

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE ECHAHID HAMMA LAKHDAR D'EL-OUED



FACULTE DES SCIENCES EXACTES
DEPARTEMENT: INFORMATIQUE

Mémoire de Fin D'étude
Présenté pour l'obtention du Diplôme de
MASTER ACADEMIQUE

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : Systèmes Distribués et Intelligence Artificielle

Réalisée par :

- ZEGHBIB AYA**
- HABI NESRINE**

Thème

**Amélioration des performances de La classification
Des tumeurs mammaires à l'aide d'une méthode de
Segmentation appropriée basée deep Learning**

Soutenue le:.....Devant le jury

Professeure MOHAMMED AMINE YAGOUB.....président

Professeure BEN BERKA KAMAL.....Rapporteur

Professeure HAMMOUD MERIEM.....Encadreur

Année Universitaire: 2022/2023

REMERCIEMENT

*La réalisation de ce mémoire a été possible grâce à Allah et au concours de plusieurs personnes à qui nous voudrions témoigner toute notre reconnaissance. Tout d'abord, nous souhaitons exprimer toute notre gratitude à notre encadrant de ce mémoire, le **Dr.Hamoud Meriem**, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à nourrir nos réflexions.*

cœur.

Un grand merci à nos familles pour leur confiance, leur aide, leur générosité et leur soutien, qui ont grandement facilité notre travail.

Nous tenons également à exprimer notre reconnaissance envers nos amis et collègues qui nous ont apporté leur soutien moral et intellectuel tout au long de notre démarche.

Dédicace

À l'âme de mon père

À ma très cher mère adorée qui m'a aidé, grâce à leur prière et à leur bénédiction.

*À tous mes frères ,a mes très chères sœurs , À ma grande famille et tous mes amis,À tous les étudiants de la faculté Informatique surtout les étudiants de la
2ème année master promotion 2023*

Aya et Nesrine

Résumé:

Une nette augmentation de l'incidence du cancer du sein au cours de la dernière décennie a été observée, ceci dit, la détection précoce grâce à des programmes de dépistage a engendré une baisse des taux de cancer du sein avancé ainsi que de mortalité. En effet, le dépistage par mammographie représente l'une des techniques d'imagerie médicale les plus utiles pour détecter et diagnostiquer précocement le cancer du sein. Cependant, selon les travaux de la littérature, la mammographie de dépistage passe à côté de 10 à 30 % des cas de cancer, ce qui entraîne un taux de faux négatifs pouvant atteindre 50 % selon le type de lésions et la densité mammaire. Dans ce sens, les recherches s'orientent vers des approches basées sur l'apprentissage automatique et l'apprentissage profond. Cependant, les méthodes d'apprentissage automatique et d'apprentissage en profondeur approuvent toujours une limitation lors de la classification des masses mammaires dans les catégories de seins denses où le tissu mammaire pathologique partage les mêmes caractéristiques avec le tissu dense normal, générant des résultats de classification erronés. Par conséquent, nous avons contribué l'approche d'apprentissage par transfert qui considère un modèle U-net pré-entraîné sur les images IRM cérébrales et nous l'avons adapté à la segmentation des mammographies denses. Par la suite ; un SVM est implémenté à des fins de classification par la prise en compte de la caractérisation de la texture des masses mammaires. Des résultats robustes ont été rapportés en atteignant une précision de 99% dans la distinction des masses malignes et bénignes.

Mots clés:

Cancer du sein, Segmentation, Modèle U-net, Apprentissage par transfert, Classification, Apprentissage profond, Apprentissage automatique, Caractérisation de texture, SVM.

Abstract:

Despite an increase in breast cancer incidence over the past decade, early detection through screening programs has led to a lowering of both advanced breast cancer and mortality rates. Indeed, mammography screening presents one of the most useful medical imaging techniques for detecting and diagnosing breast cancer early. However, according to research, screening mammography misses 10% to 30% of cancer cases, resulting in a false-negative rate of up to 50% depending on the kind of lesions and breast density. In this sense, researches are directed towards machine learning and deep learning based approaches. However, both machine learning and deep learning methods still approve a limitation when classifying breast masses within dense breast categories where the pathological breast tissue shares same characteristics with the normal dense tissue, generating a misclassified pathology. Hence, we contributed Transfer learning approach that considers a pre-trained U-net model on brain MRI images and adopts it to dense mammograms segmentation. An SVM is implemented for classification purposes by the consideration of masses texture characterization. Robust results have been reported through reaching 99% accuracy in distinguishing malignant and benign segmented masses.

Key words:

Breast cancer, Segmentation, U-net model, Transfer learning, Classification, Deep learning, Machine learning, Texture characterization, SVM.

الملخص:

زيادة واضحة في حالات سرطان الثدي خلال العقد الأخيرة تم ملاحظتها، ومع ذلك، أدى الكشف المبكر من خلال برامج الفحص إلى انخفاض معدلات حالات سرطان الثدي المتقدمة وكذلك معدلات الوفاة. فعلاً، فإن الفحص بواسطة التصوير بالأشعة الماموغرافية يمثل واحدة من أكثر تقنيات التصوير الطبي فائدة في الكشف وتشخيص سرطان الثدي في مراحله المبكرة. ومع ذلك، وفقاً للأبحاث الأدبية، يتم تفويت 10 إلى 30% من حالات السرطان عن طريق فحص الماموغرافية، مما يؤدي إلى معدل كاذب سلبي يمكن أن يصل إلى 50%. وفقاً لنوع الأورام وكثافة الثدي. في هذا السياق، تتجه الأبحاث نحو نهج يعتمد على التعلم الآلي والتعلم العميق. ومع ذلك، تظل أساليب التعلم الآلي والتعلم العميق تعاني من قيود عند تصنيف الكتل الثديية في فئات الثدي الكثيفة حيث يشترك النسيج الثدي المرضي في نفس الخصائص مع النسيج الكثيف الطبيعي، مما ينتج عنه نتائج تصنيف غير صحيحة. ولذلك، قدمنا نهج التعلم من خلال النقل، الذي يأخذ في الاعتبار نموذج U-net مُعد مسبقاً على صور الرنين المغناطيسي للدماغ وقمنا بتكييفه لتقسيم الماموغرافية الكثيفة. بعد ذلك، تم تنفيذ SVM لأغراض التصنيف من خلال مراعاة توصيف نسيج الكتل الثديية. تم تقديم نتائج قوية حيث تم الوصول إلى دقة تصل إلى 99% في تمييز الكتل الخبيثة والحميدة.

الكلمات المفتاحية:

سرطان الثدي، تقسيم، نموذج U-net، التعلم من خلال النقل، التصنيف، التعلم العميق، التعلم الآلي، توصيف النسيج، SVM.

TABLE DES MATIÈRES

Table des matières	i
Table des figures	iv
Liste des tableaux	1
Introduction générale	1
1 Les appuis majeurs de l'analyse d'image, la vision par ordinateur et l'intelligence artificielle : initiation à l'apprentissage automatique et l'apprentissage profond	4
1.1 Introduction	5
1.2 La vision par ordinateur	5
1.3 L'analyse d'image	5
1.3.1 Analyse de bas niveau d'image	6
1.3.2 Analyse de haut niveau d'image	6
1.4 Analyse de la texture d'image	6
1.5 Evocation des tendances actuelles d'apprentissage automatique et d'apprentissage profond	8
1.5.1 L'apprentissage automatique	8
1.5.2 L'apprentissage profond (Deep Learning)	12
1.6 Les modèles des réseaux de neurone convolutifs	16
1.6.1 U-NET	16
1.6.2 VGG-16	17

1.6.3	Res-Net	18
1.6.4	U-NET 3D	19
1.6.5	U-NET++	20
1.7	Conclusion	22
2	Etat de l'art sur le pronostic automatique de cancer du sein	23
2.1	Introduction	24
2.2	Préambule au cancer du sein	24
2.2.1	Définition	24
2.2.2	Influence du cancer du sein sur l'anatomie de cet organe	25
2.2.3	Les tumeurs mammaires et leurs caractéristiques	26
2.2.4	Les types de cancer du sein	27
2.2.5	Facteurs favorisant le cancer du sein	28
2.3	La mammographie de dépistage de cancer du sein	28
2.3.1	Définition et types de la mammographie	28
2.3.2	Limites de la mammographie durant le dépistage de cancer du sein	31
2.4	Diagnostic assisté par ordinateur (CAD) du cancer du sein	32
2.4.1	Détection des tumeurs assistée par ordinateur (CAdE)	34
2.4.2	Diagnostic des tumeurs assistée par ordinateur (CAdE)	34
2.5	L'apprentissage automatique pour le pronostic automatique du cancer du sein	35
2.6	L'apprentissage profond pour le pronostic automatique du cancer du sein	37
2.7	Conclusion	39
3	Problématique et approche proposée	40
3.1	Introduction	41
3.2	Problématique et plan d'action	41
3.3	Segmentation des tumeurs mammaires basée Transfer Learning	47
3.3.1	La base des mammographies DDSM	47
3.3.2	Pipeline de la segmentation des tumeurs mammaire basé Transfer Learning	49
3.4	Classification des tumeurs mammaires basée machines à vecteurs de support (SVM) :	51
3.4.1	Pipeline basé SVM pour la classification des tumeurs mammaire	51
3.5	Architecture générale de l'approche proposée	61
3.6	Conclusion	62

4	Evaluation des performances obtenues	63
4.1	Introduction	64
4.2	Evaluation de l'approche proposée de pronostic automatique de cancer de sein .	64
4.2.1	Environnement de développement	64
4.3	Outils utilisés	64
4.3.1	MATLAB (MATrix LABoratory)	64
4.3.2	Python	65
4.4	Logiciel	66
4.5	Résultats aboutis de la segmentation des tumeurs mammaires basée Transfer Learning	68
4.5.1	Cropping des tumeurs	70
4.6	Evaluation des performances de l'approche proposées de pronostic automatique de cancer du sein par les machines à vecteur de support (SVM)	71
4.7	Explication du code	75
4.8	Fonctionnement du système développé	80
4.9	Conclusion	83
5	Conclusion générale	84

TABLE DES FIGURES

1.1	Image présentant différentes résolutions texturales.	7
1.2	Exemples de textures structurales.	7
1.3	Exemples de textures aléatoires.	8
1.4	Architecture des réseaux de neurones.	10
1.5	Classification en deux classes : (a) les données sont linéairement séparables et (b) utilisation d'une séparatrice non linéaire.	11
1.6	Exemple de classification basée l'algorithme des KNN.	12
1.7	Structure générale d'un réseau de neurone convolutif (CNN).	13
1.8	Exemple d'une opération de Max pooling de taille 2×2	14
1.9	Principe de ReLU.	15
1.10	Structure du réseau U-net.	17
1.11	Architecture de VGG-16.	18
1.12	Architecture de Res-Net.	19
1.13	Architecture de U-NET 3D.	20
1.14	Architecture de U-NET++.	21
2.1	L'anatomie du sein.	26
2.2	Appareil de génération des mammographies.	29
2.3	Processus de mammographie numérique.	30
2.4	Système CAD de segmentation et classification des tumeurs mammaires.	34
3.1	Mammographies appartenant à la base de donnée DDSM avec leurs masques respectifs.	48
3.2	Résultats du pré-traitement appliqué sur les mammographies.	50

3.3	Pipeline suivi pour la classification des tumeurs mammaires.	52
3.4	Courbes de Zipf et de Zipf d'une tumeur mammaire bénigne.	54
3.5	Courbe de Zipf et Zipf inverse d'une tumeur mammaire maligne.	55
3.6	Application des filtres de Gabor dans le domaine fréquentiel.	58
3.7	Parties réelles des filtres de Gabor utilisés.	59
3.8	Parties réelles des tumeurs filtrées.	59
3.9	Exposition de la magnitude de la réponse relative au filtre de Gabor une fois avoir réalisé la convolution de la tumeur maligne avec une banque de filtres de Gabor. Par la suite, nous prenons toutes les sorties du filtre et nous calculons des statistiques locales pour générer le vecteur des descripteurs texturaux comme proposé par les auteurs dans [67].	60
3.10	Architecture de l'approche proposée de diagnostic automatique du cancer du sein.	62
4.1	logo Matlab.	65
4.2	logo python.	66
4.3	logo Kaggle.	67
4.4	Les courbes de la précision de classification ainsi que de la minimisation de loss fonction suite à l'apprentissage sur la base de données DDSM.	69
4.5	Les courbes de la précision de classification ainsi que de la minimisation de loss fonction suite à l'apprentissage sur la base de données Figshare.	70
4.6	Processus de cropping des tumeurs.	71
4.7	Matrice de confusion relative à l'approche proposée.	73
4.8	Code CLAHE.	75
4.9	Code du filtre Median.	76
4.10	Code du modèle de segmentation U-net.	76
4.11	Code Post-traitement(cropping) images.	77
4.12	Code Créer un fichier.CSV.	78
4.13	Code de ACP.	79
4.14	Code des SVM.	80
4.15	Interface login du système mmo-classification.	81
4.16	Interface principale du système mmo-classification. Avec affichage de resultat de SVM model.	82

INTRODUCTION GÉNÉRALE

Le cancer du sein est une maladie périlleuse pour les êtres humains ayant le taux d'incidence et de mortalité le plus élevé parmi tous les cancers hors cancer de la peau (mélanome)[1]. Les statistiques concernant cette maladie dangereuse sont inquiétantes :

- Le taux de létalité du cancer du sein compte pour 17 pourcent de la mortalité par cancer [2].
- Selon le rapport de l'American Cancer Society (ACS) en 2019, plus de 3 millions de femmes ont reçu un diagnostic de cancer du sein [5].
- Selon l'American Cancer Society, plus de 279 000 cas ont été signalés aux États-Unis en 2020 et on estime que 43 600 femmes mourront du cancer du sein en 2021 [3].
- Une contribution de 7 pourcent à la mortalité dans toutes les mortalités dues à tous les types de cancer dans le monde, le cancer du sein a été reconnu comme la quatrième cause de mortalité dans le monde en 2020, spécifiquement répandue chez les femmes [4].

Le cancer du sein survient généralement en raison d'une croissance anormale ou irrégulière des cellules du sein [4]. Avec une division plus rapide que les cellules ordinaires, ces cellules commencent à s'accumuler et à former une masse dans le sein, ce qui le rend trop dangereux pour la patiente. Mammographie, thermographie (TG), US, CT, IRM, EHR et séquences génomiques sont les modalités disponibles en vue d'utilisation dans le diagnostic du cancer du sein. Au cours des 30 dernières années, le taux de mortalité par cancer du sein a régulièrement diminué grâce à la détection précoce, à une sensibilisation accrue et à l'amélioration des traitements [5]. L'imagerie médicale joue un rôle clé dans la médecine moderne car elle permet la visualisation non invasive des structures internes et des processus métaboliques du corps humain en détail. Actuellement, le dépistage standard du cancer du sein est la mammographie numérique pour

les femmes présentant un risque moyen de cancer du sein où les statistiques publiées dans des revues systématiques récentes liées aux modalités d'imagerie montrent que 50 pourcent des bases de données sont relatives à la mammographie, 20 pourcent d'US, 18 pourcent d'IRM et 8 pourcent de HP [6]. Cependant, la détection précoce par mammographie est complexe et difficile pour le type de sein dense où le tissu mammaire normal est souvent diagnostiqué à tort comme des lésions anormales ou vice versa [2]. Cela induit en erreur des faux positifs et des faux négatifs et, dans certains cas, des examens supplémentaires inutiles. L'identification automatique des masses mammaires et surtout la classification de leur pathologie présentent une aide cruciale aux radiologues et aux experts dans leur prise de décision. Une telle classification peut réduire considérablement le nombre de faux positifs qui conduisent à des pratiques cliniques inutiles et à des biopsies inconfortables. L'analyse des différentes pathologies mammaires et leur classification est une tâche difficile pour les radiologues, car elle implique une détection précise des lésions ainsi qu'une prise de décision complexe. Pour répondre à ces défis, l'analyse automatisée des images médicales a connu une évolution significative pour améliorer l'interprétation de l'imagerie médicale en fournissant une "deuxième opinion" aux radiologues grâce au développement des systèmes CAD d'aide au diagnostic médical du cancer du sein. En effet, le diagnostic assisté par ordinateur est une solution prometteuse pour de nombreux programmes de dépistage de cancer du sein [4] et les technologies de l'intelligence artificielle ont récemment réussi à augmenter les chances de reconnaître une anomalie à un stade précoce afin d'éviter les biopsies inutiles et de diminuer le taux de mortalité [7]. Concrètement, le but des solutions d'IA est d'aider les cliniciens à effectuer leur travail avec plus d'efficacité et de précision, et non de les remplacer. Ce partenariat nécessite la confiance des experts cliniques, et une voie vers la confiance est la compréhension.

Une approche en multi-cascade de modèles de segmentation et de classification de tumeurs mammaires sera contribué dans ce mémoire où des algorithmes d'apprentissage en profondeur et des algorithmes d'apprentissage automatique seront combinés.

D'une part, la segmentation sera basée sur le modèle U-net consistant en des couches de convolution ainsi que de sous-échantillonnage pour classifier les pixels mammaires pendant le processus de segmentation. En effet, ces couches convolutives se composent de différents filtres extrayant les caractéristiques profondes de la tumeur mammaire. Par la suite, les tailles des cartes d'entités seront réduites par l'incorporation des couches de sous-échantillonnage et en conséquence, cela diminuera la résolution des caractéristiques spatiales des mammographies.

D'autre part, les machines à vecteurs de supports (SVM) appartenant aux modèles d'apprentissage automatique, utiliseront les régions d'intérêt obtenues lors du processus de segmentation

pour la classification.

Ce mémoire sera structuré comme suit :

Chapitre 1 : Les appuis majeurs de l'analyse d'image, la vision par ordinateur et l'intelligence artificielle : initiation à l'apprentissage automatique et l'apprentissage profond

Durant ce chapitre, les notions de base de l'analyse d'image ainsi que de la vision par ordinateur seront évoquées. De plus, le boom actuel relatif à l'apprentissage automatique et l'apprentissage profond sera détaillé.

Chapitre 2 :Etat de l'art sur le pronostic automatique de cancer du sein

Au long du chapitre II, nous détaillerons la maladie dangereuse du cancer du sein où les paradigmes de l'intelligence artificielle ont contribué au diagnostic automatique.

Chapitre 3 : Problématique et approche proposée

Le chapitre III sera dédié à la problématique de ce mémoire où l'approche proposée sera à son tour détaillée.

Chapitre 4 : Evaluation des performances obtenues

Le chapitre IV est consacré à l'évaluation ainsi que la discussion des résultats obtenus.

Nous achèverons ce mémoire par une conclusion générale tout en soulignant quelques perspectives.

CHAPITRE 1

LES APPUIS MAJEURS DE L'ANALYSE
D'IMAGE, LA VISION PAR ORDINATEUR
ET L'INTELLIGENCE ARTIFICIELLE :
INITIATION À L'APPRENTISSAGE
AUTOMATIQUE ET L'APPRENTISSAGE
PROFOND

1.1 Introduction

En s'appuyant sur les mathématiques, l'informatique, l'analyse d'image et la vision par ordinateur ; l'intelligence artificielle a bénéficié de cet ensemble de technologies importantes pour l'exécution automatique des tâches ainsi que leur compréhension, en vue de la prise de décision. Actuellement, l'intelligence artificielle est devenue une technologie jouant un rôle très important dans le processus de diagnostic des maladies notamment le cancer du sein sur lequel sera bâti ce mémoire. En effet, un aide sera fourni aux scientifiques à fin d'aboutir au diagnostic précis. Dans ce chapitre, nous évoquerons les outils proposés dans le domaine de l'analyse d'images ainsi que de l'intelligence artificielle.

1.2 La vision par ordinateur

La technologie de la vision par ordinateur désigne un outil lié à l'intelligence artificielle et effectue l'analyse des images prises par les caméras pour leur identification, manipulation et compréhension [8].

La vision humaine et la déduction sont les appuis principaux les plus importants pour la vision par ordinateur [8]. En effet, la vision par ordinateur est reliée avec l'intelligence des humains en imitant l'analyse des images perçues et la capacité du cerveau à le faire.

1.3 L'analyse d'image

Une image est un ensemble de pixels présentant une intensité lumineuse particulière [9]. Analyser une image revient à en extraire des descripteurs pertinents. Actuellement, l'analyse d'images contient de nombreux outils aussi divers que l'intelligence artificielle.

Une liaison directe s'impose entre la vision par ordinateur et l'analyse d'image, dans ce sens, une capture des images par caméra est effectuée pour les passer à un système d'analyse d'images comportant de nombreux algorithmes présentant certaines caractéristiques [10]. L'ultime but après le processus d'acquisition d'image désigne la conversion des pixels en objets pour l'interprétation du contenu de l'image.

Nous distinguons différents niveaux d'analyse d'images :

1.3.1 Analyse de bas niveau d'image

Il s'agit d'un niveau bas dont le but est de convertir l'image d'entrée en une image qui convient à l'autre niveau de traitement des objets en plus d'améliorer leur qualité. Les algorithmes qui ont été implémentés pour l'analyse de bas niveau d'image améliorent la qualité afin de supprimer toutes les informations inutiles par le processus de filtrage [11].

1.3.2 Analyse de haut niveau d'image

C'est un niveau élevé dont le but est de reconnaître et de classer des objets, citons l'exemple du processus de détection d'un visage. Ce niveau se concentre également sur le contenu et la connaissance des images, et l'une des méthodes les plus couramment utilisées dans le domaine de l'analyse et de l'interprétation des images est l'intelligence artificielle. En effet, la perception humaine est simulée dans la première étape de la vision par ordinateur à partir de l'information contenue dans l'image [11].

1.4 Analyse de la texture d'image

Bien que le concept de texture soit naturel pour certains humains. Ceci dit, durant longtemps un concept approprié n'a pu être fixé [12]. Quelques définitions ont été énoncés, citons celle que la texture désigne la répétition spatiale du même motif mais dans des directions différentes de l'espace. Harralick l'a décrit comme un phénomène bidimensionnel [12][13] : la première dimension concerne la description des éléments principaux ou primitifs (motifs) qui composent la texture tandis que la deuxième dimension concerne la description des systèmes spatiaux de ces primitives. Ce qui est important dans l'analyse des textures, c'est la précision, ce qui signifie qu'elle peut sembler différente par rapport au niveau de précision. En effet, nous pouvons distinguer ce qui génère les tissus primitifs de base comme le sable, les pierres, etc. En étant très peu précis, où les motifs de base apparaissent grossièrement par rapport aux échelles d'observation. La figure suivante illustre également le phénomène de transfert d'informations texturales [14].

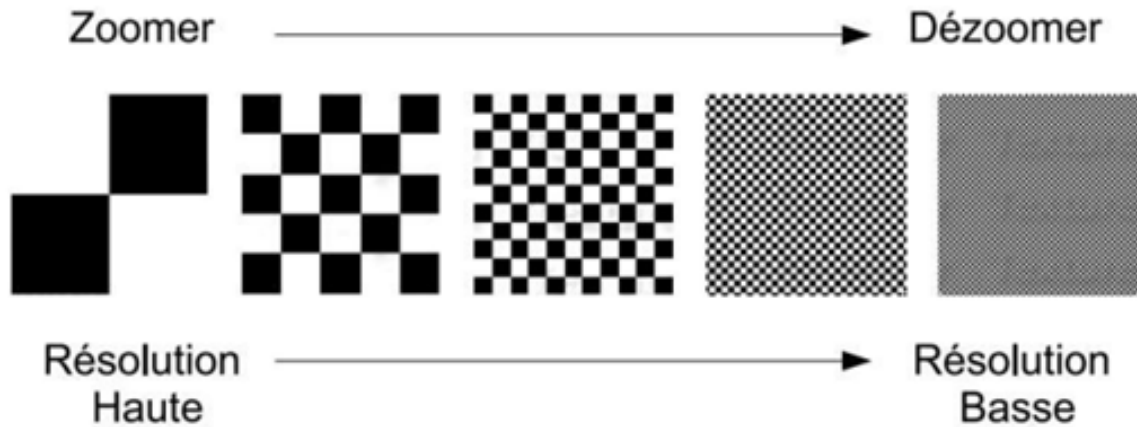


FIGURE 1.1 – Image présentant différentes résolutions texturales.

Divers types de texture existent, citons :

- **Textures structurelles**

Elle s'agit d'une répétition spatiale moins régulière que le modèle de base (primitif). Elle est d'orientations différentes de sorte que la texture cyclique est un sous-ensemble de la texture organisée, Par exemple, un mur qui montre une brique appartient à ce genre de texture [15] comme indiqué sur la figure suivante.



FIGURE 1.2 – Exemples de textures structurelles.

- **Textures aléatoires**

Il n'y a pas de modèle particulier que nous pouvons localiser ou détecter au-dedans de ces textures. En effet, une désorganisation est ressentie et aucune règle d'une disposition particulière est suivie, cependant la texture demeure homogène car la primitivité revient au niveau des

pixels. C'est ce que nous appelons le tissage fin. Citons l'exemple de la laine, de l'herbe [15]. Une texture aléatoire est exposée sur la figure suivante.

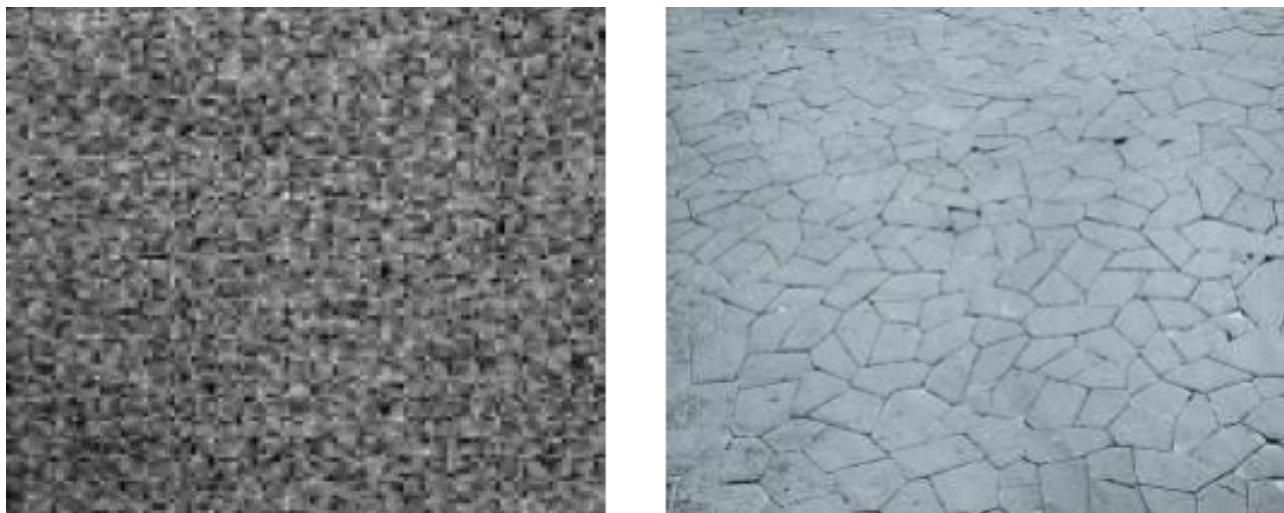


FIGURE 1.3 – Exemples de textures aléatoires.

