

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

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

جامعة سعد دحلب 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

Salem Nesrine

pour l'obtention du diplôme de Master en Électronique

Option 'Automatique

et

Informatique Industrielle'

Thème

Analyse profonde d'images ultrasonores pour la détection du cancer du sein

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

Année Universitaire 2019-2020

Remerciements

Avant tout, je rends grâce à ALLAH de m'avoir accordé la volonté et le courage pour réaliser ce mémoire.

Le travail présenté a été effectué sous la direction de Mlle Reguieg F.Zohra à qui je tiens à adresser mes plus vifs remerciements, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter ma réflexion et ses encouragements lors de la réalisation de ce mémoire. C'est très réconfortant de pouvoir partager des idées et discuter ouvertement de nouvelles façons d'aborder un problème avec une telle facilité.

Je remercie aussi, notre co-promotrice Mlle Benblidia Nadja, pour sa disponibilité, son aide qui nous a été très utile et ses encouragements.

Mes remerciements s'adressent aussi à tous les membres du jury pour l'honneur qu'ils m'ont fait en acceptant de juger ce travail.

Je remercie également tous les collègues qui m'ont accompagné tout au long de ces années, et en particulier ceux qui ont rempli ces deux dernières années de tant de joie, de rires et de camaraderie

Ces remerciements vont tout d'abord au corps professoral et administratif de l'Université Saad Dahleb Blida (USDB), plus particulièrement au corps professoral du département de l'électronique, pour la richesse et la qualité de leur enseignement et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée.

Je dédie ce mémoire :

- A mes parents, pour le soutien moral, matériel et financier durant ces années d'étude ; je ne serais pas comme cela sans vous. Que Dieu vous bénisse. Je vous aime très fort.

- A mes sœurs, mes piliers merci pour tout. Que Dieu vous bénisse et exauce toutes vos prières.

- A ma famille et mes amis.

- A tous ceux qui ont participé de près ou de loin à la réalisation de ce travail, merci du fond du cœur.

Que Dieu vous bénisse tous !

ملخص: في الوقت الحاضر، يعتبر نقل التعلم حلاً جيداً لمشاكل تعلم الشبكات العميقة من قواعد البيانات الطبية الصغيرة، إذا كانت هناك درجة عالية من التنوع البصري بين الفئات. إن تقنية نقل التعلم لا تتجاوز الأداء غير المرضي للنهج التقليدية فحسب، بل إنها تكسر حاجز العينات المحدودة لبناء الشبكات العصبية التلافيفية العميقة. في هذا العمل، تبيننا هذا المنهج، بهدف توسيع قدرات أحدث شبكات CNN، لتشمل مشكلة تصنيف الأمراض في الموجات فوق الصوتية للثدي. لقد درسنا استخدام الشبكات العصبية العميقة المُدرّبة مسبقاً وهي VGG19 و VGG16 و Inception V3؛ التي تستغل التسلسل الكلاسيكي لطبقات التلافيف والتجميع و كتل Inception. تم تنقيح النموذج المعتمد وفقاً لاستراتيجيتين: الضبط الدقيق الكامل والضببط الدقيق الجزئي. أظهرت النتائج أن Inception V3 يتفوق على VGG16 و VGG19 بدقة 94.64٪ و AUC 0.98.

الكلمات المفاتيح: تصوير الثدي بالموجات فوق الصوتية، الشبكات العصبية التلافيفية، الضبط الدقيق، Inception V3، VGG19، VGG16.

Résumé : De nos jours, l'apprentissage par transfert est considéré comme une bonne solution aux problèmes de l'apprentissage des réseaux profonds à partir des données médicales de taille réduite, en cas de présence d'une grande diversité visuelle intra-classe. La technique de l'apprentissage par transfert permet non seulement de surmonter les performances insatisfaisantes des approches traditionnelles, mais brise également l'obstacle des échantillons limités pour la construction de réseaux de neurones convolutifs profonds. Dans ce travail, nous adoptons ce paradigme, dans l'objectif d'étendre les capacités des CNN les plus récentes, au problème de la classification de pathologies en échographie mammaire. Nous avons étudié les réseaux de neurones profonds pré-entraînés en l'occurrence VGG19, VGG16 et Inception V3 ; qui exploitent l'enchaînement classique des couches de convolutions, de pooling et les blocs Inception. Le modèle adopté est affiné selon deux stratégies : le fine-tuning total et le fine-tuning partiel. Les résultats démontrent que l'inception V3, surpasse le VGG16 et le vgg19, avec une précision de 94,64% et une AUC de 0,98.

Mots clés : Echographie mammaire, Réseaux de Neurones Convolutifs, Fine-tuning, Inception V3, VGG19, VGG16.

Abstract: Today, transfer learning is considered as a good solution to the problems of learning of deep networks from small medical databases, if there is a high degree of intra-class visual diversity. The technique of transfer learning not only overcomes the unsatisfactory performance of traditional approaches, but also breaks the barrier of limited samples for the construction of deep convolutional neural network. In this work, we have adopted this paradigm, with the aim of extending the capabilities of the most recent CNNs to the problem of the classification of pathologies, in breast ultrasound. We studied the use of the pre-trained deep neural networks VGG19, VGG16 and Inception V3, which exploit the classic sequence of convolution, pooling layers and inception blocks. The adopted models are then refined according to two strategies: total fine tuning and partial fine tuning.

The results show that inception V3 outperforms VGG16 and vgg19, with an accuracy of 94.64% and a AUC of 0.98.

Keywords: Breast ultrasound, Deep convolutional neural network, Fine tuning, Inception V3, VGG19, VGG16.

Table des matières

INTRODUCTION GENERALE	1
CHAPITRE I CONTEXTE MEDICAL	3
1.1 Introduction	3
1.2 Anatomie du sein.....	3
1.3 Drainage lymphatique	5
1.4 Cancer du sein	6
1.4.1 Description des stades évolutifs des cancers du sein	7
1.4.2 Symptomatologie.....	8
1.4.3 Facteurs de risque	8
1.5 Dépistage, Diagnostic et Traitement	9
1.6 Imagerie du sein	10
1.6.1 Imagerie par résonance magnétique (IRM).....	10
1.6.2 Mammographie	11
1.6.3 Echographie mammaire	15
1.7 Anomalies mammaires	18
1.7.1 Masses.....	18
1.7.2 Calcifications mammaires.....	20
1.8 Classification des pathologies mammaires	21
1.8.1 Classification de Le Gal	21
1.8.2 Classification BI-RADS de l'ACR.....	22
1.9 Conclusion.....	23
CHAPITRE II APPRENTISSAGE PROFOND EN IMAGERIE MEDICALE	24
2.1 Introduction	24
2.2 Apprentissage Automatique	25
2.2.1 Différents types d'apprentissage automatique	25
2.2.2 Métriques d'évaluation	28
2.2.3 Fonction d'erreur	29
2.2.4 Descente stochastique de gradient.....	30
2.2.5 Quelques algorithmes d'apprentissage automatique.....	31
2.3 Réseaux de neurones artificiels.....	33
2.3.1 Principe.....	33
2.3.2 Historique	34
2.3.3 Neurone biologique.....	34

2.3.4	Neurone formel.....	35
2.3.5	Structure d'interconnexion.....	36
2.4	Apprentissage profond	39
2.4.1	Histoire de l'apprentissage profond.....	39
2.4.2	Pourquoi l'apprentissage profond ?.....	41
2.5	Architectures des réseaux de neurones en apprentissage profond.....	42
2.5.1	Réseaux de neurones récurrents	42
2.5.2	Réseaux de neurones convolutifs	44
2.6	Apprentissage profond et l'imagerie médicale	49
2.6.1	Classification	49
2.6.2	Détection et localisation.....	50
2.6.3	Segmentation.....	50
2.7	Quelques travaux réalisés en imagerie échographique du sein	50
2.8	Conclusion.....	52
CHAPITRE III CONCEPTION DU SYSTEME D'AIDE A LA DETECTION DU CANCER SEIN DANS LES IMAGES ULTRASONORES.....		53
3.1	Introduction	53
3.2	Présentation du système d'aide à la détection et la classification de pathologies en échographie du sein :	53
3.2.1	Architecture du système développé	54
3.2.2	Exploration des données	55
3.2.3	Prétraitement des images	57
3.2.4	Augmentation des données.....	57
3.3	Présentation de l'architecture VGG19.....	58
3.4	Présentation de l'architecture VGG16.....	59
3.5	Présentation de l'architecture Inception v3	61
3.5.1	Factorisation des convolutions	61
3.6	Apprentissage par transfert	65
3.6.1	Branches du Transfer Learning	65
3.6.2	Réutilisation d'un modèle pré-entraîné	66
3.7	Conclusion.....	68
CHAPITRE IV MISE EN ŒUVRE DU SYSTEME ET RESULTATS		69
4.1	Introduction	69
4.2	Environnement de travail	69
4.2.1	Matériel utilisé.....	69
4.2.2	Langage de programmation Python.....	69

4.2.3	Anaconda.....	70
4.2.4	TensorFlow	72
4.2.5	Keras.....	73
4.2.6	Numpy	74
4.2.7	Flask.....	74
4.2.8	Html.....	74
4.2.9	Tkinter	75
4.3	Base de données	75
4.4	Architecture du système 'BreastULTRASCNN'	76
4.5	Résultats et discussion.....	78
4.6	Discussion	103
4.7	Interface de BreastULTRASCNN	105
4.8	Application web.....	106
4.9	Conclusion.....	108
	Conclusion générale	109
	Références bibliographiques	111

Liste des figures

Figure 1.1 : Paroi thoracique et glande mammaire : coupe sagittale	4
Figure 1.2 : Vascularisation de la glande mammaire	5
Figure 1.3 : Niveaux du curage mammaire externe (étages ganglionnaires de Berg) et les ganglions mammaires internes.....	6
Figure 1.4 : Appareil utilisé pour l'IRM mammaire	11
Figure 1.5 : Positionnement de la patiente pour la prise d'une mammographie en face	11
Figure 1.6 : Visualisation lever du bras pour oblique médio latérale (MLO)	13
Figure 1.7 : Visualisation cranio-caudale (CC)	13
Figure 1.8 : Comparaison entre deux images de la même patiente en 2D et 3D	14
Figure 1.9 : Appareil d'ultrasonographie	15
Figure 1.10 : Différentes formes de masses.....	18
Figure 1.11 : Différents types de contour des opacités selon le BI-RADS : (a) contour circonscrit, (b) contour masqué, (c) contour micro-lobulé, (d) contour mal défini et (e) contour spiculé.....	19
Figure 1.12 : Types de densité des masses selon le BI-RADS : (a) B1 (densité < 25%), (b) B2 (densité comprise entre 25 et 50%), (c) B3 (densité comprise entre 50 et 75%) et (d) B4 (densité > 75%)	19
Figure 1.13 : Classification des microcalcifications	21
Figure 2.1 : Différents domaines de l'intelligence artificielle	25
Figure 2.2 : Schéma de principe de l'apprentissage par renforcement.....	27
Figure 2.3 : Matrice de confusion dans le cas binaire	28
Figure 2.4 : Courbe Roc.....	29
Figure 2.5 : Représentation des données d'entrée par rapport à un classificateur linéaire	31
Figure 2.6 : Exemple d'un arbre de décision avec 4 classes.....	32
Figure 2.7 : Exemple des KNN	33
Figure 2.8 : neurone biologique	35
Figure 2.9 : Schéma d'un neurone à 3 entrées	36
Figure 2.10 : Exemple d'un réseau de neurones monocouche.....	37
Figure 2.11 : Réseau à connexions locales.....	38
Figure 2.12 : Exemple d'un réseau neurone à deux couches cachées	38
Figure 2.13 : Relation entre l'intelligence artificielle, le ML et le deep learning	39
Figure 2.14 : Différence de performances entre le DL et le ML en fonction de la quantité de donnée.....	41
Figure 2.15 : Procédé de l'apprentissage classique comparé à celui du deep learning	42
Figure 2.16 : Réseaux neuronaux récurrents ont des boucles	43
Figure 2.17 : Réseau neuronal récurrent	43
Figure 2.18 : Schéma du parcours de la fenêtre du filtre de convolution sur l'image.....	45
Figure 2.19 : Exemple d'une convolution.....	46
Figure 2.20 : Exemple d'un max pooling.....	47
Figure 2.21 : Fonction ReLU	48
Figure 3.1 : Schéma du modèle adopté	54
Figure 3.2 : Utilisation des différents datasets dans l'entraînement du modèle final	57

Figure 3.3 : Architecture de VGG19.....	59
Figure 3.4 : Architecture du VGG16.....	60
Figure 3.5 : Convolution 5x5 selon deux convolution 3x3	62
Figure 3.6 : Inception module A	62
Figure 3.7 : Convolution 3x3 suivant une convolution 3x1 et une autre 1x3	63
Figure 3.8 : Inception module B	63
Figure 3.9 : Inception module C	64
Figure 3.10 : Architecture du réseau Inception V3	64
Figure 3.11 : Types de Transfer Learning	66
Figure 3.12 : Stratégies du fine-tuning	68
Figure 4.1 : Représentation de l'interface d'anaconda	71
Figure 4.2 : Comparaison entre les différentes bibliothèques du deep learning.....	72
Figure 4.3 : arXiv mentions, october 2017.....	74
Figure 4.4 : Illustration des images présentes dans la base BUSI (a) bénigne, (b) maligne, (c) normale	75
Figure 4.5 : Illustration du modèle de base	78
Figure 4.6 : Evolution de l'erreur de test avec l'algorithme momentum sur 45 époques.....	81
Figure 4.7 : Modèle 2.....	82
Figure 4.8 : Évolution de l'erreur de test avec une seule couche FC (modèle 1) et 7 couches FC (modèle 2).....	82
Figure 4.9 : Évolution de l'erreur du test et d'apprentissage	83
Figure 4.10 : Modèle 3.....	84
Figure 4.11 : Évolution de l'erreur sans dropout (modèle 2) et avec dropout (modèle 3).....	85
Figure 4.12 : Évolution de la précision sans dropout (modèle 2) et avec dropout (modèle 3).....	85
Figure 4.13 : ReLU et ELU (ELU, $\alpha = 1.0$)	86
Figure 4.14 : Le modèle 4.....	87
Figure 4.15 : Évolution de l'erreur et l'évolution du taux de bonne classification du modèle 4	87
Figure 4.16 : Modèle 5	88
Figure 4.17 : Évolution de l'erreur et l'évolution du taux de bonne classification du modèle 5	90
Figure 4.18 : Aperçu sur le code source pour batch normalisation	91
Figure 4.19 : Modèle 6.....	92
Figure 4.20 : Évolution de l'erreur et du taux de bonne classification du modèle 6.....	92
Figure 4.21 : Matrices de confusion : (a) modèles 5 et (b) modèle 6	93
Figure 4.22 : Modèle 7.....	95
Figure 4.23 : Évolution de l'erreur et du taux de bonne classification du modèle 7.....	96
Figure 4.24 : Modèle 8.....	97
Figure 4.25 : Évolution de l'erreur et du taux de bonne classification du modèle 8.....	97
Figure 4.26 : Matrices de confusion : (a) modèles 7 et (b) modèle 8	98
Figure 4.27 : Modèle 9.....	99
Figure 4.28 : Modèle 10.....	99
Figure 4.29 : Évolution de l'erreur et du taux de bonne classification du modèle 9.....	100
Figure 4.30 : Évolution de l'erreur et du taux de bonne classification du modèle 10.....	100
Figure 4.31 : Matrices de confusion : (a) modèle 9 et (b) modèle 10.....	101
Figure 4.32 : Courbe ROC du modèle final.....	105
Figure 4.33 : Interface générale de 'BreastULTRASCNN'	106
Figure 4.34 : Exemple d'application de 'BreastULTRASCNN'	106
Figure 4.35 : Page principale partie 1	107
Figure 4.36 : Page principale partie 2	107

Liste des tableaux

Tableau 1.1 : Atténuation radiologique des composants mammaires.....	12
Tableau 1.2 : Classification de Le Gal	21
Tableau 1.3 : Classification BIRADS De L'ACR	22
Tableau 2.1 : Evolution du Deep Learning	41
Tableau 3.1 : Techniques d'augmentations des données selon le type	58
Tableau 4.1 : Répartition par classe de la base de données.....	75
Tableau 4.2 : Paramètres des couches ajoutées au modèle VGG19 avec le fine-tuning total	76
Tableau 4.3 : Paramètres des couches ajoutées au modèle VGG19 avec le fine-tuning partiel	76
Tableau 4.4 : Paramètres des couches ajoutées au modèle VGG16 avec le fine-tuning total	77
Tableau 4.5 : Paramètres des couches ajoutées au modèle VGG16 avec le fine-tuning partiel	77
Tableau 4.6 : Paramètres des couches ajoutées au modèle Inception V3 avec le fine-tuning total	78
Tableau 4.7 : Paramètres des couches ajoutées au modèle Inception V3 avec le fine-tuning partiel..	78
Tableau 4.8 : Grandes lignes de l'architecture de réseau proposée	79
Tableau 4.9 : Erreur d'exécution de chaque algorithme après 45 époques	80
Tableau 4.10 : Erreur, précision et justesse de l'algorithme lors de la phase de test après 45 époques	81
Tableau 4.11 : Tableau récapitulatif des résultats précédents après 45 époques.....	82
Tableau 4.12 : Tableau récapitulatif des résultats précédent après 45 époque	85
Tableau 4.13 : Tableau récapitulatif de l'erreur d'apprentissage après 45 époques.....	88
Tableau 4.14 : Tableau récapitulatif des résultats précédents après 45 époques.....	88
Tableau 4.15 : Grandes lignes de l'architecture de réseau proposée	89
Tableau 4.16 : tableau récapitulatif des résultats précédent après 45 époque	90
Tableau 4.17 : Tableau récapitulatif des résultats précédents après 45 époques.....	92
Tableau 4.18 : Matrice de confusion avec 3 classes.....	93
Tableau 4.19 : Spécificité et sensibilité pour 3 classes	94
Tableau 4.20 : Spécificité et sensibilité des modèles 5 et 6	94
Tableau 4.21 : Grandes lignes de l'architecture du réseau proposée	96
Tableau 4.22 : Tableau récapitulatif des résultats précédents après 45 époques.....	96
Tableau 4.23 : Tableau récapitulatif des résultats précédent après 45 époques	97
Tableau 4.24 : Spécificité et sensibilité des modèles 7 et 8	98
Tableau 4.25 : Grandes lignes de l'architecture du réseau proposée	100
Tableau 4.26 : Tableau récapitulatif des résultats précédents après 45 époques.....	100
Tableau 4.27 : Spécificité et sensibilité des modèles 9 et 10.....	101
Tableau 4.28 : Comparaison des performances des modèles 'BreastULTASCNN'.....	102
Tableau 4.29 : Performances finales de notre modèle sur l'ensemble de test	104

Acronymes

ADAM Adaptive Moments.

CNN Convolutional Neural Network.

ELU Exponential Linear Units.

IA Intelligence Artificielle.

KNN K-Nearest Neighbors.

ML Machine Learning.

MLP Multi Layer Perceptron.

ReLU Rectified Linear Unit.

RNN Recurrent Neural Network.

RMSProp Root Mean Square Propagation

SGD Stochastic Gradient Descent.

SVM Support Vector Machine.

tanh Hyperbolic tangent.

Introduction générale

Ces dernières années, le cancer du sein est devenu un problème majeur de la santé dans le monde. C'est le cancer le plus fréquent chez les femmes, avec plus de 800 000 nouveaux cas diagnostiqués chaque année dans le monde [1]. Il représente la première cause de mortalité féminine dans la tranche d'âge de 35 à 55 ans, constituant donc un sérieux problème de santé publique. Il est responsable de 20% des décès dus aux cancers. En Algérie, le cancer du sein représente 30% des cas de cancer recensés, et représente près de 50 % des cancers gynécologiques chez la femme [2]. Au cours de ces 20 dernières années, l'incidence du cancer du sein a été multipliée par 3. En raison de son diagnostic tardif, il en résulte généralement un traitement lourd, mutilant et coûteux qui s'accompagne d'un taux de mortalité élevé. Dans le but d'éviter des traitements lourds, de réduire la morbidité et la mortalité par le cancer du sein, une détection précoce est nécessaire motivant ainsi des campagnes de dépistage chez les femmes d'un certain groupe d'âge, entre 40 et 50 ans selon les pays [1].

L'échographie est une technique d'imagerie médicale qui utilise des ondes ultrasonores, pour générer des images relativement précises. Cet examen est considéré, comme un examen complémentaire à la mammographie. En effet, il permet au radiologue de mieux interpréter et visualiser les anomalies, mises en évidence par une mammographie. De plus, la sensibilité de cette technique n'est pas affectée, si la densité mammaire est élevée. L'interprétation des clichés en un temps limité pour la détection précise des pathologies, est une tâche très difficile par un examen échographique ; cela nécessite alors l'utilisation d'outils d'aide à la détection par ordinateur (CAD) pour améliorer les performances du lecteur.

Notre travail consiste à réaliser un système d'aide à la détection, pour une analyse par apprentissage profond d'images ultrasonores mammaires, catégorisant les classes malignes, bénignes ou normales. En effet, le deep learning, est en train de révolutionner l'imagerie médicale, en offrant aux radiologues, la possibilité d'identifier avec précision toute anomalie, même la moins perceptible, par les méthodes classiques.

Le succès actuel des réseaux de neurones convolutifs (CNN) dans la classification d'images médicales, nous incite à les explorer dans le cadre de ce mémoire. Contrairement aux méthodes classiques d'apprentissage par la machine, où les caractéristiques sont déterminées par des méthodes traditionnelles, les CNN permettent de les extraire directement de la base de données d'apprentissage, réduisant ainsi les fausses détections. Dans ce cadre, l'approche à concevoir, s'effectue en se basant sur des modèles CNN pré-entraînés. Nous étudions à cet effet, différentes architectures des réseaux convolutifs, telles que le VGG16 et le VGG19, avec un affinement des paramètres d'apprentissage, à la recherche du modèle le plus performant, pour l'analyse des données échographiques du sein.

L'objectif final de cet outil est d'assister le radiologue, dans la tâche parfois délicate, d'interprétation des images échographiques mammaires.

Pour mener à bien ce travail, notre mémoire est organisé selon quatre chapitres :

- Le premier chapitre est consacré au contexte médical. Nous y décrivons, l'anatomie du sein, les différentes pathologies mammaires, ainsi que l'imagerie correspondante.
- Le deuxième chapitre est consacré aux concepts de l'intelligence artificielle (IA) en passant brièvement par l'apprentissage automatique, pour décrire plus en détail, l'apprentissage profond et les différents modèles des réseaux connexionnistes profonds.
- La conception de notre approche de classification de pathologies en échographie du sein, basée sur les réseaux profonds pré-entraînés est présentée dans le troisième chapitre.
- Dans le quatrième chapitre, nous présentons les expérimentations effectuées sur des données échographiques mammaires réelles.
- Une conclusion générale, clôture le mémoire.

Chapitre 1

Contexte médical

1.1 Introduction

Le cancer du sein est le cancer le plus répandu chez les femmes à travers le monde ; il constitue la cause majeure du décès féminin [1, 2]. Néanmoins, le dépistage précoce de ce cancer augmente le taux de survie et les options de traitement. L'échographie mammaire constituant l'examen de référence pour l'exploration des tissus des seins, fait suite à la mammographie pour un bilan sénologique complet.

Généralement, c'est le radiologue qui, en se basant sur son expérience, analyse et interprète les images échographiques. Cependant, des études ont montré que les radiologues manquent entre 10% et 30% des cas de cancers du sein, malgré leur expertise dans le domaine. C'est pourquoi on a besoin de recourir, à des systèmes d'aide au diagnostic. Ces derniers guident les radiologues durant le processus de dépistage, ce qui permet une meilleure analyse des cas complexes.

Dans ce chapitre, nous définirons la notion du cancer du sein ainsi que les différentes pathologies mammaires. Nous abordons par la suite le principe de l'imagerie mammographique, l'échographie et nous décrivons ainsi les anomalies détectables, pour situer le contexte de notre application.

1.2 Anatomie du sein

Le sein est une glande exocrine d'origine ectodermique. Il présente une quinzaine de canaux galactophores qui se distribuent dans autant de lobes (eux-mêmes multilobulés) et qui s'abouchent indépendamment, au niveau du mamelon et de l'aréole. Cette glande est également composée de graisse en quantité plus ou moins importante, en fonction de l'âge, du statut hormonal et du morphotype. La glande mammaire est attachée à la paroi thoracique, sur le muscle grand pectoral, par la bourse séreuse de Chassaignac.

Les rapports anatomiques en surface se font avec la peau et, en profondeur avec les muscles : grand et petit pectoral. Plus en profondeur et, surtout, à la partie inférieure, le sein est en rapport avec les côtes et les muscles inter costaux expliquant le risque de brèche pleurale, en cas de dissection trop profonde [4].

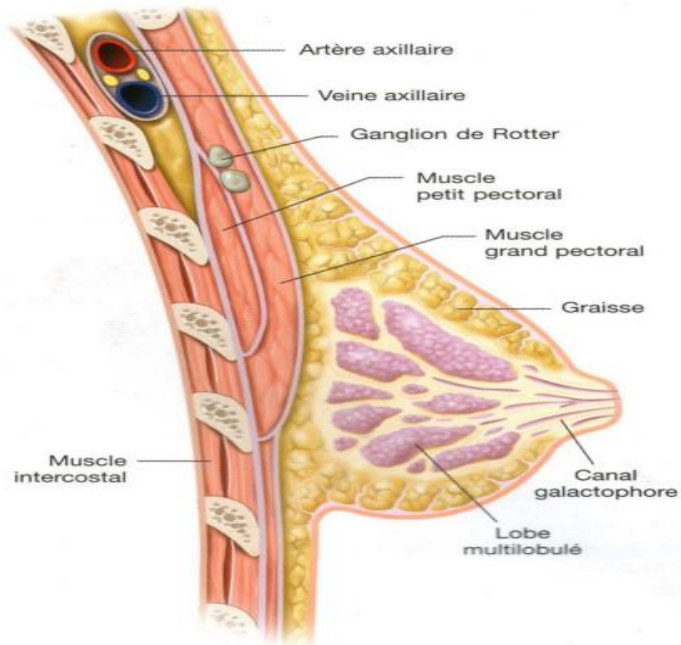


Figure 1.1 : Paroi thoracique et glande mammaire : coupe sagittale [4]

Le sein est parcouru de vaisseaux sanguins (figure 1.2) et de vaisseaux lymphatiques. Les ganglions et les vaisseaux lymphatiques, composent le système lymphatique qui aide à combattre les infections [4].

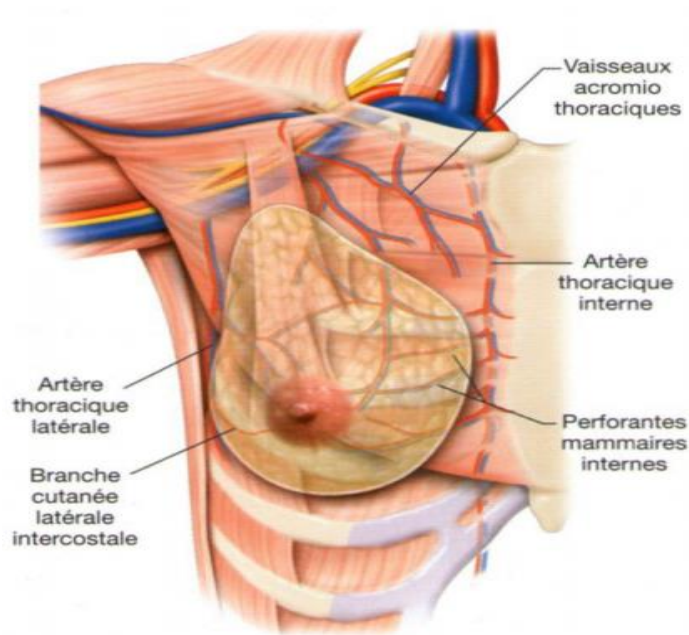


Figure 1.2 : Vascularisation de la glande mammaire [3]

1.3 Drainage lymphatique

Le drainage lymphatique est constitué :

- d'un riche réseau cutané dans la couche profonde du derme ;
- d'un réseau glandulaire profond associé aux lobes ;
- d'un réseau péri-aréolaire (cercle de Sappey), sur lequel repose l'anastomose entre les deux premiers réseaux.
- du cercle péri-aréolaire partent 4 canaux collecteurs qui se dirigent vers l'aisselle, principalement vers le groupe ganglionnaire thoracique latéral. Les ganglions lymphatiques du sein convergent donc vers 2 groupes essentiels : les ganglions axillaires et les ganglions parasternaux.

Les 3 étages ganglionnaires classiques de Berg (figure 1.3) s'intègrent dans trois niveaux. Le niveau I est en dessous du muscle pectoral, le niveau II se trouve en arrière et, le niveau III au-dessus du tendon du petit pectoral. Les ganglions axillaires et parasternaux, se jettent dans les ganglions de second niveau supraclaviculaires. L'espace inter pectoral (figure 1.1) entre les muscles grand pectoral et petit pectoral est le siège des ganglions de Rotter [3].

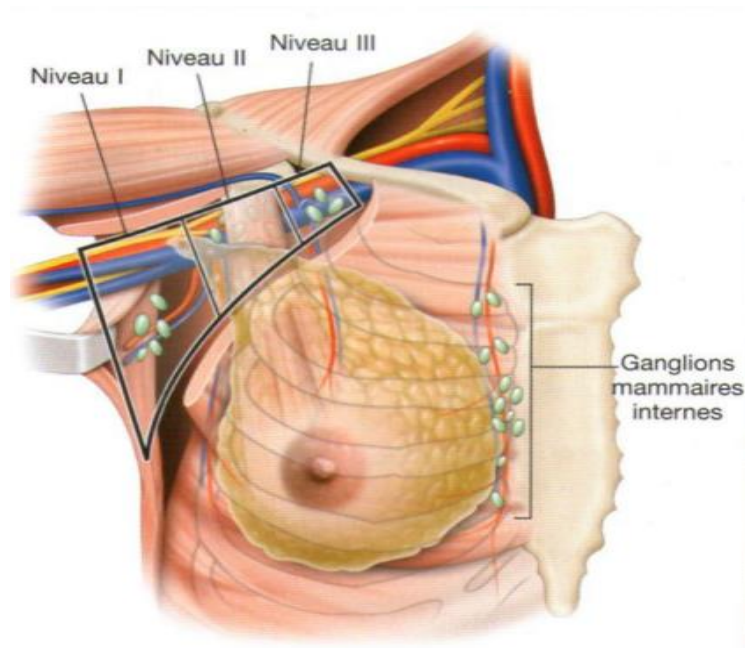


Figure 1.3 : Niveaux du curage mammaire externe (étages ganglionnaires de Berg) et les ganglions mammaires internes [3]

1.4 Cancer du sein

Le cancer du sein constitue une maladie très fréquente, caractérisée par la prolifération anormale et incontrôlée de cellules. Cette multiplication anarchique est à l'origine de la création de la tumeur. Ces cellules cancéreuses peuvent rester groupées les unes aux autres. Dans ce cas, la tumeur devient dangereuse du fait que ces cellules, provoquent la destruction de cellules normales et, mettent en danger le fonctionnement du sein.

En outre, la cellule cancéreuse [4] peut ne pas demeurer dans le sein et se détacher de la tumeur, pour migrer dans d'autres parties du corps par les vaisseaux sanguins et lymphatiques. Suite à cette propagation, des métastases, peuvent être créées dans un autre endroit, voire un autre organe. Ainsi, pour un cancer du sein métastatique dans le poumon, la tumeur secondaire est constituée de cellules de cancer du sein.

Des statistiques mondiales ont montré que les cancers du sein sont en forte hausse chez les femmes. En effet, 1,677 million nouveaux cas de cancer du sein ont été diagnostiqués dans le monde en 2012 ce qui fait plus que 25% du nombre total de nouveaux cas de tout type de cancer diagnostiqués. Ces statistiques ont montré aussi que ce type de cancer, représente également la cause de mortalité la plus fréquente chez les femmes (522000 décès) [5].

1.4.1 Description des stades évolutifs des cancers du sein

1.4.1.1 Cancers in situ

Au stade initial du cancer du sein, les cellules cancéreuses sont présentes uniquement dans le tissu d'origine (les canaux galactophores ou les lobules), sans que la tumeur infiltre les tissus voisins : on parle de carcinome canalaire in situ.

85-90 % des cancers in situ sont des carcinomes canaux et, 1 à 15 % sont des carcinomes lobulaires. Ceux-ci ne sont pas considérés comme un précurseur direct du cancer, mais comme un facteur de risque de développer un cancer du sein.

Les recherches les plus récentes tendent à définir le carcinome lobulaire in situ, comme un facteur de risque de cancer du sein et le carcinome canalaire in situ, comme un précurseur du carcinome mammaire infiltrant [6].

1.4.1.2 Maladie de Paget du mamelon

C'est un carcinome canalaire in situ, qui peut se propager au tissu du sein plus profond. Elle correspond cliniquement à un eczéma unilatéral du mamelon et, est caractérisée histologiquement par l'apparition de grandes cellules peu colorées, dans l'épiderme du mamelon [6].

1.4.1.3 Adénocarcinomes infiltrants

Lorsque les cellules cancéreuses ne restent pas confinées à leur lieu d'origine et envahissent les tissus avoisinants (les canaux galactophores et les lobules), on parle de cancer ou carcinome infiltrant [6].

1.4.1.4 Les adénocarcinomes métastasés

À un stade évolutif plus avancé, les cancers infiltrants se propagent vers les ganglions locaux (dits « ganglions axillaires » et situés sous l'aisselle) et, parfois, le reste du corps (l'os, les poumons, le foie ou, le cerveau). Elles s'accumulent alors dans les ganglions lymphatiques voisins : on dit qu'il existe un envahissement ganglionnaire. Les cellules cancéreuses ont tendance à migrer dans d'autres organes et, à y développer de nouvelles tumeurs qu'on appelle métastases. On dit dans ce cas que le cancer est métastatique [7].

Tous les cancers ne se comportent pas de la même façon. C'est la raison pour laquelle, il est nécessaire d'envisager un traitement adapté à chaque cancer. Tous les traitements, ont pour but de supprimer les cellules cancéreuses. En l'absence de traitement adapté, le cancer risque d'évoluer plus vite [8].

