

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

UNIVERSITÉ BADJI MOKHTAR - ANNABA
BADJI MOKHTAR – ANNABA UNIVERSITY



جامعة بادجي مختار – عنابة

Faculté : TECHNOLOGIE

Département : ELECTRONIQUE

Domaine : SCIENCES ET TECHNOLOGIES

Filière : RESEAUX ET TELECOMMUNICATION

Spécialité : RESEAUX ET TELECOMMUNICATION

Mémoire

Présenté en vue de l'obtention du Diplôme de Master

Thème :

Classification du cancer du sein basée sur l'apprentissage profond

Présenté par : *TAYAA HADJER*
ALLAMA AMEL

Encadrant : *BOUTERAA NADIA* M.C.A Université Badji Mokhtar Annaba

Jury de Soutenance :

Mr. BOUGHAZI	PROFESSEUR	UBMA	Président
Mme BOUTERAA NADIA	M.C.A	UBMA	Encadrant
Mme ZERMI NARIMA	M.C.A	UBMA	Examineur

Année Universitaire : 2022/2023

REMERCIEMENT

En tout premier lieu, nous remercions Allah le Tout Puissant et le Bienveillant qui nous a accordé d'innombrables bénédictions et de nous avoir donné la force et la patience tout au long de notre travail afin de réaliser ce projet avec succès.

Par cette même occasion, nos forts remerciements vont en particulier à notre professeur BOUTERAA Nadia, pour avoir accepté notre encadrement et pour l'assistance remarquable de sa part, et les maintes explications valeureuses pour nous permettre de comprendre ce sujet très pointu et aux nombreuses heures consacrées pour arriver à ces résultats tangibles tant attendus pour faire aboutir notre présent projet.

Nos remerciements les plus chaleureux et nos respects les plus dévoués vont également aux membres du jury qui nous auront fait l'honneur de juger ce travail et qui par leur disponibilité et leurs observations nous auront permis d'enrichir ce mémoire.

Un grand merci également à tous nos enseignants au niveau du Département D'électronique pour leur dévouement et leur collaboration à nos connaissances tout au long de ce cycle universitaire .Nous vous sommes infiniment reconnaissants.

Enfin les mots sont trop pauvres pour exprimer notre reconnaissance à nos chers parents pour leur amour, leur aide, leurs sacrifices et leurs prières pour nous éduquer et nous encourager .Merci beaucoup d'être toujours à nos côtés. Nous n'aurions pas pu faire tout cela sans vous.

Sans oublier bien sûr de remercier profondément tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.

Merci à toutes et à tous.

Dédicaces

Je dédie ce travail:

A mes chers parents qui m'ont donné confiance et motivation

A mes frères: Oussama, Islam et Yahia

Aux membre de ma grande famille

A mes amies et en particulier A mon binome Hadjer pour sa gentillesse et sa diligence.

ALLAMA AMEL

Toutes les lettres ne saurai en trouver les mots qu'il faut... Tous les mots ne sauraient exprimer la gratitude, la tendresse, Le respect, la reconnaissance... Aux êtres qui me sont chers et c'est tout simplement qu'à eux que je dédie ce mémoire :

A ma très chère mère SALMI Hafsia, honorable, tu représentes pour moi le symbole de la bonté par excellence. Aucune dédicace ne saurait être assez éloquente pour exprimer ce que tu mérites pour tous les sacrifices .Tu as fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études. Je te dédie ce travail en témoignage de mon profond amour. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur. Je t'aime maman.

A mon très chère père TAYAA Mohamed Sadek, A celui qui m'a aidé à découvrir le 'savoir' le trésor inépuisable. De tous les pères, tu as été le meilleur, tu as su m'entourer d'attention, m'inculquer les valeurs nobles de la vie, m'apprendre le sens du travail, de l'honnêteté et de la responsabilité. Ce travail est ton œuvre, toi qui m'a donné tant de choses et tu continues à le faire. Merci d'avoir été toujours là pour moi.

A mes chères sœurs « RANIA et SALMA» Les mots ne suffisent guère pour exprimer l'attachement, l'amour et L'affection que je porte pour vous.

A mon adorable binôme ALLAMA Amel, à qui j'exprime toute ma gratitude pour l'aide qu'elle m'a apportée durant mon cursus. Tu es une personne formidable et c'est un honneur de partager ce travail avec toi.

Tayaa hadjer

Résumé :

L'intelligence artificielle a pris une place considérable dans le domaine médicale, en effet et à travers ce travail, nous avons implémenté et expérimenté deux modèles pré-entraînés basés sur les réseaux de neurones convolutionnels Inceptionv3 et AlexNet grâce à un apprentissage par transfert pour la classification des images IRM du cancer du sein issues d'une base de données publique. Nous avons testé les deux modèles selon trois optimiseurs Adam, Sgdm et RMSprop pour en tirer le meilleur en terme de précision.

Mot clés : Cancer du sein, Apprentissage profond, CNN, Apprentissage par transfert, optimiseurs.

Abstract:

Artificial intelligence has taken a considerable place in the medical field. In this work , we have implemented and tested two pre-trained models based on the convolutional neural networks Inceptionv3 and AlexNet through transfer learning for the classification of breast cancer MRI images from a public database. We tested both models with three optimizers, Adam, Sgdm and RMSprop, to obtain the best accuracy.

Key words: Breast cancer , Deep learning ,Transfer learning ,Convolutional neural network, Optimizers .

ملخص

يحتل الذكاء الاصطناعي مكانة بالغة الأهمية في مجال الطب و لهذا قمنا في هذا البحث بتطبيق و تنفيذ نموذجين مدربين مسبقا للذات يرتكزان على الشبكات العصبية الالتفافية Inceptionv3 و شبكة Alexnet و هذا عن طريق التعلم الانتقالي، بهدف تصنيف صور بتقنية التصوير بالرنين المغناطيسي لسرطان الثدي مأخوذة من قاعدة بيانات عامة. اختبرنا هذان النموذجان مع ثلاث محسنات : Adam , Sgdm و RMSprop و ذلك للحصول على الأفضل من حيث الدقة

الكلمات المفتاحية : سرطان الثدي , التعلم العميق, الشبكات العصبية الالتفافية, التعلم الانتقالي , المحسنات .

Liste des acronymes

Ada Grad	Adaptive Gradient Algorithm
Adam	Adaptive Moment Estimation
CNN	Convolutional Neural Network
ANN	Artificiel Neural Network
DL	Deep Learning
FC	Fully Connected Layer
IA	Intelligence Artificielle
ML	Machine Learning
DL	Deep Learning
ADN	Acide DésoxyriboNucléique
IRM	l'imagerie par résonance magnétique
Pet-scan	Tomographie par Émission de Positrons
Les rayons X	radiographie avec des rayons X
LR	Learning Rate
CAD	Détection Assisté par Ordinateur
ROC	Reciever Operating Characterstic
AUC	Area Under the ROC
RVB	Rouge, Vert, Blue (système de codage informatique des couleurs)
CV	Cross-Validation
ReLU	Rectified linear unit
RMS prop	Root Mean Square Propagation
Sgdm	Stochastic Gradient Descent with Momentum
Inceptionv3	Inception Version 3
Googlenet	InceptionV1 GoogLeNet
GAN	Réseaux adversariaux génératifs
RNN	Réseaux neuronaux récurrent
RBFN	Réseaux à fonction de base radiale
ILSVRC	Image Net Large scale Visual Recognition Challenge

Liste des fonctions :

$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}}$	(2.1)
$\theta_{l+1} = \theta_l - \alpha \nabla E(\theta_l) + \gamma (\theta_l - \theta_{l-1})$	(2.2)
$M_l = \beta_1 m_{l-1} + (1 - \beta_1) \nabla E(\theta_l)$	(2.3)
$V_l = \beta_2 v_{l-1} + (1 - \beta_2) [\nabla E(\theta_l)]^2$	(2.4)
$\theta = \theta - \alpha m_l \sqrt{v_l} + \epsilon$	(2.5)
$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$	(3.1)
$\text{Precision} = \frac{TP}{TP+FP}$	(3.2)
$\text{Sensitivity/Recall} = \frac{TP}{TP+FN}$	(3.3)
$\text{Specificity} = \frac{TN}{TN+FP}$	(3.4)
$\text{F1-Score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$	(3.5)

Liste des figures

Figure 1.1: Cellule humaine	4
Figure 1.2: Division cellulaire	5
Figure1.3: Division cellulaire dans le cas d'un cancer.....	6
Figure 1.4: Anatomie du sein	6
Figure 1.5 : Exemple d'un cancer du sein	7
Figure 1.6 : Symptomes du cancer du sein.....	9
Figure 1.7 : Mégration d'un cancer.....	10
Figure 1.8: Appareil de mammographie.....	11
Figure 1.9: Exemple d'échographie mammaire	12
Figure 1.10 : Appareil utilisé pour IRM mammaire.....	12
Figure1.11: Exemple d'imagerie IRM mammaire.....	13
Figure 2.1: Schéma d'organisation d'apprentissage automatique.....	18
Figure 2.2: Diagramme de l'apprentissage supervisé.....	19
Figure 2.3: Principe de l'apprentissage non-supervisé	19
Figure 2.4: L'apprentissage semi-supervisé.....	20
Figure 2.5: Apprentissage par renforcement	21
Figure 2.6: Artificial Intelligence vs Machine learning vs Deep learning	22
Figure 2.7: Neurone Biologique.....	23
Figure 2.8: Réseaux de Neurones Artificiels.....	24
Figure2.9: La similitude entre le neurone biologique et le neurone artificiel.....	24
Figure2.10: les réseaux de neurones.....	25
Figure 2.11: Exemple d'une architecture CNN.....	25
Figure 2.12: Opération de convolution.....	27
Figure 2.13: La couche d'activation ReLU.....	27
Figure 2.14: Exemple applicatif des différents Pooling.....	29
Figure 2.15: Le flattening.....	29
Figure 2.16: Exemple de couche fully-connected.....	30
Figure 2.17: Architecture AlexNet.....	32
Figure 2.18: Diagramme générale de googlenet.....	34

Figure 2.19: Module de base de inceptionv1.....	35
Figure 2.20: Inception module A.....	36
Figure 2.21: Inception module B.....	37
Figure 2.22: Inception module C.....	37
Figure2.23: Architecture du réseau Inceptionv3.....	38
Figure3.1: Image de type tumeur bénigne.....	44
Figure 3.2: Image de type tumeur maligne.....	44
Figure 3.3: Fonctionnement de la méthode k-fold CV.....	45
Figure 3.4: Images redimensionnées	46
Figure 3.5: Architecture globale du modèle proposé.....	48
Figure 3.6: Trois dernières couches du modèle inceptionv3.....	48
Figure 3.7: Trois dernières couches avec Transfer Learning.....	49
Figure 3.8: Trois dernières couches du modèle AlexNet.....	49
Figure 3.9: Trois dernières couches du modèle Alexnet avec Transfer Learning.....	50
Figure 3.10: Matrice de confusion.....	51
Figure 3.11: Courbe ROC.....	52
Figure 3.12: Courbes d'évaluation de l'apprentissage(Accuracy and Loss) pour le split 1,2,3,4 et 5 avec Inceptionv3.....	55
Figure 3.13: Matrice de confusion globale pour l'ensemble de test (Inceptionv3 et option Adam).....	56
Figure 3.14: La courbe ROC inceptionv3.....	57
Figure 3.15: Matrice de confusion globale avec sgd.....	58
Figure 3.16: Matrice de confusion globale avec rmsprop.....	58
Figure3.17: Courbes d'évaluation de l'apprentissage(Accuracy and Loss) pour le split 1,2,3,4 et 5 avec AlexNet.....	62
Figure3.18: Matrice de confusion globale avec AlexNet.....	62
Figure3.19: La courbe ROC AlexNet.....	63
Figure 3.20: Image tests.....	64

Liste des tableaux

Tableau 2.1: Différences entre les types d'apprentissage automatique.....	21
Tableau 2.2: Tableau récapitulatif de l'architecture Alexnet.....	33
Tableau 3.1: Taux d'apprentissage et temps d'exécution en changeant le nombre d'époque	53
Tableau 3.2: Mesure de performance (inceptionv3).....	56
Tableau 3.3: tableau récapitulatif des mesures de performances selon les trois optimiseurs.....	59
Tableau 3.4: mesure de performances avec Alexnet.....	63
Tableau 3.5: Tableau comparatif des valeurs des indicateurs de performances d'Inceptionv3 et Alexnet.....	63
Tableau 3.6: Résultats de validation.....	64

Table des matières

Remerciements	II
DédicaceIII
Résumé	IV
Liste des acronymes	V
Liste de fonctions	VI
Liste des figures	VII
Liste des tableaux	IX
Table des matières	X
Introduction générale	1
CHAPITRE 1 : Le cancer du sein	3
1.1 Introduction	4
1.2 Cellule humain	4
1.3 La division cellulaire	5
1.4 La division de cellules cancéreuses	5
1.5 Anatomie du sein chez la femme	6
1.6 Le cancer du sein	7
1.6.1 Tumeurs Bénignes	8
1.6.2 Tumeurs Malignes	8
1.7 Les Symptomes	8
1.8 La Métastase	10
1.9 Le diagnostique	11
1.9.1 Bilan sanguin	11
1.9.2 Imagerie médicale	11
1.9.2.1 la mammographie	11
1.9.2.2 Echographie	11
1.9.2.3 IRM (Imagerie par résonance magnétique)	12

1.9.2.4	Le pet-scan	13
1.9.3	La biopsie	13
1.10	Les facteurs de risque	14
1.11	Objectifs et motivations de l'intelligence artificielle	15
1.12	Conclusion	15
CHAPITRE 2 : Généralités sur l'apprentissage automatique		16
2.1	Introduction	17
2.2	L'Intelligence Artificielle	17
2.2.1	Algorithme à base de règles	17
2.2.2	Algorithme à base d'apprentissage	17
2.3	L'Apprentissage automatique (Machine Learning)	17
2.3.1	Apprentissage automatique (ML) supervisé.	18
2.3.2	Apprentissage automatique (ML) non supervisé.	19
2.3.3	L' Apprentissage automatique semi-supervisé	20
2.3.4	L' Apprentissage automatique par renforcement	20
2.4	L'Apprentissage profond	22
2.5	Les réseaux de neurones	23
2.5.1	Neurone Biologique	23
2.5.2	Réseaux de Neurones Artificiels	23
2.6	Les Réseaux de neurones convolutionnels	25
2.7	Les Différentes couches des réseaux de neurones convolutifs (CNN)	26
2.7.1	Couche de convolution	26
2.7.1.a	Les filtres	26
2.7.1.b	Opération de convolution	26
2.7.2	La couche de conection RELU	27
2.7.3	Cross channel normalization	27
2.7.4	Couche de pooling	28
2.7.4.1	Average - pooling	28
2.7.4.2	Max – pooling	28
2.7.4.3	Min – pooling	28
2.7.5	Le flattening (mise à plat)	29

2.7.6	Couches entièrement connectées (fully connected FC)	30
2.7.7	La couche dropout	30
2.7.8	Couche prob <<softmax>>	30
2.8	Fonctionnement des CNN	31
2.9	Quelques architectures des CNNs	32
2.9.1	AlexNet	32
2.9.2	Googlenet	33
2.9.3	Inceptionv3	34
2.9.3.1	Architecture du modèle Inceptionv3.	34
2.9.3.2	Factorisation des convolutions	35
2.10	Les options d'apprentissage	38
2.10.1	Sgdm	38
2.10.2	AdaGrad	39
2.10.3	RMSprop	39
2.10.4	Adam	40
2.11	Conclusion	41
CHAPITRE 3 : Implementation et Résultats		42
3.1	Introduction	43
3.2	Les outils matériels et logiciel	43
3.2.1	Configuration matériel utilisée dans l'implémentation	43
3.2.2	Logiciel utilisée dans l'implémentation	43
3.3	Description de la base de données	44
3.4	Prétraitement de la base de données	44
3.4.1	Divisions du dataset (Dataset splitting).	44
3.4.2	Redimensionnement des images	46
3.4.3	Augmentation de données (Data augmentation).	46

3.5 Classification avec le modèle d'apprentissage	47
3.5.1 Application de l'apprentissage par transfert (Transfert Learning)	47
3.5.2 Architectures du modèle utilisé dans l'implémentation	47
3.5.3 Hyper paramètres d'apprentissage	50
3.6 Les indicateurs de performance	50
3.6.1 Matrice de confusion	50
3.6.2 La courbe ROC	51
3.7 Résultats et discussion	52
3.7.1 Tests selon les options d'apprentissage	57
3.7.2 Validation du modèle Inceptionv3.....	64
3.8 Conclusion	65
Conclusion générale	66
ANNEXE	67
Bibliographie	69

Introduction générale :

Le cancer du sein est le cancer féminin le plus grave et le plus fréquent, c'est la première cause de décès chez les femmes de 35 à 55 ans.

En 2020, on a recensé 2,3 millions de femmes atteintes du cancer du sein et 685 000 décès par cancer du sein dans le monde. Fin 2020, 7,8 millions de femmes en vie s'étaient vues diagnostiquer un cancer du sein au cours des cinq années passées, ce qui fait du cancer du sein le cancer le plus courant à l'échelle du globe. À l'échelle mondiale, les femmes perdent plus d'années de vie (espérance de vie corrigée de l'incapacité) en raison du cancer du sein que de n'importe quel autre type de cancer.

Dans les pays à revenu élevé, le taux de mortalité par cancer du sein comparatif par âge a chuté de 40 % entre les années 1980 et 2020. Les pays qui sont parvenus à réduire la mortalité par cancer du sein ont réussi à atteindre une réduction de la mortalité annuelle par cancer du sein de 2 à 4 % par an. Si la mortalité annuelle baissait chaque année de 2,5 % dans le monde, on éviterait 2,5 millions de décès par cancer du sein entre 2020 et 2040 [1]. Dans le but de réduire la morbidité et la mortalité par le cancer du sein, une détection précoce est nécessaire.

L'imagerie par résonance magnétique IRM est une technique d'imagerie médicale qui tire profit de l'effet d'un champ magnétique sur les molécules d'eau qui composent notre corps.

Coûteuse, l'IRM permet d'obtenir des images en 3D d'une précision inégalable, en particulier lorsqu'elle est associée à un produit de contraste.

L'interprétation des clichés en un temps limité pour la détection précise des pathologies, est une tâche très difficile ; cela nécessite alors l'utilisation d'outils d'aide à la détection par ordinateur (CAD) pour améliorer les performances du lecteur.

Dans notre travail nous avons introduit l'apprentissage profond dans le but de classer les tumeurs cancéreuses du sein de type maligne ou bénigne en appliquant deux architectures de réseaux convolutifs : inceptionv3 et AlexNet avec un apprentissage par transfert (transfert learning), ces modèles seront appliqués sur une base de données conçue au cancer du sein.

Notre mémoire est ainsi reparti en trois chapitres :

- Le premier chapitre décrit brièvement l'anatomie du sein, les types de tumeurs et les différentes techniques d'imagerie médicales.
- Le deuxième chapitre est consacré aux concepts de l'intelligence artificielle où nous décrivons ses sous ensembles : l'apprentissage automatique et l'apprentissage profond y compris les réseaux de neurones convolutifs de plus, quelques architectures de modèles pré entraînés ont été présentées.
- Le troisième chapitre comporte l'implémentation de notre modèle, les résultats que nous avons obtenus suivie d'une discussion mettant le point sur le meilleur modèle retenu en terme de précision.

CHAPITRE 1

Le cancer du sein

1.1.Introduction

Le cancer est une maladie des plus redoutable de l'histoire de l'humanité, elle ne cesse de mettre en épreuve la capacité de la science à la combattre puisqu'elle est dotée d'une particularité unique : celle-ci vient de nous-mêmes.

A travers ce chapitre, nous définirons le cancer en général et celui du sein en particulier sous ses différents types. Pour conclure, nous étudierons les étapes qui découlent de son diagnostic afin d'en déduire les motivations ayant déterminé l'introduction de l'IA au cœur de ce domaine spécifique de la médecine. Pour bien comprendre le cancer, il est nécessaire de définir la notion de cellule.

1.2. La cellule humaine

Une cellule constitue la partie la plus élémentaire de l'être. Comme illustré dans la figure ci-après, elle se présente sous forme de membrane enfermant un noyau où se loge L'ADN : c'est le manuel de production de protéines.

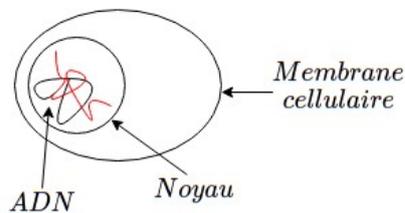


Figure 1.1. Cellule humaine

1.3. La division cellulaire

Afin de se reproduire, une cellule se subdivise pour engendrer deux autres cellules filles. Cette opération est effectuée pour remplacer 100 milliards de cellules par jour chez l'homme. La figure suivante illustre ce processus.

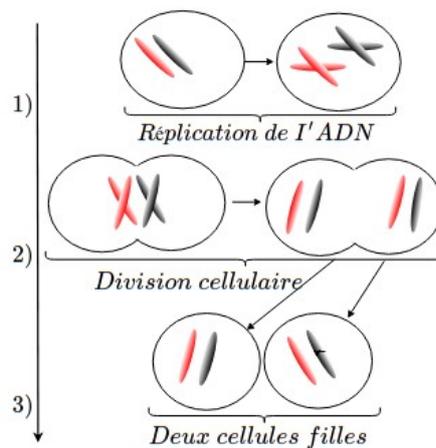


Figure 1.2. Division cellulaire

Au moment de la division, l'ADN est recopié à l'identique dans les deux autres cellules résultantes : c'est ici qu'un potentiel cancer peut se développer.

1.4. La division de cellules cancéreuses

Dans certains cas de reproduction, l'ADN se recopie faussement vers les cellules filles, c'est la mutation. Quand une cellule mute, elle s'autodétruit seulement, quand la molécule d'ADN est suffisamment altérée, la cellule perd sa capacité à s'éliminer et continue ainsi de proliférer en créant d'autres mutations formant un agglomérat de cellules qui grandit sans limite : c'est la tumeur. La figure suivante illustre ce phénomène.