- **Textures directionnelles**

La texture directionnelle n'est pas complètement aléatoire où ne nous distinguons pas d'éléments de la structure de base. En effet, cette texture a ses propres orientations [15].

1.5 Evocation des tendances actuelles d'apprentissage automatique et d'apprentissage profond

L'énorme progrès dans le développement des modèles d'apprentissage automatique (Machine Learning) et de l'apprentissage profond (Deep Learning) a conduit à de grandes avancées dans le traitement d'images. En effet, suite à l'explosion de la quantité de données acquises par jour en continuant à croître, cela a imposé les chercheurs à réfléchir au processus d'analyse de cette énorme quantité de données et à la mise en œuvre de modèles d'apprentissage automatique ainsi que d'apprentissage profond [16].

1.5.1 L'apprentissage automatique

L'un des objectifs de l'intelligence artificielle est de simuler l'intelligence humaine, en particulier l'apprentissage de nombreuses tâches, deux manières d'apprentissage existent, le premier consistant à mémoriser clairement tous les exemples possibles afin de pouvoir les répéter et le second consistant à apprendre à généraliser en extrayant des règles implicites de nombreux

exemples dans le but de les appliquer à de nouvelles situations qui n'existaient pas auparavant [17]. L'apprentissage par généralisation est relativement difficile du fait qu'il nécessite l'utilisation de règles qui ne sont pas rencontrées dans les exemples, ceci désigne l'apprentissage automatique [17] appartenant à l'intelligence artificielle dans lequel s'effectue l'analyse statistique des données d'entraînement.

Le développement de machines dotées de la capacité d'apprendre sans être explicitement programmées pour apprendre une tâche désigne une définition adéquate de l'apprentissage automatique [17].

Les algorithmes d'apprentissage automatique utilisent deux étapes comme suit :

Phase d'apprentissage : où un nombre important d'exemples est traité pour l'extraction des règles implicites construisant ces données. À cette phase, le modèle est utilisé [17].

Phase d'inférence : où le modèle formé sera appliqué sur de nouvelles entrées. En effet, suite à l'extraction des règles implicites, le modèle peut généraliser à des entrées inconnues [17].

Nous allons exposer quelques algorithmes d'apprentissage automatique dans ce qui suit :

- **Les réseaux de neurones**

C'est une sous-technologie de l'intelligence artificielle. Les neurones sont des programmes architecturaux informatiques extraits de la structure des cellules neurobiologiques et ces cellules artificielles sont liées les unes aux autres et disposées en couches afin que les données puissent être identifiées à l'aide d'exemples et diffusées afin d'aider à la prise de décision [11]. Nous présentons sur la figure 1.4 l'architecture des réseaux de neurones.

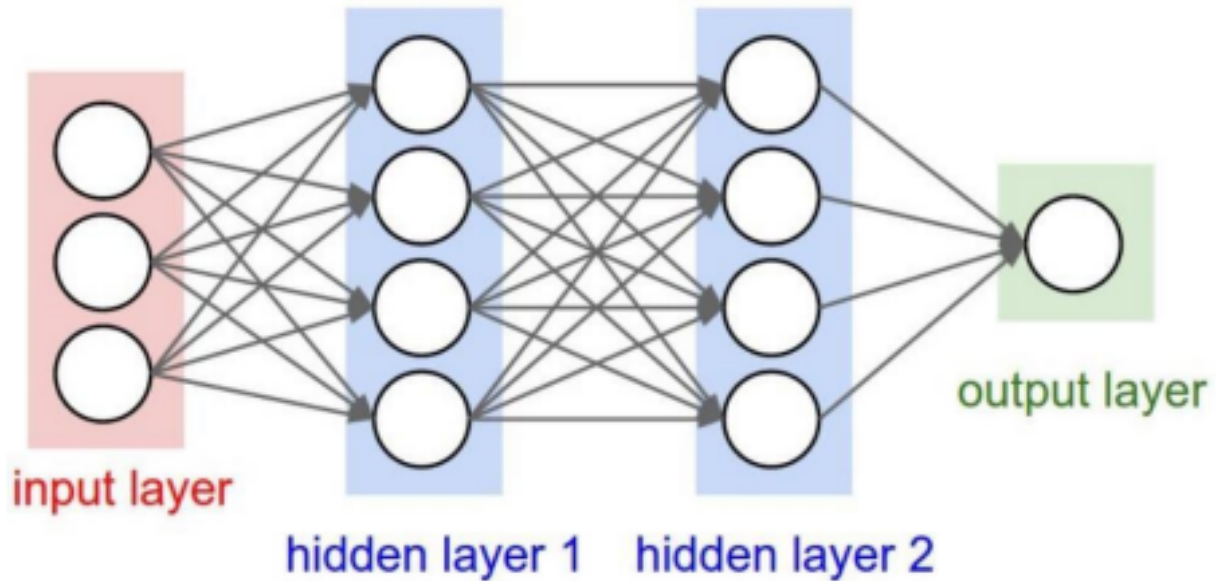


FIGURE 1.4 – Architecture des réseaux de neurones.

- **Les machines à vecteurs de support (SVM)**

Introduits par Vladimir au début des années 90, ils sont les vecteurs de soutien pour le développement de la recherche sur les techniques d'apprentissage humain et sont liés à la théorie mathématique solide par opposition aux méthodes de réseaux neuronaux. Les SVM ont été largement utilisées dans la régression et la classification binaire. Actuellement, les SVM ont intégrées plusieurs domaines de l'ingénierie et de la recherche, comme le diagnostic médical, la biologie, le marketing et la reconnaissance de l'écriture manuscrite [18]. C'est le modèle d'apprentissage automatique le plus célèbre se reposant sur deux choses fondamentales [19], à savoir, la fonction du noyau ainsi que les marges maximales. En général, l'objectif principal de SVM est de fixer des limites entre diverses classes [20].

Les SVM peuvent résoudre des problèmes de classification de données linéairement séparables et celles non linéairement séparables comme exposé sur la figure 1.5 [19].

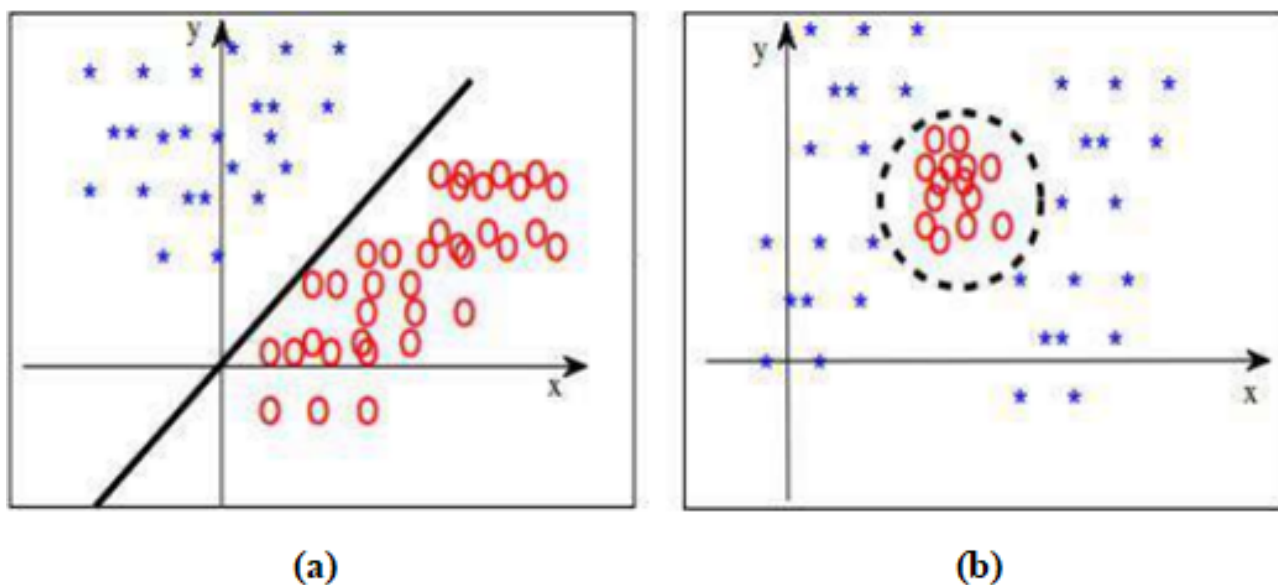


FIGURE 1.5 – Classification en deux classes : (a) les données sont linéairement séparables et (b) utilisation d'une séparatrice non linéaire.

Les SVM présentent des avantages comme le fait qu'ils présentent peu de paramètres à ajuster, leur vitesse élevée d'apprentissage, être efficace lorsque la dimension de la zone est supérieure au nombre d'échantillons d'apprentissage, l'utilisation de la fonction du noyau où les problèmes linéaires ainsi que non linéaires sont traités [19].

Ainsi que des inconvénients comme la diminution des performances dans le cas où le nombre d'échantillons est inférieur au nombre de descripteurs [19].

• **Les arbres de décision**

L'objectif principal du processus d'induction de l'arbre de décision est de construire automatiquement un ensemble de règles conditionnelles incompatibles avec un échantillon d'entraînement afin qu'il s'agisse d'un arbre bidimensionnel composé de plusieurs règles pouvant être interprétées simplement par un expert humain [21] [22].

L'un des objectifs de l'apprentissage automatique pour ces règles est de séparer les objets de manière itérative de l'échantillon d'apprentissage pour séparer les échantillons d'objets, à condition que la plupart d'entre eux aient idéalement la même valeur pour les variables de sortie. Une fois le modèle construit, nous pouvons déduire le nouvel objet en le répartissant dans l'arborescence de haut en bas à l'aide de tests sur ses attributs d'entrée.

• **Les K plus proches voisins (KNN : K Nearest Neighbours)**

C'est l'un des algorithmes d'apprentissage qui sont soumis à la supervision et dépendent de la mesure et de la comparaison directe entre le vecteur de l'instance de la bibliothèque d'apprentissage et le vecteur caractéristique de l'instance que nous voulons classer et ceci en se basant

sur le calcul d'une distance entre ces cas, puis la classification sera effectuée en attribuant la catégorie majoritaire parmi les catégories similaires [23] comme indiqué sur la figure 1.6.

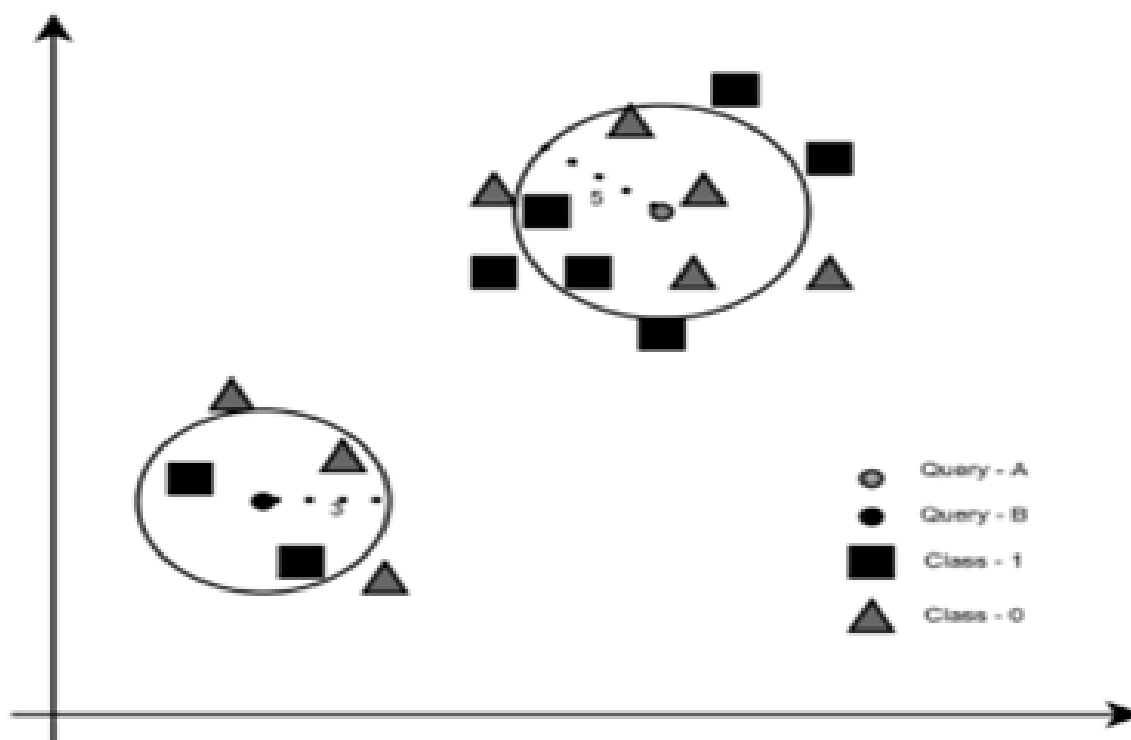


FIGURE 1.6 – Exemple de classification basée l'algorithme des KNN.

1.5.2 L'apprentissage profond (Deep Learning)

L'apprentissage profond est un domaine qui a été jugé avancé par rapport à l'apprentissage automatique en possédant de fortes capacités basées sur l'analyse. Il est utilisé dans des applications telles que la vision par ordinateur et le traitement du langage et a pu offrir une amélioration très significative des performances par rapport à l'apprentissage automatique. En effet, l'apprentissage profond utilise à son tours des fonctionnalités pour détecter et résoudre des problèmes sans l'intervention du facteur humain [24].

L'apprentissage profond est basé sur "les réseaux de neurones artificiels", constitués de milliers de "neurones" effectuant chacun des opérations diverses. Il est utile de mentionner que les résultats de la première couche de neurones servent d'entrée aux calculs d'une seconde couche et ainsi de suite.

Il existe de nombreux types de structures liées aux réseaux de neurones [25] : le réseau de neurones récurrents (RNN) qui est spécifique aux tâches complexes et fut utilisé sur un large

éventail de domaines tels que l'écriture manuscrite, la reconnaissance du langage ou bien la prédiction. D'autre part, nous distinguons les réseaux de neurones convolutifs (CNN) fonctionnant dans un large domaine tel que la reconnaissance d'images et de vidéos. Dans ce sens, nous nous intéressons particulièrement à ce type de réseaux de neurones.

• **Réseaux de neurone à convolution (CNN)**

Le réseau neuronal « CNN » est une technique d'apprentissage profond composée d'un groupe de plusieurs couches et d'un champ local, utilisant la communication locale et le partage du poids dans le but d'améliorer l'efficacité et les performances. En effet, la structure profonde aide les réseaux à apprendre de nombreuses caractéristiques ou descripteurs complexes qu'aucun réseau simple ne peut apprendre. Le principe de CNN est d'obtenir des entrées (images) dans les rangs supérieurs et de les fusionner en entrées complexes dans les rangs inférieurs [27]. Nous présentons sur la figure 1.7 la structure générale d'un réseau de neurone convolutif.

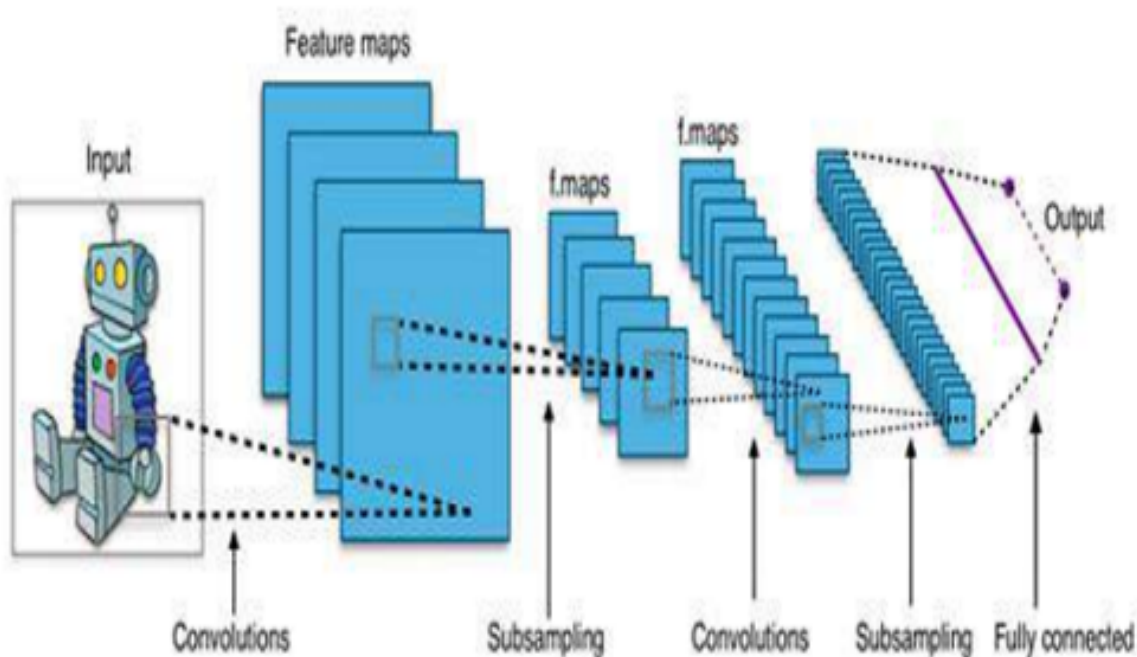


FIGURE 1.7 – Structure générale d'un réseau de neurone convolutif (CNN).

Nous mentionnons les principales couches de traitement d'un CNN [25] :

1. **Couche de convolution (CONV) :**

Les couches convolutives sont le cœur des réseaux convolutifs et sont constituées d'un ensemble de neurones rectangulaires plus un petit champ récepteur qui s'étend le long de la profondeur de l'espace d'entrée. En effet, la couche de convolution n'est pas seulement une convolution de l'image par rapport à la couche précédente, les morgues convolutives

sont déterminées par des poids, et le rôle de la couche convolutive est de déterminer les produits des neurones connectés à la région en calculant leurs poids et en se connectant. La fonction d'activation ReLU est conçue pour appliquer le processus d'activation dans le sens du nombre d'activations résultant de la couche précédente [26].

2. Couche de pooling :

Il s'agit d'une couche responsable de l'échantillonnage ainsi que de la réduction de la taille spatiale des cartes d'entités.

La valeur maximale de la partie image est renvoyée par max pooling en se basant sur le principe de la prise d'échantillon au maximum et nous pouvons prendre tous les éléments du noyau, qui est l'agrégation de la moyenne [26].

Notons que le max pooling est le plus utilisé vu son succès durant les expérimentations. La taille du noyau est de 2 x 2, générant une réduction des dimensions d'entrée des processus d'activation et la profondeur reste constante [26]. Nous exposons sur la figure I.8 un exemple du processus max pooling.

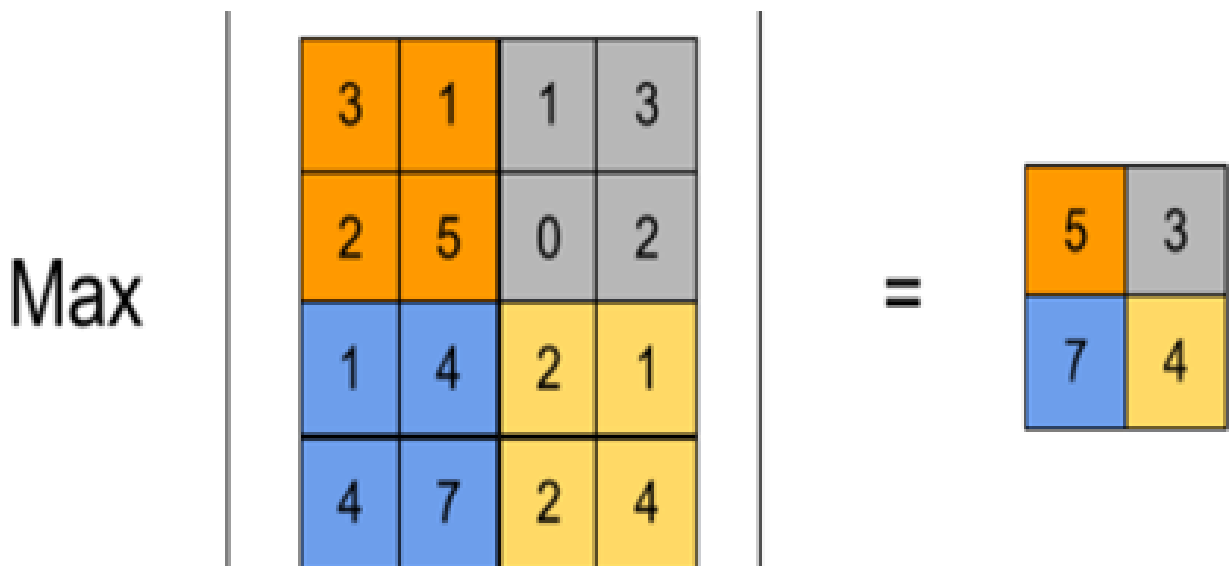


FIGURE 1.8 – Exemple d'une opération de Max pooling de taille 2 × 2.

L'échantillonnage a comme but de réduire partiellement la taille des données d'entrée tout en permettant de [26] :

- La gestion facile ainsi que la réduction de la représentation des entrées.
- La réduction de la taille des données pour ne pas nécessiter une grande puissance de calcul lors du traitement des données.

- Le maintien de l'efficacité du processus d'apprentissage en extrayant uniquement les caractéristiques nécessaires.

3. Couche de correction ReLU :

Elle est souvent utilisée après la couche convolutive et se présente sous la forme d'une couche séparée [27]. Généralement, la correction ReLU est préférable, vu le fait qu'il en résulte un réseau neuronal plusieurs fois plus rapide en effet, nous appliquons la fonction ReLU sur chaque pixel de l'image après convolution où l'ultime but consiste à remplacer chaque valeur négative par un 0 comme exposé sur la figure 1.9.

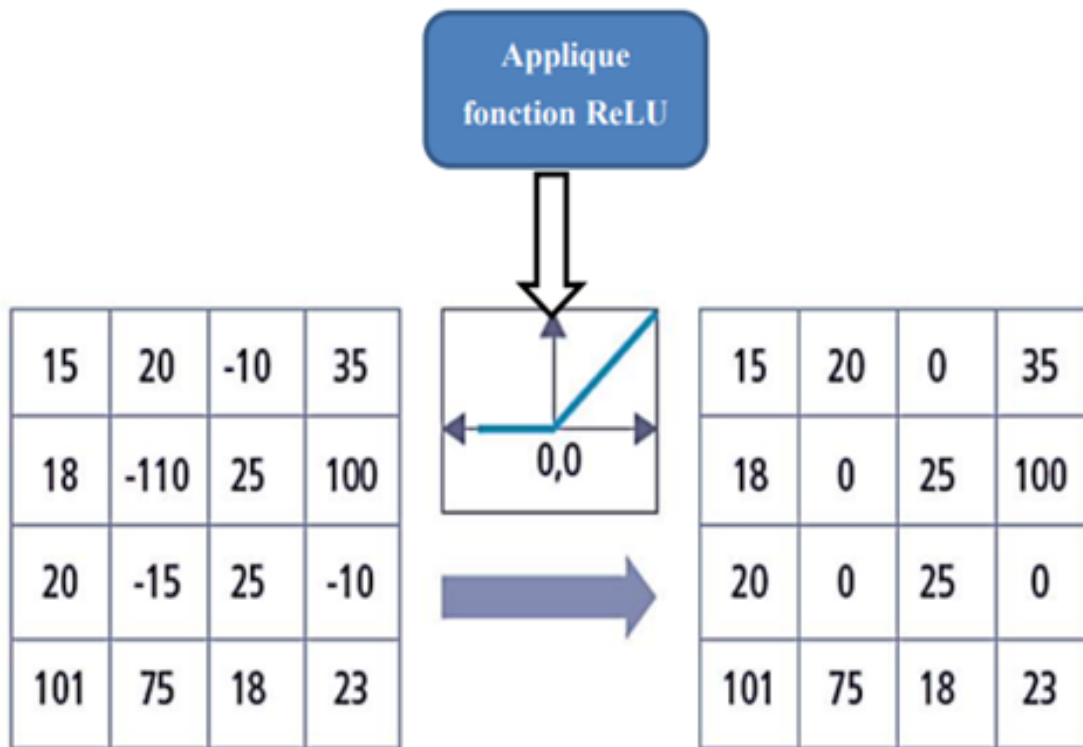


FIGURE 1.9 – Principe de ReLU.

4. Couche entièrement connectée :

C'est l'inférence avancée pour les réseaux neuronaux et se fait au moyen d'un ensemble de couches entièrement connectées après un ensemble de couches de convolution et d'agrégation. La couche entièrement connectée (fully-connected) est un Perceptron multicouche traditionnel qui utilise une fonction d'activation softmax dans la couche de sortie [26].

Notons que cette couche finale permet de classifier l'image. En effet, elle renvoie un vecteur de taille N où éventuellement N désigne le nombre de classes formulant notre problème de classification d'images. L'ultime but de la couche entièrement connectée est l'utilisation

des descripteurs pour classifier l'image d'entrée dans la classe adéquate en considérant comme référence l'ensemble de données d'apprentissage.

1.6 Les modèles des réseaux de neurone convolutifs

1.6.1 U-NET

Il s'agit d'un modèle d'encodage à quatre blocs accompagnés d'un chemin d'expansion de quatre blocs de décryptage [28] [29].

U-NET se compose de deux chemins comme nous pouvons le constater sur la figure 1.14. Le premier désigne le chemin de l'encodeur où nous l'utilisons pour la capture du contexte d'une image, ceci, à travers l'effectuation de couches de convolution ainsi que de max pooling. En effet, une carte de caractéristiques relative à une image sera générée permettant la réduction de la taille de l'image et la diminution du nombre de paramètres du réseau.

Le second chemin désigne le décodeur permettant une localisation précise de l'objet à segmenter grâce à l'effectuation d'une convolution transposée.

Il est utile de mentionner que U-net applique la fonction d'activation ReLU suite à chaque convolution pour détecter les non-linéarités entre les données tout en accélérant l'apprentissage.

