



République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la
recherche scientifique

Université Larbi Tébessi - Tébessa

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie

Département : Mathématiques et Informatique



كلية العلوم المحيطة وعلوم الطبيعة والبيئة
FSES NV
E=MC²
FACULTÉ DES SCIENCES EXACTES
ET DES SCIENCES DE LA NATURE ET DE LA VIE

Mémoire de fin d'étude
Pour l'obtention du diplôme de MASTER
Domaine : Mathématiques et Informatique
Filière : Informatique
Option : Réseaux et sécurité informatique
Thème

**DJAGNOSTIC DU CANCER DU SEIN -
APPROCHES, TECHNIQUES, ET
APPLICATIONS : ETUDE DE CAS : LES «CAD
» : (Computer Aided Diagnosis) BASEES SUR
LA TECHNIQUE : DEEP LEARNING**

Présenté Par

Oufar Amira

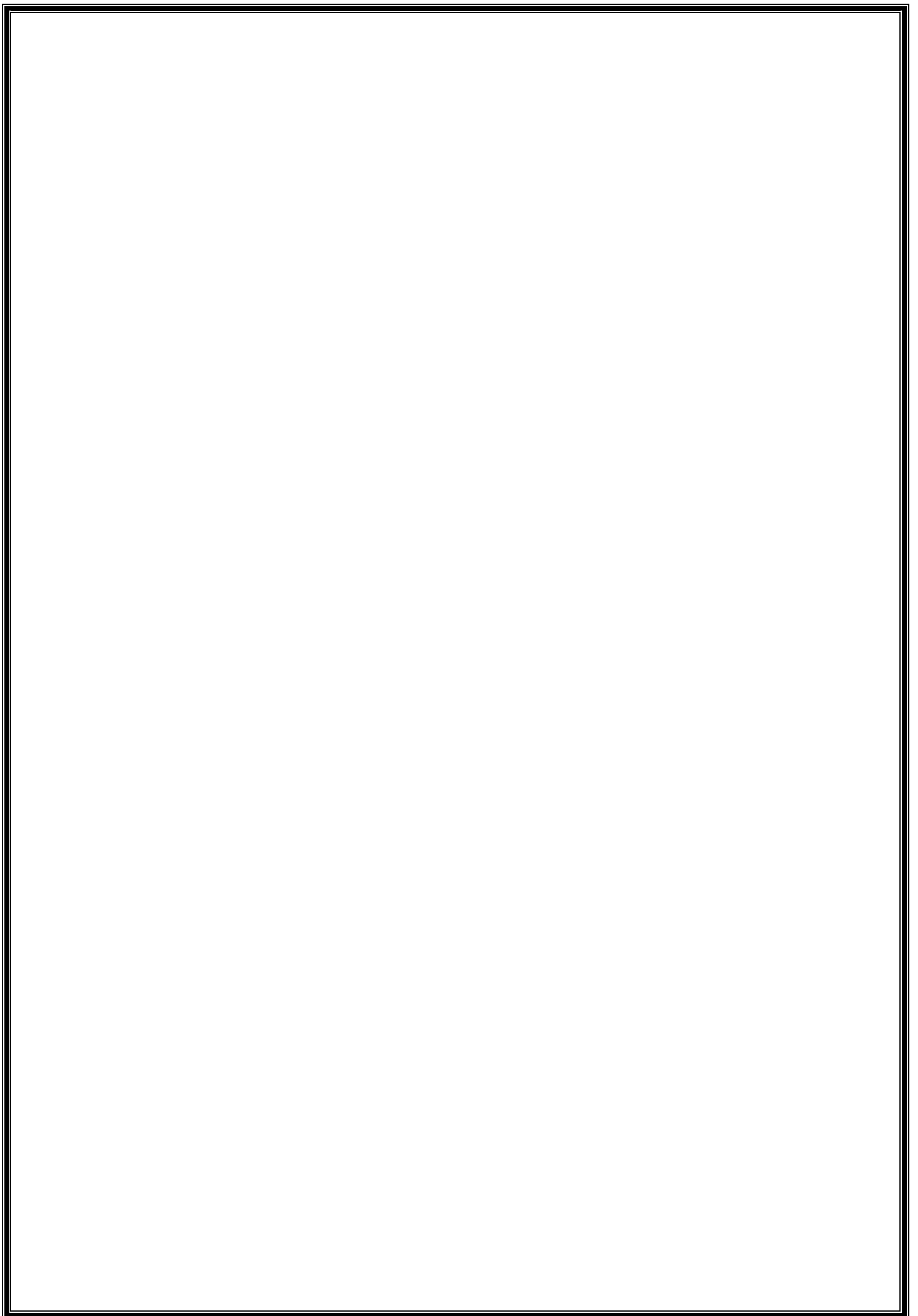
Devant le jury

Dr. Djeddi Chawki MCA Université Larbi Tébessi Président

Dr. Benour Akrem MCA Université Larbi Tébessi Examineur

Dr. Bouroqaa Salima MCB Université Larbi Tébessi Encadreur

Date de soutenance : 22/06 /2021





REMERCIEMENTS

J'adresse mes remerciements à Dieu puis à ma directrice de thèse, Mme Bourogaa Salima, je tiens à lui exprimer ma gratitude pour la confiance qu'il m'a accordée et j'espère que ce travail sera à la hauteur de ses attentes. Je le remercie également pour son encadrement exemplaire, son accessibilité et sa grande disponibilité.

Je tiens également à exprimer ma gratitude à Monsieur le Président du jury, Djeddi Shawki, et à Monsieur examinateur Benour Akram, pour l'honneur qu'ils m'ont fait en acceptant d'être porteur de ce modeste ouvrage dont les commentaires et suggestions permettront d'améliorer la qualité de ce manuscrit.

Je tiens également à remercier tous les professeurs du Département de mathématiques et d'informatique.

Je remercie également ma famille, mes amis et mes collègues.



DÉDICACE

Je dédie ce modeste travail

À Mon Cher père,

À ma chère mère,

À mes frères,

À mes enseignants,

À mes amis,

Et à mes collègues.

Résumé

Le cancer du sein est le type de cancer caractéristique chez les femmes, et le taux de mortalité de ce type augmente chaque année. Malheureusement, il n'y a aucun moyen de prévenir cette maladie, et les symptômes de cette maladie varient et n'apparaissent généralement que tard dans la maladie, donc un diagnostic précoce de la maladie aide à réduire les taux de mortalité.

Il existe un besoin urgent de développer des méthodes efficaces et intelligentes pour la détection précoce, le diagnostic et le traitement du cancer du sein.

Il ne fait aucun doute que l'expertise et la prise de décision sont des facteurs très importants. Cependant, grâce aux campagnes de prévention qui sont menées chaque année, le résultat est un ensemble de données qui nécessite du temps, des efforts et la capacité de prendre la décision finale.

Dans l'espoir d'apporter une contribution collaborative entre les médias robotiques et le domaine médical concerné par le diagnostic du cancer du sein à travers le développement de systèmes de diagnostic assisté par ordinateur (CAD), dans cette mémoire, nous avons présenté un modèle basé sur le réseau de neurones convolutifs de Deep Learning (CNN) et le Wisconsin Diagnostic Dataset pour distinguer les porteurs de cancer des autres, dont la précision était Le modèle est bon, équivalent à 96,49 %.

Mot clés : Cancer de sein, Deep Learning, CNN, Wisconsin, intelligence artificielle, CAD.

Abstract

Breast cancer is the characteristic type of cancer in women, and the death rate of this type is increasing every year. Unfortunately, there is no way to prevent this disease, and the symptoms of this disease vary and usually do not appear until late in the disease, so early diagnosis of the disease helps reduce mortality rates.

There is an urgent need to develop effective and intelligent methods for early detection, diagnosis and treatment of breast cancer.

There is no doubt that expert assessment and decision-making are very important factors. However, thanks to the prevention campaigns that are conducted every year, the result is a body of data that requires time, effort and the ability to make the final decision.

Hoping to make a collaborative contribution between robotic media and the medical field concerned with breast cancer diagnosis through the development of computer-aided diagnostic (CAD) systems, in this these we presented a model based on the convolutional neural network and the Wisconsin Diagnostic Dataset to distinguish cancer carriers from others, the accuracy of which was the model is good, equivalent to 96.49%.

Keywords:Breast cancer, Deep Learning, CNN, Wisconsin, artificial intelligence, CAD.

الملخص

سرطان الثدي هو النوع المميز للسرطان عند النساء، حيث تتزايد نسبة الوفيات من هذا النوع كل عام. لسوء الحظ، لا توجد طريقة للوقاية من هذا المرض، وتختلف أعراض هذا المرض وعادة لا تظهر إلا في وقت متأخر من المرض، لذلك فإن التشخيص المبكر للمرض يساعد في تقليل معدلات الوفيات.

هناك حاجة ملحة لتطوير طرق فعالة وذكية للكشف المبكر عن سرطان الثدي وتشخيصه وعلاجه.

ليس هناك شك في أن تقييم الخبراء واتخاذ القرار من العوامل الهامة للغاية. ومع ذلك، وبفضل حملات الوقاية التي يتم إجراؤها كل عام، فإن النتيجة هي مجموعة من البيانات التي تتطلب وقتاً وجهداً وقدرة على اتخاذ القرار النهائي.

على أمل تقديم مساهمة تعاونية بين الوسائط الآلية والمجال الطبي المعني بتشخيص سرطان الثدي من خلال تطوير أنظمة التشخيص بمساعدة الكمبيوتر (CAD)، قدمنا في هذه المذكرة نموذجاً يعتمد على الشبكة العصبية الالتفافية ومجموعة بيانات Wisconsin التشخيصية لتمييز حاملي السرطان عن الآخرين، حيث كانت دقة النموذج جيدة، تعادل 96.49%.

الكلمات المفتاحية: سرطان الثدي، التعلم العميق، الشبكة العصبية الالتفافية، Wisconsin، الذكاء الاصطناعي، CAD.

Remerciement

Dédicace

Résumé

ABSTRACT

الملخص

Sommaire

Liste des tableaux

Liste des figures

Liste d'abréviations

Introduction Générale 1

Chapitre 1 : Le cancer de sein

1.	Introduction	6
2.	La notion du cancer.....	6
2.1.	Définition.....	7
2.2.	Les causes de cancer.....	8
2.3.	Les types de cancer	9
3.	Le cancer de sein.....	10
3.1.	Définition.....	11
3.2.	Le sein.....	11
3.3.	Les facteurs de risques.....	13
3.4.	Les symptômes.....	14
3.5.	Développement de cancer de sein.....	15
3.5.1.	Stade 0	15
3.5.2.	Stade I.....	16
3.5.3.	Stade II	16
3.5.4.	Stade III.....	17
3.5.5.	Stade IV	18
3.6.	Les types	19
3.6.1.	Cancer mammaire non-invasif	19
3.6.2.	Cancer mammaire invasif.....	20
3.7.	Traitement.....	20
3.7.1.	La chirurgie.....	21
3.7.2.	La radiothérapie.....	21

3.7.3.	La chimiothérapie.....	22
3.7.4.	Les thérapies ciblées.....	22
3.7.5.	Traitement du récepteur HER2.....	22
3.7.6.	Hormonothérapie.....	23
4.	Conclusion.....	23

Chapitre 2 : Deep Learning

1.	Introduction.....	25
2.	Apprentissage automatique.....	26
2.1.	Classification.....	26
2.2.	Types de classification.....	27
2.2.1.	Classification binaire.....	27
2.2.2.	Classification multi classe.....	27
3.	Apprentissage profond.....	28
3.1.	Fonctionnement de l'apprentissage profond.....	28
4.	Les réseaux de neurones.....	30
4.1.	Les réseaux de neurones convolutifs (CNN).....	30
4.1.1.	L'opération de convolution.....	31
4.1.2.	Les couches de CNN.....	33
4.1.3.	Les paramètres de CNN.....	35
4.2.	Les réseaux de neurones récurrents (RNN).....	36
5.	Les domaines d'application d'Apprentissage profond.....	37
5.1.	Reconnaissance d'image :.....	37
5.2.	Diagnostic médical :.....	37
6.	Conclusion.....	37

Chapitre 3 : Les travaux connexe

1.	Introduction.....	40
2.	Les CADs.....	40
3.	Les approches classiques.....	41
3.1.	Auto-examens des seins et examens cliniques des seins.....	41
3.2.	Mammographie.....	42

3.3.	IRM mammaire	43
3.4.	Échographie mammaire.....	44
3.5.	PET scan.....	44
4.	Les approches intelligentes	47
4.1.	Les applications d'aide de diagnostic	47
4.1.1.	Segmentation.....	47
4.1.2.	Détection.....	48
4.1.3.	Classification.....	48
4.2.	Dataset	49
4.3	Les travaux connexes	51
5	Les problèmes et les défis.....	56
6	La synthèse.....	57
7.	Conclusion.....	58

Chapitre 4 : Conception & Réalisation

1.	Introduction	61
2.	Objective	61
3.	Conception	62
3.1.	Data set.....	64
3.1.1.	Caractéristique de dataset	64
3.1.2.	Attributs de dataset	65
3.2	Division des données	Erreur ! Signet non défini.
3.3.	Prétraitement.....	66
3.4.	Le model CNN	67
3.5.	Evaluation	68
4.	Les outils d'implémentation	70
4.1.	Google colab.....	70
4.2.	Python.....	70
4.3.	Sklearn.....	71
4.4.	Tensorflow.....	71
4.5.	Pandas	71
4.6.	Numpy	71

4.7. Keras	71
4.8. Matplotlib	72
5. Résultat et discussion	72
5.1. Résultat final	74
5.2. Classement de notre résultat	74
6. Conclusion	75
Conclusion générale & perspectives	77
Référence bibliographies	78

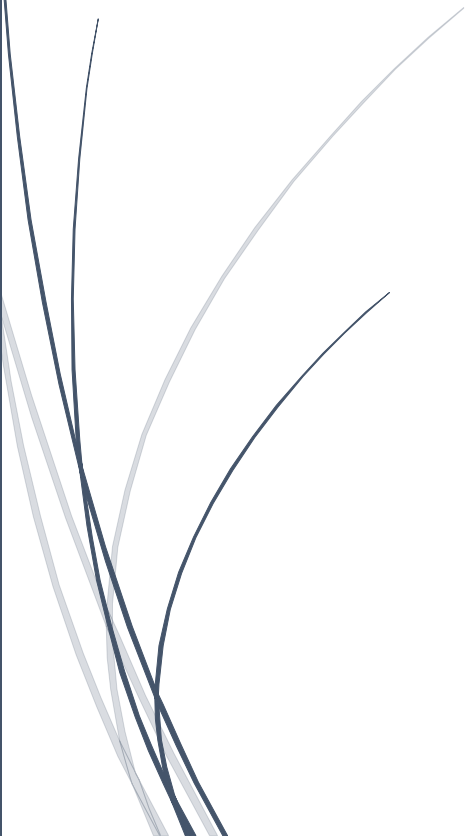


Tableau 1 : Analyse comparative des techniques d'aide de diagnostic de cancer de sein.....	46
Tableau 2 : Les jeux de données de cancer de sein.....	50
Tableau 3 :Travaux connexe	56
Tableau 4 :Les caractéristique de dataset.....	64
Tableau 5 :Les attribues de dataset	65
Tableau 6 : Comparaison des résultats.....	75

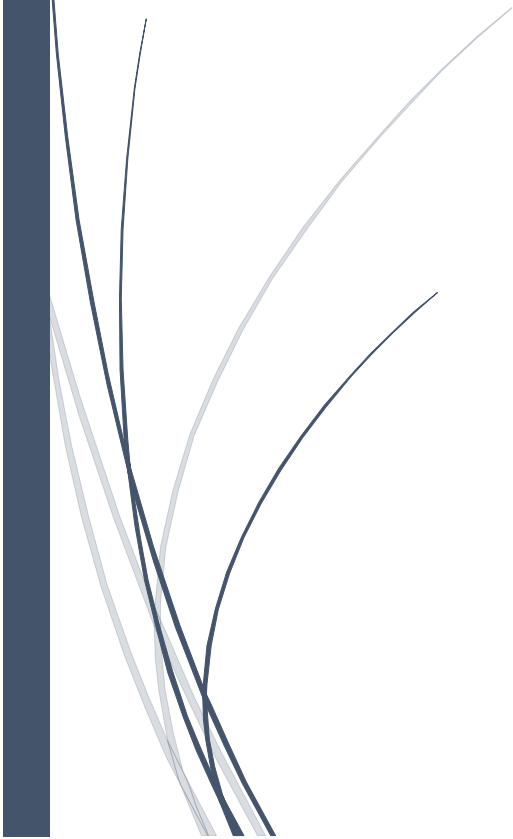


Figure 1: Taux d'incidence de tous les types de cancer pour l'année de 2020.[2]	7
Figure 2: Le processus de division cellulaire au niveau d'une cellule saine et cancéreuse [3].....	8
Figure 3: Les taux de cancer les plus élevés de tous les pays pour 2020 [2].	10
Figure 4 : Nombre de nouveaux cas chez femmes de tous âges en Algérie 2020 [2].	11
Figure 5: les composant de sein chez les femmes [8].	12
Figure 6: Stade I du cancer de sein [10].....	16
Figure 7: Stade II du cancer de sein [10].....	17
Figure 8: Stade III du cancer de sein [10].	18
Figure 9: Stade IV du cancer de sein.[11]	19
Figure 10: Les techniques de traitements du cancer se sein [14].	21
Figure11: La relation entre IA,DL &ML [15].....	25
Figure 12: La classification avec la machine Learning [16].	27
Figure 13: Classification avec le deep learning [16].....	29
Figure 14: Réseau artificielle vs réseau biologique [17].....	30
Figure 15: Architecture du CNN [18].	31
Figure 16: L'opération de convolution [18].....	32
Figure 17: Convolution avec une argument maximal [18].....	32
Figure 18 : Pooling avec un filtre 2x2 et un pas de 2 [18].	34
Figure 19 : Les fonctions d'activations.	35
Figure 20 : Architecture de RNN [21].....	36
Figure 21: Test de mammographie.....	42
Figure 22: Test IRM.	43
Figure 23: Test d'échographie.....	44
Figure 24: Test PET scan.	45
Figure 25: Processus d'atteindre notre objective	62
Figure 26: Architecture proposé	63
Figure 27: code python de division de donnés 80:20	66
Figure 28: code python de standardisation	67
Figure 29: code python de model CNN.....	68
Figure 30: Code python d'évaluation.....	69
Figure 31: Code python de matrice de confusion.....	70
Figure 32: Les outils utilisé.	72
Figure 33: Précision de model.....	73
Figure 34: Erreur de modèle.....	74
Figure 35: Matrice de confusion.....	74



Liste d'abréviations

CAD	Computer aide décision
CBE	Clinical Breast Examinations
ESB	Breast Self-Examinations
TEP	Tomographie par émission de positons
FDG	Le fluorodoxy glucose
DL	Deep Learning
DCNN	Deep Convolution Neural Network
CNN	Convolution Neural Network
SVM	Support Vector Machine
CSV	Comma-separated values
DM	Digital mammography
DBT	Digital breast tomosynthesis
UC	Université of Chicago
CNN-CT	Convolution Neural Network -Computer Tomographie
ROI	Region of interest



Introduction générale

1. Contexte et motivation

Le cancer du sein est le type distinctif de cancer chez les femmes, car le taux de mortalité de ce type augmente chaque année et représente des taux élevés partout dans le monde.

En Algérie, cette maladie représente une source de terreur pour les femmes, car elle représente environ 50% des cancers qui touchent les femmes.

Malheureusement, les programmes réguliers de dépistage du cancer du sein ne sont pas toujours disponibles et il n'y a aucun moyen de l'empêcher.

Les symptômes de cette maladie varient et n'apparaissent généralement qu'à un stade avancé de la maladie, de sorte que le diagnostic précoce et la prise en charge de la maladie à ses débuts ont été considérés comme contribuant à bien réduire les taux de mortalité.

Afin de mieux lutter contre toute maladie, en particulier le cancer, il est urgent de développer des méthodes efficaces et intelligentes de détection précoce, de diagnostic et de traitement du cancer du sein.

Il ne fait aucun doute que l'évaluation et la prise de décision par des experts sont des facteurs très importants. Cependant, grâce aux campagnes de prévention menées chaque année, il en résulte un ensemble de données qui nécessite du temps, des efforts et des capacités pour prendre la décision finale.

