

**People's Democratic Republic of Algeria**  
**Ministry of Higher Education and Scientific Research**  
**University of 8 May 1945-Guelma-**  
**Faculty of Mathematics, Computer Science and Science of Matter**  
**Department of Computer Science**



## Master Thesis

Specialty: Computer Science

### Option:

Science and technology of information and communication

## Theme

---

# Classification des images mammographiques selon leurs densités

---

**Presented by:**

BENMAAMAR Oussama

**Jury Members:**

N	Full Name	Quality
1	Dr MADI Leyla	Chairman
2	Dr FERKOUS Chokri	Supervisor
3	Dr FAREK Lazhar	Examiner

June 2023.

# Résumé

---

Le cancer de sein est une pathologie grave chez les femmes, où la détection précoce devient primordiale, car elle permet d'augmenter considérablement les chances de survie. La mammographie est une technique essentielle dans ce processus, mais son interprétation est complexe en raison de la variabilité de la densité mammaire.

Actuellement, les techniques d'apprentissage automatique, notamment l'apprentissage profond (Deep Learning) sont largement utilisées pour la classification d'images. Le but principale de ce projet de fin d'étude est la classification des mammographies numériques selon leurs densités (B ou C). Les expérimentations de ce projet ont été effectuées sur la base de données mammographique RSNA (The Radiological Society of North America), cette base est composée de 24350 mammographies. Pour réduire le bruit, quelques techniques de prétraitement ont été appliquées, et la classification a été faite en se basant sur le transfert d'apprentissage à travers des modèles pré-entraînés tels que VGG16, ResNet50, MobileNetV2 et EfficientNetB0.

Parmi ces modèles, EfficientNetB0 a été élu comme meilleur modèle de classification en raison de ces bons résultats. Ce modèle a atteint un taux de reconnaissance de 85,64%, une précision de 86%, un rappel de 88% et un score F1 de 86%.

**Mots-clefs :** transfert d'apprentissage, mammographie, classification, prétraitement.

## ملخص

يعتبر سرطان الثدي من الأمراض الخطيرة لدى النساء، حيث يصبح الكشف المبكر عنه ضروريًا، لأنه يزيد بشكل كبير من فرص البقاء على قيد الحياة. يعتبر التصوير الشعاعي للثدي تقنية أساسية في هذه العملية، لكن تفسيرها معقد بسبب تنوع كثافة الثدي. حاليًا، تُستخدم تقنيات التعلم الآلي، وخاصة التعلم العميق، على نطاق واسع لتصنيف الصور. الهدف الرئيسي لمشروع نهاية الدراسة هذا هو تصنيف صور الثدي الرقمية وفقًا لكثافتها (B أو C). أجريت تجارب هذا المشروع على قاعدة بيانات التصوير الشعاعي للثدي RSNA (الجمعية الإشعاعية لأمريكا الشمالية)، وتتكون قاعدة البيانات هذه من 24350 صورة شعاعية للثدي. لتقليل الضوضاء، تم تطبيق بعض تقنيات المعالجة المسبقة، وتم التصنيف بناءً على نقل التعلم من خلال نماذج مدربة مسبقًا مثل VGG16 و ResNet50 و MobileNetV2 و EfficientNetB0.

من بين هذه النماذج، تم اختيار EfficientNetB0 كأفضل نموذج تصنيف بسبب النتائج الجيدة. حقق هذا النموذج معدل التعرف 85.64% ودقة 86% واسترجاع 88% ودرجة F1 86%.

الكلمات المفتاحية: نقل التعلم، التصوير الشعاعي للثدي، التصنيف، المعالجة المسبقة.

# Abstract

---

Breast cancer is a serious pathology in women, where early detection is of prime importance, as it considerably increases the chances of survival. Mammography is an essential technique in this process, but its interpretation is complex due to the variability of breast density.

Currently, machine learning techniques, notably Deep Learning, are widely used for image classification. The main aim of this final-year project is to classify digital mammograms according to their density (B or C). The experiments in this project were carried out on the RSNA (The Radiological Society of North America) mammography database, consisting of 24350 mammograms. To reduce noise, some pre-processing techniques were applied, and classification was done based on transfer learning through pre-trained models such as VGG16, ResNet50, MobileNetV2 and EfficientNetB0.

Among these models, EfficientNetB0 was elected as the best classification model because of these good results. This model achieved a recognition rate of 85.64%, a precision of 86%, a recall of 88% and an F1 score of 86%.

**Keywords** : transfer learning, mammography, classification, pre-processing.

# REMERCIEMENTS

---

En tout premier lieu, je remercie Dieu, tout puissant, de m'avoir donné la force pour survivre, ainsi que l'audace pour dépasser toutes les difficultés.

Je commence par exprimer ma profonde gratitude envers mon encadreur Dr. FERKOUS Chokri, qui a accepté de superviser mon travail de mémoire. Je suis extrêmement honoré d'avoir bénéficié de son encadrement, et je tiens à le remercier chaleureusement pour sa patience, sa disponibilité et ses précieux conseils.

Je remercie tous les enseignants du Département d'informatique pour leur aide et leur soutien durant l'année scolaire.

Je tiens à remercier aussi ma mère la raison de mon existence, et mes proches, qui m'ont soutenu durant les moments de doute et d'abandon, qui ont plus cru en moi.

**BENMAAMAR Oussama**

# DEDICACE

---

*Je dédie cet ouvrage*

*A ma très chère mère, à qui je ne pourrai jamais exprimer ma gratitude suffisamment. Ta tendresse m'enveloppe, ta bienveillance me guide et ta présence à mes côtés a toujours été ma source de force face aux obstacles.*

*A mes chères sœurs, je dédie ce travail pour leur soutien moral et leurs précieux conseils tout au long de mes études.*

*A mes amis, notamment à Saïd Bouteldja, qui m'ont constamment encouragé et à qui je souhaite encore plus de succès.*

*Et à tous ceux que j'aime, je vous exprime ma reconnaissance.*

*Merci !*

**BENMAAMAR Oussama**

# Table des matières

---

RESUME .....	I
ملخص .....	II
ABSTRACT .....	III
REMERCIEMENTS .....	IV
DEDICACE .....	V
TABLE DES MATIERES .....	1
LISTE DES FIGURES .....	6
LISTE DES TABLEAUX .....	8
ABREVIATIONS ET ACRONYMES .....	9
CONDITIONAL GENERATIVE ADVERSARIAL NETWORK .....	10
INTRODUCTION GENERALE .....	11
<b>CHAPITRE I: INTELLIGENCE &amp; VISION ARTIFICIELLE .....</b>	<b>13</b>
I.1. INTRODUCTION .....	14
I.2. APPRENTISSAGE AUTOMATIQUE.....	14
I.3. LES APPROCHES D'APPRENTISSAGE .....	15
<i>I.3.1. Apprentissage supervisé.....</i>	<i>15</i>
I.3.1.1. Machines à vecteurs de support (SVM).....	15
I.3.1.2. Arbres de décision .....	16
I.3.1.3. Régression logistique .....	16
I.3.1.4. Les réseaux neuronaux .....	17
<i>I.3.2. Apprentissage non-supervisé .....</i>	<i>18</i>
I.3.2.1. K-Means.....	18
I.3.2.2. Plus proche voisin (K-NN).....	19
<i>I.3.3. Apprentissage par renforcement .....</i>	<i>20</i>
I.4. DOMAINES D'UTILISATION .....	20
I.5. CONCLUSION .....	21
<b>CHAPITRE II: LES RESEAUX DE NEURONES ET APPRENTISSAGE PROFOND.....</b>	<b>22</b>
II.1. INTRODUCTION.....	23
II.2. LES RESEAUX DE NEURONES.....	23
<i>II.2.1. Les neurones biologiques .....</i>	<i>23</i>

II.2.2.	<i>Le Neurone formel</i> .....	24
II.2.3.	<i>Réseaux de Neurones Artificiels (RNA)</i> .....	25
II.2.3.1.	Réseau monocouche.....	25
II.2.3.2.	Réseau multicouche.....	26
II.2.4.	<i>Les fonctions d'activations</i> .....	26
II.2.4.1.	La fonction sigmoïde.....	27
II.2.4.2.	La fonction tangente hyperbolique (Tanh).....	28
II.2.4.3.	La fonction Softmax.....	28
II.2.4.4.	La fonction Softsign.....	29
II.2.4.5.	La fonction de l'unité linéaire rectifiée (ReLU).....	30
II.2.4.6.	La fonction Leaky ReLU (LReLU).....	31
II.2.4.7.	La fonction Softplus.....	31
II.2.4.8.	La fonction unités linéaires exponentielles (ELUs).....	32
II.2.4.9.	La fonction Maxout.....	33
II.2.4.10.	La fonction ELiSH.....	33
II.2.5.	<i>Les fonctions de perte</i> .....	34
II.2.5.1.	Entropie croisée (Cross-entropy CE):.....	34
II.2.5.2.	Erreur moyenne au carré (squared mean error MSE):.....	35
II.2.5.3.	Erreur absolue moyenne ( Mean absolute error MAE ):.....	35
II.2.5.4.	Charnière-perte.....	35
II.2.6.	<i>Les algorithmes d'optimisation</i> .....	35
II.2.6.1.	Batch Gradient Decent.....	35
II.2.6.2.	Stochastic Gradient Decent.....	36
II.2.6.3.	Mini-Batch Gradient Decent.....	36
II.2.6.4.	Descente de gradient avec Momentum.....	37
II.2.6.5.	Nesterov Accelerated Gradient.....	37
II.2.6.6.	Adagrad.....	37
II.2.6.7.	Adadelata.....	37
II.2.6.8.	RMSprop.....	38
II.2.6.9.	Adam.....	38
II.3.	APPRENTISSAGE PROFOND.....	39
II.3.1.	<i>Types de réseaux d'apprentissage en profondeur</i> .....	39
II.3.1.1.	Réseaux de neurones récurrents (RNN).....	39
II.3.1.2.	La mémoire longue à court terme (LSTM).....	40
II.3.1.3.	Les Réseaux de neurones convolutifs (CNN).....	41
II.3.2.	<i>Les couches de réseau neuronal convolutif (CNN)</i> .....	41
II.3.2.1.	La couche convolutive.....	42
II.3.2.1.1.	Le noyau.....	42
II.3.2.1.2.	Opérations de convolution.....	42
II.3.2.2.	La Couche de regroupement (Pooling).....	43
II.3.2.3.	Fonctions d'activation.....	44

---

II.3.2.4.	Couche entièrement connectée (Fully Connected Layer) .....	44
II.3.3.	<i>Quelques architecture basée CNN</i> .....	45
II.3.3.1.	Groupe de géométrie visuelle (VGG 16).....	45
II.3.4.	<i>Resnet 50</i> .....	46
II.3.5.	<i>MobileNet V2</i> .....	47
II.3.6.	<i>EfficientNet</i> .....	47
II.4.	TRANSFERT D'APPRENTISSAGE (TA) .....	49
II.5.	CONCLUSION.....	49
<b>CHAPITRE III: LA CLASSIFICATION DES MAMMOGRAPHIES NUMERIQUES .....</b>		<b>50</b>
III.1.	INTRODUCTION.....	51
III.2.	ANATOMIE DE SEIN.....	51
III.3.	L'IMAGERIE MEDICALE DEDIEE AUX ANOMALIES MAMMAIRES.....	52
III.3.1.	<i>L'échographie</i> .....	52
III.3.2.	<i>L'imagerie par résonance magnétique (IRM)</i> .....	53
III.3.3.	<i>La mammographie</i> .....	54
III.4.	LES PATHOLOGIES MAMMAIRES.....	55
III.4.1.	<i>Les microcalcifications (Mcs)</i> .....	55
III.4.2.	<i>Les masses</i> .....	55
III.5.	LA DENSITE DES MAMMOGRAPHIES.....	56
III.5.1.	<i>Densité de type A</i> .....	57
III.5.2.	<i>Densité de type B</i> .....	57
III.5.3.	<i>Densité de type C</i> .....	57
III.5.4.	<i>Densité de type D</i> .....	57
III.6.	LA CLASSIFICATION DES PATHOLOGIES MAMMAIRES .....	58
III.6.1.	<i>La classification de Legal</i> .....	58
III.6.2.	<i>La classification BI-RADS</i> .....	59
III.6.2.1.	BI-RADS 0.....	59
III.6.2.2.	BI-RADS 1 .....	59
III.6.2.3.	BI-RADS 2.....	59
III.6.2.4.	BI-RADS 3.....	60
III.6.2.5.	BI-RADS 4.....	61
III.6.2.6.	BI-RADS 5.....	61
III.6.2.7.	BI-RADS 6.....	62
III.7.	LES SYSTEMES DE DETECTION/DIAGNOSTIC ASSISTES PAR ORDINATEUR (DAO) .....	62
III.7.1.	<i>Détection Assistée par Ordinateur (DAOe)</i> .....	62
III.7.2.	<i>Diagnostic Assistée par Ordinateur (DAOx)</i> .....	63
III.7.3.	<i>Métriques d'évaluation des CAD</i> .....	63
III.7.3.1.	Matrice de confusion .....	64

## Table des matières

---

III.7.3.2.	Le taux de reconnaissance .....	64
III.7.3.3.	Sensibilité ou rappel .....	64
III.7.3.4.	Spécificité.....	65
III.7.3.5.	La Précision.....	65
III.7.3.6.	F1-Score .....	65
III.8.	REVUE DE LITTERATURE .....	65
III.8.1.	<i>Une méthode d'apprentissage profond pour classer les catégories de densité mammographique du sein.....</i>	65
III.8.2.	<i>Segmentation et classification entièrement automatisées de la densité mammaire à l'aide de l'apprentissage profond.....</i>	65
III.8.3.	<i>Évaluation de la densité mammographique du sein à l'aide de l'apprentissage profond : mise en œuvre clinique .....</i>	66
III.8.4.	<i>Estimation automatisée de la densité mammographique à l'aide d'un réseau entièrement convolutif</i> 66	
III.8.5.	<i>Détermination de la densité mammographique du sein à l'aide d'un réseau neuronal convolutionnel profond.....</i>	66
III.8.6.	<i>Classification de la densité mammographique à l'aide d'informations d'histogramme multirésolution .....</i>	67
III.9.	CONCLUSION.....	67
<b>CHAPITRE IV: CLASSIFICATION DE LA DENSITE MAMMOGRAPHIQUE.....</b>		<b>68</b>
IV.1.	INTRODUCTION.....	69
IV.2.	ENVIRONNEMENT DE DEVELOPPEMENT DE SYSTEME.....	69
IV.2.1.	<i>Environnement matériel.....</i>	69
IV.2.2.	<i>Langage de programmation et bibliothèque utilisée .....</i>	69
IV.2.2.1.	Langage Python.....	70
IV.2.2.2.	Tensorflow .....	70
IV.2.2.3.	Keras .....	70
IV.2.2.4.	Numpy.....	70
IV.2.2.5.	Matplotlib.....	71
IV.2.2.6.	PyDicom.....	71
IV.2.2.7.	Pandas .....	71
IV.2.2.8.	Sklearn.....	72
IV.2.2.9.	Tkinter .....	72
IV.3.	LA BASE DE DONNEES UTILISEE .....	72
IV.4.	ARCHITECTURES GENERALE DE TRAVAIL.....	73
IV.4.1.	<i>Prétraitement des mammographies.....</i>	74
IV.4.1.1.	Normalisation des mammographies .....	74
IV.4.1.2.	Suppression des annotations.....	74
IV.4.1.3.	Redimensionnement des mammographies.....	75

---

## Table des matières

---

<i>IV.4.2. Préparation des données</i> .....	76
IV.4.2.1. Partition des données.....	76
IV.4.2.1.1. La base d'apprentissage .....	76
IV.4.2.1.2. La base de validation.....	76
IV.4.2.1.3. La base de test .....	76
IV.4.2.2. Augmentation de la base des données .....	76
<i>IV.4.3. Apprentissage par transfert</i> .....	77
IV.5. LES RESULTATS .....	78
IV.6. DEPLOIEMENT DE L'APPLICATION.....	82
IV.7. CONCLUSION.....	84
<b>CONCLUSION GENERALE</b> .....	<b>85</b>
<b>REFERENCES BIBLIOGRAPHIQUES</b> .....	<b>86</b>