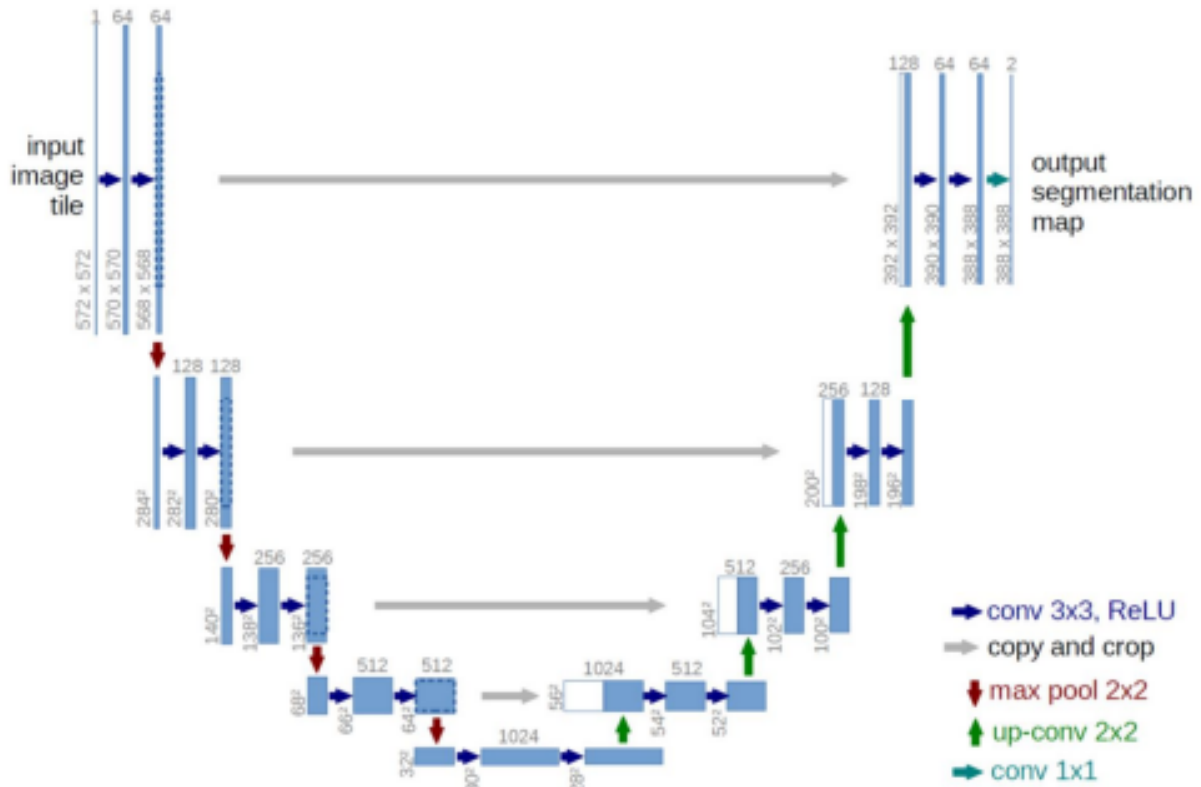


FIGURE 1.10 – Structure du réseau U-net.

1.6.2 VGG-16

Il a été proposé par Simonian et Zeisserman de consister en une structure régulière (3×3) de couches convolutives et (2×2) de couches d'agrégation (max pooling) chacune 3 couches nécessitant environ 123 millions de poids et couches convolutives et nécessitant également un rapport de 15 millions de poids [26].

Le VGG-16 global contient 31 milliards de processus de traitement d'images dans les opérations de prévision. La précision du modèle VGG-16 qui a été appliquée sur image-net était de l'ordre de 88,7pourcent[26]. Nous exposons sur la figure I.15 l'architecture de VGG-16.

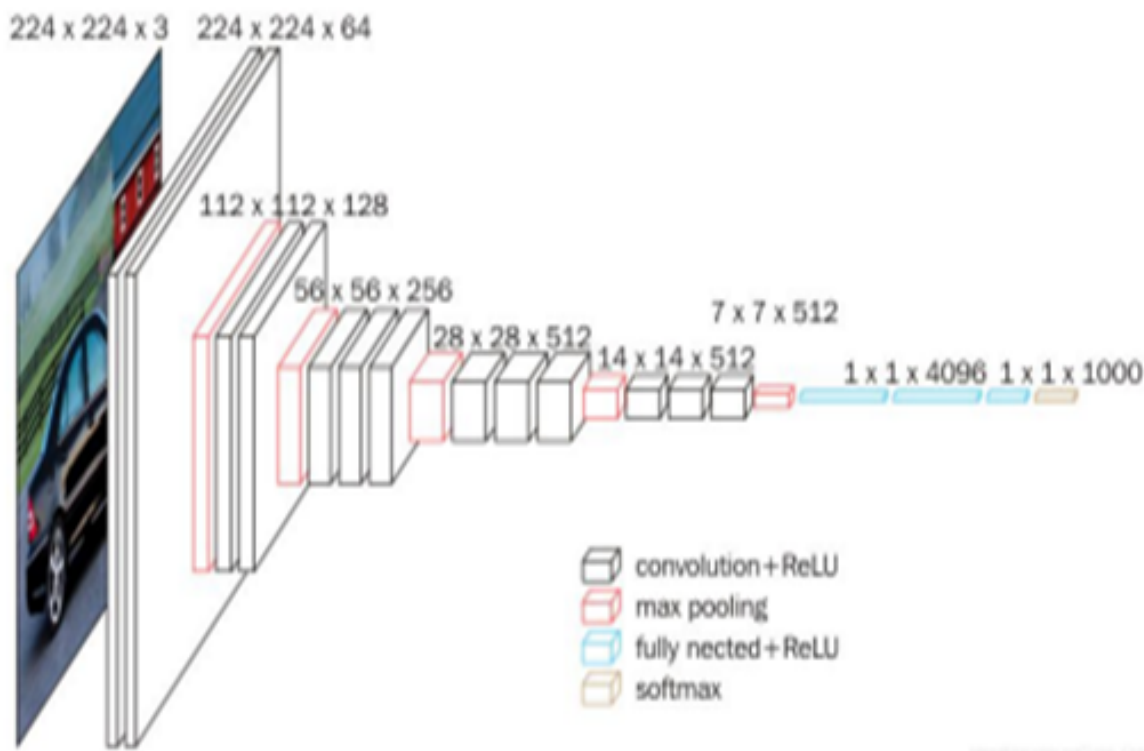


FIGURE 1.11 – Architecture de VGG-16.

1.6.3 Res-Net

ResNet, ou Réseau de neurones résiduels, est une architecture de réseau de neurones profonds utilisée pour la reconnaissance d'images et d'autres tâches de vision par ordinateur. En effet, ce réseau est généré à partir de blocs de construction appelés modules de résidus permettant au réseau de maintenir une performance élevée même dans le cas où le nombre de couches est relativement augmenté.

L'ultime innovation de ResNet réside dans l'utilisation de connexions résiduelles entre les couches du réseau permettant à l'information de sauter au-dessus de certaines couches dans le réseau. En effet, l'avantage de ces connexions résiduelles est de résoudre le problème de la disparition du gradient, pouvant survenir lorsque l'on entraîne un réseau de neurones relativement profond. Suite à l'utilisation des connexions résiduelles, le modèle ResNet peut réaliser un apprentissage pertinent avec plus de 100 couches, ce qui a permis d'aboutir à des performances de pointe en ce qui concerne les tâches de vision par ordinateur, citons la classification d'images et la détection d'objets [30]. Nous présentons sur la figure 1.12 l'architecture de ResNet.

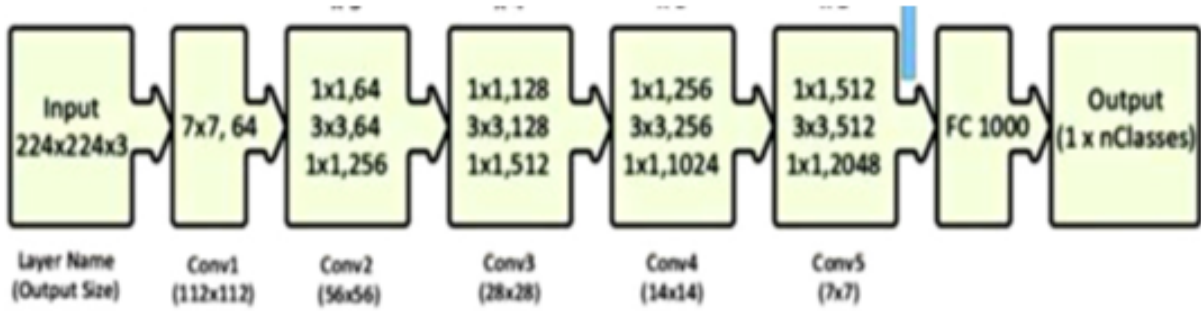


FIGURE 1.12 – Architecture de Res-Net.

1.6.4 U-NET 3D

U-Net 3D est une architecture de réseau de neurones profonds utilisée pour la segmentation d'images médicales en trois dimensions (3D). Cette architecture est basée sur le réseau U-Net, qui est conçu pour la segmentation d'images en deux dimensions (2D).

Le réseau U-Net 3D utilise une architecture d'encodeur-décodeur, dans laquelle l'image d'entrée est progressivement réduite de taille à travers une série de couches d'encodeur effectuant l'extraction des caractéristiques de l'image. En effet, par la suite ces caractéristiques seront utilisées à travers une série de couches de décodeur pour la génération d'une carte de segmentation ayant la même taille que celle de l'image d'entrée.

L'architecture U-Net 3D comprend également des connexions résiduelles pour aider à résoudre le problème de la disparition du gradient, surgissant lors de la phase d'apprentissage des réseaux de neurones profonds.

U-Net 3D est fortement recommandé dans les applications de segmentation d'images médicales, citons la segmentation des tumeurs ainsi que la segmentation des organes dans les images de tomographie (CT) et d'imagerie par résonance magnétique (IRM) [24]. Nous présentons sur la figure ci-dessous l'architecture de U-Net 3D.



FIGURE 1.13 – Architecture de U-NET 3D.

1.6.5 U-NET++

U-Net++ est une architecture de réseau de neurones profonds utilisée pour la segmentation d'image sémantique. En effet, c'est une extension de l'architecture U-net, qui est populaire pour sa capacité à effectuer la segmentation d'image de manière précise en utilisant un petit nombre de données lors de la phase d'apprentissage.

L'architecture U-Net++ consiste en une approche de réseaux de neurones en pyramide, au sein de laquelle les blocs de convolution présents dans les couches descendantes sont connectés aux blocs de convolution correspondants dans les couches ascendantes. Ainsi, une communication multi-échelle entre les niveaux différents de résolution de l'image sera présente. En effet, cette approche permet de mieux capturer les détails fins de l'image et de ce fait améliore la précision de segmentation [32]. Nous présentons sur la figure 1.14 l'architecture de U-Net ++.



FIGURE 1.14 – Architecture de U-NET++.

1.7 Conclusion

Durant ce chapitre, nous avons discuté quelques concepts de base tels que (types, définitions, principe) relatifs aux apprentissages profonds ainsi que l'apprentissage automatique. En effet, nous nous intéressons à ces deux concepts fondamentaux pour l'élaboration de notre travail. Quant au deuxième chapitre, nous évoquerons tout ce qui concerne le diagnostic automatique du cancer du sein. En effet, l'accent sera mis sur la formulation associée aux modèles d'intelligence artificielle discutés durant ce chapitre.

Le flou dans les images est un phénomène essentiellement convolutif. Il est en grande partie, dû au fait que la profondeur de champ d'un appareil photographique ne peut être infinie. De plus, les caméraphones sont souvent fabriqués avec des focales fixes, seule la vitesse d'obturation est commandée. En outre, les capteurs utilisés dans les caméraphones sont très limités dans des conditions de luminosité faible. Dans ces conditions, on récupère souvent des images bruitées et floues.

CHAPITRE 2

ETAT DE L'ART SUR LE PRONOSTIC
AUTOMATIQUE DE CANCER DU SEIN

2.1 Introduction

L'intégration et la disponibilité des études réalisées sur des images antérieures sont des éléments clés de pour la pratique des radiologues. En effet, durant le diagnostic du cancer, les radiologues examinent régulièrement les études précédentes pour détecter les éventuels changements tissulaires au fil du temps. Les données antérieures peuvent être incorporées pour enrichir les informations cliniques et ainsi améliorer la précision du diagnostic. Dans cette optique, les techniques d'apprentissage automatique et d'apprentissage en profondeur ont été développées pour aider à analyser ces vastes quantités de données et améliorer la précision du diagnostic. Récemment, le domaine le plus actif en imagerie médicale est la conception et l'implémentation de systèmes d'aide au diagnostic médical basés sur l'apprentissage automatique ainsi que l'apprentissage profond, également connus sous le nom de CAD (Computer-Aided Diagnosis). L'objectif ultime de ces systèmes est de fournir un deuxième avis aux radiologues pendant le processus de diagnostic.

Les systèmes CAD utilisent des algorithmes d'apprentissage automatique et profond pour analyser des images médicales et aider les radiologues à détecter et à diagnostiquer des anomalies ou des maladies. Ces systèmes peuvent aider à réduire les erreurs de diagnostic et à améliorer la précision du diagnostic, ce qui peut conduire à des résultats plus précoces et plus précis pour les patients.

La capacité des systèmes CAD à être performants repose sur leur aptitude à discriminer et généraliser tout au long du processus de classification de nouvelles structures, tout en réduisant les efforts des médecins et en améliorant le taux de survie des patients.

2.2 Préambule au cancer du sein

2.2.1 Définition

Le cancer est caractérisé par la prolifération désorganisée de cellules qui peut perturber le fonctionnement normal des organes et éventuellement conduire à la mort. Chaque organe et chaque tissu ont des types de cellules spécifiques qui sont disposées de manière structurée. Le cancer peut survenir à partir d'une seule cellule qui a acquis des mutations favorisant sa survie et sa division, formant ainsi une masse de tissu anormal appelée tumeur. Au fur et à mesure que le cancer progresse, la tumeur envahit les tissus sains voisins. La formation de métastases se produit lorsque certaines cellules cancéreuses se détachent de la tumeur d'origine, la tumeur primaire, et pénètrent dans la circulation sanguine ou lymphatique pour former de nouvelles

tumeurs dans d'autres parties du corps. Cette propagation du cancer rend le traitement plus difficile car les thérapies qui ciblent une zone spécifique ne sont plus aussi efficaces [33].

Le cancer est une maladie périlleuse dans laquelle les cellules anormales se divisent de manière incontrôlable et tentent de détruire les tissus corporels, ce qui la rend très dangereuse et potentiellement mortelle pour la vie humaine [4]. Le cancer a été reconnu comme la principale cause de décès dans le monde en raison de son nombre élevé de décès (10 millions) et de son nombre d'incidences (19,3 millions) en 2019 et selon les estimations du centre international de recherche sur le cancer (CIRC), le nombre approximatif d'incidence et le nombre de décès devrait augmenter à 29,5 et 16,4 millions respectivement en 2040. Le cancer est également appelé tumeur qui doit être détectée ou diagnostiquée à son stade initial pour guérir ses patients correctement et en temps opportun. La plupart des décès signalés dans le monde désignent le cancer du poumon (19 pourcent), suivi du cancer du côlon (10 pourcent), du cancer de l'estomac (9 pourcent), du cancer du sein (7 pourcent), du cancer du foie (6 pourcent), du cancer du pancréas (5 pourcent), cancer du cerveau (3pourcent), cancer de la peau (1pourcent).

Un des types de cancer mortel est le cancer du sein, qui peut entraîner la mortalité chez les femmes. Il est principalement causé par une croissance anormale des tissus mammaires et est généralement connu sous le nom de cellules bénignes et malignes. Il est très difficile de diagnostiquer le patient avec des cellules malignes à un stade précoce en raison de la faible traçabilité de la tumeur de petite taille.

2.2.2 Influence du cancer du sein sur l'anatomie de cet organe

L'anatomie du sein est composée de différents vaisseaux sanguins, tissus conjonctifs, canaux galactophores, lobules et vaisseaux lymphatiques comme mentionné sur la figure 2.1. Une tumeur se forme dans les canaux galactophores ou les lobules lorsque les tissus mammaires se développent anormalement et que la division cellulaire devient incontrôlée. La tumeur développée peut être bénigne ou maligne. Les tumeurs bénignes sont produites en raison des modifications structurelles mineures du sein et sont classées comme des tumeurs non cancéreuses. En revanche, les tumeurs malignes sont classées comme cancéreuses et sont en outre classés comme invasifs (carcinome invasif) ou non invasifs (carcinome in situ) [6]. Les tumeurs invasives du sein se propagent dans les organes environnants et créent des complications alors que les tumeurs non invasives restent confinées dans leur région et n'envahissent pas les organes voisins.

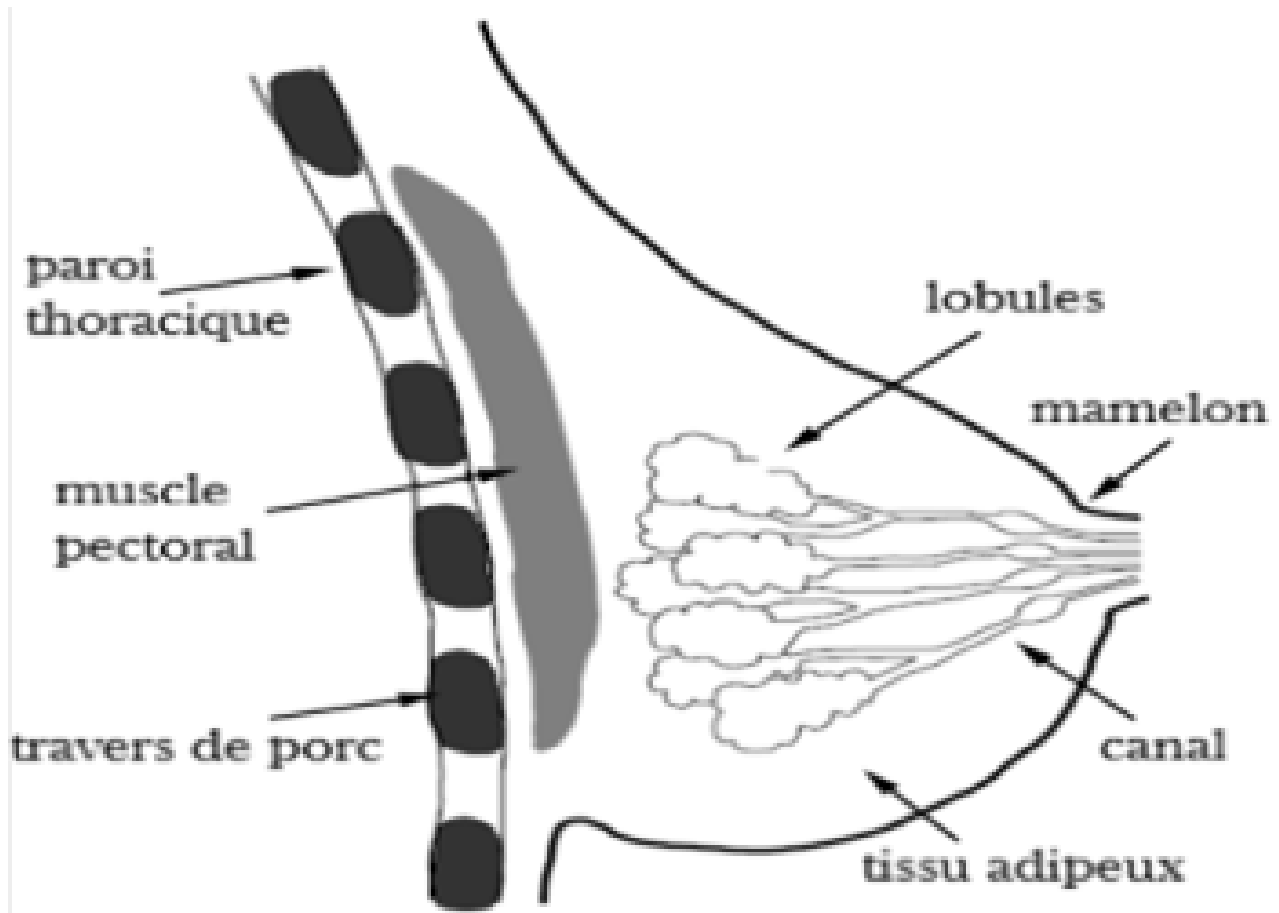


FIGURE 2.1 – L'anatomie du sein.

2.2.3 Les tumeurs mammaires et leurs caractéristiques

La tumeur existe dans le sein sous forme de masse qui peut être maligne ou bénigne et l'identification du type de pathologie de masse fait partie des processus de base qui doivent être retenus dans le diagnostic sur les mammographies de dépistage [34][35].

- Parmi les caractéristiques des tumeurs malignes figure le fait qu'elles ont une forme de contour irrégulier partiellement arrondi. En plus de cela, les masses malignes apparaissent généralement plus blanches que tout tissu qui les entoure. Tandis que les tumeurs bénignes se caractérisent par leurs formes rondes ou ovales. Les tumeurs malignes se propagent aux organes adjacents après l'ablation et sont accumulées [36].
- Les masses bénignes apparaissent généralement sous forme de régions denses dans les mammographies qui ont des limites lisses et bien circonscrites et les masses malignes ont

des limites floues et rugueuses et les deux sont accumulées dans une région spécifique avec de grandes tailles par rapport aux tailles des calcifications. Elles sont composées de plusieurs cellules similaires aux cellules normales [37]. Alors que les microcalcifications sont soit des calcifications lobulaires quasi bénignes soit des calcifications intracanales suspectes de malignité. Cependant, les deux peuvent exister dans le sein sous de nombreux formats à travers des points dispersés dans tout le sein, ou dispersés régionalement dans un plus grand volume de tissu mammaire, ou regroupés occupant un petit volume de tissu, ou existant sous forme de dépôts de calcium dans les conduits. La plupart de ces cas peuvent être classés comme bénins et ne nécessitent généralement aucun suivi. De plus, les micro-calcifications sont très petites par rapport à l'ensemble de la mammographie. Parmi les tumeurs bénignes se trouve la tumeur glandulaire[36].

2.2.4 Les types de cancer du sein

Il existe de nombreux types différents de cancer du sein [38] :

1. **Invasif vs in-situ** : Les cancers qui ont envahi les vaisseaux sanguins et lymphatiques et ont la capacité de se propager sont appelés "invasifs" ou "infiltrants". Le cancer du sein invasif, qui peut être sous forme de carcinome lobulaire invasif ou de carcinome canalaire invasif, peut se développer dans le tissu mammaire environnant, et peut être classé en stades 1 à 4. En revanche, les pré-cancers, également appelés "in situ", sont localisés et n'ont pas la capacité de se propager. Le carcinome canalaire in situ est un exemple de cancer in situ, qui est également appelé stade zéro.
2. **Sous-types histologiques** : Le sous-type histologique d'un cancer se réfère à la cellule d'origine du cancer. Les deux principaux sous-types histologiques de cancer du sein sont les cancers canalaire et lobulaire. Environ 80 à 85 pourcent des cancers du sein proviennent des canaux, tandis que les lobules du sein représentent environ 10 à 15 pourcent des cas. Dans les cancers canaux, les cellules se regroupent pour former une touffe.
3. **Sous-types moléculaires** : Cette classification est fondée sur l'expression de récepteurs spécifiques par les cellules cancéreuses, tels que le récepteur des œstrogènes (ER), le récepteur de la progestérone (PR) et le récepteur her2 (tyrosine kinase).

4. **Grade** : Cela est utilisé pour identifier l'apparence microscopique des cellules cancéreuses et déterminer leur degré de différenciation.
5. **Stade** : Le terme "stade" est employé pour décrire l'évolution d'un cancer. Il est déterminé en prenant en compte plusieurs caractéristiques telles que la taille de la tumeur, l'atteinte des ganglions lymphatiques et la présence de métastases à distance, ce qui permet d'évaluer la propagation du cancer.

2.2.5 Facteurs favorisant le cancer du sein

Nous mentionnons certains des risques conduisant à l'atteinte par le cancer du sein [38] :

- **Âge** : Les femmes âgées de 45 à 55 ans sont souvent considérées comme étant à risque de cancer du sein, car cette tranche d'âge présente une incidence plus élevée de cette maladie par rapport aux autres groupes d'âge.
- **Antécédents familiaux** : Le fait qu'un parent proche (au premier degré) d'une femme soit atteint d'un cancer du sein augmente son risque de développer cette maladie.
- **Gènes** : La présence de certains gènes tels que BRCA1 et BRCA2 augmente considérablement le risque de développer un cancer du sein.
- **Mode de vie** : Des habitudes de vie courantes telles que l'obésité, la diminution de l'activité physique, le tabagisme, la consommation excessive d'alcool et l'utilisation de contraceptifs oraux peuvent augmenter le risque de développer un cancer du sein.

2.3 La mammographie de dépistage de cancer du sein

2.3.1 Définition et types de la mammographie

Depuis sa découverte en 1913, la mammographie est considérée comme une clé essentielle pour la détection et le diagnostic précoces des lésions précieuses. Le dépistage par mammographie a aidé les radiologues à identifier le cancer du sein et plusieurs études ont montré son impact pour une réduction significative du taux de mortalité [3]. Étant donné que la mammographie de dépistage réduit la mortalité par cancer du sein de 20 à 22 pourcent, la mammographie est la modalité la plus largement utilisée pour détecter le cancer du sein [7].

Dans cette technique, les images de chaque sein sont généralement acquises à l'aide de deux différentes vues [1] : cranio-caudale (CC) imagée de haut en bas et blique médio-latérale (MLO) de gauche à droite.

La mammographie numérique a été introduite par une entreprise américaine en 2000. Elle a remplacé la mammographie standard par un détecteur numérique qui enregistre les rayons X et les signaux électriques sous forme de nombres. Les avantages de ce détecteur sont sa précision et sa clarté. La mammographie numérique est l'outil de dépistage privilégié pour les femmes ayant des seins denses [39]. Nous présentons sur la figure 2.2, l'appareil de génération des mammographies.



FIGURE 2.2 – Appareil de génération des mammographies.

La mammographie est une technique d'imagerie médicale qui utilise des rayons X pour produire des images des seins. Les rayons X ont la capacité de se propager à travers le corps humain et de traverser les tissus. Les images produites par la mammographie sont des images complètes qui sont ébauchées dans la radiographie, ainsi que la production d'une autre radiation qui réduit et diminue l'image radiographique [40].

Nous distinguons trois types de mammographies :

- **La mammographie diagnostique** : La mammographie diagnostique est une méthode utilisée pour détecter les problèmes mammaires tels qu'une masse ou une zone suspecte. Ce type de mammographie peut être réalisé si vous-même ou votre médecin détectez un problème lors d'un examen clinique des seins ou d'une mammographie de dépistage. Contrairement à

la mammographie de dépistage, la mammographie diagnostique prend plus de temps car elle consiste en la prise d'images détaillées et de vues de la poitrine sous différents angles pour examiner de près une zone spécifique de la poitrine. Les deux seins sont généralement examinés pour permettre aux médecins de comparer le tissu mammaire des deux seins [42].

- **La mammographie de dépistage** : La mammographie de dépistage est un examen médical utilisé pour détecter le cancer du sein chez les femmes qui ne présentent aucun symptôme ou problème mammaire. Cet examen consiste à examiner les deux seins afin de repérer des anomalies ou des masses de tissu mammaire qui pourraient être trop petites pour être détectées par palpation. La mammographie de dépistage permet de surveiller l'évolution du sein au fil du temps et de détecter le cancer du sein à un stade précoce [44].

