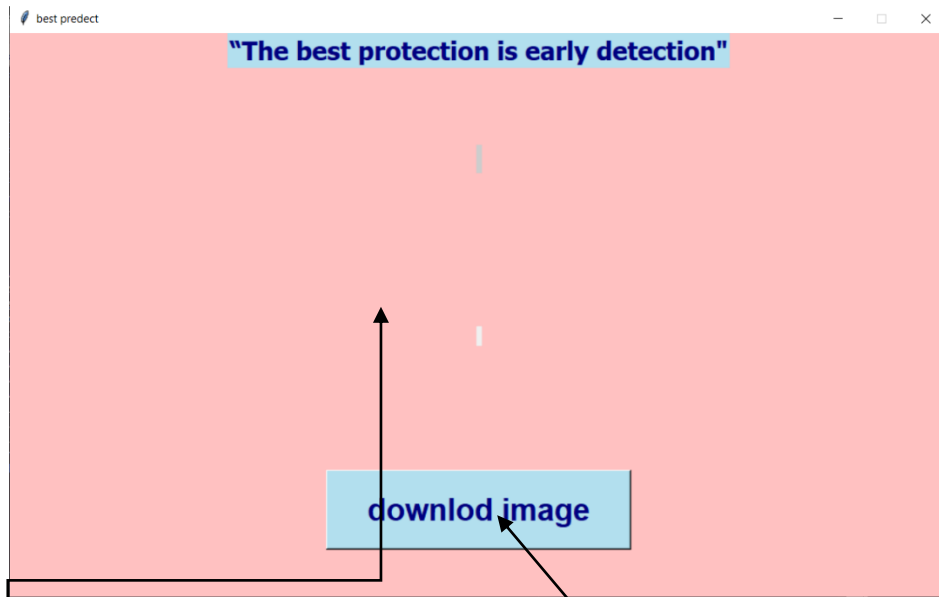


Figure 102 : Interface principale de " BreastCNNPredict"

Zone d'affichage de
de l'image

téléchargement

Le bouton Downlod Image, permet d'importer et de lire un cliché échographique de l'ensemble du test (figure 4.58).

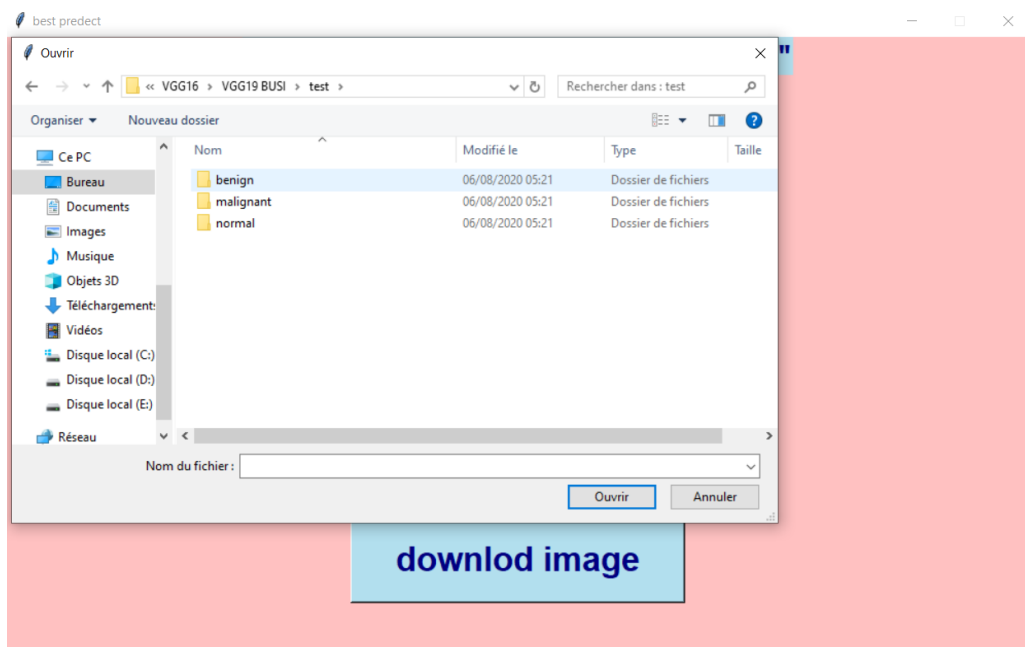


Figure 103 Choix d'un cliché

Dès que le cliché est importé, le deuxième bouton « classify image » apparait (figure 4.59).