1.4.2 Symptomatologie

Durant les premiers stades de développement d'un cancer du sein, les symptômes sont frustrants. À un stade avancé, la symptomatologie peut être la suivante :

- une induration au niveau du sein ou de l'aisselle, sans ou avec douleur ;
- une inflammation, une ulcération, une déformation, ou une rétractation de la peau au niveau d'un des seins ;
- un écoulement au niveau du mamelon, un eczéma ou un érythème du mamelon ou de l'aréole.

Dans 10 % des cas, le cancer du sein est découvert par un examen clinique, faisant suite à des signes d'appels : maladie de Paget du mamelon, masse palpable, écoulement unipore séro-sanglant mamelonnaire [6, 9].

1.4.3 Facteurs de risque

Parmi les facteurs [9, 10] qui sont responsables du cancer du sein, nous citons :

- a) **Cancers du sein liés à l'âge** : les femmes de moins de 50 ans ; ont un risque relatif de cancer du sein inférieur à 4, par rapport aux femmes de plus de 50 ans.
- b) **Cancers du sein liés aux antécédents familiaux de cancer invasif** : avec le nombre d'antécédents familiaux de cancer du sein ; Le surrisque du cancer du sein augmente.
- c) **Première menstruation précoce (ménarche).**
- d) **Ménopause tardive.**
- e) **Obésité ou surpoids.**
- f) **Cancers liés à une mutation génétique** : ces cancers concernent 5 % à 10 % des cancers du sein. On estime qu'environ 2 femmes sur 1 000 sont porteuses d'une mutation du BRCA2 (gène localisé sur le chromosome 13) ou du BRCA1 (gène localisé sur le chromosome 17).
- g) **Radiothérapie du thorax.**
- h) **Consommation d'alcool.**

i) **Certains types de traitements contraceptifs hormonaux.**

1.5 Dépistage, Diagnostic et Traitement

L'incidence du cancer du sein reste croissante. Diverses études ont confirmé qu'une détection précoce du cancer, peut améliorer le pronostic vital. Dans tout changement de la taille ou de la forme du sein, des modifications de la peau et du mamelon, présence d'une masse dure et la présence des ganglions durs et parfois douloureux au niveau de l'aisselle ; un examen médical approfondi du sein est nécessaire [3].

Le dépistage précoce du cancer du sein est obligatoire tous les deux ans, pour les femmes de plus 40 ans. L'examen utilisé est la mammographie. Le médecin demandera d'autres examens (des incidences complémentaires, échographie, biopsie), pour confirmer le diagnostic du cancer.

Le diagnostic du cancer du sein est basé sur le triplet diagnostique clinique, la mammographie et l'anatomopathologique.

- L'examen clinique est un examen effectué avant et après la mammographie, afin de mettre en évidence d'éventuelles anomalies des certaines zones et, expliquer certains résultats. Une corrélation est ainsi, entre l'examen clinique et l'imagerie est réalisée.
- Mammographie est un examen indispensable pour l'exploration de la glande mammaire, qui peut être complété par une échographie mammaire.
- L'anatomopathologie fournit des informations précises sur le type de cancer du sein. Utile l'analyse au microscope des cellules et des tissus enlevés au niveau d'une anomalie du sein.

Le traitement : est local, systémique ou les deux rangés. La radiothérapie et la chirurgie agissent localement sur les cellules cancéreuses, situées dans le sein ou, dans les ganglions.

- **La chirurgie** consiste à enlever les éventuelles greffes cancéreuses et, la tumeur. Il existe plusieurs interventions possibles : la chirurgie conservatrice qui conserve le sein et la mastectomie totale.
- **La radiothérapie** a pour but d'éliminer toute trace du cancer, qui pourrait rester après une chirurgie à l'aide d'appareils, qui émettent des rayons.
- **La chimiothérapie** a pour but d'éliminer les cellules cancéreuses, ou de les éviter de s'accroître.

1.6 Imagerie du sein

L'imagerie médicale est un domaine de la médecine, qui regroupe un ensemble de techniques d'acquisition et de restitution d'images, de différents organes du corps humain. Ce domaine a connu d'importants progrès permettant non seulement un meilleur diagnostic, mais offrant aussi de nouvelles possibilités de traitement pour plusieurs maladies. La sénologie radiologique ou l'imagerie du sein, est primordiale pour le diagnostic et la caractérisation d'anomalies visualisées en examen de dépistage. La mammographie est l'examen paraclinique, généralement couplée à une échographie mammaire. Parfois, une étude complémentaire est nécessaire. Par conséquent, un bilan IRM (Imagerie par Résonance Magnétique) peut être demandé. Nous présentons dans ce qui suit, ces différents outils.

1.6.1 Imagerie par résonance magnétique (IRM)

L'imagerie par résonance magnétique (Figure 1.4) est un examen sans radiation, réalisé grâce à un appareil qui produit un champ magnétique très puissant, projeté sur le sein permettant de visualiser avec une très grande précision, la structure intérieure du sein. En effet, l'IRM permet une exploration approfondie des pathologies, dans tous les plans de l'espace. Cet examen est généralement effectué après une biopsie positive pour un cancer, où on a besoin de plus d'informations sur l'étendue de la maladie. Cependant, cette technique ne constitue pas un examen systématique du diagnostic du cancer du sein. Il s'agit plutôt d'un outil utilisé, pour compléter l'investigation de l'imagerie standard (mammographie ou échographie). Des recherches récentes, ont montré que l'IRM permet de localiser certaines petites lésions mammaires, parfois omises par la mammographie [11].

Elle peut également aider à détecter le cancer du sein chez les femmes, ayant des implants mammaires et chez les jeunes femmes qui, ont tendance à avoir des tissus mammaires denses. La figure 1.4 montre l'appareil utilisé pour l'IRM ainsi que son déroulement.



Figure 1.4 : Appareil utilisé pour l'IRM mammaire [11]

1.6.2 Mammographie

La mammographie est un examen radiologique mammaire basé sur les rayons x, ayant pour le but de visualiser l'intégralité du sein. C'est l'examen de référence pour le dépistage du cancer du sein et pour le diagnostic de certains symptômes qui touchent le sein comme les douleurs, la modification de la peau, l'inflammation et l'écoulement du mamelon.

Pour réaliser cet examen, on utilise un appareil qui se compose d'un générateur de rayons X (la dose de radiations étant généralement très faible), d'un système de compression et d'une cassette. L'examen consiste à positionner et, comprimer le sein entre deux plaques transparentes. Les plaques sont fixées à un appareil photo hautement spécialisé, qui prend deux images du sein dans deux directions. Ensuite, le technologue répète la technique sur le sein opposé. Pour certaines femmes, plus de deux images peuvent être nécessaires, pour inclure le plus de tissu possible. La radiographie est réalisée sur des films argentiques ou, sur des systèmes de radiologie digitale de haute qualité [12].

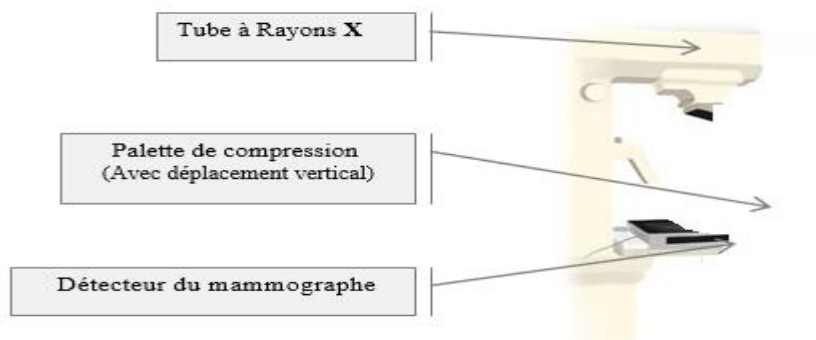


Figure 1.5 : Positionnement de la patiente pour la prise d'une mammographie en face [12]

La mammographie peut être douloureuse pour certaines femmes, mais pour la plupart, elle est légèrement inconfortable et la sensation, ne dure que quelques secondes. L'ensemble de la procédure dure environ 20 minutes.

1.6.2.1 Corrélation entre l'anatomie et les images de la mammographie

Il est d'abord nécessaire d'établir le lien, entre l'atténuation apportée aux rayons X et la structure des tissus mammaires. Par exemple la graisse est considérée comme une zone radio transparente, donc elle apparaît très sombre sur un cliché ; par contre, le tissu fibro-glandulaire qui, est un composant principal des lésions mammaires et le calcium, sont des éléments radio-opaques qui apparaissent clairs sur un cliché de mammographie. Pour ce qui est de l'eau, c'est une matière radio-opaque qui apparaît claire sur l'image [13]. Pour les substances principales dans le sein, nous obtenons le tableau de correspondance suivant :

composant	Atténuation radiologique	Aspect sur la mammographie
Graisse	Radio transparent	Très sombre
Eau	Légèrement radio	Sombre
Tissu conjonctif	opaque	Claire
Calcium	Radio opaque Très radio opaque	Très claire

Tableau 1.1 : Atténuation radiologique des composants mammaires [13]

1.6.2.2 Incidences en mammographie

En raison de la complexité de l'anatomie du sein, la mammographie est réalisée sous différents angles appelés incidences ; pour visualiser la majeure partie du tissu mammaire en l'étalant le plus possible sur la plaque radiographique. Il existe de nombreuses incidences [14] qui peuvent être divisées en deux groupes :

1.6.2.2.1 Incidences de base

Les incidences de base sont les incidences bilatérales cranio caudales (CC) et médio latérales obliques (MLO), généralement utilisées pour toutes les patientes du dépistage de routine. Chacune de ces incidences permet d'explorer une région bien précise : l'incidence CC

offre une bonne exploration des régions centrales et internes du sein, dont la réalisation implique que le tube radiogène et les détecteurs soient horizontaux. Pour ce qui est de la MLO, elle requiert une inclinaison de 40 à 60° du détecteur, selon la morphologie du patient. Ce qui permet d'explorer la totalité de la glande mammaire, sur une seule incidence.

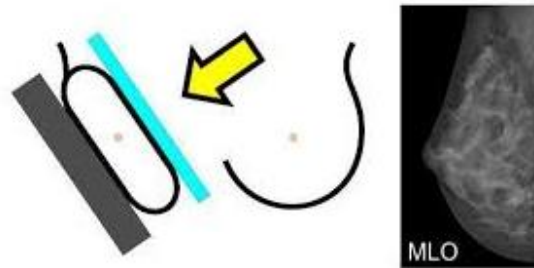


Figure 1.6 : Visualisation lever du bras pour oblique médio latérale (MLO) [14]

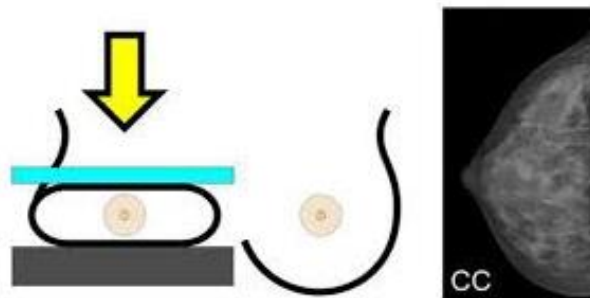


Figure 1.7 : Visualisation crano-caudale (CC)[14]

1.6.2.2 Incidences complémentaires

Les incidences complémentaires à effectuer, se font selon l'examen et la morphologie de la patiente (ex : CCX, CCXE, CCXI, DEC).

1.6.2.3 Limites de la mammographie

La mammographie présente certaines limites [14,15] que nous pouvons citer ci-dessous :

- ❖ **Les seins denses** : il faut souligner la difficulté voire l'impossibilité de la mammographie à déceler, un cancer infraclinique non calcifié dans des seins denses. Ceci-dit ; il y a le risque de passer à côté d'un cancer.
- ❖ **Les zones muettes** : il y'a des zones muettes sur la mammographie, où les incidences classiques n'arrivent pas à atteindre ce qui nécessite des incidences particulières, surtout sur les zones supéro--interne et retro aréolaire.

1.6.2.4 Nouveautés en mammographie

1.6.2.4.1 Tomosynthèse

Également appelée « mammographie 3D », la tomosynthèse est une nouvelle modalité d'imagerie qui réduit l'effet de superposition. Elle utilise plusieurs clichés pour générer une image en trois dimensions, avec une vision plus claire et plus fine des seins. Elle permet aux radiologues, une analyse plus précise de la taille, de la forme, de la localisation et du nombre des anomalies (une performance de diagnostic élevée) [16].

La différence entre la mammographie et l'appareil à tomosynthèse (figure 1.8), se résume dans le tube à rayons X qui reste statique dans la mammographie tandis que, dans la tomosynthèse, il effectue un arc de cercle.

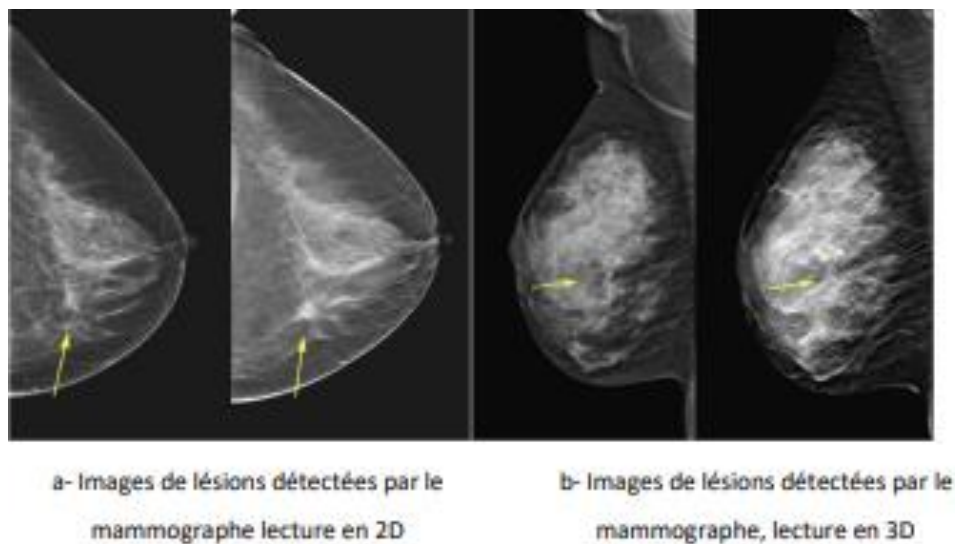


Figure 1.8 : Comparaison entre deux images de la même patiente en 2D et 3D [17]

1.6.2.4.2 Angio mammographie

L'angio mammographie, 'angio' pour vaisseaux sanguins [18], est un développement récent de la mammographie numérique. Le procédé consiste à injecter un produit de contraste iodé (par voie intraveineuse), puis chaque sein est examiné selon plusieurs incidences sous compression douce.

Cette technique permet de visualiser les vaisseaux nourriciers d'une tumeur, donc détecter dans des situations particulières des cancers, qui ne sont pas visibles en mammographie ou en échographie.

1.6.3 Echographie mammaire

L'échographie est une technique d'imagerie médicale qui utilise des ondes ultrasons, pour générer des images relativement précises [17]. Plus particulièrement, l'échographie mammaire fournit des images en temps réel, de l'intérieur des seins. Cet examen est considéré comme un examen complémentaire, à la mammographie. En effet, il permet au radiologue de mieux interpréter et visualiser les anomalies, mises en évidence par une mammographie. L'échographie mammaire ne présente aucun risque sur la patiente du fait qu'elle n'utilise pas les rayons X ; elle constitue l'examen de référence pour les femmes enceintes.

De plus, la sensibilité de cette technique n'est pas affectée, si la densité mammaire est élevée. En effet, un sein dense apparaît échogène en échographie, ce qui le rend plus facile à explorer en utilisant cette technique.

1.6.3.1 Principe de fonctionnement

Les techniques ultrasonographiques du diagnostic médical, se distinguent des autres techniques d'imagerie médicale, par leur caractère non vulnérant. Elles sont aussi remarquables par leur très haute résolution spatiale et, leur très haute résolution temporelle, avec des performances très supérieures à l'ensemble des autres techniques. En outre, l'ultrasonographie apporte, lors du même examen, des renseignements morphologiques (principalement par l'échographie) et fonctionnels (principalement en mode Doppler). Dans la très grande majorité des cas, un appareil d'ultrasonographie (**Figure 1.9**) associe le mode échographique et le mode Doppler [19].



Figure 1.9 : Appareil d'ultrasonographie [19]

Le fonctionnement de l'échographie est basé sur l'utilisation des ultrasons (avec une fréquence supérieure à 20 000 Hz). L'échographie a pour but d'envoyer des ultrasons dans le corps, puis de les réceptionner afin de permettre d'établir un diagnostic. Pour effectuer l'émission et la réception des ultrasons, le médecin doit se munir d'une sonde pourvue d'un émetteur et d'un récepteur, puis d'un système informatique afin de transformer les signaux reçus en images pour visualiser la zone étudiée.

Un gel est appliqué entre la peau et la sonde pour supprimer la présence d'air car celui-ci influe sur la vitesse des ultrasons, il les rend beaucoup plus lents. Ce gel nous permet donc de garder une image nette [20].

1.6.3.2 Signes sémiologiques échographiques des lésions malignes

Durant ces dernières années, le taux de détection de l'échographie des cancers infra cliniques, a augmenté avec l'arrivée des sondes linéaires de haute fréquence de 10 à 13 MHz. Ce qui rend l'échographie indispensable pour la découverte des cancers infiltrants, non détectables par la clinique ou la mammographie, souvent à un stade infraclinique [15, 21].

Pour la détection de ces cancers, l'échographie se base sur des signes sémiologiques qui sont prédictifs de malignité.

- **L'hypoéchogénéité** : l'échostructure nettement hypoéchogène, a une grande valeur prédictive de malignité. C'est un signe apprécié par rapport à l'échostructure de la graisse et, non celle du tissu fibreux plus échogène.
- **Les contours irréguliers** : les contours d'une masse sont une caractéristique essentielle, pour différencier entre une lésion maligne et bénigne, allant d'une marge bien définie (une caractéristique bénigne) aux marges microlobulées ou angulaires (caractéristiques malignes) [21].
- **L'atténuation du faisceau postérieur** : l'atténuation du faisceau postérieur, varie selon le stroma réactionnel du cancer. Dans l'étude de Stravros et al [22], ce caractère était présent dans 76% des cas, essentiellement dans les cancers tubuleux. Par contre les cancers médullaires, les formes hémorragiques ou nécrotiques ont plus souvent un faisceau ultrasonore normal ou rehaussé [15].
- **La direction de l'axe principal de la tumeur** : ce signe est primordial pour déterminer la forme et, juger de la nature de la lésion. Les lésions bénignes se développent dans

un axe antéropostérieur ; cependant elles sont plus larges que hautes. Par contre les lésions malignes ont un développement transversal, elles sont plus hautes que larges. Ce signe est suspect lorsqu'il suggère un développement perpendiculaire aux plans des tissus, à l'inverse des lésions bénignes qui ont un développement parallèle [15].

- **Les microcalcifications isolées** : quelques études ont analysé la performance de l'échographie, dans la détection des microcalcifications non associées à des masses ou des distorsions architecturales, à la mammographie. Lorsque les microcalcifications isolées ou en foyer sont visualisées à l'échographie, il s'agit le plus souvent de lésions malignes [23].

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

1.6.3.3 Limites de l'échographie

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

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

1.7 Anomalies mammaires

1.7.1 Masses

En imagerie du sein, on distingue deux grands types de lésions ; on a d'un côté les microcalcifications et de l'autre les masses [13,20]. Par définition, une masse est une opacité importante, occupant un espace et vue sur deux incidences différentes.

Dans le cas où, la masse serait observée uniquement sur une seule incidence, elle est appelée asymétrie jusqu'à ce que sa nature tridimensionnelle soit confirmée [46]. Si elle est visible sur une seule incidence, il s'agit d'une asymétrie. On la décrit selon :

- **la forme** : ronde, ovale, lobulée ou irrégulière (figure 1.10).



Figure 1.10 : Différentes formes de masses [13]

- **Le contour**

En plus de leur forme, les opacités varient aussi au niveau de leur contour. La classification BIRADS dénombre cinq classes de contour que sont les contours (figure 1.11) ; circonscrits, microlobulés, masqués, mal définis et spiculés [20].

- Contours circonscrits** : ce sont des contours bien définis où la frontière entre la masse et le fond de l'image est claire et correspond généralement à une lésion bénigne.
- Contours masqués** : les contours sont dissimulés par la superposition de tissus normaux, faisant penser à une masse circonscrite dont une partie du contour est cachée.
- Contours micro-lobulés** : ce sont des contours qui contiennent de petites ondulations.
- Contours mal définis** : ce sont des contours qui permet d'envisager l'existence d'une infiltration.
- Contours spiculés** : ce sont des contours comportant des structures filiformes, qui rayonnent vers l'extérieur depuis le centre de masse et qui, suggèrent fortement une malignité.

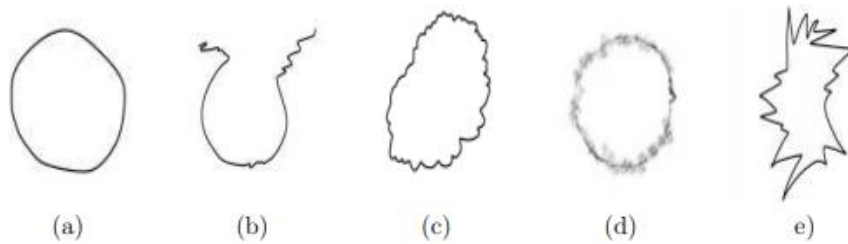


Figure 1.11 : Différents types de contour des opacités selon le BI-RADS
 (a) contour circonscrit, (b) contour masqué, (c) contour micro-lobulé, (d) contour mal défini
 (e) contour spiculé [13]

➤ **La densité par rapport au tissu fibro-glandulaire normal**

Face aux différents modèles existants et dans le but de standardiser les comptes rendus des mammographies, l'ACR [24] a mis au point une classification BI-RADS définissant quatre classes relatives à la composition mammaire (figure 1.12).

- a) **B1** : le sein est radio transparent, entièrement graisseux et homogène et, facile à interpréter pour les radiologues (la densité occupe moins de 25% de la glande mammaire).
- b) **B2** : le sein est graisseux et hétérogène. On remarque la présence d'opacités fibro-glandulaires dispersées dans le parenchyme (la densité occupe entre 25 et 50% du sein).
- c) **B3** : le sein est dense et hétérogène (la densité occupe entre 50 et 75% de la glande mammaire).
- d) **B4** : le sein est hyperdense et homogène ; par conséquent, l'interprétation est très difficile car la densité peut masquer la masse (la densité occupe plus de 75% du sein).

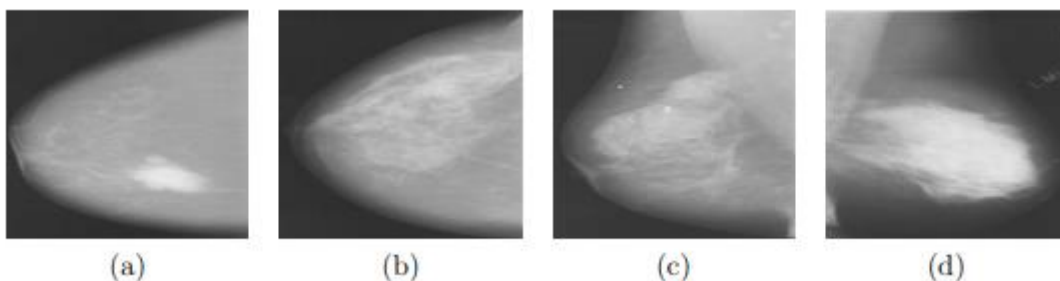


Figure 1.12 : Types de densité des masses selon le BI-RADS : (a) B1 (densité < 25%), (b) B2 (densité comprise entre 25 et 50%), (c) B3 (densité comprise entre 50 et 75%) et (d) B4 (densité > 75%) [24]

1.7.2 Calcifications mammaires

Les calcifications mammaires sont des dépôts de calcium, dans le tissu mammaire. Bien que les calcifications mammaires soient généralement non cancéreuses (bénignes), certains types de calcifications ; comme des amas serrés de forme irrégulière et d'apparence fine, peuvent indiquer un cancer du sein, ou des modifications précancéreuses du tissu mammaire. Sur une mammographie, les calcifications mammaires peuvent apparaître sous forme de macrocalcifications ou de microcalcifications [13, 20].

1.7.2.1 Macrocalcifications

Les macrocalcifications se présentent sous la forme de gros points blancs ou de tirets. Elles sont presque toujours non cancéreuses et ne nécessitent aucun autre test ou suivi.

1.7.2.2 Microcalcifications

Les microcalcifications (Mcs, figure 1.13) sont des dépôts de sels de calcium de taille inférieure, à 1 mm qui se forment dans le tissu du sein. Elles apparaissent sur une mammographie comme des « petits points blancs », situés dans la glande mammaire. Elles peuvent constituer un point d'appel radiologique d'un cancer du sein, d'où toute l'importance de leur description et/ou de leur analyse.

La séparation des microcalcifications bénignes et malignes se base principalement sur plusieurs critères :

- **taille des microcalcifications** : en général, les grosses microcalcifications d'environ un millimètre sont plus souvent bénignes. Il existe cependant des exceptions, car les calcifications hétérogènes ou dystrophiques de plus d'un millimètre, peuvent s'associer à des lésions malignes.
- **Nombre de microcalcifications** : le nombre élevé de microcalcifications dans un groupe, est un signe de nécrose.
- **Siège des microcalcifications** : elles peuvent être externes, internes ou, rétro-aréolaires.
- **Forme** : Les calcifications arrondies ou ovales, sont probablement bénignes. À l'inverse, plus elles sont irrégulières, plus elles ont de risque d'être maligne.

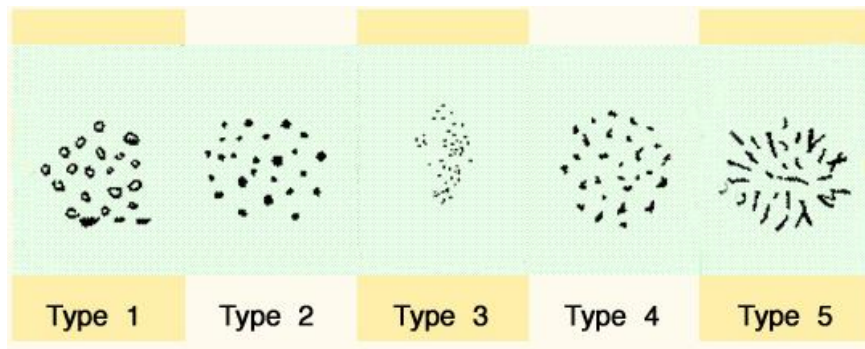


Figure 1.13 : Classification des microcalcifications [20]

1.8 Classification des pathologies mammaires

Plusieurs lexiques standards et des classifications ont été proposés, pour permettre aux radiologues d’avoir un canevas, leur permettant d’interpréter les mammographies. Dans ce cadre, le radiologue peut étudier convenablement les lésions mammaires et, faciliter une seconde lecture. Jusqu’à présent, les classifications les plus répandues et les plus couramment utilisées par les radiologues, sont celles de Le Gal et de BI-RADS.

1.8.1 Classification de Le Gal

En 1976, le Gal du service de radiodiagnostic de l’institut Curie à Paris [13], a proposé la classification dite de Le Gal. Cette classification décrit cinq types morphologiques, qui ont une valeur prédictive de malignité croissante (tableau 1.4). Elle se base uniquement sur les microcalcifications et ne prend pas en compte la morphologie des masses et des distorsions architecturales.

Type 1	Mcs annulaires, arciformes ou polyédriques. risque de cancer du sein quasi nul.
Type 2	Mcs rondes et de tailles variables. Risque de carcinome : 22%.
Type 3	Mcs poussiéreuses, pulvérulentes. Risque de cancer : 36%.
Type 4	Mcs irrégulières associées `a un risque de cancer : 56%
Type 5	Mcs vermiculaires ou branchées. Risque de carcinome : 90%.

Tableau 1.2 : Classification de Le Gal [13]

1.8.2 Classification BI-RADS de l'ACR

La classification BI-RADS (Breast Imaging Reporting And Data System) est une classification internationale établie par l'ACR (American college of Radiology) [24, 25]. Elle permet de classer les images mammographiques, en six catégories (tableau 1.3) en fonction du degré de suspicion de leur caractère pathologique et d'identifier précisément, les masses malignes et bénignes.

CATEGORIE	DESCRIPTION	Risque de cancer	CONDUITE A TENIR
ACR 0	Investigation incomplète		Des investigations complémentaires sont nécessaires
ACR 1	Normale	0%	Retour au dépistage
ACR 2	Anomalie bénigne	0%	Retour au dépistage
ACR 3	Anomalie probablement bénigne	> 0 % mais ≤ 2%	Suivi 6 mois
ACR 4	Anomalie demandant une biopsie Faiblement suspecte Modérément suspecte Très suspecte	> 2 % mais < 95 % > 2 % à ≤ 10 % > 10 % à ≤ 50 % > 50 % à < 95 %	Biopsie
ACR 5	Anomalie fortement suspecte d'un cancer	≥ 95 %	Biopsie

Tableau 1.3 : Classification BIRADS De L'ACR [25]

1.9 Conclusion

Dans ce chapitre, nous avons présenté la notion du cancer du sein ainsi que les outils de l'imagerie médicale utilisés, pour le diagnostic de ce type de lésion. Nous avons aussi, consacré une grande attention à l'étude des pathologies mammaires (les masses et les microcalcifications). Cette étude est associée à la présentation des critères adoptés par les radiologues, pour classifier les lésions mammaires en bénignes ou malignes.

Beaucoup de lésions cancéreuses ont un faible contraste et souvent, elles peuvent être confondues avec les tissus de la glande mammaire. Par conséquent l'interprétation, devient une tâche très difficile.

Dans ce cadre, notre projet se rapporte, à la classification des images échographiques mammaires en utilisant l'intelligence artificielle ; afin d'apporter une aide automatique aux médecins radiologues dans leur diagnostic. Le second chapitre, est consacré à l'apprentissage profond en imagerie médicale.

Chapitre 2

Apprentissage profond en imagerie médicale

Chapitre II Apprentissage profond en imagerie médicale

2.1 Introduction

Le but de la recherche en intelligence artificielle (IA) est de doter un système informatique, de capacités de réflexion similaires à celles des humains. Il y a donc un défi dans la compréhension du raisonnement humain, mais surtout dans la modélisation et dans la reproduction de celui-ci [26].

L'intelligence artificielle (figure 2.1) est devenue un sujet en vogue dans les médias et les magazines scientifiques, en raison des progrès accomplis dans le domaine d'apprentissage automatique. Les grandes entreprises comme IBM, Google, Facebook, Microsoft ainsi que les constructeurs automobiles tels que Toyota, Volvo et Renault, sont très actifs dans la recherche en IA et, prévoient d'y investir davantage encore à l'avenir. Plusieurs scientifiques spécialisés dans l'IA, dirigent désormais les laboratoires de recherche de ces grandes entreprises. La recherche en IA a permis de réaliser d'importants progrès, dans la dernière décennie et ce dans différents secteurs. Les avancées les plus connues sont celles réalisées dans l'apprentissage automatique, grâce notamment au développement d'architectures d'apprentissage profond [27].

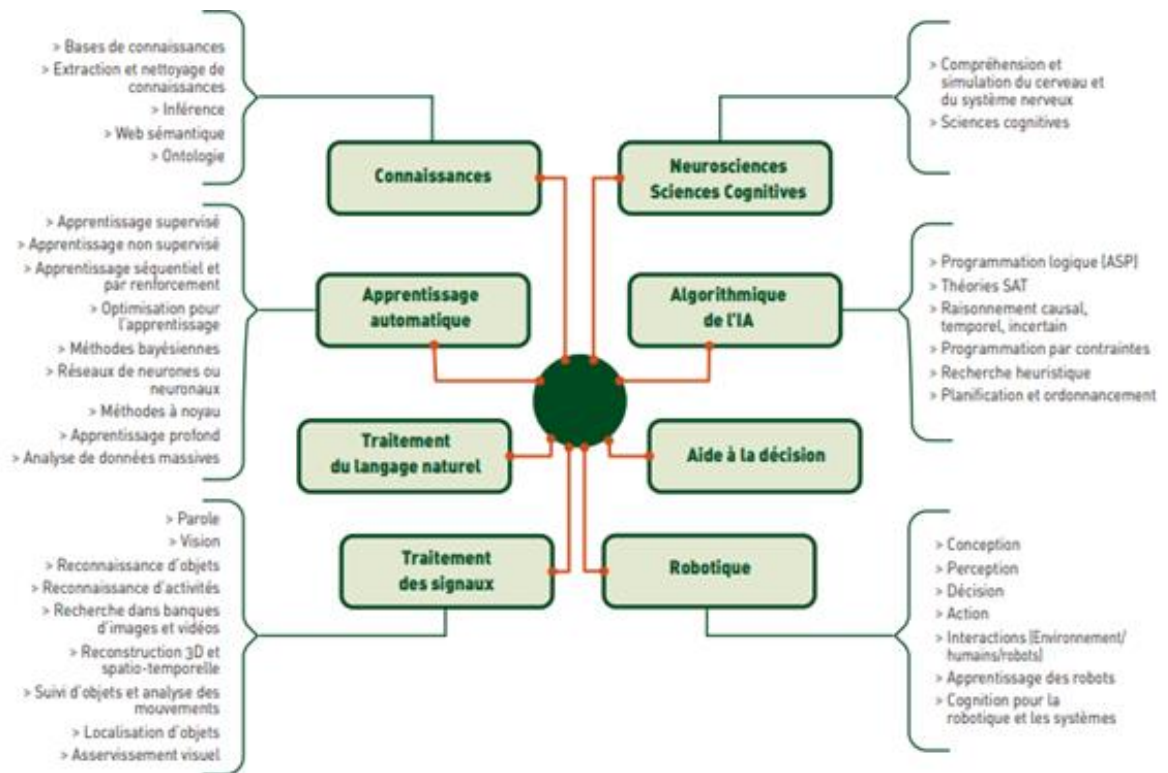


Figure 2.1 : Différents domaines de l'intelligence artificielle [27]

Dans ce chapitre, nous présentons brièvement les différents types d'apprentissage automatique, pour décrire plus en détail l'apprentissage profond.