- **Mammographie numérique (FFDM)** : La mammographie numérique est une technique d'imagerie qui permet de produire une image électronique du sein, contrairement à la mammographie traditionnelle qui utilise un film radiographique. Le processus de la mammographie numérique est similaire à celui de la mammographie standard, mais les images sont stockées dans un ordinateur plutôt que sur un film.

En général, deux projections sont réalisées pour chaque sein lors d'une mammographie numérique : la première est une projection cranio-caudale (CC) et la seconde est une projection medio-latérale oblique (MLO). La figure 2.3 présente schématiquement le processus de la mammographie numérique.

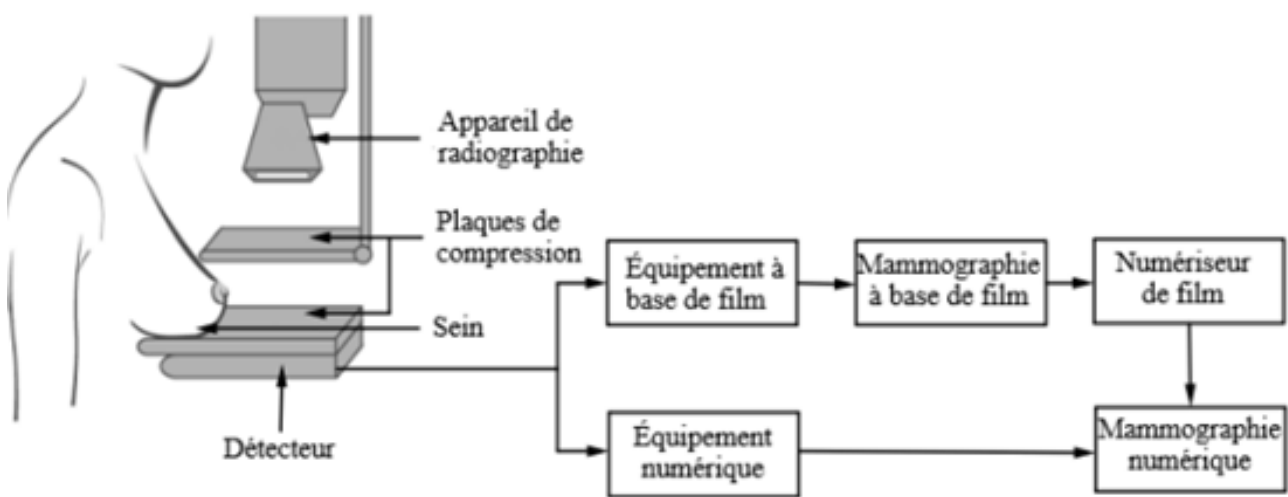


FIGURE 2.3 – Processus de mammographie numérique.

2.3.2 Limites de la mammographie durant le dépistage de cancer du sein

Il est important de noter que la mammographie ne détecte pas tous les cancers du sein et que pour les femmes ayant des seins denses, comme les jeunes femmes et les personnes atteintes de maladies fibrokystiques, la visualisation d'un cancer peut être difficile [45]. En effet, les femmes jeunes ayant des seins très denses présentent une réduction significative de la sensibilité de la mammographie, avec des taux de détection de l'ordre de 30 pourcent, par rapport à environ 80 pourcent pour les femmes plus âgées. Il est donc important de prendre en compte ces facteurs lors de l'évaluation d'un résultat de mammographie et de poursuivre des examens supplémentaires si nécessaire pour confirmer un diagnostic de cancer [39].

La détection d'anomalies suspectes dans les mammographies est une tâche fastidieuse et répétitive pour les radiologues. Sur mille mammographies analysées, seulement trois ou quatre peuvent être cancéreuses, ce qui peut conduire à des erreurs de diagnostic si une anomalie est négligée [40].

De plus, dans de nombreux cas, il peut y avoir un chevauchement entre les caractéristiques bénignes et malignes sur les images, ce qui rend difficile pour les radiologues de les différencier avec précision [40].

La mammographie est une méthode de dépistage courante pour détecter les cancers du sein, mais sa précision varie de 68 pourcent à 92 pourcent [40]. Selon des études, la mammographie peut ne révéler que 75 pourcent des cancers du sein [45]. Ce manque de précision peut être causé par plusieurs facteurs, tels que des défaillances techniques, la fatigue visuelle des radiologues entraînant des erreurs d'observation, un mauvais positionnement ou une compression insuffisante lors de la réalisation de la mammographie, la nature subtile de certaines caractéristiques mammographies des cancers du sein précoces, la superposition de structures dans les images, le grand nombre de cas normaux dans les programmes de dépistage, ainsi que les limites intrinsèques de la mammographie elle-même.

Il y a environ une chance sur dix qu'une mammographie produise un résultat faussement positif. Cela signifie que le test peut indiquer la présence de cancer du sein, même si la patiente ne présente pas réellement la maladie. D'autre part, une mammographie faussement négative peut sembler normale, mais en réalité, elle peut contenir des cellules cancéreuses sous-jacentes. En fait, la mammographie de dépistage peut manquer une tumeur sur six.

2.4 Diagnostic assisté par ordinateur (CAD) du cancer du sein

Des systèmes de diagnostic assisté par ordinateur (CAD) ont été développés pour aider les radiologues à trouver des lésions suspectes à la mammographie [3]. La détection assistée par ordinateur (CAD) est approuvée par la Food and Drug Administration (FDA) des États-Unis FDA pour une utilisation en mammographie depuis 1998 [7]. Son objectif est d'aider le radiologue à identifier les signes de cancer tels que les microcalcifications et l'augmentation de la densité. Cependant, il existe un taux élevé de faux positifs dans les CAD traditionnels, ce qui entraîne une anxiété accrue des patients, une exposition inutile aux radiations et des biopsies, ainsi qu'une augmentation des coûts. Compte tenu de cela, de nouveaux développements au sein de l'intelligence artificielle (IA) ont pris de l'importance au cours de la dernière décennie. Les réseaux de neurones artificiels tels que les réseaux de neurones convolutifs (CNN) ont réussi à aider à de nombreuses tâches.

Les systèmes de CAD traditionnels sont basés sur l'intervention humaine. Ils sont entraînés à détecter les microcalcifications en recherchant des pixels de haute intensité en forme de bâtonnet. Ils détectent les masses en observant une forme, une texture, un dégradé et un niveau de gris spécifiques. Les systèmes de CAD sont divisés en détection assistée par ordinateur et systèmes de diagnostic. Les systèmes de détection identifient les anomalies suspectes, mais laissent les décisions concernant la prise en charge des patients au radiologue. Les systèmes de diagnostic estiment la probabilité de maladie sur la base des caractéristiques de l'anomalie. Avec les systèmes de détection et de diagnostic, les CAD traditionnels ont atteint une performance similaire à celle des lecteurs humains pour la détection de masse. Cependant, comme ils dépendent de l'interprétation par un radiologue de nombreuses invites affichées, ils entraînent une spécificité plus faible et des temps de lecture plus longs.

Il a été démontré que l'IA ajoute une valeur significative aux systèmes de CAD traditionnels. Plus précisément, plutôt que de se fier uniquement aux caractéristiques d'entrée pour les algorithmes, les programmes basés sur CNN utilisent des réseaux de neurones pour apprendre des données et créer différentes représentations des données. De plus, au lieu de souligner les résultats uniquement dans certaines parties de la mammographie identifiées par le radiologue, les systèmes CNN utilisent des modèles hybrides où d'autres résultats potentiellement significatifs sur le plan clinique, tels que des calcifications suspectes, des asymétries focales, des asymétries ou des masses, sont également identifiés [7]. L'une des tâches les plus importantes de l'IA

en mammographie est d'améliorer les performances diagnostiques. Ceci est souvent mesuré en termes d'aire sous la courbe (AUC), de sensibilité et de spécificité. L'AUC tient compte à la fois de la sensibilité et de la spécificité tout en générant un nombre compris entre 0 et 1. Une valeur plus élevée équivaut à une meilleure précision diagnostique dans la détection de la maladie. En mammographie numérique, l'IA s'est révélée prometteuse pour améliorer les performances diagnostiques. Par rapport au CAD conventionnel, le CAD assisté par CNN a montré une réduction des faux positifs, une sensibilité améliorée et une augmentation de près de 7 pourcent de l'AUC [7]. Étant donné que les systèmes CNN-CAD ont des taux de faux positifs inférieurs à ceux des systèmes de CAD traditionnels, ils peuvent être utilisés pour réduire les taux de rappel. Dans certains cas, le taux de rappel a été réduit de 10–20 pourcent en raison d'une plus grande capacité à distinguer les images bénignes des images malignes. Cela peut réduire l'anxiété et la détresse du patient associées au rappel.

Dans certaines études, le logiciel autonome d'apprentissage en profondeur a surpassé de manière statistiquement significative les lecteurs ou AUC des programmes d'IA qui ont été obtenues dans ces études, variaient de 0,876 à 0,940, par rapport aux AUC des lecteurs, qui variaient de 0,778 à 0,81.

Nous présentons sur la figure 2.4, un exemple d'un système CAD.

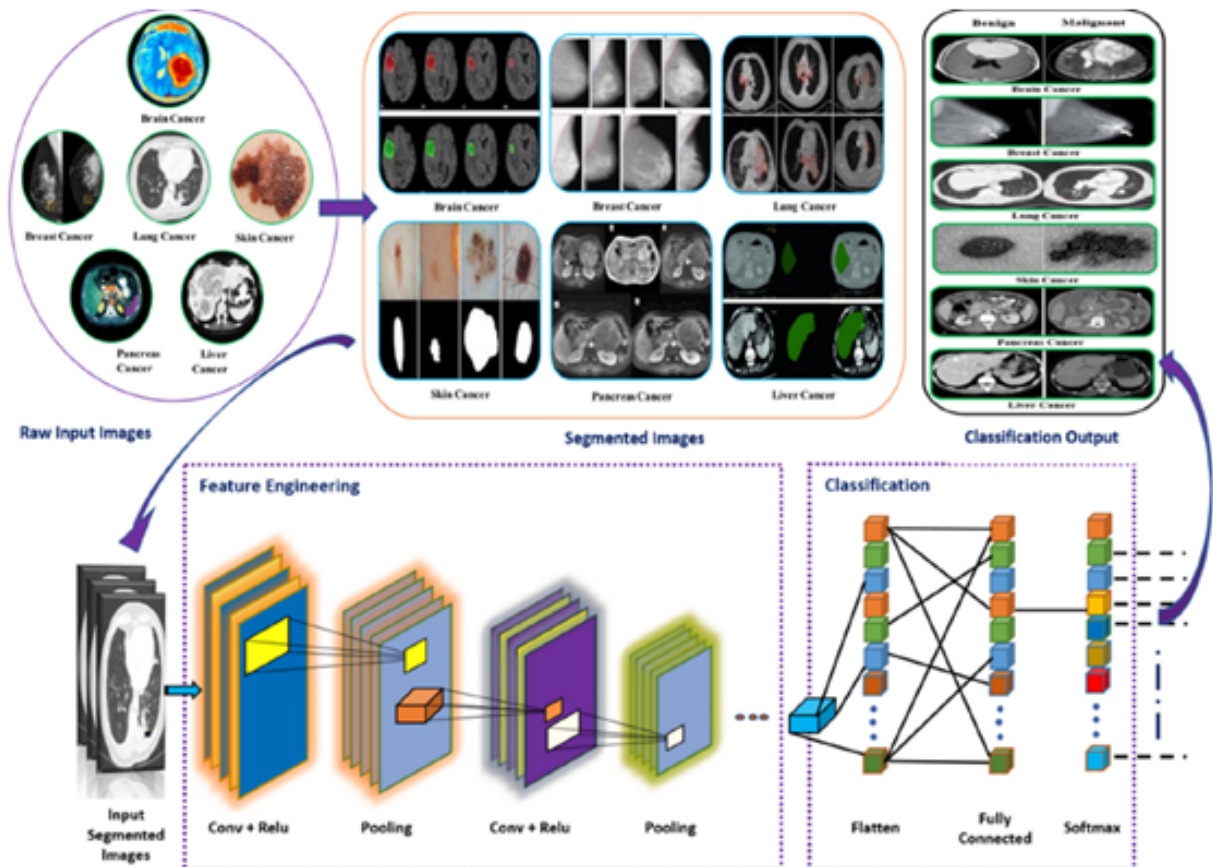


FIGURE 2.4 – Système CAD de segmentation et classification des tumeurs mammaires.

2.4.1 Détection des tumeurs assistée par ordinateur (CADE)

Cette étape clé de détection de tumeurs assistée par ordinateur (CADE) consiste en la segmentation tumorale dans l'IRM qui désigne une condition préalable à l'analyse radiomique quantitative car elle facilite le calcul de certaines mesures quantitatives de biomarqueurs d'imagerie [47]. Malheureusement, la segmentation manuelle traditionnelle par les radiologues nécessite une expertise importante, du temps et est sujette à l'erreur. Par conséquent, de nombreuses méthodes basées sur l'apprentissage ont été proposées pour la segmentation et la détection automatiques des tumeurs mammaires dans les images.

2.4.2 Diagnostic des tumeurs assistée par ordinateur (CADE)

La phase de diagnostic assisté par ordinateur (CADx) consiste en l'extraction de descripteurs discriminants sur les tumeurs segmentées. En effet, ces descripteurs seront des variables prédictives d'entrée pour un classifieur, et un modèle prédictif est bâti par l'ajustement des poids relatifs aux descripteurs se basant sur les propriétés statistiques d'éléments d'appren-

tissage pour l'estimation de la probabilité qu'une tumeur segmentée appartienne à une classe pathologique.

2.5 L'apprentissage automatique pour le pronostic automatique du cancer du sein

L'apprentissage automatique est un terme inventé par Arthur Samuel en 1959 qui l'a décrit comme un sous-domaine de l'IA. C'est une technique qui reconnaît des modèles à partir d'entrées données, apprend ces modèles sans être explicitement programmé et résout les problèmes basés sur des entrées [6]. En cas d'imagerie médicale, les caractéristiques d'image extraites manuellement (définies en termes de caractéristiques d'image distinctives telles que les formes, la région d'intérêt, la texture et l'histogramme des pixels d'image à partir des images médicales d'entrée) servent d'entrée aux algorithmes de l'apprentissage automatique. Les caractéristiques extraites peuvent ensuite être traitées par des algorithmes de sélection de caractéristiques pour sélectionner les caractéristiques les plus discriminants et enfin, la tâche des algorithmes ML est de fusionner les caractéristiques d'entrée sélectionnées en une seule valeur telle qu'une signature tumorale qui pourrait représenter une probabilité d'un état pathologique. Nous allons évoquer sur le tableau 2.1 les diverses approches CAD basées apprentissage automatique, de diagnostic de cancer du sein tout en mentionnant les techniques d'extraction de descripteurs et les classifieurs ainsi que l'évaluation des résultats obtenus.

TABLE 2.1 – Approches CAD basées apprentissage automatique de diagnostic de cancer du sein

Référence	Base de données utilisée	Contribution	Evaluation des performances
[48]	MIAS	Une approche de diagnostic automatique extrayant les descripteurs par les techniques : NSST/GLCM/Cropping où une classification des tumeurs basée : SVM/NB/MLP/KNN/LDA	Acc = 59.1%
[49]	Une base locale	Une approche extrayant les descripteurs : GLCM/PCA/LDA/MDS à partir des tumeurs où une classification basée DCNN a été réalisée	Acc = 82.43%
[50]	DDSM	Les descripteurs de la GLCM ont été extraits et donnés comme entrée à un classifieur BP-NN	Acc = 97%
[51]	DDSM	Les descripteurs issus de : FCMRG/LBP-GLCM/LPQ/mRMR/ Cropping ont été extraits et donnés comme entrée à un de classifieurs : DT/KNN.	Acc = 94.4 %
[52]	MIAS	Classification des tumeurs mammaires par l'extraction des descripteurs : Hough Transform/Canny Edge Detection/ Maximization Estimation où une classification basée SVM/LDA a été réalisée.	Acc = 94%/86%

[53]	DDSM	Classification des tumeurs mammaires comme bénignes ou malignes basée SVM.	Acc = 83.53%
[54]	DDSM	Approche de classification des tumeurs mammaires basée Fuzzy Gaussian Mixture Model (FGMM).	Acc = 93%
[55]	DDSM	Distinction entre les tumeurs malignes et bénignes basée Extreme Learning Machine (ELM-ANN).	Acc=95.73%
[56]	MIAS	Classification de la densité des tumeurs mammaires basée SVM.	Acc=91.51%
[57]	DDSM	Classification des tumeurs mammaires basée SVM.	Acc=94.44%

2.6 L'apprentissage profond pour le pronostic automatique du cancer du sein

Pour remédier aux limites des algorithmes de l'apprentissage automatique, des recherches récentes se sont orientées vers des approches basées sur l'apprentissage profond [6]. Les algorithmes DL traitent les données brutes et ne nécessitent aucune extraction manuelle et explicite des caractéristiques des données d'entrée pour produire une sortie/prédiction. Ces algorithmes ont la capacité de s'adapter à toutes sortes de caractéristiques des données d'entrée (image, texte, audio) et fournissent donc des résultats plus robustes dans les problèmes de segmentation et de classification par rapport aux méthodes ML conventionnelles. Les algorithmes d'apprentissage profond sont issus du réseau de neurones artificiels (ANN), qui est un sous-ensemble d'algorithmes d'apprentissage automatique.

Nous allons évoquer sur le tableau 2.2 les diverses approches CAD basées apprentissage profond, de diagnostic de cancer du sein tout en mentionnant les modèles utilisés ainsi que l'évaluation des résultats obtenus.

TABLE 2.2 – Approches CAD basées apprentissage profond de diagnostic de cancer du sein

Référence	Base de données utilisée	Contribution	Evaluation des performances
[58]	MIAS	Une approche de diagnostic automatique du cancer du sein par : CNN : 5 Conv. Layers +2 Fully Connected Layers + Softmax Classifier	Acc = 90.5 %
[59]	DDSM	Une approche de diagnostic automatique du cancer du sein basée : DCNN : (AlexNet[5 Conv. Layers +3 Pooling Layers + 2 Fully Connected Layers] + SVM Classifier).	Acc = 87.2%
[60]	DDSM	Une approche de classification des tumeurs mammaires basée : InceptionV3, DenseNet121, ResNet50, VGG16 and MobileNetV2.	Acc = 98.8 %
[61]	DDSM	Une approche de classification des tumeurs segmentées basée : DCNN : 5 Conv. Layers +3 Fully Connected Layers + SVM Classifier.	Acc = 96.7 %
[62]	DDSM + BCDR +	Classification des tumeurs mammaires basée InceptionV3, ResNet50, VGG16.	Acc=94%/86%
[63]	DDSM	Classification des tumeurs mammaires selon 4 classes : Irregular, Lobular, Oval, Round basée CNN : 3 Conv. Layers +2 Fully Connected Layers + Softmax Classifier.	Acc = 80 %

2.7 Conclusion

Les domaines de l'apprentissage automatique et profond, qui émergent de l'intelligence artificielle font ravage et connaissent une croissance remarquable, notamment dans l'analyse de l'imagerie médicale et le diagnostic du cancer du sein. Ce domaine exige un grand flux de données, qui a été rendu possible grâce aux diverses techniques de sauvegarde et de stockage d'imageries médicales, en vue d'une éventuelle mise en œuvre de modèles de prédiction de nouveaux cas à partir de ces données.

Dans le prochain chapitre, nous exposerons la problématique liée à ce domaine d'actualité, ainsi que le cadre méthodologique suivi pour mettre en œuvre l'approche proposée.

CHAPITRE 3

PROBLÉMATIQUE ET APPROCHE
PROPOSÉE

3.1 Introduction

Tout au long de ce chapitre, nous détaillerons la problématique de ce travail désignant l'implémentation de modèles en vue du diagnostic automatique du cancer du sein, se basant sur les paradigmes de l'intelligence artificielle, à savoir, l'apprentissage automatique et l'apprentissage profond.

3.2 Problématique et plan d'action

Le diagnostic précoce du cancer du sein est très important pour faire recours à cette dangereuse maladie. En effet, l'émergence de techniques basées sur l'apprentissage automatique et l'apprentissage profond a enrichi le domaine du diagnostic du cancer avec son efficacité débordante dans le calcul de problèmes complexes avec un taux d'erreur inférieur à celui des humains. Une révolution mesurable a été observée dans le développement du système assisté par ordinateur basé sur ces deux paradigmes pour la détection du cancer au cours de la dernière décennie. Cependant, il existe encore divers défis associés au diagnostic :

1. Le dépistage par mammographie est l'un des outils d'imagerie médicale les plus efficaces pour la détection et le diagnostic précoces du cancer du sein, et il peut réduire les taux de cancer du sein avancé et mortel à ses premiers stades. Cependant, il a été prouvé qu'environ 10 à 30 pourcent des cas de cancer passent à côté de la mammographie de dépistage, ce qui génère un taux de faux négatifs jusqu'à 50 pourcent selon le type de lésions et la densité mammaire [3]. Avec l'augmentation des suivis et des dépistages ultérieurs pendant la période de diagnostic, il a été démontré qu'environ 50 pourcent des mammographies antérieures présentaient des lésions visibles rétrospectivement. Par conséquent, cela a amené les radiologues à se demander si des mammographies antérieures normales sans signes clairs de tout type de lésions pouvaient en réalité contenir des informations cachées indiquant un risque futur d'apparition de tumeurs, spécialement pour le type de sein dense où le tissu mammaire normal est souvent diagnostiqué à tort comme des lésions anormales ou vice versa. Cela induit en erreur des faux positifs et des faux négatifs et, dans certains cas, des examens supplémentaires inutiles [2].

2. Grâce à l'adoption généralisée du système de dossiers de santé électroniques, la taille des données d'imagerie radiologique ne cesse de croître et, par conséquent, il est maintenant au-delà de la capacité des radiologues d'analyser manuellement une telle quantité de données d'imagerie [6]. Des études révèlent que, dans certains cas, un radiologue doit analyser en moyenne une image radiologique toutes les 3 à 4 s au cours d'une journée de travail de 8 h pour répondre aux exigences de la charge de travail. La croissance continue des données d'imagerie médicale à un rythme disproportionné aggrave encore cette exigence de charge de travail. Étant donné que l'analyse des données d'imagerie radiologique implique une perception visuelle élevée et des capacités cognitives robustes des radiologues, les décisions des radiologues liées à l'imagerie médicale restent toujours incertaines. L'incertitude dans les décisions augmente encore lorsque les radiologues souffrent de conditions de charge de travail ingérables et contraintes.
3. L'inspection manuelle des images mammaires pour une éventuelle détection du cancer du sein est une méthode largement utilisée, cependant, certains faits incontournables liés à l'inspection manuelle des images peuvent conduire à une détection inexacte et prolonger le processus de diagnostic [6]. Par exemple : l'indisponibilité des experts dans les pays sous-développés, l'indisponibilité d'experts ayant une connaissance suffisante du domaine pour analyser avec précision des images multi-classes (images avec d'éventuelles caractéristiques d'affections multiples), l'inspection quotidienne d'un grand nombre d'images médicales peut être une pratique exhaustive et fastidieuse, la nature subtile de la tumeur mammaire et la structure complexe des tissus mammaires rendent l'analyse manuelle plus difficile, le niveau de concentration des experts médicaux et d'autres fatigues rendent le diagnostic plus difficile et un processus qui prend du temps.
4. Outre l'indisponibilité d'ensembles de données de grande taille sur les maladies nécessaires au développement de modèles basés sur l'IA, la disponibilité d'un nombre limité d'ensembles de données entièrement étiquetés est considérée comme un défi majeur dans la tâche de diagnostic du cancer. Les chercheurs avaient utilisé leurs méthodes basées sur ML/DL soit sur des ensembles de données où seul un nombre limité d'échantillons étaient correctement étiquetés, soit ils dépendaient de la décision d'un expert pour la validation des étiquettes de données, où le chercheur n'avait pas de données étiquetées en main. Dans de tels cas, les méthodes ML/DL semi-supervisées peuvent aider le chercheur jusqu'à un

certain point.

5. Les méthodes basées sur ML/DL nécessitent une grande quantité d'échantillons de données pour une généralisation correcte du modèle de classification. Cependant, dans le domaine du diagnostic du cancer, la disponibilité et l'accès à un ensemble de données aussi volumineux peuvent ne pas être réalisables la plupart du temps en raison de problèmes de confidentialité des patients ayant participé à l'ensemble de données ou en raison des politiques de partage de données des organisations détenant cet ensemble de données. L'intervention du gouvernement ou de l'organisme de réglementation dans les politiques de partage de données de cette organisation peut explorer davantage d'alternatives pour une meilleure amélioration de l'efficacité du modèle prédictif en employant des dispositions appropriées pour la distribution sécurisée et préservant la confidentialité d'ensembles de données de grande taille entre les chercheurs et ces organisations[4].
6. Le déséquilibre des données (data imbalance) est également une préoccupation majeure au moment du diagnostic du cancer à l'aide des méthodes ML/DL car il est très difficile d'obtenir des ensembles de données équilibrés dans le domaine médical. Un ensemble de données équilibré fait référence à un ensemble de données contenant une part presque égale d'instances de chaque classe, c'est-à-dire la classe de patients et la classe saine. En règle générale, ce type d'ensemble de données peut ne pas être obtenu pour la raison que ; les institutions impliquées dans les tests ou le dépistage des patients ne prennent pas la peine de conserver les données des personnes en bonne santé. Des données déséquilibrées, en tant qu'entrées du modèle ML/DL, peuvent entraîner l'émergence d'un biais dans la prédiction du modèle utilisé.
7. Les réseaux CNN ont des limites et des inconvénients fondamentaux. L'orientation des différents composants d'une image et les relations spatiales relatives entre ces composants ne sont pas prises en compte dans un réseau basé sur CNN [2]. En fait, la couche convolutive est la couche principale d'un réseau CNN. Elle permet de détecter et d'extraire des caractéristiques simples des images de couches plus profondes et de les combiner avec des caractéristiques plus complexes de couches supérieures. Ces caractéristiques sont ensuite introduites dans des couches denses pour effectuer des prédictions de classification. Le pooling maximal est une solution qui a traversé les différentes couches du réseau CNN

- et pour détecter les caractéristiques complexes des couches convolutives supérieures dans de grandes régions des images. En effet, max pooling réalise de très bonnes performances dans de nombreux domaines et présente étonnamment une bonne solution ; cependant, comme il diminue la taille spatiale des données, il perd également des informations précieuses à partir des données.
8. En particulier les algorithmes DL sont de nature "boîte noire" [6], c'est-à-dire qu'il n'y a pas de justification ou d'explication appropriée des décisions/prédictions faites par ces algorithmes. Dans le diagnostic médical comme le diagnostic du cancer du sein, les décisions de la boîte noire (comme les recommandations/décisions prises par les systèmes automatisés basés sur l'IA) ne sont généralement pas préférées car les radiologues/médecins sont principalement intéressés à savoir et à comprendre comment la décision a été prise et sur quels facteurs elle a été prise. En raison d'une explicabilité minimale et de plusieurs autres facteurs tels que la perte de contrôle sur la prise de décision autonome, les systèmes automatisés basés sur l'IA pour la prédiction/le diagnostic des maladies (systèmes d'aide à la décision) sont souvent considérés comme une menace ou une perte de contrôle par les médecins.
 9. Un autre défi dans le diagnostic du cancer à l'aide des méthodes ML/DL réside dans les différentes variations disponibles dans les données d'entrée. Ces variations peuvent inclure différentes modalités de données présentes dans des ensembles de données comme les tomodensitogrammes et les IRM enregistrés pour un type de cancer spécifique, les mêmes scans du patient capturés par différents organismes institutionnels ou laboratoires ayant différents matériels et logiciels, et des variations de données perceptibles en raison de différentes conditions d'éclairage utilisées lors de la capture de données. En dehors de cela, les variations dues aux différentes données utilisées au moment de la phase de développement et de déploiement du modèle restent un défi majeur dans le diagnostic du cancer basé sur ML/DL. Les chercheurs développent généralement leur modèle à l'aide d'ensembles de données standard librement disponibles à des fins de recherche et de bien social, qui sont généralement dans des formats traités ou standard requis par les algorithmes ML/DL.