Espérant apporter une contribution collaborative entre les médias informatiques et le domaine médical impliqué dans le diagnostic du cancer du sein à travers le développement de systèmes de diagnostic assisté par ordinateur (CAD), qui représentent un domaine croissant d'analyse de données médicales dans le but d'aider les cliniciens à poser le bon diagnostic. Il s'agit donc d'un sujet de recherche capable de créer de nouvelles innovations.

2. Objective de mémoire

Dans cette même optique, notre objectif est de développer un système d'assistance automatisé (CAD) pour détecter et classer les porteurs du cancer du sein. Ce système est basé sur L'utilisation de l'algorithme d'apprentissage en profondeur « le CNN » pour traiter la dataset de « Breast Cancer Wisconsin (Diagnostic) ».

3. Organisation de mémoire

Afin de décrire le travail effectué tout d'abord, on va commencer par une introduction générale qui explique le domaine et le but de notre travail on suite on a devisé le mémoire en :

1. Partie 1 :

- + Chapitre 1 : Le cancer de sein
- + Chapitre 2 : Le Deep Learning
- + Chapitre 3 : Les travaux connexes

2. Partie 2 :

- + Chapitre 4 : Conception & réalisation

Et on termine par une conclusion générale et perspectives.

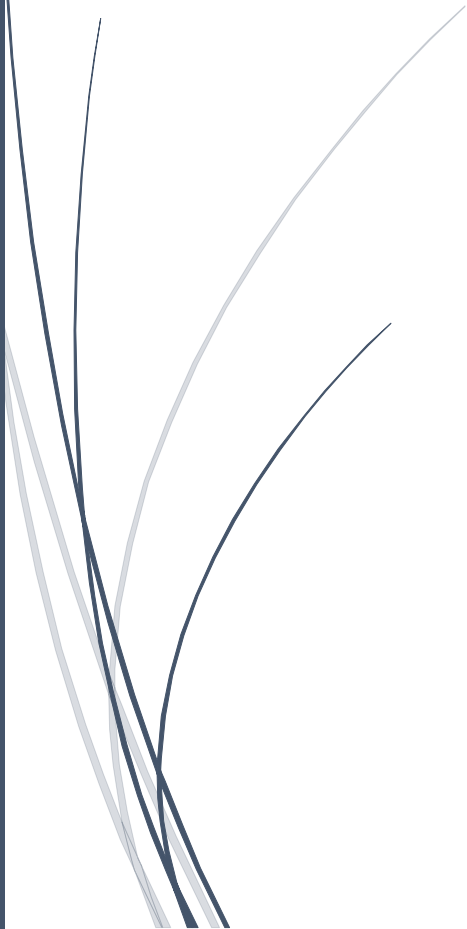


Partie 1



Chapitre 1

Le cancer de sein



1. Introduction

[Le corps humain se caractérise par contenir environ un million de milliards de cellules, qui sont constamment renouvelées et chacune d'elles a un rôle spécifique. Ils partagent la tâche d'organiser et de remodeler les tissus ou les organes. En fait, une partie des cellules meurent chaque jour et sont remplacées par de nouvelles cellules identiques. Les maladies cancéreuses sont causées par des changements dans l'une des cellules et leur prolifération se traduit par des millions de cellules présentant la même déformation. Parmi les cancers, les maladies respiratoires sont les plus courantes, suivies des cancers du côlon et du sein, trois types de cancers évitables et donnant d'excellents résultats. [1]

Le cancer du sein est le type de cancer le plus fréquemment diagnostiqué chez les femmes du monde entier. Cependant, ce n'est pas le plus mortel : le cancer du poumon reste la première cause de décès chez les femmes. En Algérie, le cancer du sein représente près de 50% des cancers féminins chez la femme et au cours des 15 dernières années, l'incidence du cancer du sein a triplé [1]. Une femme sur 9 risque de développer un cancer du sein au cours de sa vie. Rarement, les hommes sont également affectés par moins de 1%. Grâce aux progrès des diagnostics pour identifier le cancer du sein, le taux de mortalité a considérablement diminué. Dans ce chapitre, nous présentons un aperçu complet du cancer du sein, ses causes, ses phases de développements, ses types, ses éléments de risques et ses méthodes de diagnostic.

2. La notion du cancer

Le cancer est une cause majeure de décès et un obstacle important à l'augmentation de l'espérance de vie dans tous les pays du monde. Selon les estimations de l'Organisation mondiale de la santé (OMS) en 2020 (figure 1), le cancer se classe 1er et 2ème pour les décès à 70 ans dans 112 pays sur 183 et 3e et 4e pour 23 autres pays. Ces pourcentages reflètent en partie la baisse marquée des taux de mortalité due à d'autres maladies mortelles telles que les accidents vasculaires cérébraux et les maladies cardiaques dans de nombreux pays [2].

Taux d'incidence estimés normalisés selon l'âge (Monde) en 2020, tous cancers, deux sexes, tous âges

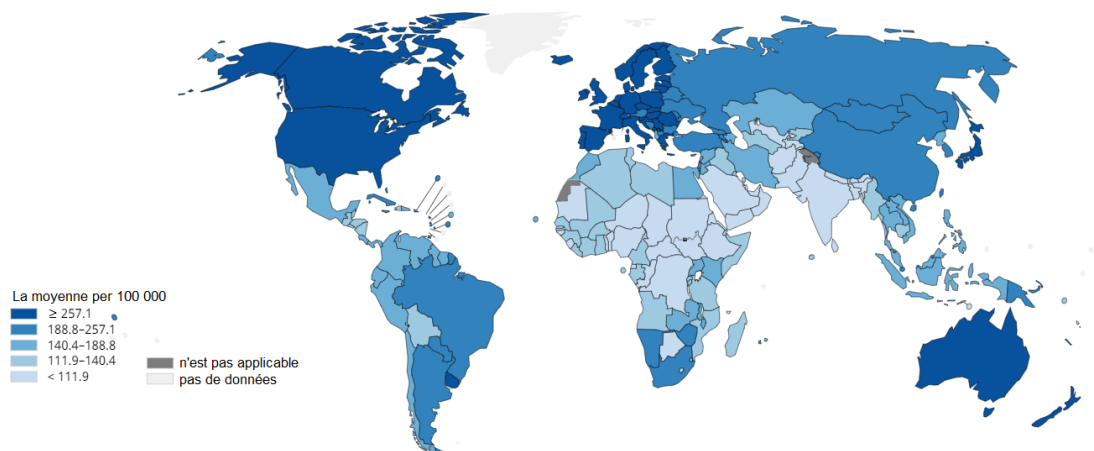


Figure 1: Taux d'incidence de tous les types de cancer pour l'année de 2020.[2]

2.1.Définition

Le cancer est défini comme une prolifération de cellules anormales, qui peuvent être trouvées dans n'importe quel organe du corps et être incontrôlables. Ils peuvent également migrer vers d'autres organes pour provoquer des métastases [3].

La figure 2 illustre une comparaison du processus de division cellulaire au niveau d'une cellule normale et cancéreuse [3]. Une tumeur peut être bénigne (sans danger pour la santé) ou maligne (potentiellement dangereuse). Les polypes ne sont pas considérés comme précancéreux : les cellules ont un aspect proche de la normale, elles se développent lentement, sans envahir les tissus environnants et sans se propager à d'autres parties du corps. À l'inverse, les tumeurs malignes sont cancéreuses, si les cellules malignes ne sont pas contrôlées, elles peuvent éventuellement se propager au-delà de la tumeur d'origine vers d'autres parties du corps.

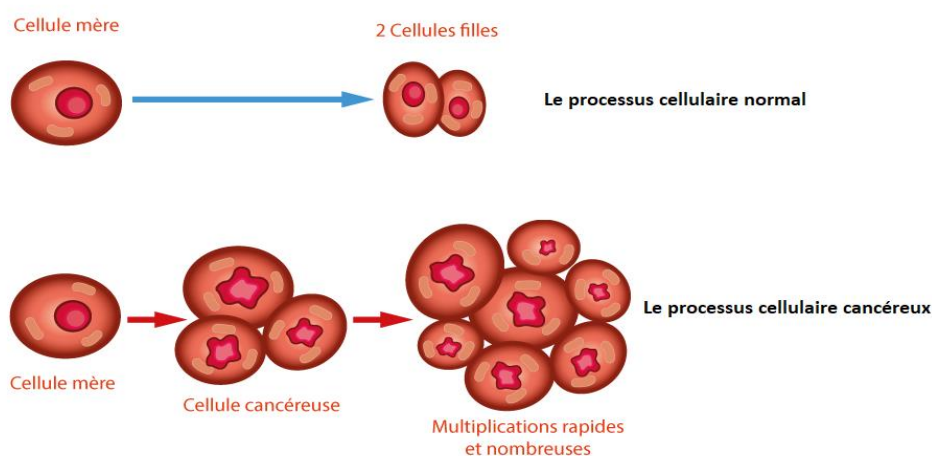


Figure 2: Le processus de division cellulaire au niveau d'une cellule saine et cancéreuse [3].

2.2. Les causes de cancer

Les causes du cancer sont nombreuses et difficiles à déterminer. Cependant, les chercheurs ont pu identifier certains facteurs qui augmentent généralement le risque de développer un cancer. Parmi eux, les facteurs non modifiables les plus importants sont :

- L'âge
- La prédisposition génétique

Certains gènes prédisposent au cancer et nous sommes plus susceptibles de développer un cancer en vieillissant. Avec l'âge, les mécanismes de reproduction cellulaire sont moins efficaces (c'est le principe du vieillissement) et cela augmente la probabilité de développer un cancer. En outre, moins de 50% des cancers sont attribués à 12 facteurs majeurs parmi les facteurs qui ont été identifiés comme potentiellement cancérogènes ou certains d'entre eux par les Centres internationaux de recherche sur le cancer [4] :

- Le tabac : 20% des cas de cancers
- L'alimentation et le surpoids et l'obésité : 10.8% des cas
- L'alcool : 8% des cas
- Les infections virales et bactériennes : 4% des cas

- Les expositions professionnelles : 3.6% des cas
- L'exposition aux Rayons ultraviolets : 3.1%
- Les radiations ionisantes : 1.9% des cas
- Une activité physique insuffisante : 0.9% des cas
- Certaines hormones : 0.8% des cas
- Le manque d'allaitement : 0.5%
- La pollution de l'air extérieur : 0.4%
- L'arsenic et le benzène : 0.1%

2.3. Les types de cancer

On parle généralement de cancer alors qu'il existe, en fait, de nombreux cancers qui diffèrent sensiblement par leur forme, leur évolution et leur traitement. Dans cette section, nous présentons les types de cancer les plus courants. En fait, il est essentiel de bien les connaître pour mieux les combattre.

Les cancers sont généralement classés en 3 catégories, à savoir les carcinomes, les sarcomes et les tumeurs hématopoïétiques [5] :

1. **Le carcinome** : les cellules cancéreuses se développent dans les tissus qui recouvrent les organes. Les tissus externes (l'épiderme) peuvent aussi être touchés. Le tissu dans lequel les cellules cancéreuses se développent et appelé épithélium. les cellules cancéreuses se développent directement dans l'épithélium d'une glande comme cela peut-être le cas avec le sein ou encore la prostate.
2. **Le sarcome** : ici, les cellules cancéreuses se développent dans les tissus de support que sont les os (ostéosarcome), la graisse (liposarcome) ou encore les muscles (rhabdomyosarcome).
3. **Le cancer hématopoïétique** : cette dernière typologie concerne les cancers du sang (leucémie) ainsi que ceux qui touchent les organes lymphoïdes (lymphome).

Certains spécialistes du cancer distinguent aussi **les tumeurs solides** des **tumeurs des cellules sanguines**. Les tumeurs solides, qui sont les plus courantes, peuvent être repérées grâce au développement d'une masse. En ce qui concerne les tumeurs des cellules sanguines, la réalité est très différente : les cellules cancéreuses sont diffusées dans le sang et la moelle osseuse.

3. Le cancer de sein

Le cancer du sein est le type de cancer le plus fréquemment diagnostiqué chez les femmes dans le monde (figure 2), représentant 1 cas de cancer sur 4. C'est le type de cancer le plus courant chez les deux sexes et la principale cause de décès par cancer chez les femmes. Selon les estimations de 2,3 millions de nouveaux cas, un cancer sur huit diagnostiqué en 2020 est un cancer du sein. En 2020, on estimait à 684 996 décès par cancer du sein, un nombre disproportionné de ces décès survenant dans des contextes sous-financés [2].

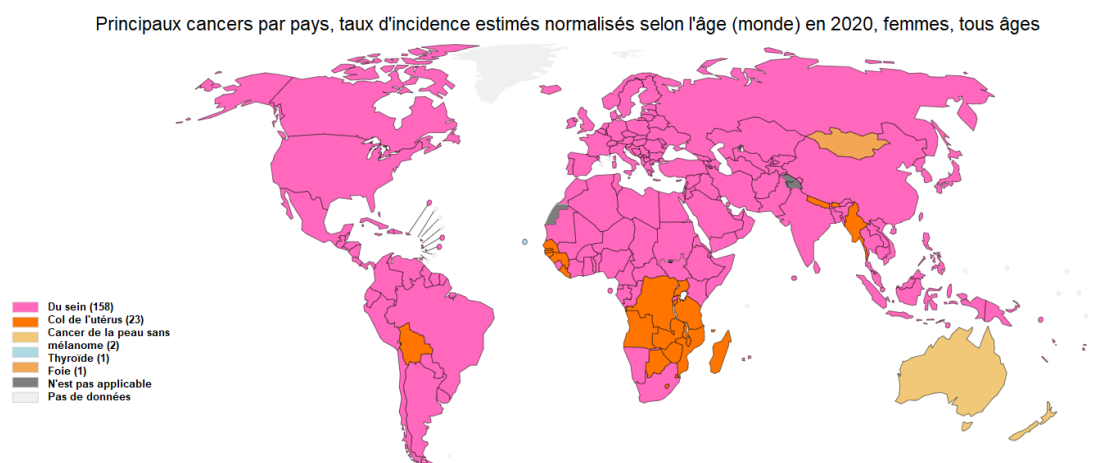


Figure 3: Les taux de cancer les plus élevés de tous les pays pour 2020 [2].

En Algérie, la même tendance est observée, le cancer de sein a classé dans la première place, à 40.3% (figure3) . L'incidence du cancer du sein est de 14,5 cas pour 100 000 habitants par an à Alger (registre des cancers d'Alger) [2]. À ce rythme, il y a 2 000 nouveaux cas chaque année ; Les deux tiers surviennent après l'âge de 45 ans. Tous stades confondus, le taux de survie à dix ans est de 50%.

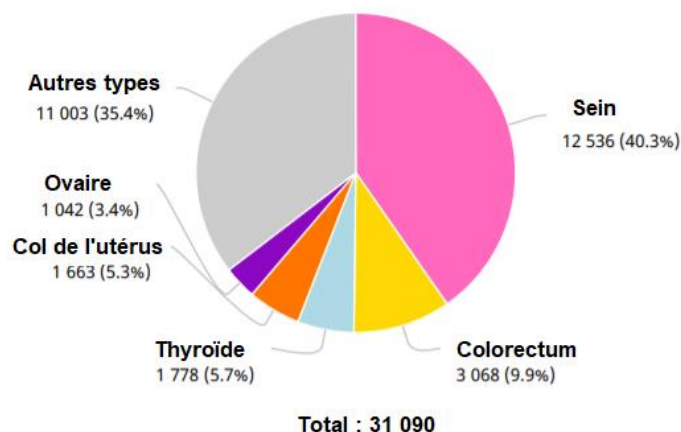


Figure 4 : Nombre de nouveaux cas chez femmes de tous âges en Algérie 2020 [2].

3.1.Définition

I.3 Description du cancer du sein

La notion de « cancer du sein » relève d'un nom générique qui désigne l'ensemble du groupe de prolifération néoplasique de la glande mammaire résultant d'une croissance anormale de cellules au niveau du tissu mammaire, qui diffèrent par leur comportement histologique et développemental [6].

Le cancer du sein est une tumeur maligne qui se développe dans les cellules de la glande. La cellule canal change son comportement habituel et commence à se diviser rapidement, car la majorité des cellules transformées à grande vitesse sont interrompues par l'apoptose. Cependant, cette mort cellulaire programmée ne se produit pas aux cellules cancéreuses. Sans apoptose, la cellule se développe de manière incontrôlable. À un stade ultérieur, les cellules atteignent les tissus voisins et risquent de se propager à d'autres parties du corps.[7]

3.2.Le sein

Les seins sont les mamelles des femmes, organe pair contenant une glande mammaire qui sécrète du lait et permet l'allaitement des nouveau-nés et nourrissons. Les seins se situent en avant des muscles pectoraux qui les soutiennent. La plus grande partie du sein est constituée

de tissu graisseux qui comble l'espace situé entre les différentes structures du sein, et il est principalement constitué de (figure 5) :

- Les ligaments (fixés au thorax) sont des bandes de tissu conjonctif soutenant les seins et traversant la peau pour joindre les muscles.
- Les lobules sont les glandes qui produisent le lait (chaque sein a de 15 à 25 lobules).
- Les canaux acheminent le lait des lobules au mamelon.
- Le mamelon, situé au centre de l'aréole, se compose de fibres musculaires.
- L'aréole, une surface ronde, entoure le mamelon.

En état normal, les tissus glandulaires subissent différents changements durant le cycle menstruel sous l'influence des hormones.

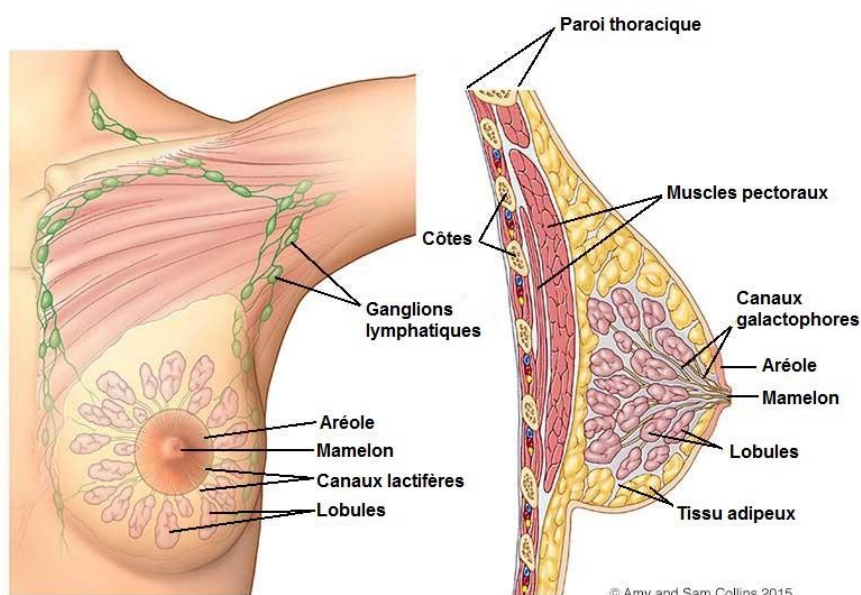


Figure 5: les composant de sein chez les femmes [8].

Le sein est composé de plusieurs vaisseaux sanguins et lymphatiques. Ces derniers sont des petits tubes qui réunissent et transportent la lymphe loin du sein et la relie à des petites masses de tissu lymphatique appelées « ganglions lymphatiques ». Ces ganglions qui entourent la région mammaire.

Le système lymphatique est composé de ganglions qui combattent les infections et qui se regroupent pour aider à évacuer la lymphe de chaque sein. On les trouve des deux côtés du corps :

- Ganglions sus-claviculaires (au-dessus de la clavicule)
- Ganglions infra claviculaires, ou sous-claviculaires (sous la clavicule)
- Ganglions mammaires internes (à l'intérieur du thorax, autour du sternum)
- Les ganglions lymphatiques axillaires situés dans l'aisselle sont échelonnés en trois niveaux selon leur proximité au muscle du thorax (grand pectoral.) Le cancer du sein se propage généralement dans les ganglions de niveau I, II et III :
 - Les ganglions de niveau I longent le côté externe du muscle (petit pectoral) qui se situe sous le grand pectoral.
 - Les ganglions de niveau II sont situés sous le petit pectoral.
 - Les ganglions de niveau III longent le côté interne du petit pectoral.