# Liste des Figures

---

<b>FIGURE I-1:</b> LES RELATIONS ENTRE L'INTELLIGENCE ARTIFICIELLE, L'APPRENTISSAGE AUTOMATIQUE ET L'APPRENTISSAGE PROFOND. .	14
<b>FIGURE I-2:</b> MACHINES A VECTEUR DE SUPPORT (SVM).	15
<b>FIGURE I-3:</b> EXEMPLE D'ARBRE DE DECISION.	16
<b>FIGURE I-4:</b> REGRESSION LOGISTIQUE.	17
<b>FIGURE I-5:</b> EXEMPLE D'UN RESEAU NEURONAL.	18
<b>FIGURE I-6:</b> EXEMPLE DE K-MEANS.	19
<b>FIGURE I-7:</b> PRESENTATION DE LA METHODE K-NN.	19
<b>FIGURE I-8:</b> APPRENTISSAGE PAR RENFORCEMENT.	20
<b>FIGURE II-1:</b> NEURONE BIOLOGIQUE.	23
<b>FIGURE II-2:</b> NEURONE FORMEL.	24
<b>FIGURE II-3:</b> NEURONE BIOLOGIQUE ET ARTIFICIEL.	25
<b>FIGURE II-4:</b> RESEAU MONOCOUCHE.	25
<b>FIGURE II-5:</b> RESEAU MULTICOUCHE.	26
<b>FIGURE II-6:</b> LA FONCTION D'ACTIVATION SIGMOÏDE.	27
<b>FIGURE II-7 :</b> REPRESENTATION DE LA FONCTION TANH.	28
<b>FIGURE II-8:</b> LA FONCTION SOFTMAX.	29
<b>FIGURE II-9:</b> LA FONCTION SOFTSIGN.	29
<b>FIGURE II-10 :</b> LA FONCTION D'ACTIVATION RELU.	31
<b>FIGURE II-11:</b> LA FONCTION D'ACTIVATION SOFTPLUS .	31
<b>FIGURE II-12:</b> LA FONCTION D'ACTIVATION ELU.	32
<b>FIGURE II-13:</b> LA FONCTION D'ACTIVATION MAXOUT.	33
<b>FIGURE II-14:</b> LA FONCTION D'ACTIVATION ELISH.	34
<b>FIGURE II-15:</b> DIAGRAMME DU RNN.	40
<b>FIGURE II-16:</b> L'ARCHITECTURE DE LSTM.	40
<b>FIGURE II-17:</b> ARCHITECTURE CNN.	41
<b>FIGURE II-18:</b> LES CALCULS PRIMAIRES EXECUTES A CHAQUE ETAPE DE LA COUCHE CONVOLUTIVE.	43
<b>FIGURE II-19:</b> LES OPERATIONS DE REGROUPEMENT ( POOLING ).	44
<b>FIGURE II-20:</b> LA COUCHE ENTIEREMENT CONNECTEE.	45
<b>FIGURE II-21:</b> L'ARCHITECTURE DE VGG 16.	46
<b>FIGURE II-22:</b> L'ARCHITECTURE RESNET DE BASE.[KIM21]	46

<b>FIGURE II-23:</b> BLOCS CONVOLUTIONNELS MOBILENETV2. ....	47
<b>FIGURE II-24:</b> REPRESENTATION SCHEMATIQUE D'EFFICIENT NET. [ATI21] .....	48
<b>FIGURE III-1 :</b> ANATOMIE DE SEIN. ....	52
<b>FIGURE III-2 :</b> EXEMPLES D'IMAGES D'ECHOGRAPHIE MAMMAIRE. A) LESION MALIGNE (SEIN DROIT) ET B) LESION BENIGNE. ....	53
<b>FIGURE III-3:</b> EXEMPLE D'IRM MAMMAIRE .....	53
<b>FIGURE III-4 :</b> LES COMPOSANTS D'UN MAMMOGRAPHE. ....	54
<b>FIGURE III-5:</b> LES DIFFERENTES FORMES D'UNE MASSE.....	55
<b>FIGURE III-6:</b> LES DIFFERENTS TYPES DE CONTOURS D'UNE MASSE. [CHE12] .....	56
<b>FIGURE III-7 :</b> DIFFERENTS TYPES DE DENSITE MAMMAIRE.....	57
<b>FIGURE III-8 :</b> ASPECT TYPIQUE DE FIBROADENOME PARTIELLEMENT CALCIFIE : BI-RADS 2.....	60
<b>FIGURE III-9 :</b> MASSES BIEN CIRCONSCRITES DISCRETEMENT LOBULEES, SANS MICROLOBULATION, NON CALCIFIEES, NON TYPIQUEMENT LIQUIDIENNES EN ECHOGRAPHIE : BIRADS 3. ....	60
<b>FIGURE III-10 :</b> IMAGES SPICULEE SANS CENTRE DENSE. ....	61
<b>FIGURE III-11 :</b> MICROCALCIFICATIONS GRANULAIRES IRREGULIERES PEU NOMBREUSES : BIRADS 4.....	61
<b>FIGURE IV-1 :</b> DISTRIBUTION DES CLASSES DE DENSITE DANS LA BASE DE DONNEES RSNA. ....	72
<b>FIGURE IV-2 :</b> L'ARCHITECTURE GENERALE DE TRAVAIL. ....	73
<b>FIGURE IV-3 :</b> TRAITEMENT DES MAMMOGRAPHIES NEGATIVES.....	74
<b>FIGURE IV-4 :</b> EXEMPLE DE SUPPRESSION DES ANNOTATIONS. (A : IMAGE ORIGINAL AVEC ETIQUETTE, B : IMAGE BINAIRE AVEC ETIQUETTE, C : IMAGE AVEC MASK , D : IMAGE PRETRAITE). ....	75
<b>FIGURE IV-5 :</b> EXEMPLE DE REDIMENSIONNEMENT.....	75
<b>FIGURE IV-6 :</b> EXEMPLE DES 3 TECHNIQUES D'AUGMENTATIONS APPLIQUEES SUR UNE MAMMOGRAPHIE DE CLASSE B. ....	77
<b>FIGURE IV-7 :</b> MATRICE DE CONFUSION DU TRANSFERT D'APPRENTISSAGE DU MODELE VGG 16.....	78
<b>FIGURE IV-8 :</b> MATRICE DE CONFUSION DU TRANSFERT D'APPRENTISSAGE DU MODELE RESNET 50.....	78
<b>FIGURE IV-9 :</b> MATRICE DE CONFUSION DU TRANSFERT D'APPRENTISSAGE DU MODELE EFFICIENTNETB0. ....	79
<b>FIGURE IV-10 :</b> MATRICE DE CONFUSION DU TRANSFERT D'APPRENTISSAGE DU MODELE MOBILENET V2. ....	79
<b>FIGURE IV-11 :</b> EVOLUTION DES TAUX D'APPRENTISSAGE ET ERREURS DE VGG 16. ....	81
<b>FIGURE IV-12 :</b> EVOLUTION DES TAUX D'APPRENTISSAGE ET ERREURS DE RESNET 50. ....	81
<b>FIGURE IV-13 :</b> EVOLUTION DES TAUX D'APPRENTISSAGE ET ERREURS D'EFFICIENTNETB0. ....	82
<b>FIGURE IV-14 :</b> EVOLUTION DES TAUX D'APPRENTISSAGE ET ERREURS DE MOBILENET V2. ....	82
<b>FIGURE IV-15 :</b> INTERFACE GRAPHIQUE. ....	83
<b>FIGURE IV-16 :</b> INTERFACE GRAPHIQUE (RESULTATS DE LA CLASSIFICATION D'UNE IMAGE DE LA CLASSE B). ....	83

# Liste des tableaux

---

<b>TABLEAU III-1</b> : CLASSIFICATION DE LE GAL. [CHE12] .....	58
<b>TABLEAU III-2</b> : MATRICE DE CONFUSION. [MON03] .....	64
<b>TABLEAU IV-1</b> : RECAPITULATIF DES DONNEES D'ENTRAINEMENT, VALIDATION ET DE TEST. ....	76
<b>TABLEAU IV-2</b> : METRIQUES D'EVALUATION SUPPLEMENTAIRES. ....	80
<b>TABLEAU IV-3</b> : LES RESULTATS OBTENUS.....	81

# Abréviations et Acronymes

ML	Machine Learning
SVM	Support Vector Machine
MCO	Moindre carré ordinaires
RNA	Réseau Neuronal Artificiel
KNN	K-Nearest Neighbour
RN	Réseau Neurones
Tanh	Tangente Hyperbolique
ReLu	Rectified Linear Unit
LReLu	Leaky Rectified Linear Unit
Elus	Exponential Linear Unit
CE	Cross Entropy
MSE	Mean Squared Error
MAE	Mean Absolute Error
SGD	Stochastic Gradient Decent
GPU	Graphics processing Unit
DL	Deep Learning
RNN	Recurrent Neural Network
PNL	Programmation Neuro-Linguistique
LSTM	Long Short Term Memory
CNN	Connvolutional Neural Network
MLP	Multi Layer Perceptron
FC	Fully Connected
RVB	Rouge , Vert , Blue
GAP	Global Average Pooling
VGG	Visual Geometry Group
TA	Transfer D'Apprentissage
IRM	Imagerie Par Résonance Magrétique
Mcs	Micro Calcifications
ACR	American College Of Radiology
BI-RADS	Breast Imaging-Reporting And Data System
DAO	Diagnostic Assistée par Ordinateur
VP	Vrai Positif
VN	Vrai Négatif
FP	Faux Positif

FN	Faux Négatif
cGAN	Conditional Generative Adversarial Network

# INTRODUCTION GENERALE

---

Selon l'organisation mondiale de la santé, le cancer du sein est le cancer le plus répandu dans le monde. En 2020 il y a eu 2,3 millions de nouveaux diagnostics de cancer du sein et 685000 décès malgré la mortalité par cancer du sein dans les pays développés ont été chutés de 40% depuis les années 1980.

Actuellement, la détection précoce du cancer du sein nécessite une expertise humaine hautement qualifiée, ce qui rend la procédure de dépistage coûteux à mener. Une pénurie imminente de radiologues dans plusieurs pays aggravera probablement ce problème. Le dépistage par mammographie entraîne également une incidence élevée de résultats faussement positifs. Cela peut entraîner une anxiété inutile, des tests d'imagerie supplémentaires et parfois la nécessité d'un prélèvement de tissus (souvent une biopsie à l'aiguille).

Il existe de nombreuses modalités d'acquisition d'images médicales, et la quantité d'images traitées dans le domaine médical ne cesse d'augmenter. Parmi ces modalités, la radiographie par les rayons X occupe une place prépondérante. Son fonctionnement repose sur l'impression des densités réfléchies d'un organe sur une radiographie. Le degré de noircissement de la radiographie varie selon l'organe examiné. La mammographie numérique, quant à elle, est un type spécifique d'examen radiographique utilisé pour le dépistage du cancer du sein. Cependant, la mammographie demeure la modalité privilégiée pour le diagnostic de ce type de cancer, où la densité des mammographies joue un rôle important pour la classification des anomalies mammaires.

Le domaine qui nous intéresse dans notre projet de fin d'étude est celui de l'analyse des images mammographiques qui représente aujourd'hui un outil efficace d'aide au diagnostic du cancer du sein. Les techniques d'apprentissage automatique pourraient aider à rationaliser le processus utilisé par les radiologues pour évaluer les mammographies de dépistage. Lors d'un examen mammographique, le radiologue ne fait que résoudre un problème de vision ; c'est à partir de ce point et de récentes techniques d'analyse et prétraitement d'image, que l'idée de réaliser un système d'aide au diagnostic devient réalisable. La recherche menée dans ce domaine actuellement est encore à un stade peu avancé.

L'analyse des mammographies numériques à travers les techniques de l'intelligence artificielle permet de fournir aux médecins des informations quantitatives et fiables, ces informations complètent l'aspect visuel et qualitatif de l'examen. Le but de notre projet de fin d'étude est de proposer un système d'aide au diagnostic pour différencier entre les deux classes de densité mammographique selon la classification BI-RADS (Breast Imaging-Reporting And Data System).

Ce mémoire de fin d'étude est organisé en quatre chapitres comme suit :

Dans le premier chapitre, nous commençons par une analyse approfondie de l'intelligence artificielle et de ses orientations de recherche. Nous accordons une attention particulière aux différents types d'apprentissage automatique, tels que les approches supervisées, non supervisées et par renforcement.

Dans le deuxième chapitre, nous approfondissons les concepts fondamentaux des réseaux de neurones artificiels et des réseaux de neurones convolutifs. L'objectif principal de ce chapitre est de renforcer notre compréhension des technologies les plus récentes utilisées dans le domaine de la classification automatique.

Dans le troisième chapitre, nous entreprenons une revue de littérature complète des systèmes de détection et de diagnostic automatique des mammographies numériques. L'objectif de cette revue est d'examiner en détail les différentes approches et les outils utilisés pour classer les images mammographiques en fonction de leurs densités.

Dans le quatrième et dernier chapitre, nous présentons l'architecture générale de notre travail et les expérimentations effectuées en utilisant la base de données RSNA (Radiological Society of North America) et à travers le transfert d'apprentissage en se basant sur quatre modèles populaires: VGG16, ResNet50, MobileNetV2 et EfficientNetB0. L'approche proposée est basée sur une base de 24 350 images mammographiques.

Enfin, dans la conclusion générale, nous présentons une synthèse de notre travail et exposons les perspectives qui en découlent.

Chapitre

---

**1**

**INTELLIGENCE & VISION  
ARTIFICIELLE**

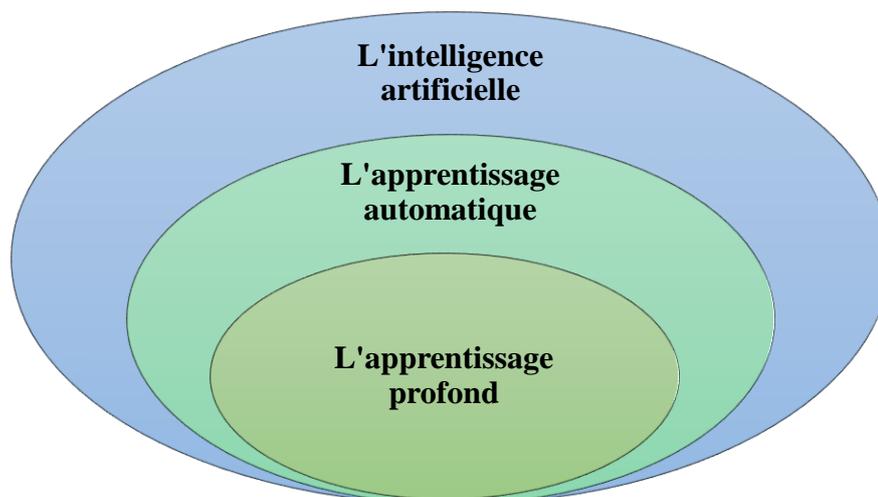
## I.1. Introduction

L'intelligence artificielle a progressé à pas de géant au cours de la dernière décennie, en particulier grâce aux techniques d'apprentissage automatique (Machine Learning ML). L'apprentissage automatique est largement utilisé dans la recherche et a été intégré à diverses applications telles que l'exploration de texte, la détection de spam, la recommandation de vidéos, la classification d'images et la recherche de concepts multimédias.

Nous présentons dans ce chapitre les concepts fondamentales pour comprendre les techniques de l'intelligence artificielle de l'apprentissage automatique.