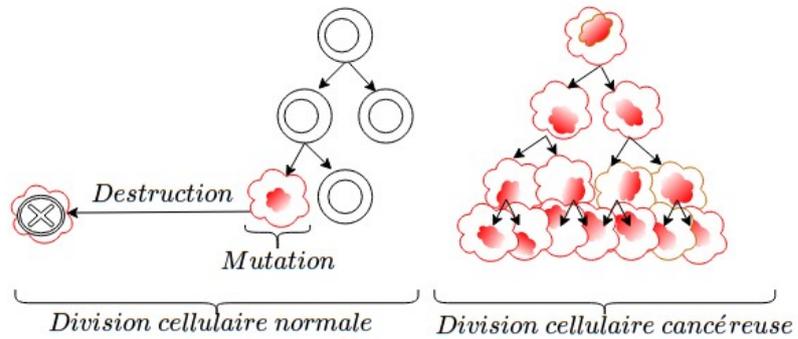


Figure 1.3 . Division cellulaire dans le cas d'un cancer

1.5. Anatomie du sein chez la femme

La fonction biologique du sein est de produire du lait afin de nourrir un nouveau-né. La structure du sein est illustrée dans la figure 1.4

- Silo graisseux : principalement de la graisse.
- Lobules : les lobules sont un groupe d'alvéoles responsables de la production de lait.
- Canal lactifère : responsable du transport de lait jusqu'au mamelon.

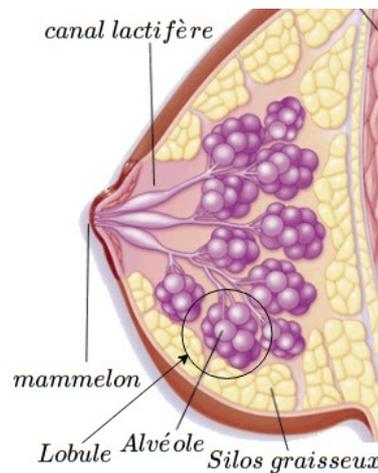


Figure 1.4. Anatomie du sein[2]

1.6. Le cancer du sein

Le cancer du sein est le premier type de cancer connu de l'histoire. Il est également l'un des plus fréquents et invasifs chez la femme. Sa présence augmente avec l'âge, mais il peut aussi toucher des femmes jeunes puisqu'un quart des cas sont diagnostiqués avant l'âge de 50 ans. Le cancer du sein prend naissance dans les cellules du sein. La tumeur cancéreuse (maligne) est un groupe de cellules cancéreuses qui peuvent envahir et détruire le tissu voisin. Elle peut aussi se propager (métastases) à d'autres parties du corps. Le premier symptôme du cancer du sein est la présence d'une boule au niveau du sein, correspondant à la tumeur. Elle peut également s'accompagner de ganglions durs au niveau de l'aisselle (ganglions axillaires) correspondant à une propagation du cancer, ainsi qu'à des modifications cutanées au niveau du sein et du mamelon (peau d'aspect capitonné et un mamelon qui entre au lieu de sortir). Le sein peut progressivement se déformer et s'ulcérer, ce qui se traduit parfois par un écoulement du mamelon, d'un seul côté. Si le cancer est diagnostiqué tardivement, la tumeur peut se propager et déclencher d'autres symptômes tels que des nausées, des vomissements, une perte de poids, une jaunisse, des douleurs osseuses ou encore des céphalées, un essoufflement ou une toux.

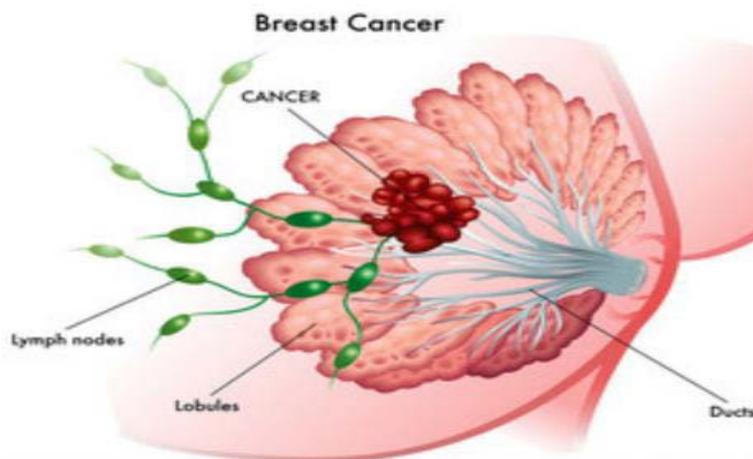


Figure 1.5.Exemple d'un cancer du sein

En bref, le cancer du sein est une tumeur maligne se développant à partir des cellules constituant la glande mammaire. Les cellules malignes se multiplient de manière anarchique et forment la

tumeur. Lorsqu'un cancer du sein n'est pas traité, les cellules tumorales se propagent localement et envahissent les organes de voisinage. Elles peuvent également se propager par voie sanguine ou lymphatique pour atteindre les organes situés à distance (métastases). Les organes les plus fréquemment atteints par les métastases sont les os, les poumons, le foie et le cerveau. Il existe deux types de tumeurs du sein : les tumeurs non cancéreuses ou "bénigne" et les tumeurs cancéreuses "maligne" [3].

1.6.1. Tumeurs bénignes

Les tumeurs bénignes ont des contours bien délimités. Elles ont une croissance lente et restent localisées dans le tissu ou l'organe dans lequel elles sont apparues. Elles n'entraînent pas de métastases dans d'autres endroits du corps. Les tumeurs bénignes sont composées de cellules qui ressemblent aux cellules normales du tissu concerné. La tumeur bénigne qui se développe le plus souvent dans le sein est le fibroadénome.

Lorsqu'une tumeur est diagnostiquée comme étant bénigne, les médecins la laissent généralement seule plutôt que de la retirer. Même si ces tumeurs ne sont généralement pas agressives envers les tissus environnants, elles peuvent parfois continuer à se développer, en appuyant sur les organes et en provoquant des douleurs ou d'autres problèmes. Dans ces situations, la tumeur est retirée, permettant à la douleur ou aux complications de s'apaiser [4].

1.6.2. Tumeurs malignes

Les tumeurs malignes sont cancéreuses et agressives car elles envahissent et endommagent les tissus environnants. Lorsqu'une tumeur est soupçonnée d'être maligne, le médecin procédera à une biopsie afin de déterminer la gravité ou l'agressivité de la tumeur.

1.7. Les symptômes

Les symptômes listés ci-dessous ne signifient pas nécessairement qu'il s'agit d'un cancer du sein. Mais si c'est le cas, il est important de le détecter le plus tôt possible.

a) Une boule dans un sein : Une boule ou une masse dans un sein est le signe d'un cancer du sein le plus couramment observé. Cette masse, en général non douloureuse, est le plus souvent de consistance dure et présente des contours irréguliers. Elle apparaît par ailleurs comme « fixée » dans le sein.

b) Des ganglions durs au niveau de l'aisselle : Une ou plusieurs masse(s) dures à l'aisselle signifient parfois qu'un cancer du sein s'est propagé aux ganglions axillaires. Les ganglions restent toutefois indolores

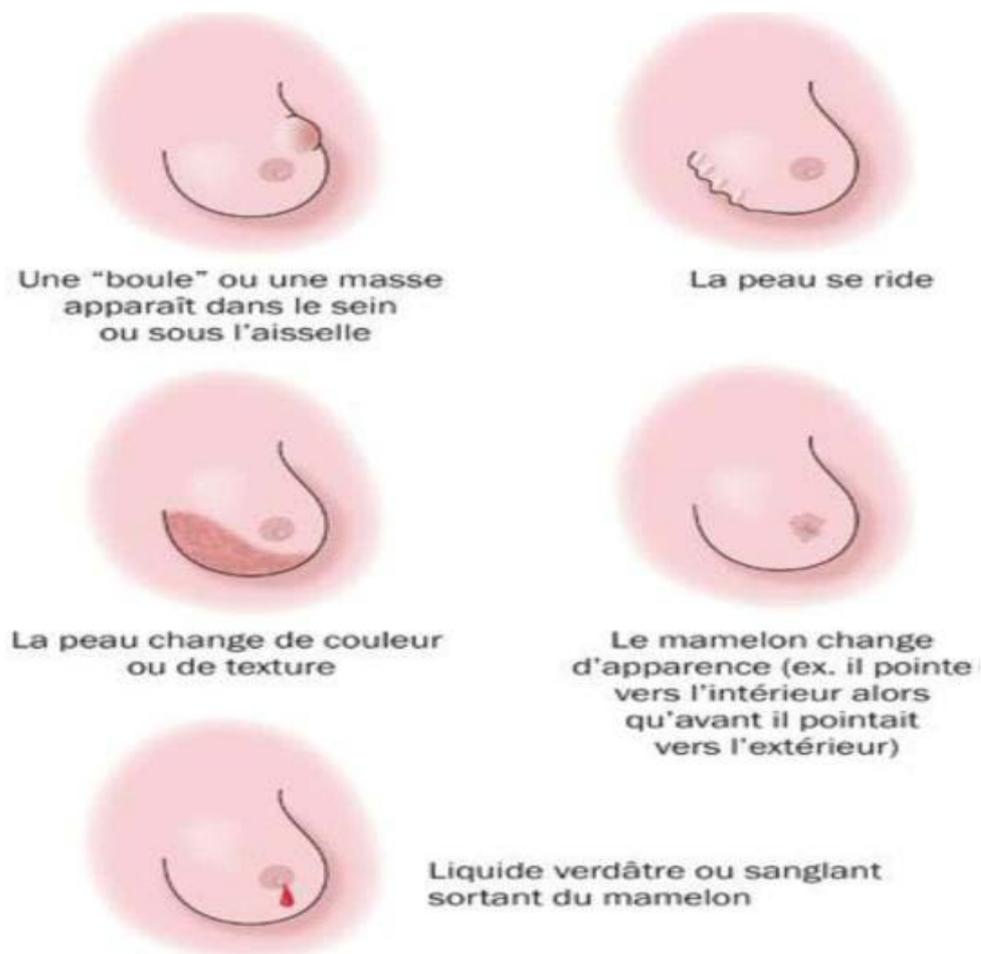


Figure 1.6. Symptômes du cancer du sein [5].

c) Des modifications de la peau du sein et du mamelon : La peau du sein peut devenir capitonnée (et prendre l'aspect d'une peau d'orange) ou plissée.

d) Le mamelon peut pointer vers l'intérieur, alors qu'habituellement, il est dirigé vers l'extérieur.

e) Le sein peut se déformer et perdre de son galbe, des rides peuvent apparaître.

f) La peau du sein peut être rouge, écorchée (ulcérée), recouverte de croûtes et celle du mamelon peut se mettre à peler.

g) Un écoulement d'un seul mamelon : peut être le signe de cancer du sein, en particulier s'il se manifeste sans qu'il n'y ait de compression du mamelon et s'il contient du sang ou s'il est verdâtre.

h) Un changement de la taille ou de la forme du sein : Une rougeur, un oedème et une chaleur importante au niveau du sein peuvent être le signe d'un cancer du sein inflammatoire (le cancer inflammatoire du sein est un cancer rare ; il représente 1 à 4 % de tous les cas de cancer du sein. Les cellules cancéreuses ont la caractéristique de se déplacer rapidement dans les vaisseaux lymphatiques de la peau du sein qu'elles finissent par bloquer. C'est ce qui provoque l'inflammation locale du sein). La constatation d'un de ces signes impose une consultation médicale sans retard. Le médecin jugera des examens complémentaires nécessaires.

1.8. La Métastase

L'organisme ayant conscience de la naissance de nouvelles cellules crée des vaisseaux sanguins afin de leurs apporter des nutriments. Les cellules cancéreuses empruntent ces vaisseaux pour migrer vers d'autres parties du corps : c'est la métastase. La membrane basale sert de moyen d'ancrage aux cellules, elle intervient comme filtre pour leurs nutrition, si celle-ci est perforée par les cellules cancéreuses, on parle de tumeur maligne [6].

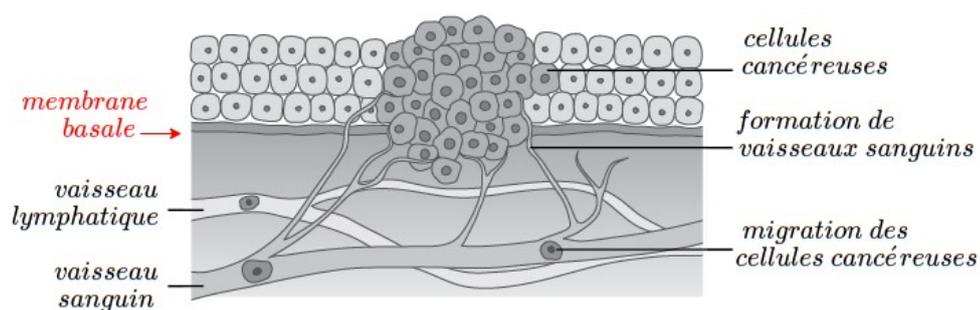


Figure1.7.Migration d'un cancer

1.9. Le diagnostique

Le diagnostic est une tâche particulièrement délicate pour le corps médical, nombreux sont les cancers qui demeurent à ce jour difficile à détecter toutefois, la pluralité des diagnostics suivent inévitablement les étapes suivantes [7].

1.9.1 Bilan sanguin

Un bilan sanguin est demandé par le médecin si celui-ci détecte une grosseur (nodules) anormale au niveau du sein ou que le/la patiente présente des symptômes inquiétants.

1.9.2 Imagerie médicale

Si une anomalie est présente sur un bilan sanguin, une imagerie médicale est sollicitée selon l'altération détectée. Parmi les techniques d'imagerie on cite la mammographie, échographie, IRM et Pet-scan.

1.9.2.1 La mammographie [8]

Une mammographie est une radiographie des seins. Elle permet d'obtenir des images de l'intérieur du sein à l'aide de rayons X et de détecter ainsi d'éventuelles anomalies. Elle est pratiquée par un radiologue.

Deux radiographies par sein sont réalisées, une de face et une en oblique, ce qui permet de comparer les deux côtés de chaque sein.

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



Figure 1.8. Appareil de mammographie

1.9.2.2 Échographie

L'échographie est une technique d'exploration de l'intérieur du corps basée sur les ultra-sons. Une sonde envoie un faisceau d'ultrasons dans la zone du corps à explorer. Selon la nature des tissus,

ces ondes sonores sont réfléchies avec plus ou moins de puissance. Le traitement de ces échos permet une visualisation des organes observés [9].

Une échographie complète certaines mammographies difficiles à interpréter, en cas de seins denses rendant son analyse très difficile car pouvant masquer de petites lésions ayant la même densité que le tissu mammaire environnant.

Inconvénients : Certains types de tumeurs ne réfléchissent pas ces ultrasons, l'échographie ne permet donc pas de les détecter.

Une échographie est très indirecte et floue, il est souvent ardu pour un médecin de distinguer une anomalie.



Figure 1.9. Exemple d'échographie mammaire [10].

1.9.2.3 IRM (Imagerie par résonance magnétique)

L'imagerie par résonance magnétique, est un moyen d'investigation dans le dépistage et le suivi thérapeutique des cancers. Proposant une visualisation de coupes anatomiques extrêmement fines, elle permet de différencier les tissus pathologiques des tissus sains et de réaliser une "cartographie" très précise et en trois dimensions des tumeurs. L'appareil utilisé pour l'IRM mammaire est représenté par la figure suivante :



Figure 1.10. Appareil utilisé pour IRM mammaire

L'IRM utilise les modifications provoquées par un champ magnétique, envoyé par la machine, dans les noyaux des atomes d'hydrogène présents dans l'organisme. Cet examen est indolore et très fiable.

L'imagerie par résonance magnétique n'utilise pas les rayons X . Parmi toutes les modalités radiologiques, l'IRM est certainement la technique la plus performante dans la recherche des lésions du sein. Rarement, une prise de contraste diffuse du tissu fibro glandulaire peut limiter l'interprétation de l'examen, voire même masquer une lésion. La spécificité de l'IRM est un peu moins bonne. Des petits foyers de prise de contraste sont difficilement caractérisables. Certaines lésions bénignes comme les fibroadénomes peuvent avoir un comportement identique à celui des tumeurs malignes. Cette situation requiert très souvent une biopsie ou une exégèse. On peut rencontrer également des situations où l'on trouve à l'IRM une prise de contraste dont l'aspect est inquiétant, et dans le même temps, ne voir aucune lésion sur les mammographies et à l'échographie (même après réexamen). Il existe alors une indication à effectuer une biopsie sous IRM.

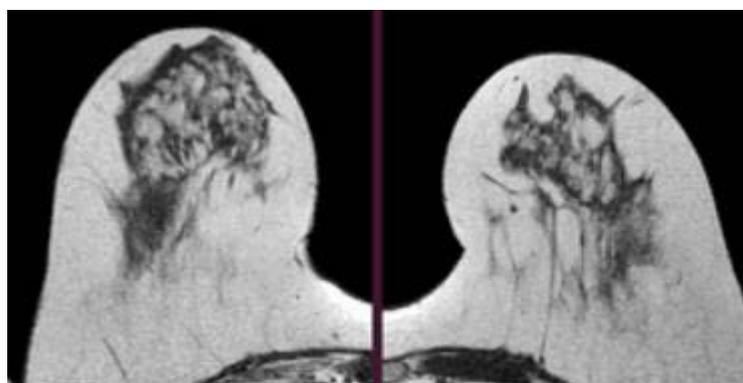


Figure1.11. Exemple d'imagerie IRM mammaire [11].

1.9.2.4 Le pet-scan

Le PET-scan permet de visualiser le fonctionnement des organes. Il consiste à injecter du glucose (un sucre faiblement radioactif) puis à analyser l'image obtenue par un scanner. Les cellules cancéreuses sont identifiables car elles ont une activité plus importante et consomment plus de glucose.

Inconvénients : certaines cellules tumorales ne consomment pas ce glucose, elles ne sont donc pas détectables par pet-scan.

1.9.3 La biopsie

Après avoir constaté une tumeur ou quelque anomalie conséquente en imagerie, un prélèvement du tissu suspecté est effectué pour le faire analyser par un anatomopathologiste.

1.10 Les facteurs de risque

Le cancer du sein est une maladie multifactorielle. Cela signifie que plusieurs facteurs influent sur le risque de sa survenue [12]

Une personne qui possède un ou plusieurs facteurs de risque peut ne jamais développer de cancer. Inversement, il est possible qu'une personne n'ayant aucun facteur de risque soit atteinte de ce cancer.

On distingue :

- **Les facteurs de risque lié à l'âge.** En effet, près de 80% des cancers du sein se développent après 50 ans.
 - **Les facteurs de risque liés à nos modes de vie** tels que la consommation d'alcool et de tabac, un surpoids ou encore pas ou peu d'activité physique peuvent favoriser l'apparition d'un cancer du sein.
 - **Les facteurs de risque liés à certains antécédents médicaux personnels et familiaux.**
Outre le sexe (plus de 99% des cancers du sein touchent les femmes), les quatre principaux facteurs de risque de cancer du sein sont :
 - L'âge ;
 - Les antécédents personnels de maladie (par exemple cancer du sein, de l'ovaire et/ou de l'endomètre) ;
 - Les antécédents familiaux de cancers (sein, ovaire, ...) ;
 - Les prédispositions génétiques au cancer du sein.
- D'autres facteurs ont également été identifiés, ce sont :
- Certains traitements hormonaux de la ménopause

- La consommation de tabac ;
- La consommation d'alcool et le surpoids ;
- Peu ou pas d'activité physique.

1.11. Objectifs et motivations de l'intelligence artificielle

1. Les thérapies utilisées pour traiter le cancer (chimiothérapie, radiothérapie, ablation chirurgicale) sont très lourdes et peuvent porter atteinte au pronostic vital du patient c'est pour quoi, un diagnostic doit être extrêmement sûr et précis.
2. Le principal outil d'un anatomopathologiste est l'œil : l'intelligence artificielle peut imiter son comportement.
3. La grande quantité de données et la complexité des histologies rendent cette tâche fastidieuse et non triviale.
4. Dans certains diagnostics délicats, un anatomopathologiste se fie uniquement à ses expériences, il fait également appel à l'expérience de ses collègues pour des échanges de dossiers, il faut donc exploiter cette notion de déjà-vu. Un ordinateur peut accumuler une mémoire à très grande échelle par rapport à un humain.

1.12. Conclusion

Dans ce chapitre, nous avons défini le cancer du sein, ses symptômes ainsi que les différentes techniques d'imageries nécessaires pour la recherche des lésions mammaires, malheureusement l'identification de ces lésions devient une tâche très difficile.

Dans ce contexte, notre projet consiste à classer le cancer du sein en tumeur bénigne et maligne en utilisant l'intelligence artificielle. Le chapitre suivant sera consacré à l'apprentissage profond.

CHAPITRE 2

Généralités sur

l'apprentissage

automatique

2.1.Introduction :

Dans le cadre de ce deuxième chapitre, nous allons présenter les notions de base sur l'intelligence artificielle, l'apprentissage automatique (Machine Learning) et l'apprentissage profond (Deep Learning) en utilisant les réseaux de neurones artificiels.

2.2 L'intelligence artificielle (AI) :

À l'origine, une AI est simplement un algorithme dont le but est de prendre des décisions relevant d'une certaine forme d'intelligence, c'est une technique qui permet aux machines d'agir comme des humains en reproduisant leur comportement et leur nature : ; telles que la perception visuelle, la reconnaissance de la parole, l'identification des personnes et l'aide à la prise de décisions dans les domaines de transports aériens et terrestres et dans les applications médicales telles que le diagnostic de différentes maladies.

L'intelligence artificielle se base sur deux formes d'algorithmes :

2.2.1 Algorithme à base de règles :

C'est un algorithme qui prend des décisions en fonction d'un ensemble de règles prédéfinies, l'ordinateur n'apprend pas, il se contente d'exécuter les ordres qu'on lui donne ; On parle de l'AI symbolique.

2.2.2 Algorithme à base d'apprentissage :

L'ordinateur calcule la ligne d'action optimale. Il est libre de s'adapter aux circonstances et de déterminer le meilleur plan d'action. Pour atteindre ce degré de jugement, l'ordinateur doit d'abord apprendre à partir d'exemples. On parle de l'apprentissage automatique (*Machine Learning*).

2.3 L'Apprentissage automatique (Machine Learning) :

L'apprentissage automatique ou Machine Learning est un sous ensemble d'intelligence artificielle qui apprend seul grâce à l'analyse d'un jeu de données. Pour cela, il utilise des procédés automatiques issus des statistiques et des probabilités. Son objectif est d'aider les humains à exploiter de grands volumes de données en leur faisant gagner du temps dans des activités simples.