3.3. Les facteurs de risques

Il existe de nombreuses causes de cancer du sein, les facteurs de risque les plus courants sont [9]:

- **Sexe féminin** : un pourcentage quasi inexistant (moins de 1%) des cas de cancer du sein touche les hommes de plus de 60 ans.
- **Âge** : relation positive entre l'âge de la femme et le risque de cancer du sein. Plus la femme est âgée (environ 85% des cas surviennent à l'âge de cinquante ans ou plus), plus le risque d'infection est grand. La maladie touche rarement les femmes de moins de 30 ans.
- **Facteur génétique** : La présence d'un membre féminin de la famille atteint d'un cancer du sein ou d'un cancer de l'ovaire augmente le risque d'en hériter d'une génération à l'autre. Environ 10% des cas de cancer du sein sont dus à une anomalie transmise par hérédité.
- **Incidence précoce** : Avoir un cancer du sein dans le passé augmente le risque de développer un deuxième cancer.

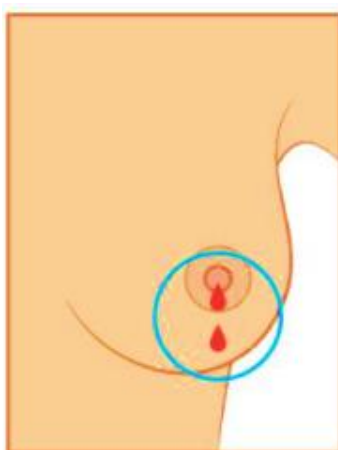
- **Maladie du sein** : les femmes présentant une lésion à risque, telle qu'une hyperplasie épithéliale intra-canaulaire atypique ou un carcinome lobulaire local, sont plus susceptibles de développer un cancer du sein à l'avenir.

3.4. Les symptômes

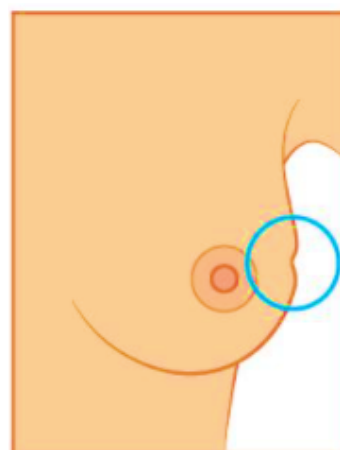
Les symptômes du cancer du sein sont presque inexistants aux premiers stades, puis ils les commencent à apparaître dans les stades avancés et se présentent généralement sous la forme d'un nodule (masse) dans le sein, d'autres symptômes se manifestent par :



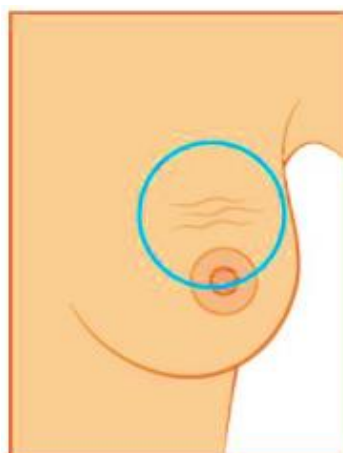
Sécrétions du mamelon autres que le lait maternel



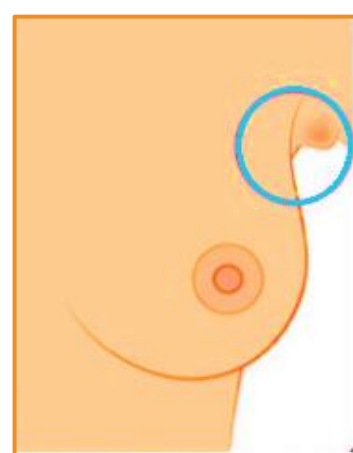
Sécrétions sanglantes du mamelon



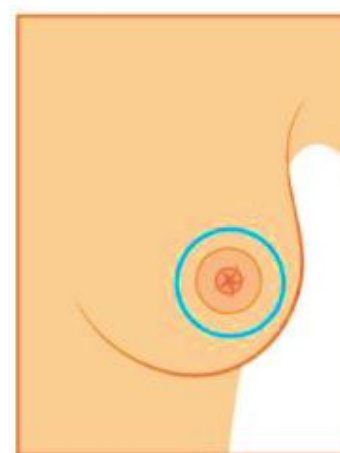
une épaisseur de tissu



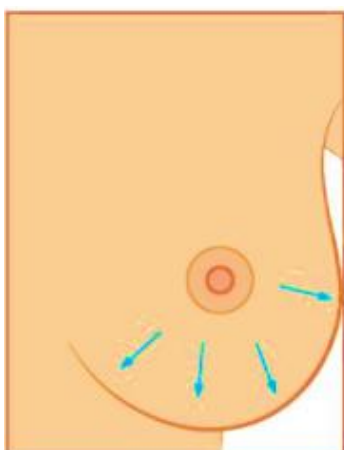
Modifications de l'aspect de la peau du sein



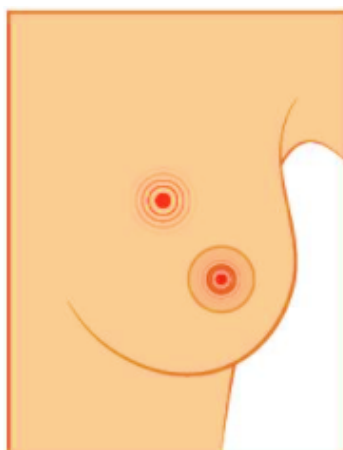
Gonflement des aisselles



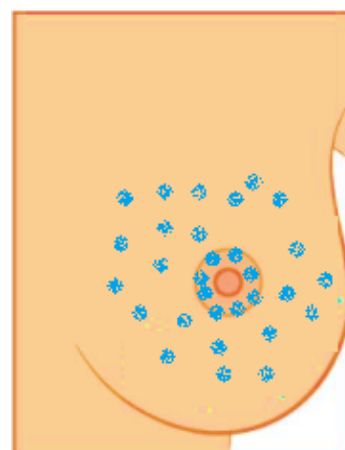
Mamelon inversé



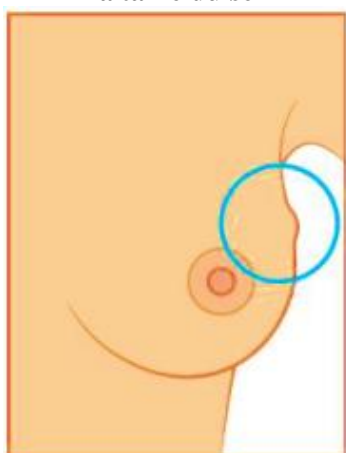
Un changement soudain et inexplicable de la forme et de la taille du sein



Douleur au sein ou au mamelon



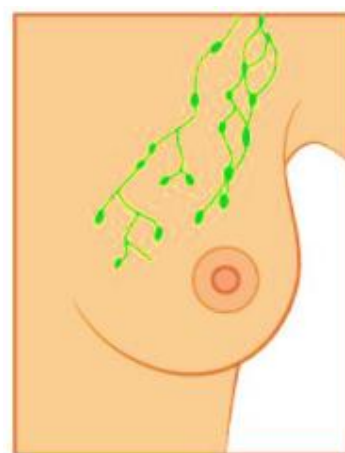
Peeling de la peau du sein ou du mamelon



Une masse mammaire



Rougeur de la peau sur tout le sein



Changement des ganglions lymphatiques

Tous ces symptômes ne sont pas synonymes de cancer car ils coïncident avec des maladies bénignes, mais leur présence nécessite l'avis d'un spécialiste.

3.5. Développement de cancer de sein

Le cancer redéveloppe 5 stages (du stage 0 a stade 4) :

3.5.1. Stade 0

Le stade 0 décrit les cancers du sein qui ne sont pas invasifs, comme le CCIS (carcinome canalaire in situ). À ce stade, il n'y a aucune preuve que les cellules cancéreuses ou les cellules cancéreuses anormales se détachent de la partie du sein où elles ont commencé.

3.5.2. Stade I

Le stade I décrit le cancer du sein invasif (les cellules cancéreuses pénètrent ou envahissent le tissu mammaire environnant normal), ce stade est divisé en sous-catégories :

Le stade IA : décrit le cancer du sein invasif dans lequel :

- La tumeur mesure jusqu'à 2 centimètres (cm)
- Et le cancer ne s'est pas propagé à l'extérieur du sein.

Le stade IB décrit un cancer du sein invasif dans lequel :

- Il n'y a pas de tumeur au sein mais de petits groupes de cellules cancéreuses ($> 0.2\text{mm}$ et $< 2\text{mm}$) se trouvent dans les ganglions lymphatiques

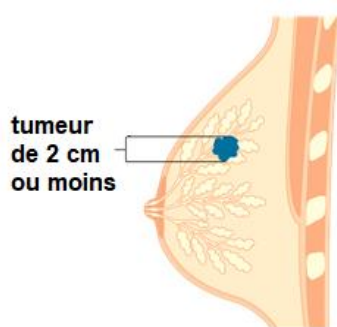


Figure 6: Stade I du cancer de sein [10]

3.5.3. Stade II

Le stade II est divisé en sous-catégories :

Le stade IIA : décrit un cancer du sein invasif dans lequel l'un des trois symptômes suivants apparaît :

- Aucune tumeur ne peut être trouvée dans le sein, mais un cancer (supérieur à 2 mm) se trouve dans 1 à 3 ganglions lymphatiques axillaires (les ganglions lymphatiques sous le bras) ou dans les ganglions lymphatiques près du sternum.
- La tumeur mesure 2 cm ou moins et s'est propagée aux ganglions lymphatiques axillaires
- La tumeur mesure dans la plage de (2 à 5 cm) et ne s'est pas propagée aux ganglions lymphatiques axillaires

Le stade IIB : décrit un cancer du sein invasif dans lequel l'un des trois symptômes suivants apparaît :

- La tumeur mesure plus de 2 cm mais pas plus de 5 centimètres ; de petits groupes de cellules cancéreuses du sein (>0.2 mm et < 2 mm)se trouvent dans les ganglions lymphatiques.
- La tumeur mesure plus de 2 cm mais pas plus de 5 cm ; le cancer s'est propagé à 1 à 3 ganglions lymphatiques axillaires ou aux ganglions lymphatiques près du sternum (détecté lors d'une biopsie du ganglion sentinelle)
- La tumeur mesure plus de 5 cm mais ne s'est pas propagée aux ganglions lymphatiques axillaires

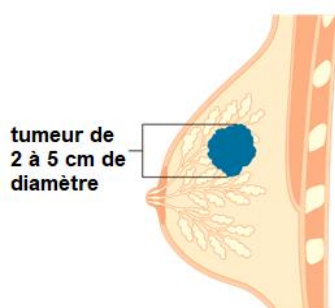


Figure 7:Stade II du cancer de sein [10].

3.5.4. Stade III

Le stade III est divisé en sous-catégories :

Le stade IIIA : décrit un cancer du sein invasif dans lequel l'un des trois symptômes suivants apparaît :

- Aucune tumeur n'est trouvée dans le sein ou la tumeur peut être de n'importe quelle taille ; le cancer se trouve dans 4 à 9 ganglions lymphatiques axillaires ou dans les ganglions lymphatiques proches du sternum (détecté lors de tests d'imagerie ou d'un examen physique)
- La tumeur mesure plus de 5 centimètres (cm) ; de petits groupes de cellules cancéreuses du sein (plus de 0,2 millimètre [mm] mais pas plus de 2 mm) se trouvent dans les ganglions lymphatiques

- La tumeur mesure plus de 5 cm ; le cancer s'est propagé à 1 à 3 ganglions lymphatiques axillaires ou aux ganglions lymphatiques près du sternum (détecté lors d'une biopsie du ganglion sentinelle)

Le stade IIIB (inflammatoire) : Les caractéristiques typiques de ce cancer comprennent :

- Rougeur d'une grande partie de la peau du sein
- Le sein est chaud et peut être enflé
- Les cellules cancéreuses se sont propagées aux ganglions lymphatiques et peuvent être trouvées dans la peau

Le stade IIIC : décrit un cancer du sein invasif dans lequel :

- Il peut n'y avoir aucun signe de cancer du sein ou, s'il y a une tumeur, elle peut être de n'importe quelle taille et peut s'être propagée à la paroi thoracique et / ou à la peau du sein
- Le cancer s'est propagé à 10 ganglions lymphatiques axillaires ou plus
- Le cancer s'est propagé aux ganglions lymphatiques axillaires ou aux ganglions lymphatiques près du sternum

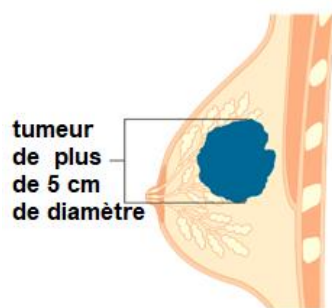


Figure 8: Stade III du cancer de sein [10].

3.5.5. Stade IV

Le stade IV (ou bien le stade avancé) (ou bien le métastase) : décrit un cancer du sein invasif qui s'est propagé au-delà du sein et des ganglions lymphatiques voisins vers d'autres organes du corps, tels que les poumons, les ganglions lymphatiques distants, la peau, les os, le foie ou le cerveau.

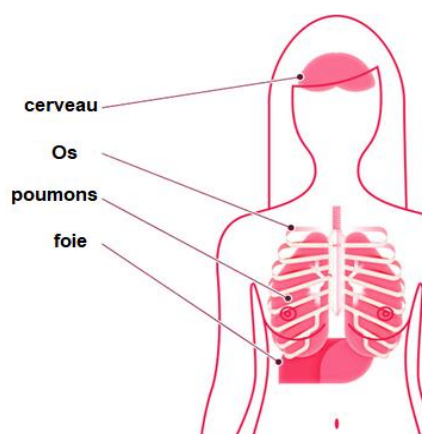


Figure 9: Stade IV du cancer de sein.[11]

3.6. Les types

Généralement, les types de cancer du sein sont divisés en deux types : Tant que les cellules cancéreuses ne traversent pas la paroi du canal, on parle de « cancer in situ » ou de « cancer du sein localisé » ou de « cancer du sein non invasif », une fois que les cellules cancéreuses migrent à travers la paroi du canal, on parle de cancer du sein invasif [12] [13].

3.6.1. Cancer mammaire non-invasif

On peut distinguer des sous types de cancer de sein non-invasifs, les plus connus sont Le carcinome canalaire in-situ (CCIS) et le carcinome lobulaire in-situ (CLIS). Ils sont caractérisés par une prolifération de cellules malignes dans la glande mammaire et les canaux, sans atteinte de la membrane basale.

Le CCIS est souvent diagnostiqué par la présence de micro calcifications à la mammographie. Il est précurseur d'un carcinome invasif et devrait donc toujours être retiré dans sa totalité. Il existe différents degrés d'agressivité qui dépendent de :

- La présence de nécrose
- Le grade de différenciation cellulaire
- La taille de la tumeur
- La présence de micro calcifications

Le CLIS n'est généralement pas palpable ni visible à la mammographie. Il est souvent découvert fortuitement dans du tissu excisé pour une raison autre. Le CLIS est généralement présent à plusieurs endroits du même sein ou même des deux seins. En présence d'un CLIS, le risque de développer un cancer invasif est d'environ 37%.

3.6.2. Cancer mammaire invasif

Les cellules des cancers invasifs ont la capacité d'envahir les tissus environnants et peuvent causer des métastases à distance dans les ganglions lymphatiques ou dans d'autres organes. Les types les plus fréquents sont :

- Le carcinome canalaire invasif (CCI) : cette tumeur se présente généralement comme une masse indurée du sein et elle est fréquemment entourée par un CCIS.
- Le carcinome lobulaire invasif (CLI) : se présente souvent comme un épaissement mal défini du sein parfois difficile à détecter, la mammographie et l'échographie ne pouvant pas toujours clairement mettre en évidence cette tumeur. Chez 90% des patients atteints d'un CLI, un CLIS est également retrouvé dans la région. Le CLI est quasiment toujours fortement hormono-sensible et présente un risque conséquent de se développer dans les deux seins.
- Les types de cancers du sein invasifs moins fréquents comprennent : le carcinome médullaire, le carcinome mucineux ou colloïde et le carcinome papillaire.

3.7. Traitement

Le traitement du cancer du sein dépend du stade de la maladie et de type de cancer. Les techniques les plus utilisées sont : la chirurgie, la radiothérapie, la chimiothérapie et les thérapies ciblées. Ces traitements peuvent être utilisés seuls ou associés les uns aux autres pour objectifs de supprimer la tumeur, de réduire le risque ou de ralentir le développement de la tumeur .

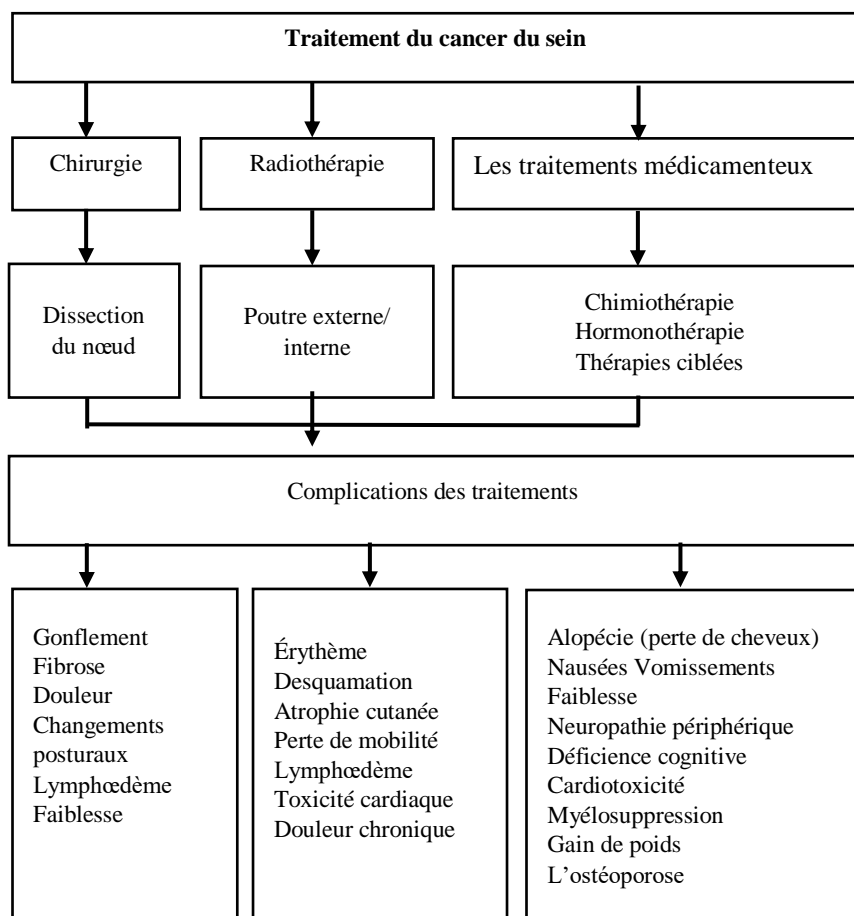


Figure 10: Les techniques de traitements du cancer se sein [14].

3.7.1. La chirurgie

La chirurgie reste l'acte principal. Dans un grand nombre de cas, l'ablation du sein (mastectomie) peut être évitée et on enlève alors largement la tumeur en gardant le reste de la glande mammaire.

3.7.2. La radiothérapie

La radiothérapie est un traitement locorégional utilisant des rayons à haute énergie pour détruire les cellules tumorales et les empêcher de se multiplier. La radiothérapie est appliquée généralement après la chirurgie pour détruire d'éventuelles cellules tumorales résiduelles et pour diminuer le risque de récurrence locale.

3.7.3. La chimiothérapie

La chimiothérapie est un traitement médicamenteux administré par voie orale ou injectable et qui utilise une ou plusieurs molécules anti-tumorales : on parle de protocole de chimiothérapie.

La chimiothérapie peut être administrée avant ou après la chirurgie.

Le plus souvent, elle est prescrite en situation adjuvante, c'est-à-dire après l'opération, pour limiter le risque de récurrence.

Cette indication dépend d'un certain nombre de facteurs qui permettent de prédire le risque de récurrence de la maladie « facteurs pronostiques » : nombre de ganglions envahis, stade d'évolution de la tumeur, âge de la patiente...

Lorsque le cancer du sein est au stade métastatique, la chirurgie n'est pas systématique et la chimiothérapie est le traitement de référence, seul ou en association avec d'autres traitements médicamenteux (hormonothérapie, thérapies ciblées).

Tous les médicaments utilisés ne provoquent pas de chute des cheveux et il faut savoir que les effets secondaires tels que nausées ou vomissements sont de mieux en mieux maîtrisés par les antiémétiques modernes.