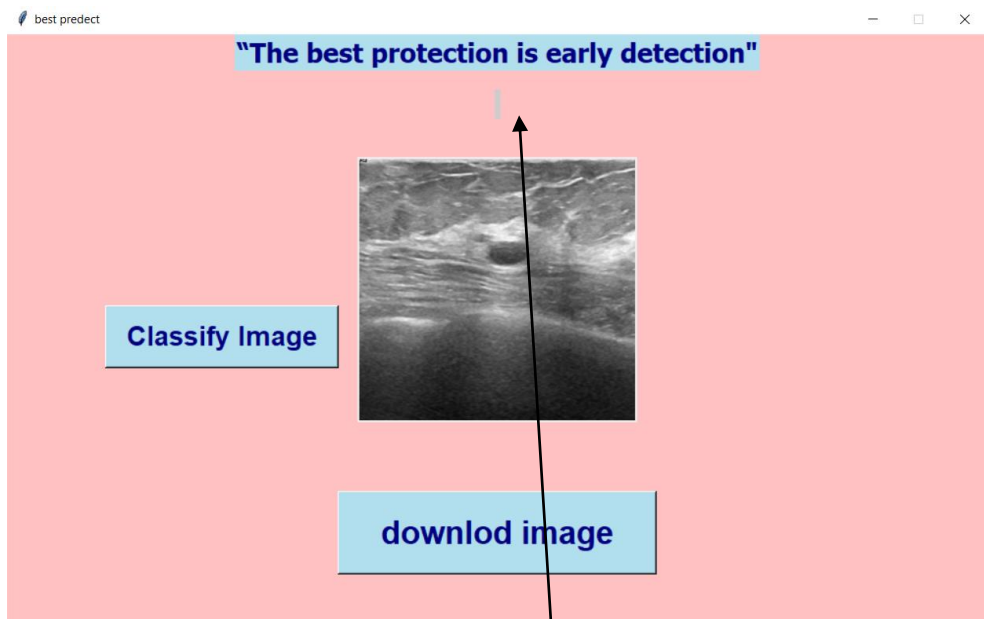


Figure 104 : Aperçu de l'exécution
Zone d'affichage du résultat

Le bouton classify image est chargé de la classification automatique de l'image en utilisant un modèle type déjà entraîné et enregistré. Les figures 4.60, 4.61 et 4.62 illustrent les résultats obtenus par l'approche 'BreastCNNPredict'.