2.2 Apprentissage Automatique

L'apprentissage automatique (en anglais machine learning), est un sous-ensemble d'intelligence artificielle. Il s'agit d'une discipline qui consiste à appliquer des algorithmes à des jeux de données, afin d'en extraire des modèles. Ceux-ci peuvent à leur tour, être appliqués sur des données similaires à des fins de prédiction. Avec suffisamment de données, il est possible de formuler une approximation de la relation entre toutes les variables d'entrée et les sorties. On peut alors appliquer cette formule sur de nouvelles données d'entrée, pour prévoir la classe associée [26, 27, 28].

2.2.1 Différents types d'apprentissage automatique

Trois grandes approches relèvent de l'apprentissage automatique : l'apprentissage supervisé, l'apprentissage non-supervisé et l'apprentissage par renforcement.

2.2.1.1 Apprentissage supervisé

L'apprentissage supervisé est l'un des paradigmes du machine learning. Son nom indique la présence d'un superviseur du processus d'apprentissage. Les données d'entrée de ce processus, sont déjà catégorisées (étiquetées). Par conséquent, les algorithmes doivent s'en servir pour prédire un résultat.

La robustesse de l'algorithme dépendra de la précision de son entraînement et l'apprentissage s'arrête lorsque l'algorithme, atteint un niveau de performance acceptable.

La plupart des problèmes d'apprentissage automatique et d'apprentissage profond, utilisent l'apprentissage supervisé. Par conséquent, il est nécessaire de comprendre comment ce mécanisme fonctionne. Parmi les applications : la reconnaissance vocale, la vision par ordinateur et la classification.

L'apprentissage supervisé est généralement applicable à deux types de problèmes. Ces tâches sont essentiellement différentes par la nature des paires (entrée, sortie) qui y sont associées.

a) Classification

On parle d'un problème de classification si la variable à prédire, prend une valeur discrète (une catégorie). La classification supervisée vise principalement à définir des règles, permettant de classer des objets dans des classes, à partir de variables caractérisant ces objets. Par exemple : la classification du spam, la classification des animaux comme : chat, chien, vache.

b) Régression

Les problèmes de régression correspondent aux situations dans lesquelles, la variable de sortie est une variable quantitative qui prend des valeurs dans un sous-domaine de l'ensemble \mathbb{R} des réels (des valeurs continues). Par exemple : le prix d'un appartement, l'évolution du climat, durée de vie d'un patient.

2.2.1.2 Apprentissage non supervisé

Contrairement à l'apprentissage supervisé, cet apprentissage se fait de façon totalement autonome.

Les données d'entrées ne sont pas étiquetées, par conséquent l'algorithme d'entraînement s'applique à trouver seul les similarités et, les différences au sein de ces données, pour regrouper celles qui partagent des caractéristiques communes. En clair, L'apprentissage non supervisé (très souvent synonyme de Clustering) cherche d'une part à former des groupes (clusters) aussi distincts que possible et, à maximiser l'homogénéité des données dans ses derniers.

L'apprentissage non supervisé est généralement utilisé pour la détection d'anomalies, y compris pour les achats frauduleux de cartes de crédit, les systèmes de recommandation qui conseillent sur les produits à acheter ou bien qui recommandent un livre ou un film à un utilisateur en fonction des goûts d'utilisateurs partageant des caractéristiques communes, les cyberattaques.

2.2.1.3 Apprentissage par renforcement

L'apprentissage par renforcement (figure 2.2) est né de la rencontre entre la psychologie expérimentale et, les neurosciences computationnelles. Cette technique populaire se base sur un cycle d'expérience/récompense, qui consiste à laisser la machine apprendre à faire une tâche de manière autonome. Elle commencera par prendre des décisions aléatoires. Si la machine se trompe, elle est pénalisée. Si elle prend la bonne décision, elle est récompensée et de récompense en récompense, elle développera ses propres méthodes pour accomplir la tâche qui lui est confiée.

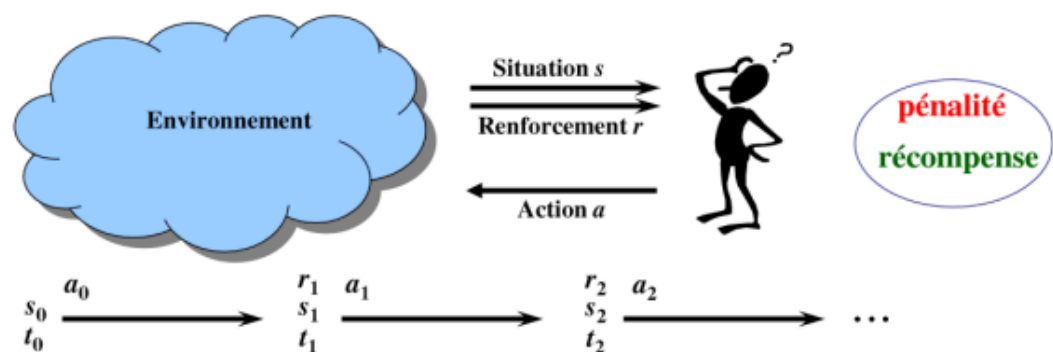


Figure 2.2 : Schéma de principe de l'apprentissage par renforcement [28]

2.2.2 Métriques d'évaluation

Les métriques d'évaluation [29,30] permettent d'estimer la qualité d'un algorithme de classification.

2.2.2.1 Matrice de confusion

C'est un tableau de contingence (figure 2.3) confrontant les classes désirées (lignes) et les classes obtenues (colonnes), pour l'échantillon. Sur la diagonale principale, se trouve donc les valeurs bien classées, hors diagonale les éléments mal classés.

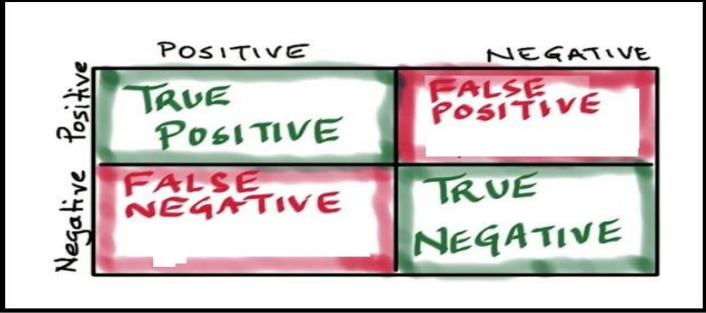


Figure 2.3 : Matrice de confusion dans le cas binaire [29]. Le tableau est divisé en quatre quadrants par une diagonale principale. Les colonnes sont étiquetées 'POSITIVE' et 'NEGATIVE' en haut. Les lignes sont étiquetées 'Positive' et 'Negative' à gauche. Les quadrants contiennent : Top-Left (vert) 'TRUE POSITIVE', Top-Right (rouge) 'FALSE POSITIVE', Bottom-Left (rouge) 'FALSE NEGATIVE', Bottom-Right (vert) 'TRUE NEGATIVE'.

	POSITIVE	NEGATIVE
Positive	TRUE POSITIVE	FALSE POSITIVE
Negative	FALSE NEGATIVE	TRUE NEGATIVE

Figure 2.3 : Matrice de confusion dans le cas binaire [29]

- VP : les vrais positifs, c'est-à-dire les individus qui ont été classés positifs, qui sont en réalité des positifs.
- FP : les faux positifs, c'est-à-dire les observations classées positives et qui sont en réalité des négatives.
- De la même manière, VN sont les vrais négatifs et les FN sont les faux négatifs.

La matrice de confusion [30] permet de déduire les paramètres suivants :

- 1) **La précision** : la précision correspond à la fréquence à laquelle le modèle prédit correctement la classe positive.

$$\text{Précision} = \frac{VP}{VP+FP} \quad (2.1)$$

- 2) **Rappel** : en d'autre terme **la sensibilité**, elle permet de répondre à la question suivante : parmi toutes les étiquettes positives possibles, combien d'entre elles le modèle a-t-il correctement identifiées ? En d'autres termes :

$$\text{Rappel} = \frac{VP}{VP+FN} \quad (2.2)$$

- 3) **La spécificité** : C'est l'équivalente du **taux de vrai négatif**. Sa formule est la suivante :

$$SP = \frac{VN}{VN+FP} \quad (2.3)$$

4) **F-mesure** : c'est la moyenne harmonique entre le rappel et la précision.

$$F - \text{mesure} = \frac{2*(\text{précision}*\text{rappel})}{\text{précision}+\text{rappel}} \quad (2.4)$$

2.2.2.2 Courbe ROC

Une courbe ROC (receiver operating characteristic) est un graphique [31] représentant les performances d'un modèle de classification, pour tous les seuils de classification. Cette courbe trace le taux de vrais positifs en fonction du taux de faux positifs (figure 2.4).

➤ **Le taux de vrais positifs (TVP)** est l'équivalent du **rappel**. Il est donc défini comme suit :

$$\text{TVP} = \frac{\text{VP}}{\text{VP}+\text{FN}} \quad (2.5)$$

➤ **Le taux de faux positifs (TFP)** est défini comme suit :

$$\text{TFP} = \frac{\text{FP}}{\text{VP}+\text{FP}} \quad (2.6)$$

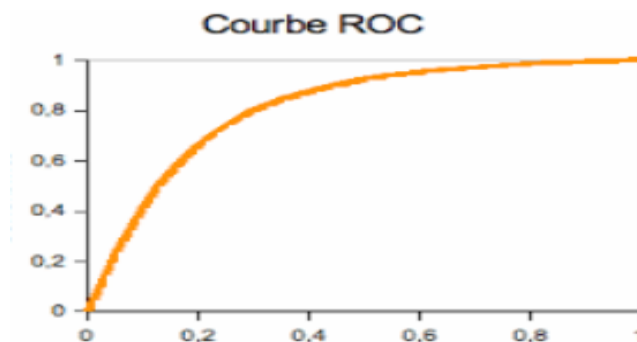


Figure 2. 4 : Courbe Roc [31]

2.2.3 Fonction d'erreur

La fonction d'erreur est une fonction mathématique qui serve à pénaliser le classifieur, en cas de mauvaise prédiction. Pour avoir un bon classifieur, il faut que les pénalités reçues par le modèle, soient minimums.

Soit L la fonction d'erreur :

$$\hat{y} = \text{model}(X, W) \quad (2.7)$$

$$\min_W L(y, \hat{y}) \quad (2.8)$$

Où \hat{y} est la prédiction du modèle,

W : hyperparamètre qui détermine la prédiction du modèle. C'est toujours la valeur de cet hyperparamètre que l'on cherche à trouver, dans les algorithmes d'apprentissage. En apprentissage automatique, le choix de la fonction d'erreur est primordial.

2.2.4 Descente stochastique de gradient

La descente stochastique de gradient (SGD) est une technique d'optimisation itérative, la plus utilisée dans le domaine du deep learning. Nous l'utilisons pour minimiser la fonction d'erreur (équation 2.7) afin de chercher les hyperparamètres [32]. Son algorithme est le suivant :

Algorithme 1 Descente stochastique de gradient

1. Initialiser l'hyperparamètre W
2. Choisir le pas d'apprentissage η

Répéter

3. Choisir aléatoirement un sous ensemble x
4. Prédire \hat{y} , $\hat{y} = model(X, W)$
5. Calculer l'erreur $L(y, \hat{y})$
6. Mettre à jour de l'hyperparamètre. $W := W + \eta \Delta W$

Jusqu'à ce que le minimum approximatif soit obtenu

Où :

ΔW : la correction de l'hyperparamètre W

η : le pas d'apprentissage.

Le "momentum" est une variante de la descente stochastique de gradient pour augmenter la vitesse de l'optimisation [56]. Il change suivant la mise à jour du gradient.

Ainsi, au lieu de faire directement $W := W + \eta \Delta W$, la mise à jour se fait en deux étapes suivant l'algorithme ci-dessous :

Algorithme 2 mise à jour des hyperparamètres

1. $v = \mu * v - \eta \Delta W$
 2. $W := W + v$
-

Où :

v : vitesse

μ : momentum est une constante.

2.2.5 Quelques algorithmes d'apprentissage automatique

2.2.5.1 Régression linéaire

C'est l'algorithme le plus basique [33] qui cherche à établir une relation linéaire, entre la variable à prédire et la variable prédictive (figure 2.5). Ceci revient à déterminer les paramètres du modèle (a, b) qui sont les coefficients de l'équation suivante :

$$Y = ax + b + \varepsilon(x) \quad (2.9)$$

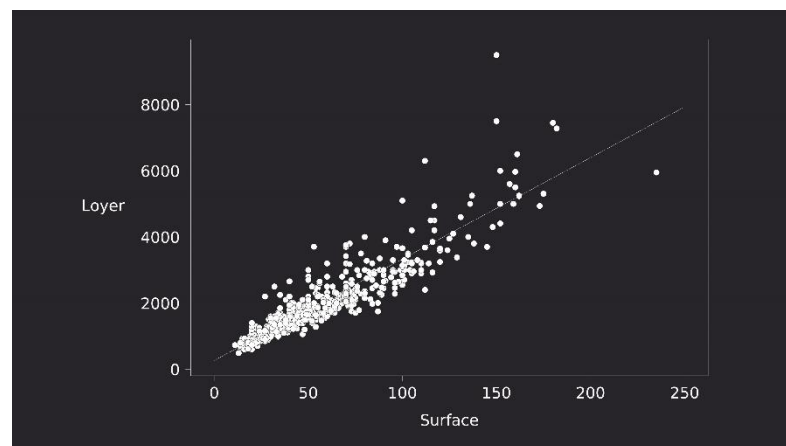


Figure 2.5 : Représentation des données d'entrée par rapport à un classificateur linéaire [34]

Où « a » est la pente, « b » est la constante et « ε » est un bruit qui permet la bonne identification de la relation.

Après avoir déterminé tous ces paramètres, nous utilisons cette équation pour prédire des nouvelles données. Le principal inconvénient de cet algorithme est qu'il n'est performant, que pour les relations simples entre les données.

2.2.5.2 Arbre de décision

L'arbre de décision est un algorithme simple d'aide à la décision, utilisée en machine learning.

Un arbre de décision est un outil d'aide à la décision représentant un ensemble de choix, sous la forme graphique d'un arbre. Les différentes décisions possibles sont situées aux extrémités des branches (les « feuilles » de l'arbre) et, sont atteintes en fonction de décisions prises à chaque étape. L'arbre de décision est un outil utilisé dans des domaines variés, tels que la fouille de données, l'imagerie médicale, etc...

Un arbre de décision, est représenté par des branches et des feuilles (figure 2.6). Chacune des branches, est affectée à un attribut. Les feuilles désignent la classe de l'objet à classer. L'objectif est de produire des groupes d'individus les plus homogènes possibles, du point de vue de la variable à prédire. Il est d'usage de représenter la distribution empirique de l'attribut à prédire sur chaque sommet (nœud) de l'arbre [35].

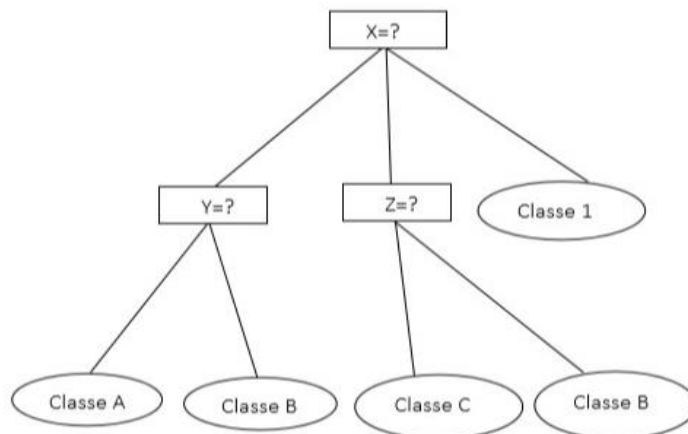


Figure 2.6 : Exemple d'un arbre de décision avec 4 classes [35]

2.2.5.3 KNN « Plus proches voisins »

La méthode des K plus proches voisins « k-NN k-nearest neighbor » est une approche simple. Elle estime la sortie associée à une nouvelle entrée, en tenant compte des K voisins les plus proches, des données connues [36].

Dans l'exemple de la figure 2.7, pour $k = 3$ la classe majoritaire du point central est la classe B, mais si on change la valeur du voisinage $k = 6$, la classe majoritaire devient la classe A.

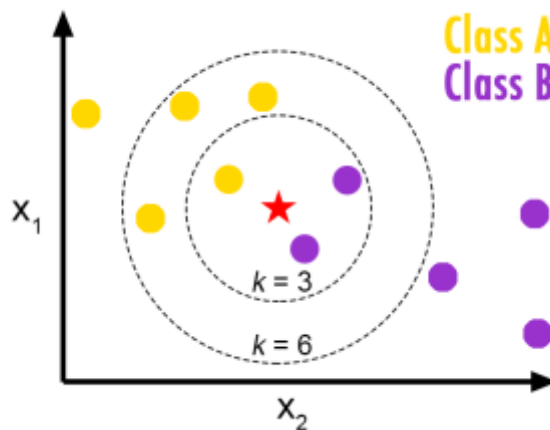


Figure 2.7 : Exemple des KNN [37]

2.3 Réseaux de neurones artificiels

2.3.1 Principe

Aujourd'hui de nombreux termes, sont utilisés pour désigner le domaine des réseaux de neurones artificiels, comme connexionnisme ou neuromimétisme. Ces deux derniers termes sont des domaines de la recherche à part entière, qui manipulent chacun des modèles des réseaux de neurones artificiels, avec des objectifs différents. L'objectif des chercheurs connexionnistes est d'améliorer la puissance de calcul, en utilisant des modèles aux composants fortement connectés. Pour leur part, les neuromiméticiens manipulent des modèles de réseaux de neurones artificiels uniquement, pour tester leurs théories biologiques du fonctionnement du système nerveux central [38,39].

❖ Définition

Les réseaux de neurones artificiels (ou Neural Network en anglais), sont des systèmes informatiques matériels et / ou logiciels. Ils sont des imitations simples des fonctions des neurones du cerveau humain, pour résoudre des problématiques d'apprentissage de la machine (Machine Learning).

❖ Pourquoi recourir à des réseaux de neurones ?

La réponse est plutôt simple dans le sens où les réseaux de neurones, s'avèrent plus performants que les techniques de régressions pour des tâches de Machine Learning.

2.3.2 Historique

De façon générale, on situe le début des réseaux de neurones artificiels en 1890, avec les travaux de W. James [38], un célèbre psychologue américain. Il introduit le concept de mémoire associative et, propose ce qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones, connue plus tard sous le nom de loi de Hebb.

En 1943, « McCulloch et Pitts » [40] ont proposé les premières notions de neurone formel. Ils montrent qu'un réseau de neurones discret, sans contrainte de topologie, peut représenter n'importe quelle fonction booléenne et donc émuler un ordinateur. Après 15 ans, Rosenblatt [38] propose le premier algorithme d'apprentissage, qui permet d'ajuster les paramètres d'un neurone.

Entre 1969-1982, les recherches se poursuivent, mais déguisées, sous le couvert de divers domaines comme, par exemple : le traitement adaptatif du signal, la reconnaissance de formes, la modélisation en neurobiologie, etc. De grands noms travaillèrent durant cette période tels que : S. Grossberg et T. Kohonen [38].

En 1985, la rétropropagation de gradient apparaît. C'est un algorithme d'apprentissage qui convient aux réseaux de neurones multicouches (MLP). Sa découverte réalisée par trois groupes de chercheurs indépendants, indique que "la chose était dans l'air". Dès cette découverte, nous avons la possibilité de réaliser une fonction non linéaire d'entrée /sortie non linéaire sur un réseau, en décomposant cette fonction en une série d'étapes linéairement séparables. Aujourd'hui, les réseaux multicouches et la rétropropagation de gradient, restent les modèles les plus étudiés et les plus productifs, au niveau des applications. Au cours des années 90, la recherche sur les réseaux de neurones connaît un essor fulgurant et, les applications commerciales de ce succès académique suivent [40].

2.3.3 Neurone biologique

Un neurone biologique [40] reçoit plusieurs informations (qui sont des neurotransmetteurs), au niveau de ses dendrites. Ces neurotransmetteurs sont libérés par les synapses.

Lorsque la quantité d'information dépasse un certain seuil, le neurone « actif » envoie un courant électrique, dans son axone ce qui, lui permet d'émettre à son tour des neurotransmetteurs via ses synapses (voir figure 2.8).

Pour résumer les différentes propriétés des réseaux de neurones :

- ✓ Les entrées d'un neurone sont les sorties d'un autre ;
- ✓ Lorsqu'un neurone reçoit une quantité d'information dépassant un seuil, il émet une erreur ;
- ✓ Les synapses gèrent la quantité d'information émise au neurone suivante.

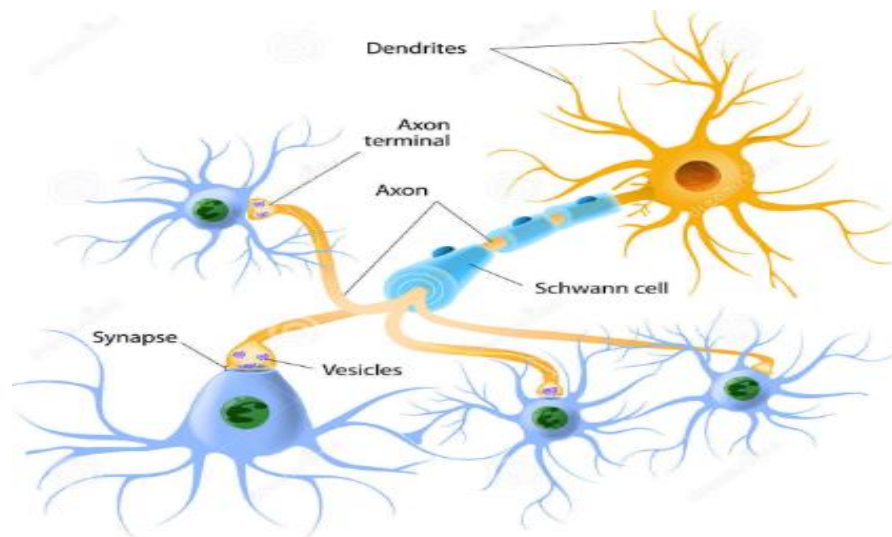


Figure 2.8 : Neurone biologique [40]

2.3.4 Neurone formel

Le modèle du neurone informatique, ou neurone formel, s'inspire d'un fonctionnement simplifié d'un neurone biologique. La modélisation simplifiée de McCulloch et Pitts datant de 1943 le conçoit comme recevant, d'autres neurones, des impulsions électriques de poids variables par ses dendrites.

Au niveau du neurone, si la somme des impulsions dépasse un certain seuil, l'influx électrique reçu est transformé, dans le corps du neurone, en une nouvelle impulsion électrique, que l'on peut modéliser comme un 1 ou un 0 (activation ou non activation). Cette activation est ensuite transmise par l'axone aux neurones connectés à celui-ci [41].

On construit, en calquant sur cette modélisation, un neurone formel où :

- les **entrées** x du neurone, se présentent comme un vecteur de dimension N ;
- un ensemble de poids W , modélisent les poids variables des différentes entrées. W est un vecteur de dimension N également. À chaque entrée, correspond un poids.
- Un **biais** b représentant le seuil d'activation du neurone.
- La somme à des N "impulsions électrique" reçues est donc obtenue par :

$$a = \sum_{i=0}^N W_i x_i + b \quad (2.10)$$

- **Une fonction d'activation** h , rendant compte de l'influx électrique de sortie, est appliquée à la somme des entrées pondérées par W et du biais b . On utilise en général la fonction sigmoïde :

$$\text{sigmoïde}(a) = \frac{1}{1+e^{-a}} \quad (2.11)$$

Muni de ces trois composants, on peut alors modéliser un neurone comme exposé dans la figure 2.9.

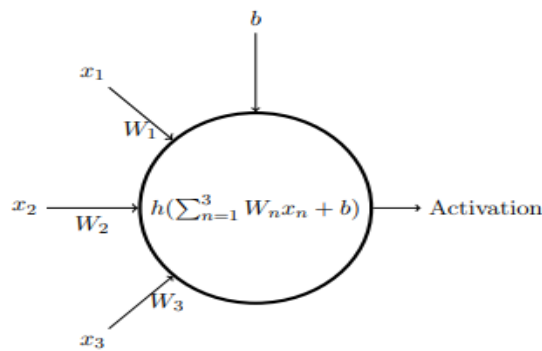


Figure 2.9 : Schéma d'un neurone à 3 entrées [41]

2.3.5 Structure d'interconnexion

Les connexions entre les neurones qui composent un réseau, décrivent la topologie du modèle. Elle peut être quelconque, mais le plus souvent, il est possible de distinguer une certaine régularité [38]. On peut généralement retenir 3 principaux RNA (réseaux de neurones artificiels).

1. Réseau de neurones monocouche
2. Réseau à connexions locales
3. Réseau multicouches

2.3.5.1 Réseau de neurones monocouche

C'est une structure de réseau [42] constituée d'un nombre de nœuds interconnectés, par des liaisons directionnelles. Le réseau le plus simple est le perceptron monocouche (figure 2.10), caractérisé de la manière suivante :

- il possède N informations en entrée ;
- il est composé de p neurones, que l'on représente généralement alignés verticalement. Chacun peut en théorie avoir une fonction d'activation différente. En pratique, ce n'est généralement pas le cas ;
- chacun des p neurones est connecté aux N informations d'entrée.

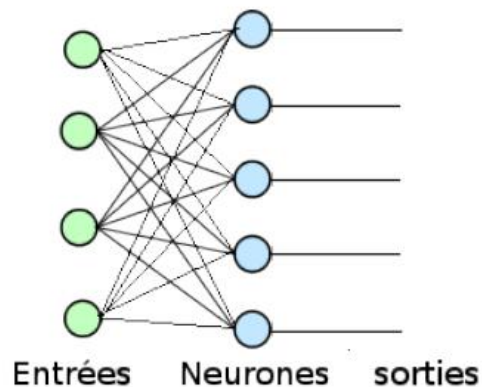


Figure 2.10 : Exemple d'un réseau de neurones monocouche [42]

2.3.5.2 Réseau à connexions locales

Il s'agit d'une structure multicouche (figure 2.11), mais qui à l'image de la rétine, conserve une certaine topologie [34].

Chaque neurone est en contact avec un nombre réduit et localisé, de neurones de la couche avale. Les connexions sont donc moins nombreuses, que dans le cas d'un réseau multicouche classique.

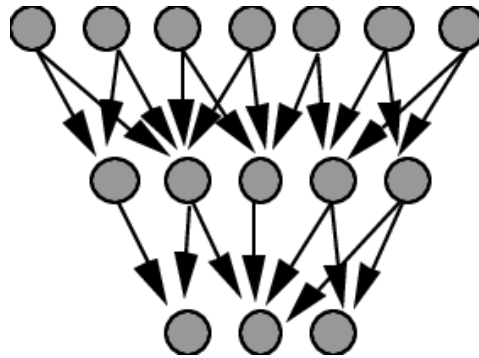


Figure 2.11 : Réseau à connexions locales [42]

2.3.5.3 Réseau multicouche

Les neurones sont arrangés par couche. Il n'y a pas de connexion entre les neurones d'une même couche et les connexions ne se font qu'avec, les neurones des couches avales. Habituellement, chaque neurone d'une couche est connecté à tous les neurones de la couche suivante. Ceci nous permet d'introduire la notion du sens de parcours de l'information (de l'activation), au sein d'un réseau et donc de définir les concepts des neurones d'entrée, et des neurones de sortie.

On appelle la couche d'entrée l'ensemble des neurones d'entrée et, la couche de sortie l'ensemble des neurones de sortie. Les couches intermédiaires n'ayant aucun contact avec l'extérieur sont appelés couches cachées (figure 2.12) ou couches fortement connectées (fully connected) [43].

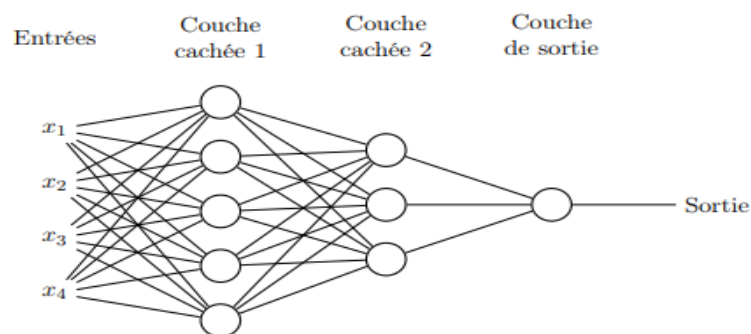


Figure 2.12 : Exemple d'un réseau neurone à deux couches cachées [43]

2.4 Apprentissage profond

L'apprentissage profond (le deep learning), est un nouveau domaine, de recherche du machine learning. Son but principal est de modéliser des relations complexes, entre les données en s'inspirant par la structure et, le fonctionnement du cerveau.

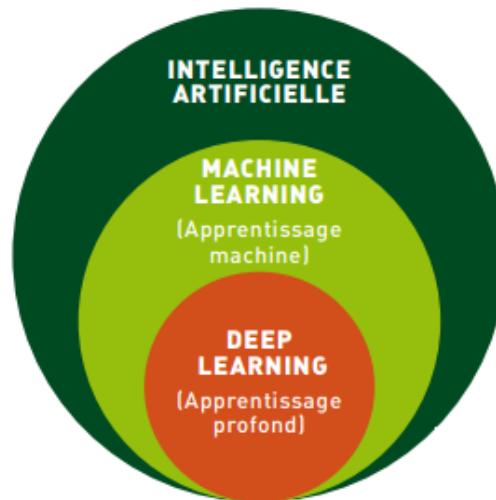


Figure 2.13 : Relation entre l'intelligence artificielle, le ML et le deep learning [44]

Le Deep Learning (DL) est basé sur l'idée des réseaux de neurones artificielles, conçu pour traiter de larges quantités de données, en ajoutant des couches au réseau. Un modèle du DL a la capacité d'extraire des caractéristiques, à partir des données brutes grâce aux multiples couches de traitement. Il permet d'apprendre sur ces caractéristiques, à travers chaque couche suivant une intervention humaine, minimale. Ces dernières années, le deep learning est devenu le domaine le plus prisé par les chercheurs [44].

2.4.1 Histoire de l'apprentissage profond

Nous résumons l'historique de l'apprentissage profond [45], dans le tableau ci-dessous :

Année	Contributeur	Contribution
300 AC	Aristote	Introduction de l'associationnisme, début de l'histoire des humains qui essaient de comprendre le cerveau.
1873	Alexander Bain	Introduction du Neural Grouping comme les premiers modèles de réseaux de neurones.

1943	McCulloch and Pitts	Introduction du McCulloch–Pitts (MCP) modèle considéré comme L’ancêtre des réseaux de neurones artificielles.
1949	Donald Hebb	Considéré comme le père des réseaux de neurones, il introduit la règle d’apprentissage de Hebb qui servira de fondation pour les réseaux de neurones modernes.
1958	Frank Rosenblatt	Introduction du premier perceptron.
1974	Paul Werbos	Introduction de la rétro propagation.
1980	Teuvo Kohonen	Introduction des cartes auto organisatrices.
1982	Kunihiko Fukushima	Introduction du Neocognitron, qui a inspiré les réseaux de neurones convolutifs.
1982	John Hopfield	Introduction des réseaux de Hopfield.
1985	Hilton and Sejnowski	Introduction des machines de Boltzmann
1986	Paul Smolensky	Introduction de ‘Harmonium’, qui sera connu plus tard, comme les machines de Boltzmann restreintes.
1986	Michael I. Jordan	Définition et introduction des réseaux de neurones récurrents.
1990	Yann LeCun	Introduction de LeNet qui montra la capacité des réseaux de neurones profonds.
1997	Schuster and Paliwal	Introduction des réseaux de neurones récurrent bidirectionnelles
1997	Hochreiter and Schmidhuber	Introduction de LSTM, qui ont résolu le problème du vanishing gradient dans les réseaux de neurones récurrents.
2006	Geoffrey Hinton	Introduction des ‘Deep belief Network’

2009	Salakhutdinov and Hinton	Introduction des Deep Boltzmann Machines.
2012	Alex Krizhevsky	Introduction de AlexNet qui remporta le challenge fac à ImageNet.

Tableau 2.1 : Evolution du Deep Learning [45]

2.4.2 Pourquoi l'apprentissage profond ?

Les approches du machine learning décrits dans la première partie, fonctionnent bien pour une grande variété de problèmes. Cependant ils n'ont pas réussi à résoudre certains problèmes majeurs d'IA, telle que la reconnaissance d'objets et la reconnaissance vocale. Vu l'échec des algorithmes traditionnels dans de telles tâches de l'IA, le deep learning a connu un développement immense. Après que les machines de calcul soient devenues plus puissantes et grâce notamment au Big Data ; on a pu comprendre le potentiel réel du Deep Learning [45].

Plus le volume de données fournie est grand, plus la performance d'un algorithme d'apprentissage profond est meilleure. Plusieurs algorithmes classiques du ML, possèdent une limite au volume de données qu'ils peuvent recevoir. Par contre les modèles du Deep Learning, peuvent même dépasser la performance humaine, dans certains domaines comme le traitement de l'image (figure 2.14).

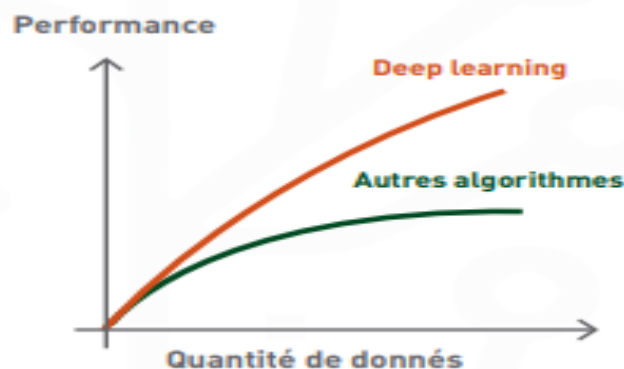


Figure 2.14 : Différence de performances entre le DL et le ML en fonction de la quantité de données [45]

L'autre différence, est l'étape de l'extraction de caractéristiques. Dans le deep learning cette étape est exécutée automatiquement par l'algorithme, alors qu'en machine learning, l'extraction de caractéristiques est faite manuellement (figure 2.15). En effet, c'est une étape coûteuse en temps, complexe, qui requiert un spécialiste en la matière.