A titre d'exemples : Examiner des images médicales, segmenter une base de clients en marketing et évaluer la probabilité qu'une machine tombe en panne, etc... [13]

Il existe 4 types d'apprentissage automatique (ML) illustrés par la figure 2.1

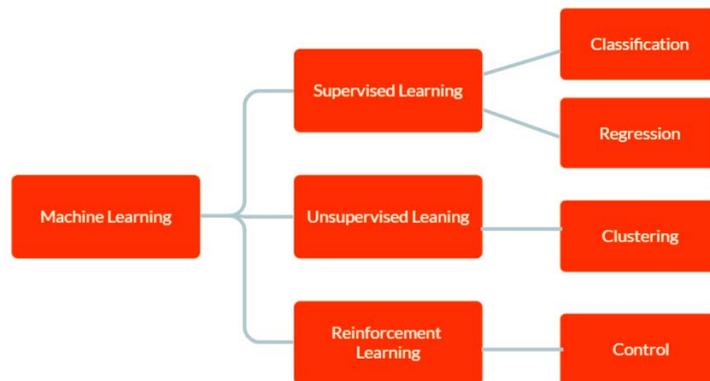


Figure 2.1 : Schéma d'organisation d'apprentissage automatique [14]

2.3.1 Apprentissage automatique (ML) supervisé :

Dans l'apprentissage supervisé, les algorithmes sont formés sur des données étiquetées. C'est-à-dire, des exemples avec des données d'entrées connues, et les sorties correspondantes connues.

L'objectif est d'apprendre la correspondance entre les entrées et les sorties afin que l'algorithme puisse faire des prédictions sur de nouvelles données.

Dans l'apprentissage supervisé on a 2 types d'algorithmes :

- Les algorithmes de régression : pour prédire une valeur continue ou une quantité.
- Les algorithmes de classification : pour prédire une classe ou une catégorie.

Quelques applications de l'apprentissage automatique :

- Diagnostic médical et étude de clientèle.
- Reconnaissance vocale, vision par ordinateur.
- Détection de spam dans les e-mails, détection de fraude, la robotique [15] etc...

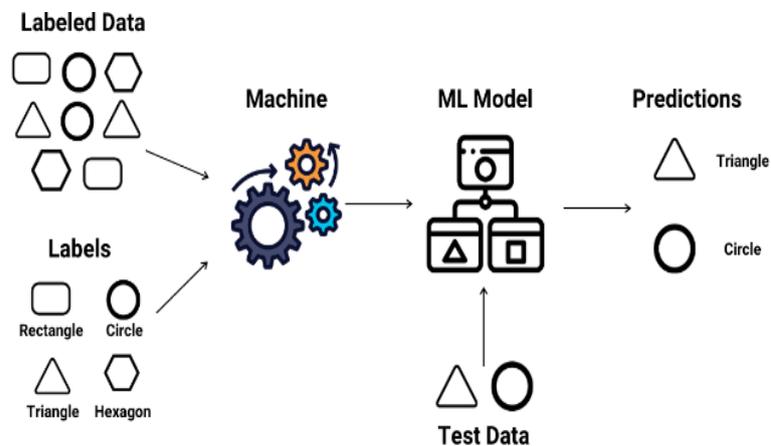


Figure 2.2 : Diagramme de l'apprentissage supervisé

2.3.2 Apprentissage automatique (ML) non supervisé :

L'apprentissage non supervisé ou Unsupervised Machine Learning consiste à faire apprendre à un algorithme d'intelligence artificielle, des informations non traitées au préalable et dont sa réaction sera le regroupement de ces données suivant la similitude de chaque groupe. Ce type d'apprentissage est principalement utilisé quand on souhaite étudier un ensemble de données non labellisés.

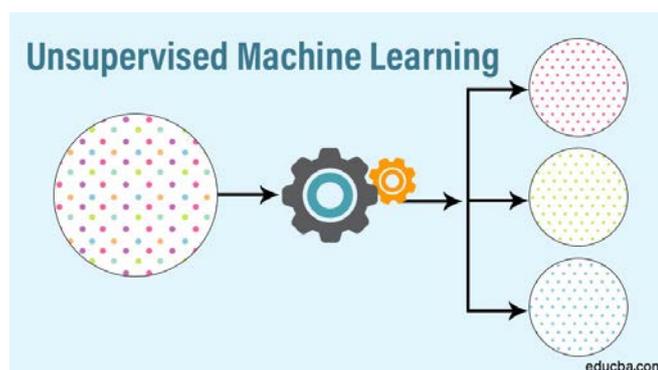


Figure 2.3 : Principe de l'apprentissage non-supervisé [16]

2.3.3 L'apprentissage automatique semi-supervisé :

C'est une approche de l'apprentissage automatique qui combine une petite quantité de données étiquetées avec une grande quantité de données non étiquetées pendant la formation.

Lorsque des données non étiquetées sont utilisées en combinaison avec une petite quantité de données étiquetées, la précision de la formation peut être considérablement améliorée.

Le coût associé au processus d'étiquetage peut donc rendre impossible la réalisation de grands ensembles d'apprentissage entièrement étiquetés, alors que l'acquisition de données non étiquetées est relativement peu coûteuse. Dans de telles situations, l'apprentissage semi supervisé peut être d'une grande valeur pratique.

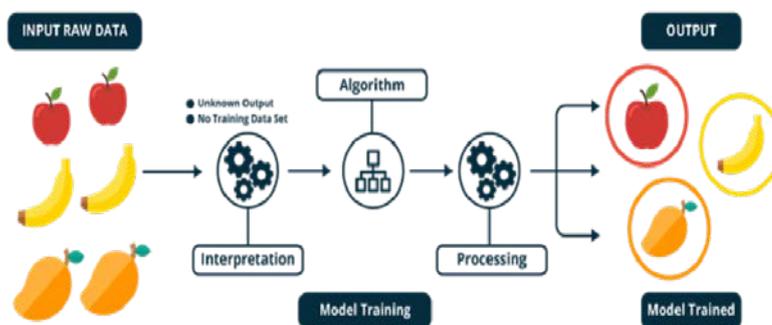


Figure 2.4 : l'apprentissage semi-supervisé[17]

2.3.4 L'apprentissage automatique par renforcement :

L'apprentissage par renforcement consiste à entraîner des modèles de l'IA d'une manière très spéciale. Dans un environnement complexe et inconnu, le modèle IA doit apprendre à réaliser un objectif en essayant toutes les solutions possibles, et même en bénéficiant des erreurs qu'il fait.

Le concepteur des modèles met en place les conditions des récompenses et des pénalités, c'est-à-dire fixer les règles du jeu.

Donc, le modèle d'apprentissage par renforcement peut effectuer de milliers de tentatives en parallèle, pour trouver la solution recherchée d'une manière aléatoire, et d'où la récompense.

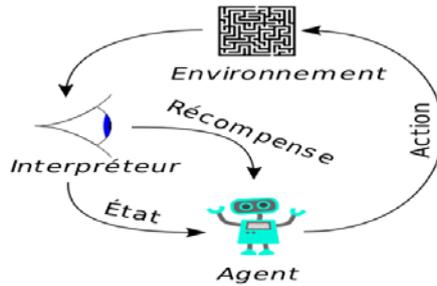


Figure 2.5: Apprentissage par renforcement[18]

Tableau 2 .1 : Différences entre les types d'apprentissage automatique [19]

	Supervisé	Non-supervisé	Renforcement	Semi-supervisé
Bases de données	Les données sont étiquetées	Les données ne sont pas étiquetées	Ne nécessite pas de données car il apprend en interagissant avec l'environnement	Combine l'apprentissage supervisé et non supervisé
Apprentissage	Basé sur les instructions et Apprend-en suivant les exemples de l'ensemble des données étiquetées.	Apprend par le raisonnement et tente d'identifier des modèles cachés afin d'extraire des informations de l'ensemble de données non étiquetées	Apprend par l'expérience en explorant et en exploitant différentes actions pour maximiser les récompenses à long terme	Apprend sur un petit nombre de données étiquetées et un grand ensemble de données non étiquetées
Objectifs	La prédiction des résultats pour de nouvelles données appartenant au même domaine	Obtenir des informations à partir des données non étiquetées qui permettent de prédire si les nouvelles données font partie de la classe ou constituent une anomalie	Orienté vers un objectif, où l'agent apprend des séquences d'actions par l'exploration et l'exploitation dans un environnement incertain, afin de maximiser les futures récompenses	Augmenter la précision avec un moindre coût. Permettre à l'être humain de se servir du modèle semi-supervisé d'apprendre.
Fonction-nements par :	Classification Régression	Regroupement, association, détection des anomalies	Basés sur la valeur, ou sur la politique, ou sur le modèle	Un grand nombre de données brut, et peu de données étiquetés.

2.4 L'Apprentissage profond :

L'apprentissage profond, également appelé Deep Learning en anglais, est un sous-domaine de l'apprentissage automatique qui consiste en un ensemble d'algorithmes capables d'imiter le fonctionnement du cerveau humain grâce à des réseaux de neurones artificiels. Ces réseaux sont composés de dizaines voire des centaines de couches de neurones. Plus le nombre de couches est élevé, plus le réseau est qualifié « profond ».

La figure 2.6 montre l'imbrication entre l'intelligence artificielle, le machine Learning et le Deep Learning [20] .

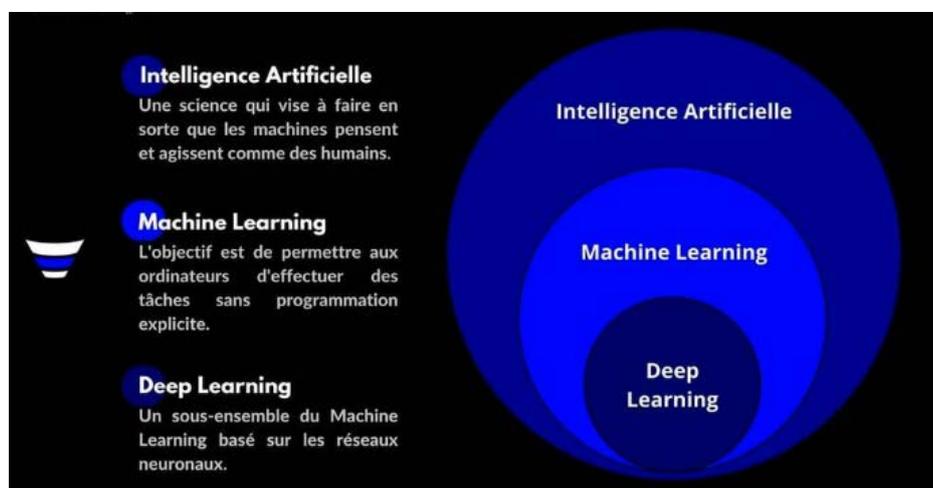


Figure 2.6:Artificial Intelligence vs Machine Learning vs Deep Learning

L'apprentissage profond est la technique la plus puissante dans le machine Learning, notamment dans la classification des images. Elle est utilisée dans :

- Le traitement du langage
- La reconnaissance d'image
- Le diagnostic médical
- Les robots

Parmi les algorithmes d'apprentissage profond les plus importants on cite [21] :

- Réseaux adversariaux génératifs (GAN)
- Réseaux neuronaux récurrent (RNN)
- Réseaux de fonction à base radiale (RBFN)
- Réseaux neuronaux convolutif (CNN) .

2.5 Les réseaux de neurones

2.5.1. Neurone biologique

L'unité de base du système nerveux est la cellule nerveuse ou neurone. Le neurone est constitué d'un grand corps cellulaire et de fibres nerveuses, l'axone et les dendrites. Le neurone biologique reçoit en entrées des signaux électriques (influx nerveux) transmis par d'autres neurones, grâce à une interaction qui se fait entre les dendrites et les synapses. Le neurone reçoit les signaux en entrée, les analyse et les traite en effectuant leur sommation, si le résultat est supérieur à un seuil d'activation, il envoie une décharge le long de son axone vers d'autres neurones. La synapse est l'espace d'interaction entre deux cellules nerveuses qui permet le passage du signal (Figure 2.7)

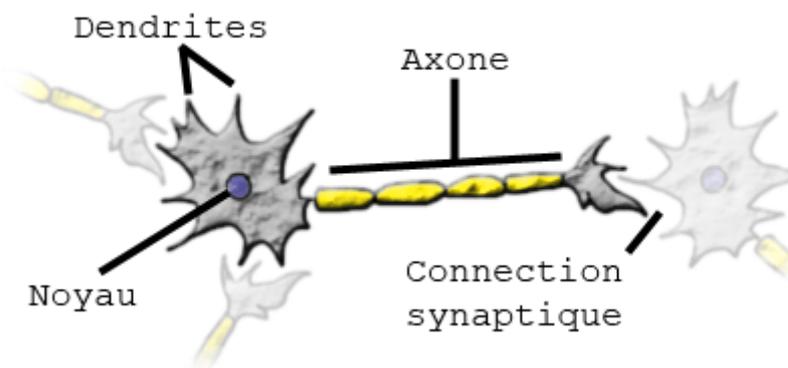


Fig . 2.7 : Neurone Biologique

2.5.2. Réseaux de Neurones Artificiels

Un Réseau de Neurones Artificiels (Artificial Neural Network) est formé d'un grand nombre de processeurs simples (neurones) qui fonctionnent en parallèle et fortement interconnectés inspiré des cellules nerveuses du cerveau humain, Dans le neurone artificiel, les poids synaptiques, fonction d'activation et élément de sortie ont des fonctions similaires aux fonctions occupées par les synapses, corps cellulaire et axone du neurone biologique.

Le neurone artificiel prend en entrée un certain nombre de variables $x = \{x_1, x_2, x_3, \dots, x_n\}$ et chaque entrée est associée à un poids (w_i) représentant la valeur de la connexion puis en ajoutant un biais (b_i). Une fonction d'activation f transforme cette somme pondérée à une valeur qui sera ensuite transmise à la couche de sortie pour être comparée à une valeur du seuil et détermine si un neurone doit être activé ou non, puis fournir une réponse non linéaire à la sortie (Figure 2.8)

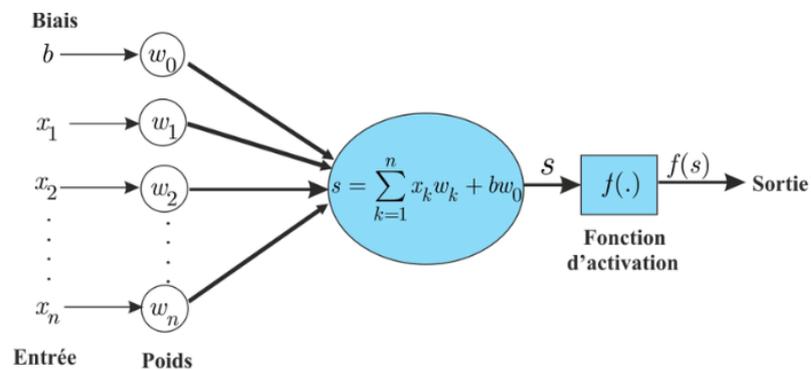


Figure .2.8: Réseaux de Neurones Artificiels

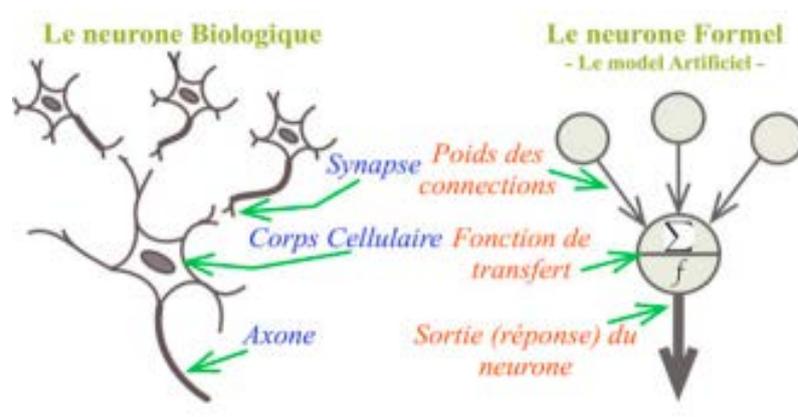


Figure. 2.9: La similitude entre le neurone artificiel et le neurone biologique

➤ Les Compositions d'un réseau neuronal

Les nœuds sont connectés entre eux en plusieurs couches, ils déplacent les données à travers le réseau neuronal dans une seule direction. il est organisé en trois couches (figure 2.10) :

- Couche d'entrée (Input layer) :

Cette couche reçoit les données, aucun calcul n'est effectué. Les nœuds ici transmettent simplement les informations à la couche cachée.

- Couches cachées (Hidden layer) :

Les nœuds de cette couche ne sont pas exposés au monde extérieur, la couche cachée effectue toutes sortes de calculs sur les entités saisies via la couche d'entrée et transfère le résultat à la couche de sortie.

- Couche de sortie :

Cette couche fait remonter les informations apprises par le réseau vers le monde extérieur.

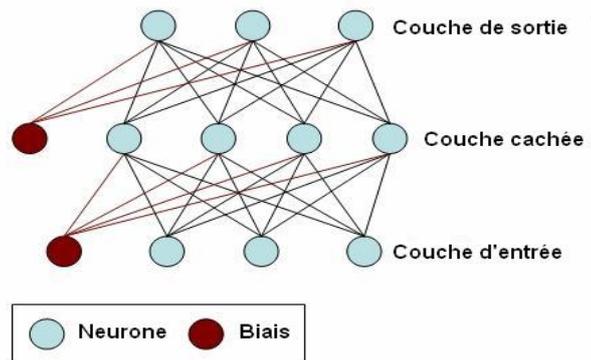


Figure 2.10. les réseaux de neurones

2.6 Les réseaux de neurones convolutionnels

Au cours des dernières décennies, l'apprentissage profond s'est révélé être un outil très puissant en raison de sa capacité à traiter de grandes quantités de données. L'intérêt d'utiliser des couches cachées a dépassé les techniques traditionnelles, en particulier dans le domaine de la reconnaissance des formes. L'un des réseaux neuronaux profonds les plus populaires est le réseau neuronal convolutif (également connu sous le nom de CNN ou ConvNet) dans le domaine de l'apprentissage profond, en particulier dans les applications de vision par ordinateur.

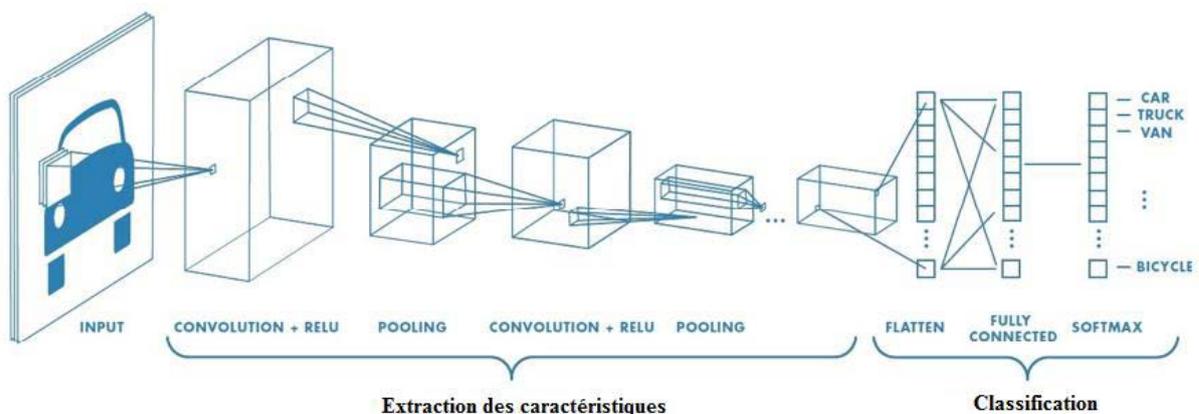


Figure 2.11: exemple d'une architecture CNN

2.7 Les différentes couches des réseaux de neurones convolutifs (CNN)

Les CNNs sont généralement formés de plusieurs couches de convolution et de pooling, suivies par une couche fully-connected qui combine les features extraites par les couches précédentes pour classifier l'image

2.7.1. Couche de convolution :

La couche de convolution est la couche de base d'un CNN, c'est là où la majorité des calculs sont effectués. Son but est de vérifier la présence d'un ensemble de caractéristiques (features) dans les images reçues en entrée et utilise pour cela des filtres.

2.7.1.a .Les filtres

Dans la pratique, les filtres d'un réseau de neurones convolutifs ont ;

- Un kernel qui représente la taille de la matrice du filtre souvent en 3×3 ,
- Un stride qui représente le décalage du kernel entre chaque calcul
- Un padding ajoute des 0 autour de l'image d'entrée, pour augmenter la taille de l'image de sortie ou éviter les dépassements.

Les valeurs des filtres sont générées aléatoirement à l'initialisation. Ensuite, lorsque le réseau apprend, les valeurs des filtres sont mises à jour pour améliorer les résultats du CNN.

2.7.1. b. Opération de convolution

L'opération de convolution consiste à faire glisser un filtre à travers toute l'image et de calculer le produit de convolution entre le filtre et chaque proportion de l'image. Une image est perçue par l'ordinateur comme une matrice de pixels avec des valeurs différentes. Une image en couleur est composée d'une matrice de pixels en 3D. Cela signifie que l'entrée aura trois dimensions (hauteur, largeur et profondeur) qui correspondent au RVB dans une image .

La figure 2.12 montre comment un filtre de taille 3×3 est appliqué sur une image de taille 5×5 . Les 3×3 premiers pixels de l'image sont sélectionnés et chaque valeur de ces pixels est multipliée par la valeur de pixel correspondant du filtre pour créer un nouveau pixel dans l'image de sortie (carte de convolution). La valeur de ce dernier est égale à la somme des résultats de multiplications. Le filtre se déplace ensuite vers la droite avec un pas, également appelé «stride», pour que la convolution soit appliquée sur la deuxième proportion et ainsi de suite. Lorsque la fin de la première ligne est atteinte, le filtre descend d'un pas et à chaque proportion rencontrée de l'image, le même calcul s'effectue jusqu'à arriver au bout de l'image.