## I.2. Apprentissage automatique

Le domaine de ML a diverses définitions formelles dans la littérature. Dans son livre fondateur, Arthur Samuel [SAM59] définit l'apprentissage automatique comme "le domaine d'étude qui permet aux ordinateurs d'apprendre sans être explicitement programmés". Dans le dictionnaire « Computer Science » de Tom Mitchell [MIT97] le décrit comme suit : "On dit qu'un programme informatique apprend de l'expérience (E) sur une classe de tâches (T) et une mesure de performance (P) si sa performance sur des tâches dans T, comme mesuré par P, s'améliore avec l'expérience E". Ethem Alpaydin [ALP14] dans son manuel définit l'apprentissage automatique comme le domaine de " la programmation d'ordinateurs pour optimiser les critères de performance à l'aide de données d'exemple ou d'expériences passées". Ces différentes définitions partagent le concept d'apprentissage de l'environnement environnant à travers des exemples répétés, demandant aux ordinateurs d'effectuer intelligemment des tâches au-delà du traitement traditionnel des chiffres. [NAQ15]



**Figure I-1:** Les relations entre l'intelligence artificielle, l'apprentissage automatique et l'apprentissage profond.

### I.3. Les approches d'apprentissage

En général, les méthodes d'apprentissage automatique sont divisées en trois catégories comme suit :

#### I.3.1. Apprentissage supervisé

Dans ce type d'apprentissage, le système reçoit un échantillon d'entrées et est mappé sur la sortie souhaitée. Chaque exemple de ce type d'apprentissage consiste en une paire qui comprend un élément d'entrée (essentiellement un vecteur) et une valeur de sortie souhaitée (signal de supervision). Les données de formation sont examinées et étudiées par un algorithme d'apprentissage supervisé, qui génère ensuite une fonction inférée à utiliser pour cartographier de nouveaux cas. La situation idéale permettra à l'algorithme d'identifier avec précision les étiquettes de classe pour les instances qui ne sont pas encore visibles. L'algorithme d'apprentissage nécessite une généralisation "raisonnable" des données d'apprentissage à des conditions inexplorées. [SOM16]

En apprentissage supervisé, il existe de nombreuses méthodes :

##### I.3.1.1. Machines à vecteurs de support (SVM)

Les machines à vecteur de support (SVM) sont étroitement liées aux réseaux neuronaux traditionnels, tels que le perceptron multicouche. Les SVM reposent sur la notion de marge, qui représente la distance entre un hyperplan séparant deux classes de données. Il a été démontré que maximiser cette marge, c'est-à-dire créer la plus grande distance possible entre l'hyperplan de séparation et les instances de chaque côté, permet de réduire l'erreur de généralisation attendue [OSI17].

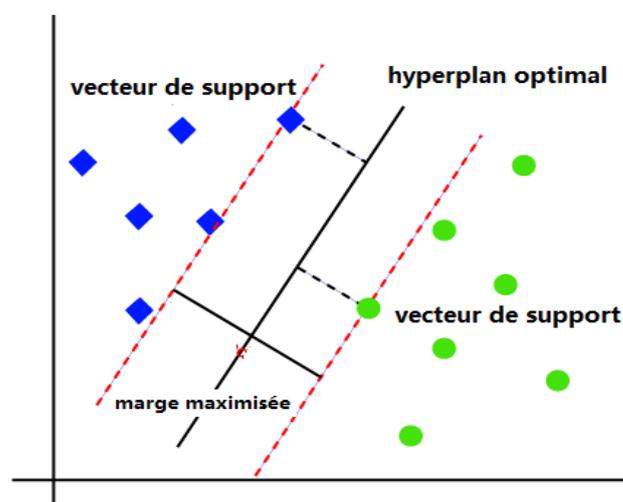


Figure I-2: Machines à vecteur de support (SVM).

### I.3.1.2. Arbres de décision

Les arbres de décision sont des structures hiérarchiques qui classent les instances en les organisant en fonction de la valeur de leurs caractéristiques. Chaque nœud d'un arbre de décision correspond à une caractéristique de l'instance à classer et chaque branche représente une valeur possible pour ce nœud. Les instances sont classées en parcourant l'arbre à partir du nœud racine et en les triant en fonction des valeurs de leurs caractéristiques. L'apprentissage par arbre de décision est une technique largement utilisée dans l'exploration de données et l'apprentissage automatique. Elle utilise un arbre de décision comme modèle prédictif qui associe les observations sur un élément à des conclusions sur la valeur cible de l'élément. Ces modèles d'arbres peuvent également être appelés arbres de classification ou arbres de régression.

Dans les classificateurs d'arbres de décision, les techniques de post-élagage sont couramment utilisées pour évaluer les performances des arbres de décision. Ces techniques consistent à élaguer l'arbre de décision sur la base d'un ensemble de validation. En supprimant un nœud, on peut lui attribuer la classe la plus fréquente parmi les instances d'apprentissage associées à ce nœud [OSI17].

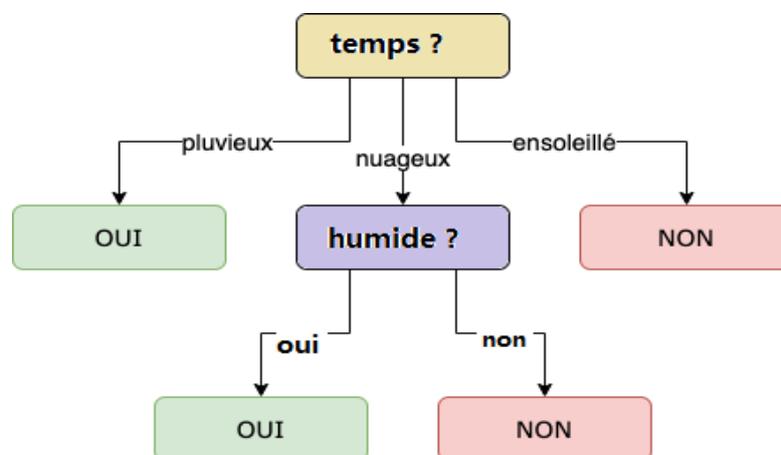
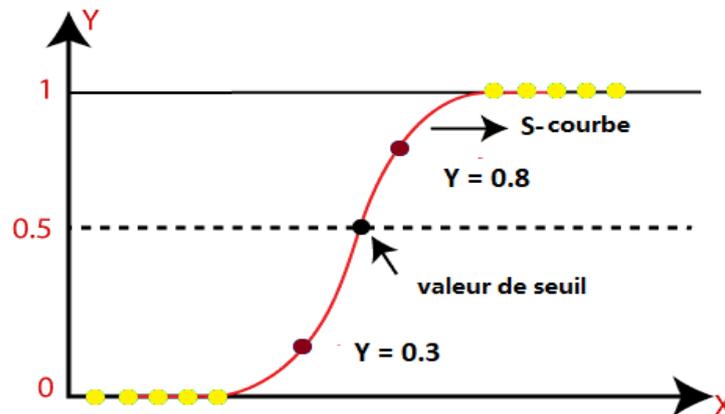


Figure I-3: Exemple d'arbre de décision.

### I.3.1.3. Régression logistique

Cette fonction est une méthode de classification qui construit un modèle basé sur les classes à l'aide d'un seul estimateur de régression logistique multinomiale. La régression logistique détermine généralement la frontière entre les classes et calcule les probabilités de classe en fonction de la distance par rapport à cette frontière d'une manière spécifique. Lorsqu'elle traite des ensembles de données plus importants, la régression logistique a tendance à s'approcher plus rapidement des extrêmes (0 et 1). Ces déclarations basées sur les

probabilités différencient la régression logistique d'un simple classificateur, car elle fournit des prédictions plus fortes et plus détaillées. Cependant, il est important de noter que ces prédictions fortes peuvent toujours être incorrectes. La régression logistique, tout comme la régression par les moindres carrés ordinaires (MCO), est une approche prédictive. Cependant, la régression logistique produit un résultat dichotomique pour les prédictions. La régression logistique est largement utilisée dans les statistiques appliquées et l'analyse des données discrètes. Elle peut être considérée comme une forme d'interpolation linéaire. [OSI17]

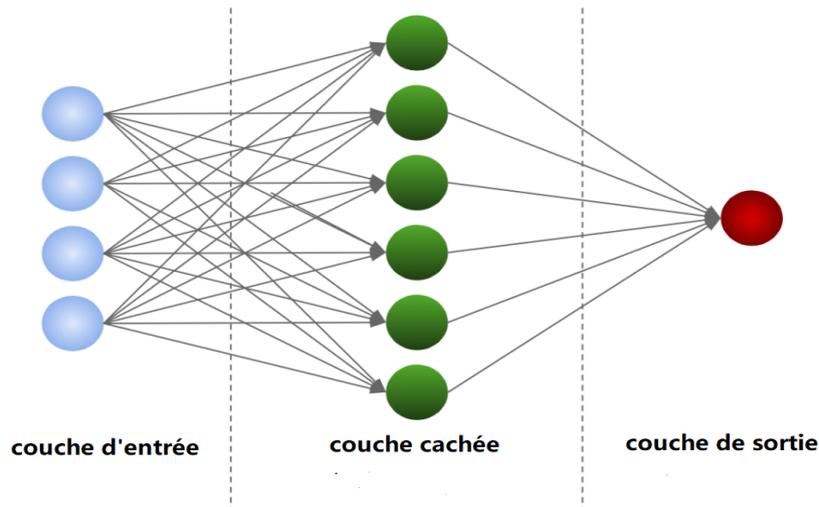


**Figure I-4:** Régression logistique.

#### **I.3.1.4. Les réseaux neuronaux**

Qui, bien que chaque réseau n'en exécute généralement qu'un seul, peuvent en fait effectuer simultanément un certain nombre de tâches de régression et/ou de classification. Par conséquent, le réseau aura généralement une seule variable de sortie, bien que dans le cas de problèmes de classification à plusieurs états, cela puisse correspondre à un certain nombre d'unités de sortie. Les fonctions d'entrée et d'activation de l'unité, la conception du réseau et le poids de chaque lien d'entrée sont les trois principaux composants qui déterminent la performance d'un réseau neuronal artificiel (RNA).

Le comportement du RNA est déterminé par les valeurs actuelles des poids, car les deux premiers aspects sont fixes. Les instances de l'ensemble de formation sont régulièrement présentées au réseau après que les poids du réseau à former ont été réglés sur des valeurs aléatoires. Les valeurs d'entrée d'une instance sont placées sur les unités d'entrée et la sortie du réseau est comparée à la sortie souhaitée de l'instance. Les valeurs de sortie du réseau sont alors légèrement modifiées de manière à les rapprocher des valeurs de sortie souhaitées. Un réseau peut être enseigné à l'aide d'une variété d'algorithmes [OSI17].



**Figure I-5:** exemple d'un réseau neuronal.

### **I.3.2. Apprentissage non-supervisé**

Quelques entrées d'exemple sont données au système pendant cette forme d'apprentissage, mais il n'y a pas de sortie d'accompagnement. Puisqu'il n'y a pas de résultat souhaité dans ce domaine, une catégorisation est effectuée pour s'assurer que l'algorithme distingue correctement les ensembles de données. Le défi consiste à créer une fonction capable d'extraire une structure cachée à partir de données non étiquetées. Étant donné que l'apprenant reçoit des échantillons ou des ensembles de formation non étiquetés, il n'y a pas d'erreur pour récompenser le signal d'envisager une solution potentielle. L'apprentissage non supervisé est distinct de l'apprentissage supervisé et de l'apprentissage par renforcement à cet égard. Il est étroitement lié au problème statistique et d'estimation de la densité [SOM16].

#### **I.3.2.1. K-Means**

L'un des algorithmes d'apprentissage non supervisé les plus simples pour résoudre le problème bien connu du regroupement est le K-Means, qui utilise un nombre prédéterminé de grappes (k grappes) pour classer un ensemble de données donné de manière simple et directe. Lorsque des données étiquetées ne sont pas disponibles, la technique K-Means est utilisée .une technique générale permettant de transformer des règles empiriques imprécises en règles de prédiction extrêmement précises. Un algorithme d'amplification peut construire un classificateur unique avec une précision très élevée, par exemple 99%, étant donné un algorithme d'apprentissage "faible" qui peut constamment trouver des classificateurs au moins marginalement meilleurs que le hasard, par exemple une précision de 55%, et des données suffisantes.[OSI17]

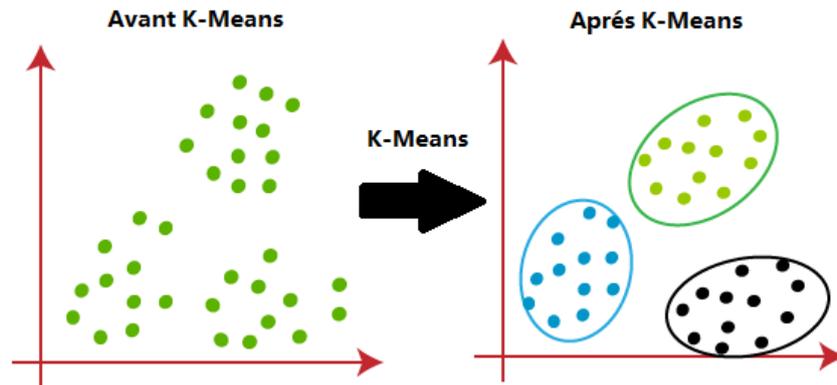


Figure I-6: Exemple de K-Means.

### I.3.2.2. Plus proche voisin (K-NN)

L'algorithme du plus proche voisin est une technique non paramétrique de classification et de régression dans le domaine de la reconnaissance des formes. Les  $K$  instances de formation les plus proches dans l'espace des caractéristiques constituent l'entrée dans les deux situations. L'algorithme K-NN est un exemple d'apprentissage basé sur les instances. Le résultat de la classification KNN est l'appartenance à une classe. La majorité des voisins choisissent de classer. La classe a un seul voisin le plus proche si  $K = 1$ . Dans une méthode de pondération classique, où  $d$  est la distance au voisin, chaque voisin se voit attribuer un poids de  $1/d$ . La distance euclidienne est la distance la plus courte qui puisse exister entre deux voisins et est toujours une ligne droite. [SIN15]

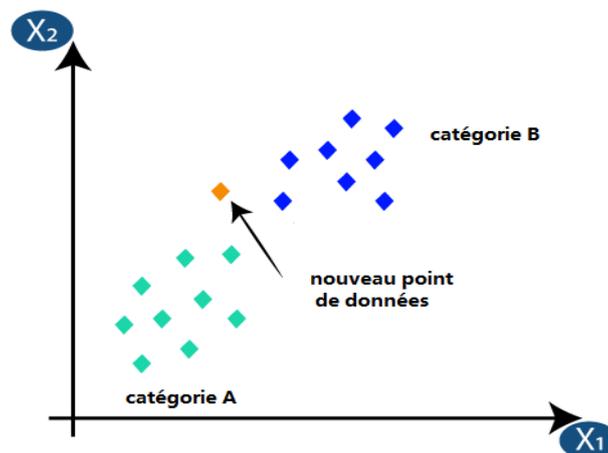


Figure I-7: Présentation de la méthode K-NN.

L'inconvénient de l'algorithme K-NN est qu'il est sensible à l'organisation locale des données. L'extraction de caractéristiques consiste à transformer les données d'entrée en une collection de caractéristiques. Avant d'utiliser la méthode K-NN dans l'espace des caractéristiques, les données brutes sont extraites. [SIN15]

### I.3.3. Apprentissage par renforcement

Une branche de l'apprentissage automatique appelée apprentissage par renforcement, qui a été influencée par la psychologie comportementale, se concentre sur la façon dont les agents logiciels doivent réagir dans un environnement donné pour maximiser une sorte de récompense cumulative. De nombreuses théories, y compris la théorie des jeux, la théorie du contrôle, la recherche opérationnelle, la théorie de l'information, l'intelligence en essaim, la statique et les algorithmes génétiques sont étudiés et en font usage [SOM16].

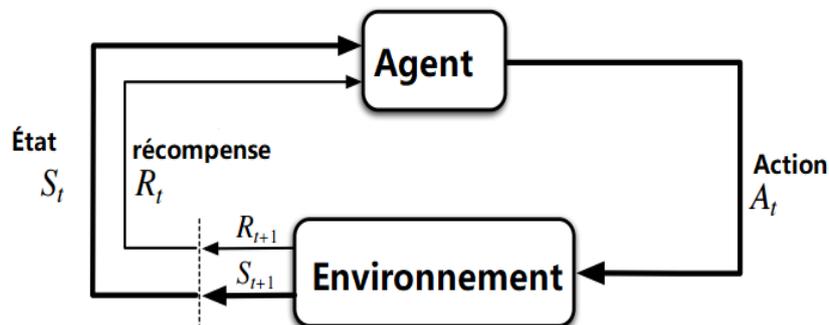


Figure I-8: Apprentissage par renforcement.

### I.4. Domaines d'utilisation

L'apprentissage automatique est utilisé dans un certain nombre de domaines, tels que [AHU20] :

- Les systèmes de reconnaissance vocale, tels que SPHINX et Siri, permettant des techniques spécifiques au locuteur pour identifier les sons et les mots dans différentes langues.
- Les véhicules autonomes pour la détection et la navigation en temps réel.
- La reconnaissance faciale pour le marquage automatique des amis sur Facebook et la publicité personnalisée.
- Les systèmes de recommandation pour des plateformes telles que Amazon, Netflix et Audible.
- Soins de santé.
- Finance pour l'analyse des investissements, la prévention des fraudes et la distribution des cartes de crédit.

## **I.5. Conclusion**

L'intelligence Artificielle est un domaine de recherche qui s'évolue rapidement, où la classification des images constitue l'outil principal de la vision par ordinateur.