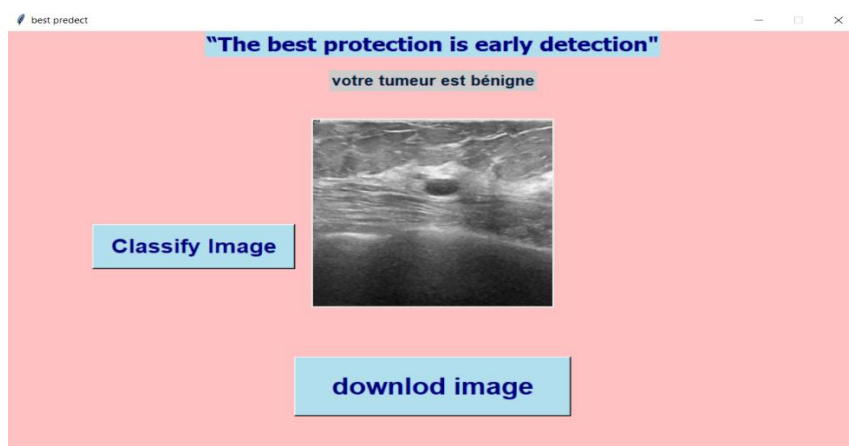


Figure 105 Classification d'une tumeur bénigne

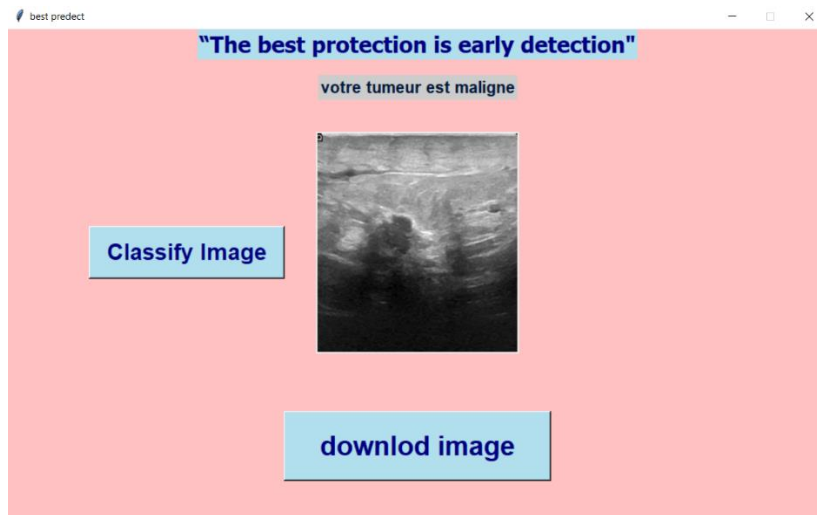


Figure 106 : Classification d'une tumeur maligne



Figure 107 : Classification d'une tumeur normale

4.12 Comparaison entre les différents modèles

Les performances des modèles développés, selon les optimiseurs et les bases échographiques adoptés, sont données dans le tableau 4.9.

DATASET	Optimiseur	ADAM	ADADELTA
	Modèle		
	XCEPTION	61%	60%

BUSI	ALEXNET	84%	68%
	VGG16	93%	91%
	VGG19	91%	87%
US	XCEPTION	62%	70%
	ALEXNET	81%	70%
	VGG16	93%	97%
	VGG19	100%	91%

Tableau 9 : Comparaison entre les différents modèles développés

L'exactitude évaluée pour chaque modèle, montre des résultats satisfaisants.