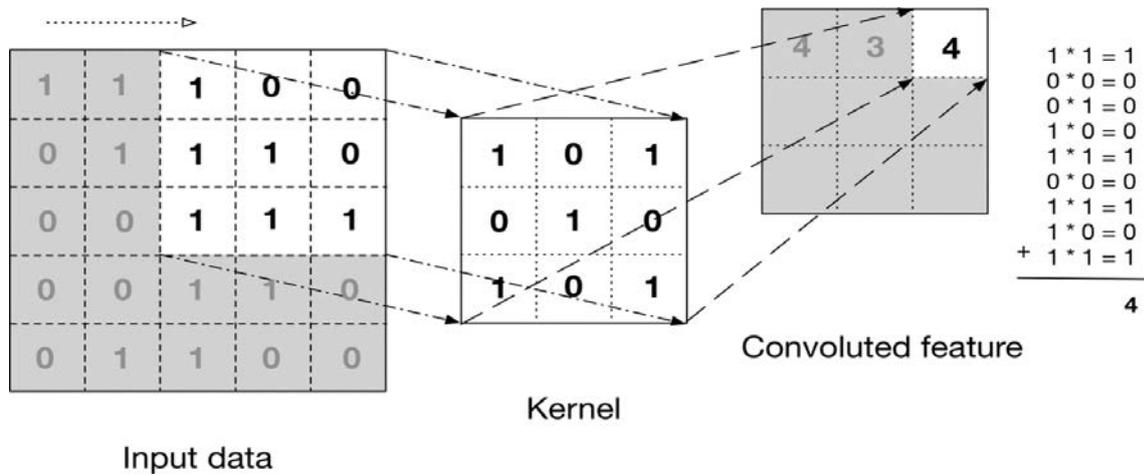


Figure. 2.12 : Opération de convolution [22]

2.7.2. La couche de correction ReLU (Rectified Linear Units) :

Une étape supplémentaire qui vient s'ajouter à la convolution, cette couche remplace toutes les valeurs négatives reçues en entrées par des zéros. L'intérêt de ces couches d'activation est de rendre le modèle non linéaire et de ce fait plus complexe [23] .

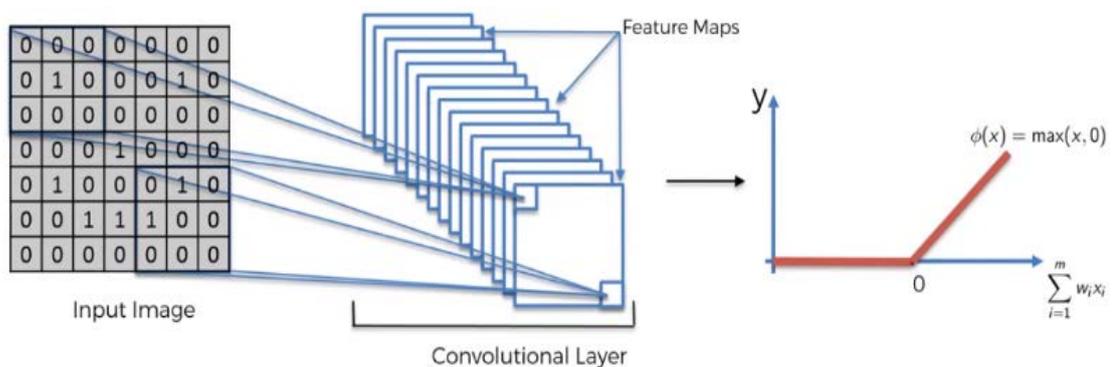


Figure 2.13 : La couche d'activation ReLU

2.7.3. cross channel normalization :

Une couche cross channel normalization effectue une normalisation par canal. Cette couche effectue une normalisation de réponse locale par canal. Il suit généralement la couche d'activation ReLU. Cette couche remplace chaque élément par une valeur normalisée obtenue à l'aide des éléments d'un certain nombre de canaux voisins (éléments dans la fenêtre de normalisation).

2.7.4. Couche de pooling [24] :

Le Pooling est un algorithme visant à réduire la dimension d'une image tout en gardant des sous-parties importantes. La fenêtre de Pooling est une matrice $m \times m$, qui parcourt l'image afin de produire un seul pixel de sortie parmi $m \times m$ pixels voisins. Plusieurs méthodes de production de pixels sont mises en place en fonction de la pertinence de l'information voulue. Parmi ces méthodes, on compte l'average-Pooling, max-Pooling et min-Pooling.

2.7.4.a. Average-Pooling

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

2.7.4.b. Max-Pooling

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

2.7.4.c. Min-Pooling

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

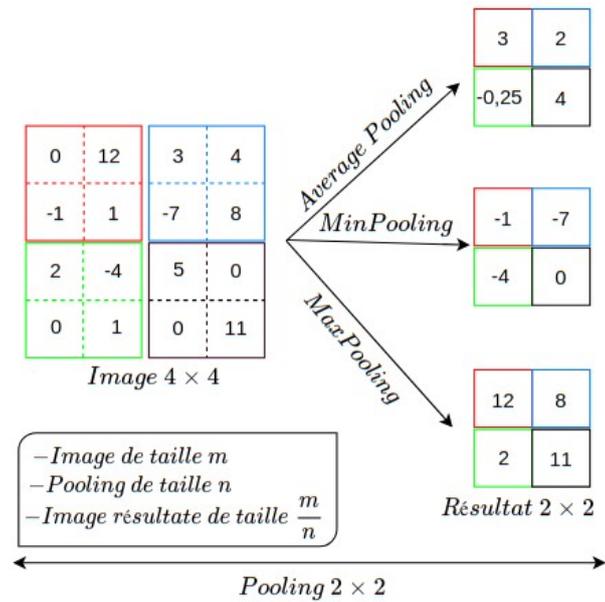


Figure 2.14 . Exemple applicatif des différents Pooling

2.7.5. Le flattening (ou mise à plat):

Dernière étape de la partie « extraction des informations », le flattening consiste simplement à mettre bout à bout toutes les images (matrices) que nous avons pour en faire un (long) vecteur. Les pixels (en réalité ce ne sont plus des images ou des pixels, mais des matrices de nombres, donc les pixels sont ces nombres) sont récupérés ligne par ligne et ajoutés au vecteur final [25] .

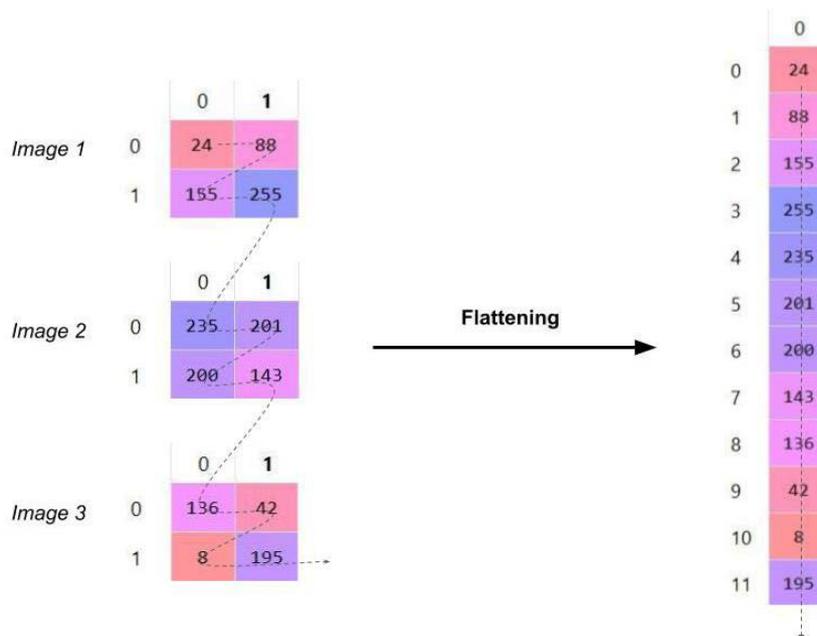


Figure 2.15 : Le flattening

2.7.6. Couches entièrement connectées (fully-connected FC) :

La couche fully connected constitue toujours la dernière couche d'un réseau de neurones, convolutif ou non, elle n'est donc pas caractéristique d'un CNN. Ce type de couche reçoit un vecteur en entrée et produit un nouveau vecteur en sortie pour cela, elle applique une combinaison linéaire puis éventuellement une fonction d'activation aux valeurs reçues en entrée. Elle permet de classifier l'image en entrée du réseau : elle renvoie un vecteur de taille N, où N est le nombre de classes dans le problème de classification d'images. Chaque élément du vecteur indique la probabilité pour l'image en entrée d'appartenir à une classe [26] .

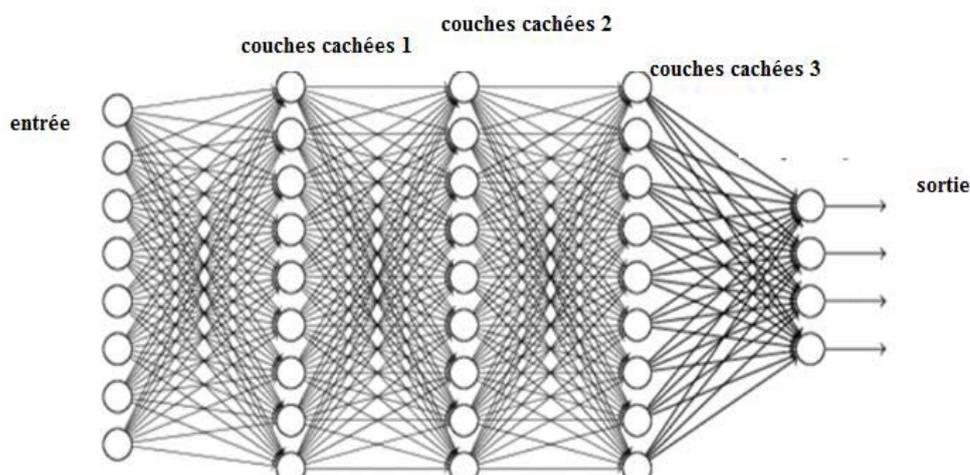


Figure.2.16 : Exemple de couches fully-connected

2.7.7. La couche dropout :

Le Dropout est une technique efficace pour éviter l'overfitting sur les données d'entraînement, ce terme permet la désactivation temporaire des neurones inactifs qui ne contribuent pas dans la propagation avant (forward pass) et la retropropagation (backward pass) pendant une itération de données de l'entraînement.

2.7.8. Couche Prob «softmax» :

La couche Softmax est mise en oeuvre via une couche de réseau de neurones juste avant la couche du résultat « output ». Permet de calculer la distribution de probabilités sur les classes de sortie. La couche Softmax doit comporter le même nombre de noeuds que la couche du résultat.

On Crée une couche softmax à l'aide de la fonction softmax Layer après la dernière couche entièrement connectée. La fonction Softmax est définie par :

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad (2.1)$$

σ : softmax

Z : vecteur d'entrée

K : nombre de classes dans la classification en classes multiples

e^{z_i} : fonction exponentielle normalisée du vecteur d'entrée

e^{z_j} : fonction exponentielle normalisée du vecteur de sortie

2.8 Fonctionnement des CNN

Dans les CNN nous distinguons deux parties :

- **La partie convolutive du modèle :**

Cette partie est la particularité de ce type de réseau. C'est la partie responsable de l'extraction des caractéristiques des images. Chaque image passe à travers une série de filtres , appelés aussi «noyaux de convolution», créant de nouvelles images appelées cartes de convolutions ou feature maps. Certain de ces filtres réduisent la taille initiale de l'image. Au final, les dernières cartes de convolutions obtenues sont concaténées dans un vecteur de caractéristiques qui sera ensuite fournie à l'entrée de la partie classification.

- **La partie classification du modèle :**

Cette partie n'est pas caractéristique de CNN, elle est présente à la fin de tous les réseaux de neurones pour la classification. Elle est composée de couches entièrement connectées (perceptron multicouche). Elle prend en entrée un vecteur, combine ses caractéristiques et renvoie un nouveau vecteur en sortie. Ce vecteur contient un élément par classe. Chaque élément est compris entre 0 et 1 et représente la probabilité d'une des classes. La dernière couche (vue précédemment) est celle qui calcule la probabilité de chaque catégorie et utilise comme fonction d'activation, une fonction logistique quand il s'agit d'une classification binaire ou une fonction softmax quand il s'agit d'une classification multi-classe.

2.9. Quelques architectures des CNNs

Les architectures CNN constituent le cadre d'apprentissage profond le plus populaire. Les CNN sont utilisés pour une variété d'applications, allant de la vision par ordinateur au traitement du langage naturel. Voici une liste des différents types d'architectures de CNN

2.9.1. AlexNet

AlexNet est un réseau neuronal convolutif de 8 couches de profondeur dont cinq sont des couches de convolution pour l'extraction des caractéristiques de l'image et trois sont des couches entièrement connectées pour la classification de cette image, il a été formé sur plus d'un million d'images à partir de la base de données ImageNet. Le réseau pré-entraîné peut classer les images en 1000 catégories d'objets. En conséquence, le réseau a appris des représentations de caractéristiques riches pour une large gamme d'images.

AlexNet utilise deux techniques pour l'accélération de l'apprentissage par retro propagation et éviter le sur-apprentissage : ReLu ou Rectified Linear Unit comme fonction d'activation des neurones et Local Response Normalization pour la normalisation.

L'entrée de ce modèle est constituée d'images de taille $227 \times 227 \times 3$.

La figure 2.17 montre l'architecture du réseau Alexnet

- 05 couches de convolution; les 1ère, 2ème et 5ème ayant des couches Max-Pooling.
- 02 couches entièrement connectées (FC) et une couche softmax à la fin pour la prédiction.

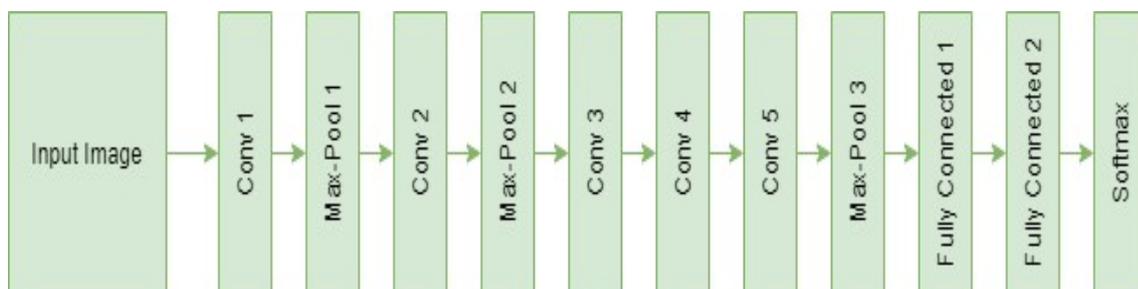


Figure 2.17 : Architecture AlexNet [27].

Tableau 2.2 : Tableau récapitulatif de l'architecture Alexnet

Couche	Entrée	Filtres (nombre – dimensions – pas) ou Nombres de sortie
Première	Image de $224 \times 224 \times 3$	96- $11 \times 11 \times 3$ -4
Seconde	Sortie 1-ère couche + Normalisation et pooling	256 - $5 \times 5 \times 48$
Troisième	Sortie 2-nde couche + Normalisation et pooling	384 - $3 \times 3 \times 256$
Quatrième	Sortie 3-ème couche	384 - $3 \times 3 \times 192$
Cinquième	Sortie 4-ème couche	256 - $3 \times 3 \times 192$
3 couches interconnectées : FC6 - FC7 - FC8	Sortie 5-ème couche + Normalization et pooling	4096
Une couche de sortie pour les 1000 classes	Sortie fully connected layer + Softmax	1000

2.9.2. Googlenet

GoogLeNet est l'architecture CNN utilisée par Google pour remporter la tâche de classification ILSVRC 2014. Elle a été développée par Jeff Dean, Christian Szegedy, Alexandro Szegedy et al. L'architecture GoogLeNet comporte 22 couches, dont 27 couches de mise en commun. Il y a 9 modules de départ empilés linéairement au total. Les extrémités des modules de départ sont connectées à la couche de mise en commun de la moyenne globale. L'architecture est plus profonde grâce à l'utilisation d'un certain nombre de techniques distinctes, notamment la convolution 1×1 et la mise en commun de la moyenne globale. L'architecture CNN de GoogLeNet est coûteuse en termes de calcul.

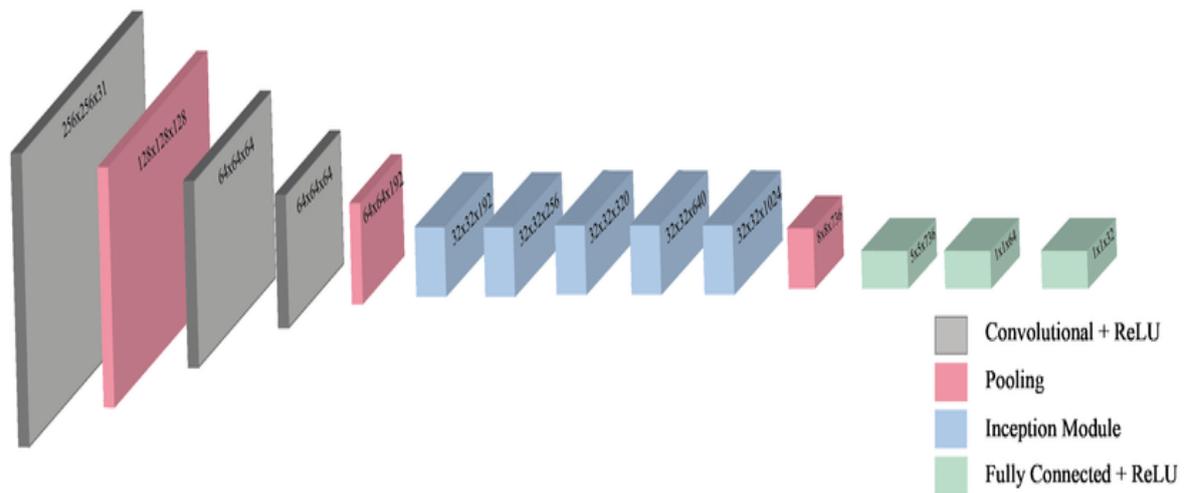


Figure 2.18 : diagramme général du googlenet [28].

2.9.3 Inceptionv3

InceptionV3 est un réseau de neurones convolutifs pré-formé de 48 couches de profondeur, qui est une version du réseau déjà formé sur plus d'un million d'images de la base de données ImageNet. Ce réseau pré-entraîné classe aussi les images en 1000 catégories d'objets, Le réseau a une taille d'entrée d'image de 299 par 299. Le modèle extrait les caractéristiques générales des images d'entrée dans la première partie et les classe en fonction de ces caractéristiques dans la deuxième partie.

2.9.3.1 Architecture du modèle Inceptionv3

L'architecture convolutionnelle profonde d' Inception, a été introduite sous le nom de Google Net (Inceptionv1). Plus tard, cette architecture a été affinée de diverses manières, d'abord par l'introduction de la normalisation par lot (Inceptionv2). Ensuite par des idées de factorisation supplémentaires dans la troisième itération, qui est appelé Inceptionv3.

Le modèle inceptionv3 est constitué de composants de base symétriques et asymétriques ; incluant les opérations de convolutions, de pooling moyen, de pooling maximale de concaténation, d'abondons et de couches entièrement connectées. La normalisation par lots (batchnorm), est amplement utilisé dans le modèle et appliquée aux entrées d'activation. La perte est calculée via la fonction Softmax.

2.9.3.2 Factorisation des convolutions

La factorisation des convolutions a pour but de réduire le nombre de paramètres sans diminuer l'efficacité du réseau.

- Factorisation en petites convolutions

L'un des atouts majeurs du modèle Inceptionv1 était la réduction généreuse des dimensions. Il est composé de quatre couches parallèles, les différentes tailles de convolutions sont effectuées pour capturer différentes tailles d'informations dans l'image.

Pour rendre le modèle Inceptionv1 encore meilleur, les plus grandes convolutions du modèle ont été factorisées en plus petites convolutions. Par exemple, considérons le module de base du module inceptionv1 représenté par la figure 2.19

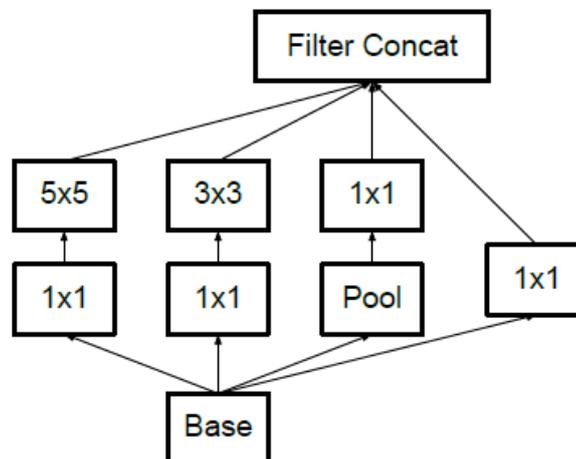


Figure 2.19 : module de base de inceptionv1

la couche convolutive 5×5 (grandes convolution) coûteuse en calcul, a été remplacée par deux couches convolutives 3×3 (petite convolution), la figure 2.20 montre la construction du nouveau modèle 'Inception module A'.

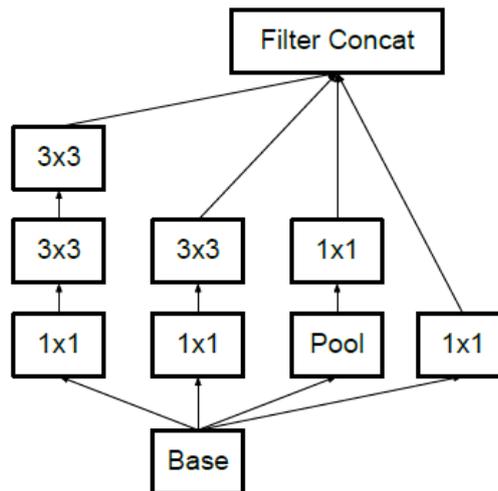


Figure 2.20 :Inception module A

Deux convolutions 3×3 réduit le nombre de paramètres. En utilisant une seule couche du filtre 5×5 , le nombre de paramètres est égale à $5 \times 5 = 25$ alors qu'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 à 28%.

- **Factorisation en convolutions asymétriques**

Une meilleure alternative pour rendre le modèle plus efficace était les convolutions asymétriques.

Les convolutions asymétriques sont de la forme $n \times 1$. Ceci consiste à remplacer les convolutions 3×3 par une convolution 1×3 suivie d'une convolution 3×1 . En utilisant un filtre 3×3 , le nombre de paramètres est $3 \times 3 = 9$.