10. La présence de bruit, la variation de la taille de la lésion, la forme irrégulière ou les limites floues de la lésion cancéreuse sont d'autres sujets de préoccupation importantes lors du diagnostic du cancer à l'aide de la modalité d'imagerie médicale. Comme ces faits peuvent augmenter la complexité de l'image, la rendant difficile à interpréter et difficile à localiser l'emplacement de la lésion, entraînant ainsi une probabilité plus élevée de mauvaise classification du modèle ML/DL développé. Par conséquent, une plus grande attention doit être portée à la phase de prétraitement du développement du modèle prédictif pour faire face à ce type de problèmes présents dans la modalité d'image.

Compte tenu des points problématiques cités ci-dessus, nous proposerons un plan d'actions concernant le recours à ces points comme suit :

- (a) Par rapport à l'inspection manuelle, l'analyse d'image automatisée basée sur l'IA évite les processus de dépistage fastidieux et chronophages et capture efficacement des informations précieuses et pertinentes à partir des données d'image massives. En effet, l'utilisation de l'IA pour l'analyse de l'imagerie médicale réduit l'incertitude et les erreurs dans la prise de décision en offrant aux radiologues la possibilité d'images présélectionnées avec des caractéristiques identifiées. Toutes ces contraintes humaines et les progrès des données d'imagerie et des ressources informatiques ont motivé les prestataires de soins de santé à utiliser la technologie de l'IA dans le domaine de l'imagerie médicale [6]. En effet, L'IA pour les applications médicales utilise les données d'imagerie médicale (radiologie) pour fournir de meilleurs services de santé en temps opportun. L'objectif principal de l'utilisation de l'IA dans l'imagerie médicale est d'atteindre une plus grande efficacité et efficacité dans les pratiques cliniques de routine et d'obtenir un soutien dans le processus de prise de décision.
- (b) La technologie de détection assistée par ordinateur (CAD) a été introduite pour améliorer la précision de l'interprétation de la mammographie. Typiquement, les CAD sont développés pour localiser les régions suspectes des lésions qui existent dans les mammographies dépistées. Nous proposons une approche CAD basée sur l'extraction de caractéristiques d'image par l'analyse de la texture des tumeurs en fusionnant deux approches, l'une régissant dans le domaine spatial, qui consiste en les lois de puissance Zipf et Zipf inverse et l'autre à la fois dans le domaine spatial et fréquentiel consistant en les filtres de Gabor pour bénéficier de leur apport complé-

mentaire durant le processus de classification basé SVM pour distinguer les tumeurs malignes de celles bénignes.

- (c) Sur la base d'une analyse effectuée sur 108 articles inclus des six dernières années (2016-2021), il a été constaté qu'environ deux tiers des articles utilisaient une technique basée sur l'apprentissage automatique lors de la détection du cancer et un tiers des articles incorporaient des méthodes basées sur l'apprentissage en profondeur pour la même chose [4]. Il montre la prédominance des techniques ML sur les méthodes DL dans la tâche de détection du cancer. Cela pourrait être dû à la nature de la boîte noire de la méthode DL en raison de leur nature opaque car l'interprétation des états internes du modèle n'est pas aussi simple qu'avec le code orienté objet standard ou les arbres de décision ; en effet, nous distinguons une capacité explicative plus élevée des méthodes ML, car la capacité explicative des méthodes appliquées est plus préférable en cas de diagnostic médical. A cet effet, nous allons utiliser l'apprentissage profond durant le processus de segmentation en réalisant un « transfert Learning » ainsi que l'apprentissage automatique durant le processus de classification des tumeurs segmentées. En effet, Les modèles d'apprentissage automatique ont été largement utilisés pour la classification des lésions dans les mammographies numériques car ils donnent de bons résultats dans la classification de petits ensembles de données. La plupart des travaux récents utilisent encore l'apprentissage automatique en améliorant les méthodes existantes, en proposant de nouvelles combinaisons de caractéristiques ou de classifieurs, ou en améliorant les images d'entrée au préalable dans la phase de prétraitement pour augmenter les performances de classification [2].
- (d) La méthode apprentissage automatique basée sur SVM et les méthodes apprentissage profond basées sur CNN se sont avérées être fréquemment utilisées ainsi que significativement efficaces que les autres méthodes basées sur ML/DL utilisées pour la tâche de diagnostic du cancer [4]. Dans ce sens, la segmentation par transfert appliqué dans ce travail se basera sur le modèle U-net que nous avons entraîné sur une base de données des IRM cérébrales du fait que les IRM présentent un grand contraste des tissus mous par rapport à la mammographie ; par la suite une classification des tumeurs en malignes et bénignes sera basée sur les SVM. En effet, l'utilisation de l'apprentissage par transfert était dû aux faits que les tumeurs mammaires se refou-

laient toujours dans les tissus denses du sein et un apprentissage sur de tels types de mammographie nous a conduit à des résultats de segmentation erronés.

3.3 Segmentation des tumeurs mammaires basée Transfer Learning

3.3.1 La base des mammographies DDSM

Nous avons choisi d'utiliser la base de données d'images mammaires DDSM, collectée par l'équipe d'experts de l'Université de Floride du Sud. Cette base de données contient 2620 cas d'images du sein de patientes, classés en 43 volumes différents. Chaque image a une taille moyenne de 3000 x 4800 pixels, une profondeur de 16 bits et une résolution spatiale de 42 microns. Chaque mammographie est accompagnée d'informations déterminées par des experts, notamment le type de cancer (bénin ou malin) et la localisation de la lésion. Il est également intéressant de noter que chaque mammographie est associée à un masque binaire qui fournit une indication générale de la position des anomalies. Cependant, ces masques ne fournissent pas de segmentation précise, ce qui nécessite la mise en œuvre d'algorithmes de segmentation pour une extraction et un diagnostic précis des tumeurs. Nous exposons sur la figure 3.1 quelques mammographies brutes avec leurs masques binaires respectifs.

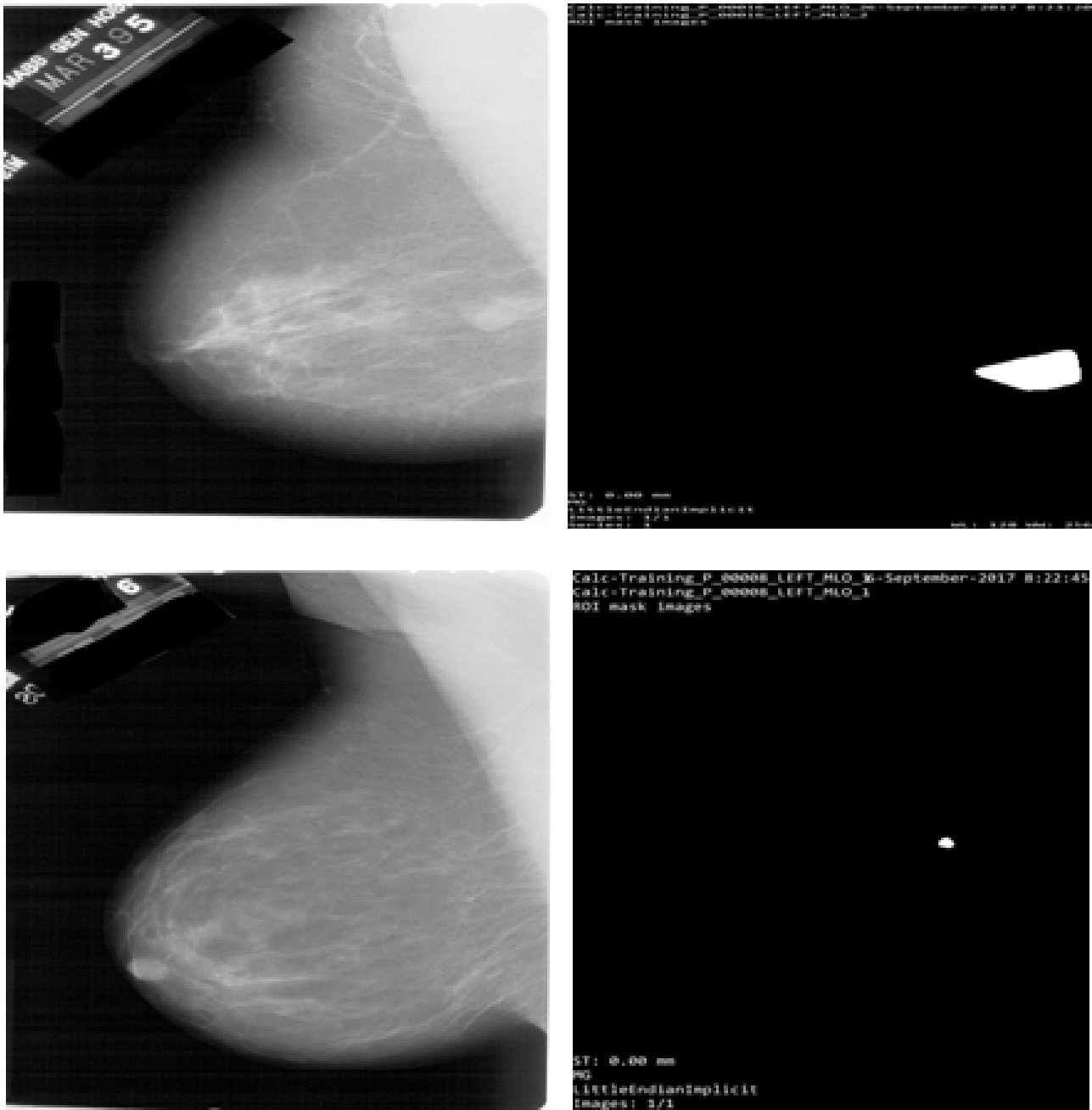


FIGURE 3.1 – Mammographies appartenant à la base de donnée DDSM avec leurs masques respectifs.

3.3.2 Pipeline de la segmentation des tumeurs mammaire basé Transfer Learning

- **Prétraitement**

Les images médicales ont un faible contraste ainsi qu'une luminosité inhomogène engendrant un défi : entre le ciblage et les organes nous distinguons une frontière floue et ambiguë se caractérisant par un contraste faible. Dans ce sens, un prétraitement jouera un rôle obligatoire pour une pertinente détection des tumeurs mammaires suite à l'élimination du bruit au sein des mammographies.

Le processus d'amélioration d'image est utilisé sur des images construites pour affiner et transformer des images afin d'améliorer l'interprétabilité des entités physiques présentes dans l'image. Il est généralement obtenu par différentes méthodes d'optimisation du contraste, de réduction du bruit, d'amélioration des contours, d'élimination des artefacts et de lissage de l'image.

Diverses approches de prétraitement ont été appliquées dans la littérature, citons le filtre gaussien les ondelettes, l'égalisation adaptative d'histogramme (CLAHE), les filtres médians et beaucoup d'autres. En effet, la CLAHE est une technique puissante pour l'amélioration du contraste que nous appliquerons en premier, suivie du filtre médian vu son rôle incontournable de la préserve des contours. Nous présentons sur la figure 3.2 quelques mammographies avant et après avoir appliqué le pré-traitement mentionné.

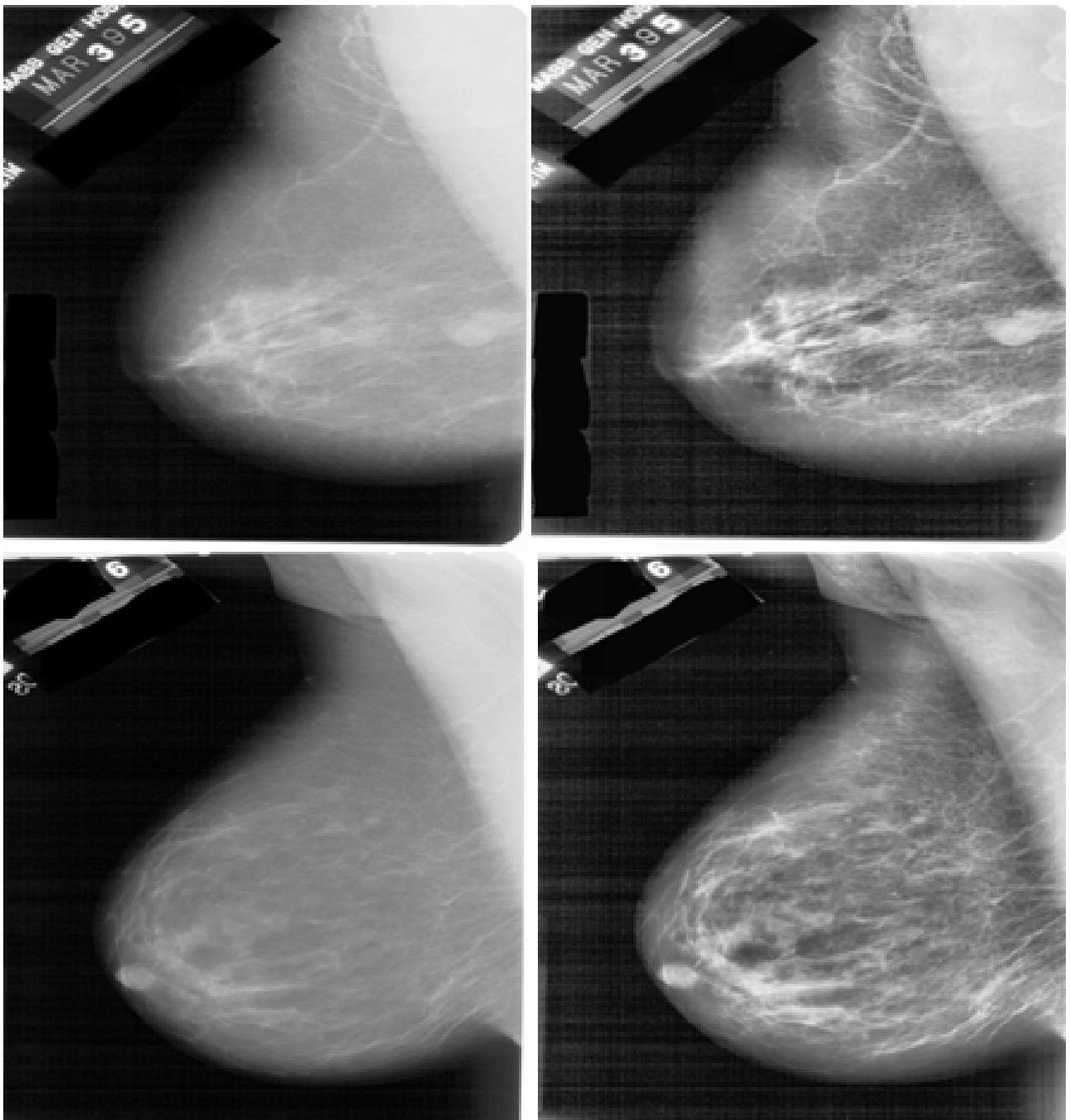


FIGURE 3.2 – Résultats du pré-traitement appliqué sur les mammographies.

- **Segmentation des tumeurs mammaires basée Transfer Learning**

Les approches basées CNN ont montré des résultats robustes pour les applications impliquant une entrée visuelle, c'est-à-dire une image. Afin de former l'architecture CNN, il est possible de réaliser l'apprentissage par transfert [6].

L'apprentissage par transfert est devenu une technique largement utilisée qui tire parti des connaissances acquises tout en résolvant un problème et en l'appliquant à un problème différent mais connexe. En effet, le recours à ce type d'apprentissage était du fait que la mammographie est connue par la densité de ses tissus où nous distinguons des tumeurs refoulées dedans où le modèle U-net appliqué sur la base des mammographies DDSM nous a généré des résultats erronés de segmentation.

En contrepartie, l'IRM cérébrale présente un grand contraste des tissus mous. Nous avons entraîné le modèle U-net sur la base des IRM cérébrale Figshar contenant 3064 IRM avec 3 types différents de tumeurs : 708 méningiomes, 1426 gliomes et 930 tumeurs hypophysaires. En effet, nous avons pris le modèle U-net pré-entraîné sur la segmentation des tumeurs cérébrales et nous l'avons adapté à notre problème de segmentation des tumeurs mammaires en apprenant uniquement les poids de la dernière couche du réseau de neurones [6].

3.4 Classification des tumeurs mammaires basée machines à vecteurs de support (SVM) :

3.4.1 Pipeline basé SVM pour la classification des tumeurs mammaire

La figure 3.3 expose le pipeline suivi pour la classification des tumeurs mammaires. Il est utile de mentionner que le prétraitement a été appliqué durant la segmentation.

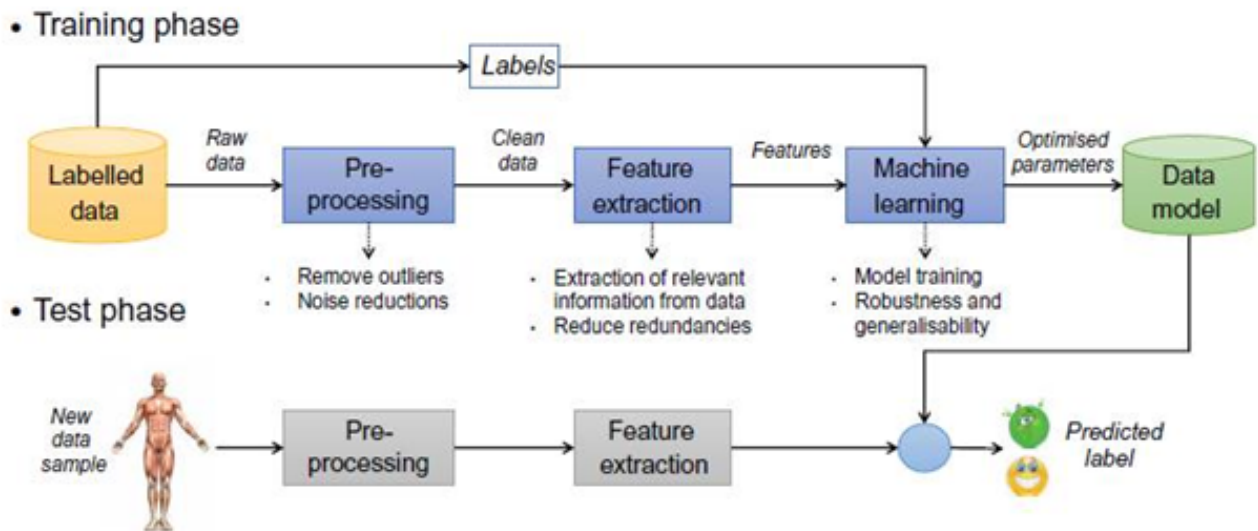


FIGURE 3.3 – Pipeline suivi pour la classification des tumeurs mammaires.

• **Extraction et sélection des descripteurs**

L'extraction et la sélection efficaces des caractéristiques constituent l'épine dorsale du développement de modèles d'apprentissage automatique / apprentissage profond optimisés, car elles minimisent les erreurs de classification par les méthodes basées sur ces approches, améliorant ainsi la précision de la classification. L'extraction et la sélection de descripteurs font référence à l'identification ainsi qu'à l'extraction des caractéristiques les plus discriminantes, utilisables et hautement corrélées à partir des données brutes disponibles. Bien que les chercheurs aient utilisé diverses méthodes pour extraire différentes caractéristiques statistiques, de texture, morphologiques et profondes à partir de données médicales dans un passé récent, mais la texture demeure le descripteur le plus discriminant pour l'analyse de l'image médicale [4]. En effet, en ce qui concerne l'utilisation des techniques d'extraction de caractéristiques dans le domaine du diagnostic médical, il existe un très grand nombre de méthodes pouvant être utilisées pendant le processus de détection du cancer. En raison du grand nombre de techniques remarquées lors d'une revue de la littérature [4], ils ont été regroupées en plusieurs catégories générales sur la base de leur tâche et analysé leur pourcentage d'utilisation dans différentes tâches de détection du cancer où la plupart des études incluses dans cette étude intégraient des méthodes d'extraction de descripteurs liées à la normalisation des caractéristiques, à l'optimisation, à la sélection, à la détection des régions d'intérêt, à la segmentation et à la détection de la texture et seules quelques-unes d'entre elles utilisaient les méthodes liées à l'étiquetage et à la détection des contours de tumeurs pendant la phase d'extraction des caractéristiques du processus de détection et de classification du cancer.

Les informations d'imagerie, telles que la texture parenchymateuse mammaire s'est révélés utiles

pour la prédiction du risque de cancer du sein [5]. En dehors de cela, au coût plus élevé de la complexité de calcul, la combinaison de plusieurs types de caractéristiques peut être utilisée pour obtenir une précision de classification optimisée par le modèle d'apprentissage automatique ou profond. Dans ce sens, nous avons appliqué les lois de Zipf et de Zipf inverse consistant en des approches perceptuelles analysant le contenu structurel des mammographies pour la caractérisation de la texture. D'autre part, les filtres de Gabor simulent les performances des cellules corticales visuelles du mammifère par l'extraction de différents traits selon diverses orientations échelles [65] en fonctionnant de manière identique que celle des neurones du cortex visuel humain.

— **Analyse et caractérisation de la texture des tumeurs mammaires par les lois de Zipf et de Zipf inverse :**

Nous analysons les tumeurs segmentées par les lois de puissance pour en extraire les courbes de Zipf et de Zipf inverse comme indiqué sur les figures 3.4 et 3.5 respectivement où les courbes de Zipf et de Zipf inverse ont été générées des tumeurs mammaires bénignes et malignes.

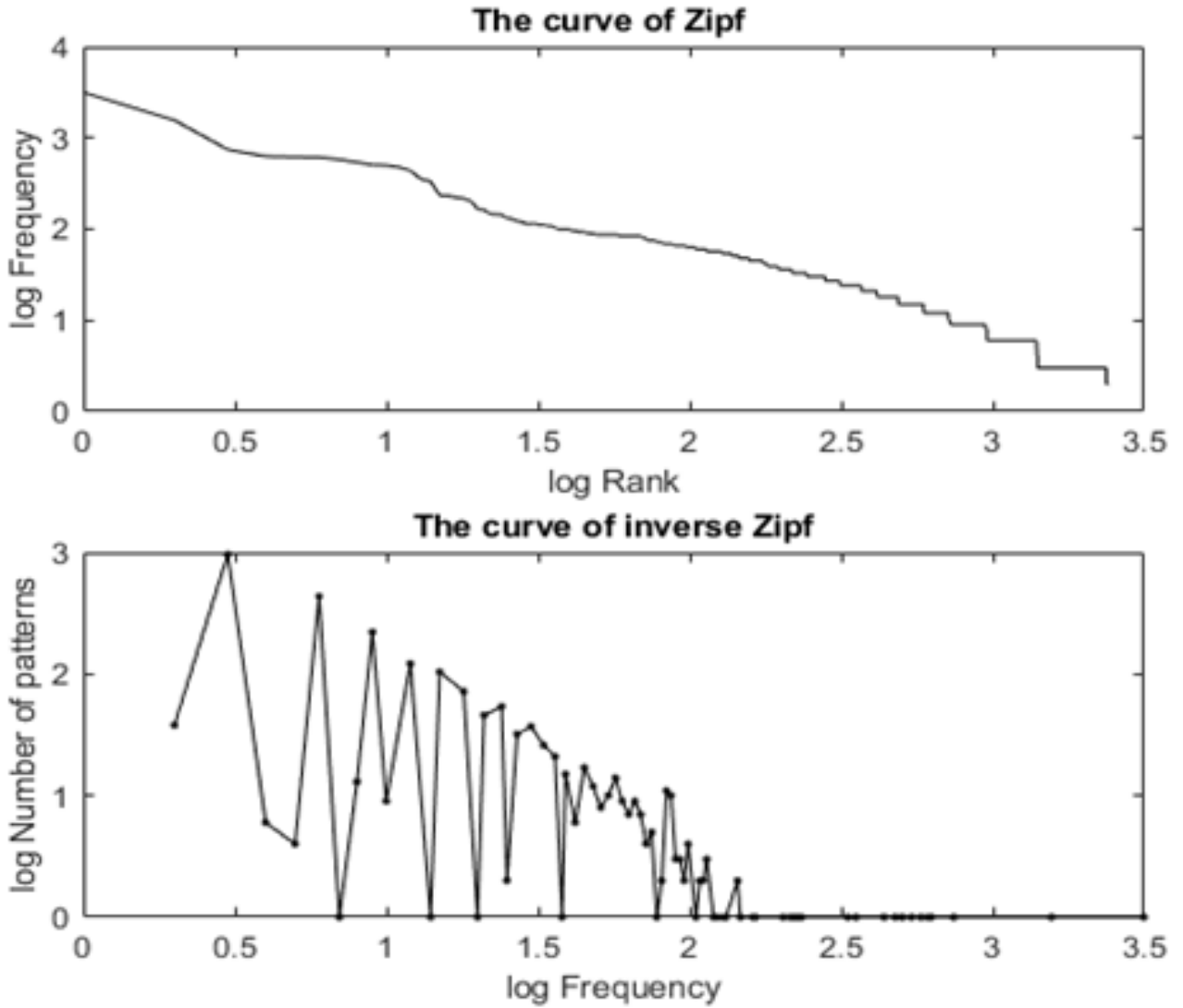


FIGURE 3.4 – Courbes de Zipf et de Zipf d’une tumeur mammaire bénigne.