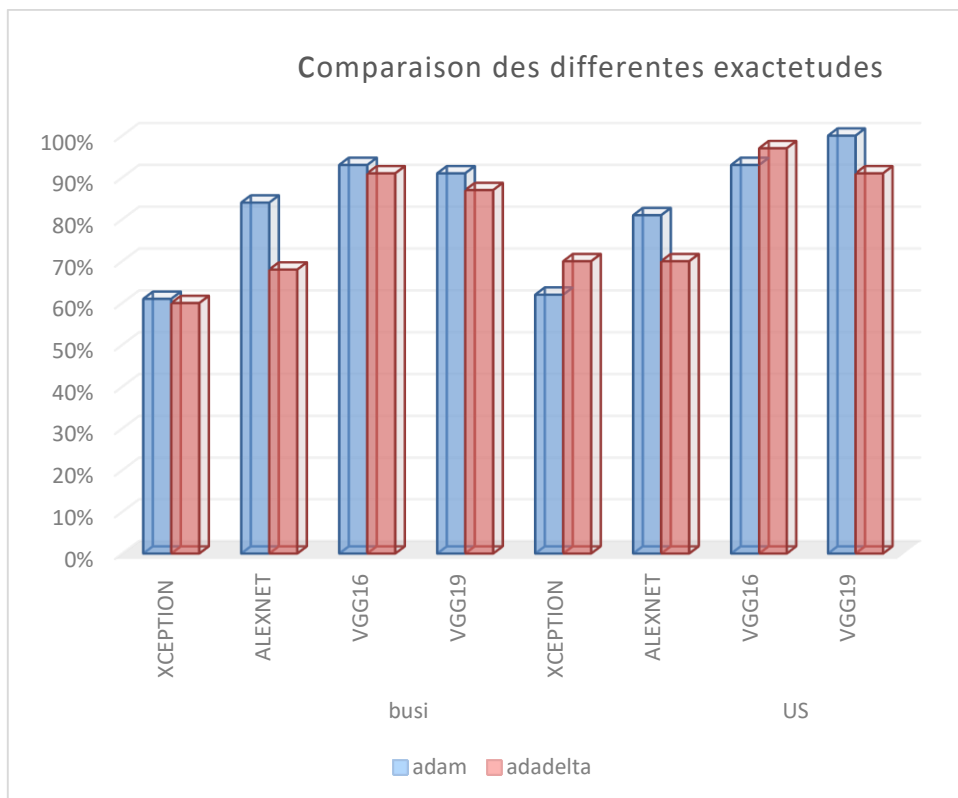


Figure 108 Exactitude des différents modèles de 'BreastCNNPredict'

4.13 Conclusion

La classification des clichés échographiques est toujours difficile, à cause de leurs natures bruitées. L'approche 'BreastCNNPredict' répond à ce souci, en utilisant quatre différentes architectures des réseaux convolutifs, appliquées à types de bases de données. L'étude comparative des performances de ces modèles, a révélé que les architectures VGG 16 et VGG19, sont les plus adaptées à notre problème.

Conclusion générale

L'interprétation des clichés échographiques représente un défi réel, pour les radiologues à cause de leurs natures bruitées. La catégorisation des tumeurs en se basant seulement sur les résultats de l'imagerie médicale, n'est pas toujours suffisante. Elle doit être complétée par un examen anatomopathologique, qui dure forcément quelques jours, voire quelques semaines. Ce qui retarderait la prise en charge médicale des patientes.

L'initiation à la recherche réalisée dans le cadre de ce mémoire, s'inscrit dans les objectifs de l'équipe, "IMAD", du laboratoire LATSI, pour la conception de systèmes d'aide au diagnostic en imagerie médicale.

Dans ce cadre, nous avons présenté et développé un système doté de l'intelligence artificielle capable de révéler la nature de la tumeur, en quelques secondes, grâce à quatre architectures différentes des réseaux de neurones convolutifs, en tenant en compte des paramètres appropriés.

L'utilisation des architectures ALEX NET et XCEPTION n'a pas apporté de résultats très satisfaisants. En effet, dans ces expérimentations, l'exactitude varie d'une valeur minimale de 20% jusqu'à une valeur maximale de 85%. Ce qui nécessite alors, l'adoption d'autres modèles tels que, VGG16 et VGG19.

Le VGG 16 appliqué sur la base BUSI, nous a permis d'obtenir, une précision de 93 % en utilisant l'optimiseur ADAM avec un taux d'apprentissage de 0.0001.

L'utilisation de VGG 19 sur la base de données « US » accorde l'obtention d'excellents résultats, retenus par une précision 100 %, en usant de l'optimiseur ADAM.

La nature des données de la base « BUSI », est plus adapté avec le modèle VGG16 qu'avec les autres réseaux. Quant à la base « US », elle est davantage compatible avec le modèle VGG19.

Les travaux futurs possibles, seraient d'hybrider deux modèles des CNN, en vue d'une multiclassification et d'explorer d'autres modalités, telles que les images IRM mammaires.