3.7.4. Les thérapies ciblées

Les thérapies ciblées forment une classe innovante de médicaments anticancéreux : à l'inverse de la chimiothérapie, ils agissent spécifiquement sur les cellules cancéreuses en ciblant une caractéristique propre à celles-ci,

Ce type de traitement permet en principe d'obtenir une efficacité, tout en réduisant le risque d'effets secondaires.

3.7.5. Traitement du récepteur HER2

Certains cancers du sein (1 sur 5 environ) sont plus agressifs, car les cellules de la tumeur ont à leur surface, des récepteurs dénommés HER2.

On dispose aujourd'hui de deux traitements ciblés sur ce récepteur HER2 : le trastuzumab et le lapatinib, qui sont capables de bloquer les récepteurs HER2.

3.7.6. Hormonothérapie

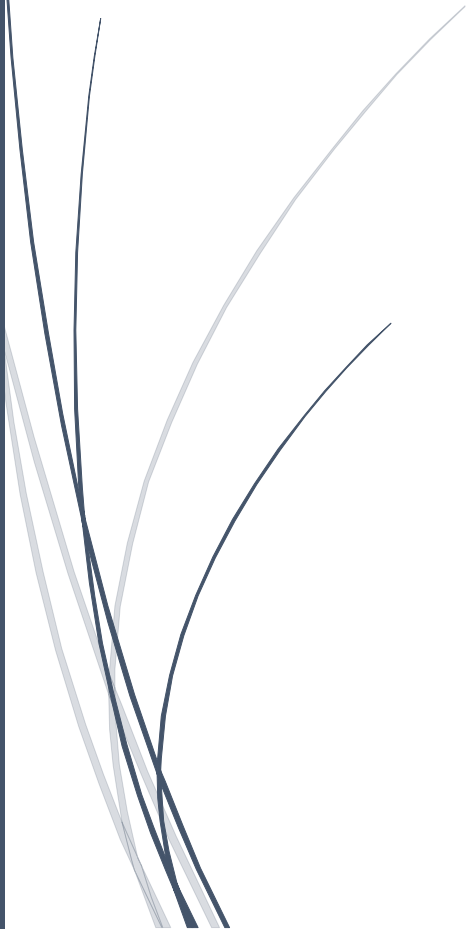
L'hormonothérapie est un traitement administré par voie orale (complète la chimiothérapie), elle a pour objectif de limiter le risque de récurrence en empêchant les cellules tumorales de recevoir les hormones (œstrogènes et/ou progestérone) nécessaires à leur développement.

4. Conclusion

Le cancer du sein est la maladie du siècle, l'incidence des femmes dans le monde augmente chaque jour. De nombreuses études dans diverses disciplines contribuent à fournir des méthodes de diagnostic et de traitement, mais la guerre contre le cancer n'est pas encore terminée. Dans ce chapitre, nous avons présenté un aperçu complet de la maladie du siècle et dans le chapitre suivant, nous traiterons d'une étude concernant la technologie « Deep Learning » afin de découvrir dans le troisième chapitre les approches basées sur cette technologie pour l'aide au diagnostic de cancer du sein.

Chapitre 2

Deep Learning



1. Introduction

Malgré la popularité des technologies d'apprentissage automatique, beaucoup d'entre nous sont incapables de faire la distinction entre l'intelligence artificielle (IA), l'apprentissage automatique (ML) et l'apprentissage en profondeur (DL), mais si les trois termes sont souvent utilisés de manière interchangeable, ils ne font pas référence aux mêmes choses. Voici une image qui tente de visualiser la distinction ou la relation entre eux :

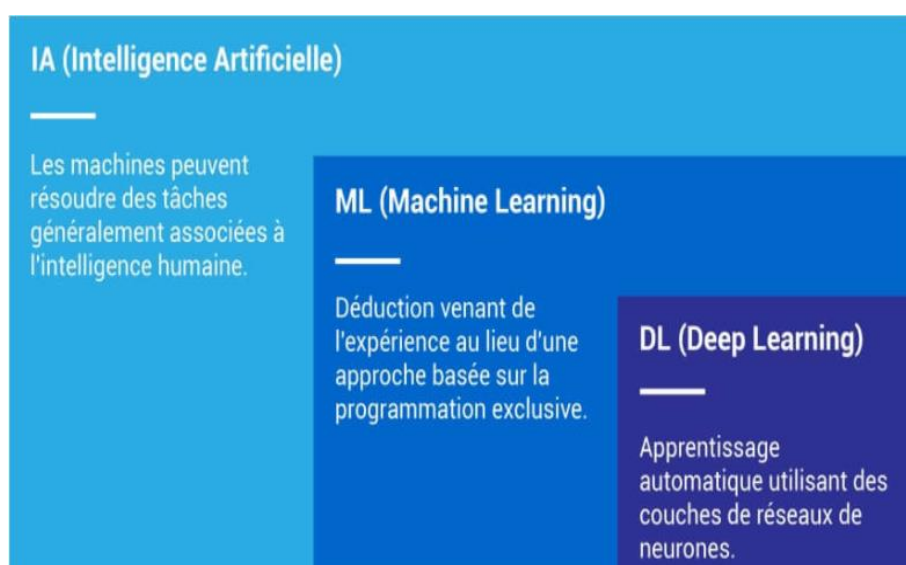


Figure 11 : La relation entre IA, DL & ML [15]

L'apprentissage automatique se concentre uniquement sur la résolution des problèmes concrets. Il faut également quelques idées de l'intelligence artificielle. L'apprentissage automatique passe par les réseaux de neurones conçus pour imiter les capacités de prise de décision de l'homme. Les outils et techniques de ML sont les deux sous-ensembles étroits centrés uniquement sur l'apprentissage en profondeur. Nous devons l'appliquer pour résoudre tout problème qui nécessite une réflexion, qu'il soit humain ou artificiel. Tout réseau de neurones profonds comprendra trois types de couches : la couche d'entrée, la couche cachée et

la couche de sorties Nous utilisons un algorithme machine pour analyser les données, en tirer des leçons et prendre des décisions éclairées en fonction de ce qu'elles ont appris. Fondamentalement, l'apprentissage en profondeur est utilisé en couches pour créer un « réseau de neurones artificiel » capable d'apprendre et de prendre des décisions intelligentes par lui-même. Nous pouvons dire que l'apprentissage en profondeur est un sous-domaine de l'apprentissage automatique.

Différentes architectures de réseau de neurone sont utilisées dans DL, chacune avec ses propres avantages et inconvénients. Les réseaux de neurones convolutionnels ou CNN sont conçus pour extraire automatiquement les caractéristiques des images et aussi des texte (1D) d'entrée, sont invariants à de légères distorsions de l'image, et implémentent la notion de partage des poids permettant de réduire considérablement le nombre de paramètres du réseau.

2. Apprentissage automatique

L'apprentissage automatique (machine Learning) est le champ de l'IA qui permet à une machine d'apprendre, d'améliorer progressivement ses performances sur une tâche spécifique en se basant sur des données, le tout sans être explicitement programmé pour résoudre cette tâche[16].

2.1. Classification

La classification est une méthode qui consiste à diviser un grand groupe de données en un certain nombre de catégories souhaitées et distinctes, dans lesquelles on peut attribuer une classification à chaque catégorie. En d'autres termes, la classification appartient à la catégorie de l'apprentissage supervisé, qui présente l'avantage de fournir les objectifs des données d'entrée. Les applications de classification sont présentées dans de nombreux domaines, notamment l'accréditation de crédit, le diagnostic médical, le marketing ciblé, etc.

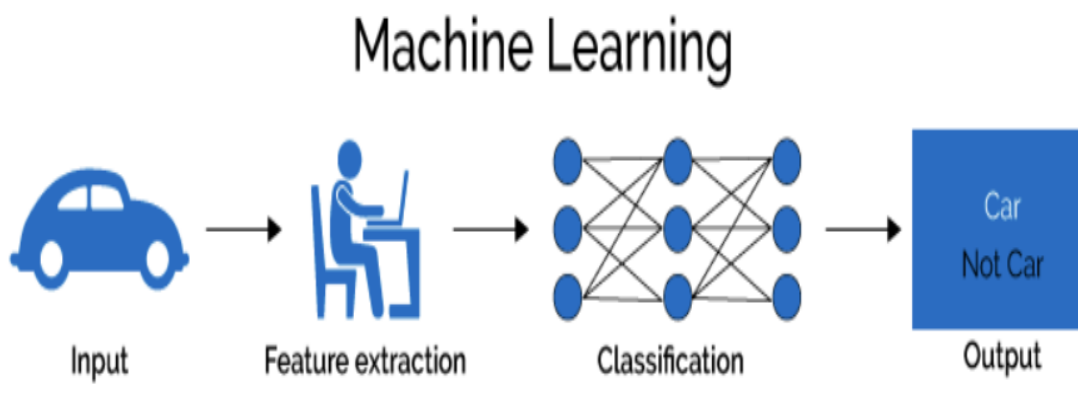


Figure 12: La classification avec la machine Learning [16].

La classification est obtenue en classant les objets dans des sous-catégories à partir d'un groupe général, mais dans machine Learning, la classification donner un problème d'identification une nouvelle observation à quel groupe de catégories (sous-populations) appartient, sur la base d'un ensemble de données de formation contenant des observations et dont les catégories sont connues.

2.2. Types de classification

2.2.1. Classification binaire

Cette classification est suivie si nous devons diviser et classer les données disponibles en deux catégories différentes (en identifiant le groupe auquel chacune appartient) et la base de la règle de reclassification. Inclut les contextes nécessitant de décider si l'élément a une propriété spécifique ou une classification binaire typique, par exemple, si une personne est malade et que l'on souhaite diagnostiquer le type de maladie, "x" s'il est infecté ou non.

2.2.2. Classification multi classe

Ici, le nombre de classes devrait être supérieur à deux, en fonction du large éventail de données

disponibles. Dans l'apprentissage automatique, la classification de classes multiples ou multi frontières est un problème de classification dans l'une ou plusieurs des trois catégories, en

dépit de l'existence d'algorithmes de classification autorisant l'utilisation de plus de deux classes.

3. Apprentissage profond

L'apprentissage en profondeur (Deep Learning) est un type d'apprentissage automatique (machine Learning) qui vise à former et à enseigner à l'ordinateur afin d'exercer des fonctions humaines telles que la distinction des objets visuels et l'identification des sons et des images. Au lieu d'organiser des données, l'apprentissage en profondeur définit ses propres paramètres de base, ce qui permet à la machine d'apprendre par elle-même, et l'intérêt actuel pour l'apprentissage en profondeur est en partie dû au bourdonnement qui entoure l'intelligence artificielle. Les techniques d'apprentissage en profondeur ont amélioré la capacité de classer, reconnaître, détecter, etc.

3.1. Fonctionnement de l'apprentissage profond

L'apprentissage en profondeur change votre façon de penser aux problèmes que vous résolvez avec l'analyse. Cela va de dire à l'ordinateur comment résoudre un problème, puis de le former à la résolution du problème lui-même. Une approche classique de l'analyse consiste à utiliser les données disponibles pour concevoir des entités afin d'en déduire de nouvelles variables, puis sélectionner un modèle analytique et enfin estimer les paramètres (ou les inconnues) de ce modèle. Ces techniques peuvent donner des systèmes prédictifs qui ne se généralisent pas bien, car leur exhaustivité et leur exactitude dépendent de la qualité du modèle et de ses caractéristiques. L'ajout de plus de données nécessite de tout recommencer. La nouvelle approche basée sur l'apprentissage en profondeur consiste à remplacer la formulation et la spécification du modèle par des caractérisations hiérarchiques (ou couches) qui apprennent à reconnaître les caractéristiques latentes des données à partir des régularités des couches. Le changement de paradigme avec l'apprentissage en profondeur passe de l'ingénierie des fonctionnalités à la représentation des fonctionnalités. Pour mieux comprendre comment le faire, prenons un exemple simple : nous voulons identifier une voiture parmi un grand groupe de voitures qui sont filmées, quelle que soit la

manière dont ils sont photographiés. À cette fin, une vaste collection d'images différentes de voitures et d'images d'autres objets doit être assemblée afin de former le réseau de neurones artificiels à la pratique de l'apprentissage en profondeur. Ces images sont converties en données et transmises au réseau après que le poids soit attribué à différents éléments par les cellules nerveuses artificielles. Les informations sont ensuite collectées par la dernière couche de neurones pour déterminer s'il s'agit d'une voiture ou non. Ensuite, le réseau de neurones artificiels comparera ce résultat à la réponse fournie. S'il y a correspondance, le réseau enregistrera ce succès, car il l'utilisera dans d'autres communes. Si le résultat est négatif, le réseau reconnaîtra le déséquilibre et ajustera le poids des différents neurones pour corriger son erreur. Le processus se fait des milliers de fois jusqu'à ce que l'image soit reconnue et appelée ce type d'apprentissage par "apprentissage supervisé". La seconde méthode est l'apprentissage non supervisé, qui repose sur des données non classifiées, c'est-à-dire que le réseau doit identifier des modèles dans les ensembles de données. Pour apprendre par eux-mêmes que tous les éléments de l'image peuvent être pertinents

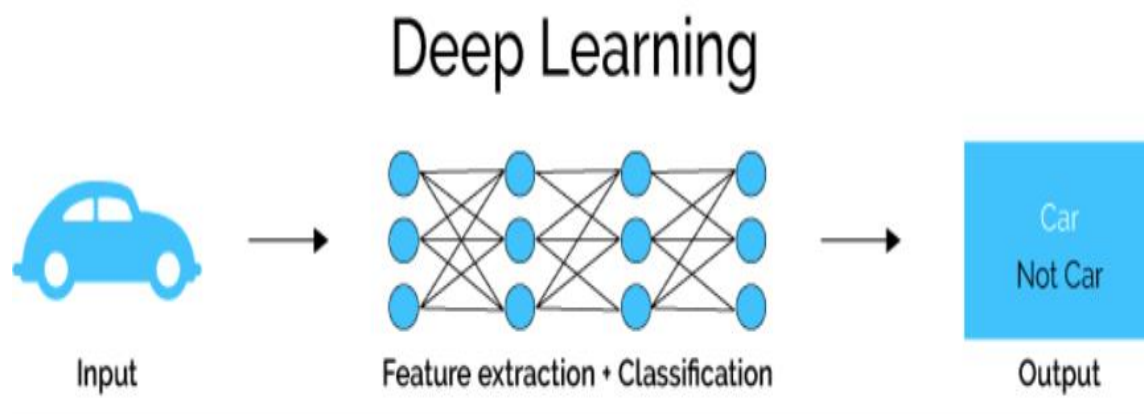


Figure 13: Classification avec le deep learning [16].

4. Les réseaux de neurones

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

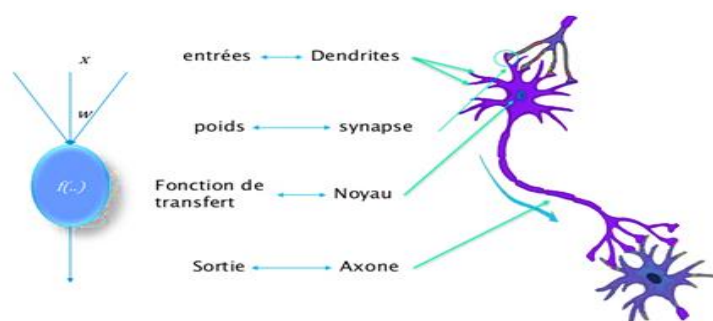


Figure 14: Réseau artificielle vs réseau biologique [17]

Deux grandes familles d'architectures les plus utilisées en réseaux de neurones profonds sont les réseaux de neurones convolutifs et Les réseaux de neurones récurrents :

4.1. Les réseaux de neurones convolutifs (CNN)

Le réseau de neurones convolutifs (CNN) est un type de réseau de neurones artificiels, il est largement utilisé dans le domaine de l'analyse d'images et la reconnaissance vidéo. CNN présente l'avantage d'utiliser des filtres de noyau (kernel) et mettent en œuvre du pooling.[18]

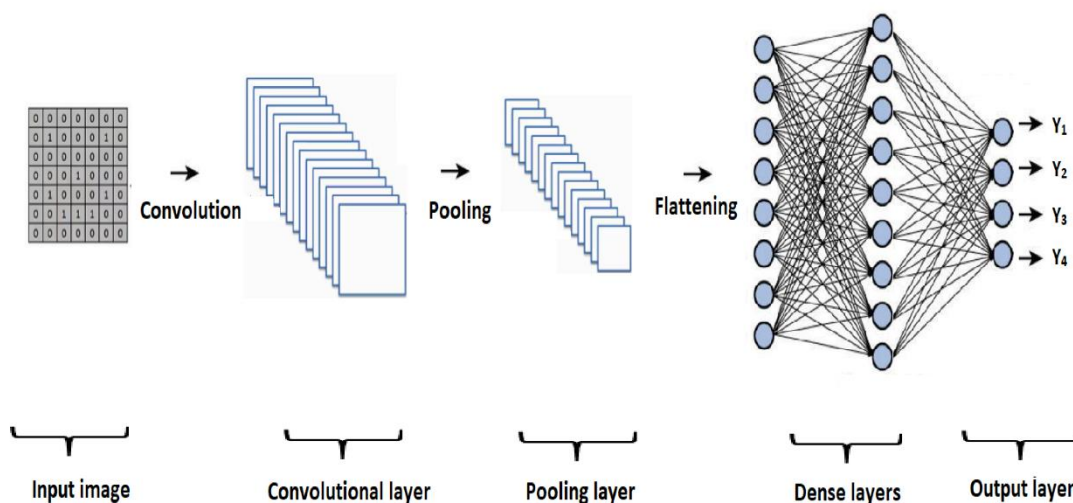


Figure 15: Architecture du CNN [18].

4.1.1. L'opération de convolution

La convolution agit comme un filtrage. On définit une taille de fenêtre qui va se balader à travers toute l'image (rappelez-vous qu'une image peut être vue comme étant un tableau). Au tout début de la convolution, la fenêtre sera positionnée tout en haut à gauche de l'image puis elle va se décaler d'un certain nombre de cases (c'est ce que l'on appelle le pas) vers la droite et lorsqu'elle arrivera au bout de l'image, elle se décalera d'un pas vers le bas, ainsi de suite jusqu'à ce que le filtre ait parcouru la totalité de l'image : [18]

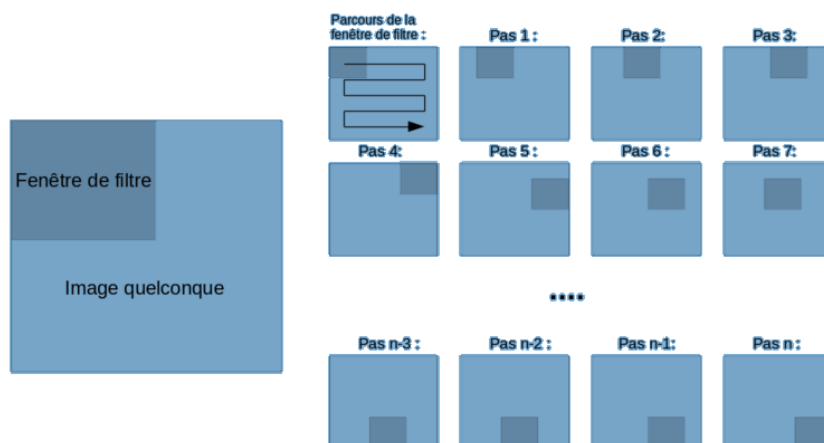


Figure 16 : L'opération de convolution [18].

Le but est de se servir des valeurs présentes dans le filtre à chaque pas. Par exemple si l'on définit une fenêtre 3 par 3, cela représentera 9 cases du tableau (9 pixels). La convolution va effectuer une opération avec ces 9 pixels. Il peut s'agir de n'importe quelle opération, par exemple on extrait l'avaleur la plus grande (soit le pixel avec la plus grande valeur).

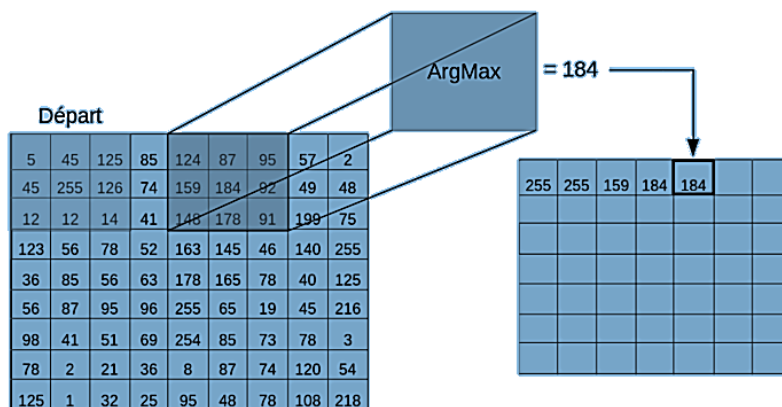


Figure 17: Convolution avec un argument maximal [18].