En utilisant les filtres 3×1 et 1×3 , le nombre de paramètres est $3 \times 1 + 1 \times 3 = 6$. Il est ainsi réduit de 33% avec cette technique, l'un des nouveaux modules Inception 'Module B' est construit selon la figure 2.21

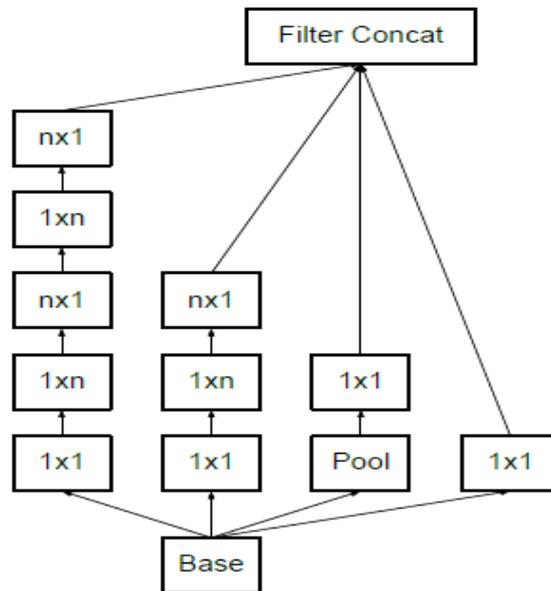


Figure. 2.21 : Inception Module B

Après avoir appliqué les deux premières techniques d'optimisation, le module C de démarrage est représenté par la figure 2.22

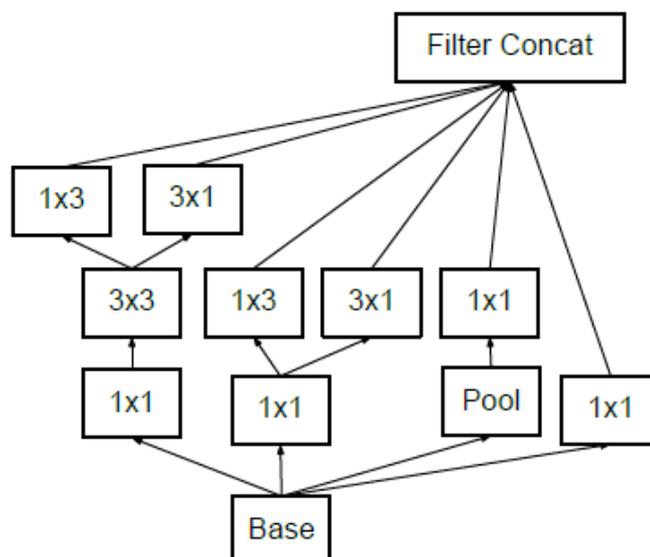


Figure 2.22 : Inception module C

Les auteurs de cette conception, proposent ces trois 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. Le cout de calcul est seulement environ 2.5 plus élevé que celui de Google Net.

Après avoir effectué toutes les optimisations, le modèle final Inception v3 est représenté par la figure 2.23

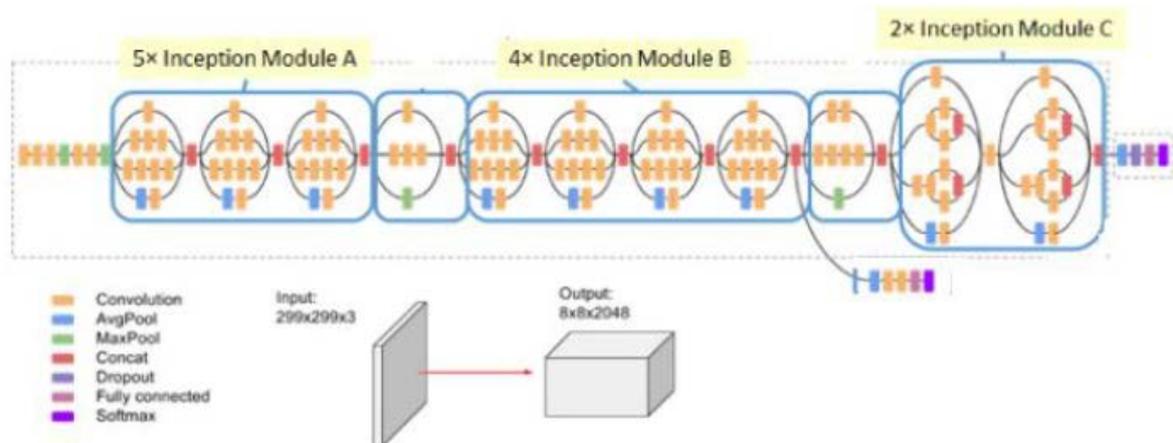


Figure. 2.23 : Architecture du réseau Inceptionv3

Il existe bien d'autres modèles très utilisés comme le modèle Xception [29] , VGG-16 [30] , Resnet50 [31] etc.....

2.10. Les options d'apprentissage :[32]

Les optimiseurs sont des algorithmes [ANNEXE] ou des méthodes modernes utilisées pour modifier les attributs du réseau de neurones tels que les poids et le taux d'apprentissage afin de minimiser les pertes.

2.10.1. Sgdm (StochasticGradient Descent with Momentum) : L'idée de base derrière l'approximation stochastique remonte à l'algorithme de Robbins - Monro des années 1950, la descente de gradient stochastique est devenue une méthode d'optimisation importante dans l'apprentissage automatique. Cet optimiseur met à jour les paramètres du réseau (poids et biais) pour minimiser la fonction de perte. Le facteur d'impulsion supplémentaire permet de réduire l'oscillation qui peut apparaître le long de la trajectoire de descente la plus abrupte vers l'optimum.

L'ajout d'un terme de Momentum à la mise à jour des paramètres est un moyen de réduire cette oscillation . Cette mise à jour est la suivante :

$$\theta_{l+1} = \theta_l - \alpha \nabla E(\theta_l) + \gamma (\theta_l - \theta_{l-1}) \quad (2.2)$$

Où γ détermine la contribution de l'étape de gradient précédente à l'itération actuelle. Vous pouvez spécifier cette valeur à l'aide de l'option de formation Momentum. Pour former un réseau neuronal à l'aide de l'algorithme de descente de gradient stochastique avec Momentum, spécifiez « Sgdm » comme premier argument d'entrée de training options et pour spécifier la valeur initiale du taux d'apprentissage α , utilisez l'option de formation Initial Learn-Rate. Cet algorithme utilise un seul taux d'apprentissage pour tous les paramètres. Il peut également obtenir de nouveaux minima.

2.10.2. AdaGrad :

L'algorithme AdaGrad ajuste individuellement le taux d'apprentissage de tous les paramètres du modèle par mise à l'échelle inversement proportionnelle à la racine carrée de la somme de toutes les valeurs quadratiques historiques du gradient. Le taux d'apprentissage du paramètre avec la plus grande dérivée partielle de la perte diminuera en conséquence rapidement, tandis que la baisse du taux d'apprentissage du paramètre avec la plus petite dérivée partielle sera relativement faible. Dans le contexte de l'optimisation convexe, l'algorithme AdaGrad présente des caractéristiques théoriques idéales. Cependant, empiriquement, pour la formation de modèles de réseaux neuronaux profonds, l'accumulation de gradients carrés dans les premiers stades de la formation entraînera une réduction prématurée et excessive du taux d'apprentissage effectif. AdaGrad convient à certains modèles d'apprentissage en profondeur, mais pas à tous.

2.10.3. RMSprop (Root Mean Square Propagation):

L'algorithme RMSprop modifie AdaGrad pour obtenir de meilleures performances dans un environnement non convexe en changeant l'accumulation de gradient en une moyenne mobile pondérée de manière exponentielle. AdaGrad est conçu pour converger rapidement lorsqu'il est appliqué à une fonction convexe. Lorsqu'elle est appliquée à une fonction non convexe pour entraîner un réseau neuronal, la trajectoire d'apprentissage peut traverser de nombreuses structures différentes et éventuellement atteindre une région qui est un bol localement convexe. AdaGrad réduit le taux d'apprentissage en fonction de toute l'histoire du gradient quadratique et peut avoir rendu le taux d'apprentissage trop faible avant d'atteindre une telle structure convexe. RMSprop utilise un moyen en décroissance exponentielle pour éliminer l'histoire du passé extrême afin qu'il puisse rapidement converger vers la fin d'un bol convexe comme s'il s'agissait d'une instance de l'algorithme AdaGrad initié dans ce bol.

2.10.4. Adam (adaptive moment estimation).

C'est un algorithme d'optimisation qui peut être utilisé à la place de procédure classique de descente de gradient stochastique pour mettre à jour les poids de réseau itératifs en fonction des données d'apprentissage. Adam décrit la combinaison des avantages des deux extensions de la descente de gradient stochastique spécifiquement (AdaGrad) et (RMSprop). Adam utilise également la moyenne des seconds moments des gradients (la variance non centrée). C'est un algorithme populaire dans le domaine de l'apprentissage profond car il obtient rapidement de bons résultats. Adam est relativement facile à configurer les ou les paramètres de configuration par défaut fonctionnent bien sûr la plupart des problèmes. Lors de l'introduction de l'algorithme, certains avantages intéressants de l'utilisation d'Adam sur des problèmes d'optimisation sont indiqués comme suit :

- Simple à mettre en œuvre.
- Efficace en termes de calcul.
- Peu de besoins en mémoire.
- Bien adapté aux problèmes importants en termes de données et / ou de paramètres.
- Convient aux objectifs non stationnaires.
- Convient aux problèmes avec des gradients très bruyants / ou clairsemés
- Invariant au redimensionnement diagonal des dégradés.
- Les hyper-paramètres ont une interprétation intuitive et nécessitent généralement peu de réglages. Concernant cette option d'apprentissage, cet optimiseur conserve une moyenne mobile par élément des gradients de paramètres et de leur valeur au carré,

$$M_\ell = \beta_1 m_{\ell-1} + (1-\beta_1) \nabla E(\theta_\ell) \quad (2.3)$$

$$V_\ell = \beta_2 v_{\ell-1} + (1-\beta_2) [\nabla E(\theta_\ell)]^2 \quad (2.4)$$

On peut spécifier les taux de décroissance β_1 et β_2 à l'aide des options d'entraînement Gradient Decay factor respectivement. Adam utilise les moyennes mobiles pour mettre à jour les paramètres du réseau comme :

$$\theta = \theta - \alpha \mathbf{m} / \sqrt{\mathbf{v}} + \epsilon \quad (2.5)$$

C'est une méthode qui calcule les taux d'apprentissage adaptatifs pour chaque paramètre et stocke une moyenne à décroissance exponentielle des gradients carrés. Cet Algorithme conserve également une moyenne des gradients passés qui décroît de façon exponentielle (m_t) semblable au Momentum (M_t) et (V_t). Si les gradients sur de nombreuses itérations sont similaires,

l'utilisation d'une moyenne mobile du gradient permet aux mises à jour des paramètres de prendre de l'élan dans une certaine direction. Si les gradients contiennent principalement du bruit, alors la moyenne mobile du gradient devient plus petite, et donc les mises à jour des paramètres deviennent également plus petites. La mise à jour complète d'Adam comprend également un mécanisme permettant de corriger un biais qui apparaît au début de la formation. Cet optimiseur utilise également le concept de moment adaptatif en ajoutant des fractions de gradients précédents au gradient actuel. Enfin, Adam est devenu assez répandu, il est pratiquement accepté pour être utilisé dans les réseaux neuronaux d'entraînement et spécifié comme premier argument d'entrée.

2.11 Conclusion

Dans ce chapitre, nous avons introduit les concepts de base du machine learning ainsi que ceux de l'apprentissage profond. Nous avons présenté les réseaux de neurones convolutionnels, nous avons introduit les deux modèles Inceptionv3 et AlexNet que nous allons utiliser pour la classification de nos images. A la fin nous avons abordé les options d'apprentissage utilisant différents algorithmes (Adam, Sgdm et RMSprop).

Le prochain chapitre sera consacré à l'implémentation des deux modèles suivant l'apprentissage par transfert ainsi que les résultats expérimentaux de classification des tumeurs bénignes et malignes du sein obtenus en terme de performance.

CHAPITRE 3

Implémentation et Résultats

3.1.Introduction

Notre travail consiste à classer les images du cancer du sein en deux classes afin de pouvoir faciliter leur identification (tumeur maligne ou bénigne) et ce en se basant sur les réseaux de neurones convolutifs profonds. Les modèles d'apprentissage en profondeur près entraînés Inceptionv3 et Alexnet sont utilisés pour classer ces images.

Ce chapitre contiendra donc, les outils logiciels et matériels qui nous ont permis d'implémenter notre approche, ensuite nous décrivons notre base de données utilisée pour tester et évaluer notre travail et nous présentons l'architecture des réseaux CNN utilisés dans l'implémentation en se basant sur le Transfer Learning (apprentissage par transfert) ainsi que les indicateurs de performance que nous avons utilisé.

Enfin de ce chapitre, nous allons présenter et discuter des résultats de classification obtenus avec l'application de différentes méthodes d'optimisation.

3.2. Les outils matériels et logiciel

3.2.1. Configuration matérielle utilisée dans l'implémentation :

Lors de notre travail, nous avons utilisé un PC portable dont les caractéristiques sont les suivantes :

- Un PC portable lenovoT430 CPU 2.60 GHZ.
- RAM de taille 8 GO
- Disque dur de taille 224 GO SSD
- Système d'exploitation Windows 10 professionnel (64 bits).
- Processeur: Intel(R) core(TM) i5-3320M

3.2.2. Logiciel utilisé dans l'implémentation

Pour mettre en œuvre notre application, nous avons utilisé le langage Matlab R2020a. MATLAB est un environnement interactive qui permet de résoudre d'énorme problèmes scientifiques et d'ingénierie, il a un grand ensemble de bibliothèque et des boites à outils prédéfinis, il peut importer et exporter des diverses données, qualifier de faire des opérations mathématiques très puissantes, ainsi il peut intégrer des environnements extérieurs facilement, il peut exécuter des algorithmes complexes...etc.

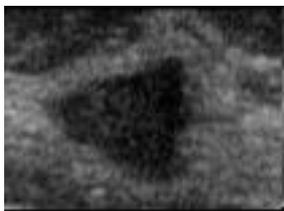
MATLAB est un outil solide dans le domaine d'apprentissage profond, il nous permet de créer des modèles avec un simple code, implémenter des modèles près entraînés, ainsi il a une

flexibilité de changer facilement des paramètres et de remplacer des différentes couches. Il nous permet la visualisation des résultats et il est capable d'atteindre de hautes précisions

3.3. Description de la base de données

Nous avons choisi la base de données Mendelej DATA-Breast ultrasound qui a été publiée le 1^{er} janvier 2018, avec la collaboration de Paulo Sergio Rodrigues [33]. Cette base de données contient un ensemble d'images IRM en niveau de gris avec une taille de 3MB elles sont de résolution différentes, de type bmp (bitmap), Elle se compose de 100 images dont la tumeur est bénigne et 150 images dont la tumeur est maligne.

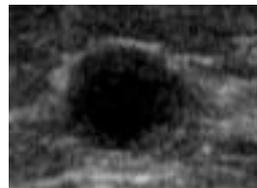
Voici quelques exemples d'images de type tumeur bénigne



105*77 taille:9.17Mo



95*75 taille:8.08Mo



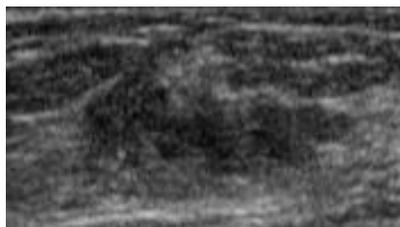
95*69 taille:7.52Mo

Figure 3.1 :Image de type tumeur bénigne

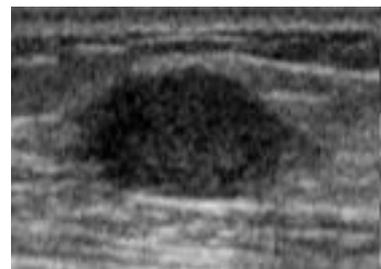
Et d'autres de tumeur maligne



137*113 taille:16.5Mo



177*99 taille:18.4Mo



139*99 taille:14.5Mo

Figure 3.2 : Image de type tumeur maligne

3.4. Prétraitement de la base de données

3.4.1.Divisions du dataset (Dataset splitting)

Fréquemment, la base de données est divisée en données d'entraînement, de validation et de test. Mais dans le cas de projets avec un nombre limité de données, cette méthode n'est pas vraiment recommandée. Vu que, la réservation d'une partie de la base de données pour la validation

reviendrait à réduire la quantité déjà faible de données dont on dispose. Et même si on faisait ce sacrifice, les données de test seraient trop peu pour être représentatives de l'ensemble de données. Pour pallier ces problèmes, on utilise la validation croisée (cross-validation en anglais (CV)). La validation croisée est une méthode statistique qui permet d'évaluer la capacité de Généralisation d'un modèle. Lorsqu'on parle de cross-validation, on fait référence généralement au k-fold cross-validation, qui est sa variante la plus populaire. Avec cette méthode, nous profitons de toutes les données dont on dispose en les divisant en k parties égales (folds) sur lesquelles on effectue l'entraînement et le test d'un modèle pendant k itérations. A chaque itération, le modèle est entraîné sur k-1 folds et est testé sur le fold restant. La figure suivante aide à mieux comprendre le principe de fonctionnement de CV.

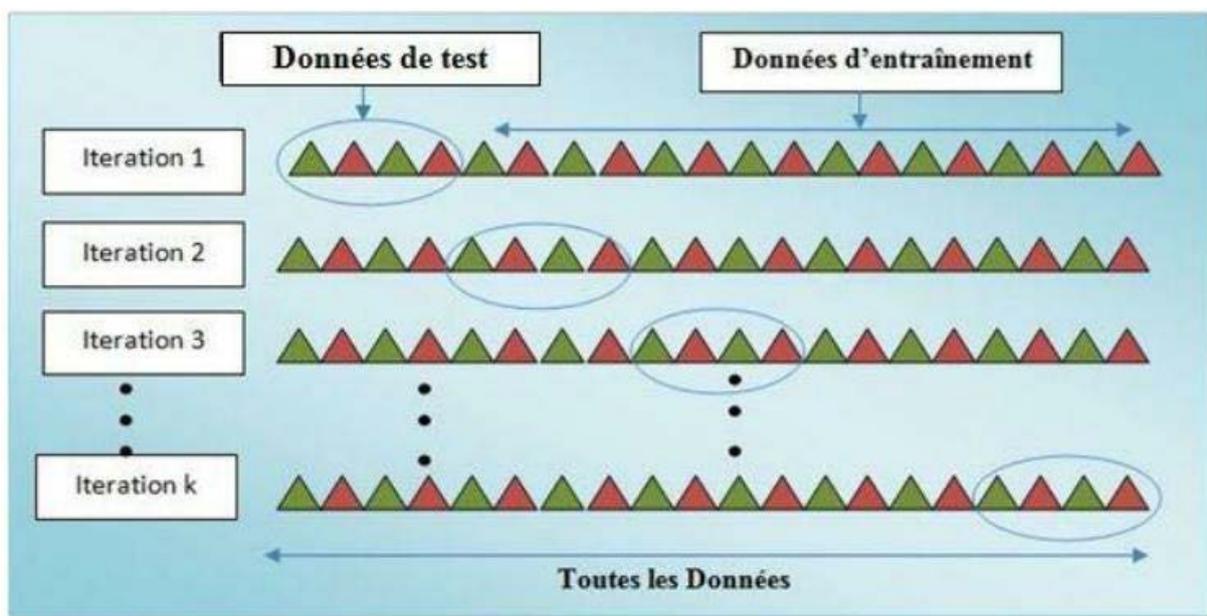


Figure 3.3 : Fonctionnement de la méthode k-fold CV.

Dans le cadre de notre mémoire, on travaille avec la méthode de cross validation.

L'ensemble des images est divisé en 5 parties égales (folds) sur lesquelles on effectue l'entraînement et le test d'un modèle pendant 5 itérations.

Dans chaque itération, un modèle est entraîné sur 4 folds et est testé sur le fold restant.

L'entraînement se fait sur 80% de l'ensemble des images et le test se fait sur les 20% restants.

3.4.2 Redimensionnement des images

Cette étape consiste à redimensionner les images à 299×299 et 227×227 pour être compatible avec le réseau Inceptionv3 et respectivement le réseau Alexnet. Nos données numériques sont constituées d'images en niveau de gris de sorte que la taille du canal est 1, pour l'image couleur, la taille du canal est 3 correspondant aux valeurs RVB va être convertie en une image en niveaux de gris de taille $299 \times 299 \times 1$ (respectivement $227 \times 227 \times 1$).

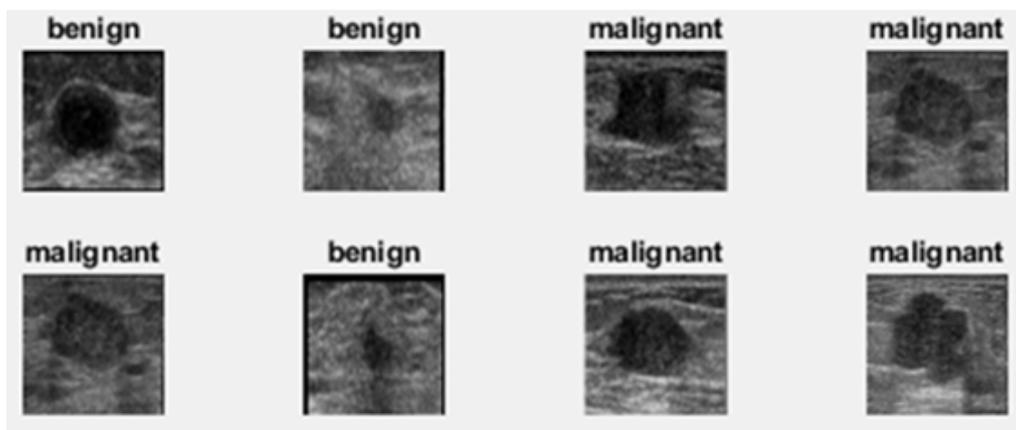


Figure 3.4 : Images redimensionnées

3.4.3. Augmentation de données (Data augmentation)

Quand le nombre d'images d'entraînement est faible, nous devons procéder à une augmentation de données pour générer de nouvelles images et prévenir ainsi le risque d'over fitting (surapprentissage).

Il existe plusieurs transformations qui augmentent le nombre d'images, nous citerons :

- le flip horizontal et/ou vertical
- la rotation
- la variation de luminosité et de contraste
- le redimensionnement...etc. Ainsi le volume de données est artificiellement multiplié.