Les méthodes de classification dépendent de l'objectif visé (description globale ou locale) et du type de l'objet à analyser (caractère, son, visage, image médicale ...etc.). Dans ce chapitre nous avons présenté les concepts fondamentaux de l'intelligence artificielle. Dans le chapitre suivant, nous présenterons en détail les approches basées sur les réseaux de neurones artificiels (ANN) et l'apprentissage profond (Deep Learning DL), afin de les utilisées par la suite dans les expérimentations de ce projet pour la classification des densités mammographiques.

# Chapitre

---

# 2

# Les réseaux de neurones et Apprentissage profond

## II.1. Introduction

Les réseaux de neurones artificiels et à convolution sont largement utilisés comme modèle pour l'apprentissage automatique, ils ont une longue histoire de développement et de nombreuses réalisations issues de travaux de recherche. Il existe de nombreuses références disponibles qui se concentrent uniquement sur les réseaux de neurones artificiels. Parallèlement au regain d'intérêt croissant pour les réseaux de neurones à convolution.

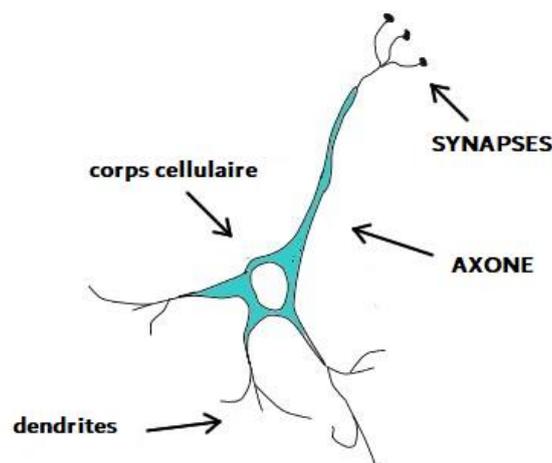
Nous présentons dans ce chapitre les concepts fondamentales des réseaux de neurones artificiels et à convolution.

## II.2. Les réseaux de neurones.

La raison pour laquelle on parle de réseau de neurones (RN) est qu'il est composé d'éléments interconnectés. L'étude du système nerveux biologique a eu un impact sur ces composants. Un réseau de neurones est une tentative de construire une machine fonctionnelle dans notre environnement, pour le dire autrement. Semblable au cerveau humain en ce sens que nous construisons ces robots à partir de composants qui fonctionnent de manière similaire aux neurones biologiques. [PIC94]

### II.2.1. Les neurones biologiques

Le nombre de cellules de soutien dans le système nerveux central se chiffre en centaines de milliards. Chaque neurone reçoit les impulsions nerveuses entrantes par ses dendrites, les combine avec les impulsions entrantes et les transmet aux neurones voisins via la cellule nerveuse. Il a un biais axial, comme on le voit sur la **Figure II-1**. [SOU20]



**Figure II-1:** Neurone biologique.

La membrane qui sépare l'intérieur du neurone de l'extérieur détermine la forme de cette cellule vivante. Cette cellule peut prendre différentes formes, notamment pyramidale,

sphérique ou allongée. Ces neurones sont considérés comme les blocs de construction qui constituent l'unité primaire de traitement cérébral. [SOU20]

Un neurone biologique se compose de [SOU20] :

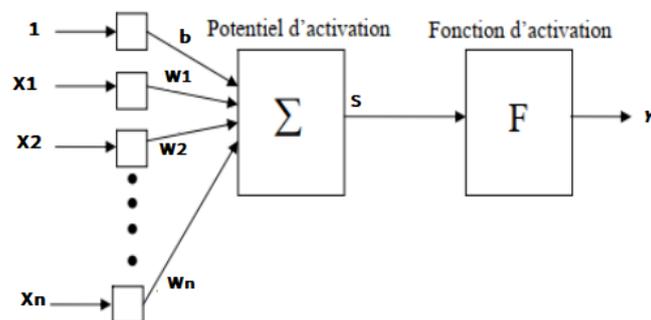
**Axone** : Le transfert des signaux neuronaux vers d'autres cellules est une extension unique. Par conséquent, il est démontré comment il interagit avec d'autres neurones. Il existe de nombreuses autres cellules avec lesquelles il peut être partagé.

**Corps cellulaire** : Il est constitué d'un nœud cellulaire qui effectue les changements biochimiques nécessaires à la vie neuronale, et se divise en dendrites qui forment des connexions avec les autres cellules.

**Synapses** : Un élément de communication appelé synapse maintient la connexion entre les membranes des neurones voisins et le cytoplasme d'un neurone. Elle est essentielle au processus de transduction du signal.

### II.2.2. Le Neurone formel

L'idée fondamentale du fonctionnement des neurones biologiques est traduite par la représentation mathématique d'un neurone artificiel dans la Figure II-2. Reconnaisant que, d'un point de vue biologique, toutes les synapses n'ont pas la même "valeur" (Les connexions entre les neurones sont plus ou moins fortes.) Pour cette raison, les chercheurs ont développé un algorithme qui divise le nombre total d'entrées par les poids synaptiques (facteurs de pondération). En général, un neurone formel est un composant de traitement doté d'une ou plusieurs sorties et de  $n$  connexions entrantes (ou connexions à des neurones externes) de  $x_1$ ,  $x_2$ , ...,  $x_i$ , ..., et  $x_n$ . La méthode utilisée pour le traiter consiste à déclencher la fonction d'activation, également connue sous le nom de fonction de somme des carrés, afin de produire le résultat souhaité. [SOU20]

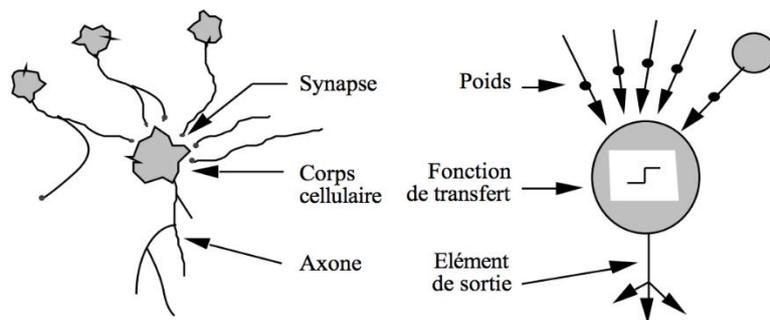


**Figure II-2:** neurone formel.

( $x_i$ : les entrées,  $S$  est le potentiel d'activation,  $w_i$  représentent les poids,  $y_i$  est la sortie du réseau).

### II.2.3. Réseaux de Neurones Artificiels (RNA)

Un réseau de neurones est un système opérateur connecté de manière non linéaire qui reçoit des signaux du monde extérieur par des entrées et les transmet. Ces métaphores (RN) décrivent la structure du cerveau, le fonctionnement du traitement parallèle et de la distribution de l'information, ainsi que l'exécution de ces fonctions. Les nombreuses unités de traitement présentes dans le cerveau sont appelées neurones. Chaque neurone fonctionne indépendamment, ce qui en fait un système parallèle à forte connectivité. Un réseau interconnecté de neurones distribue les informations dont il dispose entre ses membres ; il ne stocke jamais d'informations sous forme de symboles dans une quelconque partie de la mémoire. Aucun réseau de neurones n'a jamais été programmé pour effectuer telle ou telle tâche. Il s'implique et se développe en un mécanisme d'apprentissage qui interagit avec les composants du réseau pour réaliser au mieux la tâche souhaitée. Les traits des neurones naturels et synthétiques, tels qu'ils sont représentés dans la Figure II-3. [SOU20]

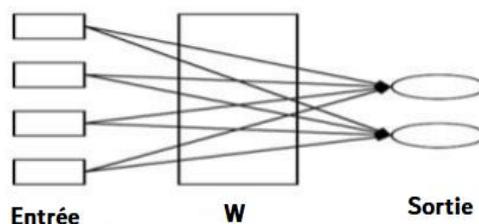


**Figure II-3:** Neurone biologique et artificiel.

Le nombre de cellules dans un réseau influence plusieurs facteurs, notamment la structure des connexions entre les neurones. De nombreux types d'architecture peuvent être répertoriés:

#### II.2.3.1. Réseau monocouche

L'organisation d'un réseau monocouche ou perceptron est telle que les neurones organisés en entrées sont complètement connectés aux autres neurones organisés en sorties via des couches à poids variable. [SOU20]



**Figure II-4:** Réseau monocouche.

### II.2.3.2. Réseau multicouche

Les neurones sont disposés en couches, les neurones de la première couche étant connectés au monde extérieur et recevant des vecteurs d'entrée, tandis que les neurones de la deuxième couche sont en fait les sorties de leurs prédécesseurs. Un réseau de neurones peut avoir une ou plusieurs sorties.

Dans un réseau multicouche, les connexions entre les neurones d'une même couche ne sont pas établies ; au contraire, chaque neurone de la couche supérieure est connecté à chaque neurone de la couche inférieure. [SOU20]

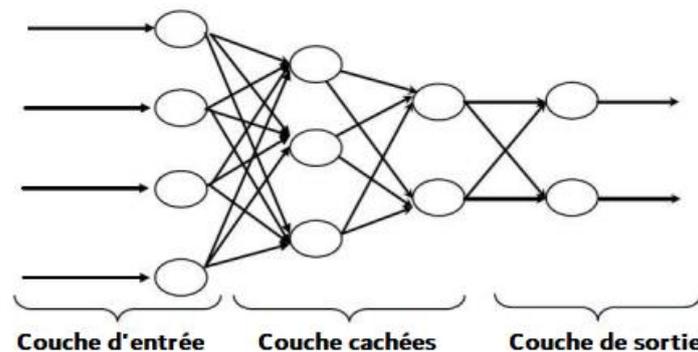


Figure II-5: Réseau multicouche.

**Couche d'entrée :** Contenant tous les neurones d'entrée. Cette couche est passive sans aucun traitement effectué par ses neurones.

**Couche de sortie :** il contient tous les neurones de sortie.

**Couche cachée :** une couche intermédiaire cachée et sans contact avec l'extérieur. [SOU20]

### II.2.4. Les fonctions d'activations

Les capacités d'activation sont des capacités utilisées dans les réseaux de neurones pour calculer la somme pondérée des entrées et des biais, qui sont utilisées pour déterminer si un neurone peut être déclenché ou non. Il manipule les statistiques fournies à travers quelques Gradients de traitement normalement la descente de gradient et produit ensuite une sortie pour le réseau de neurones, qui se compose des paramètres à l'intérieur des statistiques. Ces fonctions d'activations sont souvent appelées une fonction de commutation dans un peu de littérature. La fonctionnalité d'activation peut être à la fois linéaire ou non linéaire en fonction de la fonctionnalité qu'elle représente, et est utilisée pour manipuler les sorties de nos réseaux de neurones [NWA18]. Voici les principales fonctions d'activations que l'on peut trouver dans le domaine des réseaux de neurones:

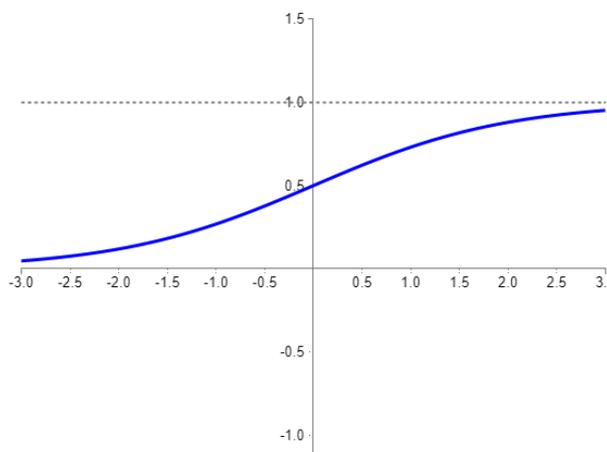
### II.2.4.1. La fonction sigmoïde

La fonction logistique ou fonction d'écrasement sont d'autres noms de la fonction sigmoïde qui ont été utilisés dans certaines littératures. Les trois versions de la fonction d'activation sigmoïdes utilisées dans les applications DL sont les résultats de l'étude sur la fonction sigmoïde. Une fonction d'activation non linéaire principalement utilisé dans les réseaux de neurones à anticipation est le sigmoïde. C'est une fonction réelle différentiable bornée, définie pour des valeurs d'entrée réelles, avec des dérivées positives partout et une certaine régularité. Donnée par la relation :

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad \text{Équation II-1}$$

La fonction sigmoïde est utilisée dans les couches de sortie des architectures DL et est utilisée pour prévoir la sortie basée sur la probabilité. Il a été utilisé efficacement dans des tâches de classification binaire, des tâches de modélisation de régression logistique et d'autres domaines de réseaux neuronaux. Selon Neal, les principaux avantages de la fonction sigmoïde sont sa simplicité de compréhension et son utilisation principalement dans les réseaux peu profonds. De plus, lors du démarrage du réseau de neurones avec des poids aléatoires modestes, Glorot et Bengio [GLO10] proposent d'éviter la fonction d'activation sigmoïde.

Cependant, la fonction d'activation sigmoïde présente des problèmes importants, notamment des gradients d'humidité nets lors de la rétropropagation des couches cachées plus profondes vers les couches d'entrée, une saturation du gradient, une convergence lente et une sortie centrée non nulle, ce qui entraîne la propagation des mises à jour du gradient dans différentes directions. Pour remédier à certaines de ces lacunes de la fonction sigmoïde, d'autres formes, telles que la fonction tangente hyperbolique, ont été proposées. [NWA18]



**Figure II-6:** La fonction d'activation sigmoïde.

### II.2.4.2. La fonction tangente hyperbolique (Tanh)

Un autre type de fonctions d'activation utilisé est la fonction de tangente hyperbolique, qui existe également dans certaines versions. La fonction tanh a une plage centrée sur zéro plus lisse [1] entre -1 et 1. Par conséquent, la sortie de la fonction tanh est représentée par :

$$f(x) = \left( \frac{e^x - e^{-x}}{e^x + e^{-x}} \right) \quad \text{Équation II-2}$$

En raison de ses performances d'entraînement supérieures pour les réseaux de neurones multicouches par rapport à la fonction sigmoïde, la fonction tanh est devenue la fonction privilégiée. Le problème de fuite du gradient auquel les fonctions sigmoïdes étaient également confrontées n'a cependant pas pu être résolu par la fonction tanh. Le principal avantage de la fonction est qu'elle génère une sortie centrée sur zéro, ce qui facilite le processus de rétropropagation.

La fonction tanh a la particularité de ne pouvoir atteindre un gradient de 1 que lorsque la valeur d'entrée est 0 ou lorsque x est égal à zéro. En conséquence, le calcul de la fonction tanh entraîne la mort de certains neurones. Le poids d'activation est rarement utilisé dans le cas de neurones morts en raison d'un gradient nul. La fonction d'activation de l'unité linéaire rectifiée (Relu) a été créée à la suite de recherches supplémentaires sur les fonctions d'activation pour remédier à l'inconvénient de la fonction tanh. Pour les tâches impliquant la reconnaissance de la parole et le traitement du langage naturel, les réseaux de neurones récurrents ont principalement utilisé les fonctions tanh . [NWA18]

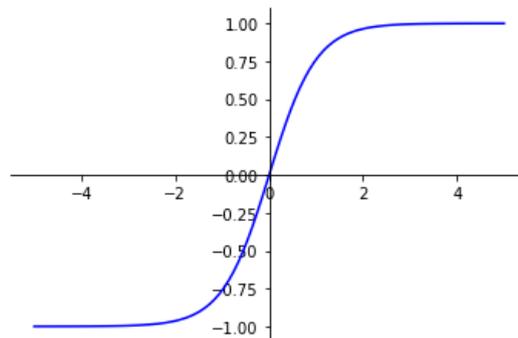


Figure II-7 : Représentation de la fonction Tanh.