Bibliographie

- [1] OMS. Cancer du sein : prévention et lutte contre la maladie, <https://www.who.int/topics/cancer/breastcancer/fr/index3.html>, date de consultation, novembre 2019.
- [2] Anatomie du sein, cancer du sein, <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Anatomie-du-sein/>, consulté en janvier 2020.
- [3] E. Sebban, L'anatomie du sein: Les différents types de tumeurs bénignes et malignes, <https://www.docteur-eric-sebban.fr/cancer-du-sein/diagnostic-cancer-sein/anatomie-et-pathologies-du-sein/>, mars 2019, date de consultation janvier 2020.
- [4] Médecine facile. Le sein : structure et vascularisation, <https://medecinefa.wordpress.com/le-sein-structure-et-vascularisation/>, 2015 date de consultation, janvier 2020.
- [5] Lésions bénignes du sein, <https://eurekasante.vidal.fr/maladies/cancers/cancer-sein.html?pb=lesions-benignes>, date de consultation janvier 2020.
- [6] A. Le Corgne. Rôle du pharmacien d'officine dans la prise en charge du cancer du sein après chirurgie mammaire, Thèse de doctorat en pharmacie. Université de Bourgogne. Dijon France, 2016.
- [7] OMS. Cancer today, <http://gco.iarc.fr/today/home>, date de consultation, janvier 2020.
- [8] Cancer du sein, Quels signes et symptômes ? Fiches santé et conseils médicaux, <https://sante.lefigaro.fr/sante/maladie/cancer-sein/ce-que-je-ressens>, date de consultation, janvier 2020.
- [9] Stades et types du cancer du sein, vivre avec un cancer du sein métastatique, <https://www.vivreavec.eu/cancer-du-sein-metastatique/comprendre-la-maladie/stades-et-types-de-cancer-du-sein.html>, date de consultation, janvier 2020.

- [10] S. Verbeke. Etude des voies de signalisation du récepteur p75NTR impliquées dans la croissance des cellules du cancer du sein, thèse de doctorat en biologie, Université des sciences et technologie de Lille, France, 2010.
- [11]. Société canadienne du cancer. Stades du cancer du sein, <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/breast/staging/?region=on>, dates de consultation, mai 2020.
- [12] Institut national du cancer. Facteurs de risque, <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Facteurs-de-risque>, date de consultation, janvier 2020.
- [13] Cancer du sein, <https://cancer-du-sein.ooreka.fr/comprendre/depistage-cancer-sein>, date de consultation, janvier 2020.
- [14] A. André, P. Thierry, A. KAS. Biophysique. Edition Lavoisier, 2013, <https://books.google.tn/books?id=B2qkAgAAQBAJ>, date de consultation janvier 2020.
- [15] Mammographie, Diagnostic, <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Diagnostic/Mammographie>, date de consultation février 2020.
- [16] Présentation de l'échographie <http://tpe-echographie-obstetricale.e-monsie.com/pages/presentation-de-l-echographie.html>, date de consultation janvier 2020.
- [17] Verywell Health. Why Don't We Have Annual Ultrasounds Instead of Mammograms? <https://www.verywellhealth.com/why-not-annual-ultrasounds-instead-of-mammograms-430185>, date de consultation janvier 2020.
- [18] IRM - Diagnostic, <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Diagnostic/IRM>, date de consultation, février 2020.
- [19] IRM des seins, https://sites.artilab.com/radiologie_provence/893-IRM_des_seins_ou_mammaire, date de consultation janvier 2020.

- [20] Types de lésions mammaires mises en évidence en échographie, <https://info-radiologie.ch/echographie-cancer-sein.php#:~:text=%20R%C3%B4le%20de%20l%E2%80%99%C3%A9chographie%20dans%20le%20d%C3%A9pistage%20du,facteur%20de%20grossissement%20g%C3%A9om%C3%A9trique.%20l%E2%80%99%C3%A9chographie%20permet...%20More%20>, date de consultation janvier 2020.
- [21] Classification BI-RADS, <http://www.depistagesein.ca/classification-bi-rads/#.XuVtIUUzZPY>, date de consultation janvier 2020.
- [22] Les Traitements médicaux du cancer du sein, <https://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=cancer-sein-pm-traitements-medicaux-du-cancer-du-sein>, date de consultation janvier 2020.
- [23] Histoire de l'intelligence artificielle, <https://www.coe.int/fr/web/artificial-intelligence/history-of-ai>, date de consultation février 2020.
- [24] OpenClassrooms. Plongez-vous dans la peau d'un Data scientist, <https://openclassrooms.com/fr/courses/4011851-initiez-vous-au-machine-learning/4025156-plongez-vous-dans-la-peau-d-un-data-scientist>, date de consultation mars 2020.
- [25] Big Data et Data Sciences : quelles différences ? <https://www.soft-concept.com/surveymag/big-data-et-data-sciences-quelles-differences.html>, date de consultation, février 2020.
- [26] B. Le Blanc. Alan Turing : les machines à calculer et l'intelligence, La Revue 2014/1 N° 68, pages 123 à 126, Hermès, 2014.
- [27] Histoire de l'intelligence artificielle, <https://siecldigital.fr/2018/08/20/histoire-intelligence-artificielle/>, date de consultation, mars 2020.
- [28] Deux types d'Intelligence Artificielle, <https://sites.google.com/site/iatpe2011/1ere-partie-fonctionnement-et-caracteristiques-de-l-deux-types-d-ia>, date de consultation, mars 2020.