Dans notre application, nous avons utilisé le redimensionnement ainsi que des transformations géométriques telles que: la rotation, la réflexion et le cisaillement (shear) sur les données d'entraînement selon le code suivant:

```
73 % Data Augumentation
74 - augmenter = imageDataAugmenter( ...
75     'RandRotation', [-5 5], 'RandXReflection', 1, ...
76     'RandYReflection', 1, 'RandXShear', [-0.05 0.05], 'RandYShear', [-0.05 0.05]);
77
78 % Resizing all training images to [229 229] for inceptionv3 architecture
79 - auimds = augmentedImageDatastore([299 299], imdsTrain, 'DataAugmentation', augmenter);
80
81 % Training
82 - netTransfer = trainNetwork(auimds, lgraph, options);
83
84 % Resizing all testing images to [299 299] for inceptionv3 architecture
85 - augtestimds = augmentedImageDatastore([299 299], imdsTest);
```

3.5. Classification avec le modèle d'apprentissage.

Les réseaux pré entraînés Inceptionv3 et Alexnet sont conçus pour résoudre le problème de la classification des images où l'entrée est une image de l'une des 1000 classes différentes et la sortie est un vecteur de 1000 éléments. Pour afficher une visualisation interactive de l'architecture du réseau et des informations détaillées sur les couches réseau nous utiliserons la commande : `analyzeNetwork(net)`

3.5.1. Application de l'apprentissage par transfert (Transfer Learning) :

La plupart des applications de l'apprentissage profond utilisent la méthode d'apprentissage par transfert, ou Transfer Learning, qui consiste à mettre au point un modèle pré-entraîné. Le processus commence avec un réseau existant, tels que InceptionV3 et Alexnet, qu'il faut enrichir avec de nouvelles données contenant des classes auparavant inconnues du réseau. Une fois quelques ajustements effectués au réseau, nous pouvons effectuer une nouvelle tâche, telle que la classification de la tumeur du sein.

L'avantage de la technique du Transfer Learning réside dans l'utilisation d'une base de données faible et un temps de calcul réduit.

3.5.2 Architectures du modèle utilisé dans l'implémentation

Nous avons utilisé les réseaux près entraînés Inceptionv3 et Alexnet en tant qu'extracteur de caractéristiques, la figure 3.5. montre l'architecture globale du modèle proposé.

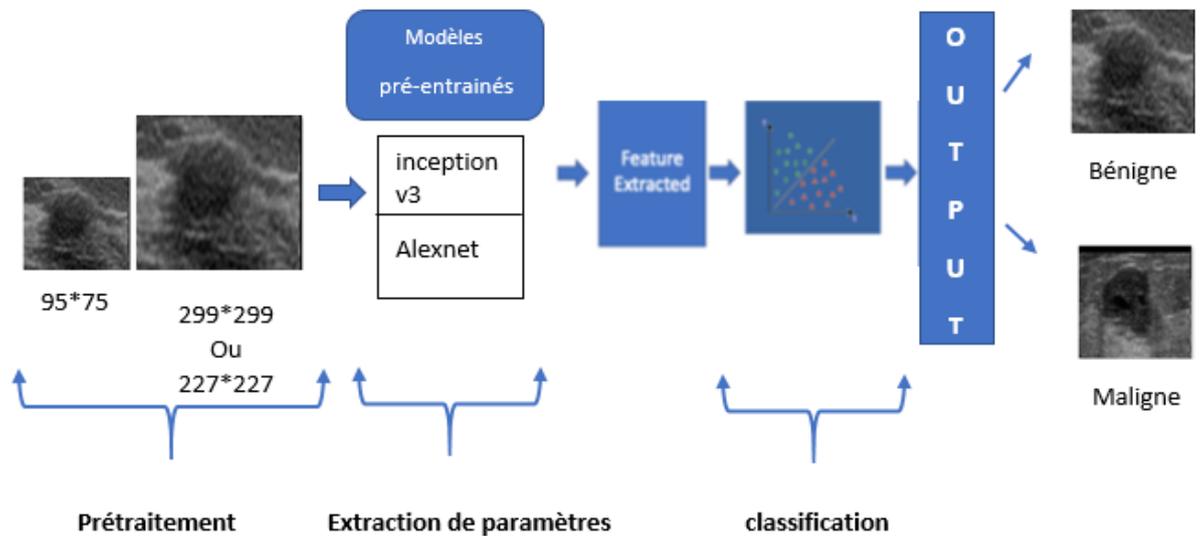


Figure 3.5 : Architecture globale du modèle proposé

L'architecture détaillée des 2 modèles pré-entraînés a été mentionnée dans le chapitre 2, pour cela on ne fait qu'apparaître que les 3 dernières couches dont nous avons besoin selon la figure 3.6.

➤ **Pour le modèle Inceptionv3**

313	predictions 1000 fully connected layer	Fully Connected	1×1×1000
314	predictions_softmax softmax	Softmax	1×1×1000
315	ClassificationLayer_predictions crossentropyex with 'tench' and 999 other classes	Classification Output	-

Figure 3.6 : Trois dernières couches du modèle inceptionv3

Nous avons gardé toutes les couches du réseau telle quelles à l'exception des trois dernières qui sont configurées pour 1000 classes. Ces trois couches doivent être adaptées pour le nouveau problème de classification (deux classes bénigne et maligne). Alors ces trois dernières on les a remplacées par :

- une couche entièrement connectée (new_fc)
- une couche softmax (new_softmax)
- une couche de sortie de classification (new_classoutput) selon la figure 3.7.

313	new_fc 2 fully connected layer	Fully Connected	1×1×2
314	new_softmax softmax	Softmax	1×1×2
315	new_classoutput crossentropyex with classes 'benign' and 'malignant'	Classification Output	-

Figure 3.7 : Trois dernières couches avec Transfer Learning

Nous avons spécifié les options de la nouvelle couche entièrement connectée en fonction des nouvelles données en la définissant pour qu'elle ait la même taille que le nombre de classes dans la nouvelle situation, alors nous l'avons configuré pour 2 classes. Comme il est mentionné dans la partie du code suivante :

```

45 net=inceptionv3;
46 lgraph = layerGraph(net);
47 clear net;
48
49 % Number of categories
50 numClasses = numel(categories(imdsTrain.Labels));
51
52 % Replacing the last layers with new layers
53
54 newFCLayer = fullyConnectedLayer(numClasses,'Name','new_fc','WeightLearnRateFactor',10,'BiasLearnRateFactor',10);
55 lgraph = replaceLayer(lgraph,'predictions',newFCLayer);
56 newClassLayer = softmaxLayer('Name','new_softmax');
57 lgraph = replaceLayer(lgraph,'predictions_softmax',newClassLayer);
58 newClassLayer = classificationLayer('Name','new_classoutput');
59 lgraph = replaceLayer(lgraph,'ClassificationLayer_predictions',newClassLayer);

```

➤ Pour le modèle Alexnet

Les 3 dernières couches du modèle Alexnet sont illustrées par la figure 3.8

23	fc8 1000 fully connected layer	Fully Connected	1×1×1000
24	prob softmax	Softmax	1×1×1000
25	output crossentropyex with 'tench' and 999 other classes	Classification Output	-

Figure 3.8 : Trois dernières couches du modèle Alexnet

Avec Transfert Learning, les 3 dernières couches du modèle Alexnet sont représentées par la figure 3.9

23	fc 2 fully connected layer	Fully Connected	1×1×2
24	softmax softmax	Softmax	1×1×2
25	classoutput crossentropyex with classes 'benign' and 'malignant'	Classification Output	-

Figure 3.9 : Trois dernières couches du modèle Alexnet avec Transfer Learning

3.5.3 Hyper paramètres d'apprentissage

- **Mini Batch Size (Batch ou lot) :** le Mini Batch size signifie que vous ne prenez qu'un sous-ensemble de toutes vos données au cours d'une itération pour évaluer le gradient de la fonction de perte et mettre à jour la pondération.
- **Epoques (epoch) :** Epoch signifie un passage sur l'ensemble d'entraînement complet. Le nombre d'époques est un hyper paramètre de descente de gradient qui contrôle le nombre de passages complets dans l'ensemble de données d'apprentissage. 'MaxEpochs' c'est le nombre maximal d'époques à utiliser pour l'apprentissage spécifié.
- **Learning rate :** (taux d'apprentissage). Dans le domaine de l'apprentissage automatique, le taux d'apprentissage consiste à un paramètre de réglage, qui détermine la taille du pas à chaque itération tout en visant le minimum d'une fonction de perte. Il représente métaphoriquement la vitesse à laquelle un modèle d'apprentissage automatique apprend.

3.6. Les indicateurs de performance

Dans le but d'évaluer les performances du modèle InceptionV3 (respectivement Alexnet) que nous allons utiliser pour l'implémentation, nous allons calculer la précision, l'accuracy (exactitude), la sensibilité, la spécificité, et le F1 score. Nous présenterons la matrice de confusion qui regroupe tous ces paramètres ainsi que la courbe ROC

3.6.1 Matrice de confusion.

La matrice de confusion englobe les résultats de prédiction pour un problème particulier de classification (bonne prédictions sur sa diagonale). Elle compare les données réelles pour une variable cible à celles prédites par un modèle. Il y a quatre (04) cas de prédiction que nous pouvons présenter dans la matrice de confusion :

- **True Positive (TP) :** la prédiction et la valeur réelle sont positives.
- **True Negative (TN) :** la prédiction et la valeur réelle sont négatives.
- **False Positive (FP) :** la prédiction est positive alors que la valeur réelle est négative.
- **False Negative (FN) :** la prédiction est négative alors que la valeur réelle est positive.

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{TP + FN}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{TN + FP}$
		Precision $\frac{TP}{TP + FP}$	Negative Predictive Value $\frac{TN}{TN + FN}$	Accuracy $\frac{TP + TN}{TP + TN + FP + FN}$

Figure 3.10 : Matrice de confusion [34]

➤ **Mesure de performance**

- Exactitude ‘Accuracy’ : Taux de bonnes prédictions.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (3.1)$$

- Précision : Taux faux positif

$$\text{Précision} = \frac{TP}{TP+FP} \quad (3.2)$$

- Sensitivité ‘Recall’ : Taux vrai positif Probabilité d’un cas positif si la prédiction est positive.

$$\text{Sensitivité Ou Recall} = \frac{TP}{TP + FN} \quad (3.3)$$

- Spécificité : Taux vrai négatif :

$$\text{Spécificité} = \frac{TN}{TN + FP} \quad (3.4)$$

- F1-Score :

$$\text{F1-Score} = \frac{2 * \text{précision} * \text{recall}}{\text{précision} + \text{recall}} \quad (3.5)$$

3.6.2.La courbe ROC

La courbe ROC (Reciever Operating Characterstic) et AUC (Area Under the ROC) sont des mesures graphiques des performances de modèle, qui nous aident à résoudre le problème de classification, la courbe ROC correspond à la sensibilité en fonction de 1-spécificité a différent

seuils de classification, où la sensibilité représente le taux de vrais positifs et 1-specificité représente le taux de faux positifs.

L'AUC quantifie la capacité d'un modèle à différencier entre les instances positifs et les instances négatifs, plus l'AUC est proche de 1 plus le modèle est capable de classer correctement, AUC de 0.5 indique que le modèle est non informatif.

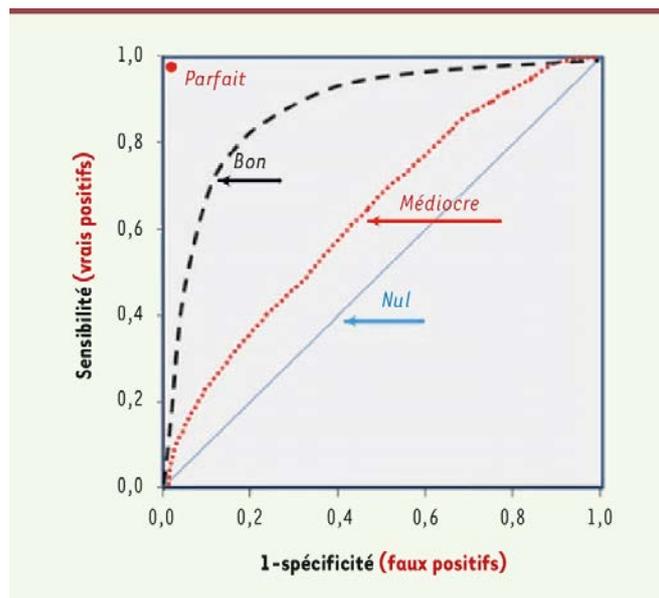


Figure 3.11 : Courbe ROC [35]

3.7. Résultats et discussions

Dans cette section nous allons présenter et discuter les résultats obtenus suite à l'application du modèle InceptionV3 et respectivement le modèle Alexnet pour la classification du cancer du sein en tumeur bénigne et maligne.

- **Modèle Inceptionv3**

En premier lieu, nous allons commencer d'abord par changer le nombre d'époques pour voir son impact sur l'accuracy. Après avoir initialisé les hyper paramètres comme suite :

Learning Rate =0.0001

Mini Batch size =10

Option Adam

Et en faisant varier le nombre d'époques de 6, 10 jusqu'à 15, nous avons obtenu les résultats illustrés dans le tableau 3.1

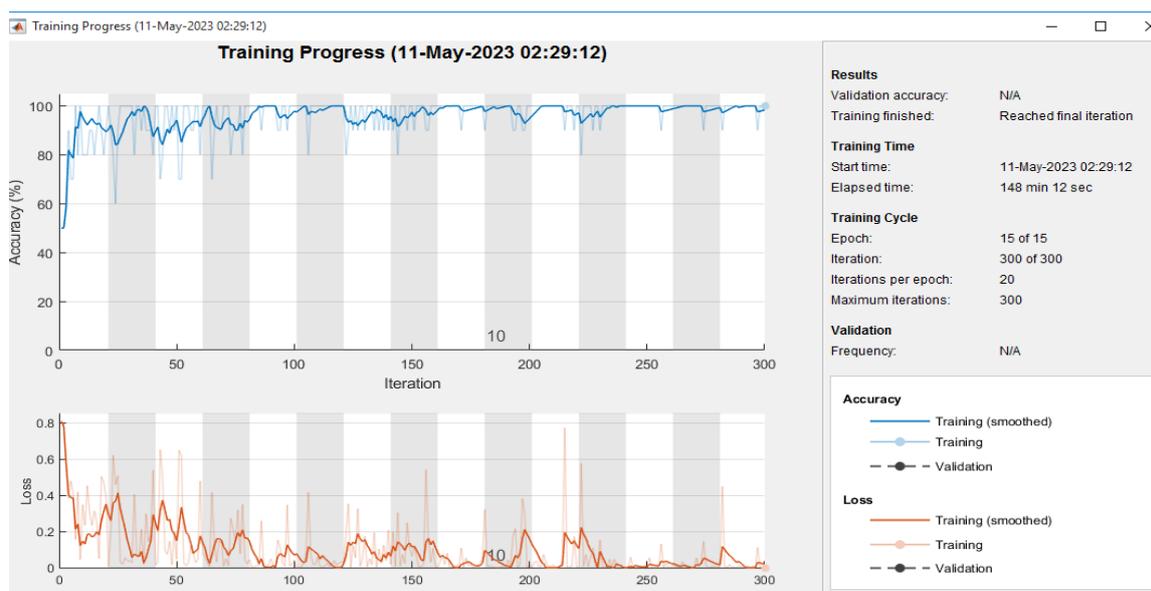
Tableau 3.1 : Taux d'apprentissage et temps d'exécution en changeant le nombre d'époque.

MaxEpochs	6	10	15
Accuracy(%)	96.4%	98%	99.6%
Temps d'exécution	296min125s	473min159s	738min186s

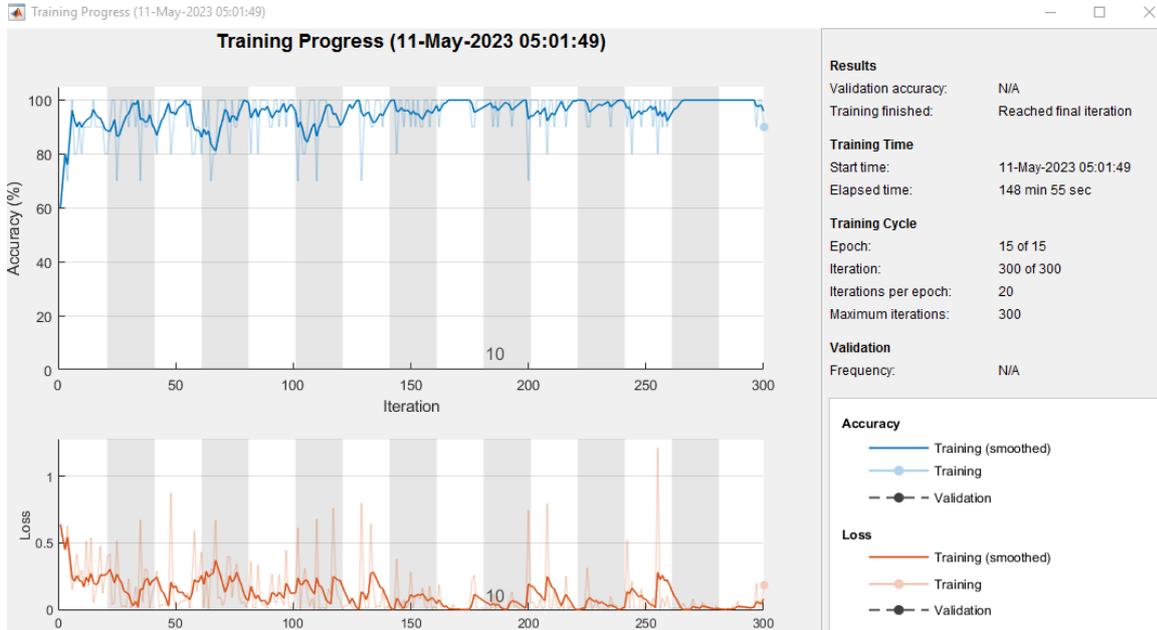
Les résultats obtenus dans le tableau 3.1 montrent qu'en augmentant le nombre d'époques jusqu'à 15, le taux l'accuracy augmente à **99.6%** ce qui prouve qu'après chaque époque le modèle apprend plus d'informations au détriment d'un temps d'exécution trop long (**738min186s**) et ce à cause du matériel utilisé.

Les courbes d'évolution de l'apprentissage (accuracy and loss) sur les données d'entraînement du modèle inceptionv3 sont données par les figures suivantes pour un nombre d'époque égale à 15, Learning Rate =0.0001, Mini Batch size =10 et l'option Adam selon les 5 split:

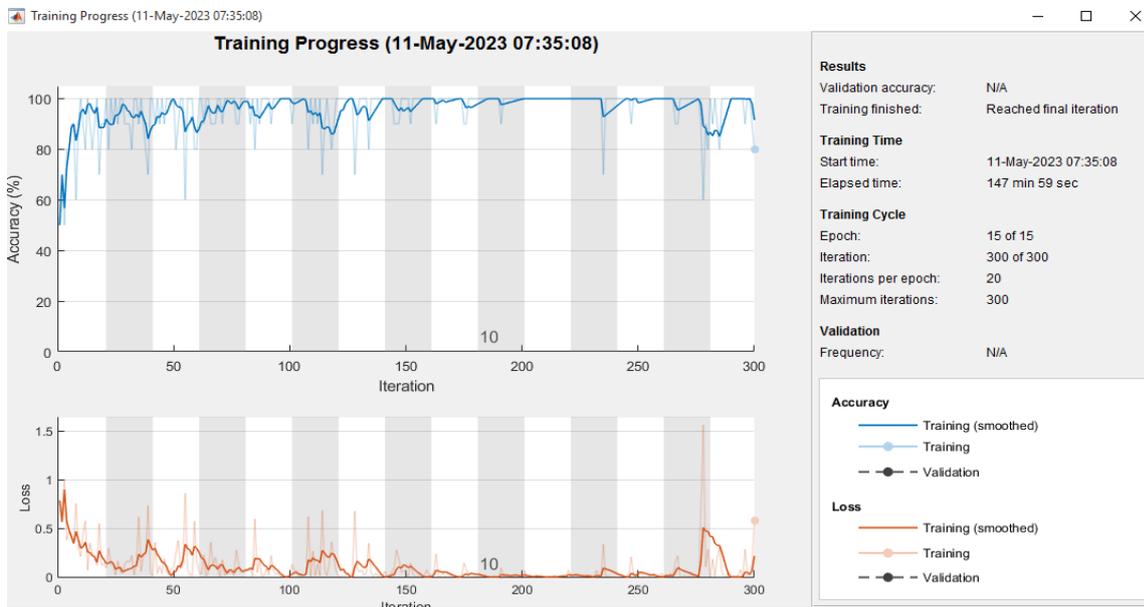
Split 1



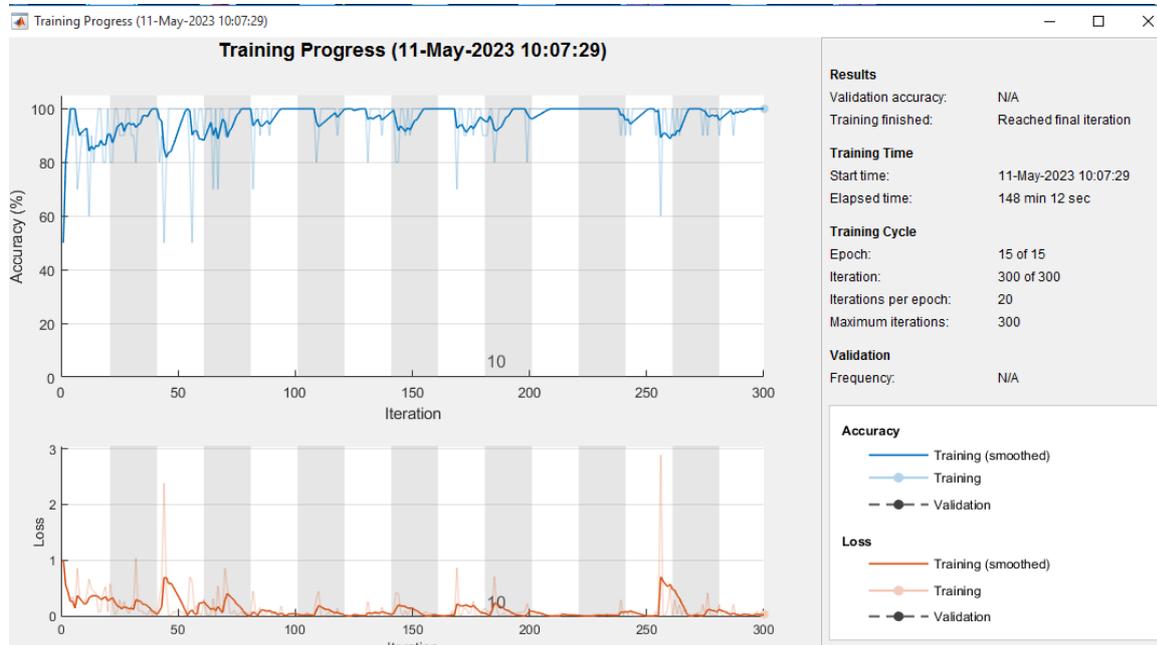
Split2



Split3



Split4



Split5

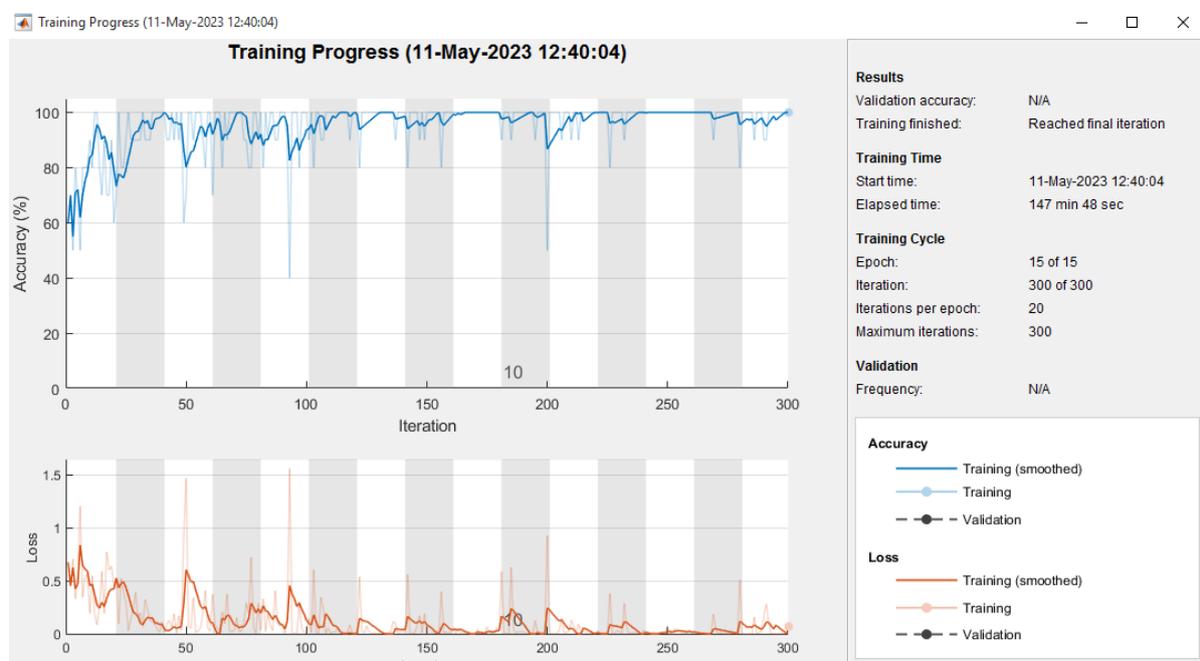


Figure 3.12 : Courbes d'évolution de l'apprentissage (accuracy and loss) pour les split 1, 2, 3, 4 et 5 avec Inceptionv3

D'après ces figures, nous remarquons bien qu'en augmentant le nombre d'époques, la précision d'entraînement est en effet en train d'augmenter alors que la perte d'entraînement diminue.