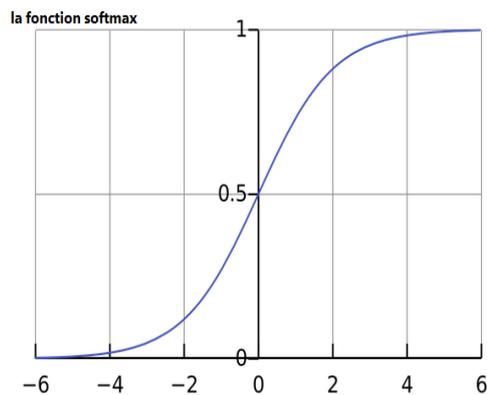
### II.2.4.3. La fonction Softmax

La fonction Soft max est l'un des nombreux types de fonctions d'activation utilisés dans l'informatique neuronale. A l'aide d'un vecteur de nombres réels, il calcule la distribution de probabilité. La sortie de la fonction Soft max est une plage de valeurs entre 0 et 1, où la somme des probabilités est égale à 1. La fonction Soft max est calculée en utilisant :

$$f(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_i)} \quad \text{Équation II-3}$$

Dans les modèles multi-classes, la fonction Soft max est utilisée pour renvoyer les probabilités pour chaque classe, la classe cible ayant la probabilité la plus élevée. La fonction Soft max est principalement utilisée dans presque toutes les couches de sortie des architectures d'apprentissage en profondeur.

La principale distinction entre le Soft max et le Sigmoid est que le Soft max est utilisé pour les tâches de classification multivariées, tandis que le Sigmoid est utilisé pour les tâches de classification binaire. [NWA18]

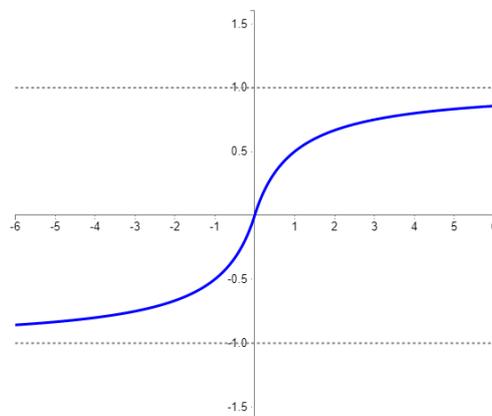


**Figure II-8:** La fonction softmax.

#### II.2.4.4. La fonction Softsign

Un autre type de fonction utilisé dans le calcul des réseaux neuronaux est le softsign. Une autre fonction non linéaire utilisée dans les applications DL est la fonction Softsign . Polynôme quadratique, la fonction Softsign est donnée par :

$$f(x) = \left( \frac{x}{|x| + 1} \right) \quad \text{Équation II-4}$$



**Figure II-9:** La fonction softsign.

La fonction Softsign converge sous forme polynomiale par opposition à la convergence exponentielle de la fonction tanh, qui est la principale distinction entre les deux fonctions.

Le Softsign a été principalement utilisé pour des problèmes impliquant le calcul de régression, mais a également été utilisé dans des tests basés sur DL de systèmes vocaux, les auteurs rapportant des résultats encourageants. [NWA18]

### II.2.4.5. La fonction de l'unité linéaire rectifiée (ReLU)

Depuis la proposition de Nair et Hinton en 2010 [NAI10], la fonction d'activation de l'unité linéaire rectifiée (Rectified Linear Unit Function ReLU) a été la fonction d'activation la plus souvent utilisée pour les applications d'apprentissage en profondeur avec les meilleurs résultats à ce jour. Le ReLU est une fonction à apprentissage plus rapide, s'est imposé comme la fonction la plus efficace et la plus souvent utilisée. Par rapport aux fonctions d'activation Sigmoid et tanh, il offre de meilleures performances et généralisation dans l'apprentissage en profondeur. ReLU conserve les caractéristiques des modèles linéaires qui les rendaient simples à optimiser à l'aide de méthodes de descente de gradient puisqu'il représente une fonction à peu près linéaire.

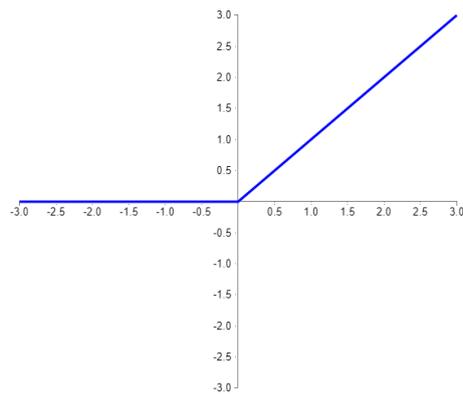
Chaque élément d'entrée est soumis à une opération de seuil par la fonction d'activation ReLU, qui fixe les valeurs inférieures à zéro à zéro. Il écrit comme suite [NWA18]:

$$F(x) = \max(0, x) = \begin{cases} x_i, & \text{si } x_i \geq 0 \\ 0 & \text{si } x < 0 \end{cases} \quad \text{Équation II-5}$$

En pilotant les valeurs d'entrée inférieures à zéro vers zéro, cette fonction rectifie les valeurs d'entrée inférieures à zéro et résout le problème de gradient de fuite qui était présent dans les types de fonctions d'activation antérieurs.

ReLU a été utilisé dans les unités cachées des réseaux de neurones profonds avec une autre fonction d'activation dans les couches de sortie du réseau. Les applications typiques incluent la catégorisation d'objets, et la reconnaissance vocale.

Le principal avantage de l'utilisation d'unités linéaires rectifiées dans les calculs est qu'elles garantissent un calcul plus rapide car elles ne calculent pas les divisions et les exponentielles, ce qui augmente la vitesse de calcul globale. Comme il comprime les valeurs entre zéro et maximum, le ReLU introduit également de la parcimonie dans les unités cachées. Bien que la stratégie de décrochage ait été mise en œuvre pour atténuer l'influence de la tendance des ReLU à sur-adapter par rapport à la fonction sigmoïde, les réseaux corrigés ont amélioré les performances des réseaux de neurones profonds. [NWA18]



**Figure II-10 :** La fonction d'activation Relu.

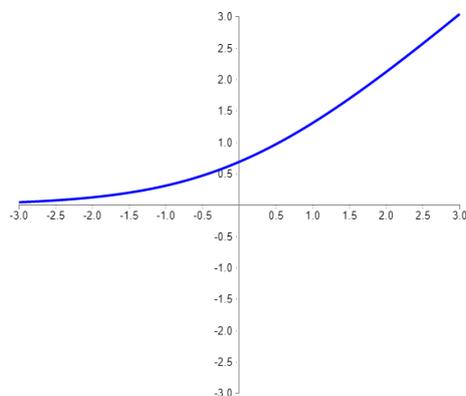
### II.2.4.6. La fonction Leaky ReLU (LReLU)

Afin de maintenir et de maintenir les mises à jour de poids tout au long de la phase de propagation, le ReLU qui fuit a été proposé en 2013 en tant que fonction d'activation. Le paramètre alpha a été développé pour remédier aux problèmes de neurones morts avec la fonction ReLU afin que les gradients ne soient jamais nuls pendant l'entraînement. Pour le gradient négatif, le LReLU calcule le gradient en utilisant une très petite valeur constante, souvent de l'ordre de 0,01. [NWA18] et donnée comme ça :

$$f(x) = ax + x = \begin{cases} x & \text{si } x > 0 \\ ax & \text{si } x \leq 0 \end{cases} \quad \text{Équation II-6}$$

### II.2.4.7. La fonction Softplus

Le Softplus est une variante fluide de la fonction ReLU qui a des qualités de gradient et de lissage non nulles, ce qui améliore les performances et la stabilisation des réseaux de neurones profonds créés avec les unités Softplus.



**Figure II-11:** La fonction d'activation softplus .

Une comparaison de la fonction Softplus avec les fonctions ReLU et Sigmoid par Zheng et al. [ZHE15] ont révélé des performances améliorées avec moins d'époques de convergence

pendant l'entraînement, en utilisant la fonction Softplus. Et la fonction Softplus est une primitive de la fonction sigmoïde. Il donner par la relation [NWA18]:

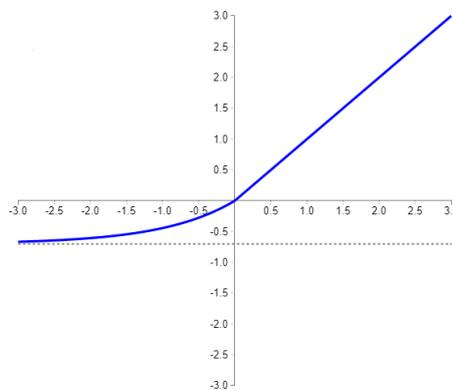
$$f(x) = \log(1 + \exp^x) \quad \text{Équation II-7}$$

Les fonctions ReLU et sigmoïde ont été les fonctions prédominants dans les systèmes de reconnaissance vocale qui utilisent le Softplus pour réaliser la reconnaissance vocale automatique. [NWA18]

### II.2.4.8. La fonction unités linéaires exponentielles (ELUs)

Une autre sorte de fonction suggérée par Clevert et al. [CLE15], les unités linéaires exponentielles (Exponential Linear Units **ELUs**), est utilisé pour accélérer la formation des réseaux de neurones profonds. L'avantage fondamental des ELU est qu'ils peuvent améliorer les fonctionnalités d'apprentissage et résoudre le problème du gradient de fuite en utilisant l'identité pour les valeurs positives. Ils ont des valeurs négatives, ce qui permet de rapprocher l'activation unitaire moyenne de zéro, de diminuer la complexité de calcul et d'accélérer l'apprentissage. L'ELU est un bon substitut au ReLU car il réduit les décalages de biais en entraînant l'activation moyenne à zéro. ELU donner par la relation suivante [NWA18] :

$$f(x) = \begin{pmatrix} x, & \text{si } x > 0 \\ \alpha \exp(x) - 1, & \text{si } x \leq 0 \end{pmatrix} \quad \text{Équation II-8}$$



**Figure II-12:** La fonction d'activation ELU.

Les ELU ont un plateau de saturation distinct dans leur régime négatif, ce qui conduit à l'apprentissage de représentations plus robustes. Ils apprennent également plus rapidement et ont une meilleure généralisation que ReLU et LReLU, en particulier au-dessus de cinq couches, et ils garantissent des résultats de pointe par rapport aux variantes ReLU. L'ELU paramétrique a été suggérée comme solution à ce problème car l'ELU a un grave défaut en ce sens qu'elle ne centre pas les valeurs à zéro. [NWA18]

### II.2.4.9. La fonction Maxout

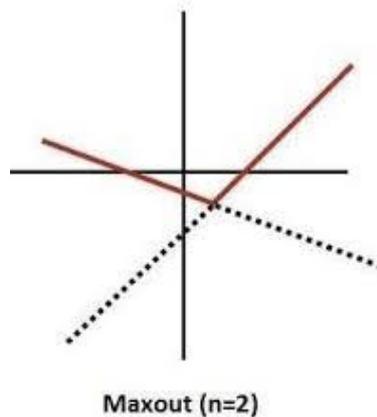
Le Maxout est une fonction qui utilise la non-linéarité comme produit scalaire des poids des données et du réseau neuronal. Le Maxout est une généralisation de ReLU et ReLU qui fuient où le neurone hérite des traits de ReLU et de ReLU qui fuit lorsqu'il n'y a pas de neurones morts ou de saturation dans le calcul du réseau. La formule de la fonction Maxout est :

$$f(x) = \max(w_1^T x + b_1, w_2^T x + b_2) \quad \text{Équation II-9}$$

w représente les poids, b les biais et T la transposition.

Les applications de reconnaissance téléphonique ont testé avec succès la fonction Maxout.

Le principal inconvénient de la fonction Maxout est qu'elle est coûteuse en calcul car elle double les paramètres utilisés dans chaque neurone, augmentant ainsi le nombre de paramètres que le réseau doit calculer. [NWA18]



**Figure II-13:** La fonction d'activation Maxout.

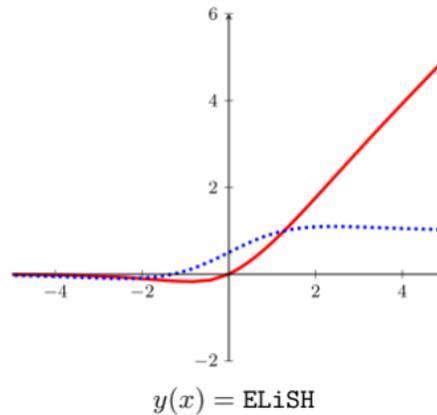
### II.2.4.10. La fonction ELiSH

L'un des fonctions d'activation les plus récents était la fonction d'écrasement linéaire exponentiel, ou fonction ELiSH, que Basirat et Roth ont introduit en 2018 [BAS18]. La fonction Swish et l'ELiSH ont des caractéristiques similaires. Les fonctions ELU et Sigmoid se combinent pour former la fonction ELiSH. La relation donner par :

$$f(x) = \begin{cases} \frac{x}{1 + e^{-x}} , & x \geq 0 \\ \frac{e^x - 1}{1 + e^{-x}} , & x < 0 \end{cases} \quad \text{Équation II-10}$$

Les restrictions déterminent comment les caractéristiques de la fonction ELiSH varient dans les sections négatives et positives. La composante sigmoïde de la fonction ELiSH

améliore le flux d'informations, tandis que la composante linéaire résout les problèmes de gradient de fuite. Plusieurs architectures convolutionnelles profondes ont été utilisées pour appliquer avec succès la fonction ELiSH à l'ensemble de données ImageNet . [NWA18]



**Figure II-14:** La fonction d'activation ELiSH.

## II.2.5. Les fonctions de perte

Les fonctions de perte sont des éléments essentiels dans les environnements d'apprentissage supervisés. Il comprend la métrique qu'un apprenant minimise pour obtenir une bonne hypothèse (modèle) pour le problème à résoudre. Formellement, soit  $P|Y|$  représentent le simplexe de probabilité  $|Y|$ -dimensionnel et  $Y$  représente l'espace des cibles (par exemple,  $Y = \{+1, -1\}$  pour la classification binaire). Une fonction de perte

$L : P \times Y \rightarrow R^+$  convertit une paire de prédictions cibles en un scalaire non négatif dans une tâche de classification typique. Sur un certain jeu de données d'entraînement  $D$ , nous cherchons à minimiser. La variable de décision dans ce problème de minimisation est constituée de paramètres du modèle (dans la version paramétrique). [DUA21]

### II.2.5.1. Entropie croisée (Cross-entropy CE) :

La fonction de perte la plus utilisée pour les problèmes de classification est de loin l'entropie croisée (EC). Cette perte est la sortie d'un softmax paramétrant le logarithme négatif d'une vraisemblance catégorielle. Supposons que pour un échantillon d'apprentissage dont l'étiquette réelle à un codage one-hot  $y$ ,  $p$  est la prédiction convertie softmax du modèle. Nous pouvons exprimer la fonction CE catégoriquement comme suit [DUA21] :

$$L_{CE}(y, p) = - \sum_{i=1}^c y_i \log p_i \quad \text{Équation II-11}$$

### II.2.5.2. Erreur moyenne au carré (squared mean error MSE):

La fonction MSE peut être utilisée pour des tâches de classification, bien qu'il soit plus fréquemment utilisé pour la régression. La perte MSE peut être exprimée comme suit lors de l'utilisation de  $y$  comme étiquette cible dans une représentation ponctuelle [DUA21] :

$$L_{MSE}(y, p) = \sum_{i=1}^c (y_i - p_i)^2 \quad \text{Équation II-12}$$

### II.2.5.3. Erreur absolue moyenne ( Mean absolute error MAE ):

La fonction MAE est fréquemment utilisée dans les problèmes de régression, tout comme le MSE. Le MAE se rapporte à la norme L1, tandis que le MSE représente la norme L2 entre les prévisions et les objectifs [DUA21]:

$$L_{MAE}(y, p) = \sum_{i=1}^c |y_i - p_i| \quad \text{Équation II-13}$$

### II.2.5.4. Charnière-perte

Dans les approches à grande marge pour la classification binaire, les fonctions de perte de charnière sont cruciales [DUA 21] :

$$l_{hinge}(y, p) = \sum_{i=1}^c \max(0, \gamma + p_i - p_j^*) \quad \text{Équation II-14}$$

## II.2.6. Les algorithmes d'optimisation

Les méthodes d'optimisation les plus importantes et les plus appréciées sont les algorithmes de descente de gradient. L'ancienne approche décente de gradient par lots habituelle et la méthode décente de gradient stochastique ne sont pas en mesure d'effectuer la formation de manière efficace et efficiente en raison de l'ensemble de données à grande échelle et de la mémoire de calcul restreinte, en particulier sur les GPU. De plus, pendant le processus de formation, les techniques d'ajustement de la taille des pas sont très importantes. De nombreuses modifications différentes de la méthode du gradient ont été suggérées comme solutions à ce problème. [PIJ17]

### II.2.6.1. Batch Gradient Decent

Après avoir parcouru l'ensemble de l'apprentissage, la méthode de descente de gradient par lots met à jour les paramètres  $x$ .

$$x_{k+1} = x_k - t_k \Delta f(x_k)^{(1:n)} \quad \text{Équation II-15}$$

Pour les problèmes convexes, la descente de gradient par lots est garantie pour converger vers le minimum global ; pour les problèmes non convexes, elle converge vers un minimum local. Le balayage de l'ensemble d'apprentissage complet pour calculer le gradient dans les tâches d'apprentissage profond correspondantes prendra beaucoup de temps, car l'ensemble d'apprentissage contient des millions, voire des milliards d'échantillons. Par conséquent, la mise à jour d'un seul paramètre est trop lente.