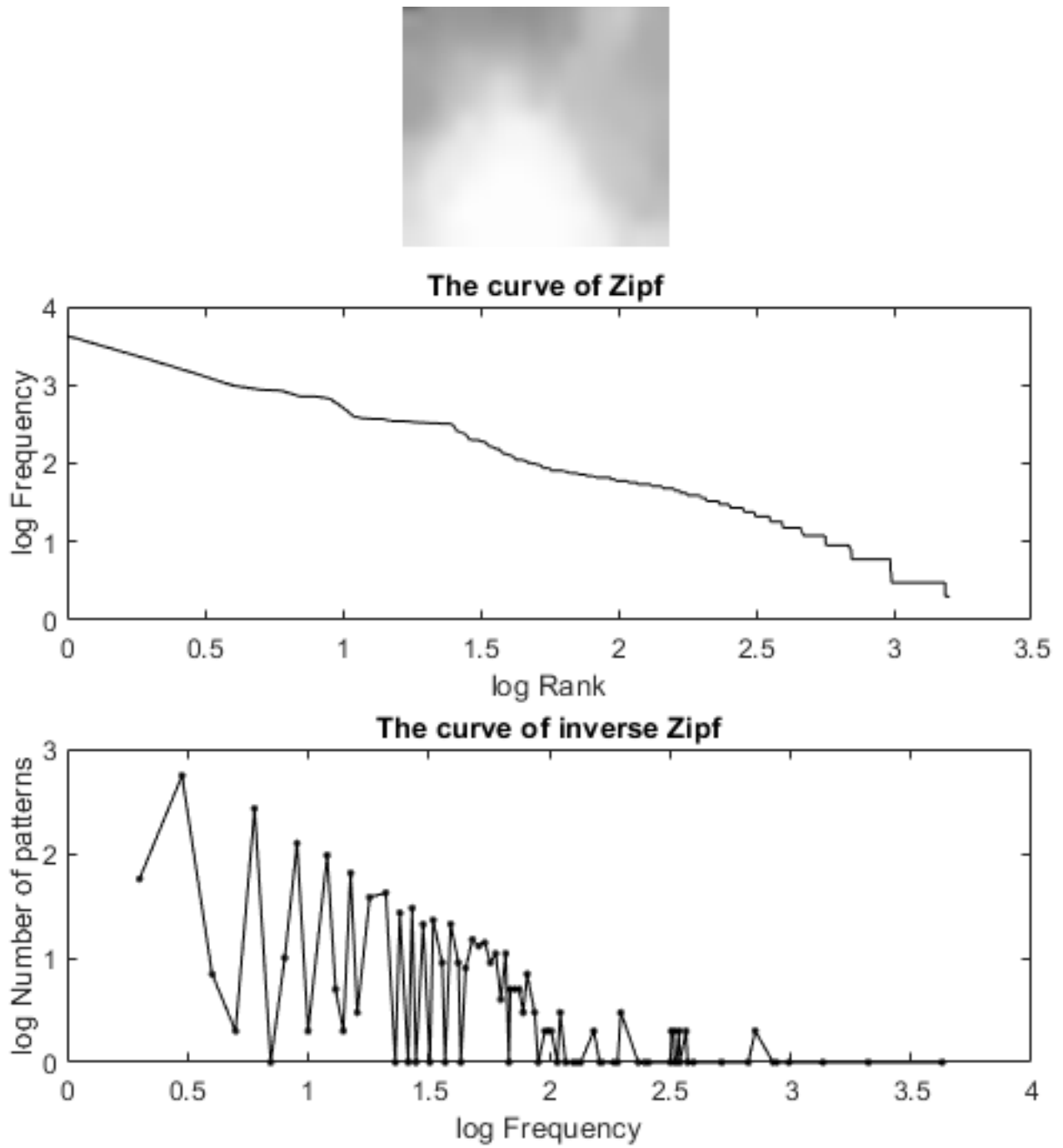


FIGURE 3.5 – Courbe de Zipf et Zipf inverse d’une tumeur mammaire maligne.

Les descripteurs dérivés des courbes de Zipf et de Zipf inverses sont les suivants [39] :

- **Les pentes des courbes de Zipf et de Zipf inverse**

La pente moyenne d'une courbe est le coefficient directeur de la droite des moindres carrés. Elle est donnée par la formule 3.1 :

$$p = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (3.1)$$

- **L'aire délimitée par la courbe de Zipf :**

Nous calculons l'aire délimitée par la courbe de Zipf à partir des courbes de Zipf obtenues suite au codage de l'image par le codage des rangs généraux. Soit n le nombre de motifs de la courbe, f_i la fréquence et r_i le rang du motif i , l'aire de la courbe est donnée par la formule 3.2 :

$$A = \sum_{i=1}^{n-1} \frac{(f_i + f_{i+1})(r_{i+1} - r_i)}{2} \quad (3.2)$$

- **Entropie1 de la courbe de Zipf :**

L'entropie relative aux motifs des images mammographies est définie par la formule 3.3 :

$$H_w = - \sum_{r=1}^R \frac{f(r)}{T} \log_R \left(\frac{f(r)}{T} \right) \quad (3.3)$$

Dans cette formule, $f(r)$ représente la fréquence du motif pour la ligne r , T représente le nombre total de motifs différents, et nous utilisons un logarithme avec la base R .

- **Entropie2 de la courbe de Zipf**

L'entropie relative à la fréquence d'apparition des motifs est définie par la formule suivante :

$$H_f = - \sum_{f=1}^F \frac{I(f)}{R} \log_F \left(\frac{I(f)}{R} \right) \quad (3.4)$$

Dans cette formule, $I(f)$ représente le nombre de motifs distincts ayant une fréquence d'apparition égale à f et F représente le nombre entier d'occurrences des motifs dans l'image.

- **Les ordonnées à l'origine des courbes de Zipf et Zipf inverse**

- **La constante alpha de la courbe de Zipf**

La loi de Zipf est fortement exprimée de la façon suivante : Quel que soit un motif appartenant à une image, la fréquence d'apparition de ce motif * son rang dans une liste ordonnée décroissante des fréquences d'apparition des motifs = constante.

— **Analyse et quantification de la texture des tumeurs mammaires basée filtres de Gabor :**

Les filtres de Gabor agissent dans les deux domaines spatial et fréquentiel en réalisant une décomposition multi-résolution.

Dans le cas des images 2D, le filtre de Gabor est généré suite à une gaussienne bidimensionnelle modulée par une fonction sinusoidale complexe [66]. La formule du filtre de Gabor 2D est la suivante [66] :

$$h(x, y) = \frac{1}{\sigma_x \sigma_y \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \exp(j2\pi u_0(x \cos \theta + y \sin \theta)) \quad (3.5)$$

L'enveloppe Gaussienne :

$$g(x, y) = \frac{1}{\sigma_x \sigma_y \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \quad (3.6)$$

Avec :

$$X = x \cos O + y \sin O$$

$$Y = -x \sin O + y \cos O$$

Où x et y sont les écarts-types respectivement le long des axes x et y (ou les constants spatiales du filtre), elles déterminent la largeur du filtre.

U_0 : la fréquence centrale.

O : l'angle de rotation de $[x, y]$ par rapport à (x, y) , il donne l'orientation de l'enveloppe Gaussienne $g(x, y)$.

Les filtres de Gabor désignent un type de filtrage passe-bande qui conservent l'information prise sur le spectre d'une image via la bande de fréquence sélectionnée. Effectivement, ces filtres analysent des propriétés locales des tumeurs. Par conséquent, la configuration fréquemment utilisée d'extraction de données spatio-fréquentielles à partir des textures est générée suite à la convolution de la région d'intérêt avec une banque de filtres [66] où chacun se retrouvera centré sur une fréquence et une orientation pour couvrir l'intégralité du domaine fréquentiel avec succès. Dans ce sens, chaque pixel de la tumeur générera une réponse à chaque filtre.

Concernant la classification, nous analysons la relation existante entre les réponses aboutie par les filtres de Gabor de chaque tumeur.

Nous mentionnons sur la figure suivantes, l'analyse d'une tumeur par les filtres de Gabor où les sorties seront utilisées pour l'extraction des descripteurs.

Les paramètres pertinents du filtre de Gabor sont la fréquence radiale ainsi que l'orientation. Ils définissent la localisation du canal dans le plan fréquentiel. En effet, chaque image de taille $N \times N$ aura des fréquences significatives dans l'intervalle [4] et à la puissance 2.

Nous avons appliqué une banque des filtres de Gabor avec une taille de 39×39 selon 5 fréquences et 8 orientations comme mentionné sur la figure 3.6 :

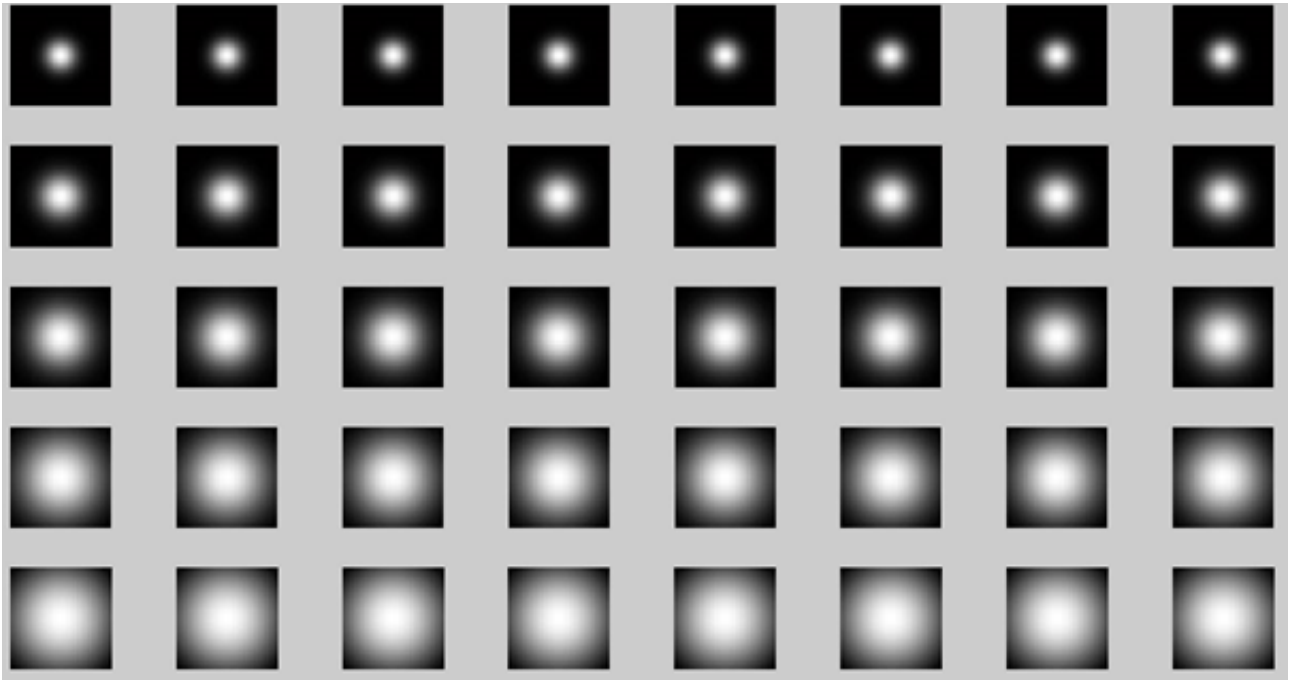


FIGURE 3.6 – Application des filtres de Gabor dans le domaine fréquentiel.

Suite à l'utilisation d'un ensemble riche en variation de filtres de Gabor, ceci permettra une couverture plus étendue de l'espace fréquentiel en détectant davantage d'orientations, et aboutissons donc à l'extraction de tous les contours des régions d'intérêt. La figure 3.7, démontre la partie réelle de ces filtres.

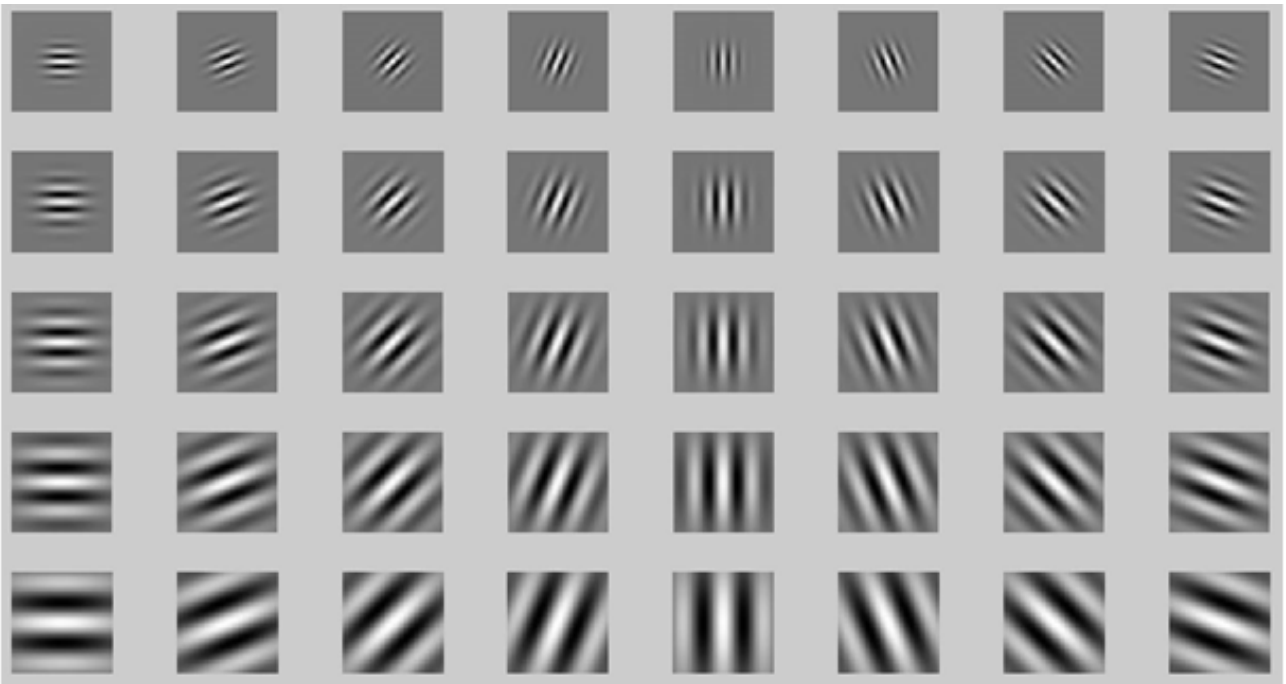


FIGURE 3.7 – Parties réelles des filtres de Gabor utilisés.

Nous allons démontrer sur la figure 3.8, la partie réelle des régions d'intérêt filtrées :

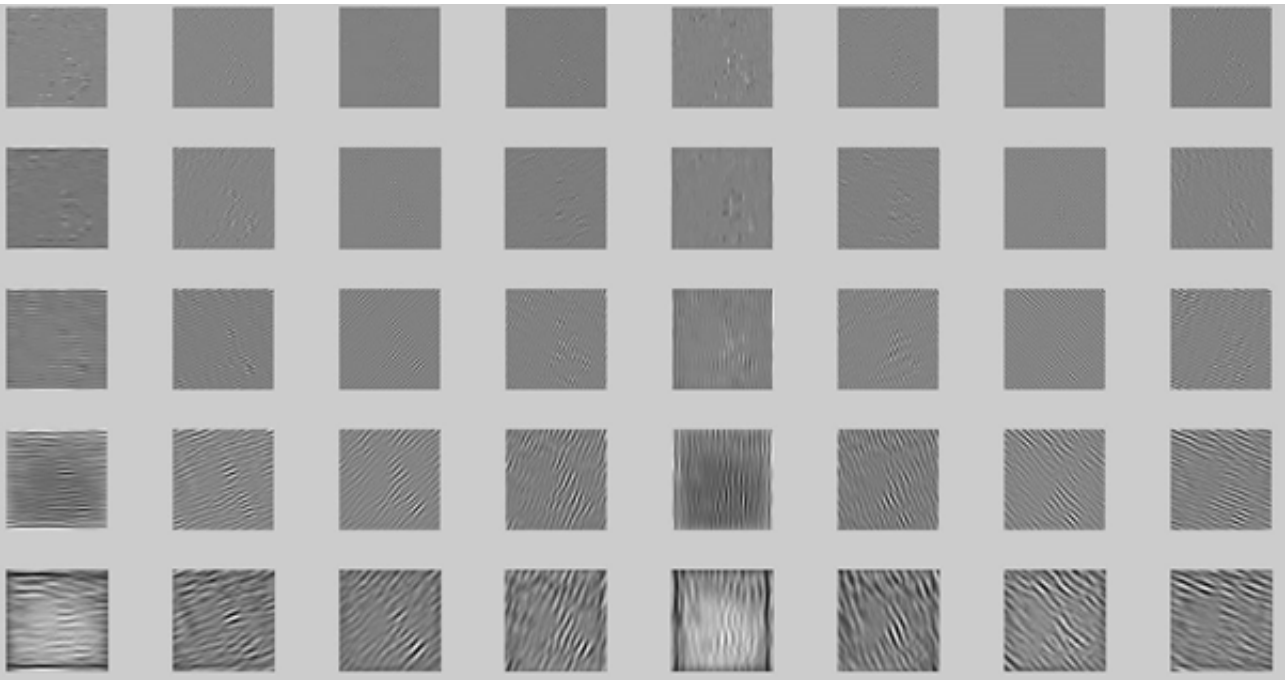


FIGURE 3.8 – Parties réelles des tumeurs filtrées.

La figure 3.9 démontre la magnitude de la réponse des tumeurs filtrées.

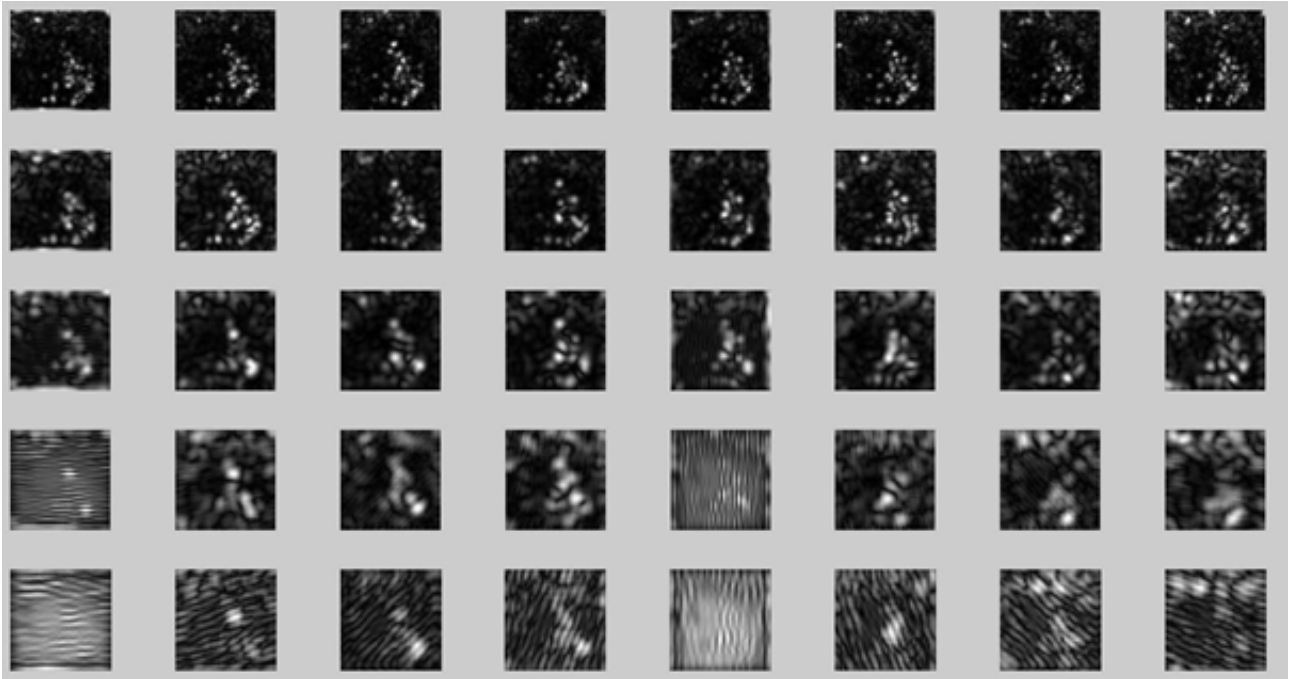


FIGURE 3.9 – Exposition de la magnitude de la réponse relative au filtre de Gabor une fois avoir réalisé la convolution de la tumeur maligne avec une banque de filtres de Gabor. Par la suite, nous prenons toutes les sorties du filtre et nous calculons des statistiques locales pour générer le vecteur des descripteurs texturaux comme proposé par les auteurs dans [67].

Une fois le vecteur des descripteurs est obtenu, nous en sélectionnons les descripteurs les plus discriminants par l'ACP qui désigne une technique de réduction de dimensionnalité populaire dans le domaine de l'apprentissage automatique et généralement utilisée pour réduire la dimensionnalité d'un ensemble de données contenant un grand nombre de caractéristiques ou de variables [67]. en effet, la réduction de la dimensionnalité aide à rendre les algorithmes d'apprentissage automatique moins complexes, ce qui leur permet de surmonter les problèmes de l'overfitting et de les maintenir avec une généralisabilité et des performances améliorées. L'objectif principal du processus d'extraction d'entités basé sur l'ACP est de créer un nouveau groupe d'entités indépendantes en utilisant des entités autonomes ou une combinaison de différentes entités existantes présentes dans l'ensemble de données tout en ne conservant que les entités les plus importantes pour prédire l'entité cible, préservant ainsi une représentation significative de la base de données initiale même après avoir transformé les entités d'origine en composants de principe indépendants. Le processus ACP commence par la normalisation de la gamme de caractéristiques initiales, suivie du calcul de la matrice de covariance pour déployer la corrélation entre les caractéristiques. L'étape suivante consiste à calculer et à organiser le

vecteur propre et les valeurs propres respectives de la matrice de covariance dans l'ordre décroissant, où le vecteur propre et les valeurs propres respectives appréhendent respectivement la composante principale et la mesure de la variance.

La dernière étape consiste à sélectionner des paires propres avec des valeurs propres maximales à partir du vecteur de caractéristiques généré lors de l'étape précédente, représentant le premier composant principal qui préserve le plus d'informations de l'ensemble de données initial. ACP améliore les performances des algorithmes d'apprentissage automatique en éliminant les descripteurs non contributifs de l'espace des caractéristiques.

• Classification des tumeurs mammaires basée SVM

La méthode d'apprentissage automatique basée sur SVM et les méthodes d'apprentissage profond basées sur CNN se sont avérées être fréquemment utilisées ainsi que significativement efficaces que les autres méthodes durant le diagnostic du cancer [4].

SVM est l'algorithme de classification basé sur l'apprentissage automatique le plus utilisé en raison de sa capacité à produire une précision de prédiction significative avec une puissance de calcul réduite [4]. Trouver un hyperplan optimal dans un espace de caractéristiques à N dimensions avec la plus grande marge possible autour de la limite de décision afin de classer les points de données distinctement, est l'objectif clé de l'algorithme SVM.

L'hyperplan généré par l'algorithme SVM est une frontière de décision, qui aide à la classification des points de données entre différentes classes. Les points de données de chaque côté de l'hyperplan peuvent appartenir à différentes classes. La dimension de l'hyperplan dépend du nombre d'entités prises en compte pour la classification des points de données, c'est-à-dire que si seuls deux nombres d'entités sont pris en compte, l'hyperplan dessiné est uniquement une ligne et si trois nombres d'entités sont pris en compte, l'hyperplan sera un plan bidimensionnel. Les raisons de la popularité de SVM résident dans ses avantages, c'est-à-dire son efficacité de traitement des données structurées et non structurées, son efficacité dans les espaces de grande dimension, son efficacité de la mémoire, son astuce du noyau pour gérer les problèmes complexes et le faible risque de l'overfitting.

3.5 Architecture générale de l'approche proposée

Nous réaliserons une segmentation des tumeurs mammaires par le Transfer learning qui sera suivie d'une classification basée SVM des tumeurs en : tumeur maligne et tumeur bénigne. L'architecture de l'approche proposée est présentée sur la figure 3.10.

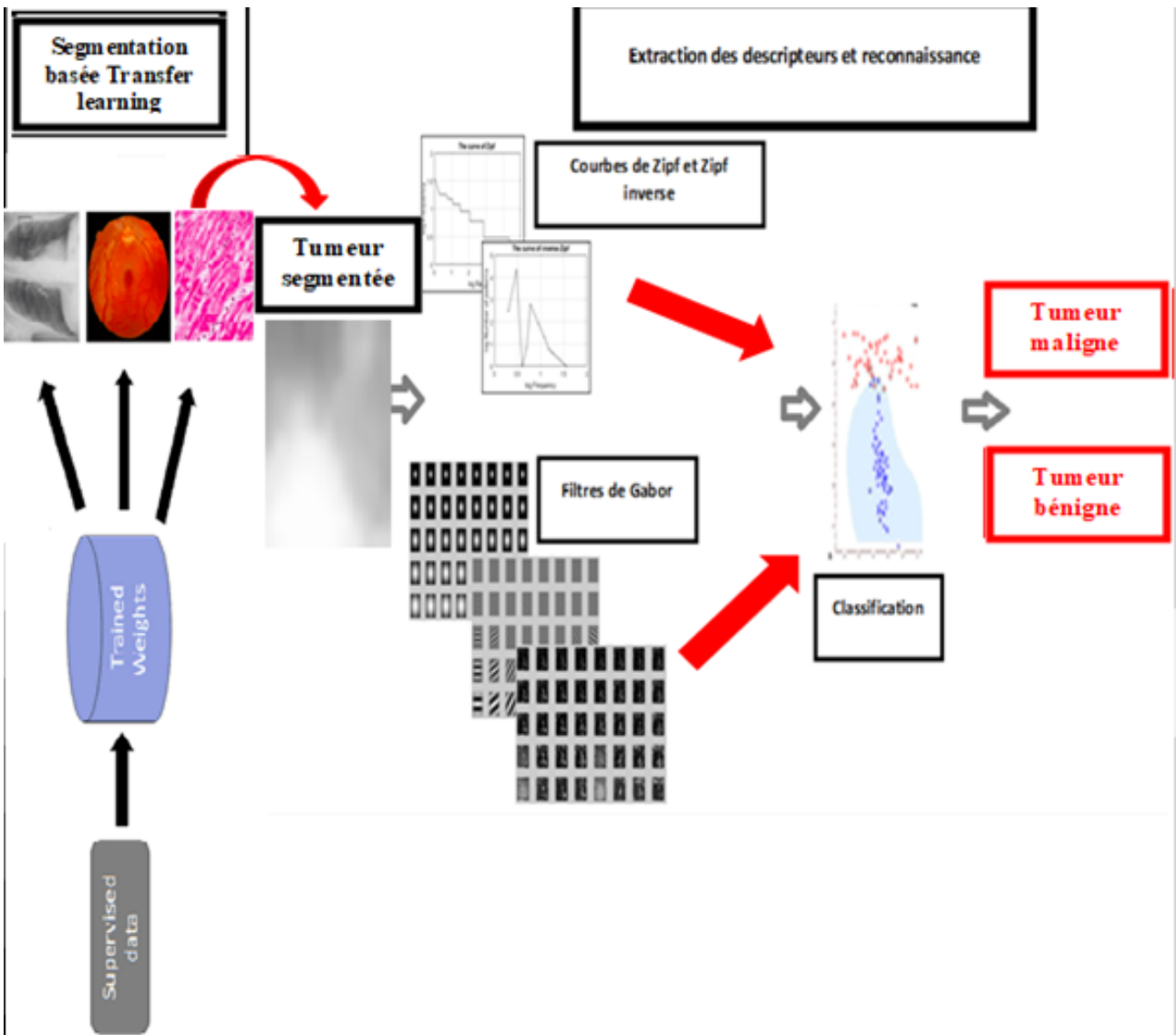


FIGURE 3.10 – Architecture de l’approche proposée de diagnostic automatique du cancer du sein.