La figure 3.13 représente la matrice de confusion pour l'ensemble de test, rappelons que cette matrice permet d'évaluer notre modèle.

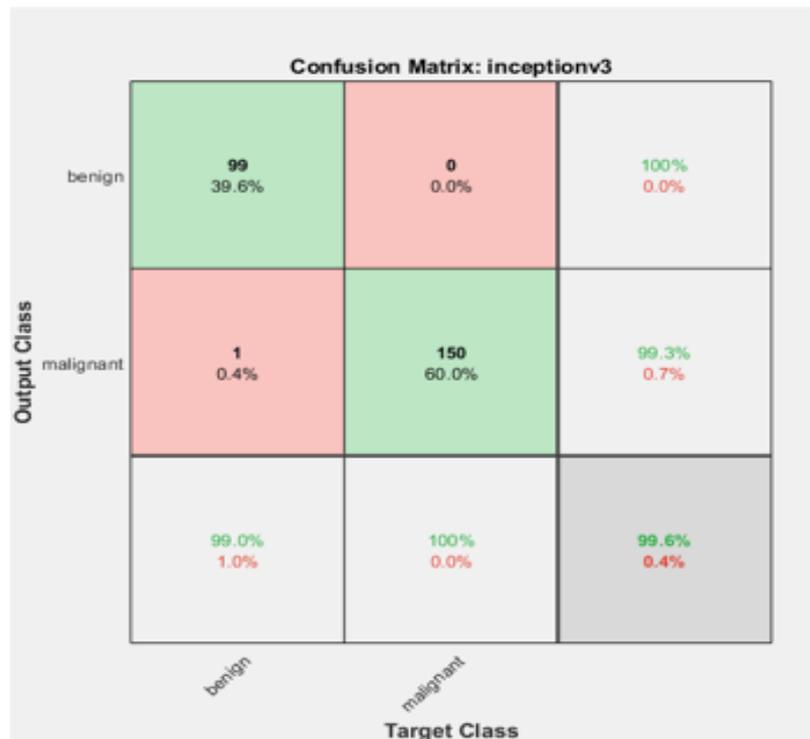


Figure 3.13 : Matrice de confusion globale pour l'ensemble de test (Inceptionv3 et option Adam)

Inceptionv3 a obtenu un taux de précision de 99.6% avec un taux d'erreur de 0.4%.

Le tableau 3.2 regroupe les indicateurs de performances du model Inceptiov3.

Tableau 3.2 : mesures de performances (Inceptionv3)

Modèle	Accuracy	Sensitivity	Spécificity	Précision	F1 score
Inceptionv3	99.6%	99%	100%	100%	99.4%

La figure 3.14 représente la courbe ROC, remarquons qu'elle est très proche de 1 (AUC=0.9999) ce qui prouve la performance du classifieur.

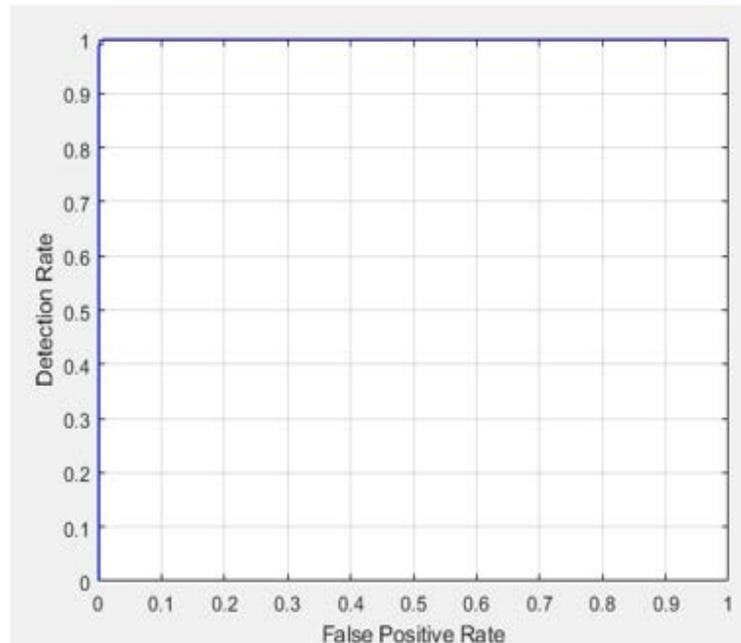


Figure 3.14 : La courbe ROC inceptionv3

3.7.1 Tests selon les options d'apprentissage

La deuxième étape de notre travail consiste à tester plusieurs méthodes d'optimisation en changeant l'optimiseur adam, en sgd et rmsprop afin de déterminer l'option la plus adaptée pour notre modèle.

L'optimiseur est un algorithme chargé de modifier les poids dans les couches du modèle en fonction des erreurs lors du fonctionnement.

Les différents optimiseurs que nous voulons tester, ont été cités dans le chapitre précédent

En initialisant les hyper paramètres à :

- MaxEpochs = 15
- Learning Rate =0.0001
- MiniBatchSize = 10

Et après avoir utilisé l'optimiseur Adam ci-dessus, appliquons à présent l'optimiseur Sgdm et RMSprop simultanément. Les figures 3.15 et 3.16 montrent leurs matrices de confusion.

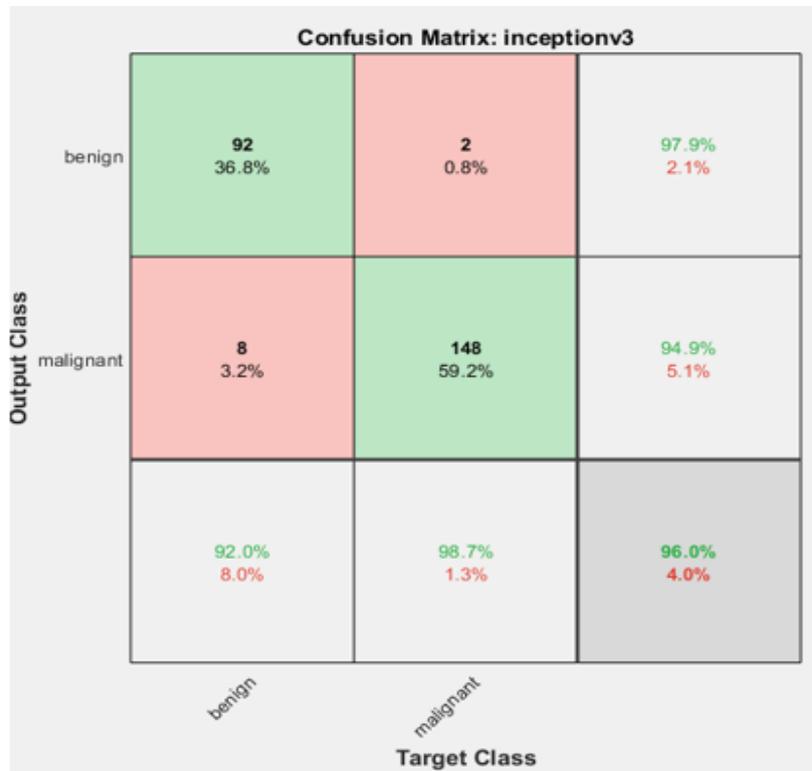


Figure 3.15 : Matrice de confusion globale avec sgd

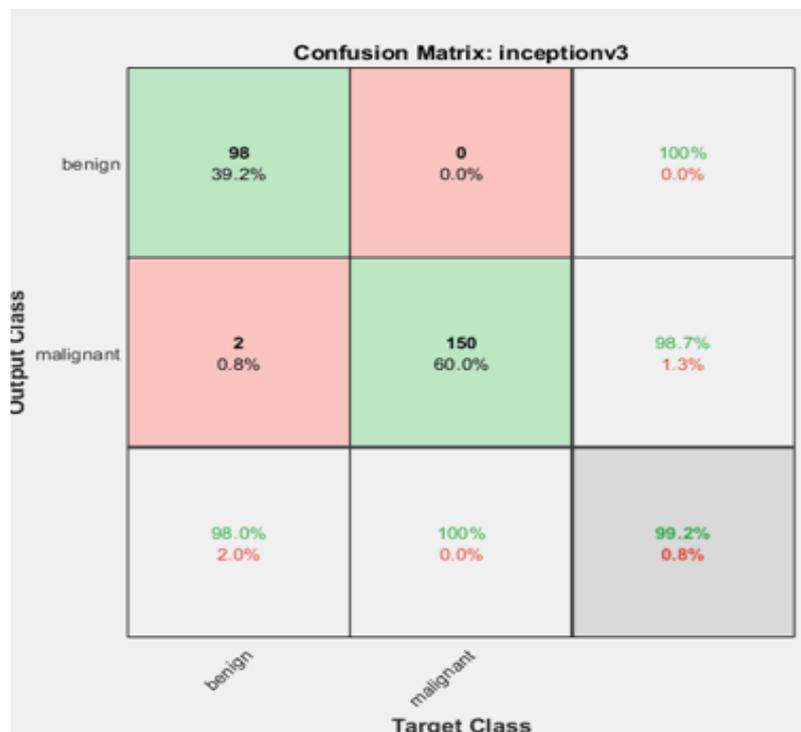


Figure 3.16 : Matrice de confusion globale avec RMSprop

Les performances de classification du modèle Inceptionv3 pour le problème de classification binaire (maligne /bénigne) sont fournies dans le tableau

Tableau 3.3 : tableau récapitulatif des mesures de performances selon les trois optimiseurs.

	Epoch6 Adam	Epoch10 Adam	Epoch15 Adam	Epoch6 RMSprop	Epoch10 RMSprop	Epoch15 RMSprop	Epoch6 Sgdm	Epoch10 Sgdm	Epoch15 Sgdm
AUC	99.44%	99.98%	99.99%	99.41%	99.86%	99.99%	98.01%	99.13%	99.63%
Accuracy	96.4%	98%	99.6%	95.6%	97.6%	99.2%	92%	93.6%	96%
Précision	97.9%	100%	100%	100%	100%	100%	94.4%	94.7%	97.9%
spécificity	98.7%	100%	100%	100%	100%	100%	96.7%	96.7%	98.7%
sensitivity	93%	95%	99%	89%	94%	98%	85%	89%	92%
F1 score	95.3%	97.4%	99.4%	94.1%	96.9%	98.9%	89.4%	91.7%	94.8%

- Pour 15 époques, l'application de l'optimiseur sgdm nous donne une Accuracy de 96% et une perte de 4% alors que le RMSprop et l'adam nous donnent respectivement une accuracy de 99.2% (99.6%) et une perte de 0.8% (0.4%).

Notons que l'accuracy obtenue avec l'optimiseur adam dépasse considérablement celle obtenue avec l'optimiseur sgdm mais elle est sensiblement égale à l'optimiseur RMSprop. L'AUC avec adam et RMSprop est la même (AUC=99.99) et se rapproche de 1, alors que pour le sgdm l'AUC est nettement inférieure (99.63).

- Pour 6 et 10 époques, le modèle donne un accuracy de 96.4% et 98% respectivement, qui sont inférieurs par rapport à 15 époques.

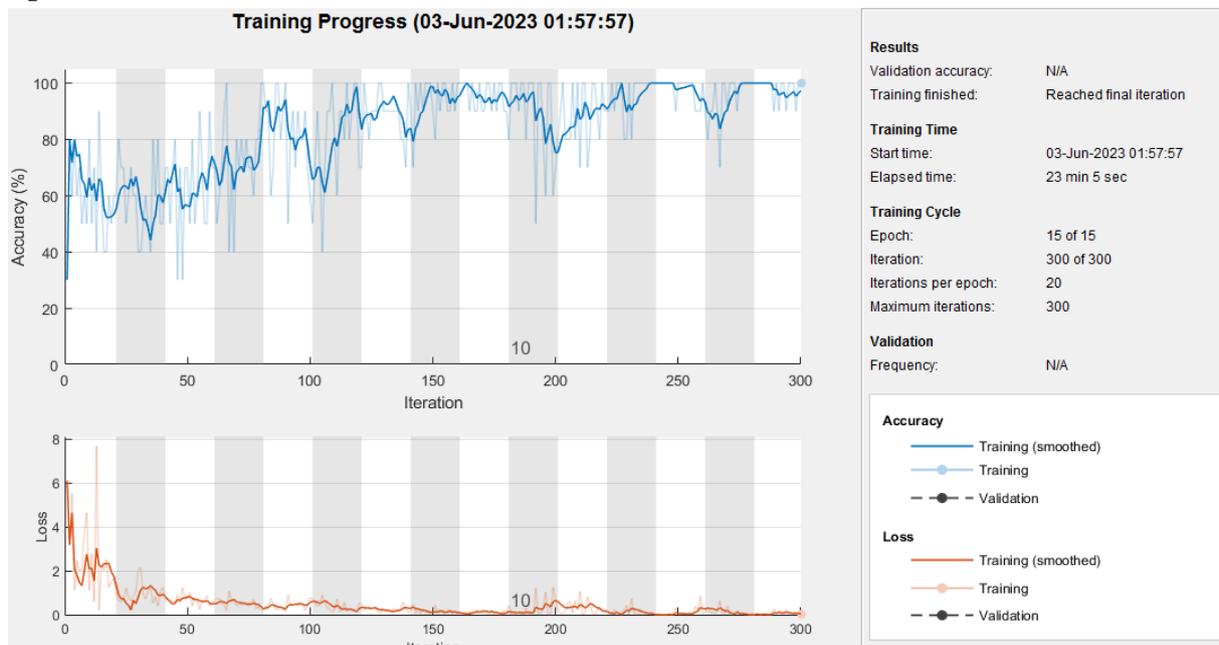
- Le modèle avec **15 époques** et l'optimiseur **Adam** donne le meilleur résultat : un Accuracy qui vaut **99.6%**

Tous ces tests ont été réalisés dans le but de voir leurs impacts sur l'efficacité du modèle.

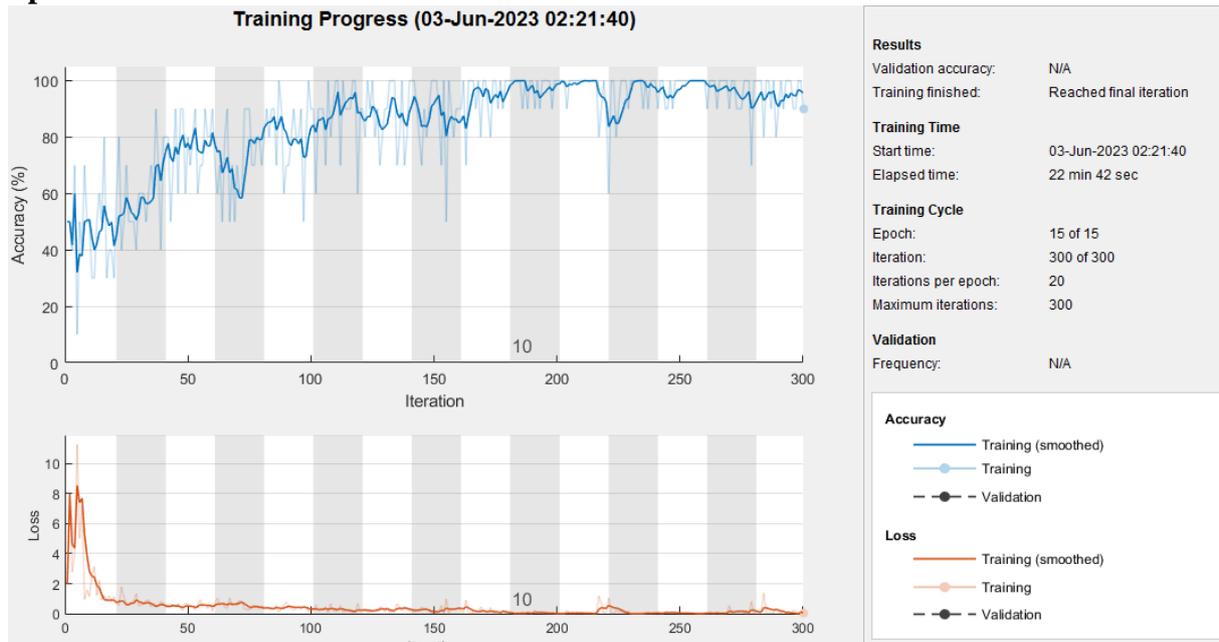
➤ **Modèle Alexnet**

Nous allons dans cette partie, essayer la classification avec le modèle Alexnet pour les paramètres suivants : MaxEpoch= 15, MiniBatchsize= 10, Learning Rate= 0.0001, optimiseur Adam.