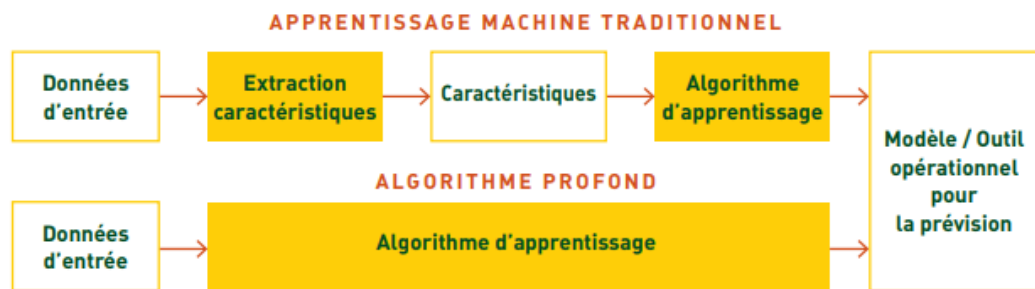


Figure 2.15 : Procédé de l'apprentissage classique comparé à celui du deep learning [45]

2.5 Architectures des réseaux de neurones en apprentissage profond

L'apprentissage profond est un domaine en évolution rapide. Toutes les semaines, de nouvelles architectures apparaissent. Comme elles ne sont pas toutes évaluées sur les mêmes ensembles de données, il n'est pas toujours possible de comparer leurs performances.

Nous décrivons dans ce qui suit, les principaux modèles des réseaux connexionnistes dans l'apprentissage profond.

2.5.1 Réseaux de neurones récurrents

L'humain ne commence pas sa réflexion à partir de zéro chaque seconde. En lisant ce paragraphe, nous comprenons chaque mot, en fonction de notre compréhension des mots précédents. Nos pensées ont de la persévérance.

Les réseaux de neurones traditionnels ne peuvent pas le faire, ce qui semble être une lacune majeure. Par exemple, supposons que nous souhaitions classer, le type d'événement qui se produit à chaque étape d'un film. On ne sait pas comment un réseau de neurones traditionnel, pourrait utiliser son raisonnement sur des événements précédents, dans le film pour les informer plus tard. Les réseaux de neurones récurrents résolvent ce problème [46, 47].

Ce sont des réseaux avec des boucles, qui permettent à l'information, de se maintenir. Dans la figure 2.16, un segment du réseau neuronal ; "A" avec une entrée x_t , fournit une valeur h_t . Cette boucle permet de transmettre des informations, d'une étape du réseau à l'autre.

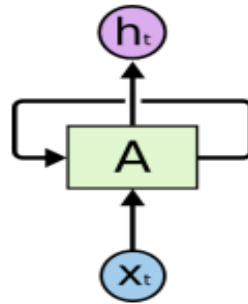


Figure 2.16 : Réseaux neuronaux récurrents ont des boucles [46]

Un réseau de neurones récurrent est considéré comme des copies multiples, du même réseau ; chacune transmettant un message à son successeur, comme le montre la figure 2.17.

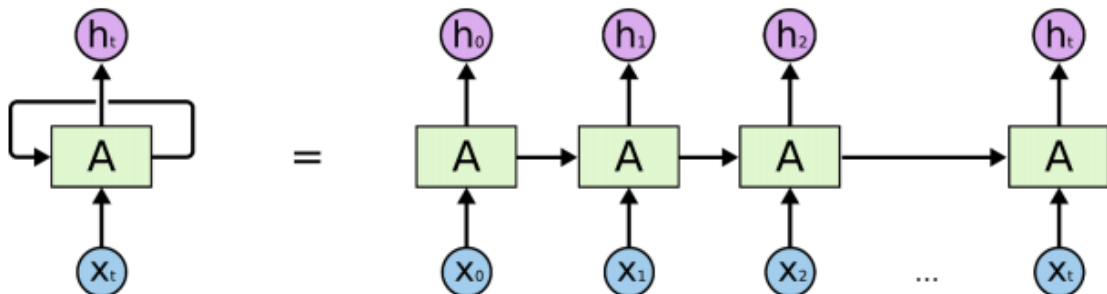


Figure 2.17 : Réseau neuronal récurrent [46]

Ces dernières années, les RNN ont connu un grand succès, appliqués, à une variété de problèmes comme la reconnaissance de la parole, la modélisation du langage, la traduction et le sous-titrage des images [47].

2.5.2 Réseaux de neurones convolutifs

Désignés par l'acronyme CNN (Convolutional Neural Network), les réseaux convolutifs sont des réseaux de neurones artificiels, utilisés pour la reconnaissance et le traitement d'images. Ce type de réseaux est une puissante application de l'intelligence artificielle (IA) au traitement des images, qui s'appuie sur l'apprentissage profond pour effectuer des tâches descriptives et génératives [48].

Dans les CNN, les « neurones » sont disposés comme ceux du lobe frontal, où les humains et les animaux traitent les stimulus visuels. Le système utilisé par eux, est comparable aux perceptrons multicouches. Cependant, son objectif de conception est de réduire la quantité de calcul. On distingue deux parties ; la première est la partie convolutive du modèle. Elle fonctionne comme un extracteur de caractéristiques des images et, la seconde est la classification du modèle constituée de couches entièrement connectées (perceptron multicouches (Multi Layers Perceptron MLP).

Les réseaux convolutifs exploitent généralement la vision artificielle, notamment la reconnaissance de la vidéo et des images, ainsi que le traitement automatique du langage naturel (TALN).

2.5.2.1 Architecture de réseaux de neurones convolutifs

La structure des CNN [49] se compose d'une série de couches :

- couche de convolution ;
- couche de regroupement ;
- couche de correction ReLU ;
- couche entièrement connectée ;
- couche de perte.

a. Couche de convolution (CONV)

La couche convolutive est un élément clé, des réseaux de neurones convolutifs. Elle constitue toujours au moins, leur première couche. Son but est de repérer l'existence d'un ensemble de caractéristiques, dans l'image reçue en entrée.

Pour cela, on effectue un filtrage par convolution (figure 2.18). On glisse le masque sur l'image et, on calcule le produit de convolution entre le filtre et chaque portion de l'image balayée [50].

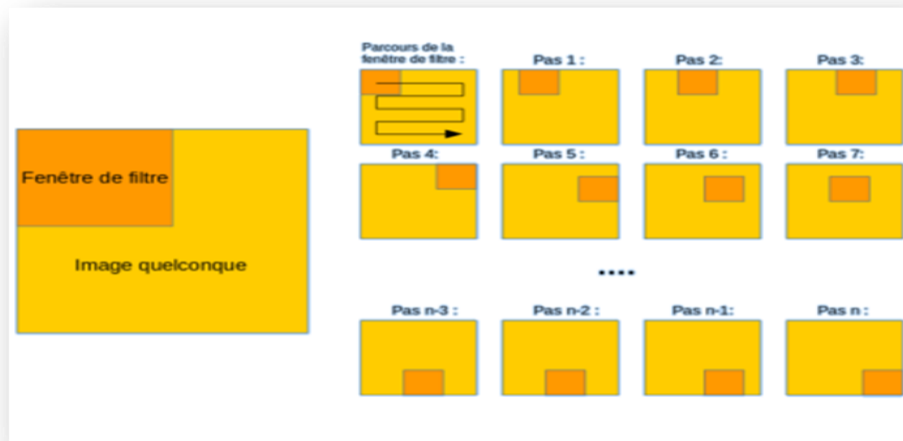


Figure 2.18 : Schéma du parcours de la fenêtre du filtre de convolution sur l'image [51]

L'image passe ainsi, à travers une série de filtres (noyaux de convolution). On obtient une carte d'activation ou 'feature map' pour chaque paire (image, filtre), qui nous indique la position des caractéristiques dans l'image. Dans ce cas, plus la valeur est élevée, plus l'endroit correspondant dans l'image, est affecté aux attributs pertinents.

Trois hyperparamètres déterminent le volume de la couche de convolution :

- ❖ La profondeur de la couche « K (nombre de noyaux de convolution) » ;
- ❖ Le pas « S » contrôle le chevauchement des champs récepteurs ;
- ❖ Le zero padding « P » : permet de contrôler la dimension spatiale du volume de sortie.

Le calcul de la convolution (figure 2.19) est relativement simple : pour une convolution de taille 3×3, on va sélectionner dans l'image les 3×3 premiers pixels pour en créer un nouveau, dans l'image de sortie. La valeur de ce nouveau pixel, est égale à « pixel1 de la sélection dans l'image » * « pixel1 du filtre » + « pixel2 de la sélection dans l'image » * « pixel2 du filtre » + etc... jusqu'au neuvième pixel.

Pour calculer la deuxième valeur de l'image de sortie, on utilise le paramètre « stride » de la convolution. Il représente le nombre de pixels à décaler, dans l'image d'entrée, pour réappliquer la convolution [51].

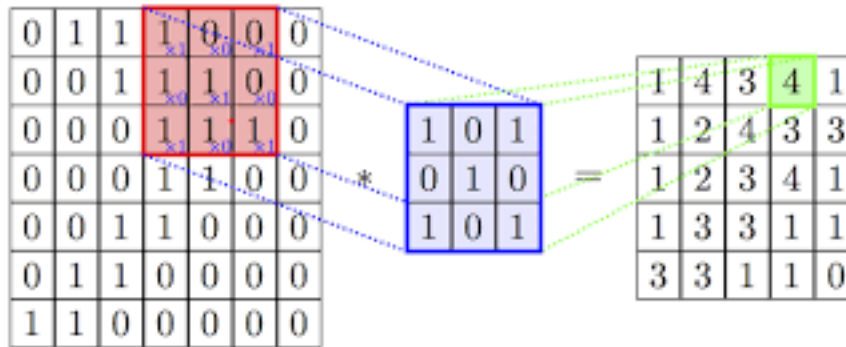


Figure 2.19 : Exemple d'une convolution

Cependant il peut arriver que la convolution dépasse l'image, si on choisit mal le kernel et le stride. Par exemple, pour une image 4x4, un kernel de 2x2 et un stride de 3, on aura un problème après la première convolution. De plus, on remarquera que le résultat de la convolution, est une image plus petite que l'image d'entrée. A ces deux problèmes, le paramètre « padding » utilisé dans les convolutions, ajoute simplement des 0, autour de l'image d'entrée, pour ajuster la taille de l'image de sortie à celle de l'entrée. En fait, le filtre du réseau neuronal convolutionnel présente un 'kernel', 'stride' et un 'padding' prédéfinis, mais leurs valeurs sont générées de manière aléatoire lors de l'initialisation. Dans ce cas, lorsque le réseau réalise l'apprentissage, les valeurs des filtres sont mises à jour, afin d'améliorer les résultats du CNN. Les valeurs des filtres correspondent aux variables (poids, biais...) que le réseau change, pendant l'apprentissage [52].

b. Couche de pooling (POOL)

Le Pooling est un autre outil intéressant, utilisé par les CNNs. Une forme de sous-échantillonnage qui permet de prendre une grande image et, de réduire sa taille tout en conservant les informations les plus importantes qu'elle contient ; réduisant ainsi la quantité de paramètres et de calcul dans le réseau. Par conséquent, une couche de mise en commun, est souvent insérée entre deux couches convolutionnelles consécutives de l'architecture CNN, pour contrôler le sur-apprentissage (l'overfitting). L'opération de pooling crée également une forme d'invariance par translation. Ainsi, la couche de pooling rend le réseau moins sensible, à l'emplacement des caractéristiques [52].

Le fait que les caractéristiques soient plus hautes ou plus basses, ou même qu'elles aient une orientation légèrement différente, ne devrait pas entraîner de changement radical, dans la classification des images.

En pratique, on utilise généralement, des fenêtres de petite taille, pour ne pas perdre trop d'informations. Le choix les plus courants, sont des cellules adjacentes de taille 2×2 pixels qui ne se chevauchent pas. Comme la valeur de sortie, est une valeur maximale de l'entrée, nous parlons dans ce cas, de "Max-Pool 2×2 " (figure 2.20).

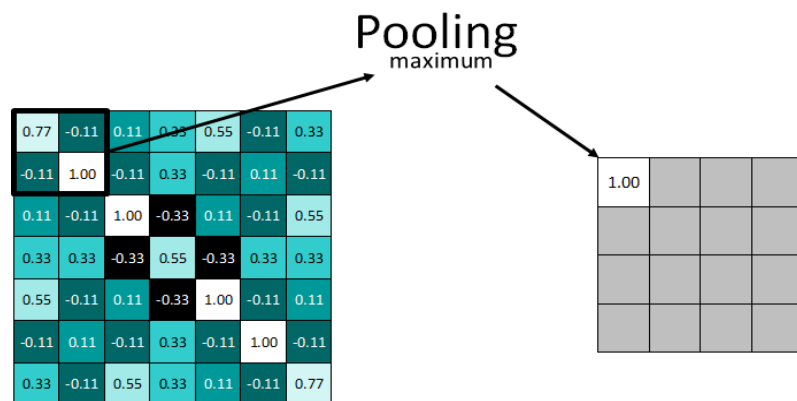


Figure 2.20 : Exemple d'un max pooling

Il est possible d'utiliser d'autres fonctions de pooling que le maximum comme :

- ❖ Le « mean pooling » (ou average pooling), soit la moyenne des pixels de la sélection. On calcule la somme de toutes les valeurs et on les divise par le nombre de valeurs de la fenêtre. On obtient ainsi, une valeur intermédiaire pour représenter ce lot de pixels.
- ❖ Le « sum pooling », c'est la moyenne sans avoir à la diviser par le nombre de valeurs (on ne calcule que leur somme).

c. Couche de correction (ReLU)

La fonction ReLU (non-linear activation function) est une fonction d'activation très couramment utilisée. Elle est formulée par la formule : $f(x) = \max(0, x)$. Elle remplace tout simplement les résultats négatifs, par zéro. Cette fonction (figure 2.21) oblige les neurones à retourner, des valeurs positives [52].

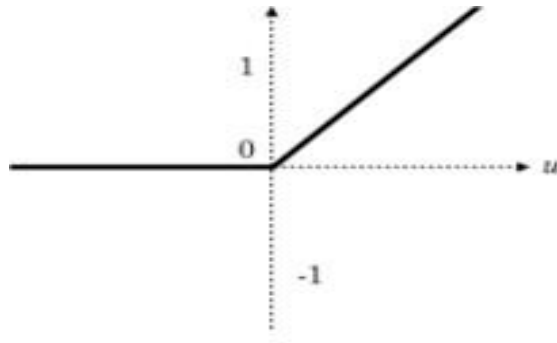


Figure 2.21 : Fonction ReLU

Souvent, la correction Relu est préférable, mais il existe d'autre formes :

- ❖ La correction par tangente hyperbolique $f(x) = \tanh(x)$,
- ❖ La correction par la tangente hyperbolique saturante : $f(x) = |\tanh(x)|$,
- ❖ La correction par sigmoïde : $f(x) = \frac{1}{1+e^{-a}}$

d. Couche entièrement connectée (FC)

Le raisonnement de haut niveau dans le réseau neuronal, se fait toujours à travers des couches entièrement connectées (Fully-connected Layers). Les neurones dans cette couche, sont connectés avec chacun des neurones, de la couche précédente. Cette couche engendre un vecteur de M dimensions, où M est le nombre de classes à prédire. Ce vecteur contient les probabilités de classement, pour chaque catégorie d'une image [52].

e. Couche de perte (LOSS)

Cette couche suit la couche entièrement connectée et gère les ajustements de poids, sur le réseau. Avant le début de l'apprentissage du réseau, les poids de la couche convolutionnelle et de la couche entièrement connectée, sont spécifiés sous forme de valeurs aléatoires. Ensuite, pendant le processus d'entraînement ; la couche de perte vérifie en permanence les suppositions de la couche entièrement connectée, par rapport aux valeurs réelles pour minimiser autant que possible, la différence entre l'estimation et la valeur réelle, en ajustant les poids, à la fois dans la couche de convolution et, dans les couches entièrement connectées. Diverses fonctions de perte adaptées à différentes tâches, peuvent y être utilisées. La fonction « Softmax » permet de calculer la distribution des probabilités sur les classes de sortie [52].

2.6 Apprentissage profond et l'imagerie médicale

Diverses techniques d'acquisition permettent d'obtenir des images internes du corps humain, comme l'échographie, l'imagerie par résonance magnétique (IRM), la mammographie et le scanner. Ces représentations numériques doivent être interprétées, afin de faire des diagnostics qui prennent beaucoup de temps aux médecins.

Tout au début, l'analyse des images médicales était faite, avec des traitements d'images basiques, comme la détection de contours, le filtrage, la binarisation, l'érosion ou la dilatation morphologiques. Il était difficile de travailler avec ces images à cause de leur résolution et de l'importance du bruit interférentiel, dû aux matériels d'acquisition utilisés.

Après l'apparition de l'apprentissage automatique, il a été possible de réaliser une classification et une segmentation, plus ou moins efficaces. Mais la procédure est encore difficile, car il faut faire manuellement beaucoup d'extraction de variables clefs, avant de les intégrer dans un classifieur (SVM, régression logistique, forêts aléatoires) pour améliorer la performance de ces algorithmes [13].

Actuellement, le deep learning est le choix le plus tendance, dans l'analyse des images médicales. Grâce à ce type d'apprentissage, la phase d'extraction des caractéristiques clefs, peut-être automatisée. Dans les sections suivantes, nous présentons quelques utilisations, de l'apprentissage profond.

2.6.1 Classification

La classification est l'application la plus directe de l'apprentissage profond, non seulement dans l'imagerie médicale mais aussi, dans d'autres applications. On utilise généralement du transfert d'apprentissage, car il n'y a pas de données suffisantes pour entraîner un modèle dans la plupart des cas. Hwang et al [53], ont obtenu une efficacité de 70.5% dans la classification d'images cytopathologiques de la tuberculose, avec un transfert d'apprentissage. D'autres travaux, ont implémentés leurs propres architectures des CNN au lieu de faire un transfert d'apprentissage. Les résultats ne sont satisfaisants, que si l'on dispose de suffisamment de données [54, 55].

2.6.2 Détection et localisation

Il s'agit de localiser un élément (organe, lésion) dans l'espace de l'image. D'habitude on ajoute un contour, comme un cadre ou un cercle. A la différence de la classification, la détection est plus difficile, car on veut des informations beaucoup plus précises. Le plus grand challenge de la détection dans l'imagerie médicale, c'est la vitesse de la détection des détails et, la pertinence de la localisation des régions d'intérêt [13, 20].

2.6.3 Segmentation

La segmentation a beaucoup d'importance en analyse d'image médicales, surtout dans des domaines spécifiques comme la cardiologie, la neurologie ou la cancérologie.

U-Net est un modèle de CNN spécialement créé, pour faire de la segmentation des images biomédicales. Son architecture est une série de couches de convolution avec différentes tailles et filtres, pour un rétrécissement puis une autre série de couches de déconvolution, pour un agrandissement et la reconstitution de l'image, à la taille initiale. Si l'on fait passer juste une image dans U-Net, on a directement l'image segmentée à la sortie [56].

2.7 Quelques travaux réalisés en imagerie échographique du sein

La technique d'imagerie par ultrasons est l'un des outils radiologiques, utilisé pour le diagnostic du cancer du sein. Face à l'augmentation du nombre d'échographie du sein pendant ces dernières années, il n'y a pas beaucoup de travaux de recherches pour la réalisation de systèmes de diagnostic assisté par ordinateur. La majorité de travaux effectués, se basent sur des images mammographiques. Nous décrivons dans cette section, quelques travaux explorés en imagerie ultrasonore du sein.

Nayeem et al [57] soulèvent le problème de la sélection de caractéristiques pour la détection du cancer du sein à partir des images échographiques. En effet, au fil des ans, plusieurs caractéristiques de haute précision, ont été proposées par les chercheurs pour distinguer les lésions malignes des lésions bénignes. Ces travaux emploient plus de cinquante de ces éléments, qui peuvent servir de référence aux chercheurs. Nayeem et al cherchent à sélectionner un sous-ensemble optimisé de cet ensemble de caractéristiques, en utilisant trois méthodes différentes.

Dans ce travail, ils avaient utilisé avec succès la méthode (MCFS) (Multi-Cluster Feature Selection), pour décrire au mieux le cancer du sein. L'outil de diagnostic assisté par ordinateur développé dans ce cadre, prend en compte, une combinaison optimale de 25 caractéristiques différentes pour différencier les tumeurs malignes des tumeurs bénignes. Le SRC (sparse representation classifier) est déployé pour classer les tumeurs. La technique proposée a été examinée sur une base de données de 504 tumeurs du sein diagnostiquées pathologiquement, dont 454 sont bénignes et 50 sont malignes. La surface sous la courbe AUC a été évaluée à 93,31 %.

Yap et al [58] proposent l'utilisation d'approches d'apprentissage profond pour la détection des lésions par échographie mammaire. L'étude se fait suivant trois méthodes différentes : un LeNet basé sur un patch, un U-Net et une approche d'apprentissage par transfert avec un FCN-AlexNet pré-entraîné. En outre, Ils comparent et mettent en contraste deux ensembles de données d'images échographiques classiques acquis, à partir de deux systèmes à ultrasons différents. L'ensemble de données A comprend 306 images (60 malignes et 246 bénignes) et, l'ensemble de données B comprend 163 (53 images malignes et 110 bénignes). Les résultats démontrent une amélioration globale des approches adoptées, lorsqu'ils sont évalués sur les deux ensembles de données, en termes de vrais positifs, de faux positifs et de F-mesure.

Zhantao et al [59] traitent la problématique de la détection et la classification des lésions mammaires échographiques, suivant des architectures d'apprentissage profond. Dans cette étude, ils analysent, explorent et évaluent différentes méthodes de détection et de classification d'objets (Fast Region-based convolutional neural networks (R-CNN), Faster R-CNN, You Only Look Once (YOLO), YOLO version 3 (YOLOv3), et Single Shot MultiBox Detector (SSD)) basées sur des architectures de type CNN (AlexNet, ZFNet, VGG, ResNet, GoogLeNet et DenseNet). Les résultats de ce travail, montrent que pour la tâche de détection des lésions, le détecteur Single Shot MultiBox avec la taille d'entrée de 300×300 (SSD300) obtient les meilleures performances, en termes de taux de précision moyen, de taux de rappel moyen et de score F1. Pour la tâche de classification, DenseNet est plus adapté au problème étudié.

Al-Dhabyani et al [60] traitent la problématique d'augmentation des données et la classification des masses mammaires en échographie du sein, suivant des approches d'apprentissage profond. A cet effet, Al-Dhabyani et al utilisent une nouvelle méthodologie pour l'augmentation des données, à l'aide du Generative Adversarial Network (GAN) [46]. Le système développé use des CNNs pré-entraînés, avec et sans augmentation des données. Les résultats obtenus, confirment une amélioration de la classification des pathologies mammaires, grâce à l'augmentation des données.

2.8 Conclusion

Tout au long du deuxième chapitre, nous avons introduit les concepts de base du machine learning ainsi que ceux de l'apprentissage profond. Nous avons décrit brièvement les applications de l'apprentissage profond dans le domaine de l'imagerie médicale, comme la détection de régions d'intérêt dans une image. À la fin, nous avons cité quelques travaux réalisés en imagerie échographique du sein.

Dans le chapitre suivant, nous appliquons ces connaissances dans la conception du système d'aide à la détection du cancer du sein, dans les images ultrasonores.

Chapitre 3
Conception du système d'aide
à la
détection du cancer sein dans les
images ultrasonores

Chapitre III Conception du système d'aide à la détection du cancer sein dans les images ultrasonores

3.1 Introduction

L'objectif de ce chapitre, est d'étudier les principales architectures des CNN, qui permettent la conception d'un système d'aide à la décision d'images ultrasonores mammaires, pour la distinction des classes, malignes, bénignes et saines. Ce type d'analyse, nécessite des méthodes d'apprentissage profond, pour atteindre des résultats efficaces.

À l'heure actuelle, dans la recherche scientifique, le deep learning suscite un réel engouement pour l'imagerie médicale. Les études abondent, tendant à montrer que l'IA pourrait, diagnostiquer plus sûrement une pathologie, qu'un radiologue [36]. Des méthodes classiques, d'apprentissage automatique, aux techniques d'apprentissage profond, les réseaux de neurones, montrent leurs performances, via les réseaux à convolution (CNN). Dans ce cadre, seront étudiés dans ce chapitre, les architectures des CNN, qui contribuent à la conception de notre système d'aide à la décision.

3.2 Présentation du système d'aide à la détection et la classification de pathologies en échographie du sein

L'objectif attendu par cette étude, est la conception d'un système d'aide à la prise de décision (CAD) dédiée à la classification de pathologies en échographie du sein. Cette approche consiste à assister le radiologue, dans sa tâche parfois délicate, pour l'interprétation des pathologies mammaires. C'est dans ce cadre, que notre conception, est basée sur différentes architectures des CNN, pour estimer les plus efficaces.

3.2.1 Architecture du système développé

Notre recherche concernant ce travail, nous ont permis de constater que toutes les solutions apportées à la classification des pathologies, sont structurées selon la même architecture globale, en deux modules principaux fonctionnant indépendamment. Ce sont les étapes qui constituent notre système (figure 3.1), représentées par : l'extraction de caractéristiques et la classification.

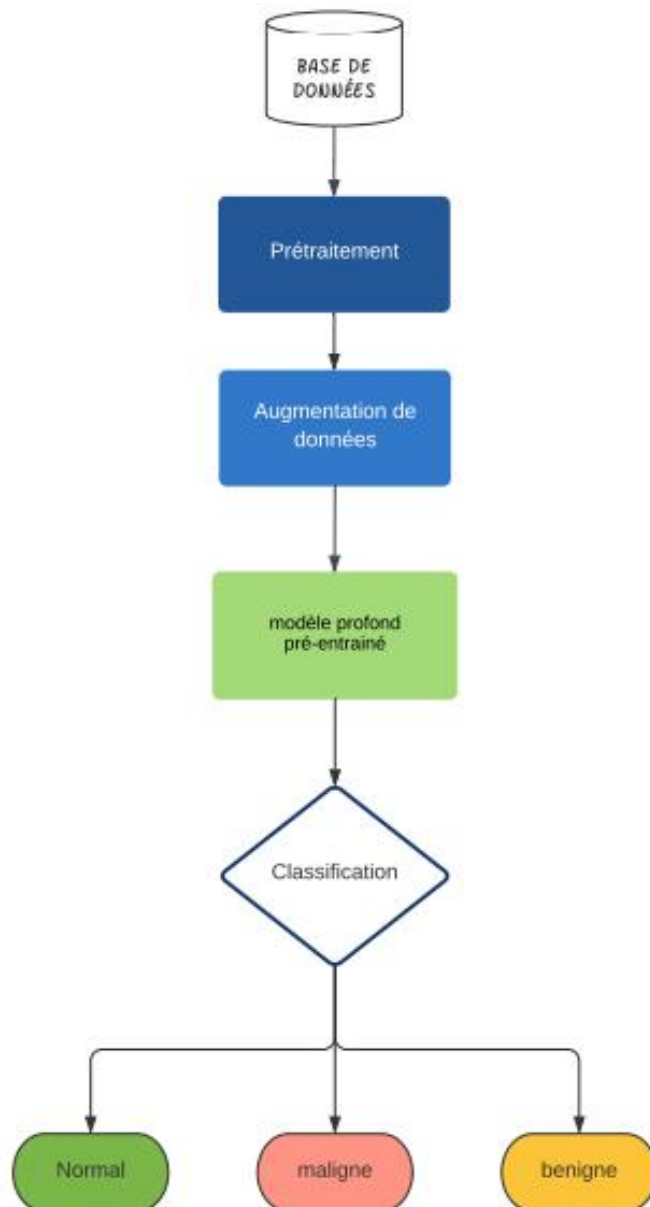


Figure 3.1 : Schéma du modèle adopté

Une fois les images introduites, le système lance le processus d'extraction des caractéristiques qui convertit les données à des configurations plus réduites et optimales, pour que la représentation extraite soit utilisée dans le processus de la classification.

Cette étape réduit les dimensions de l'image à l'entrée, en gardant les données les plus utiles, pour la classification. Ce processus est considéré comme le cœur de notre système. Il est basé sur la technologie des réseaux de neurones convolutionnels (CNN), adoptant trois architectures, le VGG19, le VGG16 et Inception V3 pour la classification des pathologies échographiques mammaires.

Après notre recherche bibliographique, nous proposons différents modèles des CNNs, qui ont été conçus avec l'idée d'utiliser le transfert learning, pour obtenir une meilleure précision en fonction de la base des données choisies.

L'utilisation du concept d'apprentissage par transfert, nous aide à améliorer la précision du modèle, réduire le temps d'entraînement et, à surmonter le problème du manque des données médicales.

Le système adopté repose sur l'idée de bénéficier de l'enchaînement classique, de couches de convolutions et de pooling (dans le cas du VGG16 et VGG19) ainsi que des blocs d'Inception, dans le but d'atteindre les meilleures performances, à la détection des pathologies. Tout d'abord, nous modifions les architectures choisies en remplaçant leurs dernières couches par un nouveau classifieur, ensuite nous les affinons selon la fine-tuning total et le fine-tuning partiel.

3.2.2 Exploration des données

La base de données utilisée est « Dataset Busi » (Breast Ultrasound Images Dataset). Il s'agit d'une base échographique mammaire égyptienne [61]. Cette base de données (Dataset) est divisée en trois parties, chacune est utilisée pour une phase de création du modèle.

- La base d'entraînement (training set) (représentée par 80% de l'ensemble de données d'origine). Cette base est utilisée, pour entraîner le modèle adopté (trouver les bons poids et biais).
- La base de validation (indiquée par 10% de l'ensemble de données d'origine). Nous utilisons ce dataset, afin de minimiser le sur-apprentissage.

- Cette base n'affectera pas directement la modification des poids ; elle permet seulement de vérifier que, toute augmentation du pourcentage de bonne classification (précision) par rapport au dataset d'apprentissage, entraîne une augmentation du pourcentage de bonne classification relativement à un ensemble de données, qui n'a pas encore été testé.
- Si la précision sur le dataset d'entraînement augmente, mais la précision sur le dataset de validation reste la même ou diminue, cela implique un sur-apprentissage, par conséquent, nous devons arrêter l'entraînement.
- Dataset de test ou Test set (10% de l'ensemble de données d'origine) : ce dataset est utilisée une fois le modèle finit l'entraînement. Nous l'utilisons uniquement pour tester la solution finale afin de confirmer la précision réelle du réseau.

Il est fortement recommandé de ne pas ignorer, la phase de test. En effet, un algorithme qui classifie bien lors de la phase de validation, ne signifie pas forcément que c'est le modèle le plus convenable, au problème traité. Au cours de la phase de test, le modèle final est utilisé afin de classifier de nouveaux échantillons. Par conséquent, si la précision du modèle est très mauvaise pendant le processus du test, tout le processus de la conception du modèle doit être remis en question.

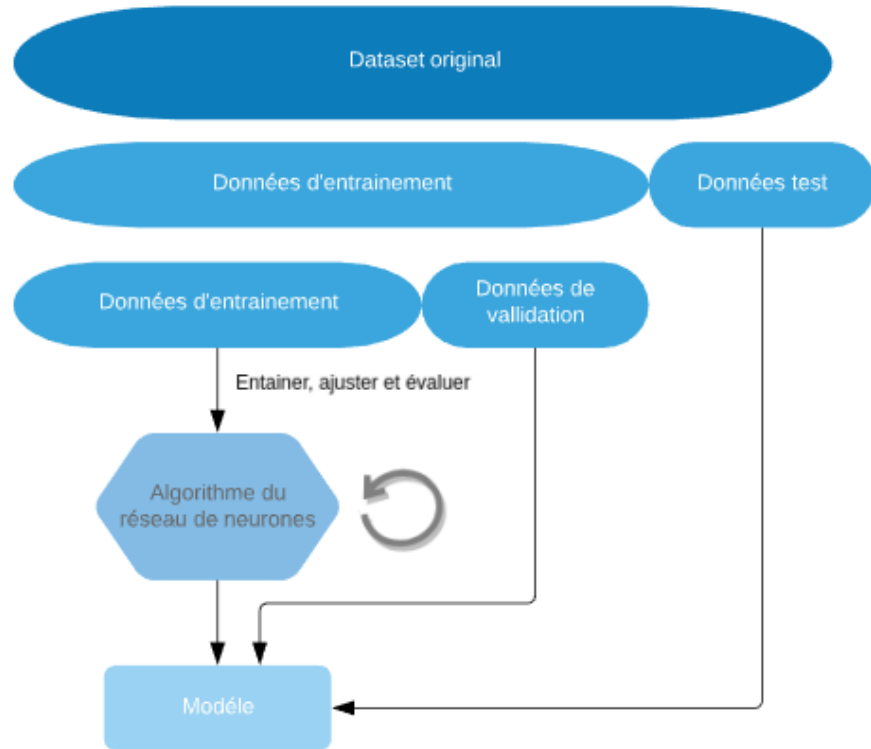


Figure 3.2 : Utilisation des différents datasets dans l’entraînement du modèle final

3.2.3 Prétraitement des images

Le prétraitement des images est une étape importante, pour le processus ; car il permet d’uniformiser les images à traiter. En effet, les images sont sur différentes formes, différentes résolutions, et en niveaux de gris. Elles sont alors, toutes redimensionnées à une taille 224*224 pixels.

Ce redimensionnement nous permettra aussi, de gagner de l’espace mémoire.

3.2.4 Augmentation des données

La quantité de données d’apprentissage est cruciale, pour estimer les réseaux convolutionnels. Or, l’augmentation de données est fondamentalement essentielle, afin d’améliorer les performances des réseaux. Elle consiste à augmenter artificiellement la taille de la base de données, en appliquant certaines transformations sur l’ensemble initial d’apprentissage, pour créer de nouveaux exemples artificiels.

Ainsi, les transformations appliquées, ne changent pas la nature de la classe détectée. L'augmentation de données permet de créer artificiellement, des invariances du réseau final et d'augmenter les performances en généralisation.