3.6 Conclusion

Nous avons présenté une étude détaillée de la conception de l’approche proposée dans ce travail, désignant le diagnostic automatique du cancer du sein où les pipelines suivis pour les processus de segmentation et de classification des tumeurs ont été présentés.

Evidemment, l’implémentation et l’évaluation de l’approches proposée seront détaillées au long du dernier chapitre.

CHAPITRE 4

EVALUATION DES PERFORMANCES
OBTENUES

4.1 Introduction

Dans ce chapitre, nous allons présenter les expérimentations effectuées sur notre modèle afin de démontrer son efficacité à détecter le cancer du sein. Nous détaillerons d'abord l'environnement de développement et outils utilisés (Python, Matlab. . . .) ainsi que la description des mesures de performances utilisées (les formules), puis nous allons présenter l'évaluation des résultats obtenus. Finalement, nous exposerons quelque fonctionnalités de l'approche proposée.

4.2 Evaluation de l'approche proposée de pronostic automatique de cancer de sein

4.2.1 Environnement de développement

Nous avons mis en œuvre notre méthode sur un ordinateur Dell fonctionnant sous le système d'exploitation Windows 10. L'ordinateur est équipé d'un processeur Intel(R) Core(TM) i5-4200U cadencé à 1,6 GHz. Il dispose de 16 Go de mémoire vive (RAM) et fonctionne sur une version premium de Windows en 64 bits.

4.3 Outils utilisés

4.3.1 MATLAB (MATrix LABoratory)

MATLAB est un environnement de calcul scientifique puissant, complet et convivial. Il offre une large gamme de fonctionnalités pour les calculs matriciels, le développement et l'exécution d'algorithmes, la création d'interfaces utilisateur (UI) et la visualisation de données. Cet environnement de calcul numérique multi-paradigme permet aux développeurs d'interagir avec des programmes écrits dans différents langages, ce qui leur permet de tirer parti des avantages uniques de chaque langage à des fins variées.

MATLAB est largement utilisé par les ingénieurs et les scientifiques dans divers domaines tels que le traitement d'images et de signaux, les systèmes de contrôle industriels, la conception de réseaux intelligents, la robotique et la finance informatique. Nous avons utilisé MATLAB pour tracer des courbes de Zipf, Zipf inverse et des figures de Gabor.



FIGURE 4.1 – logo Matlab.

4.3.2 Python

Python est un langage de programmation interprété, orienté objet et de haut niveau qui possède une sémantique dynamique. Sa richesse en structures de données intégrées de haut niveau, combinée à son typage dynamique et à sa liaison dynamique, en font un choix extrêmement attractif pour le développement rapide d'applications et son utilisation en tant que langage de script pour connecter des composants entre eux. L'implémentation de fonctionnalités avancées est facile avec Python grâce à l'accès à des bibliothèques de haute qualité. Il est particulièrement omniprésent dans les domaines de l'intelligence artificielle et de l'apprentissage automatique. Des bibliothèques telles que Pandas, ImageNet, Pascal et MLxtend prennent en charge Python et sont disponibles sur plusieurs systèmes d'exploitation.

Il existe différentes versions de Python, notamment Python 2 et Python 3. Nous avons utilisé la dernière version de Python, qui est Python 3.9.13.



FIGURE 4.2 – logo python.

4.4 Logiciel

- **Kaggle**

Kaggle est une plateforme web qui propose une vaste sélection de jeux de données, de codes, de cours et de compétitions dans le domaine de la science des données.

Lorsque vous créez un compte sur Kaggle, vous obtenez :

- Lorsque vous créez un compte sur Kaggle, vous bénéficiez d'un espace de chargement illimité pour vos ensembles de données. De plus, la plateforme offre une vitesse de téléchargement rapide lorsque vous téléchargez des fichiers au format zip.
- La fonctionnalité de sauvegarde des versions sur Kaggle permet :
 1. Exécution du code sans connexion Internet et envoi d'une notification par e-mail lorsque celui-ci est terminé.
 2. save output.
- Exécuter le code en utilisant un GPU avec une limite de quota de 30 à 45 heures par semai.

Nous avons choisi Kaggle en raison de son avantage spécifique, à savoir l'exécution par lots, qui limite le temps nécessaire pour corriger les erreurs pouvant survenir lors de l'exécution.



FIGURE 4.3 – logo Kaggle.

- **TensorFlow**

TensorFlow est une bibliothèque open-source populaire pour l'apprentissage automatique et le calcul numérique. Elle a été développée par Google et est largement utilisée dans la communauté de l'intelligence artificielle.

- **Keras**

Keras est une bibliothèque open-source très populaire pour l'apprentissage en profondeur (deep learning). Elle est conçue pour être simple, conviviale et flexible, ce qui en fait un choix populaire parmi les chercheurs et les praticiens en apprentissage automatique.

- **Open CV(Open Computer Vision)**

OpenCV (Open Source Computer Vision Library) est une bibliothèque open-source très populaire et largement utilisée dans le domaine de la vision par ordinateur et du traitement d'images. Elle fournit des outils et des algorithmes pour traiter, analyser et comprendre les images et les flux vidéo en temps réel.

- **SK learn**

Scikit-learn, également connu sous le nom de sklearn, est une bibliothèque open-source très populaire pour l'apprentissage automatique (machine learning) en Python. Elle fournit un ensemble d'outils et d'algorithmes pour les tâches courantes de classification, régression, clustering, et prétraitement des données.

- **Matplotlib**

Matplotlib est une bibliothèque open-source très populaire en Python pour la visualisation de données et la création de graphiques. Elle offre un large éventail de fonctionnalités pour créer des graphiques statiques, des graphiques interactifs et des visualisations complexes.

- **pill**

PIL (Python Imaging Library), également connue sous le nom de Pillow, est une bibliothèque open-source populaire en Python pour la manipulation d'images. Elle offre un large éventail de fonctionnalités pour ouvrir, modifier, convertir et sauvegarder des images dans différents formats.

4.5 Résultats aboutis de la segmentation des tumeurs mammaires basée Transfer Learning

C'est une étape essentielle où le modèle apprend à partir des exemples fournis afin de faire des prédictions sur de nouvelles données.

Après cette période d'étude, nous n'avons pas obtenu de bons résultats, avec une perte $loss=0.7$ et une précision $acc=0.97$. Cela est dû au fait que les femmes âgées de 25 à 30 ans présentent des seins denses qui se manifestent par des tumeurs refoulées dans le tissu du sein. En effet,

cela mènera à des résultats de segmentation erronés du fait que la tumeur est moins visible. Nous présentons sur la figure 4.4 les courbe de la précision de la classification ainsi que de la minimisation de loss function.

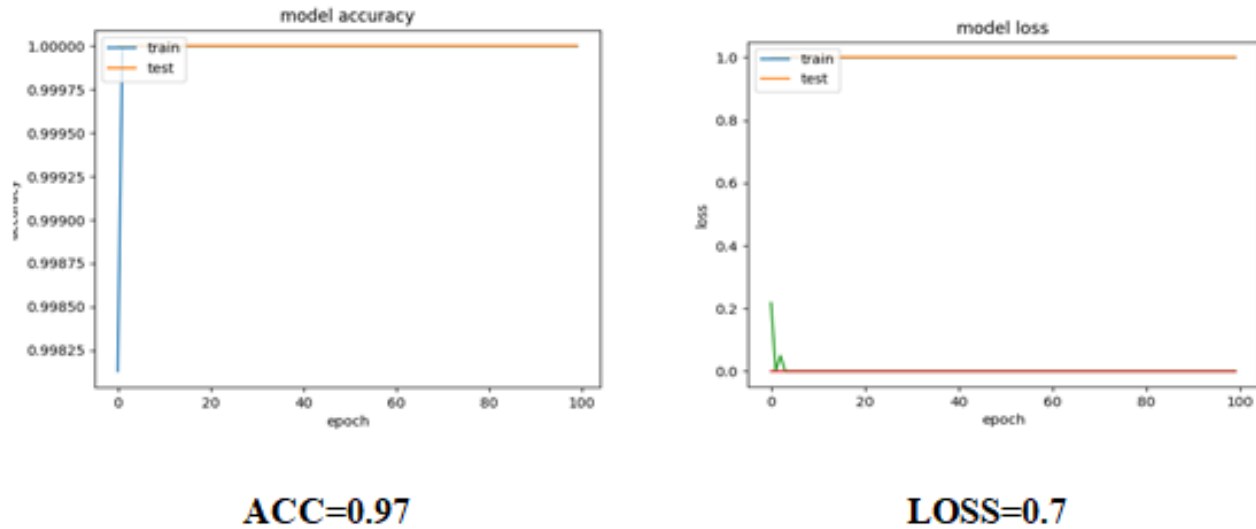


FIGURE 4.4 – Les courbes de la précision de classification ainsi que de la minimisation de loss function suite à l'apprentissage sur la base de données DDSM.

C'est ce qui nous a poussés à utiliser la méthode du transfert learning. En effet, l'IRM cérébrale IRM présentent un grand contraste des tissus mous par rapport à la mammographie. Nous avons obtenu des résultats bien meilleurs par rapport aux résultats précédents. En effet, nous avons entraîné le modèle U-net sur la base Figshare des IRM cérébrale contenant trois types différents de tumeurs. Nous présentons sur la figure 4.5 les courbes obtenues.

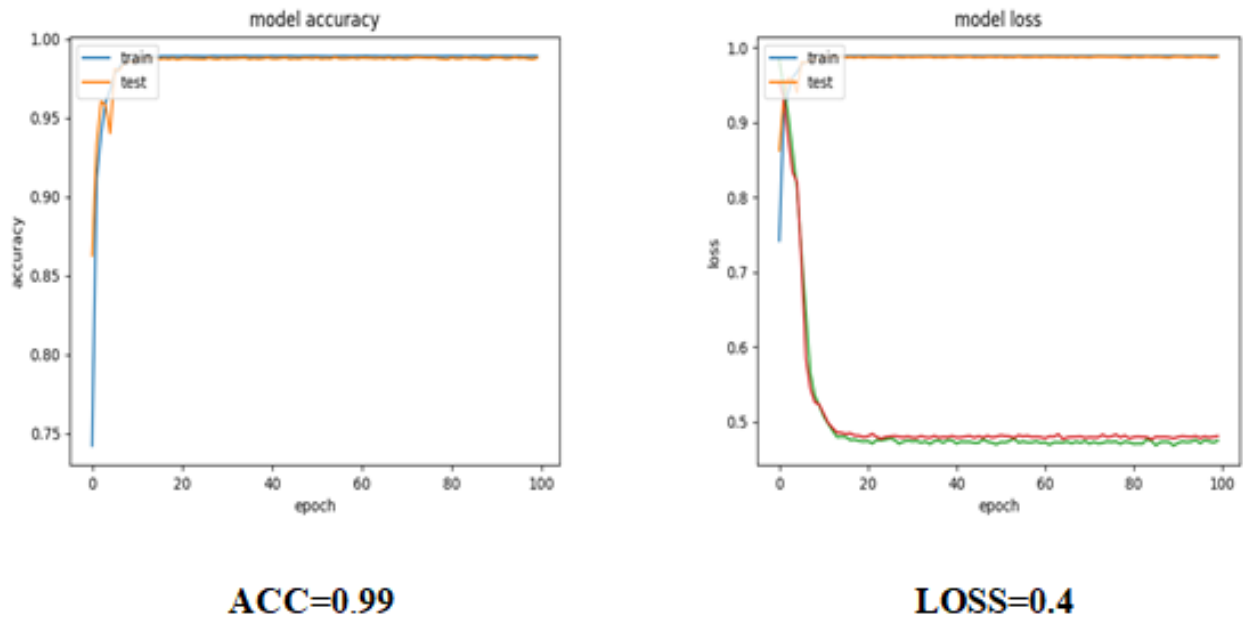


FIGURE 4.5 – Les courbes de la précision de classification ainsi que de la minimisation de loss function suite à l'apprentissage sur la base de données Figshare.

4.5.1 Cropping des tumeurs

Le "cropping" des tumeurs, également connu sous le nom de recadrage d'image, est une technique utilisée pour couper ou découper une partie spécifique d'une image, en éliminant les parties indésirables ou en se concentrant sur une région d'intérêt particulière.

Nous présentons sur la figure 4.6 le résultat du cropping.

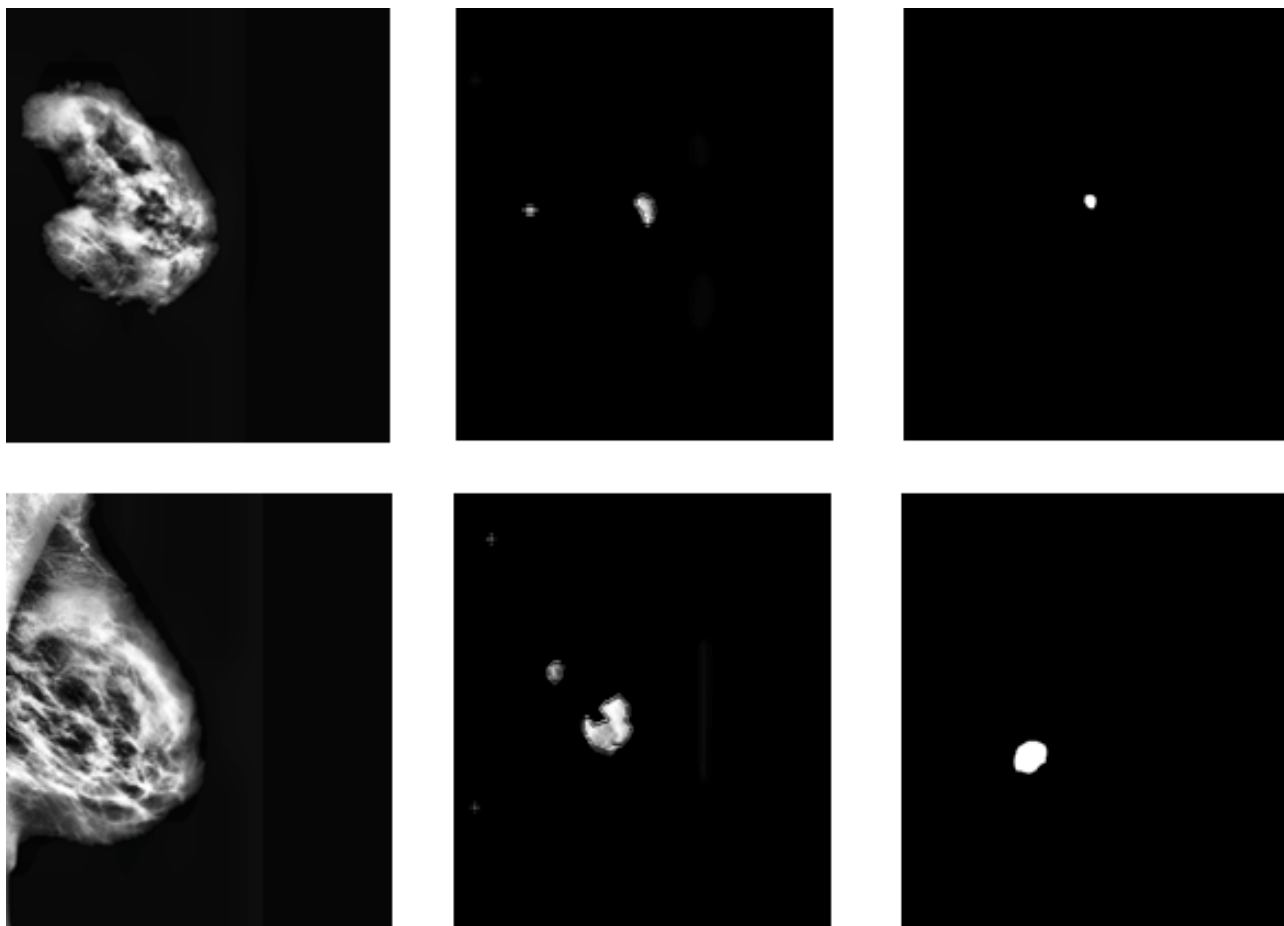


FIGURE 4.6 – Processus de cropping des tumeurs.

4.6 Evaluation des performances de l'approche proposées de pronostic automatique de cancer du sein par les machines à vecteur de support (SVM)

Dans le domaine de la santé numérique, lors de l'évaluation de la progression d'une maladie, les données sont généralement très limitées en nombre où nous disposons seulement d'un nombre restreint d'informations pronostiques concernant les patients. Dans ce sens, la validation croisée est considérée comme le remède le plus approprié.

Nous avons utilisé l'intégralité des régions d'intérêt segmentées à partir de la base de données DDSM lors du processus de classification, en distinguant les tumeurs bénignes des tumeurs malignes, afin de démontrer la capacité discriminatoire du vecteur descripteur entre les tissus pathologiques et ceux bénins.

Selon [68], l'« Accuracy » (exactitude) est la mesure d'évaluation la plus couramment utilisée pour évaluer un modèle. Elle représente la proportion du nombre total de prédictions qui étaient correctes.

Nous exposons sur le tableau 4.4 les résultats de la classification basée SVM obtenus. En effet, pour le premier modèle de pronostic automatique du cancer du sein ; nous avons obtenu une précision de classification dans les alentours de 49 pourcent. Pour améliorer les résultats, nous avons utilisé la méthode de l'ACP pour la sélection des descripteurs les plus discriminants et simplifier des ensembles de données complexes en réduisant leur dimensionnalité tout en conservant les informations essentielles. Cela peut faciliter l'analyse des données, la visualisation et améliorer les performances des modèles prédictifs. En effet, nous avons obtenu 99 pourcent de précision de la classification suite à l'application de l'ACP.

TABLE 4.1 – Evaluation des performances de l'approche proposées de pronostic automatique de cancer du sein

Modèle de classification (SVM)	Kernel	Précision (Accuracy)
Modèle1(sans ACP)	RBF	49.3 %
Modèle2(avec ACP)	RBF	99%

Les données d'apprentissage et de test sont généralement divisées selon un rapport de division standard de 80 :20 [69]. Ensuite, une métrique d'évaluation est utilisée pour évaluer l'exactitude du modèle en se basant sur les étiquettes réelles des données de test. Les paramètres d'évaluation populaires pour la classification sont basés sur la matrice de confusion, qui permet d'analyser les performances du modèle en termes de vrais positifs, faux positifs, vrais négatifs et faux négatifs [69]. Nous enrichissons l'évaluation de l'approche proposée par la matrice de confusion présentée sur la figure 4.7.

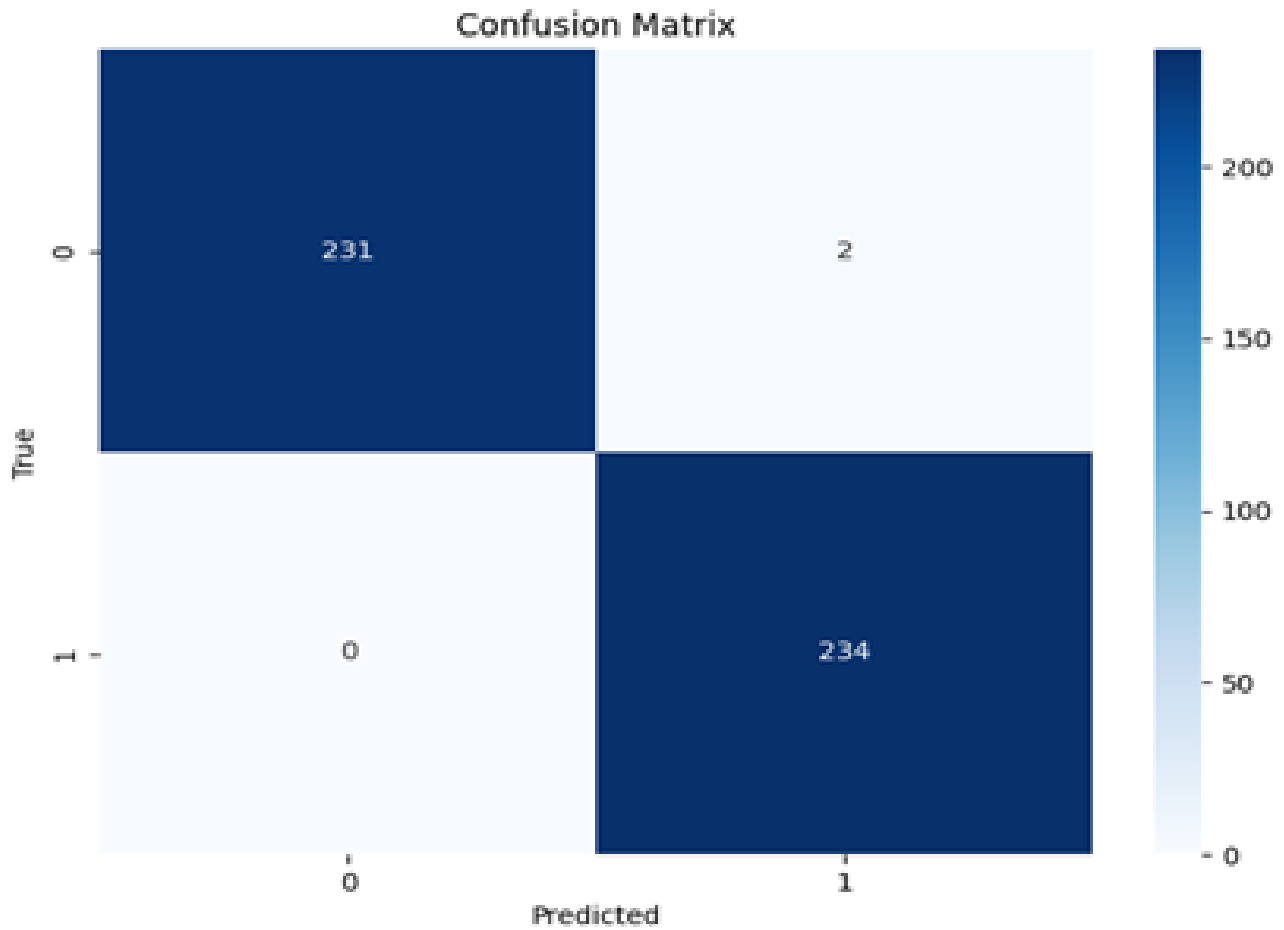


FIGURE 4.7 – Matrice de confusion relative à l’approche proposée.

La comparaison des approches dans le domaine de la santé numérique reste difficile en raison de plusieurs limitations. En effet, les chercheurs utilisent différentes bases de données, dont beaucoup font partie de projets hospitaliers impliquant des collaborations entre universités et hôpitaux. En outre, les échantillons d’images utilisés pour l’évaluation sont constamment différents. De plus, les réglages et les paramètres utilisés varient d’une approche à l’autre. Tous ces éléments contribuent à la difficulté de réaliser une comparaison équitable entre les différentes approches.

Nous présentons sur le tableau 4.2 une comparaison de l’approche proposée avec des approches de la littérature.

TABLE 4.2 – Comparaison de l’approche proposée vis-à-vis des approches de l’état de l’art pour le pronostic automatique de cancer de sein.

Référence	Magraphies utilisée	Technique d’ap- rentissage Automatique	Descripteurs	Evaluation Base des résultats
[48]	MIAS	NSST/GLCM/Cropping	SVM/NB/MLP/KNN/LDA	Acc = 59.1%
[49]	Une base locale	DCNN	GLCM/PCA/LDA/MDS	Acc = 82.43%
[50]	DDSM	BP-NN	GLCM	Acc = 97%
[51]	DDSM	DT/KNN.	FCMRG/LBP-GLCM/LPQ/mRMR/ Cropping	Acc = 94.4 %
[52]	MIAS	SVM/LDA	Hough Transform/Canny Edge Detection/ Maximization Estimation	Acc = 94%/86%
[58]	MIAS	CNN	5 Conv. Layers +2 Fully Connected Layers+ Softmax Classifier	Acc = 90.5 %
[59]	DDSM	DCNN	AlexNet[5 Conv. Layers +3 Pooling Layers + 2 Fully Connected Layers] + SVM Classifier)	Acc = 87.2%
[61]	DDSM	DCNN	5 Conv. Layers +3 Fully Connected Layers + SVM Classifier	Acc = 96.7 %
[63]	DDSM	CNN	3 Conv. Layers +2 Fully Connected Layers + Softmax Classifier	Acc = 80 %
[71]	DDSM	SVM	segmentation basée fusion de u-net et vgg-16 + Zipf+Zipf inverse+gabor	Acc=84%
Approche proposée	DDSM	SVM	segmentation basée transfer learning Zipf+Zipf inverse+gabor	Acc=99%

4.7 Explication du code

Le code suivant permet la réalisation du filtrage des mammographie par CLAHE.

```
2 import numpy as np
3 import cv2
4 import os
5 import argparse
6
7 from skimage import io
8 from tqdm import tqdm
9
10 def get_all_file_paths(directory):
11
12     # initializing empty file paths list
13     file_paths = []
14
15     # crawling through directory and subdirectories
16     print("get all file paths")
17     for root, directories, files in tqdm(os.walk(directory)):
18         for filename in files:
19             # join the two strings in order to form the full filepath.
20             filepath = os.path.join(root, filename)
21             file_paths.append(filepath)
22
23     # returning all file paths
24     return file_paths
25
26 def get_args():
27
28     parser = argparse.ArgumentParser(description = "Qata_Covid19 Segmentation" ,
29                                     formatter_class=argparse.ArgumentDefaultsHelpFormatter)
```

FIGURE 4.8 – Code CLAHE.

Le code suivant permet la réalisation du filtrage des mammographie par le filtre Median.

```
1 import cv2
2 import numpy as np
3 import os
4
5 kernel1 = np.ones((3, 3), np.float32) / 9
6
7 for image in os.listdir("images"):
8     try:
9         # result = cv2.filter2D(src=cv2.imread("images/"+image), ddepth=-1, kernel=kernel1)
10        # cv2.imwrite("new_moyenneur/"+image, result)
11        result = cv2.medianBlur(src=cv2.imread("images/"+image), ksize=3)
12        cv2.imwrite("new_median/"+image, result)
13    except:
14        pass
```

FIGURE 4.9 – Code du filtre Median.