[29] : L'IA d'Alibaba pourrait détecter le coronavirus en quelques secondes, <https://www.futura-sciences.com/tech/actualites/intelligence-artificielle-ia-alibaba-pourrait-detecter-coronavirus-quelques-secondes-79873/>, juillet 2020, date de consultation août 2020.

[30] A.Benmansour, S.Oudinat , S.Bouzoina, L'aide de l'intelligence artificielle à la prise de décision (Cas d'étude : Classification des tubes PEHD de la société Canal Plast Groupe Kherbouche) , mémoire de master en génie électrique, Université Abou Bakr Belkaid Tlencen, Tlencen, 2017.

[31] Présentation & historique des réseaux neuronaux, <https://www.futura-sciences.com/tech/dossiers/robotique-presentation-historique-reseaux-neuronaux-31/>, date de consultation, avril 2020.

[32] R. Mafouf, Amélioration de l'apprentissage d'un modèle neuronal pour la reconnaissance des anomalies cardiaques, mémoire de magistère en automatique, université Abou Bakr Belkaid Tlencen, Tlencen, 2010.

[33] Coursera. Réseaux de neurones et deep learning, <https://www.coursera.org/learn/neural-networks-deep-learning>, date de consultation, avril 2020.

[34] S. Ronaghan. Deep Learning: Which Loss and Activation Functions should I use? <https://towardsdatascience.com/deep-learning-which-loss-and-activation-functions-should-i-use-ac02f1c56aa8>, date de consultation mars 2020.

[35] Adrian Roser Brock. Deep Learning for Computer vision, 1^{er} Edition, 2017.

[36] Le big data. La matrice de confusion, l'outil de mesure des performances du machine learning, <https://www.lebigdata.fr/confusion-matrix-definition>, consulté en juillet 2020.

[37] Aquila data enabler. Classification d'images et détection d'objets par CNN, <https://www.aquiladata.fr/classification-dimages-et-detection-dobjets-par-cnn/> date de consultation, août 2020