On fait glisser la fenêtre en bleu foncé et à chaque pas on récupère la valeur la plus grande parmi les 9 valeurs de pixels. [18]. La sortie de cette opération (convolution) qui s'appelle « matrice de caractéristique » à des dimensions plus petites que celle de l'image en entrée.

4.1.2. Les couches de CNN

La structure de CNN consiste en un ensemble de couches de traitement indépendantes :

1. Couche convolution

Cette couche est le composant principal des réseaux de neurones convolutifs et constitue leur première couche. Sa fonction est de localiser un ensemble de caractéristiques (features) dans les images qui ont été données sous forme des entrées, raison pour laquelle nous effectuons un processus appelé méta-filtrage, dans lequel nous faisons glisser la fonction d'image dans une forme défenestrée calculons le produit de convolution entre la feature et chaque portion de l'image balayée. Pour mesurer le volume de la couche de convolution, nous avons besoin de 3 hyper paramètres :

1) **Profondeur de la couche :**

Le nombre de neurones qui ont une corrélation directe avec le même champ du récepteur.

2) **Le pas :**

Le pas contrôle le chevauchement des champs récepteurs. Plus le pas est petit, plus les champs récepteurs se chevauchent et plus le volume de sortie sera grand.

3) **La marge (à 0) ou zéro padding :**

Parfois, il est commode de mettre des zéros à la frontière du volume d'entrée. Cette marge permet de contrôler la dimension spatiale du volume de sortie[19].

2. Couche de pooling (POOL)

La chose la plus commune à propos de cette couche est qu'elle est placée entre deux couches de convolution, ce qui lui permet de recevoir de nombreuses feature maps en entrée, l'opération de pooling consiste à minimiser la taille de l'image et à en préserver les propriétés importantes.

Parmi les caractéristiques connues de pooling est de donner une grande puissance de calcul, et malgré la très faible représentation (et donc la perte d'informations associées), toutes les tendances actuelles consistent à utiliser des filtres de petite taille (type 2×2). Mais le risque

de sur apprentissage posait de nombreux problèmes qui permettaient très bien d'éviter la couche de pooling.

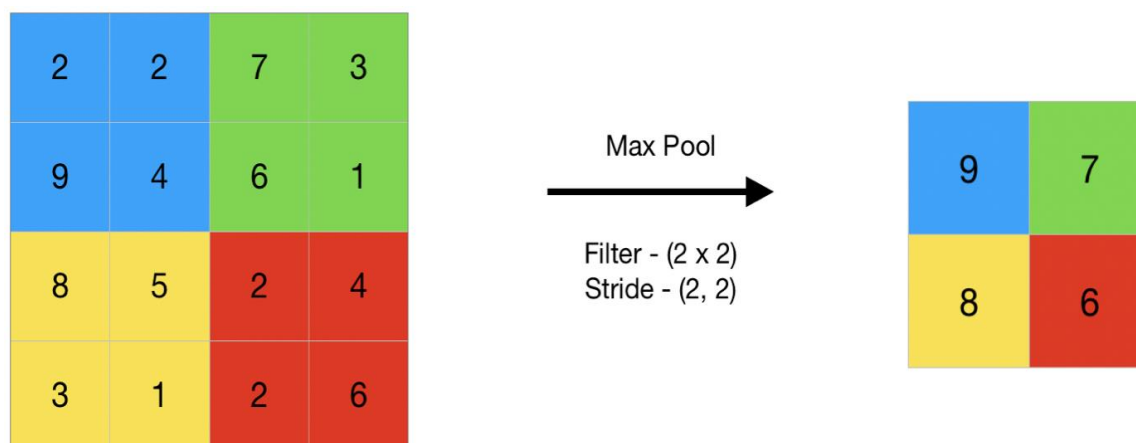


Figure 18 : Pooling avec un filtre 2x2 et un pas de 2 [18].

3. Couche de correction

Les fonctions d'activation tel que : Relu, Sigmoid, PRelu (voir Figure 19), on a but pour améliorer l'efficacité du traitement en. La fonction ReLU (Rectified Linear Units) la plus utilisé, désigne la fonction réelle non-linéaire, augmente les propriétés non linéaires de la fonction de décision elle remplace toutes les valeurs négatives reçues en entrées par des zéros.

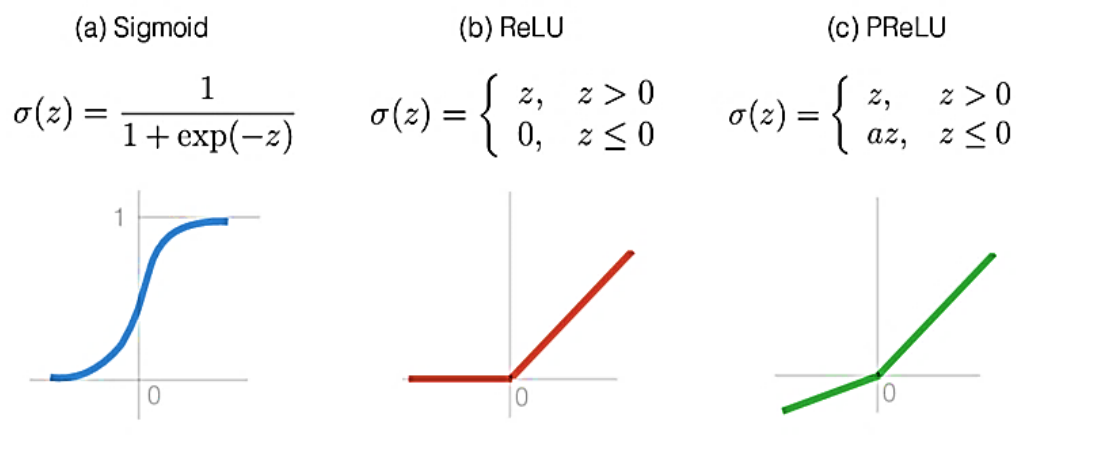


Figure 19 : Les fonctions d'activations.

4. Couche entièrement connectée

Ce type de couche reçoit un vecteur d'entrée et produit un nouveau vecteur de sortie. Pour cela, il applique une combinaison linéaire et éventuellement une fonction d'activation aux valeurs reçues en entrée.

5. Couche de perte (LOSS)

La dernière couche dans le réseau, spécifie comment l'entraînement du réseau pénalise l'écart entre le signal prévu et réel. Diverses fonctions de perte adaptées à différentes tâches peuvent y être utilisées. La perte « Softmax » est utilisée pour prédire une seule classe parmi K classes mutuellement exclusives. [20]

4.1.3. Les paramètres de CNN

1. Nombre de filtres

En raison de la réduction de la taille des images intermédiaires en fonction de la profondeur du traitement, le nombre de filtres dans les couches proches de l'entrée est réduit. Il est très important de conserver le nombre de sorties intermédiaires (c'est-à-dire de multiplier le nombre d'images intermédiaires par le nombre de positions de pixels) afin de préserver les informations d'entrée.

2. Forme du filtre

Les formes de filtrage sont souvent choisies en fonction de l'ensemble de données, en raison de la grande diversité de leurs formes dans la littérature. Les meilleurs résultats sont ceux obtenus sur MNIST (28×28) dans une plage de 5×5 dans la première couche et 2×2 pour la couche de Pooling. Des volumes d'entrée très importants peuvent justifier une mise en commun de 4×4 dans les couches inférieures. Cependant, le choix de formes plus grandes réduira considérablement la dimension du signal et peut entraîner une perte excessive d'informations

4.2. Les réseaux de neurones récurrents (RNN)

Un réseau de neurones récurrents (RNN) est une classe de réseaux de neurones artificiels où les connexions entre les nœuds forment un graphe dirigé le long d'une séquence. Cela lui permet de représenter un comportement dynamique temporel pour une séquence temporelle. Les RNN peuvent utiliser leur état interne (mémoire) pour traiter des séquences d'entrées. Cela les rend applicables à des tâches telles que la reconnaissance de l'écriture manuscrite ou la reconnaissance vocale connectée, non segmentées.

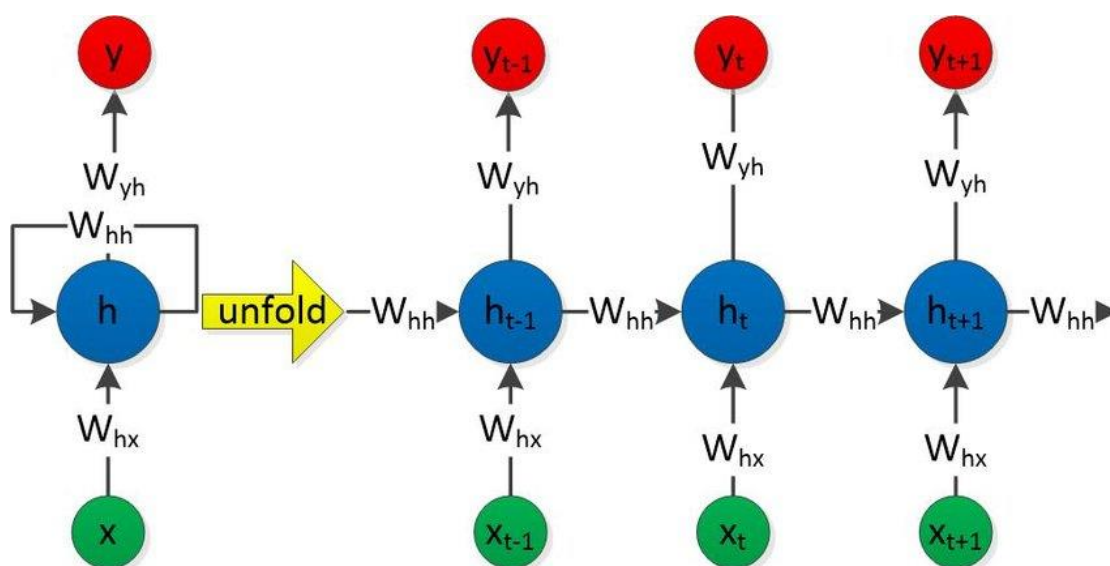


Figure 20 : Architecture de RNN [21].

Les RNN ont connu un grand succès dans les tâches d'étiquetage et de prédiction de séquences, tel que la reconnaissance de l'écriture manuscrite et la modélisation du langage. Dans la modélisation acoustique pour la reconnaissance de la parole, cependant, où les réseaux de neurones profonds (Deep Neural Network) sont l'état de la technique établi, les RNN ont récemment reçu peu d'attention au-delà des tâches de reconnaissance téléphonique à petite échelle [22].

5. Les domaines d'application d'Apprentissage profond

L'apprentissage profond est appliqué dans différents domaines, les domaines les plus proches de notre travail sont [23] :

5.1. Reconnaissance d'image :

Il est utilisé pour identifier des objets, des personnes, des lieux, des images numériques, etc. Le cas d'utilisation populaire de la reconnaissance d'image et de la détection de visage est la suggestion de marquage automatique des amis.

5.2. Diagnostic médical :

En science médicale, l'apprentissage automatique est utilisé pour le diagnostic des maladies. Avec cela, la technologie médicale se développe très rapidement et est capable de construire des modèles 3D qui peuvent prédire la position exacte des lésions dans le cerveau.

6. Conclusion

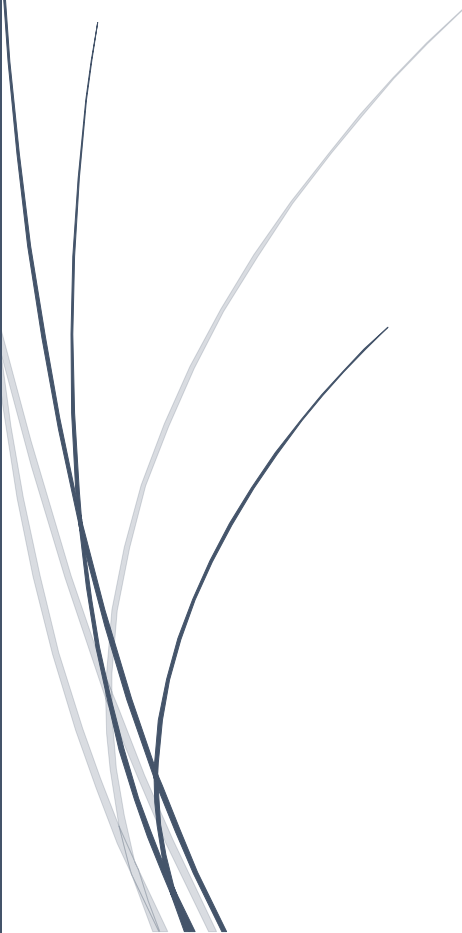
Dans ce chapitre, nous avons présenté le domaine de l'intelligence artificielle et clarifié l'apprentissage automatique, y compris la classification et les implications de la relation entre l'apprentissage automatique et l'apprentissage en profondeur, qui a surmonté les méthodes traditionnelles de manipulation d'images et de texte en rendant les systèmes plus rapides et plus faciles.

Dans le chapitre qui suit, on va présenter une étude sur les différents travaux connexes pour l'aide à la diagnostique.



Chapitre 3

*Les travaux connexes :
Approches, Techniques
et Applications pour
l'aide de diagnostic*



1. Introduction

Le cancer est une maladie mortelle due à la prolifération indésirable de cellules. Le cancer du sein est le type le plus courant et la principale cause de décès par cancer chez les femmes. Par conséquent, le diagnostic est devenu essentiel pour augmenter la survie à long terme et réduire les taux de mortalité.

Dans ce chapitre, nous passons en revue les travaux pertinents des approches, techniques et applications d'aide au diagnostic les plus populaires pour le cancer du sein, les bases de données génériques les plus citées et les types d'applications de Deep Learning dans l'analyse d'images médicales. Nous concluons enfin en notant les lacunes de la recherche et la synthèse de nos recherches.

2. Les CADs

Le traitement manuel des images médicales souffre de nombreux inconvénients, tels que l'indisponibilité des pathologistes en un seul endroit, la difficulté de l'analyse manuelle des images et le manque d'expérience professionnelle. Ainsi, pour pallier les erreurs de diagnostic aux stades précoces, une CAD « Computer aide décision » « Systèmes de détection assistée par ordinateur » est apparue, qui sert à résoudre les problèmes de classification multiple du cancer du sein [16].

Pour traiter ces photos. La coronaropathie est raisonnablement efficace pour détecter les cancers du sein invasifs à ses stades précoces, mais augmente le risque de résultats faussement positifs. Ainsi, la précision et l'efficacité du flux de travail doivent être prises en compte avant de mettre en œuvre la CAD dans la pratique clinique, et deuxièmement, pour éviter une mauvaise utilisation, la formation des utilisateurs est tout aussi importante pour comprendre les limites et les capacités des systèmes de CAD [19].

Les systèmes de détection assistée par ordinateur sont classés en deux catégories [21]:

- ❖ **Systèmes de détection assistée par ordinateur (CADE)** : assurent la détection et la localisation informatisées des anomalies dans les images médicales
- ❖ **Systèmes de diagnostic assisté par ordinateur (CADx)** : Cela classe les anomalies comme bénignes ou malignes. Il peut être utilisé comme un pas en avant pour le modèle de divulgation, pour fournir des informations supplémentaires et une aide supplémentaire dans la prise de décision.

Dans le reste de ce chapitre on va étudier les approches des CAD, à savoir les techniques classiques de détection de cancer de sein et les différentes applications intelligentes d'aide de diagnostic.

3. Les approches classiques

Afin de comprendre si les symptômes pathologiques sont causés par un cancer du sein bénin ou malin, des approches classiques basées sur les examens physiques complets sont effectués en plus d'un examen des seins pour aider à comprendre la cause des symptômes. Les tests (techniques) les plus connus d'aide au diagnostic sont :

3.1. Auto-examens des seins et examens cliniques des seins

L'examen des seins, à savoir le CBE (examens cliniques des seins : Clinical Breast Examinations) et l'ESB (Auto-examens des seins : Breast Self-Examinations), est des méthodes primordiales pour le dépistage du cancer, mais il est également nocif. Le CBE donne une spécificité élevée de 97,11%, tandis que la sensibilité est légèrement inférieure, c'est-à-dire 57,14% [4]. Il est utile uniquement pour détecter les lésions mammaires suspectes car il ne permet même pas de déterminer la malignité avec certitude.

- **Défis**
 - La sensibilité est faible.
 - Ne détermine même pas la malignité.

- Nocif en termes de biopsies supplémentaires qui sont nécessaires.

3.2. Mammographie

Une mammographie est un examen aux rayons X pour détecter des anomalies du sein. Il est relativement rapide et largement disponible. Cependant, il fonctionne bien chez les femmes ménopausées, alors qu'il n'est pas bon pour les jeunes femmes de moins de 40 ans, et l'exposition cumulative aux rayons X peut parfois être une cause de cancer. Ses limites sont liées à une plage dynamique limitée, un faible contraste et une image granuleuse. Les très petites lésions sont difficiles à visualiser chez les femmes ayant de gros implants ou des cicatrices chirurgicales [5]. En mammographie numérique, l'image est enregistrée dans un fichier électronique plutôt que dans un film comme le montre la figure 1.

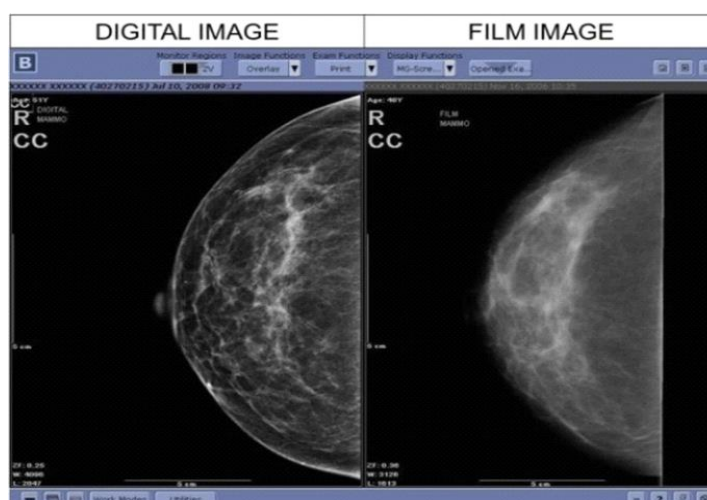


Figure 21: Test de mammographie.

- **Défis**
 - Risque pour les jeunes femmes de moins de 40 ans.
 - Plage dynamique limitée, faible contraste et image granuleuse.
 - La densité mammaire affecte la sensibilité et la spécificité.
 - Tomosynthèse
 - Mammographie numérique à contraste amélioré

3.3. IRM mammaire

L'IRM est une meilleure alternative à la, il utilise des ondes radio et un champ magnétique pour modifier l'alignement des protons dans le noyau d'hydrogène, produisant alors une image en coupe très détaillée de ce changement. L'IRM du sein fournit des images plus détaillées en raison de la résolution spatiale plus élevée, comme le montre la figure suivante. Parmi tous les avantages de l'IRM, il y a quelques limitations telles que la très faible spécificité qui conduit au test. Biopsies supplémentaires, coûteuses et non standardisées [10].



Figure 22: Test IRM.

- **Défis**

- Non normalisé et coûteux
 - La spécificité est faible.
 - Temps d'imagerie lent.
 - Non recommandé pour les patients avec des appareils métalliques.
 - Recommandé uniquement pour le dépistage des femmes à haut risque.

3.4. Échographie mammaire

Une échographie mammaire agit comme un test de mammographie à suivre pour les anomalies et fournit des lignes directrices pour une biopsie à l'aiguille de la masse. Un appareil portatif appelé transducteur doit être placé sur la partie supérieure de la peau du sein où un gel est appliqué et déplacé sur le sein pour montrer la structure du tissu sous-jacent. Ensuite, les ondes sonores sont émises et le transducteur doit capturer l'écho produit par les tissus corporels. Les échos sont ensuite convertis en une image qui doit être nommée en fonction de l'orientation et de l'emplacement dans le sein. Il produit des informations diagnostiques quantitatives et qualitatives basées sur un ensemble de méthodologies complexes [12]. La figure 3 montre une image échographique du sein.



Figure 23: Test d'échographie.

- **Défis**
 - Mauvaise résolution.
 - Une image à faible contraste est formée.