Le code suivant désigne le code du modèle U-net.

```
1 #!/usr/bin/env python3
2 import numpy as np
3 import torch
4 import torch.nn as nn
5 import torchvision.transforms as transforms
6 from QataCovDataset import QataCovDataset
7 import argparse
8 import logging
9 import os
10 import sys
11
12 import matplotlib.pyplot as plt
13
14 from torch import optim
15 from tqdm import tqdm
16 from model.unet import UNet
17 import time
18 import copy
19 from sklearn.model_selection import train_test_split
20 from sklearn.utils import shuffle
21 from DiceLoss import DiceLoss
22
23 def train_and_validate(net,criterion, optimizer, scheduler, dataloader,device,epochs, load_model = None):
```

FIGURE 4.10 – Code du modèle de segmentation U-net.

Le Code suivant permet le post-traitement(cropping)images.

```
1  #appeler les package:
2  import numpy as np
3  import cv2
4  import imutils
5  import os
6
7  folder_path = r"D:\dataset\Predict_Ground-truths"
8  output_folder = r"D:\dataset\Mask_processing"
9
10 # Iterate over all files in the folder
11 for filename in os.listdir(folder_path):
12     # Check if the file is an image
13     if filename.endswith(('.jpg', '.jpeg', '.png', '.gif')):
14         # Construct the full file path
15         file_path = os.path.join(folder_path, filename)
16
17         # Read the image using OpenCV
18         img = cv2.imread(file_path)
19
20         # Perform image processing steps
21         gray = cv2.cvtColor(img, cv2.COLOR_RGB2GRAY)
22         thresh = cv2.threshold(gray, 45, 255, cv2.THRESH_BINARY)[1]
23         thresh = cv2.erode(thresh, None, iterations=2)
24         thresh = cv2.dilate(thresh, None, iterations=2)
25
26         cnts = cv2.findContours(thresh.copy(), cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
```

FIGURE 4.11 – Code Post-traitement(cropping) images.

Le Code suivant permet l'extraction des descripteurs dans un fichier.csv.

```
1 import os
2 import csv
3
4 data_dir = r"D:\dataset\Images"
5 csv_file = r"D:\dataset\datafilter.csv"
6
7 # Open the CSV file in write mode
8 with open(csv_file, "w", newline="") as file:
9     writer = csv.writer(file)
10    writer.writerow(["index", "img", "target"]) # Write header row
11    index=0
12    # Iterate over all files in the directory
13    for filename in os.listdir(data_dir):
14        if filename.endswith(".png"):
15
16            label = 0 # Replace with the actual label for the image
17            if "Calc" in filename:
18                label = 1
19            else:
20                label = 0
21            # Write the image path and label to the CSV file
22            writer.writerow([ index,filename, label])
23            index=index+1
```

FIGURE 4.12 – Code Créer un fichier.CSV.

Le Code suivant permet la sélection des descripteurs les plus discriminants par l'ACP.

```
1 import pandas as pd
2 from sklearn.decomposition import PCA
3
4 data = pd.read_csv('/kaggle/input/data-csv/data.csv')
5
6 non_numeric_columns = ['img']
7 numeric_columns = [col for col in data.columns if col not in non_numeric_columns]
8 X = data[numeric_columns]
9
10 pca = PCA(n_components=200) # Set n_components to 100 for 100 dimensions
11
12 X_pca = pca.fit_transform(X)
13
14 df_pca = pd.DataFrame(X_pca, columns=['PC{}'.format(i+1) for i in range(200)]) # Adjust column names to PC1, PC2, ..., PC100
15
16 df_pca.to_csv('path_to_save_pca_data.csv', index=False)
```

FIGURE 4.13 – Code de ACP.

Le Code suivant est relatif au modèle SVM.

```
1 # This Python 3 environment comes with many helpful analytics libraries installed
2 # It is defined by the kaggle/python docker image: https://github.com/kaggle/docker-python
3 # For example, here's several helpful packages to load in
4
5 import numpy as np # linear algebra
6 import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
7 import matplotlib.pyplot as plt # for data visualization
8 import seaborn as sns # for statistical data visualization
9 import argparse
10 import os
11
12
13 # matplotlib inline
14
15 # Any results you write to the current directory are saved as output.
16 def get_args():
17     parser = argparse.ArgumentParser(description="U-Net for Lung Segmentation",
18                                     formatter_class=argparse.ArgumentDefaultsHelpFormatter)
19
20     # set your environment
21     parser.add_argument('--path', type=str, default=r"D:\dataset")
22     parser.add_argument('--out', type=str, default=r"D:\dataset")
23     return parser.parse_args()
24
25
26 def main():
27     # data = '/kaggle/input/predicting-a-pulsar-star/pulsar_stars.csv'
28
29     args = get_args()
```

FIGURE 4.14 – Code des SVM.

4.8 Fonctionnement du système développé

Une fois l'application lancée, l'utilisateur pourra se connecter en utilisant l'interface 'Démarrer' en entrant son nom d'utilisateur et son prénom :

The image shows a web browser window titled "Form" with standard window controls. The main content is a login interface for a breast cancer detection application. At the top, a white box contains the title "Application de détection du cancer du sein". Below this, there are two input fields: the first is labeled "Nom" and the second is labeled "prénom". A "Démarrer" button with a play icon is positioned below the "prénom" field. The background features a pink and purple illustration of a woman with long dark hair, a pink awareness ribbon, and the text "Breast CANCER AWARENESS MONTH".

FIGURE 4.15 – Interface login du système mmo-classification.

Ainsi, l'utilisateur accède à l'interface suivante :

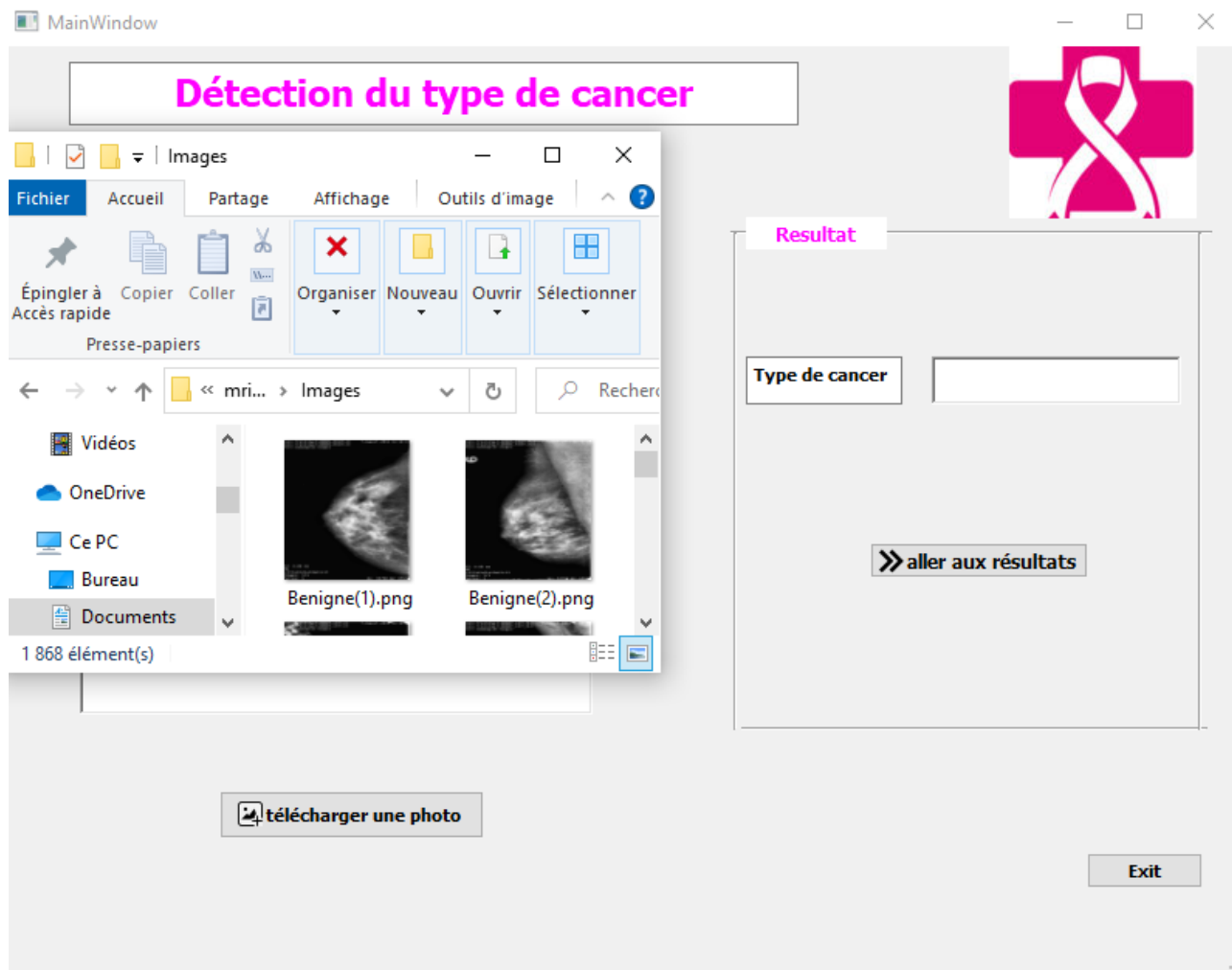


FIGURE 4.16 – Interface principale du système mmo-classification. Avec affichage de resultat de SVM model.

Après avoir sélectionné le bouton "Télécharger une photo" dans le menu principal, l'utilisateur est autorisé à parcourir les fichiers du système pour choisir la mammographie à traiter.

- Une fois que l'image de la tumeur s'affiche à l'écran, cliquez sur le bouton "Aller aux résultats".
- Après avoir cliqué sur ce bouton, le résultat de la classification (bénigne / maligne) sera affiché sous l'image.
- Pour quitter l'application, il suffit d'appuyer sur le bouton "Exit".

4.9 Conclusion

Nous avons discuté en détail de l'implémentation de l'approche proposée pour la prédiction automatique du cancer du sein, ainsi que de la validation des résultats obtenus. Cependant, la comparaison des performances de notre approche avec celles de l'état de l'art n'était pas évidente en raison de plusieurs facteurs. Ces facteurs comprennent l'utilisation de différentes bases de données de mammographie, en particulier dans le contexte des projets hospitaliers, la disparité du nombre de mammographies utilisées pour l'apprentissage et les tests, ainsi que d'autres lacunes.

CHAPITRE 5

CONCLUSION GÉNÉRALE

La quantité de données d'imagerie médicale augmente rapidement en raison des progrès du matériel, de l'augmentation de la population, de la diminution des coûts et de la prise de conscience de l'utilité des modalités d'imagerie.

Cela ajoute à la difficulté croissante pour les radiologues et les cliniciens de faire face à la charge croissante d'analyser les grandes quantités de données disponibles provenant de sources de données disparates [70]. Il s'ensuit qu'il existe un besoin évolutif d'outils pouvant aider au diagnostic et à la prise de décision.

Les progrès récents de l'intelligence artificielle (IA) ont commencé à se répandre dans les domaines de santé consistent en des modules non linéaires capables d'apprendre automatiquement plusieurs niveaux de représentations à partir de données de grande dimension sans aucun besoin d'ingénierie explicite des caractéristiques par les humains.

Les méthodes d'apprentissage en profondeur, en particulier les réseaux de neurones convolutifs (CNN), deviennent de facto la norme pour résoudre les problèmes de vision par ordinateur.

Les solutions neuronales profondes actuelles sont qualifiées de boîtes noires en raison d'un manque de compréhension des spécificités concernant le processus de prise de décision.

Traditionnellement, les pathologistes/radiologues avaient l'habitude d'observer manuellement les images mammaires et, avec le consensus d'autres experts médicaux, finalisaient leurs décisions et leurs analyses.

Cependant, l'observation manuelle d'un grand nombre d'images mammaires pour un éventuel diagnostic de cancer du sein est un processus long et fastidieux qui conduit souvent à des résultats faux positifs ou faux négatifs [6]. Par conséquent, un système automatisé est toujours

nécessaire pour accélérer le processus d'analyse des images et aider les radiologues à diagnostiquer précocement le cancer du sein. Ces systèmes automatisés offrent aux radiologues la possibilité d'obtenir un deuxième avis et, par conséquent, ils peuvent prendre des décisions plus solides, fiables et précises concernant le diagnostic du cancer du sein.

Les principaux défis de recherche abordés dans ce domaine concernent la détection du cancer du sein, avec l'ambiguïté des images mammaires denses, à l'aide de méthodes d'apprentissage automatique et d'apprentissage profond où le processus de segmentation était basé sur la technique d'apprentissage par transfert et le processus de classification était basé sur les SVM.

En effet, les résultats obtenus étaient très satisfaisants atteignant une précision de distinction entre les tumeurs malignes et bénignes de l'ordre de 99%. Nous soulignons quelques perspectives comme suit [4] :

- Bien que la précision observée des méthodologies utilisées basées sur l'apprentissage automatique et l'apprentissage profond ait atteint un taux d'erreur inférieur à celui des humains, elle est encore loin des cas d'utilisation réels, car nous ne pouvons pas croire aux résultats fournis par les méthodes d'IA. L'une des raisons pourrait être l'inexplicabilité des résultats interprétés par le modèle d'apprentissage automatique ou profond utilisé et d'autres pourraient être que la plupart des chercheurs ont appliqué ces méthodologies sur de petits ensembles de données ou n'ont pas utilisé des bases de données de référence.
- De nombreux défis à l'adoption généralisée de l'IA dans l'imagerie mammaire subsistent, notamment le coût de la mise en œuvre ainsi que les implications éthiques et juridiques du consentement éclairé, des préjugés, de la responsabilité, de la confidentialité et de la cybersécurité. Malgré ces défis, l'IA continuera probablement à s'améliorer en tant qu'outil d'assistance en imagerie mammaire avec un grand potentiel pour améliorer le dépistage du cancer du sein.
- Bien que les méthodes basées sur l'apprentissage automatique ou profond aient montré jusqu'à présent leur capacité prédictive prometteuse dans le domaine du diagnostic du cancer, néanmoins, ces solutions sont encore loin d'être acceptées par le public en raison de la nature de la boîte noire ou de l'inexplicabilité des résultats obtenus à partir de ces approches.

BIBLIOGRAPHIE

- [1] AGARWAL, Richa, et al. Deep learning for mass detection in full field digital mammograms. *Computers in biology and medicine*, 2020, 121 : 103774.
- [2] SOULAMI, Khaoula Belhaj ; KAABOUCHE, Naima ; SAIDI, Mohamed Nabil. Breast cancer : Classification of suspicious regions in digital mammograms based on capsule network. *Biomedical Signal Processing and Control*, 2022, 76 : 103696.
- [3] BACCOUCHE, Asma, et al. Early detection and classification of abnormality in prior mammograms using image-to-image translation and YOLO techniques. *Computer Methods and Programs in Biomedicine*, 2022, 221 : 106884.
- [4] PAINULI, Deepak, et al. Recent advancement in cancer diagnosis using machine learning and deep learning techniques : A comprehensive review. *Computers in Biology and Medicine*, 2022, 105580.
- [5] DADSETAN, Saba, et al. Deep learning of longitudinal mammogram examinations for breast cancer risk prediction. *Pattern Recognition*, 2022, 132 : 108919.
- [6] SHAH, Shahid Munir, et al. Artificial intelligence for breast cancer analysis : Trends directions. *Computers in Biology and Medicine*, 2022, 105221.
- [7] HA, Richard ; JAIRAM, Meghan P. A review of artificial intelligence in mammography. *Clinical Imaging*, 2022.
- [8] <https://www.journaldunet.fr/web-tech/dictionnaire-de-l-iot/1489513-computer-vision-definition-et-applications-de-la-vision-par-ordinateur/>.
- [9] BOUCHER, Arnaud. Recalage et analyse d'un couple d'images : application aux mammographies. 2013. PhD Thesis. Université René Descartes-Paris V.

- [10] DIPERT, Brian ; SHOHAM, Amit. Eye, robot : Embedded vision, the next big thing in digital signal processing. IEEE Solid-State Circuits Magazine, 2012, 4.2 : 26-29.
- [11] YOUNG, Ian T. ; GERBRANDS, Jan J. ; VAN VLIET, Lucas J. Fundamentals of image processing. 1998.
- [12] AMROUN, Fazia. Extraction de la composante texturée d'une image. 2013. PhD Thesis. Université Mouloud Mammeri.
- [13] HARALICK, Robert M. Statistical and structural approaches to texture. Proceedings of the IEEE, 1979, 67.5 : 786-804
- [14] PAULHAC, Ludovic. Outils et méthodes d'analyse d'images 3D texturées : application à la segmentation des images échographiques. 2009. PhD Thesis. Université François Rabelais-Tours.
- [15] GADA, Saliha. Décomposition Modale Empirique : Application à l'analyse de la texture. 2013. PhD Thesis. Université Mouloud Mammeri.
- [16] BOUAFIA, Nadjat. Classification efficace des vêtements de mode basée sur les approches : apprentissage automatique ML et apprentissage profond DL. 2020. PhD Thesis. FACULTE MATHEMATIQUES ET INFORMATIQUE-DEPARTEMENT INFORMATIQUE-OPTION : Informatique Décisionnelle Et Optimisation.
- [17] DE MATTEIS, Ludovic. Introduction à l'apprentissage automatique.
- [18] BOUAFIA, Nadjat. Classification efficace des vêtements de mode basée sur les approches : apprentissage automatique ML et apprentissage profond DL. 2020. PhD Thesis. FACULTE MATHEMATIQUES ET INFORMATIQUE-DEPARTEMENT INFORMATIQUE-OPTION : Informatique Décisionnelle Et Optimisation.
- [19] ALOUANE, Khadidja ; SAIDOUN, Meryem ; LAIDI, Maamar. Modélisation de l'adsorption multi-composants par la méthode machine à vecteurs de support. 2021.
- [20] NASTESKI, Vladimir ; MANEVSKA, Violeta ; SAVOSKA, Snezana. Optimizing and Verifying Model Effectiveness Using Random Forest and Logistic Regression. 2017.
- [21] BAAZI, Khaled. Classification automatique de la parole pathologique par la méthode des arbres de décisions chez des locuteurs présentant une fente palatine ou labio-palatine. 2022. PhD Thesis.
- [22] MARÉE, Raphaël. Classification automatique d'images par arbres de décision. 2005.

- [23] GHENNAM BOCHRA, SMARA Safia. Les réseaux de neurone convolutionnel (CNN) pour la classification des images associées aux places de stationnement d'un parc de véhicule. 2019.
- [24] OTOUM, Yazan ; LIU, Dandan ; NAYAK, Amiya. DL-IDS : a deep learning–based intrusion detection framework for securing IoT. *Transactions on Emerging Telecommunications Technologies*, 2022, 33.3 : e3803.
- [25] KHAN, Asif Iqbal ; SHAH, Junaid Latief ; BHAT, Mohammad Mudasar. CoroNet : A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Computer methods and programs in biomedicine*, 2020, 196 : 105581.
- [26] ZERKOUT, Mohamed Lotfi, et al. Classification des images veines de la main par un réseau de neurone convolutif. 2022.
- [27] NOVIKOV, Alexander, et al. Tensorizing neural networks. *Advances in neural information processing systems*, 2015, 28.
- [28] DU, Getao, et al. Medical image segmentation based on u-net : A review. *Journal of Imaging Science and Technology*, 2020.
- [29] TAHIR, Anas M., et al. Deep learning for reliable classification of COVID-19, MERS, and SARS from chest X-ray images. *Cognitive Computation*, 2022, 1-21.
- [30] He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [31] ÇIÇEK, Özgün, et al. 3D U-Net : learning dense volumetric segmentation from sparse annotation. In : *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016 : 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II* 19. Springer International Publishing, 2016. p. 424-432.
- [32] ZHOU, Zongwei, et al. Unet++ : A nested u-net architecture for medical image segmentation. In : *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*. Springer International Publishing, 2018. p. 3-11.
- [33] BELAI, Amira, et al. Etude des effets secondaires de la chimiothérapie : cas du cancer du sein. 2022.

- [34] HARIZ BEKKAR, Ouacila. Analyse de la texture des images mammaires par une fusion des lois de Zipf et des Ondelettes pour la classification des tumeurs mammaires via l'analyse en composantes principale. 2018.
- [35] ALY, Ghada Hamed, et al. YOLO based breast masses detection and classification in full-field digital mammograms. *Computer Methods and Programs in Biomedicine*, 2021, 200 : 105823.
- [36] TARDIVON, A. ; MALHAIRE, C. Cancer du sein. *Epidemiologie, facteurs de risque, imagerie. EMC-RADIOLOGIE ET IMAGERIE MÉDICALE : Génito-urinaire-Gynécobobstétricale-Mammaire*, 2009, 1-30.
- [37] BADEREDDINE, Belkis. Analyse de la texture des images mammaires par une fusion des lois de Zipf et des SFTA pour la classification des tumeurs mammaires via l'analyse en composantes principales. 2018.
- [38] MAHAPATRA, Debolina ; RAY, Ratula ; DASH, Satya Ranjan. Deep Learning Model for Efficient Mammogram Analysis. *Technical Advancements of Machine Learning in Healthcare*, 2021, 223-240.
- [39] HAMOUD, Meriem. Indexation et segmentation d'images basées loi de Zipf et Zipf inverse. 2015. PhD Thesis. Université Badji Mokhtar.
- [40] MANSEL, Robert E. ; WEBSTER, David JT. Hughes, Mansel Webster's benign disorders and diseases of the breast. 2009.
- [41] DAO, Adama. Apport de l'échographie et la mammographie dans le diagnostic des pathologies mammaires. 2022.
- [42] BADEREDDINE, Belkis. Analyse de la texture des images mammaires par une fusion des lois de Zipf et des SFTA pour la classification des tumeurs mammaires via l'analyse en composantes principales. 2018.
- [43] CHEIKHROUHOU, Imen. Description et classification des masses mammaires pour le diagnostic du cancer du sein. 2012. PhD Thesis. Université d'Evry-Val d'Essonne.
- [44] BEURA, Shradhananda. Development of Features and Feature Reduction Techniques for Mammogram Classification. 2016. PhD Thesis.
- [45] HARIZ BEKKAR, Ouacila. Analyse de la texture des images mammaires par une fusion des lois de Zipf et des Ondelettes pour la classification des tumeurs mammaires via l'analyse en composantes principale. 2018.

- [46] SAMPAT, Mehul P., et al. Computer-aided detection and diagnosis in mammography. Handbook of image and video processing, 2005, 2.1 : 1195-1217.
- [47] ZHOU, Lei, et al. Three-dimensional affinity learning based multi-branch ensemble network for breast tumor segmentation in MRI. Pattern Recognition, 2022, 129 : 108723.
- [48] PERUMAL, Varalakshmi, et al. Performance evaluation and comparative analysis of various machine learning techniques for diagnosis of breast cancer. Biomedical Research (0970-938X), 2016, 27.3.
- [49] SUN, Wenqing, et al. Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data. Computerized Medical Imaging and Graphics, 2017, 57 : 4-9.
- [50] MUGHAL, Bushra, et al. A novel classification scheme to decline the mortality rate among women due to breast tumor. Microscopy research and technique, 2018, 81.2 : 171-180.
- [51] SADAD, Tariq, et al. Fuzzy C-means and region growing based classification of tumor from mammograms using hybrid texture feature. Journal of computational science, 2018, 29 : 34-45.
- [52] VIJAYARAJESWARI, R., et al. Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform. Measurement, 2019, 146 : 800-805.
- [53] DE NAZARÉ SILVA, Joberth, et al. Automatic detection of masses in mammograms using quality threshold clustering, correlogram function, and SVM. Journal of digital imaging, 2015, 28 : 323-337.
- [54] AMINIKHANGHAHI, Samaneh, et al. A new fuzzy Gaussian mixture model (FGMM) based algorithm for mammography tumor image classification. Multimedia Tools and Applications, 2017, 76 : 10191-10205.
- [55] XIE, Weiying ; LI, Yunsong ; MA, Yide. Breast mass classification in digital mammography based on extreme learning machine. Neurocomputing, 2016, 173 : 930-941.
- [56] VAIDEHI, K. ; SUBASHINI, T. S. Automatic classification and retrieval of mammographic tissue density using texture features. In : 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO). IEEE, 2015. p. 1-6.
- [57] KHALAF, Aya F. ; YASSINE, Inas A. Spectral correlation analysis for microcalcification detection in digital mammogram images. In : 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). IEEE, 2015. p. 88-91.

- [58] TING, Fung Fung; TAN, Yen Jun; SIM, Kok Swee. Convolutional neural network improvement for breast cancer classification. *Expert Systems with Applications*, 2019, 120 : 103-115.
- [59] RAGAB, Dina A., et al. Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ*, 2019, 7 : e6201.
- [60] SALAMA, Wessam M.; ALY, Moustafa H. Deep learning in mammography images segmentation and classification : Automated CNN approach. *Alexandria Engineering Journal*, 2021, 60.5 : 4701-4709.
- [61] CHOUGRAD, Hiba; ZOUAKI, Hamid; ALHEYANE, Omar. Deep convolutional neural networks for breast cancer screening. *Computer methods and programs in biomedicine*, 2018, 157 : 19-30.
- [62] CHOUGRAD, Hiba; ZOUAKI, Hamid; ALHEYANE, Omar. Deep convolutional neural networks for breast cancer screening. *Computer methods and programs in biomedicine*, 2018, 157 : 19-30.
- [63] SINGH, Vivek Kumar, et al. Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. *Expert Systems with Applications*, 2020, 139 : 112855.
- [64] CUMMINS, Nicholas, et al. Machine learning in digital health, recent trends, and ongoing challenges. In : *Artificial Intelligence in Precision Health*. Academic Press, 2020. p. 121-148.
- [65] AL-KADI, Omar S. A gabor filter texture analysis approach for histopathological brain tumor subtype discrimination. *arXiv preprint arXiv :1704.05122*, 2017.
- [66] BELKACEM, Allel; LYES, Mihoubi. Analyse de texture par les filtres de Gabor et Laws. 2014. PhD Thesis. Université Mouloud Mammeri.
- [67] HAGHIGHAT, Mohammad; ZONOUZ, Saman; ABDEL-MOTTALEB, Mohamed. CloudID : Trustworthy cloud-based and cross-enterprise biometric identification. *Expert Systems with Applications*, 2015, 42.21 : 7905-7916.
- [68] ZUCKERKANDL, Emile; PAULING, Linus. Evolutionary divergence and convergence in proteins. In : *Evolving genes and proteins*. Academic Press, 1965. p. 97-166. <https://doi.org/10.1016/B978-0-12-819583-3.00002-3>.
- [69] GOSWAMI, Tilottama. Machine learning behind classification tasks in various engineering and science domains. In : *Cognitive Informatics, Computer Modelling, and Cognitive Science*. Academic Press, 2020. p. 339-356. <https://doi.org/10.1016/B978-0-12-819443-0.00016-7>.

- [70] PUTTAGUNTA, Muralikrishna ; RAVI, S. Medical image analysis based on deep learning approach. *Multimedia tools and applications*, 2021, 80 : 24365-24398.
- [71] Belhadi, N. Gobbi, A. (2021). *Classification Des Tumeurs Mammaires Basée Deep Learning* Mémoire de Master, Université Hamma Lakhdar - Eloued.