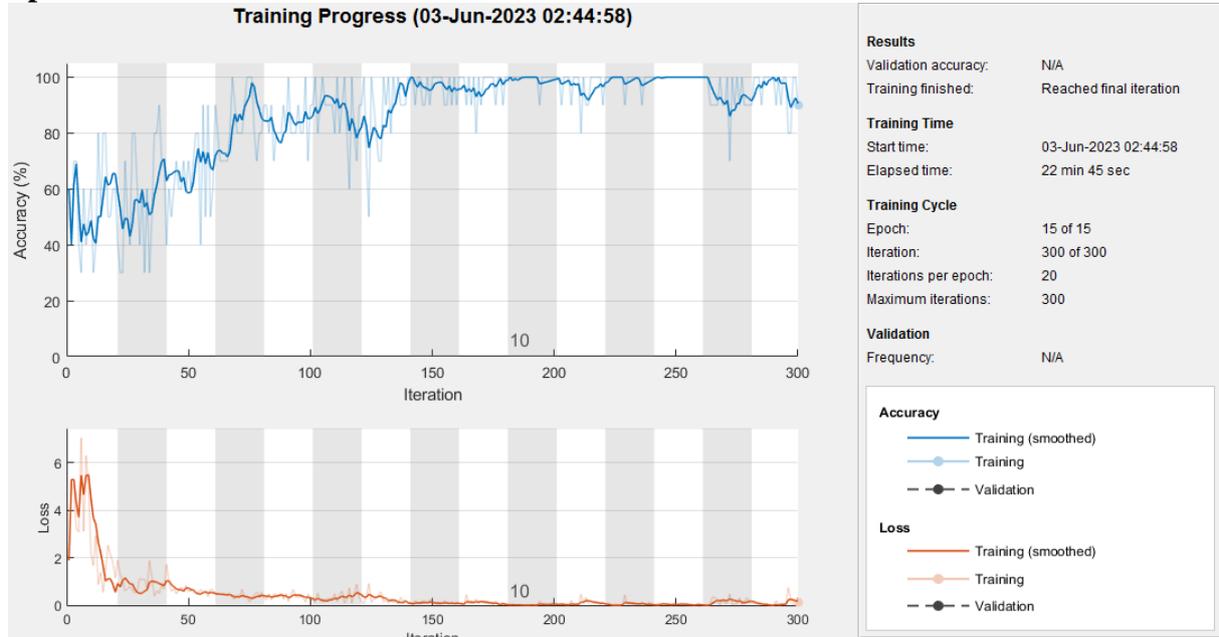
Split1



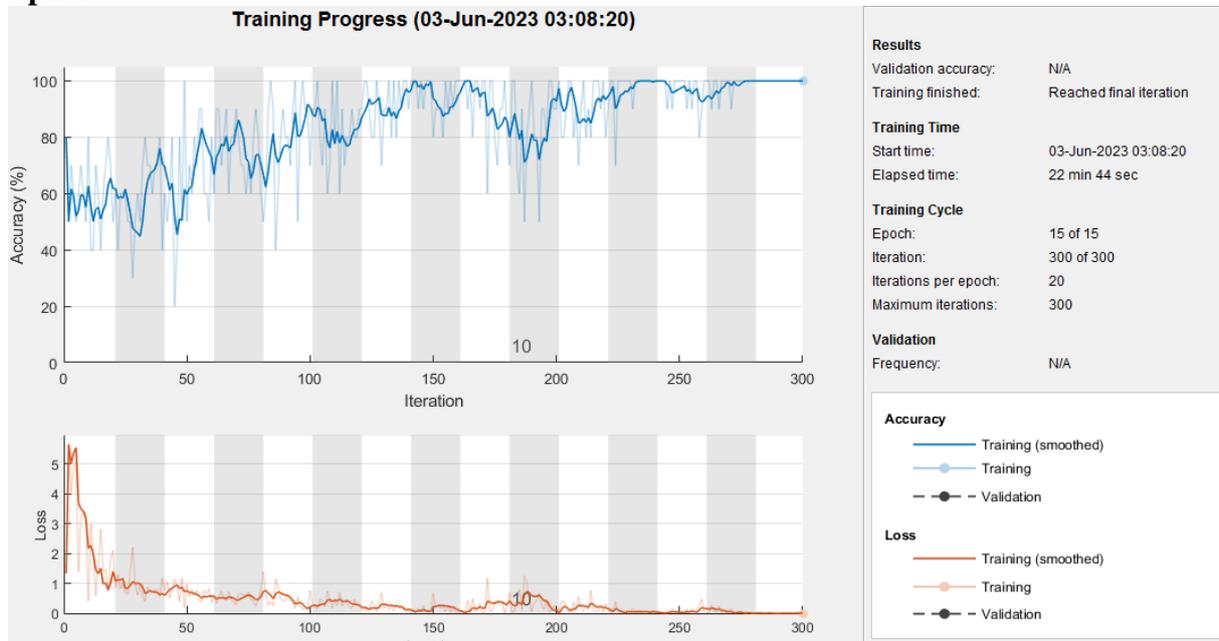
Split2



Split3



Split4



Split5

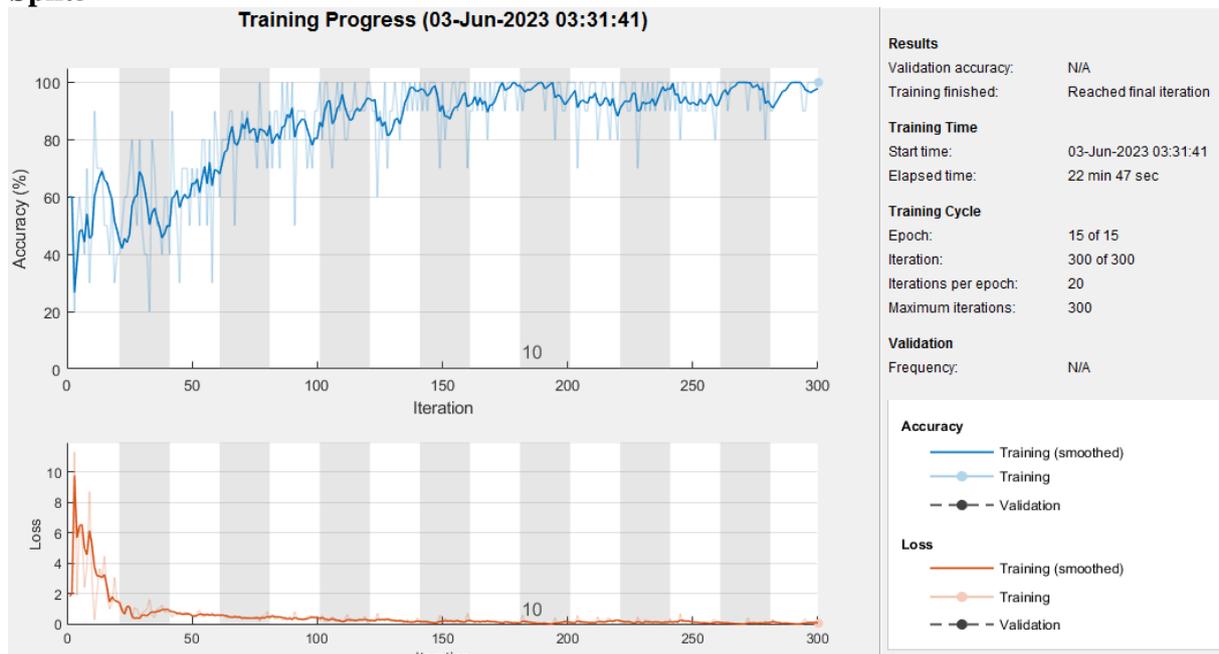


Figure 3.17 : Courbes d'évolution de l'apprentissage (accuracy and loss) pour 1, 2, 3, 4 et 5 pour Alexnet

Malgré l'augmentation du nombre d'époque à 15, nous remarquons l'instabilité des courbes d'entraînement pour le modèle Alexnet.

La figure 3.18 représente la matrice de confusion d'Alexnet pour l'ensemble de test. Cette matrice montre que ce modèle a obtenu un accuracy de 94.4% avec un taux d'erreur de 5.6%.

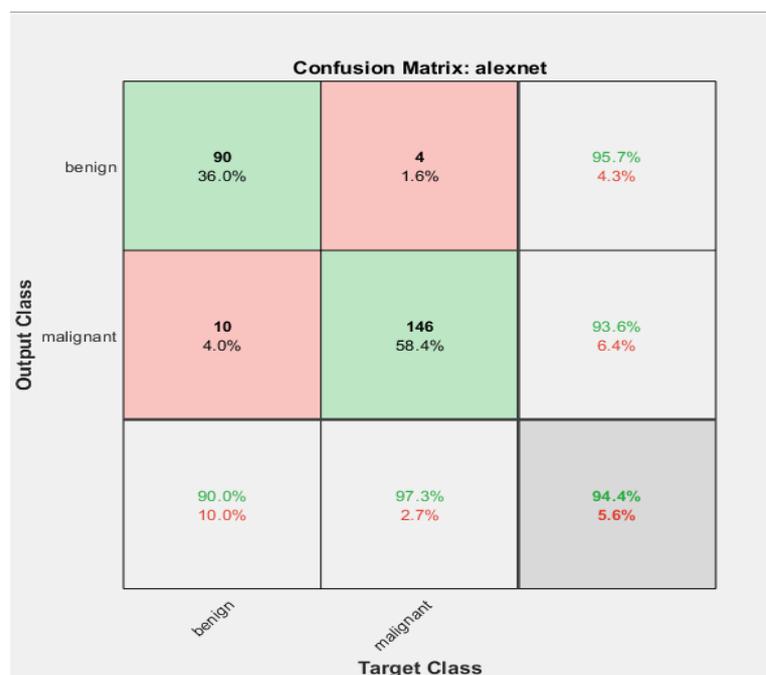


Figure 3.18 : Matrice de confusion globale avec Alexnet

La courbe ROC prend l'allure suivante :

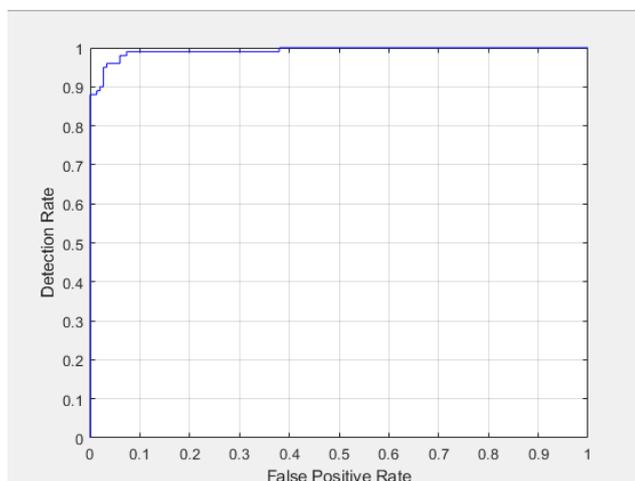


Figure 3.19 : La courbe ROC Alexnet

L'AUC obtenu pour ce modèle est de **99.23%**, comparée à l'AUC du modèle Inceptionv3 qui est égale à **99.99%**, une différence appréciable est à noter.

Le tableau 3.4 regroupe l'évaluation des performances obtenues avec le modèle Alexnet.

Tableau 3.4 : mesures de performances avec Alexnet

Modèle	Accuracy	sensitivité	Spécificité	précision	F1 score
Alexnet	94.4%	90%	97.3%	95.7%	92.7%

Enfin le tableau 3.5 regroupe les paramètres de performances des deux modèles inceptionv3 et Alexnet pour 15 époque, option Adam, Learning rate =0.0001 Mini BatchSize=10.

Tableau 3.5 : Tableau comparatif des valeurs des indicateurs de performances d'Inceptionv3 et Alexnet

Modèle	Accuracy	sensitivité	spécificité	Précision	F1 score
Alexnet	94.4%	90%	97.3%	95.7%	92.7%
Inceptionv3	99.6%	99%	100%	100%	99.4%

À partir des résultats obtenus, on voit bien que l'utilisation du modèle inceptionv3 permet d'obtenir des valeurs meilleures qu'Alexnet du point de vue de tous les indicateurs de performances Accuracy, sensibilité, spécificité, précision, et F1 score aussi pour les deux courbes ROC, on remarque que l'aire sous la courbe de Inceptiov3 est plus grande que celle sous la courbe de Alexnet.

3.7.2 Validation du modèle inceptionv3

Afin de tester le modèle inceptionv3, nous avons sélectionné des images au hasard [36] présentant un cancer du sein figure 3.20 et nous avons effectué une validation sur ce modèle, les résultats de validation sont affichés dans le tableau 3.6.représentant l’accuracy obtenue pour chaque image.

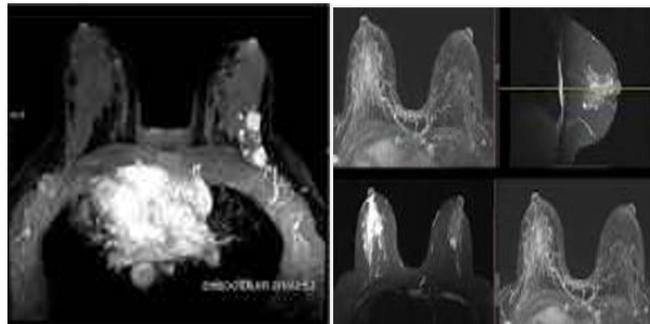
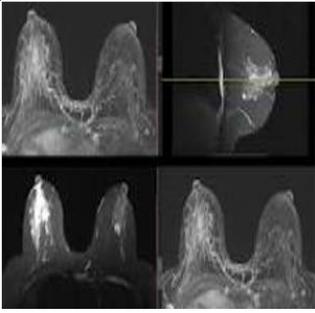
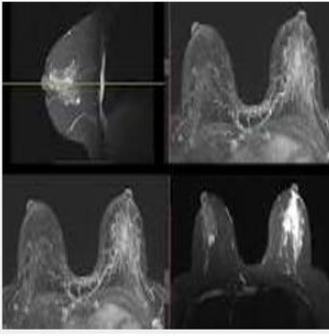


Figure.3.20 :Images tests

Tableau 3.6 : Résultats de validation

Images tests	Accuracy
 <p>Image1</p>	<p>benign0.917829</p>  <p>91,78% bénigne</p>
 <p>Image2</p>	<p>malignant0.999308</p>  <p>99,93% maligne</p>

Nous constatons que le modèle a reconnu l'image 1 bénigne avec une exactitude de 91.78%, et l'image 2 maligne à 99.93%.

3.8.Conclusion

Dans ce chapitre, nous avons décrit notre base de données, nous avons présenté deux modèles utilisés pour notre implémentation ainsi que les phases entraînement. Après ceci nous avons montré les résultats représentés en termes de performance ainsi que les courbes correspondantes. Nous avons aussi mené nos tests selon trois options d'apprentissage, nous les avons ensuite comparés dans le but de voir lequel est le meilleur pour notre étude. Enfin nous avons comparé les deux modèles et les expérimentations ont montré l'efficacité de notre approche avec le modèle Inceptionv3 dans l'identification du cancer du sein.

Conclusion générale

Le cancer du sein est la maladie la plus répandue, malgré les grandes percées dans la détection et la prévention, et il reste la principale cause de décès. Ces derniers jours, l'intelligence artificielle a explosé dans tous les domaines de la science. L'utilisation de l'intelligence artificielle dans la science médicale a amélioré l'exactitude et la précision de la prédiction de cette maladie dans les premiers stades.

Dans l'étude proposée, un apprentissage en profondeur et une imagerie par résonance magnétique sont utilisés pour classer deux types différents de cancer du sein : tumeur maligne et tumeur bénigne. Deux modèles pré-entraînés des réseaux de neurones convolutionnels Inceptionv3 et Alexnet avec le transfert Learning ont été proposés pour classer le cancer du sein. La mesure de l'efficacité de la classification faite par Inceptionv3 et Alexnet s'est basée sur deux indicateurs de performance : la matrice de confusion et la courbe ROC, de même plusieurs optimiseurs ont été comparés dans le but de conclure le meilleur algorithme d'optimisation.

Enfin, le modèle de classification Inceptionv3 avec l'optimiseur adam a atteint un taux de précision de 99.6% pour quinze époques et c'est ce qui prouve son efficacité.

PERSPECTIVES :

Plusieurs perspectives peuvent être envisagées ci-après :

- L'utilisation des ressources de calcul très puissantes tel que les GPU (Graphic Processing Unit) et les TPU (Tensor Processing Unit), pour accélérer le temps d'exécution .
- Il est recommandé d'utiliser le langage python car il contient un riche écosystème de Bibliothèque pour l'apprentissage profond. Python a une grande communauté de programmeur car il est open-source.
- Entraîner notre modèle sur une base de données plus grande.
- Utiliser des architectures des modèles plus avancé.

ANNEXE

Algorithme Sgdm

Algorithm : Stochastic gradient descent (SGD) with momentum

Require : Learning rate ϵ , momentum parameter α

Require : Initial parameter θ , initial velocity v

While stopping criterion not met do

Sample a minibatch of m examples from the training set $\{x^{(1)}, \dots, x^{(m)}\}$ with corresponding targets $y^{(i)}$.

Compute gradient estimate : $g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x^i; \theta), y^i)$

Compute velocity : $v \leftarrow \alpha v - \epsilon g$.

Apply update : $\theta \leftarrow \theta + v$

End While

Algorithme AdaGrad

Algorithm : The AdaGrad algorithm

Require : Global learning rate ϵ

Require : Initial parameter θ

Require : Small constant δ , perhaps 10^{-7} , for numerical stability

Initialize gradient accumulation variable $r = \mathbf{0}$

While stopping criterion not met **do**

Sample a minibatch of m examples from the training set $\{x^{(1)}, \dots, x^{(m)}\}$ with corresponding targets $y^{(i)}$.

Compute gradient : $g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x^i; \theta), y^i)$

Accumulate squared gradient : $r \leftarrow r + g \odot g$

Compute update : $\Delta\theta \leftarrow -\frac{\epsilon}{\delta + \sqrt{r}} \odot g$. (Division and square root applied element-wise)

Apply update : $\theta \leftarrow \theta + \Delta\theta$

End While

Algorithm RMSProp

Algorithm : The RMSProp algorithm

Require : Global learning rate ϵ , decay rate ρ

Require : Initial parameter θ

Require : Small constant δ , perhaps 10^{-6} , used to stabilize division by small numbers

Initialize gradient accumulation variable $r = \mathbf{0}$

While stopping criterion not met **do**

Sample a minibatch of m examples from the training set $\{x^{(1)}, \dots, x^{(m)}\}$ with corresponding targets $y^{(i)}$.

Compute gradient : $g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x^i; \theta), y^i)$

Accumulate squared gradient : $r \leftarrow \rho r + (1 - \rho) g \odot g$

Compute update : $\Delta\theta \leftarrow -\frac{\epsilon}{\delta + \sqrt{r}} \odot g$ ($\frac{\epsilon}{\delta + \sqrt{r}}$ applied element-wise)

Apply update : $\theta \leftarrow \theta + \Delta\theta$

End While

Algorithm Adam

Algorithm : the ADAM algorithm

Require : Step size ϵ (Suggested default : 0.001)

Require : Exponential decay rates for moment estimates, ρ_1 and ρ_2 in $[0,1)$. (Suggested defaults : 0.9 and 0.999 respectively)

Require : Small constant δ used for numerical stabilization (Suggested default : 10^{-8})

Require : Initial parameters θ

Initialize 1st and 2nd moment variables $s = \mathbf{0}$, $r = \mathbf{0}$

Initialize time step $t = 0$

While stopping criterion not met **do**

Sample a minibatch of m examples from the training set $\{x^{(1)}, \dots, x^{(m)}\}$ with corresponding targets $y^{(i)}$.

Compute gradient : $g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x^i; \theta), y^i)$

$t \leftarrow t + 1$

Update bias first moment estimate : $s \leftarrow \rho_1 s + (1 - \rho_1) g$

Update biased second moment estimate : $r \leftarrow \rho_2 r + (1 - \rho_2) g \odot g$

Correct bias in first moment : $\hat{s} \leftarrow \frac{s}{1 - \rho_1^t}$

Correct bias in second moment : $\hat{r} \leftarrow \frac{r}{1 - \rho_2^t}$

Compute update : Compute update : $\Delta\theta \leftarrow -\epsilon \frac{\hat{s}}{\sqrt{\hat{r} + \delta}}$ (operation applied element-wise)

Apply update : $\theta \leftarrow \theta + \Delta\theta$

End While

Bibliographie :

- [1] <https://www.who.int>...>principaux repères> Détail>
- [2] <https://lemondeetnous.cafe-sciences.org/2014/07/comprendre-le-cancer-du-sein-partie-i/>, consulté le 20/05/2020
- [3] “Qu’est-ce que le cancer du sein ? |”Biron.” [Online]. Available: <https://www.biron.com/fr/centre-du-savoir/petit-guide-biron/cancer-du-sein/>. [Accessed: 15-Feb-2020].
- [4] Santé mobile : des applications de qualité.fr. url : https://www.has-sante.fr/jcms/p_3106528/fr/sante-mobile-des-applications-de-qualite.
- [5] F. A. Zeiser et al., “Segmentation of Masses on Mammograms Using Data Augmentation and Deep Learning,” J. Digit. Imaging, 2020, doi: 10.1007/s10278-020-00330-4.
- [6] Christophe Ginestier, Hasan Korkaya, Gabriela Dontu, Daniel Birnbaum, Max S. Wicha, Emmanuelle Charafe-Jauffret, La cellule souche cancéreuse, Un pilote aux commandes du cancer du sein, 2007
- [7] Olivier Hermine, Cancer : le diagnostic, <https://www.fondation-arc.org/cancer-le-diagnostic>, consulté le 29/0/2020 dictionnaire médicale flammariion
- [8] <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Diagnostic/Mammographie>
- [9] Samet Sara : Une approche basée Deep Learning pour la détection du cancer du sein. Mémoire Master Université de Tébessa. Septembre 2020
- [10] “Imagerie Médicale - Toutes les infos utiles sur l’IRM mammaire.” [Online]. Available: <https://imageriemedicale.fr/examens/irm/mammaire/>. [Accessed: 15-Feb-2020].
- [11] P. Baron, L. Bergevin, G. Lépine, and J. Morin, *Mammographie normes de pratique spécifiques*. Ordre des technologues en imagerie médicale, en radio-oncologie et en électrophysiologie médicale du Québec, 2019.
- [12] <https://www.e-cancer.fr/Patients-et-proches/Les-cancers/Cancer-du-sein/Facteurs-de-risque>
- [13] https://members.loria.fr/FSur/enseignement/apprauto/poly_apprauto_FSur.pdf
- [14] <https://geekflare.com/fr/regression-vs-classification/>
- [15] <https://blent.ai/blog/a/apprentissage-supervise-definition>
- [16] <https://www.educba.com/unsupervised-machine-learning/>
- [17] <https://www.olliefgfohn.top/ProductDetail.aspx?iid=381884217&pr=39.88>
- [18] https://en.wikipedia.org/wiki/Reinforcement_learning#/media/File:Reinforcement_learning_diagram.svg

- [19] <https://arshren.medium.com/supervised-unsupervised-and-reinforcement-learning-245b59709f68#:~:text=Supervised%20learning%20requires%20a%20labeled,by%20interacting%20with%20the%20environment.>
- [20] DJOUANI AYA: Une approche Deep Learning pour le diagnostic préalable de la tuberculose . Université Mohamed Khider – BISKRA. Juin 2022
- [21] <https://www.jedha.co/formation-ia/algorithmes-deep-learning>
- [22] <https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/>
- [23] <https://datascientest.com/convolutional-neural-network>
- [24] Mesbahi Kamel, Aitouakli Thanina: Un modèle d'apprentissage automatique pour la prédiction des maladies, cas d'étude : cancer du sein. Université MOULOUD MAMMERIE DE TIZI OUZOU. septembre 2020
- [25] <http://penseeartificielle.fr/focus-reseau-neurones-convolutifs/> consulté le 15/04/2019
- [26] BELHADEF Salima et BENABDALLAH Khadidja. Segmentation des images IRM cérébrales par apprentissage profond . Université Abou Bakr Belkaïd de Tlemcen. Novembre 2020 .
- [27] <https://www.geeksforgeeks.org/ml-getting-started-with-alexnet/>
- [28] https://www.researchgate.net/figure/CNN-architecture-based-on-GoogLeNet-model_fig5_336816242
- [29] <https://maelfabien.github.io/deeplearning/xception/#>
- [30] <https://blog.paperspace.com/popular-deep-learning-architectures-alexnet-vgg-googlenet/>
- [31] <https://datagen.tech/guides/computer-vision/resnet-50/>
- [32] Baraka Hodhaifa abdelghani : Étude sur les Méthodes d'Optimisation utilisée dans l'Apprentissage Automatique . Université – Tlemcen. 27 / 10 / 2020
- [33] <https://data.mendeley.com/datasets/wmy84gzngw/1>
- [34] <https://datascientest.com> > matrice-de-confusion
- [35] <https://www.idbc.fr/tutoriel-comment-lire-une-courbe-roc-et-interpreter-son-auc/#:~:text=L'aire%20sous%20la%20courbe,que%20pour%20le%20non%2Dmalade>
- [36] <https://chirurgiefemme-paris.fr/cancer-sein/irm-dans-le-diagnostic-du-cancer-du-sein/#125>