3.5. PET scan

Est une tomographie pour laquelle tout type d'ondes pénétrantes peut être utilisé [17]. Les reconstructions en coupe sont le phénomène mathématique de base derrière la formation

d'image. Pour la tomographie par émission de positons (TEP), le fluorodéoxyglucose (FDG) est la substance radioactive / tracker la plus largement utilisée qui doit être injectée dans une veine avec le bras pour caractériser le tissu. La quantité maximale de sucre radioactif est absorbée par les tissus cancéreux, qui sont les cellules les plus actives. Ensuite, un scanner TEP spécial a développé une image des zones irradiées du corps. La figure 4 illustre le scan FDG-PET[21].

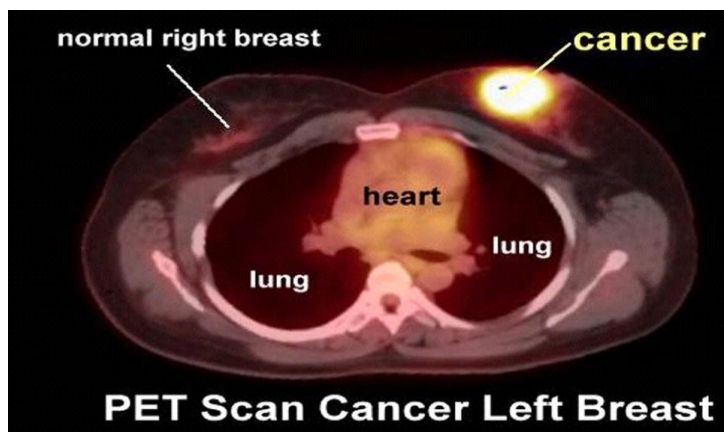


Figure 24: Test PET scan.

- **Défis**
 - Temps d'imagerie lent.
 - Couteux
 - Utilise un rayonnement ionisé.

Le tableau 1 présente l'analyse comparative de diverses techniques de détection du cancer du sein comme suit :

Tableau 1 : Analyse comparative des techniques d'aide de diagnostic de cancer de sein.

	BSE	CBE	Mammographi	Échographie	PET	MRI
Description	Auto-examen pour la détection des lésions	Examen clinique pour la détection des lésions mammaires	Examen aux rayons X pour l'évaluation et la détection des anomalies mammaires	Les ondes sonores sont introduites dans le corps et les échos sont utilisés pour former l'image	Le métabolisme du glucose est utilisé pour détecter le cancer et une image est formée par les zones radioactives du corps par un scanner TEP spécial	Des ondes radio et un champ magnétique pour modifier l'alignement des protons des noyaux d'hydrogène, puis produire une image en coupe très détaillée à partir de ce changement
Avantages	<ul style="list-style-type: none"> Utilisé pour la population à haut risque. Facile et pratique à utiliser 	<ul style="list-style-type: none"> Le taux de mortalité peut être réduit 	<ul style="list-style-type: none"> Moins de temps d'imagerie. La résolution est bonne. Appareil facilement portable 	<ul style="list-style-type: none"> Large disponibilité. Non invasif. Réduction des coûts. Sans douleur 	<ul style="list-style-type: none"> Fournit des informations anatomiques ou fonctionnelles. Bon contraste 	<ul style="list-style-type: none"> Haute résolution spatiale. Pas d'exposition aux radiations Peut mieux détecter la propagation intra-canalair du cancer
Limites	<ul style="list-style-type: none"> N'a aucun effet sur le taux de mortalité Nuisible 	<ul style="list-style-type: none"> Faible taux de mortalité. Ne détermine même pas la malignité 	<ul style="list-style-type: none"> Plage dynamique limitée, des radiations ionisantes sont utilisées. Contraste médiocre. Pas bon pour les jeunes femmes 	<ul style="list-style-type: none"> Mauvaise résolution. Mauvais contraste. Selon les l'expérience du médecin 	<ul style="list-style-type: none"> Non portable. Temps d'imagerie lent. Utilise des radiations ionisantes. Cher 	<ul style="list-style-type: none"> Temps d'imagerie lent. Non portable. Seule la vue latérale est disponible Cher. interprétation complexe
Sensibilité	Faible	Modérer	Faible (femme < 40) Haute (femme > 40)	Faible	Haute	Tré Haute
Spécificité	Modérer	Haute	Haute	Modérer	Modérer	Faible
Utilisé pour	Pré-scan	Pré-scan	Dépistage, diagnostic, pronostic	Sonoélastographie, imagerie harmonique tissulaire, tomographie échographique en spirale	Dépistage, diagnostic, pronostic	Imagerie de diffusion et de perfusion, spectroscopie par résonance magnétique du proton, élastographie RM

4. Les approches intelligentes

Des algorithmes d'apprentissage automatique en couches profondes tels que les réseaux de neurones à convolution (Deep Convolution Neural Network) ont récemment montré du succès dans diverses tâches d'analyse d'images médicales telles que la segmentation, la détection et la classification.

Ils ont commencé par la manière dont les multiples couches de l'approche DeepLearning ont aidé à permettre l'apprentissage efficace des représentations cachées dans les ensembles de données et l'obtention de gains exponentiels dans la puissance picturale de chaque entité dans les ensembles de données. Outre le coût de calcul, ils ont ajouté que le réglage fin des hyperparamètres des modèles et le choix des caractéristiques structurelles doivent encore être réalisés pour les technologies DL. Mais la disponibilité de modèles pré-entraînés a permis aux chercheurs d'extraire des caractéristiques à divers points.

4.1. Les applications d'aide de diagnostic

La classification, la segmentation et la détection automatiques des types de cancer sont une condition préalable à diverses analyses quantitatives et qualitatives en pathologie computationnelle. Au cours des dernières années, différentes approches DCNN ont été proposées et appliquées avec succès sur des problèmes d'analyse d'images médicales et montrent des performances supérieures sur différents ensembles de données de référence pour la classification, la segmentation et la tâche de détection [1]. Les applications sur l'apprentissage automatique traditionnel et la classification, la segmentation et la détection des cancer de sein basés sur l'apprentissage profond sont les suivants :

4.1.1. Segmentation

Une nouvelle approche de détection et de segmentation des cellules basée sur les contours est proposée en 2012. Cette méthode utilise des informations a priori minimales et détecte les contours indépendamment de leur forme et a obtenu des résultats de segmentation prometteurs dans [19]. La segmentation de la membrane des noyaux avec le modèle CNN est

proposée à partir d'images microscopiques dans [20]. Un autre chercheur [21] a proposé une approche basée sur CNN appelée U-Net pour la segmentation générale des images médicales dans . Une approche très simple de segmentation des noyaux basée sur un modèle CNN est proposée en 2017, nommée modèles CNN2 et CNN3 en ce qui concerne le nombre de classes de sortie. Pour le modèle à deux classes, le réseau est utilisé pour classer les pixels respectivement à l'intérieur et à l'extérieur de la région des noyaux. D'autre part, le modèle pour trois classes, ils ont été utilisés pour classer les pixels appartiennent à l'intérieur, un côté de sortie, et la limite des régions de noyaux dans [23].

Cependant, pour la segmentation générale des images médicales, une version améliorée du modèle d'apprentissage en profondeur U-Net a été proposée où des modules résiduels récurrents sont incorporés dans U-Net au lieu de couches convolutives à action directe. Le modèle a été évalué sur différentes modalités d'imagerie médicale, y compris la segmentation des vaisseaux sanguins rétiniens, la segmentation du cancer de la peau et la segmentation pulmonaire.

4.1.2. Détection

Il existe deux méthodes différentes principalement utilisées pour la détection des noyaux. Premièrement, le comptage basé sur la détection, qui nécessite une détection ou une segmentation préalable comme discuté dans [29]. D'autre part, la méthode basée sur l'estimation de la densité est utilisée pour la détection des noyaux sans utiliser la segmentation qui est expliquée dans [30]. Un cadre avec CNN à pool maximal supervisé est formé à la détection de la région de pixels de cellule qui est présélectionnée avec Support Vector Machine (SVM). Un modèle de régression basé sur CNN est utilisé pour la détection et le comptage des noyaux où un modèle de réseau de régression neurale entièrement convolutif est utilisé et capable de cartographier la densité pour une image d'entrée de taille arbitraire.

4.1.3. Classification

La classification peut être utilisée pour diverses applications. Des caractéristiques telles que la forme, la texture et la taille des noyaux étaient prises en compte pour le pléomorphisme

nucléaire dans les images de cancer du sein [12]. Autre chercheurs ont appliqué CNN pour classer les cellules mitotiques et non mitotiques en utilisant les informations de couleur, de forme et de texture dans [13]. Une autre méthode avec le classificateur de l'intensité, la morphologie et la texture sont utilisées et le principal objectif de ce travail était la segmentation des noyaux avec une approche de classification [15].

Cependant, des études récentes ont montré que les approches basées sur l'apprentissage profond produisent une précision de classification prometteuse sur des images pathologiques à grande échelle.

Une autre approche basée sur l'apprentissage en profondeur est proposée pour la classification des cellules et comparée à une méthode avec un ensemble de fonctionnalités et de représentations canoniques dans [17]. L'approche de classification des images histopathologiques est proposée et appliquée sur la machine à vecteurs de support (SVM), AdaBoost et DCNN. L'expérience est menée sur quatre ensembles de données d'images colorées, à savoir les données sur le cancer de la prostate, du sein, des cellules claires rénales et des cellules papillaires rénales.

4.2. Dataset

Des nombreux jeux de données (Dataset) ont été produites et publiées dans un référentiel différent où certaines d'entre elles étaient accessibles au public et autres sont privée. La plupart des jeux de données existent en deux formats :

- ✓ Fichier CSV (comma-separated values)
- ✓ Image (jpg, pgm, png, DICOM et jpeg)

Les dataset sont aussi par modalité que ce soit :

- ✓ **DM** : Mammographie numérique (Digital mammography) , est un type de mammographie caractérisé par une forte exposition aux rayonnements

- ✓ **DBT** : La tomosynthèse mammaire numérique ((Digital breast tomosynthesis) est une modalité d'imagerie qui produit une image 3D du sein à l'aide de rayons X à faible dose reçus sous différents angles

L'analyse d'images du cancer du sein a principalement utilisé ces bases de données. Par exemple, la base de données de Breast Cancer Wisconsin (Diagnostic) est la plus populaire et la plus applicable par de nombreux chercheurs. Il contient 569 échantillons d'images dont 208 sont normaux et 114 sont anormaux (63 cas bénins et 51 cas malins). L'autre base de données populaire est la base de données MIAS avec 322 échantillons. Le résumé des bases de données sur le cancer du sein les plus citées est présenté dans le tableau suivant :

Tableau 2: Les jeux de données de cancer de sein.

Dataset	Size	Citation	Classes	Format	Modalité
MIAS	322	304	<ul style="list-style-type: none"> N : Normal AB : Abnormal 	Pgm	Digital Mammogram Database : DM
DDSM	55,890	15	N, AB	Jpeg	DM
Breast Cancer Wisconsin (Diagnostic)	569	375	<ul style="list-style-type: none"> B : Benign M : Malignant 	CSV	DM
Breast Histopathology	277,524	160	<ul style="list-style-type: none"> IDC (-) : invasive ductal carcinoma negative IDC (+) : invasive ductal carcinoma positive 	Png	DM
NKI Breast Cancer Data	272	30	-	CSV	Gene
Breast Cancer Proteomes	64	80	-	CSV	Gene
Incidence of breast cancer (all)	-	24	-	Html	-

IRMA	14k	-	-	png	-
UM:University of Michigan Health System	-	-	-	-	-
UC :niversity of Chicago	-	-	-	-	-

4.3. Les travaux connexes

Pour la détection précoce, le diagnostic et le traitement des maladies, les images médicales sont principalement interprétées par des experts humains tels que des radiologues et des médecins. Cependant, en raison des grandes différences de pathologie et de la fatigue potentielle des experts humains, les chercheurs et les cliniciens ont récemment commencé à bénéficier des interventions assistées par ordinateur.

Les progrès des technologies d'imagerie médicale se sont récemment améliorés grâce aux techniques d'apprentissage automatique. Les domaines d'application les plus courants de la DL dans les soins médicaux comprennent : l'analyse d'images du cancer du sein, l'analyse d'images cérébrales, l'analyse d'images rétiniennes, l'analyse radiographique pulmonaire et les images musculaires et osseuses. Quelques travaux de CAD basés sur DL est les suivant :

- [24] propose un système de CAD basé sur les fonctionnalités basé sur DCNN. Dans DCNN, l'apprentissage par transfert est appliqué pour former les quatre premières couches convolutives et les trois dernières couches entièrement connectées de DCNN en utilisant uniquement des images de mammographie pour identifier les lésions et réduire les faux positifs. L'apprentissage par transfert a obtenu une formation accuracy de 0,99 et a été validé avec la dataset d'image DBT et a atteint une AUC¹ de 0,81. Cependant, après la formation avec DM (Data Mamography) uniquement, une formation supplémentaire a été effectuée à l'aide d'images DBT pour améliorer le score de validation du modèle et atteindre une AUC de 0,90. La source de l'ensemble

¹ AUC : aire sous la courbe ROC

AUC signifie "aire sous la courbe ROC". Cette valeur mesure l'intégralité de l'aire à deux dimensions située sous l'ensemble de la courbe ROC (par calculs d'intégrales) de (0,0) à (1,1).

de données d'image était le Département de radiologie du système de santé de l'Université du Michigan et de l'Université du sud de la Floride. Les caractéristiques morphologiques et structurelles ont été utilisées comme système de CAD basé sur les caractéristiques pour la détection de la masse dans le corps.

- ✚ [25] ont introduit des techniques de classification multi-classes basées sur CNN-CT (ComputerTomography) pour classer les mammographies des ensembles de données IRMA en normales, bénignes et malignes. La fusion des caractéristiques CNN et des caractéristiques les plus descriptives avec des ondelettes a bien fonctionné et a atteint une précision de 83,74% pour le classificateur SVM.
- ✚ [26] ont appliqué DCNN pour la détection et le diagnostic des tumeurs du sein. Les auteurs ont prétraité (recadrage et redimensionnement) les mammographies originales de MIAS. Ensuite, l'augmentation des données a été appliquée en faisant pivoter les images originales pour améliorer la taille des ensembles de données d'image. Ils ont extrait des caractéristiques à l'aide du modèle pré-entraîné CNN et ont donné ces caractéristiques à SVM et ont obtenu une précision de 64,52%.
- ✚ [27] ont introduit deux techniques DCNN qui classent les images mammaires en bénignes et malignes. Les images annotées ont été recadrées en utilisant une boîte de délimitation carrée autour des limites annotées et ont obtenu 891 malins et 365 bénins . Ces images ont été augmentées en utilisant la rotation (90°, 180°, 270°) et le retournement (droite-gauche, bas-haut). Dans la première approche, le CNN avec trois couches convolutives a été formé avec les ensembles de données étiquetés. Et dans la deuxième approche, les mêmes ensembles de données étiquetés ont été utilisés comme entrée pour VGGNet pré-entraîné pour extraire les entités d'une couche entièrement connectée pour effectuer une classification à l'aide de SVM. La précision, la sensibilité, la spécificité et l'accuracy de la première approche étaient de 83%, 84%, 82% et 0,91, respectivement. Et la précision, la sensibilité, la spécificité et l'ASC de la seconde approche étaient respectivement de 73%, 77%, 68% et 0,81.
- ✚ [28] ont proposé deux façons différentes d'extraire des caractéristiques dans le but de classer les images en bénignes et malignes, la première étant basée sur la segmentation et la seconde basée sur CNN. L'étude était basée sur 640 images collectées à l'aide de

la modalité d'imagerie IRM. Sur 640 images, 191 étaient bénignes et 449 étaient malignes. Dans une approche basée sur la segmentation, 38 caractéristiques de 6 catégories différentes (texture d'amélioration, taille, variance cinétique, morphologie, forme et cinétique) ont été extraites après segmentation à des fins de classification. Dans une approche basée sur CNN, les retours sur investissement de 148×148 pixels extraits ont été fournis en entrée au modèle pré-entraîné AlexNet et 4 096 vecteurs de caractéristiques ont été extraits des couches FC. Cependant, seulement 518 caractéristiques ont été utilisées pour l'analyse après avoir laissé 80% des vecteurs de caractéristiques avec une valeur nulle. Les performances de cette étude ont été évaluées à l'aide d'un classificateur LDA avec validation croisée à tour de rôle pour trois cas (fonctionnalités basées sur la segmentation (38), fonctionnalités basées sur CNN (518) et fonctionnalités fusionnées (556)). Les résultats de performance en termes d'accuracy pour les caractéristiques CNN basées sur la segmentation et les caractéristiques combinées étaient respectivement de 0,88, 0,76 et 0,91.

- ✚ [29] ont collecté trois ensembles de données en utilisant trois modalités d'imagerie (mammographie, échographie et IRM). Le nombre de patients pris en compte dans les modalités d'imagerie mammographique était de 245, 1125 patients pour l'échographie et 690 pour IRM. Cependant, le nombre de région d'intérêt (Région of intérêts :ROI) pour la modalité d'imagerie mammographique était de 739 (328 bénin, 411 malin), 2393 (1978 bénin, 415 malin) pour une échographie et 690 (212 bénin, 478 malin) pour IRM. Pour tous les ensembles de données, les caractéristiques basées sur CNN des couches FC et Max-pool de VGGNet (VGG19) et les caractéristiques conventionnelles (caractéristiques artisanales) ont été collectées. Cependant, les performances des fonctionnalités max-pool ont surpassé les fonctionnalités FC selon la comparaison effectuée sur la base de l'AUC et, par conséquent, les performances des fonctionnalités conventionnelles ont été réalisées uniquement avec les fonctionnalités max-pool. Les deux caractéristiques (caractéristiques conventionnelles et caractéristiques CNN) ont été introduites dans un SVM non linéaire avec noyau RBF gaussien et ont atteint une valeur AUC de 0,79 et 0,81 pour les images mammographiques, 0,84 et 0,87 pour les images échographiques et 0,86 et 0,87 pour

l'IRM, respectivement. La performance de SVM a également été évaluée pour les caractéristiques combinées ou fusionnées (caractéristiques conventionnelles + caractéristiques CNN) et a atteint une valeur AUC de 0,86, 0,90 et 0,89 pour les ensembles de données mammographie, échographie et IRM, respectivement.

- ✚ [30] ont extrait les caractéristiques basées sur CNN des cinq couches de pooling max en utilisant des images DCE-MRI de 690 cas. Sur la base d'un rapport de pathologistes et de radiologues, sur 690 cas, 212 cas étaient bénins et 478 cas étaient malins. Les caractéristiques extraites ont d'abord été normalisées avec la norme euclidienne avant d'être concaténées pour former des vecteurs de caractéristiques CNN fusionnés. Ensuite, ces caractéristiques ont été attribuées à la SVM linéaire pour classer les images IRM comme malignes et bénignes. Le pouvoir discriminant des fonctionnalités de les trois ROI ont été évalués à l'aide de accuracy avec 80% des fonctionnalités pour la formation et 20% pour les tests. La valeur AUC de la tranche centrale du second post-contraste, une tranche centrale du second post-contraste soustraite et MIP étaient de 0,80, 0,83 et 0,88, respectivement.
- ✚ [31] ont extrait des caractéristiques des cinq couches max-pool à l'aide de VGGNet avec 19 couches à classer comme bénignes et malignes. Sur les 703 ensembles de données d'images collectées à l'aide de la modalité d'imagerie IRM, 221 étaient bénignes et 482 étaient malignes. Ils ont extrait séparément les caractéristiques des images avant et après l'amélioration du contraste et les approches classiques. Les paramètres des classificateurs LSTM et SVM ont été ajustés sur une recherche de grille avec validation croisée (5 fois). L'efficacité des classificateurs et le pouvoir distinctif des caractéristiques ont été mesurés sur la base de l'analyse AUC. L'AUC pour le classificateur SVM était de 0,81 et 0,85 pour le classificateur LSTM.