Parmi les opérations classiques utilisées pour l'augmentation des données (tableau 3.1), on peut trouver : la translation, la rotation, l'agrandissement, et d'autres transformations plus avancées comme La modification de la luminosité ou du contraste de l'image qui peuvent être également ajoutées selon les besoins de l'utilisateur. L'injection de bruit à l'entrée d'un réseau de neurones, peut également être considérée comme une forme d'augmentation des données [52].

	Images	Voix	Texte
Technique d'augmentation	Translation, rotation, réflexion, re-dimension, rognage, extraction de zones, altération des couleurs	Manipulation du signal, manipulation du spectrogramme, injection de bruit	Substitution lexicale par des synonymes, injection de bruit, transformation de surface(expressions régulières)

Tableau 3.1 : Techniques d'augmentation des données selon le type [52]

Dans notre cas, nous effectuons une augmentation des données, en appliquant des déformations géométriques aux images, comme la symétrie horizontale, la verticale, ainsi que l'effet de miroir.

3.3 Présentation de l'architecture VGG19

Dans ce modèle, nous réalisons un transfert d'apprentissage avec Oxford VGG19 comme modèle de CNN de base.

Nous avons choisi ce modèle, car il a déjà fait ces preuves [62] avec 7.3% d'erreur pour la classification d'image sur 1000 catégories d'ImageNet Large Scale Visual Recognition Challenge (ILVRC) [62].

Comme l'illustre la figure 3.4, le CNN VGG19 contient en total 19 couches : 16 couches de convolution, 3 couches entièrement connectées, 5 couches MaxPool et 1 couche SoftMax. Le VGG19 est fait pour catégoriser des images dans 1000 classes. Nous avons donc adapté ce modèle, à la classification d'images ultrasonores, que nous souhaitons réaliser.

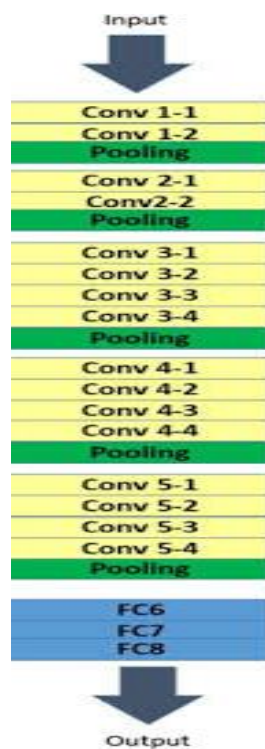


Figure 3.3 : Architecture du VGG19 [63]

3.4 Présentation de l'architecture VGG16

Le VGG16 est un modèle de réseaux de neurones convolutif proposé par Simonyan et al [55] de l'Université d'Oxford dans l'article intitulé « réseaux convolutionnels très profonds pour la reconnaissance d'images à grande échelle ». Il est constitué de plusieurs couches, dont 13 couches de convolution et 3 couches entièrement connectées [52, 55]. Ce modèle utilise des filtres de noyaux 3X3 dans les couches de convolution (qui est la plus petite taille pour capturer la notion de gauche / droite, haut / bas, centre) et un filtre de taille 2X2 en max pooling.

Toutes les couches cachées, sont équipées de la non-linéarité de rectification (ReLU). Les cartes de caractéristiques (maps features) de la couche finale, sont représentées sous forme de vecteurs avec des valeurs scalaires, suivant la fonction de classification softmax. Les couches entièrement connectées sont coûteuses en termes de calcul. A cet effet, des approches alternatives ont été proposées au cours des dernières années, comme la mise en œuvre d'une couche globale de pooling moyenne, qui aide à réduire le nombre de paramètres dans le réseau de manière significative.

Le modèle atteint une précision de 92,7% dans le top 5 des tests dans ImageNet ; qui est un jeu de données de plus de 14 millions d'images, appartenant à 1000 classes.

C'était l'un des fameux modèles soumis à la conférence ILSVRC-2014 [76]. Ses résultats sont beaucoup plus élevés, par rapport à ceux obtenus par l'architecture AlexNet [64,65].



Figure 3.4 : Architecture du VGG16 [65]

3.5 Présentation de l'architecture Inception v3

Inception v3 est un modèle de reconnaissance d'images couramment utilisé qui a démontré, sur l'ensemble de données ImageNet, une justesse supérieure à 78,1 %. Ce modèle est l'aboutissement de nombreuses idées, développées par plusieurs chercheurs au fil des ans [66].

L'architecture convolutionnelle profonde d'Inception, a été introduite sous le nom de GoogLeNet (Inception-v1). Plus tard, l'architecture Inception a été affinée de diverses manières, d'abord par l'introduction de la normalisation par lots (Inception-v2). Plus tard, par des idées de factorisation supplémentaires dans la troisième itération, qui est appelée Inception-v3 [67].

Ainsi, Inception-v1 / Inception-v2 se rapportent à une normalisation par lots, tandis qu'Inception-v3 porte sur des idées de factorisation [66].

Le modèle lui-même est constitué de composants de base symétriques et asymétriques ; incluant les opérations de convolutions, de pooling moyen, de pooling maximal, de concaténations, d'abandons et de couches entièrement connectées. La normalisation par lots (batchnorm), est amplement utilisée dans le modèle et, appliquée aux entrées d'activation. La perte est calculée via la fonction Softmax [67].

3.5.1 Factorisation des convolutions

Le but de la factorisation des convolutions est de réduire le nombre de paramètres sans diminuer l'efficacité du réseau.

3.5.1.1 Factorisation en petites convolutions

Deux convolutions 3×3 remplacent une convolution 5×5 (figure 3.5).

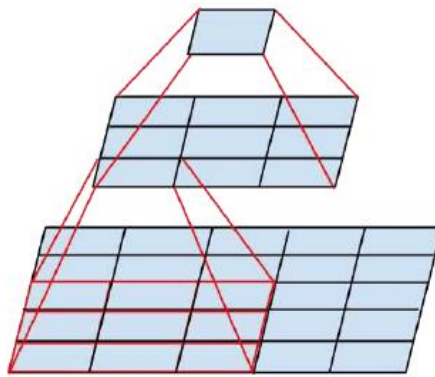


Figure 3.5 : Convolution 5x5 selon deux convolutions 3x3 [68]

- En utilisant une seule couche du filtre 5×5 , le nombre de paramètres, est égal à $5 \times 5 = 25$
- En utilisant 2 couches de filtres 3×3 , le nombre de paramètres correspond à $3 \times 3 + 3 \times 3 = 18$. Le nombre de paramètres est ainsi réduit de 28%. Une technique similaire a déjà été mentionnée dans VGGNet [68]. Il s'agit de la même technique utilisée, afin de construire l'un des nouveaux modules Inception, 'Inception Module A' (figure 3.6).

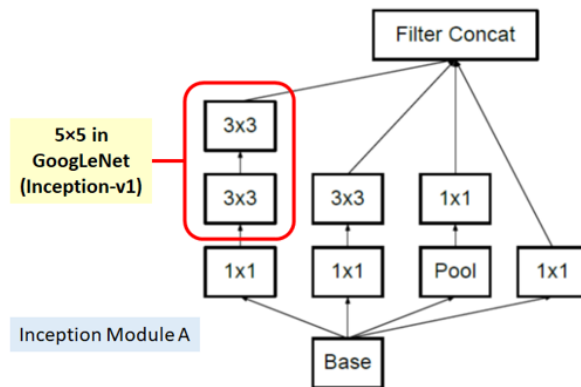


Figure 3.6 : Inception module A [67]

3.5.1.2 Factorisation en convolutions asymétriques

Une convolution 3×1 suivie d'une convolution 1×3 remplace une convolution 3×3 suivant la figure 3.7.

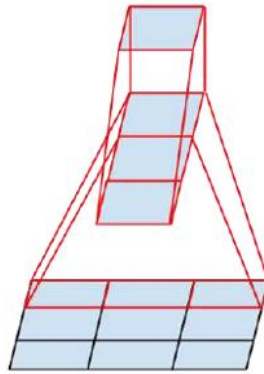


Figure 3.7 : Convolution 3x3 suivant une convolution 3x1 et une autre de 1x3 [68]

En utilisant un filtre 3×3 , le nombre de paramètres est de $3 \times 3 = 9$.

En utilisant des filtres 3×1 et 1×3 , le nombre de paramètres est de $3 \times 1 + 1 \times 3 = 6$. Il est ainsi réduit de 33% [70]. Avec cette technique, l'un des nouveaux modules Inception 'Inception Module B' est construit suivant la figure 3.8.

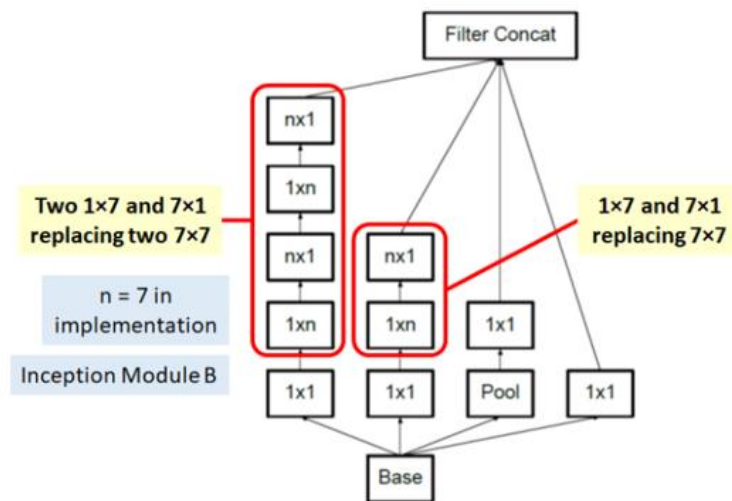


Figure 3.8 : Inception module B [67]

Le module de lancement C, est également proposé pour promouvoir des représentations de haute dimension selon les descriptions des auteurs comme suit :

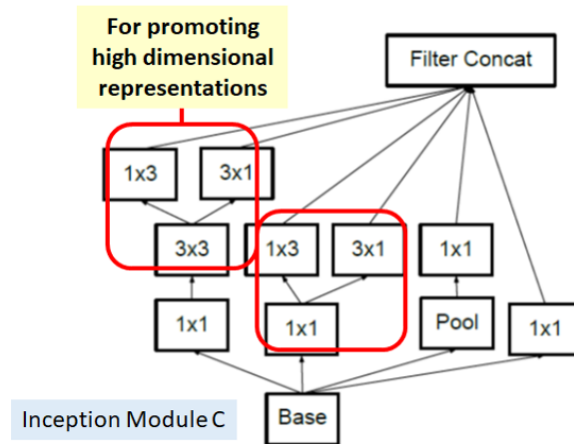


Figure 3.9 : Inception module C [67]

Les auteurs de cette conception [70], proposent ces 3 types de modules pour le modèle inception. Avec la factorisation, le nombre de paramètres est réduit pour l'ensemble du réseau, il est moins susceptible d'être sur-ajusté et par conséquent, le réseau peut être plus profond. Dans ce cadre, avec 42 couches de profondeur, le coût de calcul est seulement environ 2,5 plus élevé que celui de GoogLeNet, beaucoup plus efficace que celui de VGGNet [68].

Un diagramme général du modèle est présenté ci-dessous :

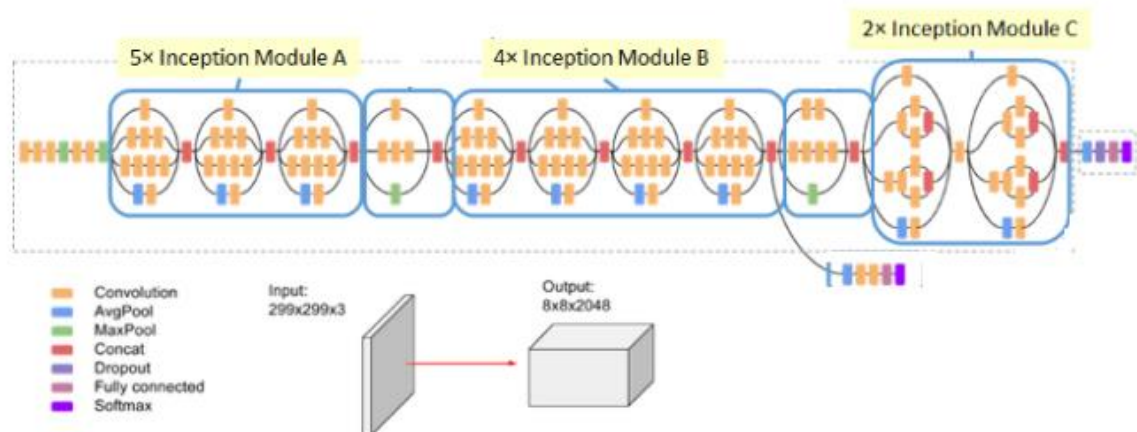


Figure 3.10 : Architecture du réseau Inception V3 [67]

3.6 Apprentissage par transfert

L'apprentissage par transfert (transfer learning) est une forme particulière d'apprentissage automatique, qui fait référence au transfert de connaissances, tirées d'une tâche à une autre [78]. Cette technique est inspirée, du comportement humain. Par exemple, il est plus facile d'apprendre à piloter une moto, si on sait déjà faire du vélo, puisque ces deux dernières requièrent les mêmes compétences. L'apprentissage par transfert (transfer Learning) des réseaux de neurones, est basé sur le même principe. Si on a entraîné un réseau de neurones pour distinguer des chiens, des chats, à partir de photos, alors on pourra s'appuyer sur ce réseau, pour deviner à quelle race appartient un chien. Encore mieux, on pourra utiliser notre réseau pour classer des loups.

“ EN TOUTE GÉNÉRALITÉ LE TRANSFER LEARNING CONSISTE À RÉUTILISER TOUT OU PARTIE D'UN MODÈLE, PRÉENTRAÎNÉ SUR UNE TÂCHE D'APPRENTISSAGE 'A', COMME BRIQUE DE BASE POUR CONSTRUIRE UN MODÈLE PRÉDICTIF POUR UNE TÂCHE 'B' EN BÉNÉFICIAINT DE LA CONNAISSANCE ACQUISE SUR LA TÂCHE 'A'.

Les domaines d'application du 'Transfer Learning', sont nombreux. Les méthodes de transfert de connaissances, sont principalement utilisées pour la reconnaissance d'image ainsi que, le traitement automatique du langage. Ces deux domaines d'apprentissage sont chronophages. C'est pourquoi l'apprentissage par transfert, apporte un souffle nouveau pour tenter d'optimiser ces traitements, en utilisant des modèles déjà entraînés [69].

3.6.1 Branches du Transfer Learning

Il existe plusieurs branches du 'transfer Learning', selon la problématique qui se pose. La dichotomie principale est la même que pour celle du machine learning : la labellisation des données. Si ces dernières sont étiquetées suivant le domaine cible, alors il s'agit de l'apprentissage inductif par transfert (inductive transfer learning), sinon, c'est l'apprentissage supervisé par transfert (unsupervised transfer learning) [69].

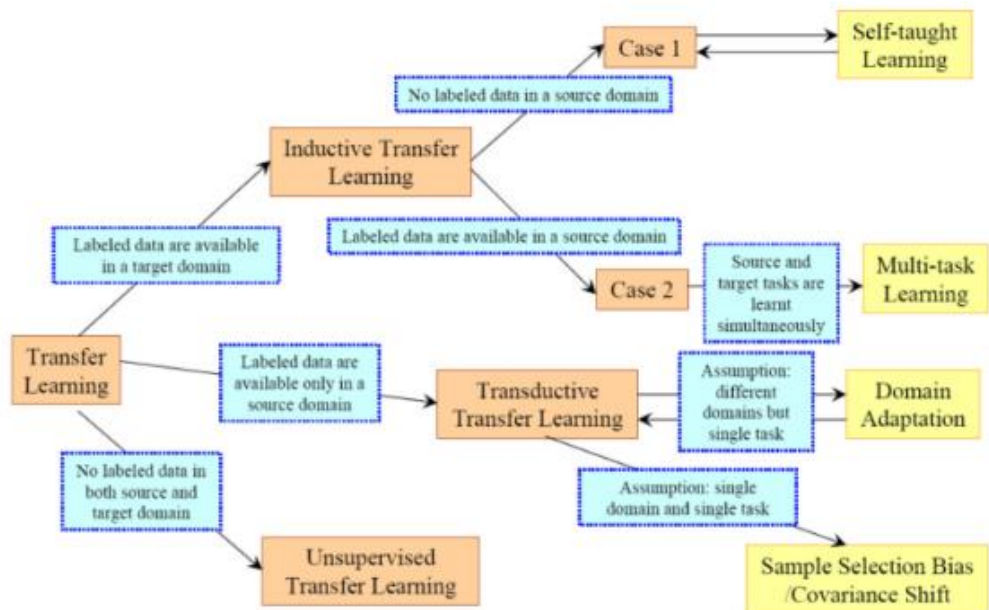


Figure 3.11 : Types de l'apprentissage par transfert [69]

3.6.2 Réutilisation d'un modèle pré-entraîné

Lorsque on réutilise un modèle pré-entraîné [70], on commence tout d'abord par supprimer le classificateur d'origine, puis on ajoute un nouveau classificateur selon les besoins. On doit affiner par la suite, le modèle selon l'une des trois stratégies :

➤ Stratégie 1 : fine-tuning total

On remplace la couche fully-connected du réseau pré-entraîné, par un classifieur adapté au nouveau problème et, initialisé aléatoirement. Toutes les couches sont par la suite, entraînées. Cette stratégie doit être utilisée, lorsque la nouvelle base de données est grande. Dans ce cas, on peut se permettre d'entraîner tout le réseau, sans risque de sur-apprentissage. De plus, comme les paramètres de toutes les couches (sauf de la dernière) sont initialement ceux du réseau pré-entraîné, la phase d'apprentissage sera terminée plus rapidement, que si l'initialisation avait été aléatoire [70].

➤ Stratégie 2 : fine-tuning partiel

Il s'agit d'un mélange des stratégies 1 et 3. On remplace à nouveau, la dernière couche 'fully- connected' par le nouveau classifieur initialisé de manière aléatoire et, on fixe les paramètres de certaines couches, du réseau pré-entraîné.

Ainsi, en plus du classifieur, on entraîne sur les nouvelles données les couches non-fixées, qui correspondent généralement aux plus hautes du réseau.

On utilise cette stratégie lorsque la nouvelle base des données est petite, mais très différente de celle du pré-entraînement. Pour cela, la stratégie 1 qui consiste à entraîner tout le réseau, n'est pas envisageable à cause du risque de sur-appariement [70].

D'autre part, on élimine également la stratégie 3, car les nouvelles données ont très peu de points communs avec les anciennes. Utiliser ainsi, les caractéristiques (features), du réseau pré-entraîné, pour les représenter n'est pas une bonne idée [70].

➤ **Stratégie 3 : extraction des caractéristiques**

Cette stratégie (figure 3.12) consiste à utiliser le réseau de neurones, comme une variable fixe d'extraction. Pour cela, on retire la dernière couche 'fully-connected' et, on gèle (fixe) les poids du modèle pour les utiliser, comme une variable fixe. On entraîne par la suite le nouveau dataset, suivant une classification linéaire (SVM linéaire ou Softmax) [70].

La stratégie 3 doit être utilisée, lorsque la nouvelle base de données est petite et, similaire aux données de pré-entraînement. En effet, entraîner le réseau sur peu d'images, est dangereux puisque le risque de sur-apprentissage, est très élevé [70].

De plus, si les nouvelles données ressemblent aux anciennes, elles peuvent alors être représentées avec les mêmes caractéristiques.

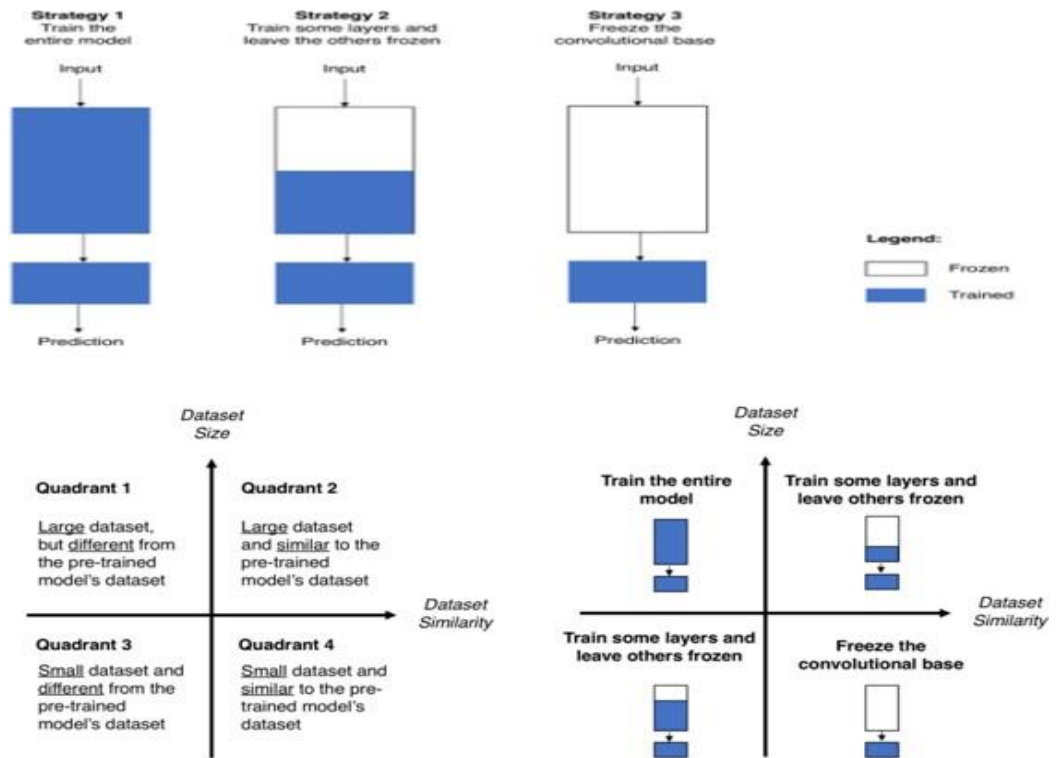


Figure 3.12 : Stratégies du fine-tuning [70]

3.7 Conclusion

Ce chapitre était consacré à l'étude conceptuelle pour la réalisation d'une approche de détection de lésions mammaires. Nous avons introduit les trois modèles VGG19, VGG16 et Inception V3 utilisés pour la classification des images, sur lesquels nous nous basons, pour concevoir le système d'analyse profonde d'images ultrasonores.

Dans le prochain chapitre, nous présentons l'utilisation de ces trois architectures pour les tester sur la classification de pathologies en échographie du sein, afin de révéler les performances du système développé, suivant l'apprentissage par transfert (transfer learning). Une comparaison entre les différents modèles, sera effectuée sur la base des métriques évaluées.

1

2

3

CHAPITRE 4

MISE EN ŒUVRE DU SYSTEME ET

RESULTATS

Chapitre IV Mise en œuvre du système et résultats

4.1 Introduction

L'objectif de ce projet de fin d'étude est de concevoir et de réaliser une application, qui permet de la classification de pathologies en échographie du sein. Nous nous sommes intéressées à l'utilisation des réseaux profonds pré-entraînés de type CNN.

Dans ce chapitre nous présentons l'implémentation de notre système et les différentes étapes de réalisation, l'environnement de développement, ainsi que les différents outils utilisés pour effectuer ce travail. Nous poursuivons ce chapitre, par la présentation des différents résultats expérimentaux obtenus, quelques captures d'écrans de notre application ainsi que les performances obtenues.

4.2 Environnement de travail

Avant de commencer l'implémentation de notre travail, nous allons tout d'abord spécifier les outils utilisés qui nous ont semblé être un bon choix, vu les avantages qu'ils offrent.

4.2.1 Matériel utilisé

L'implantation de nos codes a été réalisée, grâce à un PC portable muni d'un processeur Intel Core i5 2410U @ 2.30GHz et d'une capacité de mémoire vive « RAM de 4 Go ».

4.2.2 Langage de programmation Python

Python est un langage de programmation open source créé par le néerlandais Guido van Rossum en 1991. Il tire son nom de l'émission Monty Python's Flying Circus [71]. Il est important de noter qu'il s'agit d'un langage interprété (il n'y a pas d'étape de compilation) et orienté objet, avec une sémantique dynamique. Il est très sollicité par une large communauté de développeurs et de programmeurs. Grâce à son succès mondial, python dispose d'une vaste bibliothèque standard complète.

En tant que langage de programmation de haut niveau, Python permet aux programmeurs de se focaliser sur ce qu'ils font plutôt que, sur la façon dont ils le font.

Ainsi, écrire des programmes, prend moins de temps que dans un autre langage. Il s'agit d'un langage idéal pour les débutants. Python est un langage simple, facile à apprendre et permet une bonne réduction du cout de la maintenance des codes [71, 72, 73].

Ce langage de programmation présente de nombreuses caractéristiques intéressantes :

- Il est gratuit et multiplateforme. Depuis les mini-ordinateurs Raspberry Pi jusqu'aux supercalculateurs, Il est fonctionnel sur Windows, Mac OS X, Linux, Android, iOS.
- C'est un langage de haut niveau. Il demande relativement peu de connaissances sur le fonctionnement d'un ordinateur pour être utilisé.
- C'est un langage interprété. Un script Python n'a pas besoin d'être compilé, pour être exécuté, contrairement à des langages, comme le C ou le C++.
- C'est un langage dynamique et extensible. Il favorise la programmation structurée fonctionnelle et orientée objet. C'est-à-dire qu'il est possible de concevoir en Python des entités, qui miment celles du monde réel (une cellule, une protéine, un atome, etc.) avec un certain nombre de règles de fonctionnement et d'interactions.
- Il est relativement simple à prendre en main.
- La syntaxe de Python est très simple et, combinée à des types de données évolués (listes, dictionnaires, ...), conduit à des programmes à la fois très compacts et très lisibles.
- Il gère ses ressources (mémoire, descripteurs de fichiers...) sans intervention du programmeur.
- Enfin, il est très utilisé en bio-informatique ainsi que l'intelligence artificielle et, plus généralement en analyse de données.

4.2.3 Anaconda

Anaconda est une plateforme de gestion de paquets, d'environnement et de distribution des langages de programmation R et Python. Cette plateforme est utilisée pour le calcul scientifique, l'analyse prédictive et le traitement des données à grande échelle.

Cette dernière est destinée aux scientifiques manipulant des données, aux développeurs, aux analystes travaillant dans des entreprises ou, des organisations. Anaconda existe en version gratuite appelée « Distribution » et en version payante nommée « Enterprise ».

La version « distribution » propose deux moyens d'accéder à ses fonctions. Soit de manière graphique avec Anaconda-Navigator, soit en lignes de commande (depuis Anaconda Prompt sur Windows, ou un Terminal pour Linux ou macOS) [74,75].

4.2.3.1 Anaconda Navigator

Anaconda Navigator (figure 4.1) est une interface utilisateur graphique de bureau (GUI) incluse dans la distribution Anaconda, qui nous permet de lancer des applications et de gérer facilement les paquets, les environnements et les canaux conda. Cette interface peut rechercher des paquets sur Anaconda Cloud, ou dans un dépôt Anaconda local. Il est disponible pour Windows, macOS et Linux [75, 76].

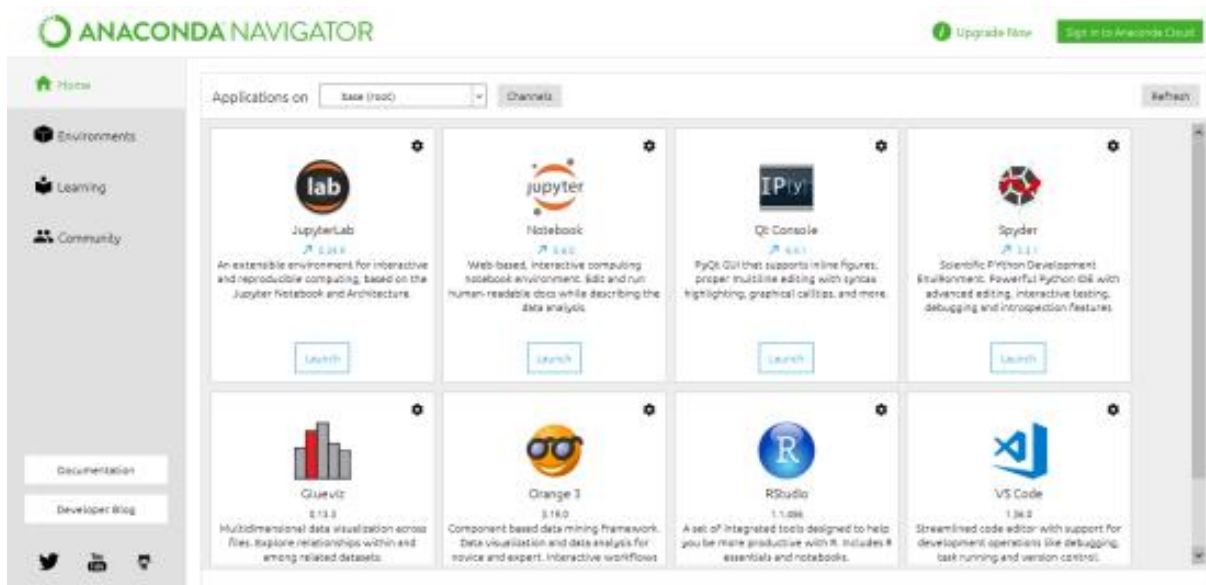


Figure 4.1 : Représentation de l'interface d'anaconda [75]

Anaconda-Navigator propose différents services (déjà installés, ou à installer). Son onglet Home permet de lancer le service désiré. L'un des principaux services à utiliser pour développer des programmes python, est :

4.2.3.2 Spyder

Spyder est un environnement scientifique puissant écrit en Python, pour Python et conçu par et pour des scientifiques, ingénieurs et analystes de données.

Il offre une combinaison unique des fonctionnalités avancées d'édition, d'analyse, de débogage et de profilage d'un outil de développement complet avec l'exploration de données, l'exécution interactive, l'inspection approfondie et les belles capacités de visualisation d'un

logiciel scientifique. Au-delà de ses nombreuses fonctionnalités intégrées, ses capacités peuvent être encore étendues grâce à son système de plugins et son API. De plus, Spyder peut également être utilisé comme bibliothèque d'extension PyQt, ce qui permet aux développeurs de s'appuyer sur ses fonctionnalités et d'intégrer ses composants, tels que la console interactive, dans leur propre logiciel PyQt [77].

4.2.4 TensorFlow

Nous avons utilisé dans notre travail TensorFlow, qui est un framework de programmation pour le calcul numérique, développé par l'équipe Google Brain et qui a été rendu Open Source, par Google en Novembre 2015. La version 1.0.0 a été lancée en février 2017 [78]. Depuis sa libération, TensorFlow n'a cessé de gagner en popularité, pour devenir très rapidement l'un des frameworks les plus utilisés pour le Deep Learning, comme le montrent les dernières comparaisons, faites par François Chollet (auteur de la librairie Keras) [79]. La figure ci-dessous, représente une comparaison entre les différentes bibliothèques utilisées pour le deep learning.

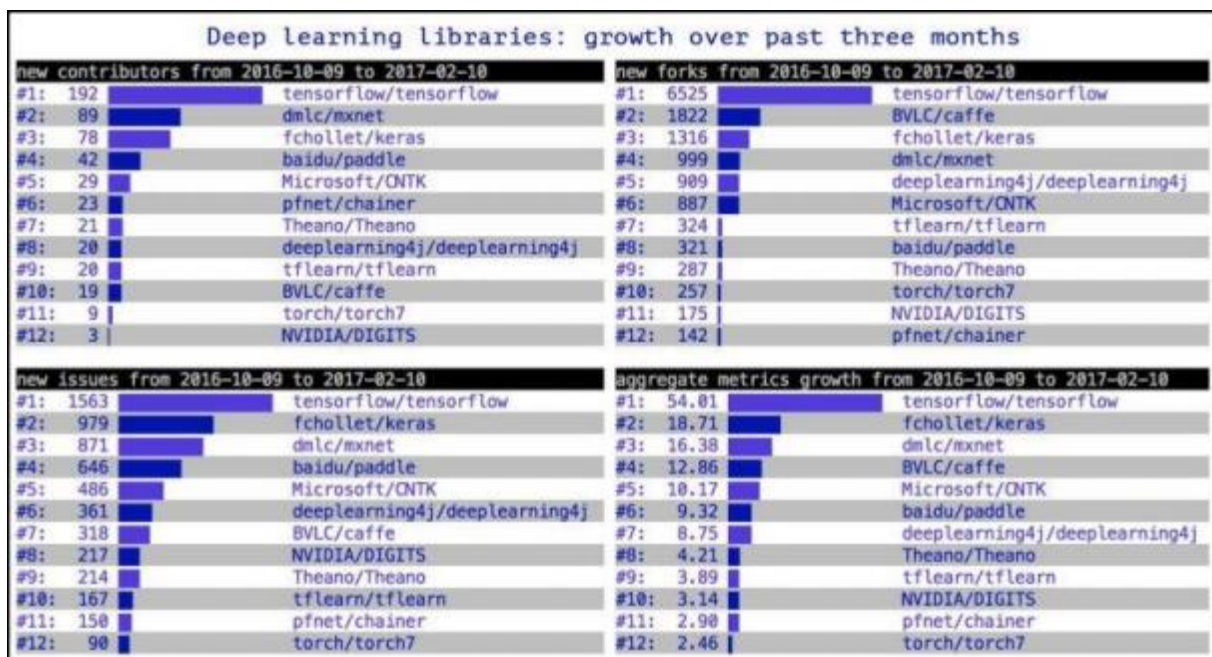


Figure 4.2 : Comparaison entre les différentes bibliothèques du deep learning [80]

TensorFlow est très populaire pour ses nombreux avantages qu'on cite ci-dessous:

- Multi-plateformes (Linux, Mac OS, et même Android et iOS).

- APIs en Python, C++, Java et Go (l'API Python est plus complète, c'est sur celle-ci que nous travaillons).
- Temps de compilation très courts.
- Support des calculs sur CPU, GPU et même le calcul distribué sur cluster.
- Documentation extrêmement bien fournie, avec de nombreux exemples et tutoriels.