En outre, la mémoire de calcul est limitée, ce qui rend difficile l'introduction de toutes les données en une seule fois dans le modèle. Par conséquent, seul un petit pourcentage de modèles d'apprentissage profond utilise la descente de gradient par lots pour résoudre le problème d'optimisation. [PIJ17]

### II.2.6.2. Stochastic Gradient Decent

Pour chaque échantillon d'apprentissage, la descente de gradient stochastique calcule le gradient et met à jour les paramètres.

$$x_{k+1} = x_k - t_k \Delta f(x_k)^{(i)} \quad \text{Équation II-16}$$

Cependant, des mises à jour fréquentes des paramètres entraînent des fluctuations importantes de la fonction objective en raison de la variance élevée des échantillons d'apprentissage. De petites tailles de pas peuvent permettre à SGD de converger vers un point souhaitable, mais elles ralentissent l'apprentissage. De plus, les transferts de données fréquents entre le GPU et la mémoire locale lors de l'utilisation des GPU pour effectuer les calculs réduisent l'efficacité [PIJ17].

### II.2.6.3. Mini-Batch Gradient Decent

Les avantages de la descente de gradient par lots et de la descente de gradient stochastique sont combinés dans la descente de gradient par mini-lots, qui met également à jour les paramètres après avoir recueilli le gradient d'un petit ensemble d'échantillons.

$$x_{k+1} = x_k - t_k \Delta f(x_k)^{(i:i+m)} \quad \text{Équation II-17}$$

Où  $m$  est la taille du mini-batch. La descente de gradient par mini-lots ne peut pas garantir une bonne convergence, et il faut aussi de l'expérience pour régler la taille du pas. C'est pourquoi plusieurs universitaires l'ont enrichi de conseils utiles et de moyens

supplémentaires pour améliorer la convergence. Par commodité, on parle aussi de descente de gradient par mini-lots (SGD) [PIJ17].

#### II.2.6.4. Descente de gradient avec Momentum

Le SGD conventionnel aura tendance à fluctuer dans la vallée étroite s'il y a un long ravin peu profond dans cette direction avec des falaises abruptes. L'un des mécanismes utilisés pour fixer la direction est le Momentum.

$$\begin{aligned}v_k &= mv_{k-1} + t_k \Delta f(x_k) \\x_{k+1} &= x_k - v_k\end{aligned}\tag{Équation II-18}$$

Où  $m$  (0, 1) spécifie le nombre d'itérations pour lesquelles les gradients précédents sont inclus dans la mise à jour actuelle. Dans la plupart des cas,  $m$  est initialement fixé à 0,5, puis porté à 0,9 ou plus une fois l'apprentissage initial stabilisé. [PIJ17]

#### II.2.6.5. Nesterov Accelerated Gradient

En estimant la direction en regardant devant soi, il est possible d'effectuer l'amélioration en utilisant le gradient accéléré de Nesterov [PIJ17].

$$\begin{aligned}v_k &= mv_{k-1} + t_k \Delta f(x_k - mv_{k-1}) \\x_{k+1} &= x_k - v_k\end{aligned}\tag{Équation II-19}$$

#### II.2.6.6. Adagrad

Adagrad le fait essentiellement en divisant le gradient actuel dans la règle de mise à jour par le total des gradients antérieurs afin d'échelonner la taille du pas pour chaque paramètre en fonction de l'historique des gradients pour ce paramètre. [PIJ17]

$$\begin{aligned}G_k &= G_{k-1} + \Delta f(x_k)^2 \\x_{k+1} &= x_k - \frac{t}{\sqrt{G_k + \epsilon}} \Delta f(x_k)\end{aligned}\tag{Équation II-20}$$

#### II.2.6.7. Adadelta

Les deux défauts fondamentaux d'Adagrad - la dégradation constante des taux d'apprentissage au cours de la formation et l'exigence d'un taux d'apprentissage global choisi manuellement - ont été corrigés par le développement d'Adadelta . Adadelta combine les avantages d'Adagrad et du Momentum.

$$\begin{aligned}
 E[\Delta f(x)^2]_k &= \rho E[\Delta f(x)^2]_{k-1} + (1 - \rho)\Delta f(x_k)^2 \\
 \hat{x}_k &= -\frac{\sqrt{E[\hat{x}^2]_{k-1} + \varepsilon}}{\sqrt{E[\Delta f(x)^2]_k + \varepsilon}} \Delta f(x_k) \\
 E[\hat{x}^2]_k &= \rho E[\hat{x}^2]_{k-1} + (1 - \rho)\hat{x}_k^2 \\
 x_{k+1} &= x_k + \hat{x}_k
 \end{aligned}
 \tag{Équation II-21}$$

Il met à l'échelle la taille du pas spécifiquement en fonction du gradient historique. Mais contrairement à Adagrad, il utilise uniquement la fenêtre temporelle la plus récente. De plus, un composant qui agit comme un terme d'accélération et compile les mises à jour précédentes est utilisé (similaire au Momentum).

Où  $\rho$  est une constante de décroissance et  $\varepsilon$  est une petite valeur (par exemple, 1e-6) pour la stabilité numérique. [PIJ17]

### II.2.6.8. RMSprop

Afin de résoudre le problème de la disparition de la taille de pas d'Adagrad, RMSprop est également proposé. De plus, elle utilise la moyenne décroissante des gradients historiques.

$$\begin{aligned}
 E[\Delta f(x)^2]_k &= \rho E[\Delta f(x)^2]_{k-1} + (1 - \rho)\Delta f(x_k)^2 \\
 x_{k+1} &= x_k - \frac{t}{\sqrt{E[\Delta f(x)^2]_k + \varepsilon}} \Delta f(x_k)
 \end{aligned}
 \tag{Équation II-22}$$

Où  $\rho$  est une constante de décroissance. [PIJ17]

### II.2.6.9. Adam

Une autre technique qui détermine la taille de pas adaptative pour chaque paramètre est Adam. Il utilise à la fois les valeurs au carré et la moyenne décroissante des gradients historiques. Il y a les étapes suivantes dans la règle de mise à jour Adam [PIJ17] :

$$\begin{aligned}
 m_k &= \beta_1 m_{k-1} + (1 - \beta_1) \Delta f(x_k) \\
 v_k &= \beta_2 v_{k-1} + (1 - \beta_2) \Delta f(x_k)^2 \\
 \hat{m}_k &= \frac{m_k}{1 - \beta_1^k}, \quad \hat{v}_k = \frac{v_k}{1 - \beta_2^k} \\
 x_{k+1} &= x_k - \frac{t}{\sqrt{\hat{v}_k + \varepsilon}} \hat{m}_k
 \end{aligned}
 \tag{Équation II-23}$$

Où  $\beta_1$  peut être 0,9,  $\beta_2$  peut être 0,999 et  $\varepsilon$  peut être  $1e - 8$ . [PIJ17]

## II.3. Apprentissage profond

L'une des principales méthodes d'apprentissage automatique s'appelle l'apprentissage en profondeur et repose sur des algorithmes inspirés du fonctionnement et de l'assemblage du cerveau. Un système informatique est obligé de s'entraîner à l'aide d'exemples précédemment disponibles à la suite d'un apprentissage approfondi produisant des résultats. La puissance de l'apprentissage en profondeur a permis d'atteindre des résultats auparavant irréalisables. La majorité des applications que nous voyons actuellement reposent toujours sur elles comme mécanisme fondamental. Commande vocale d'appareils électroniques largement utilisés, y compris les appareils intelligents, les véhicules autonomes, les téléphones et les téléviseurs. [BAS19]

L'intelligence artificielle est précieuse dans de nombreux contextes et est comparable ou supérieure aux spécialistes humains dans divers cas. Par conséquent, DL pourrait être une solution aux problèmes suivants :

- ✓ Cas où il n'y a pas de spécialistes humains disponibles.
- ✓ Situations dans lesquelles la taille du problème est énorme et dépasse les capacités de notre faible pensée.
- ✓ Les cas où les gens sont incapables de justifier les choix qu'ils ont faits en fonction de leur expertise. [ALZ21]

### II.3.1. Types de réseaux d'apprentissage en profondeur

Il existe plusieurs types d'algorithmes sont utilisés en apprentissage profond, tels que :

#### II.3.1.1. Réseaux de neurones récurrents (RNN)

Dans le domaine de l'apprentissage en profondeur, RNN est une méthode largement utilisée et fiable. Les principales applications de RNN se situent dans les domaines du traitement du langage et de la PNL. Contrairement aux réseaux conventionnels, RNN utilise

des données séquentielles à l'intérieur du réseau. Étant donné que la structure intégrée fournit des informations précieuses dans un ensemble de données, cette capacité est à la base d'une large gamme d'applications. Déterminer le sens d'un certain mot à l'intérieur d'une phrase, par exemple, nécessite de comprendre le contexte de la phrase. À la lumière de cela, nous pouvons considérer le RNN comme une unité de mémoire à court terme. Où  $x$  représente la couche d'entrée,  $y$  la couche de sortie et  $s$  la couche d'état (masquée). La Figure II-15 montre une évolution typique d'un RNN pour une séquence d'entrée donnée. Trois types différents de méthodes RNN profondes existe, y compris entrée vers caché, caché vers sortie et caché vers caché. L'introduction d'un RNN profond qui offre les avantages des RNN plus avancés basés sur ces trois techniques tout en réduisant la difficulté de formation du réseau profond. [ALZ21]

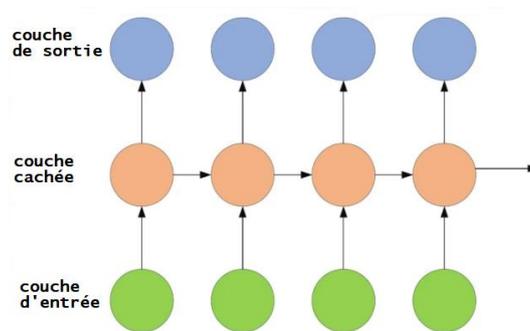


Figure II-15: diagramme du RNN.

### II.3.1.2. La mémoire longue à court terme (LSTM)

Pour résoudre le problème de la disparition des gradients, le réseau Long Short-Term Memory (LSTM) a été créé. La capacité de la cellule RNN à être entraînée pour garantir que le gradient de la fonction objectif par rapport au signal d'état - une quantité directement proportionnelle aux mises à jour des paramètres calculées lors de l'entraînement via Gradient Descent - ne disparaît pas - était l'idée fondamentale de l'architecture LSTM. [SHE20]

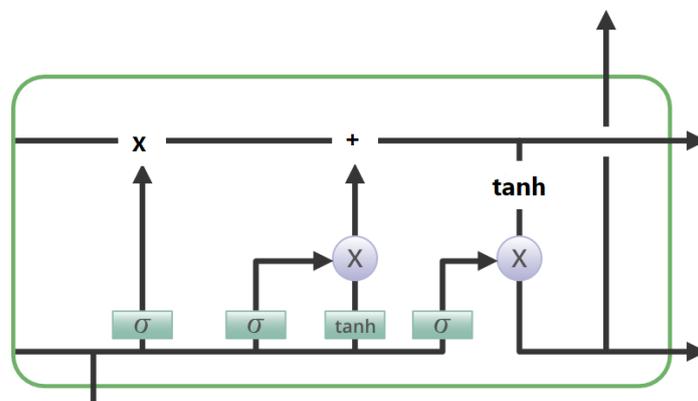


Figure II-16: L'architecture de LSTM.

### II.3.1.3. Les Réseaux de neurones convolutifs (CNN)

L'algorithme le plus connu et le plus souvent utilisé dans l'espace DL et qui nous intéresse dans notre projet de fin d'étude est CNN. Le fait que CNN reconnaisse automatiquement les fonctionnalités pertinentes sans surveillance humaine est un avantage significatif par rapport à ses prédécesseurs. Le CNN est largement utilisé dans divers domaines, notamment la reconnaissance faciale, le traitement de la parole et la vision par ordinateur. Réseaux de neurones traditionnels. Plus précisément, le cortex visuel est constitué de réseaux cellulaires complexes dans le cerveau du chat. CNN simule cette scène. Goodfellow et al. [GOO16] ont identifié trois principaux avantages de CNN. Représentation égale, communication claire et partage des paramètres. Contrairement aux réseaux conventionnels entièrement connectés (FC), CNN utilise des poids partagés et des connexions locales pour utiliser pleinement les formats de données 2D d'entrée comme les signaux d'image. Cette opération utilise un très petit nombre de paramètres, ce qui rationalise le processus de construction du réseau. Les cellules du cortex visuel sont la même chose. En particulier, seules quelques petites zones de la scène, plutôt que la scène entière, sont capturées par ces cellules (c'est-à-dire qu'elles extraient spatialement les corrélations locales disponibles à l'entrée, comme un filtre local à l'entrée). [ALZ21]

La dernière couche d'une forme commune de CNN, similaire au Perceptron multicouche (MLP), est appelée la couche FC, et elle est constituée d'un certain nombre de couches convolutives précédées de couches de sous-échantillonnage. La Figure II-17 montre une illustration de l'architecture CNN. [ALZ21]

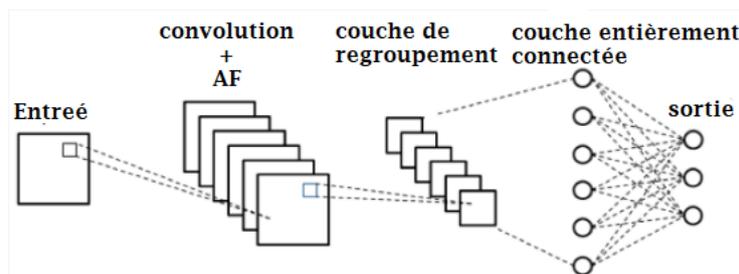


Figure II-17: architecture CNN.

### II.3.2. Les couches de réseau neuronal convolutif (CNN)

Une architecture CNN se compose de plusieurs couches. Chaque couche de l'architecture CNN, y compris ses fonctions.