Tableau 3 :Travaux connexe

Ref	Classifieur	Size	Dataset	Modalité	AUC	Acc
[24]	CNN	64	UM	DBT	0.93	-
[25]	SVM	-	IRMA	DM	-	83.74
[26]	SVM			DM	64.52	
[27]	CNN	4863	-	MRI	0.91	83
[28]	LDA (CNN)	640	UC	MRI	0.76	-
[29]	SVM (CNN)	690	UC	MRI	0.87	-
[30]	SVM	690	UC	MRI	0.83	-
[31]	LSTM	703	UC	MRI	0.85	

5. Les problèmes et les défis

Dans la plupart des applications médicales, certains types de méthodes médicales sont insuffisants ou inefficaces pour fournir la source d'information requise, ce qui réduit à son tour l'efficacité des processus nécessaires. Cet obstacle peut être surmonté grâce à diverses méthodes d'imagerie médicale qui peuvent fournir des informations raisonnables et suffisantes, mais qui souffrent toujours des problèmes suivants :

- ✚ Le stade de diagnostic du cancer du sein nécessite un examen adéquat pendant plus de 15 jours, ce qui affecte le stade du traitement.
- ✚ La radiographie permet une détection précoce de la maladie, bien qu'il s'agisse toujours d'un taux d'erreur élevé (pourcentage élevé de faux positifs et de faux négatifs) qui entraîne un traitement inutile avec d'autres effets tels que les dommages psychologiques
- ✚ Nécessite un gros budget financier

- ✚ Exposer les patients à des rayonnements, qui peuvent à leur tour être nocifs pour la santé humaine.

6. La synthèse

L'établissement de systèmes robustes d'analyse des données est essentiel pour améliorer le coût des traitements du cancer du sein, réduire la mortalité et réduire les interventions chirurgicales.

Les systèmes de diagnostic assisté par ordinateur (CAD) sont l'une des solutions automatiques proposées qui peuvent aider le radiologue à détecter efficacement les anomalies et à participer ainsi à la réduction du taux de mortalité par cancer du sein. Ces systèmes agissent comme un deuxième lecteur indépendant ou semi-indépendant pour la mammographie et attirent l'attention du radiologue sur des zones souvent manquées lors de la première lecture. Des études ont montré que ces systèmes de mammographie améliorent considérablement la précision de la détection du cancer du sein par les radiologues.

Une autre étude a montré que la performance du système CAD était similaire à la performance globale des 13 radiologues qui ont analysé les mammographies dans cette population.

L'activité d'assistance au diagnostic s'articule autour du simple processus de vision artificielle en imagerie médicale, qui permet au radiologue de confirmer sa décision ou de revoir son interprétation. Avant de commencer ce processus, choisissez les méthodes appropriées.

Nous avons d'abord examiné l'approche spécialisée la plus importante du cancer du sein et identifié les faiblesses de chaque technique. De plus, nous avons suivi l'approche de détection intelligente basée sur des algorithmes d'apprentissage profond, où nous avons découvert que les algorithmes d'apprentissage profond ont obtenu des résultats très satisfaisants dans toutes les applications de détection, de classification et de prédiction du cancer du sein. Nous suivons cette approche et nous nous efforçons de:

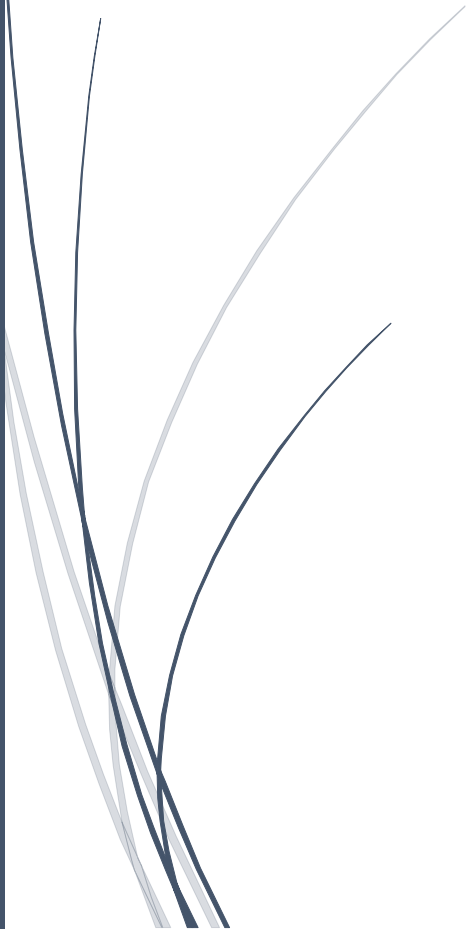
1. Réaliser une connexion entre le domaine de la médecine et l'avenir des technologies de l'information afin d'améliorer les performances des systèmes de diagnostic du cancer du sein, c'est-à-dire de construire un système de CAD
2. Proposition d'un modèle intelligent basé sur les réseaux de neurones de classification d'images (CNN: convolutional neural network) pour classer le diagnostic (sain ou blessé) à l'aide du type de base de données mammographie "Breast Cancer Wisconsin (Diagnostic)"

7. Conclusion

Dans ce chapitre, nous avons présenté en vue sur les différentes approches, techniques et applications d'aide au diagnostic les plus populaires pour le cancer du sein, à savoir les défis de chaque technique et une comparaison entre eux. Les technologies intelligentes basées sur des algorithmes d'apprentissage en profondeur ont prouvé leur force dans le diagnostic du cancer du sein, car elles représentent une solution intelligente qui joue un rôle dans le lien entre l'informatique et la médecine.

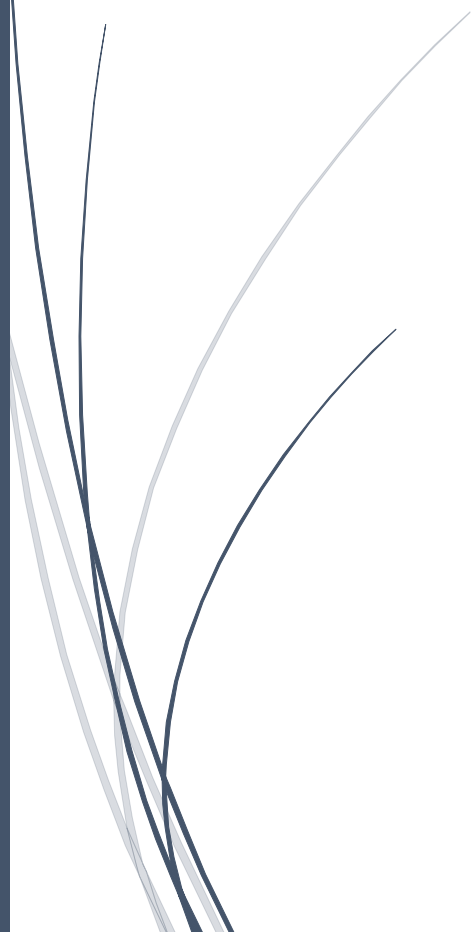
Dans le chapitre qui suit, on va présenter notre contribution, l'algorithme et la dataset choisie pour construire notre modèle d'aide au diagnostic de cancer de sein.

Partie 2



Chapitre 4

Conception & Réalisation



1. Introduction

Dans ce chapitre, nous présentons notre contribution et l'implémentation du notre modèle de classification des maladies du cancer du sein. Tout d'abord, nous donnons un aperçu du schéma de construction du modèle pour toute application d'apprentissage automatique, puis nous commençons par la formation de notre propre système, puis nous fournirons une définition détaillée de chaque étape de la construction du modèle, et nous terminons le chapitre avec le résultat et la discussion.

2. Objective

Notre objectif dans ce chapitre est de trouver une solution intelligente pour aider à diagnostiquer le cancer du sein, nous cherchons donc à former et à mettre au point un réseau de neurones convolutifs (CNN) pour classer avec précision les patientes atteintes d'un cancer du sein comme malignes ou bénignes. Certaines des études que nous avons détaillées dans le chapitre précédent suggèrent que les CNN peuvent être plus biaisés avec ce type de maladie, ce qui signifie que les réseaux de neurones profonds sont particulièrement bien adaptés à la reconnaissance de formes et peuvent obtenir de meilleurs résultats sans d'abord identifier les mêmes caractéristiques qu'un l'humain recherche.

Ainsi, comme le montre la figure suivante, nous commençons par rechercher un ensemble de données pour notre domaine d'étude, puis prétraitons les données, puis construisons un modèle CNN à partir de zéro, l'entraînons et obtenons la précision du modèle pour finalement servir de modèle prêt à l'emploi et aider à prendre des décisions concernant le cancer du sein

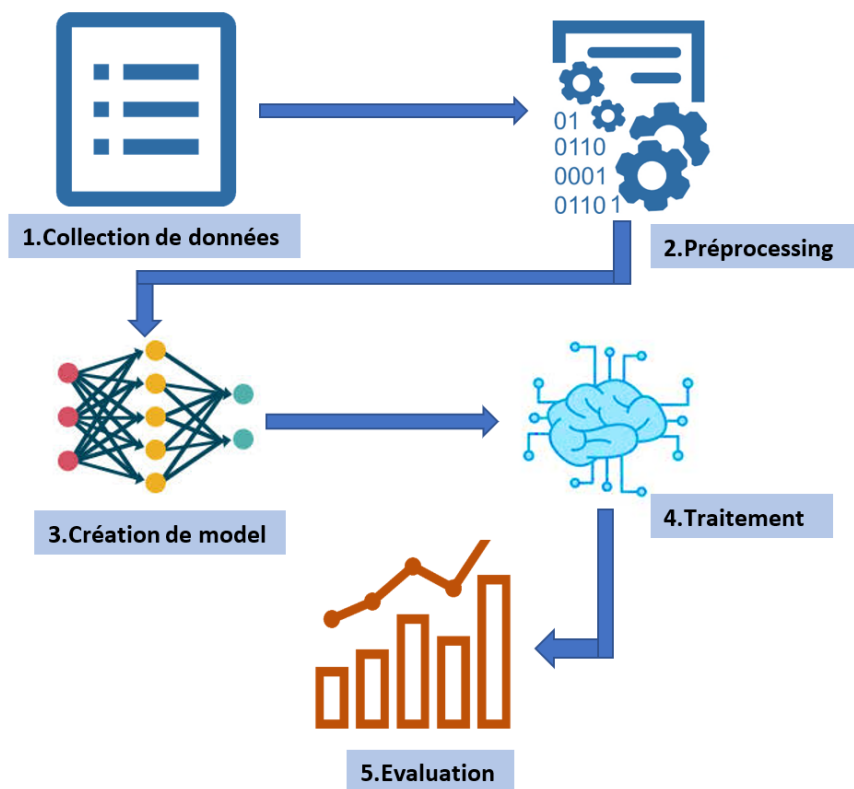


Figure 25: Processus d'atteindre notre objective

3. Conception

Pour spécifier notre contribution, nous avons élaboré l'architecture suivante qui englobe tout les étapes d'implémentation de notre modèle, puis nous détaillons chaque étape dans les sections qui suivent.

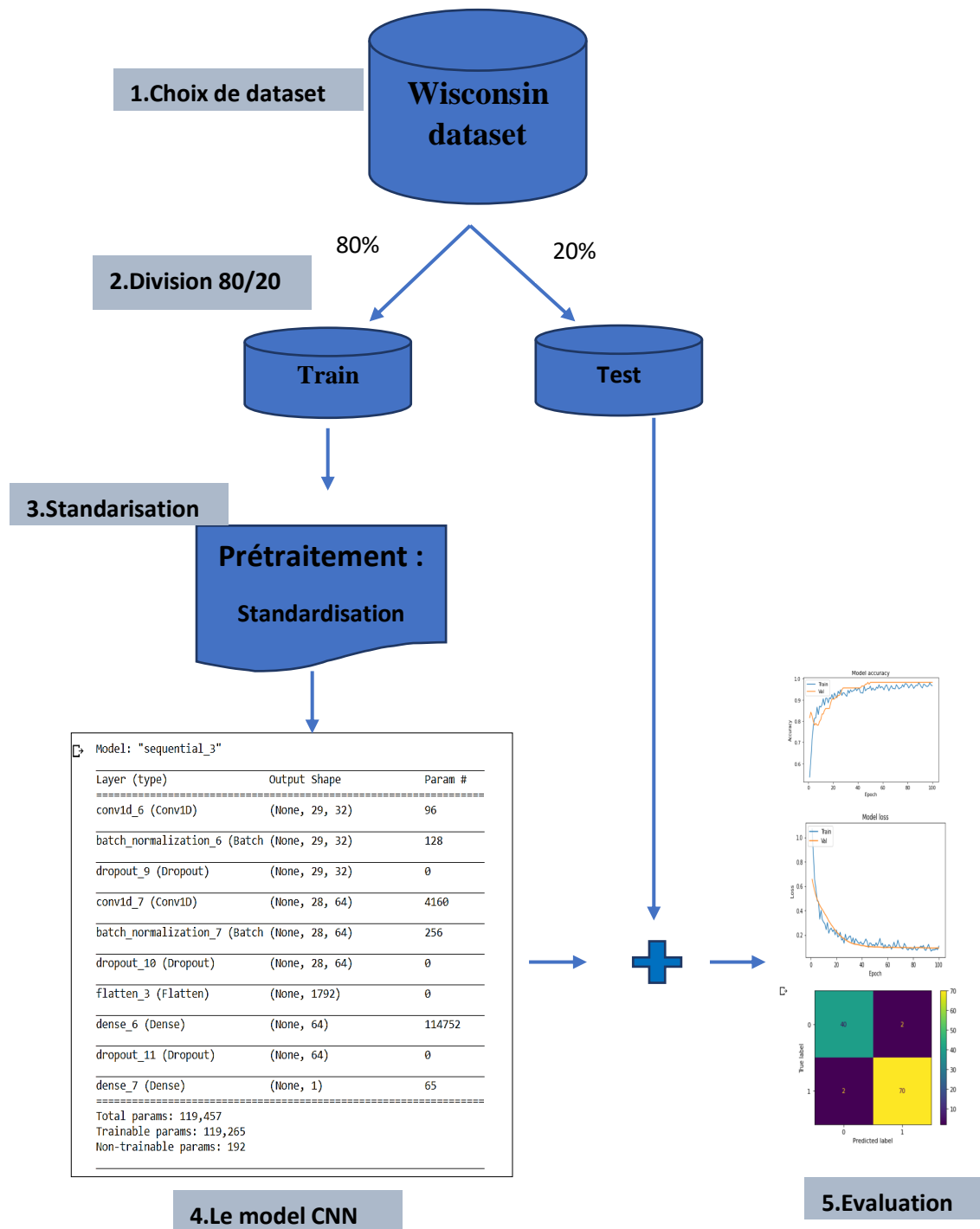


Figure 26: Architecture proposé

3.1. Data set

L'ensemble de données (Dataset) utilisé pour la mise en œuvre du travail proposé est l'ensemble de données (Dataset) Wisconsin Breast Cancer (Original). Il s'agit d'un ensemble de données disponible gratuitement dans le référentiel d'apprentissage automatique de l'Université de Californie, Irvine (UCI).

Cet ensemble de données contient 699 instances, 10 variables d'entrée, c'est-à-dire les caractéristiques et 1 variable de sortie, c'est-à-dire la réponse. La description de toutes les caractéristiques ainsi que leurs attributs sont bien détaillées dans les tableaux suivants. L'ensemble de données se compose de 458 (65,5 %) cas de tumeur bénigne et 241 (34,5 %) cas de tumeur maligne. Il y a 16 instances dans l'ensemble de données et chacune contient une seule valeur manquante. Pour plus de simplicité, dans l'attribut de réponse « classe », 2 est remplacé par « bénigne » et 4 est remplacé par « malin ».

3.1.1. Caractéristique de dataset

Tableau 4: Les caractéristiques de dataset¹

Caractéristiques de dataset :	Multivarié
Caractéristiques des attributs :	Réel
Tâches associées :	Classification
Nombre d'instances :	569
Nombre d'attributs :	32
Valeurs manquantes :	Non
Surface :	Vie
Date du don	1995-11-01
Nombre de visites Web :	1515460

¹Dr. William H. Wolberg, General Surgery Dept. University of Wisconsin, Clinical Sciences Center Madison, WI 53792
wolberg '@' eagle.surgery.wisc.edu
 UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set

3.1.2. Attributs de dataset

Tableau 5: Les attribues de dataset

N	Attributs	Description	Domaine
1	Exemple de numéro de code	L'identifiant unique de l'observation.	Numéro unique
2	Épaisseur de touffe	Les cellules regroupées en monocouches sont identifiées comme des cellules bénignes, tandis que les cellules regroupées en multicouches sont identifiées comme des cellules malignes, c'est-à-dire cancéreuses.	1-10
3	Uniformité de la taille des cellules	La taille des cellules diffère des cellules malignes aux cellules bénignes.	1-10
4	Uniformité de la forme des cellules	La forme des cellules diffère des cellules malignes aux cellules bénignes.	1-10
5	Adhérence marginale : 1 - 10	Les cellules normales sont de nature compacte tandis que les cellules cancéreuses sont de nature lâche.	1-10
6	Taille de cellule épithéliale unique	Les cellules épithéliales distendues et assez grandes peuvent être une cellule cancéreuse	1-10
7	Noyaux nus	Les noyaux des cellules qui ne sont pas entourés de cytoplasme sont appelés noyaux nus ou nus. Cela se trouve dans les tumeurs bénignes.	1-10
8	Chromatine fade	Dans les cellules bénignes, on retrouve une texture fine et uniforme du noyau tandis que la chromatine est plus grossière dans les cellules cancéreuses	1-10
9	Nucléoles normaux	Généralement, les cellules contiennent des nucléoles à peine visibles mais dans les cellules cancéreuses, les nucléoles sont proéminents.	1-10
10	Mitoses	Le processus de division des cellules normales varie de celui des cellules cancéreuses.	1-10
11	Classe	Attribut de réponse	2 pour bénigne, 4 pour maligne

3.2.1 Division des données

Pour entraîner n'importe quel modèle d'apprentissage automatique, quel que soit le type d'ensemble de données utilisé, l'ensemble de données doit être divisé en données d'entraînement et données de test.

Lors du fractionnement d'un ensemble de données, il existe deux problèmes concurrents :

- ✚ Si vous disposez de moins de données d'apprentissage, vos estimations de paramètres ont une plus grande variance.
- ✚ Et si vous avez moins de données de test, vos statistiques de performances auront plus de variance.

Les données doivent être segmentées de manière à ce qu'aucune d'entre elles ne soit trop élevée, ce qui dépend d'avantage de la quantité de données dont vous disposez. Si vos données sont trop petites, aucune division ne vous donnera un contraste satisfaisant, vous devrez donc effectuer une validation croisée, mais si vos données sont énormes, une meilleure division est de 80 :20.

Diviser la dataset en train 80 % et test 20%

```
[ ] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state = 0, stratify = y)
```

Figure 27: code python de division de données 80:20

3.3. Prétraitement

Comme une étape de prétraitement nous avons appliqué la standardisation, c'est un processus de Standardiser les fonctionnalités en supprimant la moyenne et en adaptant la variance à l'unité. Le score standard d'un échantillon x est calculé comme suit : $z = (x - u) / s$ où u est la moyenne des échantillons d'apprentissage, et s est l'écart type des échantillons d'apprentissage. Le centrage et la mise à l'échelle se produisent indépendamment sur chaque caractéristique en calculant les statistiques pertinentes sur les échantillons de l'ensemble d'apprentissage. La moyenne et l'écart type sont ensuite stockés pour être utilisés sur des données ultérieures à l'aide de la transformation. La standardisation d'un ensemble de données est une exigence courante pour de nombreux estimateurs d'apprentissage automatique : ils

peuvent se comporter mal si les caractéristiques individuelles ne ressemblent pas plus ou moins à des données standard distribuées normalement.

StandardScaler: Standardiser les fonctionnalités en supprimant la moyenne et en mettant à l'échelle la variance unitaire

```
[ ] scaler = StandardScaler()
    X_train = scaler.fit_transform(X_train)
    X_test = scaler.transform(X_test)
```

Figure 28 :code python de standardisation

3.4. Le model CNN

Notre CNN modèle que nous présentons dans la figure suivante c'est un modèle CNN-1D , il est composé de 2 couches deconvolution et 2 couches batch_normalization , une couche de flatten , une couche caché et une couche de sortie :

Les données de l'entrée sont de taille de 29 éléments (Total de caractéristiques) est passé d'abord à la première couche de convolution. Cette couche est composée de 32 filtres de taille 2, ensuite on applique la batch_normalization qui permet de réduction du phénomène de overfitting , on applique aussi un dropout de 30% pour éviter le problème de sur-apprentissage.

Une autre fois on applique une couche de convolution. Cette fois est composée de 64 filtres de taille 2, en suite on applique batch_normalization, on applique aussi un dropout de 50% pour éviter le problème de sur apprentissage.

Puis en passe par la couche flatten qui prend des données de forme bidimensionnelles et génère des données de dimension unidimensionnelles (c'est l'entrée de réseau de neurone).