TensorFlow est aujourd'hui particulièrement utilisé pour le Deep Learning et, donc les réseaux de neurones. Son nom est notamment inspiré du fait que les opérations courantes sur des réseaux de neurones, sont principalement faites via des tables de données multidimensionnelles, appelées Tenseurs (Tensor). Un Tensor à deux dimensions est l'équivalent d'une matrice [77]. Aujourd'hui, les principaux produits de Google sont basés sur TensorFlow, tels que : Gmail, Google Photos, etc...

4.2.5 Keras

Keras est une API de réseaux de neurones de haut niveau, écrite en Python et capable de fonctionner sur TensorFlow ou Theano.

Keras a été développé en mettant l'accent sur l'expérimentation rapide. Être capable d'aller de l'idée à un résultat avec le moins de délai possible, est la clé pour faire de bonnes recherches. Keras a été développé dans le cadre de l'effort de recherche du projet ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System). Son principal auteur et mainteneur est François Chollet, un ingénieur Google. En 2017, l'équipe TensorFlow de Google a décidé de soutenir Keras dans la bibliothèque principale de TensorFlow. Chollet a expliqué que Keras a été conçue, comme une interface plutôt que comme un cadre d'apprentissage end to end. Cette bibliothèque présente un ensemble d'abstractions de niveau supérieur et plus intuitif, qui facilitent la configuration des réseaux neuronaux indépendamment de la bibliothèque informatique de Backend. Microsoft travaille également à ajouter un Backend CNTK à Keras aussi [81].

Keras est également l'un des outils favoris des chercheurs, il s'agit de la deuxième mention la plus populaire dans les papiers scientifiques mis en ligne sur arXiv.org (figure 4.3). Elle est utilisée par des chercheurs de grandes institutions scientifiques, comme le CERN et la NASA [82].

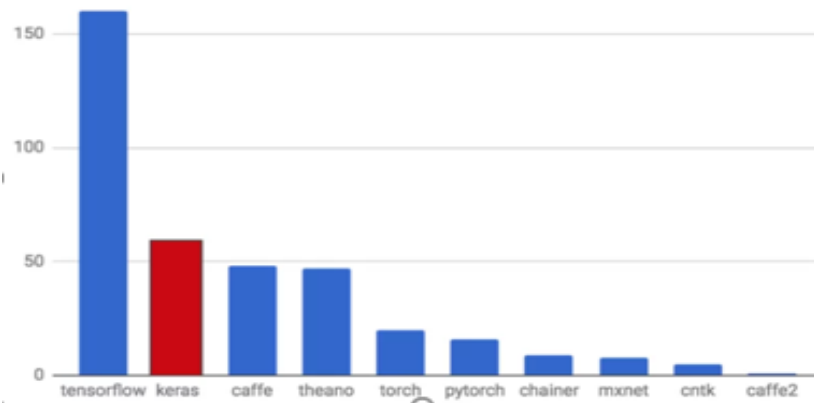


Figure 4.3 : arXiv mentions, october 2017 [82]

4.2.6 Numpy

Numpy [72, 83] est une librairie Python et un projet open-source créé par Travis Oliphant. Elle offre la pièce fondamentale, pour faire des calculs scientifiques avec Python. Elle est sous licence BSD, ce qui permet à chacun de l'utiliser librement pour ses besoins (académiques, privés et professionnels).

Elle est majoritairement écrite en C (pour les parties clé à optimiser) et Python. Quelques portions sont également écrites, en C++ et Fortran.

Elle étend les capacités de Python, pour travailler sur des tableaux et matrices à n dimensions de façon bien plus optimisée. Elle offre ainsi, des fonctions mathématiques de haut niveau sur ces objets.

4.2.7 Flask

Flask est un micro-framework python facile et simple, qui permet de faire des applications web évolutives. Flask dépend de la boite à outils WSGI de Werkzeug et du moteur de templates Jinja [84].

4.2.8 Html

Le HTML est un langage qui a été créé en 1991 [85]. Les sigles « HTML » sont l'abréviation de « HyperText Markup Language ». Il désigne un type de langage informatique descriptif.

Il s'agit plus précisément d'un format de données utilisé dans l'univers d'Internet, pour la mise en forme des pages Web. Il permet, entre autres, d'écrire de l'hypertexte, mais aussi d'introduire des ressources multimédias dans un contenu [86].

4.2.9 Tkinter

Tkinter (Tk interface) est un module intégré à la bibliothèque standard de Python, bien qu'il ne soit pas maintenu directement par les développeurs de Python. Il offre un moyen de créer des interfaces graphiques via Python. Tkinter est disponible sur Windows et la plupart des systèmes Unix. Les interfaces que nous pourrions développer auront donc toutes les chances, d'être portables d'un système à l'autre [87].

4.3 Base de données

La base de données utilisée est « Dataset Busi » (Breast Ultrasound Images Dataset) [61]. Elle contient des images échographiques du sein chez les femmes âgées de 25 à 75 ans. Ces données ont été collectées en 2018 et sont catégorisées en 3 classes : normales, bénignes et malignes (tableau 4.1).

L'ensemble de données se compose de 780 images et leurs masques au format PNG avec une taille d'image moyenne de 500 * 500 pixels. Ces images échographiques du sein (figure 4.4) peuvent produire d'excellents résultats dans la classification, la détection, la prédiction et la segmentation du cancer du sein en combinaison avec l'apprentissage automatique.

Bégnine	maligne	normal
887	430	266

Tableau 4.1 : Répartition par classe de la base de données [61]

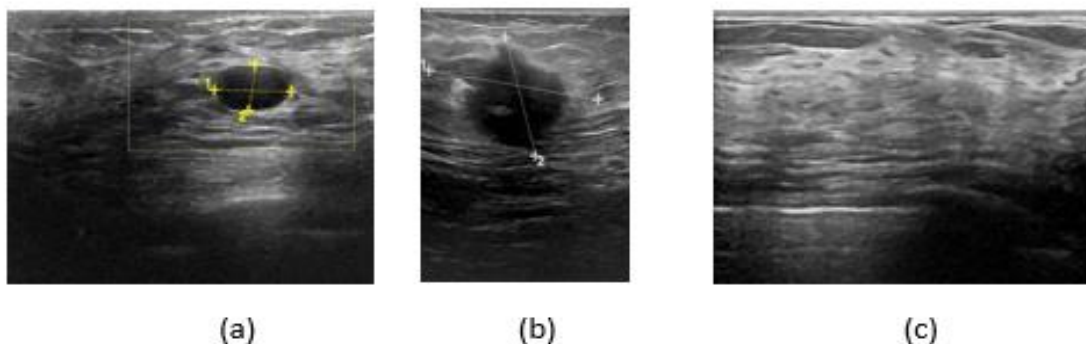


Figure 4.4 : Illustration des images présentes dans la base BUSI (a) bénigne, (b) maligne, (c) normale

4.4 Architecture du système 'BreastULTRASCNN'

Du fait du nombre limité d'images de la base utilisée, nous avons choisi de développer notre algorithme en nous basant sur des modèles profonds pré-entraînés en l'occurrence VGG19, VGG16 et Inception V3. Nous avons dans un premier temps légèrement modifié l'architecture des modèles pré-entraînés, en remplaçant leurs dernières couches par des couches complètement connectés (CC). Pour chaque architecture, nous affinons selon deux stratégies :

1. Fine-tuning total : toutes les couches sont entraînées.
2. Fine-tuning partiel : on fixe les paramètres de certaines couches du réseau pré-entraîné. En plus du classifieur, on entraîne alors, sur les nouvelles données les couches non fixées.

Les tableaux ci-dessous décrivent l'architecture des couches ajoutées sur chaque modèle.

Nom Couche (Type)	Taille	Paramètres
CC	3	Activation : Softmax

Tableau 4.2 : Paramètres des couches ajoutées au modèle VGG19 avec le fine-tuning total

Nom Couche (Type)	Taille	Paramètres
Dense_1 (cc)	4000	Activation : Relu
Dropout_3 (Dropout)	4000	
Dense_2(CC)	1003	Activation : Relu
Dence_3 (CC)	1003	Activation : Relu
Dence_4 (CC)	1003	Activation : Relu
Dense_5(CC)	1003	Activation : Relu
Dropout_4 (Dropout)	1003	
Dense_6(CC)	1003	Activation : Relu
Dense_7(CC)	3	Activation : Softmax

Tableau 4.3 : Paramètres des couches ajoutées au modèle VGG19 avec le fine-tuning partiel

Nom Couche (Type)	Taille	Paramètres
Dense_6 (CC)	4000	Activation : Relu
Dense_7(CC)	1003	Activation : Relu
Dence_8 (CC)	1003	Activation : Relu
Dence_9 (CC)	1003	Activation : Relu
Dence_11 (CC)	3	Activation : Softmax

Tableau 4.4 : Paramètres des couches ajoutées au modèle VGG16 avec le fine-tuning total

Nom Couche (Type)	Taille	Paramètres
Dense_6 (cc)	4000	Activation : Relu
Dense_7(CC)	1003	Activation : Relu
Dence_8(CC)	1003	Activation : Relu
Dence_9 (CC)	1003	Activation : Relu
Dense_10(CC)	1003	Activation : Relu
Dense_11(CC)	1003	Activation : Relu
Dropout_4 (Dropout)	1003	
Dense_12(CC)	3	Activation : Softmax

Tableau 4.5 : Paramètres des couches ajoutées au modèle VGG16 avec le fine-tuning partiel

Nom Couche (Type)	Taille	Paramètres
Dense_3 (cc)	1000	Activation : Relu
Dense_4(CC)	1000	Activation : Relu
Dence_5 (CC)	512	Activation : Relu
Dropout_1 (Dropout)	512	
Dence_6(CC)	256	Activation : Relu
Dropout_2 (Dropout)	256	
Dense_7(CC)	128	Activation : Relu
Dropout3_(Dropout)	128	
Dense_8(CC)	64	Activation : Relu

Dense_9(CC)	3	Activation : Softmax
-------------	---	----------------------

Tableau 4.6 : Paramètres des couches ajoutées au modèle Inception V3 avec le fine-tuning total

Nom Couche (Type)	Taille	Paramètres
Dense_3 (cc)	1000	Activation : Relu
Dense_4(CC)	1000	Activation : Relu
Dence_5 (CC)	512	Activation : Relu
Dence_6(CC)	256	Activation : Relu
Dense_7(CC)	128	Activation : Relu
Dense_8(CC)	64	Activation : Relu
Dense_9(CC)	3	Activation : Softmax

Tableau 4.7 : Paramètres des couches ajoutées au modèle Inception V3 avec le fine-tuning partiel

4.5 Résultats et discussion

a) Modèle VGG19 avec le fine-tuning total

Tout d'abord, nous commençons par entraîner un réseau, qui se base sur un modèle profond pré-entraîné VGG19. Nous avons modifié son architecture en remplaçant son MLP, par une seule couche entièrement connectée. Ce réseau possède 20,048,963 paramètres entraînaibles. L'architecture est résumée dans la figure 4.5.

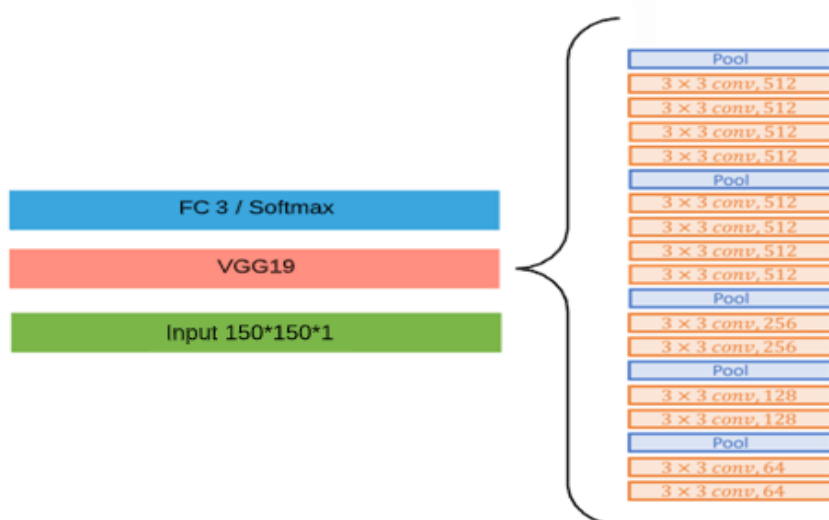


Figure 4.5 : Illustration du modèle de base

Dans le tableau (Input : 150x150x1) qui suit, nous présentons les paramètres utilisés au niveau de chaque couche de convolution et de max-pooling du réseau.

Type	Patch size x nombre de convolution / Stride	Output
Conv	3x3x64/1	150x150x64
Conv	3x3x64/1	150x150x64
Pool	2x2x64/2	75x75x64
Conv	3x3x128/1	75x75x128
Conv	3x3x128/1	75x75x128
Pool	2x2x128/2	37x37x128
Conv	3x3x256/1	37x37x256
Conv	3x3x256/1	37x37x256
Conv	3x3x256/1	37x37x256
Conv	3x3x256/1	37x37x256
Pool	2x2x256/2	18x18x256
Conv	3x3x512/1	18x18x512
Conv	3x3x512/1	18x18x512
Conv	3x3x512/1	18x18x512
Conv	3x3x512/1	18x18x512
Pool	2x2x512/2	9x9x512
Conv	3x3x512/1	9x9x512
Conv	3x3x512/1	9x9x512
Conv	3x3x512/1	9x9x512
Conv	3x3x512/1	9x9x512
Pool	2x2x512/2	4x4x512

Tableau 4.8 : Grandes lignes de l'architecture de réseau proposée

Dans le but de choisir un algorithme d'apprentissage, nous expérimentons les paramètres d'optimisation « SGD », « Rsmprop », « Adadelata », « Adam » et « momentum » avec le même

modèle décrit par la figure 4.5. Les poids issus de l'entraînement sont directement récupérés sur ImageNet (weights='imagenet'). Le nombre d'époque sera le même pour les 5 algorithmes.

Les hypers paramètres de RMSprop, Adadelta et Adam sont initialisés à leurs valeurs par défaut. Le taux d'apprentissage est fixé à 0.009. Le momentum est fixé à 0.9. La taille du batch est de 16 pour l'apprentissage et 128 pour le test. La fonction coût est l'entropie croisée.

Les résultats obtenus pour ces paramètres, sont indiqués dans le tableau 4.9.

	erreur
Adam	0.9856
SGD	0.9693
Rmsprop	1.0053
Adadelta	0.9778
momentum	0.9501

Tableau 4.9 : Erreur d'exécution de chaque algorithme après 45 époques

Les 3 algorithmes « rmsprop », « Adam » et « Adadelta » procèdent à un gradient adaptatif. Une accumulation restreinte des gradients utilisés, progresse rapidement et minimise l'erreur de l'apprentissage dès le début du processus. Mais malheureusement ils n'arrivent pas à minimiser l'erreur comme le « momentum ». De la correction du « momentum » résulte une convergence plus rapide que « SGD ».

Les expérimentations sur les autres architectures, concernent la sélection de bons hyperparamètres. Cependant, ils génèrent des résultats presque pareils.

Toutes les prochaines expérimentations (figure 4.6) se feront avec « momentum » comme algorithme d'apprentissage.

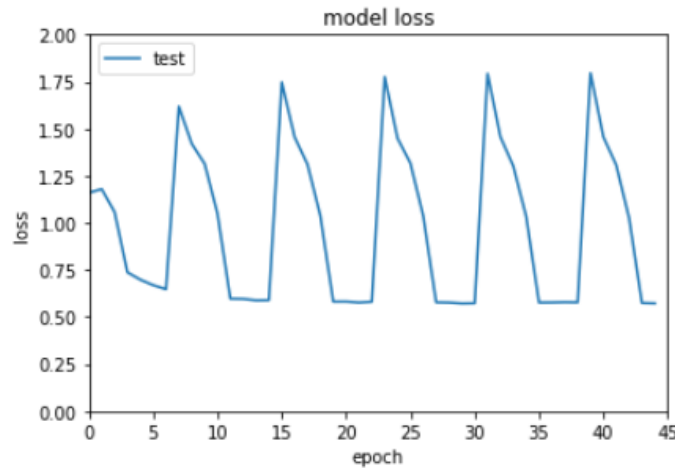


Figure 4.6 : Evolution de l’erreur de test avec l’algorithme momentum sur 45 époques

Au tout début, l’erreur est bien minimisée et atteint une faible valeur après 6 époques. Elle diverge par la suite, complètement et l’erreur commence à augmenter. La courbe se stabilise, et au même moment l’erreur de l’apprentissage, continue de diminuer.

	erreur	Précision%
Modèle 1	1.8000247478485107	55.52050471305847

Tableau 4.10 : Erreur, précision et justesse de l’algorithme lors de la phase de test après 45 époques

Afin d’améliorer la performance de notre modèle, quelques modifications sont réalisées au niveau de son architecture.

➤ **Ajout de 6 couches entièrement connectées au modèle**

Pour augmenter la précision du réseau, augmenter le nombre de couches entièrement connectées, semble être la bonne décision. Le premier niveau détecte les points robustes, le second détecte les lignes, le troisième, la courbe et ainsi de suite. Le fait d’ajouter plus de couches aide à mieux extraire, la relation globale entre les caractéristiques (features).

Dans notre cas, nous avons ajouté plus de 6 couches entièrement connectées (FC) avec une non-linéarité, pour nous assurer que ces nœuds interagissent bien et tiennent compte de toutes les dépendances possibles au niveau des caractéristiques. Pour un apprentissage plus rapide, la fonction d’activation sera ReLU (figure 4.7).

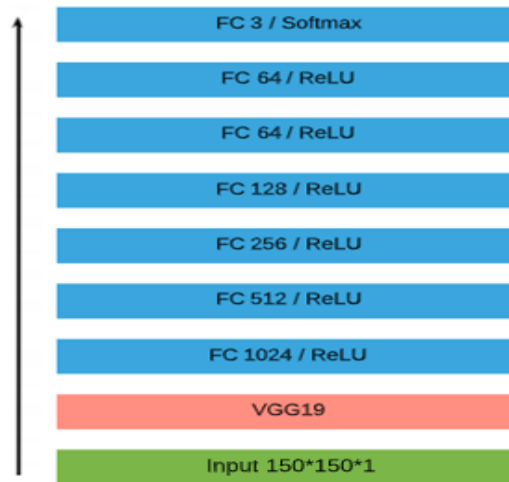


Figure 4.7 : Modèle 2

A travers les modifications apportées, la figure 4.8, montre leur impact sur la performance du modèle.

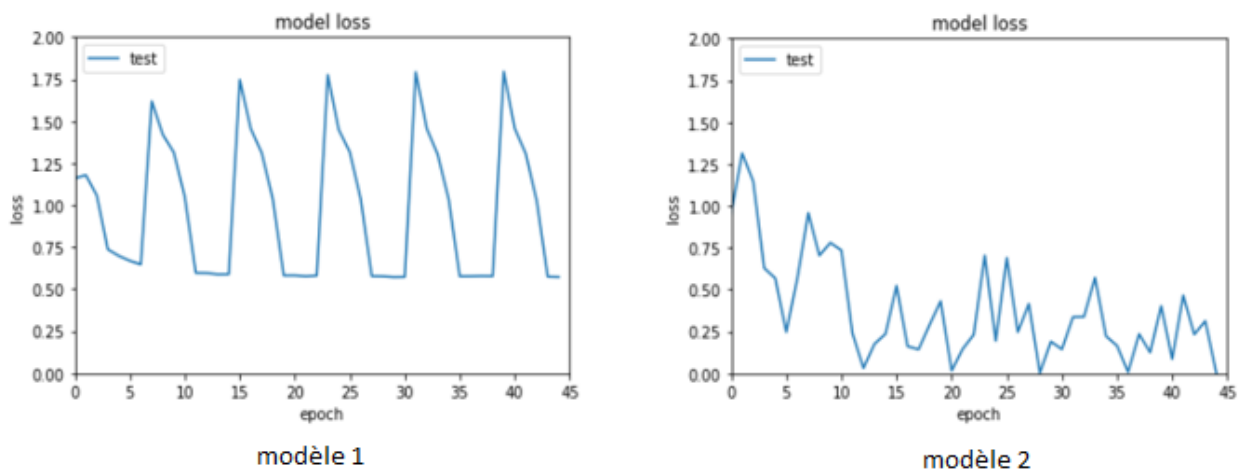


Figure 4.8 : Évolution de l'erreur de test avec une seule couche FC (modèle 1) et 7 couches FC (modèle 2)

Il est à remarquer que le modèle 2, indique une erreur plus faible que celle du premier (tableau 4.11).

	Erreur	Précision %
Modèle 1	1.8000247478485107	55.52050471305847
Modèle 2	0.3641234159469604	92.42902398109436

Tableau 4.11 : Tableau récapitulatif des résultats précédents après 45 époques

A partir de la figures 4.8, ainsi que du tableau 4.11, nous remarquons que le modèle avec 7 couches entièrement connectées, réalise une performance largement meilleure que le modèle avec une seule couche. L'effet se voit tout aussi bien sur la précision, cette dernière réussit à atteindre 92.42%.

En fait, malgré l'utilisation de la validation croisée « leave one out cross-validation » -qui est très utile pour de petites bases d'exemples, la courbe d'erreur de test (figure 4.9) montre des mouvements bruyants et, la courbe d'erreur d'entraînement ressemble à un bon ajustement. L'ensemble de données de test est non représentatif. Ce dernier signifie que l'ensemble de données de test, ne fournit pas suffisamment d'informations. La seule solution est d'ajouter plus d'observations, à notre ensemble de données de test. Mais ce n'est pas le cas, car nous sommes limitées par le nombre d'observations. Nous avons essayé aussi de repartitionner l'ensemble de données (60% pour l'entraînement, 20% pour la validation et 20% pour le test), on retombe sur un autre problème « L'ensemble de données d'entraînement est non représentatif ».

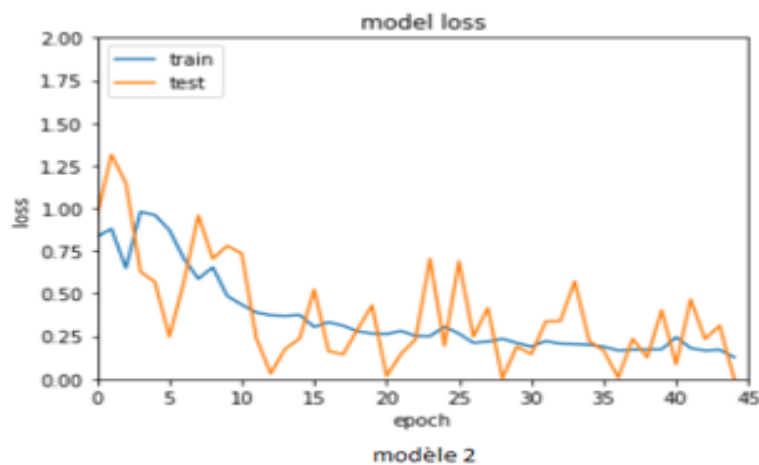


Figure 4.9 : Évolution de l'erreur du test et d'apprentissage

Nous essayons alors d'améliorer la précision et diminuer l'erreur, en introduisant quelques modifications au niveau de l'architecture.

➤ **Introduire le Dropout**

C'est une technique où des neurones sélectionnés au hasard, sont temporairement ignorés pendant l'apprentissage. Pendant la phase de propagation, leur contribution à l'activation des neurones qui leur succède est temporairement supprimé et, toutes les mises

à jour de poids ne sont pas appliquées au neurone, pendant la phase de retro-propagation. Lors de l'apprentissage lorsque des neurones sont supprimés au hasard du réseau, les autres neurones devront intervenir et gérer la représentation requise pour faire des prédictions pour les neurones manquants. Pendant la phase d'apprentissage, pour chaque itération, un neurone est gardé ou supprimé avec une probabilité p .

Nous introduisons alors, le dropout dans notre réseau (figure 4.10) et voir comment il impacte les performances. Il s'agit de la même architecture que le modèle 2, mais régularisé avec la technique du Dropout.

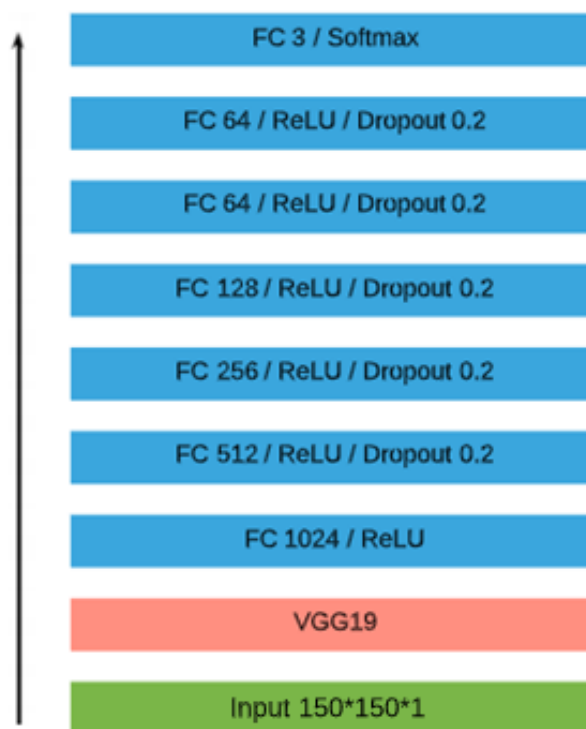


Figure 4.10 : Modèle 3

Nous allons comparer les performances du modèle 3 (figure 4.10) et celui du modèle 2 (figure 4.7) et voir quels sont les changements, qui vont se produire. Les nouveaux résultats sont illustrés dans les figures 4.11, 4.12 ainsi que le tableau 4.12.

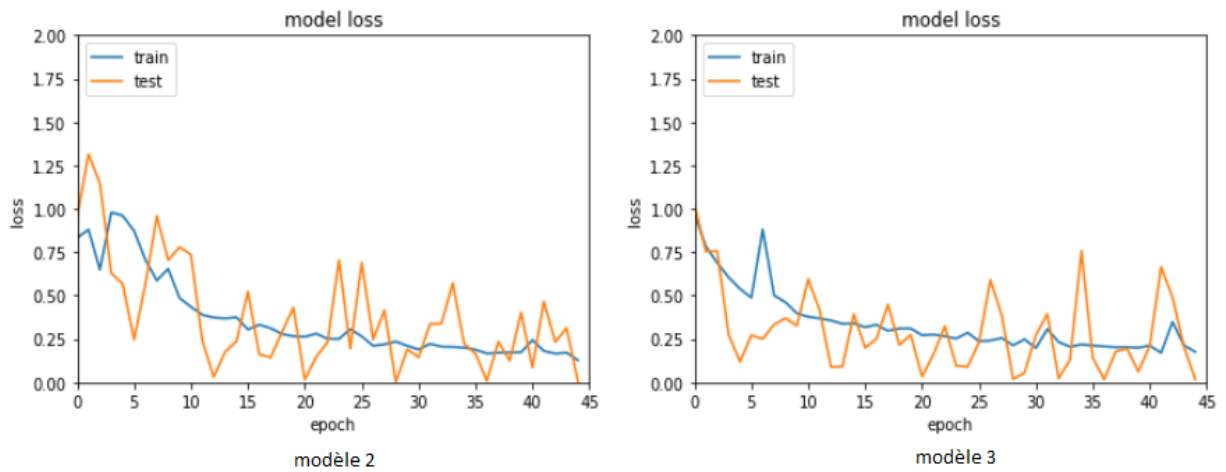


Figure 4.11 : Évolution de l'erreur sans dropout (modèle 2) et avec dropout (modèle 3)

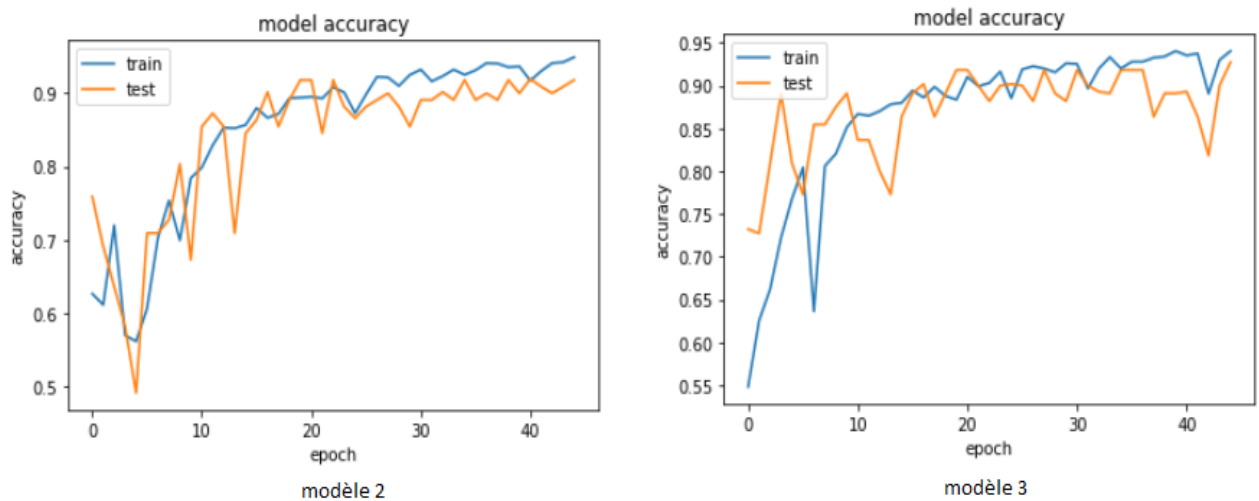


Figure 4.12 : Évolution de la précision sans dropout (modèle 2) et avec dropout (modèle 3)

	Erreur	Précision%
Modèle 2	0.3641234159469604	92.42902398109436
Modèle 3	0.7926584482192993	90.22082090377808

Tableau 4.12: Tableau récapitulatif des résultats précédent après 45 époque

A partir des figures 4.11, 4.12 et le tableau 4.12, nous remarquons tout de suite que le modèle avec dropout ne réussit pas de réaliser une meilleure performance, que le modèle sans dropout. C'est vrai qu'au début de l'apprentissage, avec le modèle 3 (figure 4.10), l'erreur est tout aussi rapidement minimisée même avec des neurones en moins, lors de l'apprentissage (l'effet du dropout), mais il n'y a pas de différence : l'erreur est minimisée et

atteint une faible valeur après 4 époques, puis la courbe diverge complètement et montre un comportement bruyant.

➤ Exponential Linear Units

ReLU est considérée comme la fonction d'activation la plus populaire pour les RNs. Elle permet de régler le problème du vanishing gradient, tout comme la fonction 'ELU' (Exponential Linear Units) grâce à l'identité, pour les entées positives. Néanmoins, ELU améliore l'apprentissage car elle a des valeurs négatives, qui lui permettent de pousser les activations moyennes de l'unité plus près de zéro, ce qui accélère l'apprentissage et permet d'avoir une meilleure précision lors de la classification. La fonction ELU est :

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha (\exp(x) - 1), & x \leq 0 \end{cases}$$

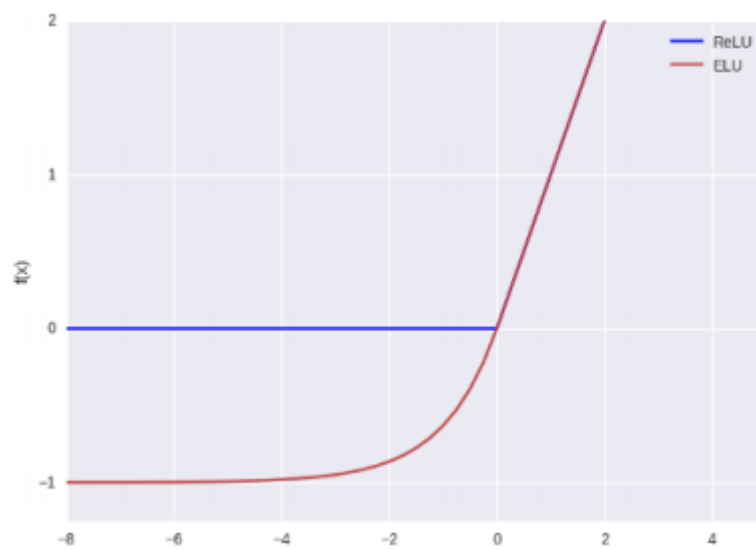


Figure 4.13 : ReLU et ELU (ELU, $\alpha = 1.0$)

Nous allons introduire ELU dans notre réseau selon le modèle de la figure 4.14.

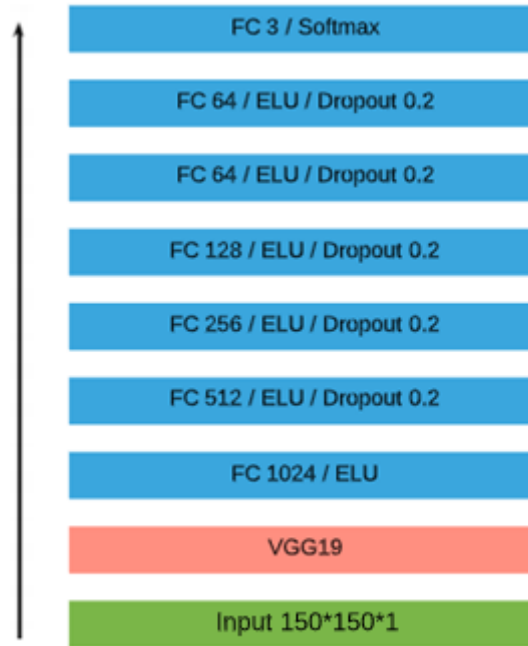


Figure 4.14 : Le modèle 4

Nous comparons dans ce cadre, les performances du modèle 3 et celles du modèle 4 (figure 4.14) pour voir, quels sont les changements qui vont se produire.

ELU diminue le décalage des biais avec ses sorties négatives, car les activations moyennes sont plus proches de zéro. Un décalage des biais moindre accélère l'apprentissage et, améliore la performance du modèle. C'est absolument ce qui se passe avec notre modèle. Quoique l'erreur de test soit diminuée, mais la courbe reste bruyante (figure 4.15, tableaux 4.13 et 4.14).