### **II.3.2.1. La couche convolutive**

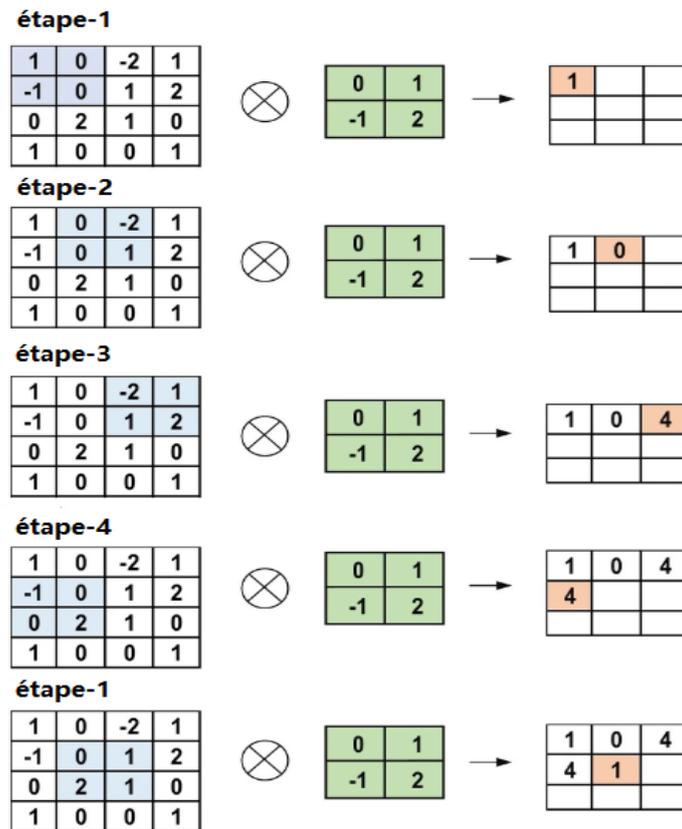
Dans l'architecture CNN, le composant le plus important est la couche convolutive. Il consiste en un ensemble de filtres convolutifs (appelés noyaux). Une image d'entrée, représentée sous la forme d'un volume à N dimensions, est convoluée avec ces filtres pour produire une carte de caractéristiques de sortie. [ALZ21]

#### **II.3.2.1.1. Le noyau**

C'est une grille de nombres ou de valeurs discrètes décrit le noyau. Chaque valeur est appelée un poids de noyau. Au début du processus de formation CNN, des nombres aléatoires sont attribués comme pondérations pour les noyaux. De plus, il existe plusieurs méthodes différentes pour initialiser les poids. Ensuite, ces poids sont ajustés à chaque époque d'apprentissage ; ainsi, le noyau apprend à extraire les caractéristiques importantes. [ALZ21]

#### **II.3.2.1.2. Opérations de convolution**

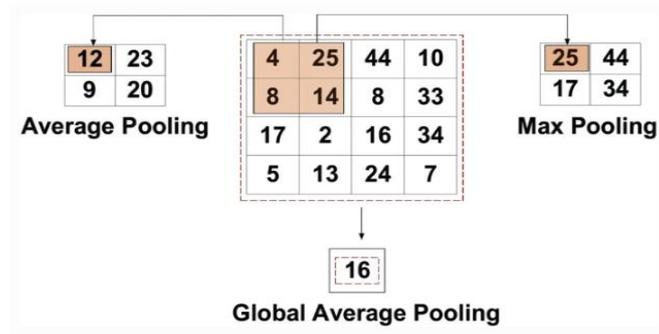
Tout d'abord, le format d'entrée CNN est décrit. Les formats vectoriels sont l'entrée des réseaux de neurones traditionnels, tandis que les images multicanaux sont l'entrée des CNN. Par exemple, un seul canal est un format d'image en niveaux de gris, tandis que le format d'image RVB est à trois canaux. Pour comprendre l'opération de convolution, prenons comme exemple une image  $4 \times 4$  en niveaux de gris avec un noyau  $2 \times 2$  initialisé avec des poids aléatoires. Tout d'abord, le noyau fait glisser toute l'image horizontalement et verticalement. De plus, le produit scalaire entre l'image d'entrée et le noyau est déterminé, leurs valeurs correspondantes sont multipliées, puis additionnées pour créer une seule valeur scalaire qui est calculée simultanément. Ensuite, répétez tout le processus jusqu'à ce que vous ne puissiez plus glisser. Notez que les valeurs calculées du produit scalaire représentent les cartes d'entités de sortie. La Figure II-18 illustre graphiquement les principaux calculs effectués à chaque étape. Dans ce chiffre, le vert clair représente les noyaux  $2 \times 2$ , tandis que le bleu clair représente des régions de taille similaire de l'image d'entrée. Les deux sont multipliés ; le résultat final (marqué en orange clair) après avoir additionné les valeurs multipliées résultantes représente la valeur d'entrée pour la carte d'entités en sortie. [ALZ21]



**Figure II-18:** Les calculs primaires exécutés à chaque étape de la couche convolutive.

### II.3.2.2. La Couche de regroupement (Pooling)

La tâche principale de la couche de regroupement est le sous-échantillonnage de la carte des entités. Ces cartes sont générées par les opérations de convolution suivantes. En d'autres termes, cette méthode réduit l'échelle des cartes d'entités à grande échelle pour créer des cartes d'entités plus petites. En même temps, il préserve la plupart des informations (ou caractéristiques) dominantes à chaque étape de l'étape de mise en commun. De manière similaire à l'opération de convolution, les pas et les tailles initiales des noyaux sont alloués avant d'effectuer l'opération de regroupement. Il existe plusieurs types de méthodes de regroupement disponibles pour différentes couches de regroupement. Ces méthodes incluent la mise en pool arborescente, la mise en pool fermée, la mise en pool moyenne, la mise en pool min, la mise en pool max, la mise en pool moyenne globale (GAP) et la mise en pool max globale. Les méthodes de mise en commun les plus connues et les plus couramment utilisées sont la mise en commun maximale, la mise en commun minimale et la mise en commun GAP. La Figure II-19 illustre les trois opérations de regroupement. [ALZ21]



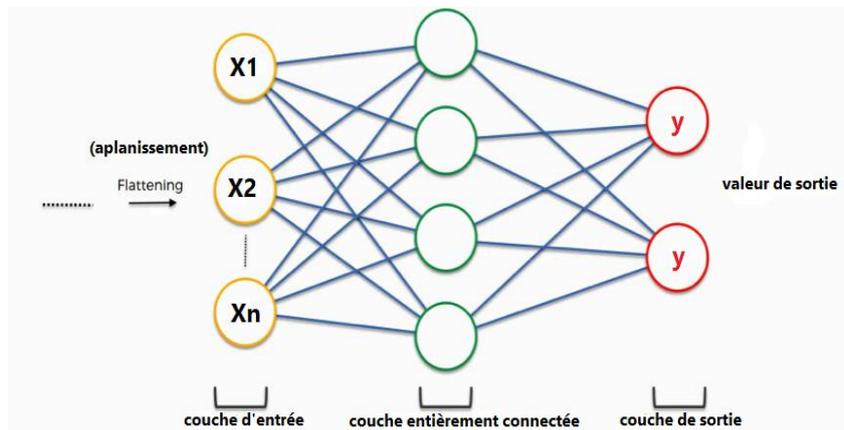
**Figure II-19:** Les opérations de regroupement ( pooling ).

### II.3.2.3. Fonctions d'activation

Le mappage de l'entrée à la sortie est une fonction centrale de tous les types de fonctions d'activation dans tous les types de réseaux de neurones. La valeur d'entrée est déterminée en calculant une somme pondérée de l'entrée du neurone et de son biais. Cela signifie que la fonction d'activation décide d'activer ou non un neurone en fonction d'une entrée particulière en créant une sortie correspondante. Utilisez des couches d'activation non linéaires après toutes les couches avec des poids dans l'architecture CNN. L'activation de cette propriété non linéaire des couches signifie que le mappage de l'entrée à la sortie sera non linéaire ; de plus, ces couches permettent aux CNN d'apprendre des choses très complexes. La fonction d'activation doit également être discriminante, une propriété extrêmement importante car elle permet de former le réseau en utilisant la rétropropagation des erreurs. [ALZ21]

### II.3.2.4. Couche entièrement connectée (Fully Connected Layer)

Ordinairement, cette couche se trouve à la fin de chaque architecture CNN. Au sein de cette couche, chaque neurone est connecté à tous les neurones de la couche précédente, une approche dite entièrement connectée (FC). Il est utilisé comme classificateur CNN. Il suit l'approche de base du réseau de neurones perceptron multicouche traditionnel car il s'agit d'un réseau de neurones à anticipation. L'entrée de la couche FC provient de la dernière couche de regroupement ou de convolution. Cette entrée se présente sous la forme d'un vecteur, qui est créé à partir de la carte des entités après aplatissement. La sortie de la couche FC représente la sortie CNN finale. [ALZ21]



**Figure II-20:** La couche entièrement connectée.

### II.3.3. Quelques architecture basée CNN

Différentes architectures de réseau sont utilisé pour les taches de classification d'image, parmi les plus notables, on retrouve VGG16, ResNet50, MobileNetV2, EfficientNetB0.

#### II.3.3.1. Groupe de géométrie visuelle (VGG 16)

L'une des architectures d'apprentissage en profondeur les plus populaires s'appelle VGG 16. Il a 41 couches qui ont été organisées comme suit : 13 couches convolutionnelles (Conv), 16 couches de poids et 3 couches FC. Toutes les couches Conv dans VGG 16 utilisent un petit noyau 3x3 (filtre) avec un pas de 1. Les couches Conv sont toujours suivies par les couches Max Pooling. Trois canaux de taille fixe de 224x224 pixels constituent l'entrée pour VGG 16. Les trois couches FC dans VGG 16 ont différentes profondeurs. Le dernier FC a une taille de canal de 1000, ce qui correspond au numéro de l'étiquette de classe dans le jeu de données imageNet, alors que les deux précédents ont la même taille de canal (4096). La couche softmax, qui est en charge de la probabilité spécifiée pour l'image d'entrée, sert de couche de sortie. La Figure II-21 affiche l'architecture préminente pour VGG 16 [ALB21].

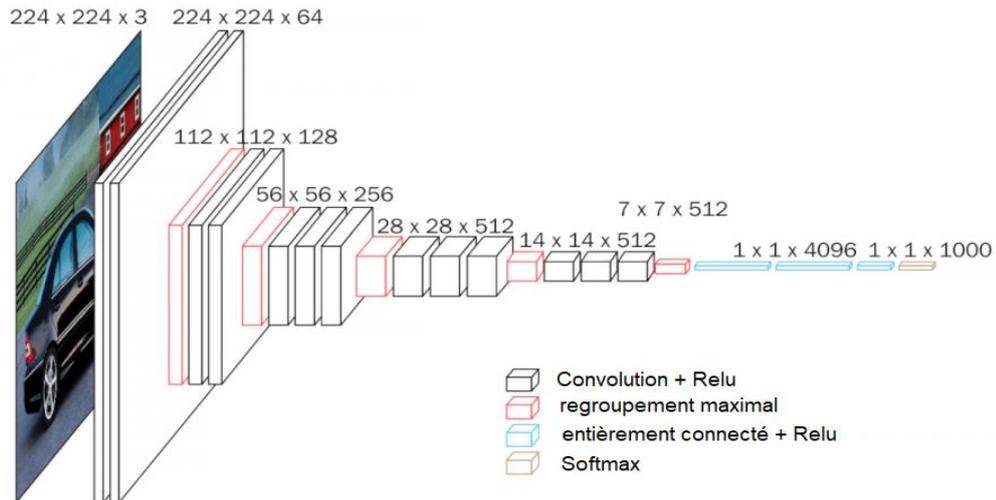


Figure II-21: L'architecture de VGG 16.

### II.3.4. Resnet 50

ResNet (Residual Network) a été créé par He et al [ZHA16]. Contrairement aux réseaux précédents, leur objectif était de créer un réseau ultra-profond immunisé contre le problème du gradient de fuite. En fonction du nombre de couches (commençant à 34 couches et continuant jusqu'à 1202 couches), différentes variantes de ResNet ont été développées. ResNet 50, qui comprenait 49 couches convolutionnelles et une couche FC, était le type le plus populaire. Alors qu'il y avait 3,9 millions de MAC au total, il y avait 25,5 millions de poids de réseau au total. La notion innovante de ResNet est son utilisation de l'idée de voie de contournement, qui a été utilisée dans Highway Nets en 2015 pour résoudre le problème de la formation d'un réseau plus profond. [ALZ21]

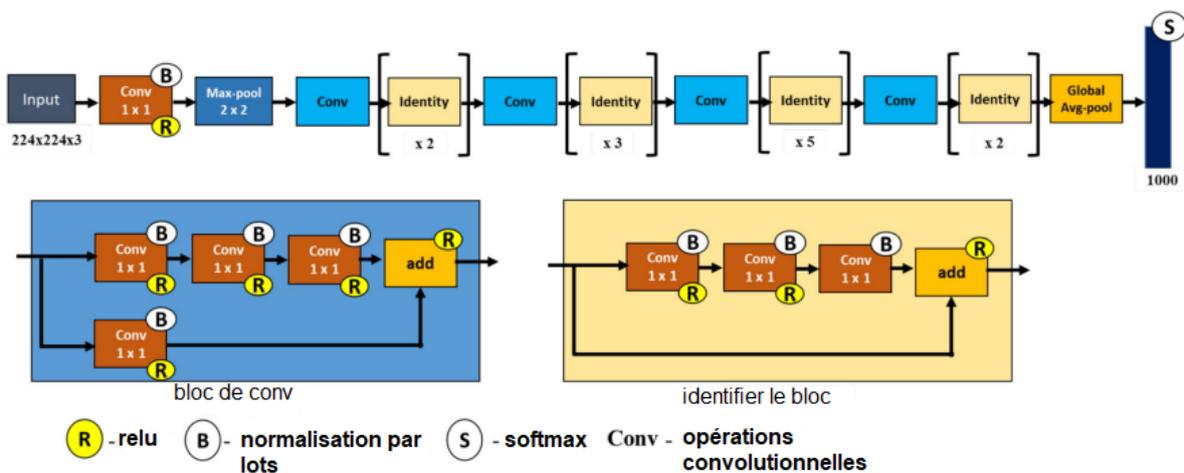


Figure II-22: L'Architecture ResNet de base.[KIM21]

### II.3.5. MobileNet V2

Une architecture d'apprentissage en profondeur appelée MobileNet est axée sur la plateforme mobile, qui a des ressources informatiques limitées. Google lance plus tard MobileNet V2 [ZEN18], une version améliorée qui n'apporte que des modifications mineures à la première. La convolution séparable continue de servir de bloc de construction fondamental du réseau. Les caractéristiques des images de fruits ont été extraites dans [SAN18] à l'aide de la version 2 de MobileNet qui avait été précédemment formée sur des ensembles de données ImageNet. Les paramètres utilisés sont passés de 4,24 millions à seulement 3,47 millions, mais avec une précision améliorée. [ZAK20]

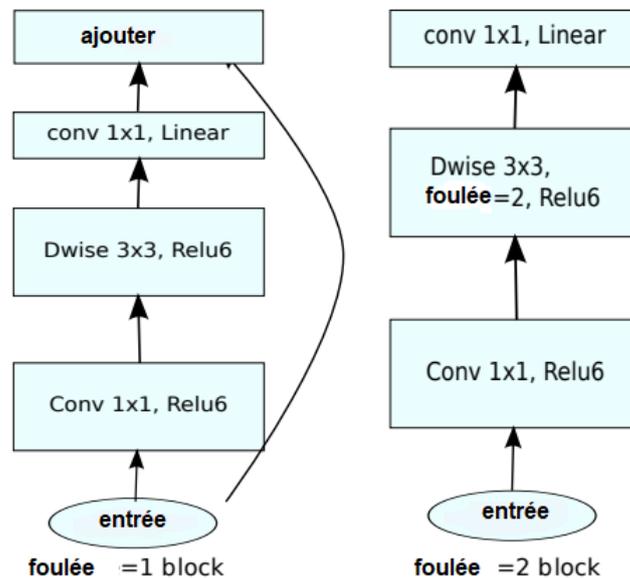


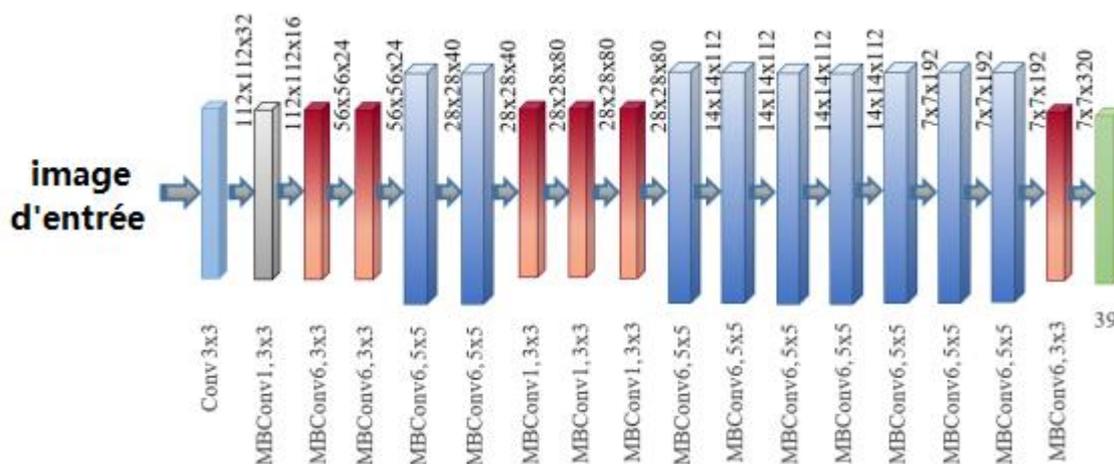
Figure II-23: Blocs convolutionnels MobileNetV2.