Finalement nous passons par une couche cache de 64 neurones et une couche de sortie avec un seule neurone et une fonction d'activation 'sigmoid' pour limiter le résultat entre 0 et 1.

Creation de model

```

▶ epochs = 100
model = Sequential()
model.add(Conv1D(filters=32, kernel_size=2, activation='relu', input_shape = (30,1)))
model.add(BatchNormalization())
model.add(Dropout(0.3))

model.add(Conv1D(filters=64, kernel_size=2, activation='relu'))
model.add(BatchNormalization())
model.add(Dropout(0.5))

model.add(Flatten())
model.add(Dense(64, activation='relu'))
model.add(Dropout(0.5))

model.add(Dense(1, activation='sigmoid'))

```

Figure 29: code python de model CNN.

3.5.Evaluation

Afin d'évaluer les performances du modèle, nous avons choisi de tracer à la fois les graphiques de précision et de perte du modèle. De plus, nous avons dessiné une matrice de confusion afin de distinguer le nombre d'échantillons attendus, vrai et faux, ce qui nous permet de calculer l'accuracy de model

+ Accuracy

La précision (Accuracy) est la mesure de performance la plus intuitive et il s'agit d'un rapport entre les observations correctement prédites et les observations totales.

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

Où

- ❖ TP est le nombre de vrais positifs
- ❖ FP le nombre de faux positifs.
- ❖ TN est le nombre de vrais négatifs
- ❖ FN le nombre de faux négatifs.

▾ Evaluer le model

```

▶ def plot_learningCurve(history, epoch):
    # accuracy
    epoch_range = range(1, epoch+1)
    plt.plot(epoch_range, history.history["accuracy"])
    plt.plot(epoch_range, history.history["val_accuracy"])
    plt.title('Model accuracy')
    plt.ylabel('Accuracy')
    plt.xlabel('Epoch')
    plt.legend(['Train', 'Val'], loc='upper left')
    plt.show()

    # loss
    plt.plot(epoch_range, history.history['loss'])
    plt.plot(epoch_range, history.history['val_loss'])
    plt.title('Model loss')
    plt.ylabel('Loss')
    plt.xlabel('Epoch')
    plt.legend(['Train', 'Val'], loc='upper left')
    plt.show()

```

Figure 30:Code python d'évaluation

```

▶ import matplotlib.pyplot as plt
from sklearn.datasets import make_classification
from sklearn.metrics import plot_confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC

X, y = make_classification(random_state=0)
#X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0)
clf = SVC(random_state=0)
clf.fit(X_train, y_train)
SVC(random_state=0)
plot_confusion_matrix(clf, X_test, y_test)
plt.show()

```

Figure 31: Code python de matrice de confusion

4. Les outils d'implémentation

Les outils que nous avons utilisé dans l'implémentation de notre model est les suivant, note que tout le model est implémenté dans un pc portable Lenovo i7 -8500U 8GO de ram avec un nvidia GEFORCE Mx110 de 2GO.

4.1. Google colab

Google colab ou Colaboratory est un service cloud, offert par Google (gratuit), basé sur Jupyter Notebook et destiné à la formation et à la recherche dans l'apprentissage automatique. Cette plateforme permet d'entraîner des modèles de Machine Learning directement dans le cloud sans avoir besoin d'installer quoi que ce soit sur notre ordinateur à l'exception d'un navigateur.

4.2. Python

Python est le langage de programmation open source le plus employé par les informaticiens. Ce langage s'est propulsé en tête de la gestion d'infrastructure, d'analyse de données ou dans le domaine du développement de logiciels

4.3. Sklearn

Scikit-learn est une bibliothèque libre Python destinée à l'apprentissage automatique. Nous avons l'utiliser pour nous permettre de ploter les graphe de accuracy , de erreur et la matrice de confusion .

4.4. Tensorflow

Est une bibliothèque de logiciels libres et open source pour l'apprentissage automatique. Il peut être utilisé dans une gamme de tâches, mais se concentre particulièrement sur la formation et l'inférence des réseaux de neurones profonds.

4.5. Pandas

Pandas est une bibliothèque logicielle écrite pour le langage de programmation Python pour la manipulation et l'analyse de données. En particulier, il propose des structures de données et des opérations pour manipuler des tableaux numériques et des séries chronologiques.

4.6. Numpy

NumPy est une bibliothèque pour le langage de programmation Python, ajoutant la prise en charge de grands tableaux et matrices multidimensionnels, ainsi qu'une grande collection de fonctions mathématiques de haut niveau pour opérer sur ces tableaux

4.7. Keras

Keras est une bibliothèque logicielle open source qui fournit une interface Python pour les réseaux de neurones artificiels. Keras agit comme une interface pour la bibliothèque TensorFlow.

4.8. Matplotlib

Matplotlib est une bibliothèque du langage de programmation Python destinée à tracer et visualiser des données sous formes de graphiques.



Figure 32: Les outils utilisé.

5. Résultat et discussion

D'après l'analyse des résultats obtenus, On constate les remarques suivantes :

D'après la figure de précision et d'erreur suivantes, la précision de l'apprentissage et validation augmente avec le nombre d'époque, ceci reflète qu'à chaque époque le modèle apprend plus d'informations.

De même, l'erreur d'apprentissage et de la validation diminue jusqu'à stable.

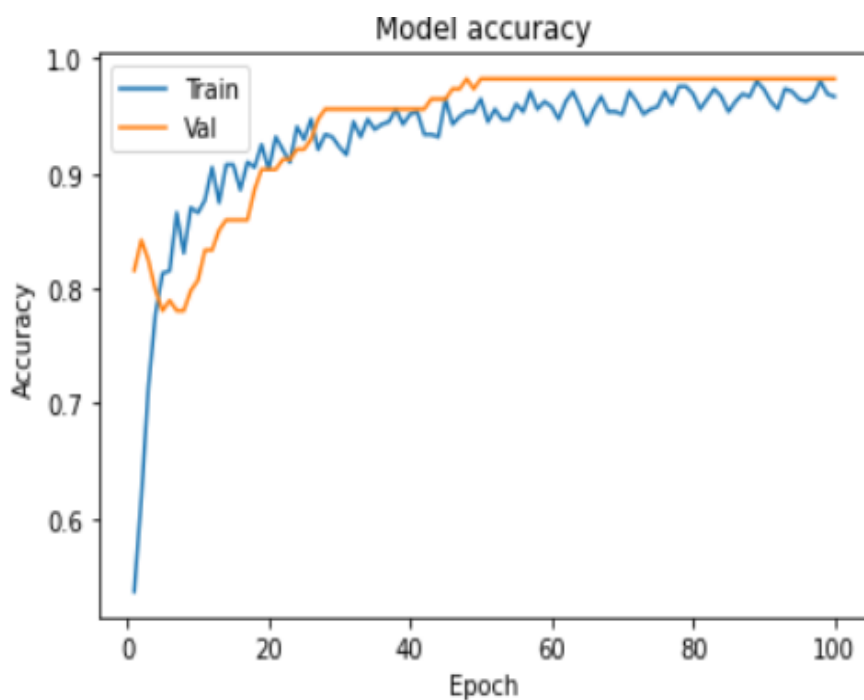


Figure 33: Précision de modèle

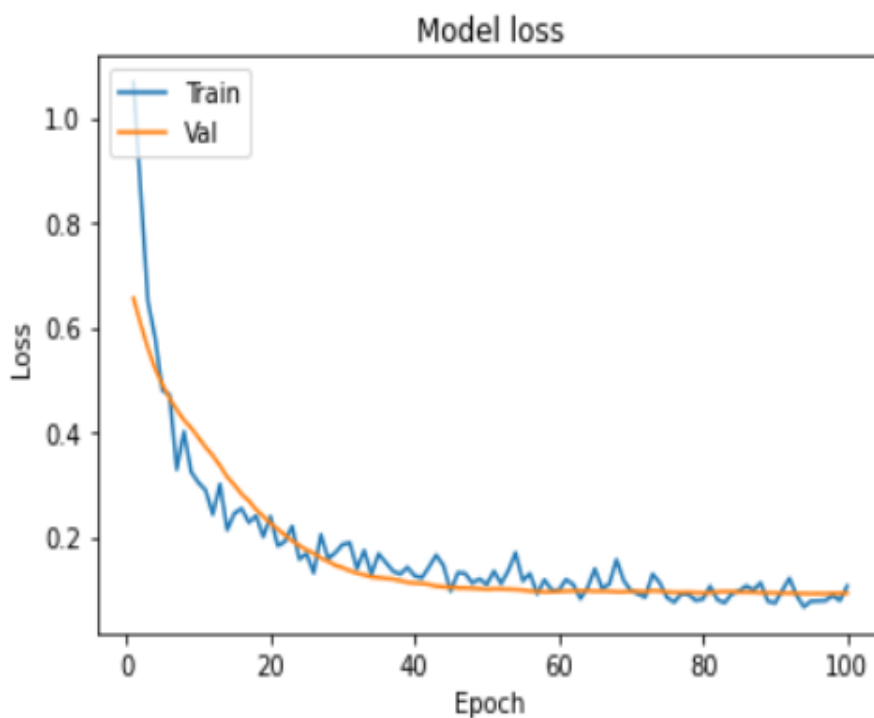


Figure34: Erreur de modèle

D'après la matrice de confusion suivante nous remarquons que le modèle est bien classé 110 (70+40) lignes de données et il est mal classé seulement 4 (2+2) lignes de données.

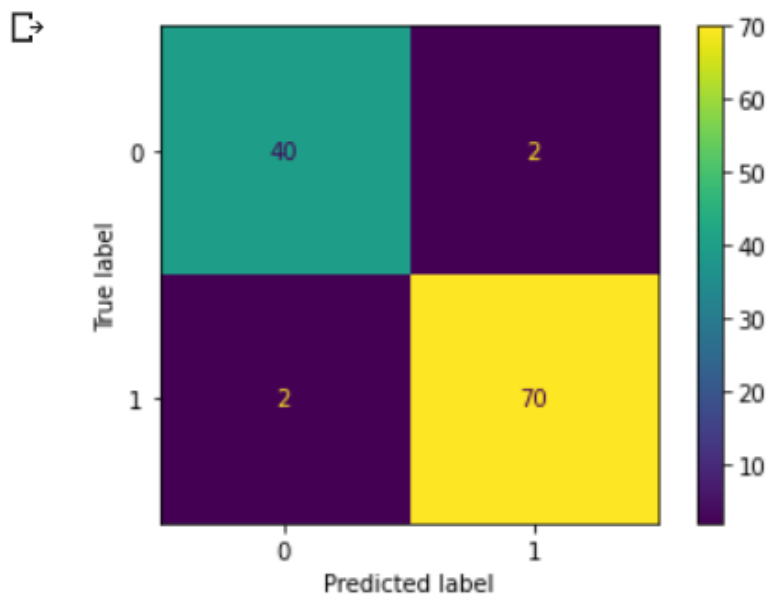


Figure 35: Matrice de confusion

5.1. Résultat final

Au final, le modèle que nous avons construit a donné des résultats très honorables, car après avoir calculé la précision du modèle, le résultat était de 96.49%

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} = \frac{70 + 40}{70 + 2 + 2 + 40} = 0,96491228$$

5.2. Classement de notre résultat

De la table suivante nous présentons une petite comparaison de notre résultat avec les résultats de travaux connexes.

Tableau 6: Comparaison des résultats

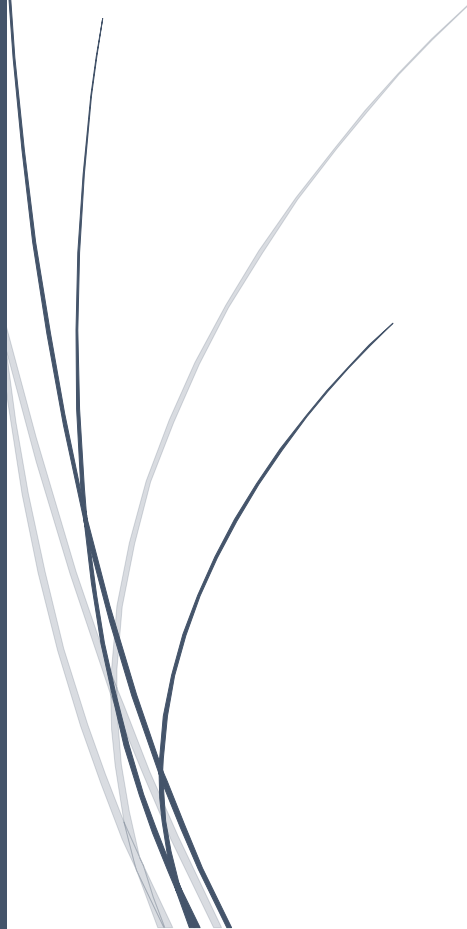
Ref	Accuray
[25]	83.74%
[27]	83%
Notre model	96.49%

6. Conclusion

Nous avons présenté dans ce chapitre notre contribution de classification base sur le réseau CNN pour l'aide à la diagnostic de cancer de sein , pour cela on a utilisé crée un modèle depuis le zéro avec différents couches et on a montré les différents résultats. Le résultat obtenue a montré que le model CNN a donnée meilleurs résultat avec ce type de donne de Wisconsin.

A dark blue vertical bar runs down the left side of the page. A blue arrow points from the right side of this bar towards the text.

Conclusion générale & perspectives



Dans ce travail, nous nous sommes intéressés à l'étude des systèmes automatisés de diagnostic assisté par ordinateur pour le diagnostic du cancer du sein, en particulier la description et la classification des masses mammaires basées sur la base de données de diagnostic du cancer du sein du Wisconsin.

Les entreprises visant à construire des systèmes d'assistance informatique se concentrent sur l'amélioration des résultats de classification. Dans ce cadre, nous avons mené une étude sur les ouvrages contextuels en termes de méthodes, de techniques et d'applications, ainsi que d'en extraire les lacunes de certains travaux et de choisir la voie de notre approche.

La base de données de diagnostic du cancer du sein du Wisconsin (WDBC) est un ensemble de données structuré qui est décrit par des caractéristiques cytologiques. Dans cet mémoire, notre premier objectif était de chercher à identifier des moyens d'améliorer les performances de classification dans un ensemble de données basé sur les réseaux de neurones convolutifs (CNN).

Bien que ces réseaux soient conçus pour des données non structurées, en particulier pour des données d'images, ils ont fait leurs preuves dans le domaine de la reconnaissance d'images. Afin de profiter de CNN pour améliorer les performances de classification des données structurées, nous avons suggéré d'utiliser le type CNN-1D pour essayer d'atteindre de bons résultats. Les résultats ont montré que CNN pouvait obtenir de bonnes performances de classification sur l'ensemble de données sélectionné, où la précision atteignait 96.49%. Les résultats obtenus sont intéressants, et ils peuvent être considérés comme des résultats compétitifs avec les méthodes proposées dans la littérature.

Ce mémoire a permis d'identifier plusieurs perspectives de vue que nous résumons en ces quelques lignes :

1. Chercher à améliorer les résultats par le développement du modèle proposé.
2. Testez le modèle proposé sur une autre dataset.



REFERENCES
BIBLIOGRAPHIES

- [1] : JP, G. (2020). L'évaluation : moteur du processus de planification sanitaire" Plan National Cancer 2015-2019". Algerian Journal of Health Sciences, 2(3), 53-61.
- [2]: <https://gco.iarc.fr/today/online-analysis-map> (12/03/21)
- [3] : <https://larvlortho.com/cancer-vads/le-cancer> (12/03/21)
- [4] : <https://youmatter.world/fr/causes-cancer-augmente-risques-cancerigene> (13/03/21)
- [5] : <https://www.cos-eu.com/sinformer-sur-le-cancer/mieux-comprendre-les-differents-types-de-cancer> (13/03/21)
- [6] : HAS, I. (2010). Tumeur maligne, affection maligne du tissu lymphatique ou hématopoïétique, Cancer du sein. 2010.
- [7] : HAS, I. (2010). Tumeur maligne, affection maligne du tissu lymphatique ou hématopoïétique, Cancer du sein. 2010.
- [8] : <https://www.biron.com/fr/centre-du-savoir/petit-guide-biron/cancer-du-sein/>
- [9] : Horde P, <http://www.e-cancer.fr/les-cancers/generalites/les-facteurs-de-risque.html>,2010.
- [10]: Khan, S., & Ahmad, S. (2019). Role of Angiotensin Converting Enzyme (ACE) Gene Polymorphism in Breast Cancer among North Indian Population. Annals of International Medical and Dental Research, 5(6), 17-23.
- [11]: Simos, D., Catley, C., van Walraven, C., Arnaout, A., Booth, C. M., McInnes, M., ... & Clemons, M. (2015). Imaging for distant metastases in women with early-stage breast cancer: a population-based cohort study. CMAJ, 187(12), E387-E397.
- [12] : <https://www.pactonco.fr/les-differents-types-de-cancers-du-sein> (12/04/2021)
- [13] : <https://www.nationalbreastcancer.org/types-of-breast-cancer> (23/03/21)
- [14]: Betty Smoot PT, D. P. T. S., Meredith Wampler PT, D. P. T. S., & Topp, K. S. (2009). Breast cancer treatments and complications: implications for rehabilitation. Rehabilitation Oncology, 27(3), 16.
- [15]: <https://blogs.nvidia.com/blog/2016/07/29/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai>
- [16] : A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," in IBM Journal of Research and Development, vol. 3, no. 3, pp. 210-229, July 1959, doi: 10.1147/rd.33.0210.
- [17] : <https://towardsdatascience.com/the-differences-between-artificial-and-biological-neural-networks-a8b46db828b7> (04/05/21)

- [18] Uijlings, J. R. R., van de Sande, K. E. A., Gevers, T. & Smeulders, A. W. M. (2013). Selective Search for Object Recognition. *International Journal of Computer Vision*, 104 (2), 154–171. doi :10.1007/s11263-013-0620-5
- [19] Ren, S., He, K., Girshick, R. B. & Sun, J. (2015). Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks. *CoRR*, abs/1506.01497. arXiv : 1506.01497. Récupéré à partir de <http://arxiv.org/abs/1506.01497>
- [20] [https://R%C3%A9seau neuronal convolutif](https://R%C3%A9seau_neuronal_convolutif) (16/04/21)
- [21] REVIEW/ Deep learning/ Yann LeCun 1,2, Yoshua Bengio3 & Geoffrey Hinton4,5 / 28 MAY 2015 | VOL 521 | NATURE /doi:10.1038/nature14539/25/04/21
- [22] <https://towardsdatascience.com/automated-machine-learning-hyperparametertuning-in-python-dfda59b72f8a> , Will Koehrsen , Jul 3, 2018 (21/04/21)
- [23] : A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," in *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210-229, July 1959, doi: 10.1147/rd.33.0210.
- [24]: Samala RK et al (2016) Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. *Med Phys* 43(12):6654–6666.
- [25]: Jadoon MM et al (2017) Three-class mammogram classification based on descriptive CNN features. *Hindawi Biomed Res Int*. [https:// doi.org/10.1155/2017/3640901](https://doi.org/10.1155/2017/3640901)
- [26]: Gallego-Posado JD et al (2016) Detection and diagnosis of breast tumors using deep Convolutional Neural Networks. In: *Research Group on Mathematical Modeling School of Mathematical Sciences Universidad EAFIT Medellin, Colombia*, pp 115-121
- [27]: Amit G et al (2017) Classification of breast MRI lesions using small size training sets: comparison of deep learning approaches. In: *Proceedings of SPIE 10134, medical imaging 2017: computer aided diagnosis, 101341H*. <https://doi.org/10.1117/12.2249981>
- [28]: Antropova N, Huynh B, Giger Maryellen (2017) Performance comparison of deep learning and segmentation-based radiomic methods in the task of distinguishing benign and malignant breast lesions on DCE-MRI. In: *Proceedings of SPIE 10134, medical imaging 2017: computer-aided diagnosis, 101341G*. <https://doi.org/10.1117/12.2255582>
- [29]: Antropova N et al (2017) A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med Phys* 44(10):5162–5171

[30]: Antropova N, Huynh B, Giger M (2018) Recurrent neural networks for breast lesion classification based on DCE-MRIs. In: Proceedings of SPIE 10575, medical imaging 2018: computer-aided diagnosis, 105752M. <https://doi.org/10.1117/12.2293265>

[31]: Antropova HA, Giger ML (2018) Use of clinical MRI maximum intensity projections for improved breast lesion classification with deep convolutional neural networks. J Med Imaging 5(1):014503. <https://doi.org/10.1117/1.JMI.5.1.014503>