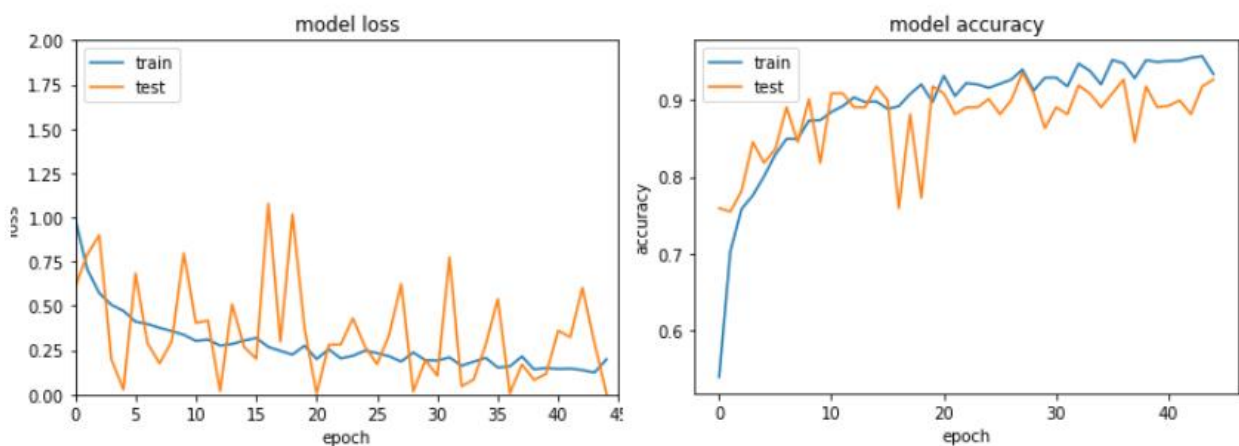


Figure 4.15 : Évolution de l'erreur et l'évolution du taux de bonne classification du modèle 4

erreur	
ReLU	0.7926584482192993
ELU	0.1932217925786972

Tableau 4.13 : Tableau récapitulatif de l'erreur d'apprentissage après 45 époques

Précision %	
Modèle 3	90.22082090377808
Modèle 4	92.1135663986206

Tableau 4.14 : Tableau récapitulatif des résultats précédents après 45 époques

Nous décidons de changer carrément l'architecture. Dans ce qui suit, les images sont toutes redimensionnées suivant une taille 224*224 pixels. Nous avons aussi changé le nombre de neurones, dans chacune des couches entièrement connectées. Le modèle 5 est représenté par la figure 4.16.

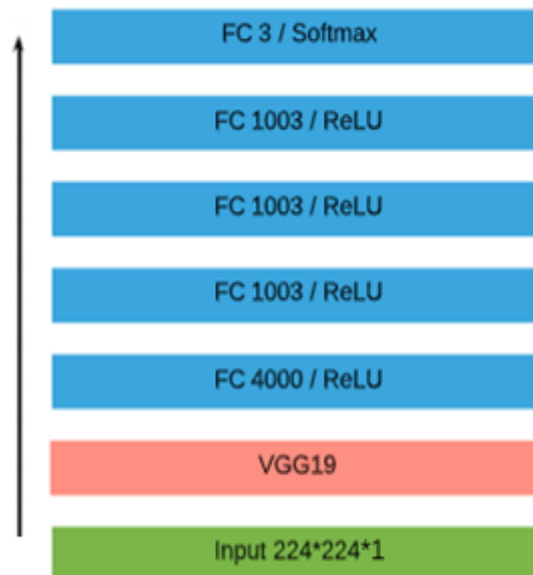


Figure 4.16 : Modèle 5

Le tableau suivant montre les résultats obtenus (nous parlons de la dimension d'output) après chaque couche de convolution et de max-pooling. Input : 224x224x1

Type	Patch size x nombre de convolution / Stride	Output
Conv	3x3x64/1	224x224xx64
Conv	3x3x64/1	224x224xx64
Pool	2x2x64/2	112x112x64
Conv	3x3x128/1	112x112x128
Conv	3x3x128/1	112x112x128
Pool	2x2x128/2	56x56x128
Conv	3x3x256/1	56x56x256
Conv	3x3x256/1	56x56x256
Conv	3x3x256/1	56x56x256
Conv	3x3x256/1	56x56x256
Pool	2x2x256/2	28x28x256
Conv	3x3x512/1	28x28x512
Conv	3x3x512/1	28x28x512
Conv	3x3x512/1	28x28x512
Conv	3x3x512/1	28x28x512
Pool	2x2x512/2	14x14x512
Conv	3x3x512/1	14x14x512
Conv	3x3x512/1	14x14x512
Conv	3x3x512/1	14x14x512
Conv	3x3x512/1	14x14x512
Pool	2x2x512/2	7x7x512

Tableau 4.15 : Grandes lignes de l'architecture de réseau proposée

La figure 4.17 et le tableau 4.16 montrent les résultats des modifications, apportées par le modèle 5.

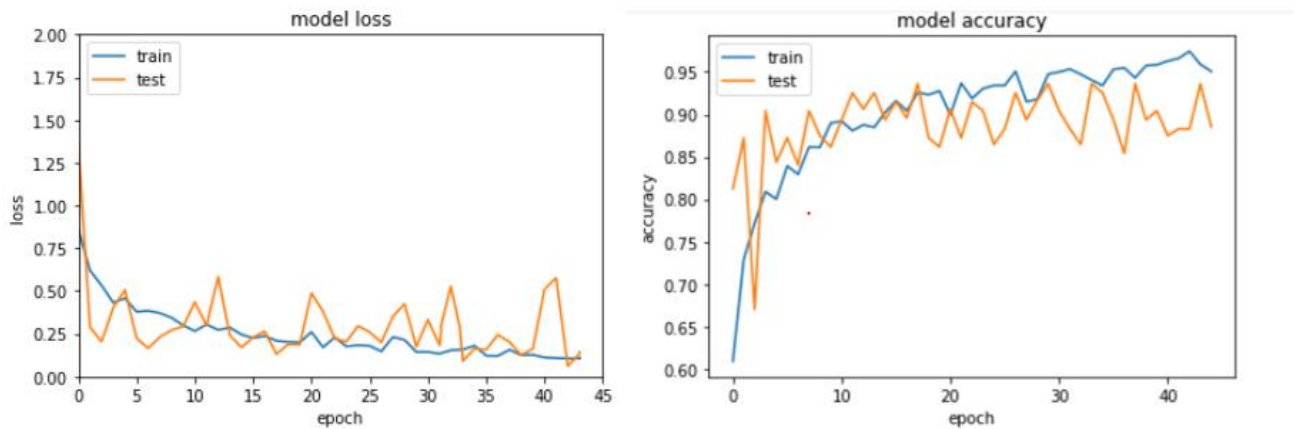


Figure 4.17 : Évolution de l'erreur et l'évolution du taux de bonne classification du modèle 5

	erreur	Précision%
Modèle 2	0.3641234159469604	92.42902398109436
Modèle 4	0.1932217925786972	92.1135663986206
Modèle 5	0.21301056444644928	0.9337539672851562

Tableau 4.16 : tableau récapitulatif des résultats précédent après 45 époque

C'est vrai qu'étendre la résolution d'image de 150 x150 à 224x224 et, augmenter le nombre des neurones, améliore un peu plus la précision du modèle, mais diminue aussi la vitesse d'apprentissage.

Une image de 224x224 prendra environ 6 fois plus de mémoire et de temps de traitement, qu'une image de 150 x150. Nous observons que la courbe d'erreur du test, devient un peu moins bruyante par rapport aux modèles précédents.

b) Modèle VGG19 avec le fine-tuning partiel

Dans le cadre de ces traitements, nous améliorons les résultats des modèles précédents, avec le fine tuning partiel. En pratique, nous créons un modèle VGG19 sans les 'top layers' avec les poids (weights) pré-entraînés sur image Net. Ensuite, nous indiquons au modèle que nous ne voulons pas modifier les poids existants en les gelant (freezing). Dans ce cas, l'apprentissage

ne les modifiera pas pour les 4 premières couches du modèle. À la fin, nous obtenons le classifieur souhaité.

Il faut bien mentionner que lors du fine-tuning, si une couche de batch normalisation (BN) est fixée (gelée), elle utilise les statistiques du mini-batch. Cela peut conduire à une précision réduite, en particulier lorsque nous utilisons l'apprentissage par transfert. Une meilleure approche dans ce cas, serait d'utiliser les valeurs de la moyenne et de la variance. Nous décidons donc, de mettre à jour la couche de batch normalisation afin d'utiliser les statistiques apprises, si elles sont fixées pendant l'entraînement.

```
199  ▼ for layer in vgg_model.layers[:4]:
200      layer.trainable = False
201
202  ▼ for layer in vgg_model.layers:
203      ▼ if "BatchNormalization" in layer.__class__.__name__:
204          layer.trainable = True
205
```

Figure 4.18 : Aperçu sur le code source pour batch normalisation

Pour clarifier les choses, supposons que nous utilisons l'un des CNN pré-entraînés de Keras et que nous souhaitons l'ajuster (fine-tune). Malheureusement, nous n'avons aucune garantie que la moyenne et la variance de notre nouvel ensemble de données, à l'intérieur des couches BN seront similaires à celles de l'ensemble de données d'origine. En conséquence, si nous affinons les couches supérieures, leurs poids seront ajustés à la moyenne / variance du nouvel ensemble de données. Néanmoins, pendant l'inférence, les couches supérieures recevront des données qui sont mises à l'échelle en utilisant la moyenne / variance de l'ensemble de données d'origine. Cet écart peut conduire à une précision réduite.

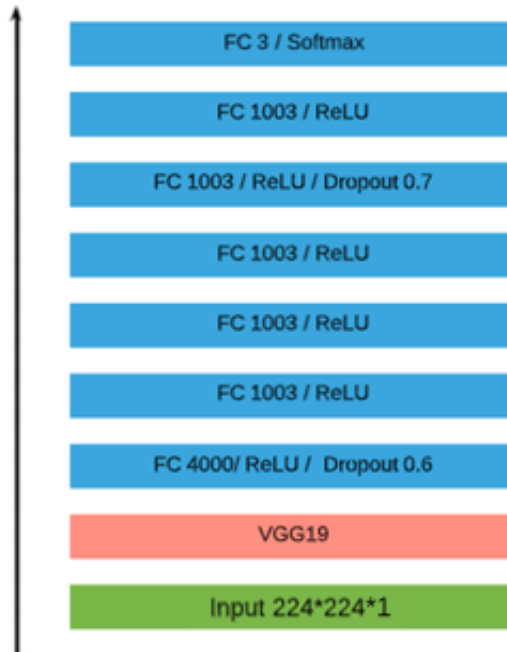


Figure 4.19 : Modèle 6

Nous observons les courbes d'erreur et de précision du test et d'apprentissage du modèle 6, dans la figure 4.20 et le tableau 4.17.

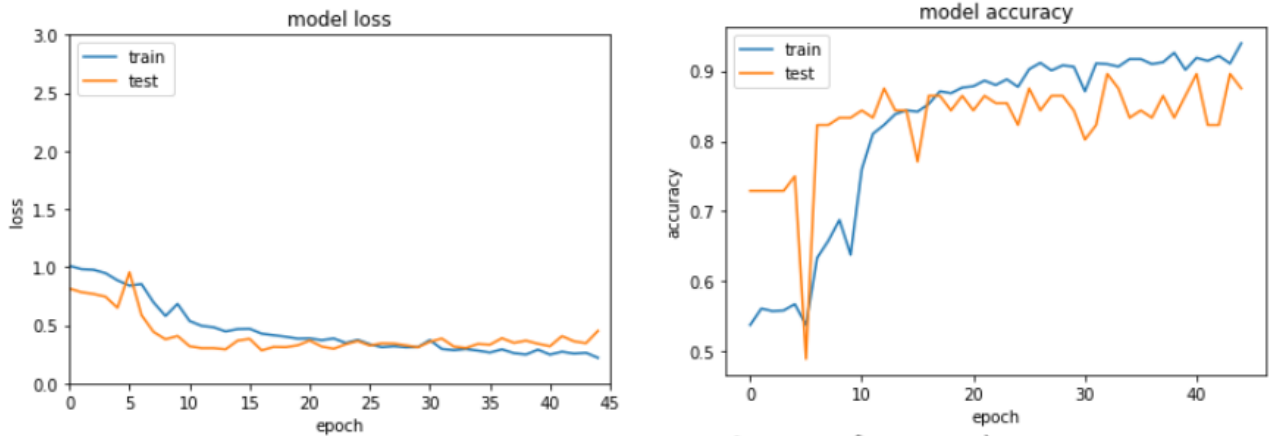


Figure 4.20 : Évolution de l'erreur et du taux de bonne classification du modèle 6

	erreur	Précision%
Modèle 5	0.21301056444644928	93.375396728515621
Modèle 6	0.307094007730484	90.530627848625183

Tableau 4.17 : Tableau récapitulatif des résultats précédents après 45 époques

À partir de la figure 4.20, nous constatons tout de suite que la courbe d'erreur de ce réseau, constitue un bon ajustement des données. Néanmoins le modèle VGG19 fine-tuning partiel, ne réussit pas à atteindre une précision meilleure que le VGG19 fine-tuning total même après 65 époques.

Pour la matrice de confusion du modèle 5 (figure 4.21), il y a aussi une confusion entre les classes mais moins que celle du modèle 6. Par contre, ce modèle trouve aussi une petite difficulté à reconnaître les deux classe bénigne et normale par rapport au modèle 5. Cela peut être dû au faible nombre d'échantillons, dans la base. Les matrices de confusion, sont confirmées par les valeurs, de la sensibilité et de la spécificité des modèles 5 et 6 (tableau 4.20), obtenues pour chaque classe.

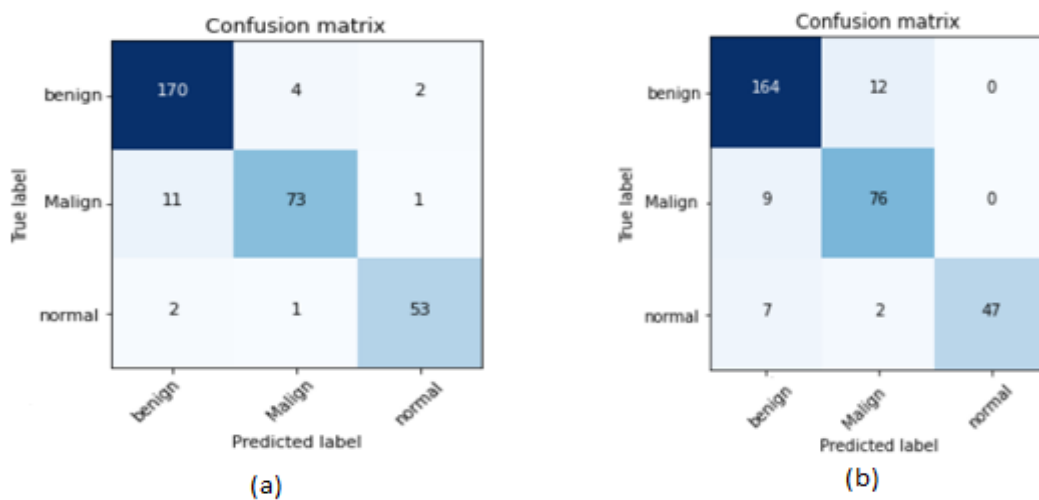


Figure 4.21 : Matrices de confusion : (a) modèles 5 et (b) modèle 6

Pour calculer la sensibilité et la spécificité, Les formules restent les mêmes, c'est-à-dire $VP / (VP + FN)$ pour calculer la sensibilité et $VN / (VN + FP)$ pour calculer la spécificité. Il vous suffit de l'appliquer individuellement sur les trois classes.

	Classe 1	Classe 2	Classe 3	FN
Classe 1	N11	N12	N13	(N12+N13)
Classe 2	N21	N22	N23	(N21+N23)
Classe 3	N31	N32	N33	(N31+N32)
FP	(N21+N31)	(N12+N32)	(N13+N23)	

Tableau 4.18 : Matrice de confusion avec 3 classes

La sensibilité et la spécificité seraient déterminées en calculant ce qui suit :

	Spécificité	sensibilité
Classe 1	$(N22+N33)/(N21+N31+N22+N33)$	$N11/(N11+N12+N13)$
Classe 2	$(N11+N33)/(N12+N32+N11+N33)$	$N22/(N21+N22+N23)$
Classe 3	$(N11+N22)/(N13+N23+N11+N22)$	$N33/(N31+N32+N33)$

Tableau 4.19 : Spécificité et sensibilité pour 3 classes

Donc les valeurs de la spécificité et la sensibilité des modèles 5 et 6 sont :

	VGG19 complètement entraîné		VGG19 fine-tuning partiel	
	Spécificité	Sensibilité	Spécificité	Sensibilité
bénigne	0.90	0.97	0.88	0.93
maligne	0.98	0.86	0.94	0.89
normale	0.99	0.95	1	0.84

Tableau 4.20 : Spécificité et sensibilité des modèles 5 et 6

c) Modèle VGG16 avec le fine-tuning total

Dans cette partie, nous allons entraîner un réseau qui se base sur un modèle profond pré-entraîné VGG16. Il est constitué de plusieurs couches, treize (13) (après l'élimination des trois dernières couches entièrement connectées), dont les couches de convolution utilisant des filtres de tailles 3x3, en appliquant une normalisation par lot (BatchNormalization).

Nous ajoutons par la suite, cinq couches entièrement connectées et des corrections ReLU sont utilisées, pour éliminer les valeurs négatives. A la fin, la fonction Softmax est appliquée pour calculer le taux des 3 Classes (bénigne, maligne et normale). Elle renvoie donc un vecteur de taille 1x1x3, qui contient les probabilités d'appartenance à chacune des classes. Ce réseau possède 121,100,727 paramètres entraînaibles. Dans cette partie, nous allons entraîner tout le réseau. L'architecture est résumée dans la figure 4.22.

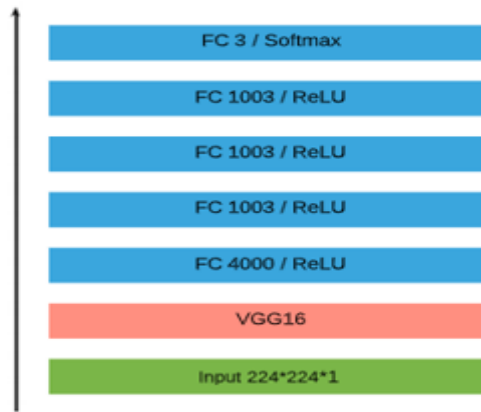


Figure 4.22 : Modèle 7

Les résultats obtenus (les dimensions des outputs) après chaque couche de convolution et de max-pooling, sont résumés dans le tableau 4.21.

Input : 224x224x1

Type	Patch size x nombre de convolution / Stride	Output
Conv	3x3x64/1	224x224x64
Conv	3x3x64/1	224x224x64
Pool	2x2x64/2	112x112x64
Conv	3x3x128/1	112x112x128
Conv	3x3x128/1	112x112x128
Pool	2x2x128/2	56x56x128
Conv	3x3x256/1	56x56x256
Conv	3x3x256/1	56x56x256
Conv	3x3x256/1	56x56x256
Pool	2x2x256/2	28x28x256
Conv	3x3x512/1	28x28x512
Conv	3x3x512/1	28x28x512
Conv	3x3x512/1	28x28x512
Pool	2x2x512/2	14x14x512
Conv	3x3x512/1	14x14x512
Conv	3x3x512/1	14x14x512
Conv	3x3x512/1	14x14x512

Pool	2x2x512/2	7x7x512
-------------	-----------	---------

Tableau 4.21 : Grandes lignes de l'architecture du réseau proposée

Le modèle 7 génère l'erreur et la justesse (figure 4.23), suivant des époques, allant jusqu'à 45. Ce modèle montre une erreur de 0.21 et une précision de 94% (tableau 4.22).

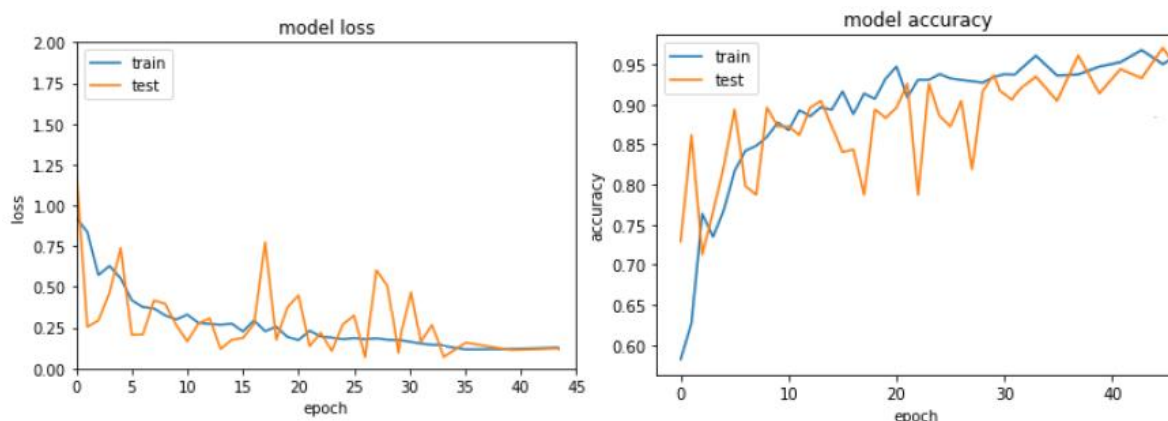


Figure 4.23 : Évolution de l'erreur et du taux de bonne classification du modèle 7

	erreur	Précision%
Modèle 7	0.216222709273608153	94.01026933154803211

Tableau 4.22 : Tableau récapitulatif des résultats précédents après 45 époques

D'après la figure 4.23 nous remarquons clairement, qu'après un comportement bruyant au début de l'apprentissage, la courbe d'erreur a réussi à se stabiliser après 35 époques. L'erreur est bien minimisée et a atteint une faible valeur. Le pourcentage de bonne classification atteint 94.01% après 45 époques. Cette précision continue d'augmenter doucement, jusqu'à atteindre 94.23% après 55 époques.

d) Modèle VGG16 avec le fine-tuning partiel

Afin de créer ce modèle, nous avons réalisé les mêmes opérations utilisées précédemment que dans le VGG19 fines-tuning partiel. L'architecture est résumée dans la figure 4.24.

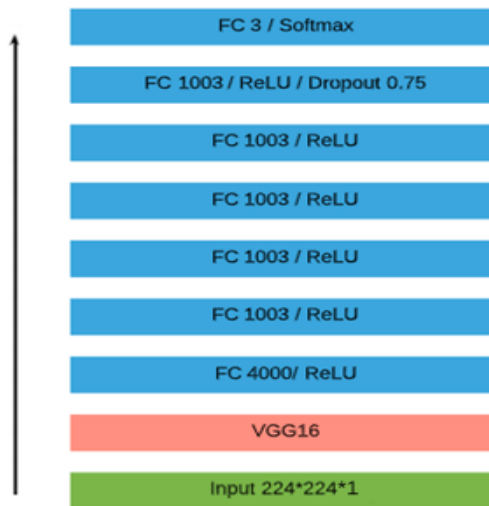


Figure 4.24 : Modèle 8

Les courbes de l'erreur et de la justesse engendrée par le modèle 8, sont illustrées par la figure 4.25.

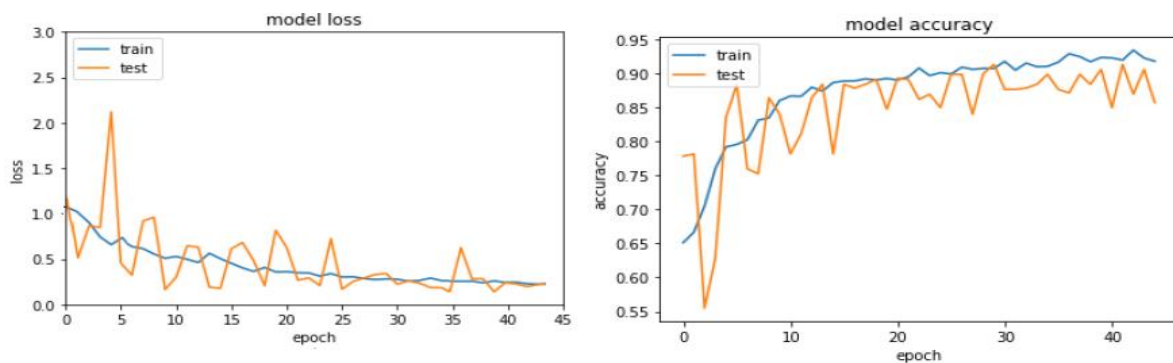


Figure 4.25 : Évolution de l'erreur et du taux de bonne classification du modèle 8

D'après les résultats obtenus suivant les graphes précédents et bien que le nombre des paramètres traités dans le modèle 7(121,100,727 paramètres) est moins que celui du modèle 8 (123,114,751 paramètres) ; cela permet de bénéficier d'un gain en temps considérable, si nous appliquons l'apprentissage sur une base de données avec plus d'échantillons. La performance du modèle 8 obtenue sur la base BUSI, est relativement faible (91.80%) par rapport à la performance du modèle 7, qui est de 94.01% sur la même base (tableau 4.23).

	erreur	Précision%
Modèle 7	0.216222709273608153	94.01026933154803211
Modèle 8	0.271802013488579137	91.800251489710362

Tableau 4.23 : Tableau récapitulatif des résultats précédent après 45 époques

Nous constatons à travers la matrice de confusion du modèle 8 (figure 4.26), qu'il y a 12 images dans la classe bénigne qui sont reconnues comme des images, de la classe maligne. Les matrices de confusion, sont confirmées par les valeurs, de la sensibilité et de la spécificité des modèles 7 et 8 (tableau 4.24), obtenues pour chaque classe. Mais le modèle reste performant par rapport aux modèles précédents.

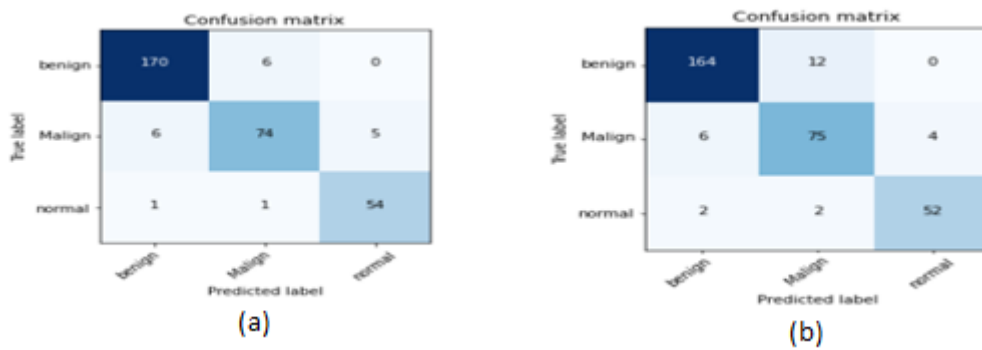


Figure 4.26 : Matrices de confusion : (a) modèles 7 et (b) modèle 8

	VGG16 complètement entraîné		VGG16 fines-tuning partiel	
	Spécificité	Sensibilité	Spécificité	Sensibilité
bénigne	0.95	0.97	0.94	0.93
maligne	0.97	0.87	0.94	0.88
normale	0.98	0.96	0.98	0.93

Tableau 4.24 : Spécificité et sensibilité des modèles 7 et 8

e) Modèle Inception V3

Dans cette partie, nous allons entraîner deux réseaux, qui se basent sur un modèle profond pré-entraîné Inception V3.

Pour le premier réseau, nous avons modifié l'architecture inception V3 en remplaçant son MLP (le perceptron multicouches) par 7 couches entièrement connectées. Ce réseau possède 74,689,971 paramètres. Pour chaque itération, nous avons fixé la taille des lots à 32. Nous avons alors entraîné notre modèle, complètement par « momentum » comme algorithme d'apprentissage. Le taux d'apprentissage est fixé à 0.009. Le momentum est fixé à 0.9. L'architecture est résumée dans la figure 4.27.

En ce qui concerne le deuxième réseau, nous construisons un modèle inception v3, sans les tops layers avec les poids/weight pré-entraînés sur image Net. Nous indiquons par la suite, au modèle que nous ne voulons pas modifier les poids existants, en les gelant juste pour les premières couches du modèle. Le classifieur est alors créé, en ajoutant 7 couches entièrement connectées comme le montre la figure 4.28.

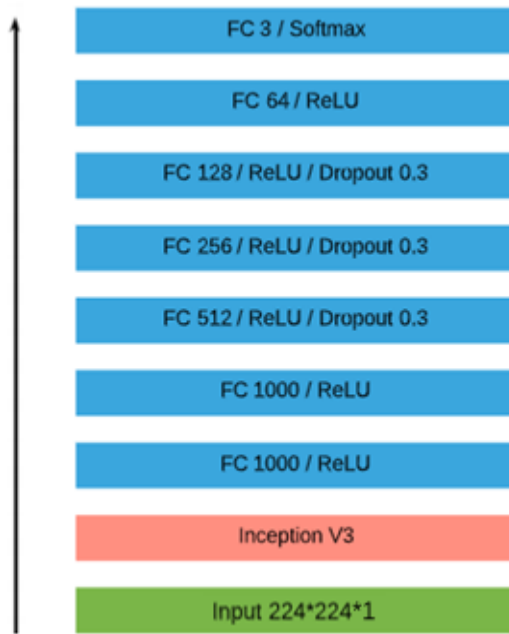


Figure 4.27 : Modèle 9

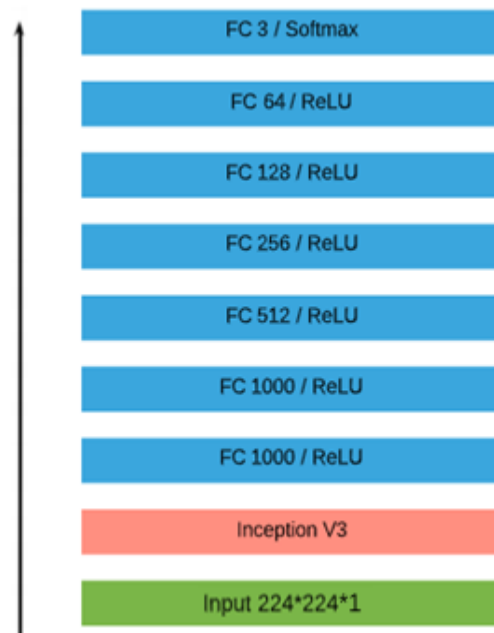


Figure 4.28 : Modèle 10

Les dimensions des outputs après chaque couche de convolution et de max-pooling pour Inception V3, sont résumées dans le tableau suivant, où les entrées correspondent à 'Input : 224x224x1'.

Type	Patch size x nombre de convolution / Stride	Output
Conv	3x3x32/1	111x111x32
Conv	3x3x32/1	109x109x32
Conv padded	2x2x64/2	109x109x64
pool	3x3x64/1	54x54x64
Conv	3x3x80/1	52x52x80
Conv	2x2x192/2	25x25x192
Conv	3x3x288/1	25x25x288
Inception Ax3	-----	12x12x768

Inception Bx5	-----	5x5x1280
Inception Cx2	-----	5x5x2048

Tableau 4.25 : Grandes lignes de l'architecture du réseau proposée

Le réseau créé, indique les variations de l'erreur et de la justesse, dans les graphes des figures 4.29 et 4.30.

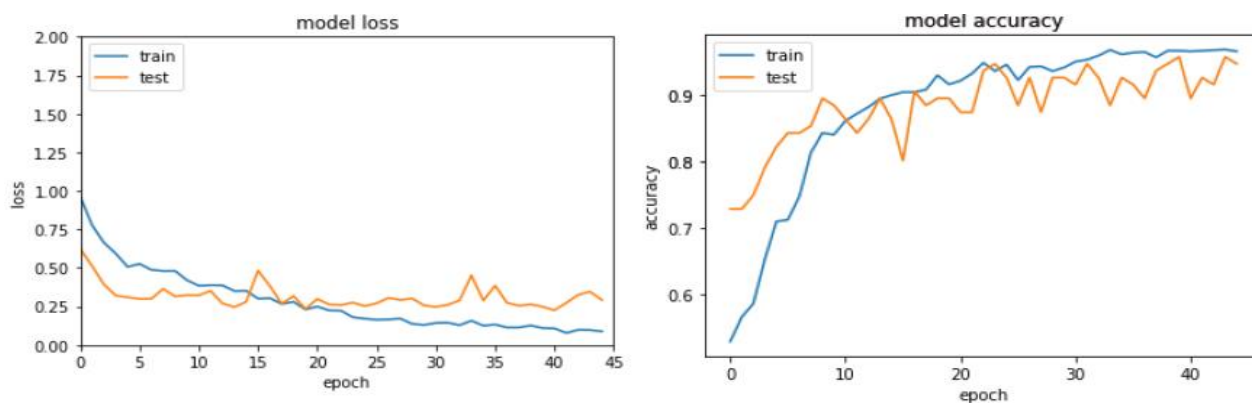


Figure 4.29 : Évolution de l'erreur et du taux de bonne classification du modèle 9

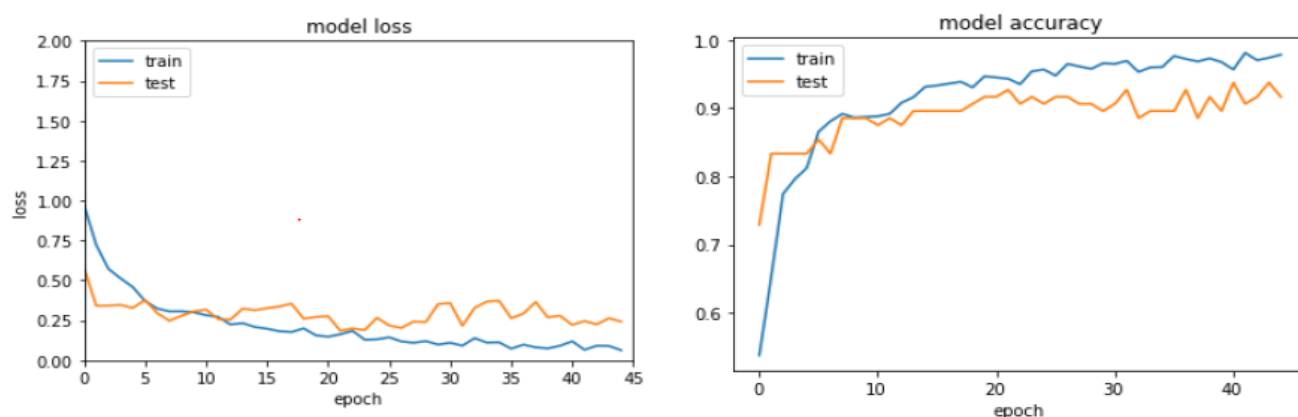


Figure 4.30 : Évolution de l'erreur et du taux de bonne classification du modèle 10

Nous distinguons à travers ces graphes, que pour 45 époques, les 2 modèles présentent une précision de 94%. L'erreur s'affaiblit pour le modèle 10 (tableau 4.26).