- [38] OpenClassrooms. Analysez le filtrage spatial et la convolution par masque, <https://openclassrooms.com/fr/courses/5060661-initiez-vous-aux-traitements-de-base-des-images-numeriques/5217251-analysez-le-filtrage-spatial-et-la-convolution-par-masque>, date de consultation, mai 2020.
- [39] Openclassrooms. Découvrez les différentes couches d'un CNN, <https://openclassrooms.com/fr/courses/4470531-classez-et-segmentez-des-donnees-visuelles/5083336-decouvrez-les-differentes-couches-dun-cnn>, consulté en juillet 2020.
- [40] ImageNet, <http://www.image-net.org>, date de consultation mai 2020.
- [41] Siddharth Das et Analytics Vidhya, CNN Architectures : LeNet, AlexNet, VGG,GoogLeNet, ResNet and more, 2017, [https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5#:~:text=Top%20highlight-.CNN%20Architectures%3A%20LeNet%2C%20AlexNet%2C%20VGG,%2C%20GoogLeNet%2C%20ResNet%20and%20more%E2%80%A6&text=A%20Convolutional%20Neural%20Network%20\(CNN,pixel%20images%20with%20minimal%20preprocessing](https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5#:~:text=Top%20highlight-.CNN%20Architectures%3A%20LeNet%2C%20AlexNet%2C%20VGG,%2C%20GoogLeNet%2C%20ResNet%20and%20more%E2%80%A6&text=A%20Convolutional%20Neural%20Network%20(CNN,pixel%20images%20with%20minimal%20preprocessing) , juillet 2020.
- [42] Understanding AlexNet, <https://www.learnopencv.com/understanding-alexnet/>, date consultation, juin 2020.
- [43] CNN VGG 16 and VGG 19, Master Data Science, <http://datahacker.rs/deep-learning-vgg-16-vs-vgg-19/>, date de consultation, mai 2020.
- [44] Xception Model and Depthwise Separable Convolutions, <https://maelfabien.github.io/deeplearning/xception/>, date de consultation juillet 2020.
- [45] Keras, the documentation, <https://keras.io/applications/>, date de consultation, juin 2020.
- [46] M. El Aidouni. Understanding YOLO and YOLOv2, manalelaidouni.github.io/Understanding%20YOLO%20and%20YOLOv2.html, date de consultation, juin 2020.

- [47] P. Marcelino. Transfer learning from pre-trained models, <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>, consulté en juillet 2020.
- [48] W. Al-Dhabyani, M. Goma, H. Khaled, A. Fahmy. Deep Learning Approaches for Data Augmentation and Classification of Breast Masses using Ultrasound Images. Int J Adv Comput Sci Appl; Vol. 10, N°5, 2019.
- [49] BUSI dataset. <https://scholar.cu.edu.eg/?q=afahmy/pages/dataset>, consulté en janvier 2020.
- [50] Z. Cao, L. Duan, G. Yang, T. Yue Q. Chen. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. BMC Med Imaging <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6604293/>, juillet 2019.
- [51] Yap MH, Pons G, Martí J, Ganau S, Sentís M, Zwigelaar R, et al. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. IEEE J Biomed Health Inform. juill 2018;22(4):1218-26.
- [52] M. A. R. Nayeem, M.A. M. Joadder, S. A. Shetu, F. R. Jamil, A. Al Helal. Feature selection for breast cancer detection from ultrasound images <https://www.researchgate.net/publication/269032490> Feature selection for breast cancer detection from ultrasound images.
- [53] Machine Learning. Quelle est la meilleure bibliothèque Python et pourquoi faire ? <https://fr.bitdegree.org/tutos/bibliotheque-python>, date de consultation, juillet 2020.
- [54] Anaconda. Your data science toolkit, <https://www.anaconda.com/products/individual>, consulté en janvier 2020.
- [55] Spyder, <https://www.spyder-ide.org/>, consulté en janvier 2020.
- [56] R. Vickery. Beginners Guide to Deep Learning with TensorFlow, <https://towardsdatascience.com/beginners-guide-to-deep-learning-with-tensorflow-ca85969b2f2>, consulté en mai 2020.

[57] S. Pattanayak. Pro Deep Learning with TensorFlow : A Mathematical Approach to Advanced Artificial Intelligence in Python. A press, 2017.

[58] S. Bancal. Introduction à Numpy, <https://enacit.epfl.ch/cours/python/scientifique/numpy.html#numpy>, consulté en janvier 2020.

[59] Scikit-learn, machine learning in python, <https://scipy.org>, date de consultation, juin 2020.

[60] 15 bibliothèques Python pour la science des données, <https://moncoachdata.com/blog/python-pour-la-science-des-donnees>, date de consultation, juin 2020.

[61] P. S. Rodrigues. Breast Ultrasound Image, Mendeley Data, <http://dx.doi.org/10.17632/wmy84gzngw.1>, (2017)

[62] Understanding Categorical Cross-Entropy Loss, Binary Cross-Entropy Loss, Softmax Loss, Logistic Loss, Focal Loss and all those confusing names, https://gombru.github.io/2018/05/23/cross_entropy_loss/, date de consultation, Janvier 2020.