	Erreur	Précision
Modèle 9	0.2447115033864975	0.9463722109794617
Modèle 10	0.21076686680316925	0.9400631189346313

Tableau 4.26 : Tableau récapitulatif des résultats précédents après 45 époques

Nous constatons à travers la matrice de confusion du modèle 10 (figure 4.31), qu'il y a 5 classes d'images dans la classe bénigne, qui sont reconnues comme des images de la classe maligne. Les matrices de confusion, sont confirmés par les valeurs, de la sensibilité et de la spécificité des modèles 9 et 10 (tableau 4.27), obtenues pour chaque classe. Le modèle réalisé, montre des résultats encourageants.

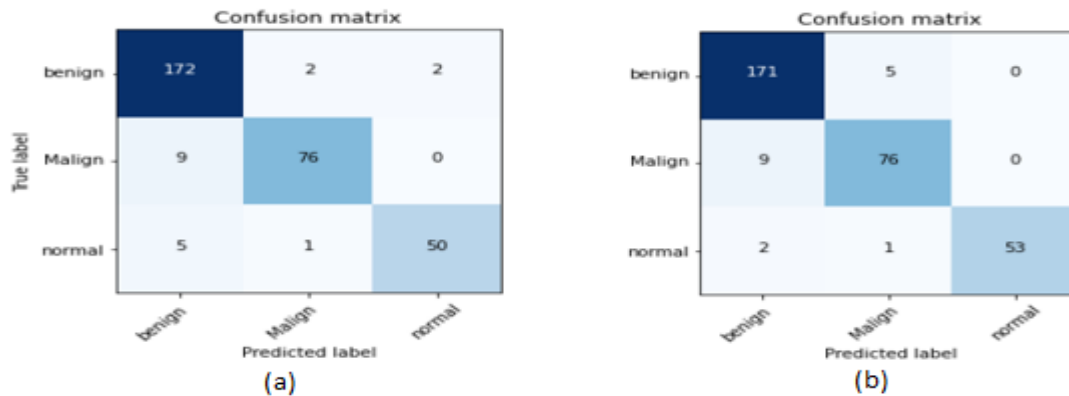


Figure 4.31 : Matrices de confusion : (a) modèle 9 et (b) modèle 10

	Inception V3 complètement entraîné		Inception V3 fines-tuning partiel	
	Spécificité	Sensibilité	Spécificité	Sensibilité
bénigne	0.92	0.97	0.90	0.98
maligne	0.97	0.89	0.99	0.89
normal	1	0.95	0.99	0.89

Tableau 4.27 : Spécificité et sensibilité des modèles 9 et 10

Les performances des modèles VGG16 et VGG19 obtenues sur la base BUSI sont relativement faibles par rapport à celle du modèle Inception V3, qui est de 94% (en moyenne pour le fine tuning total et partiel) sur la même base ; bien que le nombre de paramètres traités dans Inception V3, est beaucoup moins important, que celui du VGG16 et VGG19. Si nous réalisons l'apprentissage sur une base de données avec plus d'échantillons, cela nous permet de bénéficier d'un gain en temps considérable. Ceci s'explique par le fait que le CNN est plus performant dans le cas d'un bon choix de techniques de réduction de paramètres.

Une récapitulation des résultats obtenus, est dressée dans le tableau 4.28.

	Nombre De paramètres	Nombre De paramètres entraînable	Nombre de paramètres non entraînable	Le nombre de couche de convolution	La taille de noyau de convolution utilisé	Le nombre d'unités cachés dans la couche connectée	Taille de batch	Précision %	Spécificité moyenne	Sensibilité moyenne	Erreur de test
Modèle 1	20,048,963	20,048,963	0	16	3x3	3	16	55.52	0.000	0.3333	1.80
Modèle 2	29,115,651	29,115,651	0	16	3x3	1024+512+256+128+64+64+3 = 2051	16	92.43	0.9366	0.9133	0.36
Modèle 3	29,115,651	29,115,651	0	16	3x3	2051	16	90.22	0.9500	0.9000	0.80
Modèle 4	29,115,651	29,115,651	0	16	3x3	2051	16	92.11	0.9500	0.9133	0.19
Modèle 5	126,410,423	126,410,423	0	16	3x3	4000+(1003x3) +3 = 5009	32	93.37	0.9566	0.9366	0.21
Modèle 6	128,424,447	128,385,727	38,720	16	3x3	4000+(1003x5) +3 =7015	32	90.53	0.9400	0.8866	0.30
Modèle 7	121,100,727	121,100,727	0	13	3x3	5009	32	94.01	0.9666	0.9333	0.21
Modèle 8	123,114,751	123,076,031	38,720	13	3x3	7015	32	91.80	0.9533	0.9133	0.27
Modèle 9	74,689,971	74,655,539	34,432	32	3x3 , 1x3 , 5x5 , 7x7 , 1x7	2963	32	94.63	0.9633	0.9366	0.24
Modèle 10	74,689,971	74,654,675	35,296	32	3x3,1x3,5x5 , 7x7 , 1x7	2963	32	94.00	0.9600	0.9200	0.21

Tableau 4.28 : Comparaison des performances des modèles 'BreastULTASCNN'

4.6 Discussion

Diverses tailles de lots (batch) sont testées, afin de classifier les différentes catégories de la base (BUSI). Le même taux d'apprentissage, est utilisé pour chaque apprentissage. Ces expériences montrent que la taille du batch dans ce contexte, a un impact important sur les performances de la classification des pathologies mammaires. En effet, l'utilisation de grands batches altère les performances. Nous avons alors employé des batches de tailles 16 et 32 qui ont montré des bonnes performances. La raison principale est que l'estimation du gradient pour la descente, bénéficie plus de l'utilisation de batches de petites tailles. En revanche, un grand batch permet de traiter plus d'exemples, donc l'apprentissage sera plus rapide.

A partir du tableau 4.28, nous observons que les réseaux qui possèdent le plus nombre de paramètres, sont les modèle 5 et 6 (VGG19). A partir de cela, nous concluons que le nombre de couches de convolutions et le nombre d'unités dans les couches entièrement connectées, ont une influence importante sur la répartition des poids du réseau. Notons également, que l'augmentation du nombre d'unités cachées, améliore considérablement les performances. L'utilisation de grandes couches connectées avec 4000 et 1003 permettent d'obtenir la complexité nécessaire, mais ces couches possèdent plus de poids à estimer. Le problème de la classification de pathologies, nécessite l'utilisation de grandes couches connectées. Il est important de mentionner que les couches connectées représentent près de 80% de l'ensemble des poids, à estimer pour un réseau.

Le bloc primordial des réseaux neurones convolutifs est le bloc de convolutions, qui possède également des paramètres à déterminer tels que, la taille des convolutions utilisées. En effet, la taille des convolutions peut prendre les valeurs 1, 3 et 7. L'influence de la taille des convolutions sur le nombre total de poids à estimer, est négligeable comparée à la dimension des couches connectées. Cependant, nous pouvons également voir que l'utilisation des différentes tailles de convolution, dans les modèles 9 et 10 améliore considérablement les performances finales, malgré le faible nombre de paramètres.

La structure du CNN est supposée être le paramètre déterminant, quoique le choix des paramètres de l'algorithme d'optimisation, change les performances finales. Chaque hyper paramètre, a un effet sur la structure et en particulier, sur le nombre de poids à estimer.

Les choix des structures sont les plus à même d'être contestés, car notre travail a prouvé que des blocs tels que Inception, peuvent montrer de meilleures performances, que l'enchaînement classique de couches de convolutions et de pooling (dans le cas de VGG16 et VGG19).

Enfin, il faut bien mentionner que les choix conceptuels étudiés, ont été réalisés pour simplifier les recherches et les premiers tests. Le monde de l'apprentissage profond est un monde basé sur des démonstrations empiriques et dont la constante du temps est extrêmement rapide. L'intervalle de temps des nouvelles techniques pour arriver dans le monde du « deep learning » est entre deux à trois mois. Il est donc difficile de rester à la pointe de la technique, en permanence. La conséquence est l'ignorance d'une méthodologie globale de résolution des problèmes. En effet, les expériences sont réalisées sur des cas d'utilisation spécifiques. Par conséquent, elles ne peuvent être transférées à d'autres cas.

Notre modèle final est bien évidemment, le modèle 9 Inception V3. En effet, l'apprentissage est prolongé et les paramètres sont sauvegardés à chaque époque, jusqu'à ce que les performances de l'ensemble de test ne s'améliorent plus. Le modèle est par la suite, restauré à l'état où il a offert les meilleures performances. Du fait que l'entraînement dure longtemps, nous baissons les probabilités du dropout, dans le but d'augmenter la capacité de la généralisation du modèle.

Les performances finales, obtenues pour un apprentissage de 300 époques, sont données par le tableau 4.29, suivant les paramètres ci-dessous :

- Méthode d'apprentissage : momentum
- Taux d'apprentissage : 0.09

	Erreur	Précision%
Modèle final	0.20109581685066223	94.95267868041992

Tableau 4.29: Performances finales de notre modèle sur l'ensemble de test

L'évaluation du système 'BreastULTRASCNN', est réalisée par la courbe ROC (figure 4.32). Nous remarquons, bien les performances, obtenues par le modèle sélectionné.

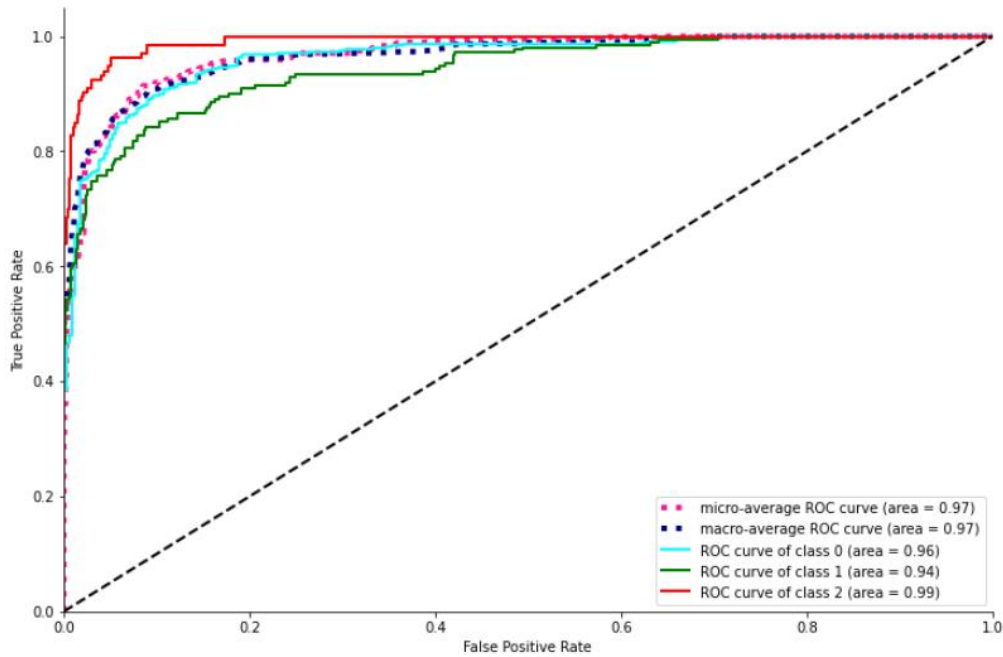


Figure 4.32 : Courbe ROC du modèle final

4.7 Interface de BreastULTRASCNN

Afin de pouvoir visualiser les performances de notre réseau de neurones facilement sans être obligé de passer à chaque fois par les lignes de code, nous l'avons introduit dans une interface graphique (figures 4.33 et 4.34). Le rôle de cette interface, est de générer les informations présentées ci-dessous :

1. Affichage de l'image échographique.
2. Téléchargement de l'image sur laquelle tester le réseau.
3. Prédiction de la classe.
4. Fermeture de la fenêtre d'application.



Figure 4.33 : Interface générale de 'BreastULTRASCNN'



Figure 4.34 : Exemple d'application de 'BreastULTRASCNN'

4.8 Application web

Le modèle 'BreastULTRASCNN', a été aussi réalisé sur une application web. Les maquettes des figures 4.35 et 4.36 représentent les premières pages de notre système, qui contiennent les principales fonctionnalités « select image », « Predict ». Le maquettage reste une étape importante à définir afin, d'avoir un rendu visuel de notre application web.



Figure 4.35 : Page principale partie 1

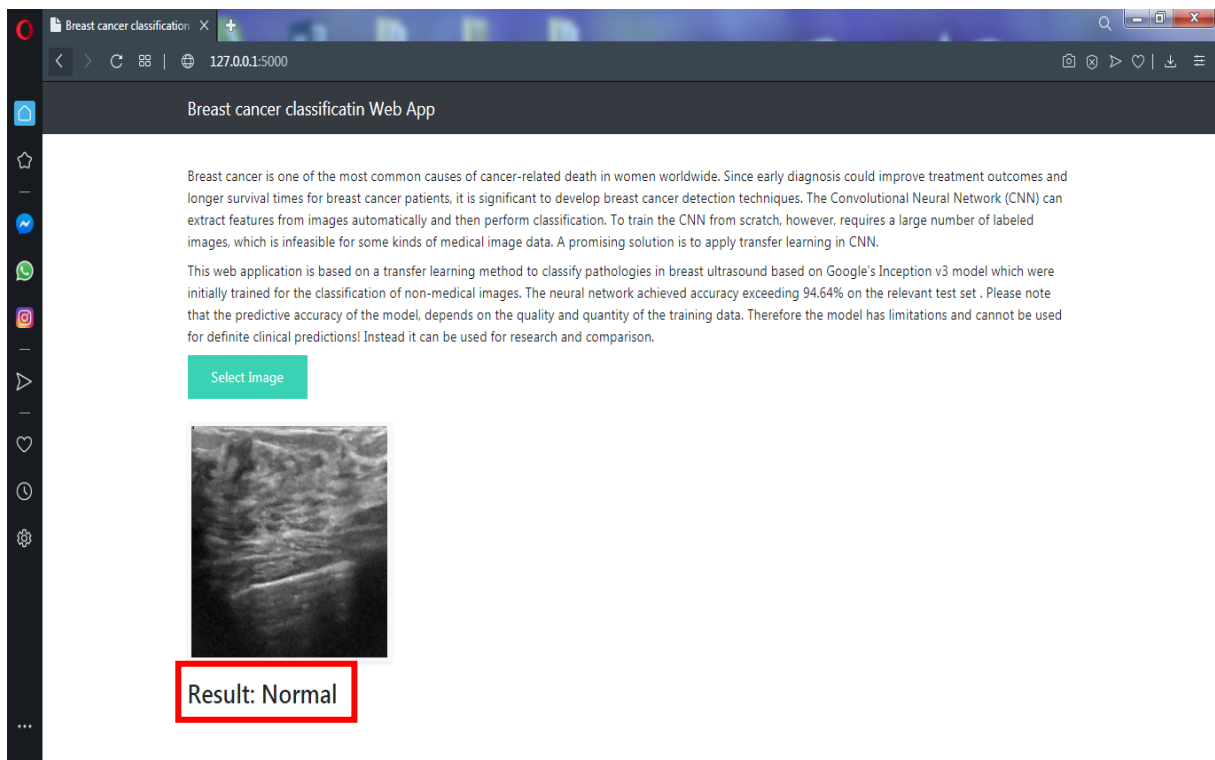


Figure 4.36 : Page principale partie 2

4.9 Conclusion

Dans ce chapitre, nous avons présenté notre système de la classification de pathologies en échographie du sein basée sur des réseaux profonds pré-entraînés de types CNN, en utilisant trois architectures différentes. La première est le VGG19, la deuxième est le VGG16 et la troisième, c'est l'architecture Inception V3.

Des résultats, décrits et discutés dans les sections précédentes, nous avons pu analyser les défauts des différents modèles entraînés, en observant leurs comportements, grâce aux variations des erreurs et des justesses. Le réseau qui a engendré les meilleurs résultats, est le modèle inception V3. La précision obtenue à cet effet, dépasse 94%.

Conclusion générale

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

Les objectifs fixés par la présente étude, se rapportent à l'analyse profonde d'images ultrasonores, pour la détection du cancer du sein suivant les réseaux de neurones convolutifs. En effet, ces réseaux ont révolutionné la vision par ordinateur, ces dernières années. L'histoire de leur conception, a commencé avec des modèles qui étaient de simples piles de convolutions, pour les opérations d'extraction de caractéristiques et de pooling maximal, pour le sous-échantillonnage spatial. Ces idées ont été affinées par la suite dans d'autres architectures, dans laquelle les opérations de convolution étaient répétées plusieurs fois, entre les opérations de max-pooling, permettant ainsi au réseau d'apprendre des caractéristiques plus riches à toutes les échelles spatiales. C'est dans ce cadre, que le système développé a été conçu suivant différentes architectures, parmi lesquelles, le modèle VGG16, le VGG19 et l'architecture Inception V3. Le paradigme étudié a été affiné selon deux stratégies ; le fine-tuning total et le fine-tuning partiel, afin de construire des modèles basés sur l'apprentissage par transfert pour les tester, sur la base de données échographiques réelles BUSI.

L'étude réalisée par l'utilisation des réseaux de neurones profonds pré-entraînés, qui exploitent les blocs inception en profondeur et, l'enchaînement classique de couches de convolutions et de pooling, a montré des résultats encourageants.

Nous avons montré comment les blocs inception ajoutés dans le CNN Inception V3, pourraient améliorer la précision, en diminuant énormément de temps.

La richesse des couches qui constituent les réseaux VGG16 et VGG19 permet de traiter, un maximum de paramètres. Néanmoins cette technique a abouti à une précision plus faible, que celle d'Inception V3.

Nous avons vu que le fine tuning d'Inception-v3, peut améliorer efficacement la précision de la classification des images échographiques du sein. Cependant, si la sélection du réseau d'apprentissage par transfert est inappropriée, le problème du transfert négatif peut survenir, ce qui entraînera une diminution de la précision et une augmentation du temps d'apprentissage.

Dans nos travaux futurs, nous visons l'exploration d'autres modèles profonds, tels que RCNN (Region Based CNN) de [103], et YOLO (You Only Look Once) aux images échographiques 3D, ou à d'autres modalités.

Références bibliographiques

- [1] Ligue du cancer. Cancer du sein, https://www.ligue-cancer.net/article/26094_cancer-du-sein, avril 2020, consulté en mai 2020.
- [2] Global health organization. Global Cancer Observatory, <https://gco.iarc.fr>, consulté en mai 2020.
- [3] N. Benachenhou, Elargissement et amélioration des résultats du traitement conservateur dans le cancer du sein par les techniques d'oncoplastie, thèse de doctorat en médecine, Université d'Alger, 2012.
- [4] R. Givaudan, J. Serres, A. Dupuy, E. Thuane. Le cancer et les cellules cancéreuses, <http://lecancerdusein-tpe.blogspot.com>, consulté en juillet 2020.
- [5] Globocan 2012. Estimatiéd cancer incidence, mortality and prevalence worldwide in 2012, <http://globocan.iarc.fr>, consulté en juillet 2020.
- [6] Institut national du cancer de France. Cancers du sein, <http://www.e-cancer.fr/soins/recommandations/cancers-du-sein>, consulté en juillet, 2020.
- [7] Eureka santé, Vidal. Cancer du sein, <https://eurekasante.vidal.fr/maladies/cancers/cancer-sein.html>, consulté en juillet 2020.
- [8] Fédération nationale des centres de lutte contre le cancer (FNCLCC). Comprendre le cancer du sein, SOR Savoir Patient, 2007.
- [9] Fondation ARC pour la recherche sur le cancer. Cancers du sein : les facteurs de risque. <http://www.fondation-arc.org/Les-cancers-du-sein/detecter-un-cancer-du-sein.html>, consulté, juillet /2020.
- [10] Institut national du cancer de France. <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Facteurs-de-risque/Predispositions-genetiques>, consulté en juillet, 2020.
- [11] Blogue IRM Québec, IRM du sein, <https://irmquebec.com/tag/irm-du-sein>, consulté en juillet 2020.
- [12] Ooreka, Santé. Mammographie. <https://cancer-du-sein.ooreka.fr/astuce/voir/639875/mammographie>, consulté en juillet 2020.

[13] I. Cheikhrouhou. Description et classification des masses mammaires pour le diagnostic du cancer du sein, thèse de doctorat en informatique, Université d'Evry Val d'Essonne, juin 2012.

[14] Rustvorsorge.at Mammographie, <http://www.breastcare.at/brustkrebsvorsorge/mammographie>, consulté en juillet 2020.

[15] A. Zaki. Apport du couple écho-mammographie dans le diagnostic des cancers du sein A propos de 100 cas, thèse de doctorat en médecine, Université Cadi Ayyad de Marrakech, Maroc, 2015

[16] Centre d'imagerie Duroc. La tomosynthèse mammaire ou mammographie 3D, <https://www.imagerieduroc.com/examens/centre-specialise-en-senologie/la-tomosynthese-mammaire-ou-mammographie-3d>, consulté en juillet 2020.

[17] P. DAVID Valeur primordiale de l'échographie en aval de la mammographie de dépistage du cancer du sein, Imagerie de la Femme, Vol. 14, Supp. 1, Pages 2-18, 2004.

[18] C. Balleyguier. L'angiommammographie : nouvelle technique d'imagerie diagnostique des lésions du sein, Contrast mammography: principes, applications, la Lettre du Sénologue N° 73, 2016.

[19] M. Dautat. Les bases de l'échographie, <https://ultrasonographie-vasculaire.edu.umontpellier.fr/files/2013/09/MD-Bases-Echographie-Abr-2013A.pdf>, CHU de Nîmes, 2013, consulté en juillet 2020.

[20] Sègbédji Goubalan. Contributions à l'analyse d'images médicales pour la reconnaissance du cancer du sein, thèse de doctorat en informatique, Université de Paris-Saclay, Décembre 2016.

[21] T. C. Chao, H. H. Chao, S. C. Chen, M. F. Chen, Sonographic Features of Breast Hamartomas, journal of ultrasound in medicine, Vol.26, pages 447-452, 2007.

[22] A. T. Stavros, D Thickman, C L Rapp, M A Dennis, S H Parker, G A Sisney. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Journal of Radiology Vol. 196, N°1, pages 123-134, 1995.

[23] I. Leconte. Quelle est la place de l'échographie dans le dépistage du cancer du sein ? Imagerie de la Femme Vol. 16, N°2, pages 95-100, 2006.

[24] M. ALAOUILAMRANI MOULAY YOUSSEF ; corrélations histopathologie des lésions mammaires classes bi-rads 4, 5 et 6 thèse de doctorat en médecine, Université M. Ben abdellah, Maroc, juin 2012.

[25] Programme québécois de dépistage du cancer du sein. Classification BI-RADS, <http://www.depistagesein.ca/classification-bi-rads/#.Xxnic0TPzcc>, consulté en juillet 2020.

- [26] C. Beaulac. Intelligence artificielle avec apprentissage automatique pour l'estimation de la position d'un agent mobile en utilisant les modèles de markov cachés, Mémoire de Maîtrise en mathématiques. Université du Québec, Montréal, 2015.
- [27] B. Braunschweig. Intelligence artificielle : Les défis actuels et l'action d'Inria- livre blanc, rapport d'activités, pages 1–21, <https://www.inria.fr/actualite/actualites-inria/livre-blanc-sur-l-intelligence-artificielle>, 2016, consulté en juin 2020.
- [28] Le blog des octos. Apprentissage par renforcement, de la théorie à la pratique. <https://blog.octo.com/apprentissage-par-renforcement-de-la-theorie-a-la-pratique>, consulté en juillet 2020.
- [29] 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.
- [30] Openclassrooms. Évaluez les performances d'un modèle de machine learning, <https://openclassrooms.com/fr/courses/4297211-evaluez-les-performances-dun-modele-de-machine-learning/4308256-evaluez-un-algorithme-de-classification-qui-retourne-des-valeurs-binaires>, consulté en juillet 2020.
- [31] Classification : ROC et AUC, <https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc>, consulté en juillet 2020.
- [32] L. Bottou. F. E. Curtis, J. Nocedal. Optimization Methods for Large-Scale Machine Learning, *Siam Reviews*, Vol. 60, N°.2, pages 223-311, 2018.
- [33] Le big data. Analyse de données, <https://www.lebigdata.fr/analyse-de-donnees-algorithmes#:~:text=La%20régression%20linéaire%20est%20,»%20ou%20»%20variable%20i ndépendante%20«%20>, consulté en juillet 2020.
- [34] Openclassrooms. Programmez votre première régression linéaire, <https://openclassrooms.com/fr/courses/4011851-initiez-vous-au-machine-learning/4121986-programmez-votre-premiere-regression-lineaire>, consulté en septembre 2020.
- [35] R. Rakotomalala. Arbres de décision, tutoriel, Revue MODULAD, N°.33, pages 163-184, Laboratoire ERIC, Université Lumière Lyon 2, 2005.
- [36] K. Hanifi Bien comprendre l'algorithme des K-plus proches voisins (Fonctionnement et implémentation sur R et Python, <https://medium.com/@kenzaharifi/bien-comprendre-lalgorithme-des-k-plus-proches-voisins-fonctionnement-et-implémentation-sur-r-et-a66d2d372679>, consulté en juillet 2020.
- [37] L. QOLBIYATUL Klasifikasi K-NN Menggunakan Data Tooth Growth di R, <https://medium.com/@16611110/klasifikasi-k-nn-menggunakan-data-tooth-growth-di-r-ec18148f1cd9>, consulté en juillet 2020.
- [38] C. Touzet. Les réseaux de neurones artificiels, introduction au connexionnisme : cours, exercices et travaux pratiques, ouvrage, Université d'Aix Marseille, 1992.

- [39] Juri'Predis, le blog. Démystifier le Machine Learning, Partie 2 : les Réseaux de Neurones artificiels, <https://www.juripredis.com/fr/blog/id-19-demystifier-le-machine-learning-partie-2-les-reseaux-de-neurones-artificiels>, consulté en juillet 2020.
- [40] Wikistat. Réseaux de neurones, <https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-rn.pdf>, consulté en juillet 2020.
- [41] A. Cohen-Hadria, Estimations de descriptions musicales et sonores par apprentissage profond, thèse de doctorat en informatique, Université de la Sorbonne, 2019.
- [42] Réseaux de neurones, <https://docplayer.fr/58425981-3-6-reseaux-de-neurones.html>, consulté en juillet 2020.
- [43] Y. Yactine. Réseaux neuronaux, <http://mapdata.science/~stpierre/reseaux-neuronaux/node9.html>, 1999, consulté en juillet 2020.
- [44] J. Schmidhuber. Deep learning in neural network: An overview, Neural networks, Vol 61, pages 85-117, 2015.
- [45] H. Wang, B. Raj, E. P. Xing. On the origin of deep learning, arXiv:1702.0780, 2017.
- [46] R.Tavenard. keras & réseaux de neurones récurrents, https://rtavenar.github.io/teaching/neuralnets_td/html/rnn.html, consulté en juillet 2020.
- [47] M. Bouaziz. Réseaux de neurones récurrents pour la classification de séquences dans des flux audiovisuels parallèles, thèse de doctorat en Informatique, Université d'Avignon et des pays de Vaucluse, 2017.
- [48] Le Magit. Réseau neuronal convolutif, <https://www.lemagit.fr/definition/Reseau-neuronal-convolutif#:~:text=Les%20réseaux%20convolutifs%20sont%20une,des%20tâches%20descriptives%20et%20génératives>, consulté en juillet 2020.
- [49] 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.
- [50] Openclassrooms. Classez et segmentez des données visuelles, <https://openclassrooms.com/fr/courses/4470531-classez-et-segmentez-des-donnees-visuelles/5083336-decouvrez-les-differentes-couches-dun-cnn>, consulté en juillet 2020.
- [51] Datasciencetoday. Deep learning, <https://www.datasciencetoday.net/images/2018/08/cnn1/image004.png>, consulté en juillet 2020.
- [52] Pensée artificielle. Les réseaux de neurones convolutifs, <https://penseeartificielle.fr/focus-reseau-neurones-convolutifs/>, consulté en juillet 2020.
- [53] S. Hwang, H.E. Kim, J.Jeong, H.J. Kim.. A novel approach for tuberculosis screening based on deep convolutional neural networks, Proc. SPIE 9785, Medical Imaging 2016.

- [54] S. J. Pan, Q. Yang. A survey on transfer learning, IEEE Transactions on knowledge and data engineering, Vol. 22, N°.10, pages 1345–1359, 2010.
- [55] K. Simonyan, A. Zisserman. Very deep convolutional networks for largescale image recognition. arXiv:1409.1556, 2015.
- [56] O. Ronneberger, P. Fischer, T. Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pages 234–241. Springer, 2015.
- [57] M.A. Nayeem, Md. A. Mannan Joadder, S. A. Shetu, F. R. Jamil, A. Helal. Feature selection for breast cancer detection from ultrasound images. International Conference on Informatics, Electronics and Vision, ICIEV, 2014.
- [58] M. H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentís, R. Zwigelaar, A. Davison, R. Martí. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. IEEE Journal of Biomedical and Health Informatics. Vol.22, N°.4, pages. 1218-1226, 2018.
- [59] Z. Cao, L. Duan, G. Yang, Duan, T. Yue, Q. Chen. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. BMC Medical Imaging. Vol. 19, N°. 51 <https://doi.org/10.1186/s12880-019-0349-x>, 2019.
- [60]. W. Al-Dhabyani, M. Goma, H. Khaled, A. Fahmy. Deep Learning Approaches for Data Augmentation and Classification of Breast Masses using Ultrasound Images. International Journal of Advanced Computer Science and Applications. Vol. 10, N°.5, pages 618-627, 2019.
- [61] BUSI dataset. <https://scholar.cu.edu.eg/?q=afahmy/pages/dataset>, consulté en janvier 2020.
- [62] M. Randrianarivony. Détection de concepts et annotation automatique d'images médicales par apprentissage profond, thèse de doctorat en informatique, Université d'Antananarivo, 2018.
- [63] Özyurt, Fatih. Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures. The Journal of Supercomputing. Pages 1-19. [10.1007/s11227-019-03106-y](https://doi.org/10.1007/s11227-019-03106-y), 2019.
- [64] K. Ameer, S. Azzouz. Authentification faciale par apprentissage profond, mémoire de Master en électronique Université de Saad Dahleb- Blida 1, octobre 2019.
- [65] Neurohive. VGG16 – Convolutional Network for Classification and Detection, <https://neurohive.io/en/popular-networks/vgg16>, consulté en septembre 2020.
- [66] Guide avancé d'Inception v3 sur Cloud TPU, <https://cloud.google.com/tpu/docs/inception-v3-advanced?hl=fr>, consulté en juillet 2020.
- [67] S. H. Tsang. Review: Inception-v3, <https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-image-classification-in-ilsrvc-2015-17915421f77c>, consulté en juillet 2020.

- [68] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna. Rethinking the inception architecture for computer vision. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [69] A. Fausto. Les méthodes de Transfer Learning, <https://meritis.fr/ia/methodes-de-transfer-learning>, 2018, consulté en juillet 2020.
- [70] P. Marcelino. Transfer learning from pre-trained models, <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>, consulté en juillet 2020.
- [71] TIC magazine. Python : tout savoir sur le principal langage Big Data et Machine Learning, <https://ticmagazine.bf/python-tout-savoir-sur-le-principal-langage-big-data-et-machine-learning>, consulté en janvier 2020.
- [72] P. Fuchs, P. Poulain. Programmation en Python pour les sciences de la vie, Edition Dunod, 2020.
- [73] Présentation du langage Python, <http://www.linux-center.org/articles/9812/python.html>, consulté en janvier 2020.
- [74] Anaconda, <https://anaconda.fr.softonic.com>, consulté en janvier 2020.
- [75] Académie de Normandie. Distribution Anaconda pour Python, <https://numerique-sciences-informatiques.discip.ac-caen.fr/distribution-anaconda-pour-python>, consulté en janvier 2020.
- [76] Anaconda. Your data science toolkit, <https://www.anaconda.com/products/individual>, consulté en janvier 2020.
- [77] Spyder, <https://www.spyder-ide.org/>, consulté en janvier 2020.
- [78] EDUCBA. Introduction to Tensorflow, <https://www.educba.com/introduction-to-tensorflow>, consulté en janvier 2020.
- [79] R. Vickery. Beginners Guide to Deep Learning with TensorFlow, <https://towardsdatascience.com/beginners-guide-to-deep-learning-with-tensorflow-ca85969b2f2>, consulté en Janvier 2020.
- [80] Y. Benoit. <https://blog.engineering.publicissapient.fr/2017/03/01/tensorflow-deep-learning-episode-1-introduction>, consulté en janvier 2020.
- [81] Keras documentation, <https://keras.io/#keras-the-python-deep-learning-library>, consulté en juin 2020.
- [82] Actuaia. Pourquoi utiliser Keras ? <https://www.actuaia.com/keras/pourquoi-utiliser-keras>, consulté en juin 2020.
- [83] S. Bancal. Introduction à Numpy, <https://enacit.epfl.ch/cours/python/scientifique/numpy.html#numpy>, consulté en janvier 2020.

[84] A. Oukahere. Démarrer avec Flask - Un micro framework Python, <https://www.kaherecode.com/tutorial/demarrer-avec-flask-un-micro-framework-python>, consulté en juillet 2020.

[85] Web & Tech. HTML HyperText Markup Langage. <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203255-html-hypertext-markup-langage-definition-traduction/>, consulté, en juillet 2020.

[86] Définitions et usages du HTML et du CSS. <https://www.pierre-giraud.com/html-css-apprendre-coder-cours/definition-utilisation/>, consulté en juillet 2020.

[87] Openclassrooms. Créez des interfaces graphiques avec Tkinter, <https://openclassrooms.com/fr/courses/235344-apprenez-a-programmer-en-python/234859-creez-des-interfaces-graphiques-avec-tkinter>, consulté en juillet 2020.