### II.3.6. EfficientNet

Depuis 2012, le succès s'est accru en même temps que la complexité des modèles employés dans l'ensemble de données ImageNet, mais beaucoup d'entre eux sont inefficaces en termes de charge de calcul. Une combinaison de modèles CNN peut être considérée comme le modèle EfficientNet, qui fait partie des modèles les plus avancés puisqu'elle a atteint une précision de 84,4 % avec 66 millions de paramètres dans le défi de classification ImageNet. Le groupe EfficientNet se compose de 8 modèles entre B0 et B7, et au fur et à mesure que le nombre de modèles augmente, la précision augmente considérablement alors que le nombre de paramètres calculés n'augmente pas de manière significative. Au lieu de la fonction d'activation Rectifier Linear Unit (ReLU), EfficientNet utilise une toute nouvelle

fonction d'activation appelée Swish [TAN19]. Les architectures d'apprentissage profond sont conçues pour fournir des méthodes plus efficaces en utilisant moins de modèles. EfficientNet, contrairement à d'autres modèles de pointe, produit des résultats plus efficaces en augmentant constamment la profondeur, la largeur et la résolution tout en réduisant l'échelle du modèle. Trouver une grille pour représenter la connexion entre les différentes dimensions de mise à l'échelle du réseau de base sous une limitation de ressources définie est la première étape de l'approche de mise à l'échelle composée. Les dimensions de profondeur, d'étendue et de résolution sont mises à l'échelle de manière appropriée à l'aide de cette méthode. Le réseau cible est mis à l'échelle à l'aide de ces paramètres [TAN19] pour créer le réseau de référence. [ATI21]

Le principal composant de construction d'Efficient Net est le goulot d'étranglement inversé MB Conv, qui a été initialement présenté dans MobileNet V2 [MAR18], bien qu'il soit utilisé un peu plus que MobileNet V2 en raison du budget FLOPS (opérations en virgule flottante par seconde) plus élevé. Des connexions directes sont utilisées entre les goulets d'étranglement qui relient beaucoup moins de canaux que les couches d'expansion, car les blocs de MB Conv sont constitués d'une couche qui étend d'abord les canaux, puis les compresse. Par rapport aux couches conventionnelles, les convolutions séparables en profondeur de cette architecture doublent presque l'efficacité de calcul d'un facteur  $k^2$ , où  $k$  est la taille du noyau, qui indique les dimensions de la largeur et de la hauteur de la fenêtre de convolution 2D [MAR18]. La **Figure II-24** présente une illustration schématique du modèle EfficientNet B0. [ATI21]



**Figure II-24:** représentation schématique d'Efficient Net. [ATI21]

## II.4. Transfert d'apprentissage (TA)

Selon des recherches récentes, les CNN profonds, qui offrent un support révolutionnaire pour résoudre de nombreux problèmes de classification, sont largement utilisés. Les modèles Deep CNN nécessitent généralement une quantité importante de données pour bien fonctionner. L'absence de données de formation est un problème fréquent lors de l'utilisation de tels modèles. Il n'existe actuellement aucun moyen efficace d'accomplir la tâche laborieuse de collecter une quantité importante de données. Par conséquent, la technique transfert d'apprentissage, qui est très efficace pour résoudre le problème de pénurie de données d'apprentissage, est utilisée pour gérer le problème de l'ensemble de données sous-dimensionné. De grandes quantités de données sont utilisées pour former le modèle CNN dans le cadre du processus TA. Le modèle est ajusté pour la formation sur un ensemble de données limité de demandes à l'étape suivante. [ALZ21]

Dans le but de la reconnaissance d'images, de nombreux modèles CNN ont été formés sur de grands ensembles de données comme ImageNet. Ensuite, ces modèles peuvent être utilisés pour identifier une activité différente sans nécessiter de nouvelle formation. Mis à part quelques éléments d'apprentissage, les poids sont également inchangés. Ces modèles sont très utiles dans les situations où il n'y a pas suffisamment d'échantillons de données. L'utilisation d'un modèle pré-entraîné présente de nombreux avantages. Premièrement, une puissance de traitement coûteuse est nécessaire pour former des modèles complexes sur des ensembles de données importants. Deuxièmement, la formation de grands modèles peut prendre plusieurs semaines et prendre beaucoup de temps. Un modèle pré-entraîné peut également accélérer la convergence et aider à la généralisation du réseau. [ALZ21]

## II.5. Conclusion

Dans ce chapitre nous avons présenté les concepts fondamentales des réseaux de neurones artificiels et à convolution, en commençant par la relation entre les neurones biologiques et artificielles, les différents types des fonctions d'activation (Sigmoid/tanh/Relu/leakyrelu,... etc.), ainsi que les différentes fonctions de perte, les algorithmes d'optimisations (Descente de gradient/Adagard/Adadelta/Adam ,...etc). Ensuite, nous avons abordé le domaine de l'apprentissage en profondeur, en discutant les différents types des réseaux (RNN/CNN,...etc.). Nous avons également présenté quelques architectures basé sur CNN (VGG16, ResNet50,...etc.), nous avons abordé aussi le concept de transfert d'apprentissage que nous voulons utiliser pour la classification des mammographies.

# Chapitre

---

# 3

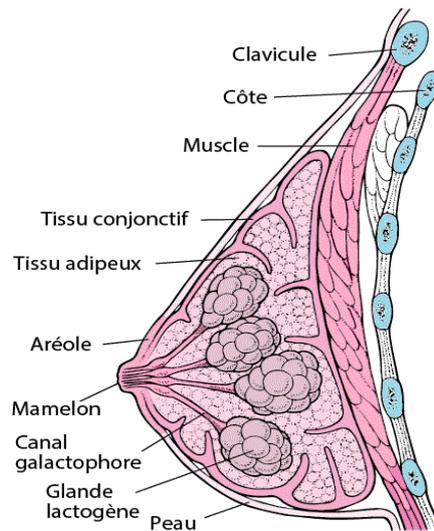
# LA CLASSIFICATION DES MAMMOGRAPHIES NUMERIQUES

### **III.1. Introduction**

Le cancer du sein est la forme de cancer la plus répandue chez les femmes dans le monde. La mammographie est une technique d'imagerie par rayons X à faible dose qui joue un rôle crucial dans la détection et le diagnostic précoces du cancer du sein. Cependant, la densité mammaire est un facteur important qui peut influencer sur la précision du dépistage par mammographie. La densité mammaire fait référence à la quantité de tissu glandulaire et conjonctif dans le sein par rapport à la quantité de graisse. Les femmes ayant une densité mammaire plus élevée ont un risque plus élevé de développer un cancer du sein, car un tissu mammaire dense peut rendre plus difficile la détection de petites tumeurs sur les mammographies. Le système de classification BI-RADS est utilisé pour classer la densité mammaire en quatre catégories, ce qui peut aider les radiologues à déterminer le protocole de dépistage approprié pour chaque patient. La compréhension de la densité mammaire est cruciale pour la détection et le diagnostic précoces du cancer du sein et peut aider à améliorer l'efficacité du dépistage par mammographie. Dans ce chapitre nous couvrirons les concepts fondamentaux de l'anatomie de sein et la mammographie, nous présenterons également l'imagerie médicale dédiée au dépistage du cancer du sein et les pathologies mammaires. Ensuite les méthodes de classification des images mammographiques.

### **III.2. Anatomie de sein**

La majorité du tissu qui compose l'âme est un tissu adipex-graisseux, c'est lui qui lui donne forme et volume. Vingt et un lobules composent le canal mammaire, qui est noué dans le tissu conjonctif et devient actif pendant l'allaitement. Au niveau du cantaloup, des galactophores séparés sont utilisés pour détourner le lait séparé. La peau qui recouvre l'abdomen et les fibres musculaires de la poitrine maintiennent en place la myriade de vaisseaux sanguins qui sillonnent sa surface. Plus d'informations sur ces différents composants du sein sont fournies dans la **Figure III-1**. [FAT17]



**Figure III-1** : anatomie de sein.

Le cancer du sein est une tumeur maligne qui se développe à partir des cellules de la glande mammaire et qui se caractérise par une multiplication anarchique des cellules malignes. Si elle n'est pas traitée, cette tumeur peut se propager localement aux organes voisins (extension locale puis extension régionale), ainsi que par voie sanguine ou lymphatique (extension générale) vers des organes situés à distance (métastases). Les tumeurs connues par deux grandes catégories sont : bénignes et malignes. [FAT17]

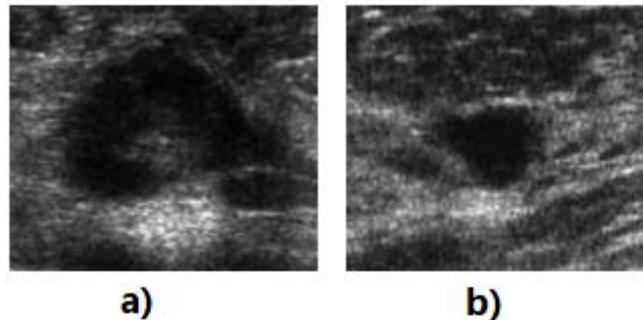
### **III.3. L'imagerie médicale dédiée aux anomalies mammaires**

Plusieurs méthodes de diagnostic sont actuellement utilisées fréquemment dans l'industrie médicale pour étudier le corps humain. Chaque méthode a des utilisations pour divers organes et est sensible à un type spécifique de contraste. De plus, différentes approches peuvent offrir des connaissances complémentaires sur le même tissu. L'imagerie par ultrasons, l'imagerie par résonance magnétique (IRM) et la mammographie sont les techniques d'imagerie médicale utilisées pour la détection et le traitement du cancer du sein. (Imagerie par rayons X). Les différentes méthodes actuellement utilisées, ainsi que leurs caractéristiques, sont présentées dans l'écriture qui suit. [CHE12]

#### **III.3.1. L'échographie**

L'idée de base derrière l'échographie est de placer une sonde sur l'épiderme au-dessus de la zone cible. Les vibrations ultrasonores émises par cette sonde traversent les tissus et sont renvoyées à la sonde. Une fois ce signal recueilli, il est traité par un système informatique qui envoie alors une image en direct sur un écran vidéo. La Figure III-2 montre deux exemples

d'images d'échographie mammaire, l'une avec une tumeur cancéreuse et l'autre avec une lésion bénigne. [CHE12]



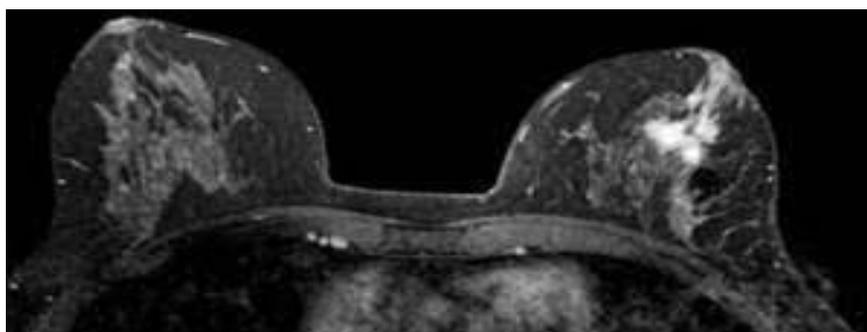
**Figure III-2 :** Exemples d'images d'échographie mammaire. a) Lésion maligne (sein droit) et  
b) Lésion bénigne.

### III.3.2. L'imagerie par résonance magnétique (IRM)

Une méthode de dépistage médical relativement nouvelle est l'imagerie par résonance magnétique. (Début des années 1980). Cette technique repose sur l'utilisation d'un rayonnement RF et d'un aimant pour créer le champ magnétique. Les atomes d'hydrogène dans l'organisme humain vibrent subtilement comme principe de base.

Tous les atomes d'hydrogène s'alignent de la même manière en présence d'un puissant champ magnétique. Puis, pendant un temps très bref, des signaux radio les stimulent.

Ils sont prétendument mis en résonance. Les atomes déchargent l'énergie accumulée en générant un signal à l'issue de cette stimulation. Un dispositif informatique enregistre et convertit ces informations en une image. [CHE12]



**Figure III-3:** Exemple d'IRM mammaire

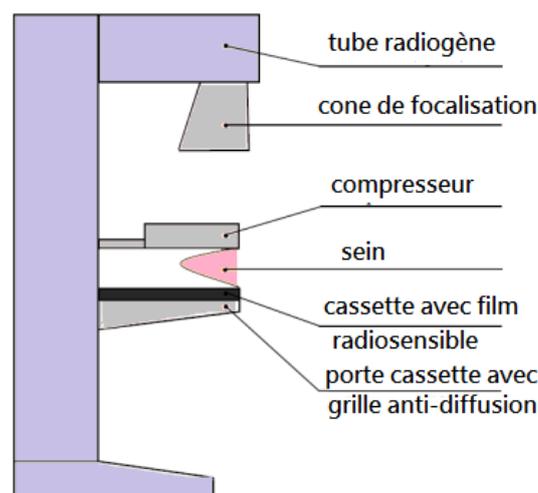
L'IRM permet la collecte d'images et d'études de haute qualité dans toutes les dimensions spatiales. De plus, il offre une image haute résolution qui permet une analyse approfondie de la maladie. Cette méthode n'est utilisée que sur des patients sans prothèses métalliques en raison de son coût élevé. Par conséquent, il ne peut être utilisé que pour fournir des informations sur une anomalie déjà identifiée par mammographie ou échographie.

La mammographie apparaît comme la technique la plus appropriée pour une utilisation générale dans le suivi ou le diagnostic du cancer du sein compte tenu des circonstances générales des différentes méthodes d'imagerie médicale (limites de l'examen échographique, coût de l'examen IRM). [CHE12]

### III.3.3. La mammographie

La mammographie est une technique de radiographie particulièrement adaptée pour examiner les seins des femmes. Elle a pour objectif de détecter les anomalies le plus tôt possible avant qu'elles ne provoquent des symptômes visibles. La mammographie est utilisée non seulement dans les campagnes de dépistage du cancer du sein, mais également pour le diagnostic et la localisation des lésions lors d'interventions chirurgicales telles que les ponctions. Cette technique permet d'observer l'ensemble du tissu mammaire à partir d'une ou deux incidences seulement. [CHE12]

Le mammographe est l'appareil utilisé pour réaliser une mammographie (**Figure III-4**), Il est équipé d'un tube radiogène produisant des rayons X de faible énergie (entre 20 et 50 keV) et d'un système de compression du sein. Les deux seins sont comprimés successivement, ce qui permet d'étaler les tissus mammaires et de faciliter la visualisation des structures internes tout en réduisant la dose de rayonnement. Ensuite, les deux seins sont exposés à une faible dose de rayons X, produisant une image sur un détecteur plan. Les images de la glande mammaire sont analysées en distinguant l'atténuation des rayons X par les différents types de tissus. Dans la section suivante, nous détaillons l'anatomie du sein pour établir la relation entre la nature du tissu mammaire et la pénétration des rayons X. [CHE12]



**Figure III-4 :** les composants d'un mammographe.

### III.4. Les pathologies mammaires

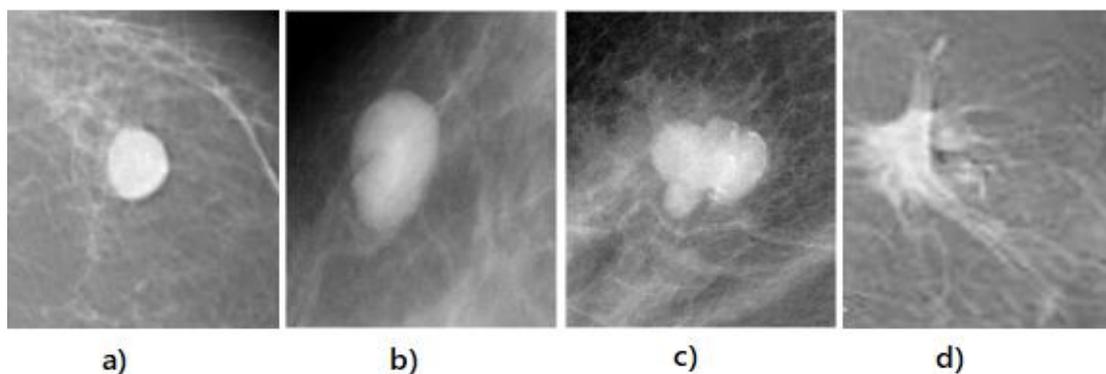
#### III.4.1. Les microcalcifications (Mcs)

Une microcalcification est une accumulation d'ions calcium, qui sont composés des éléments  $\text{Ca}_3(\text{PO}_4)_2$ ,  $\text{CaCO}_3$  et  $\text{Mg}_3(\text{PO}_4)_2$ . En raison de leur radio-opacité élevée, ces composés apparaissent sur les images mammographiques sous forme de minuscules taches lumineuses. Le fort contraste et la taille minuscule (0,5 mm) des microcalcifications les distinguent des autres éléments. On parle alors de macro-calcifications, souvent anodines, lorsqu'elles sont supérieures à 1 mm. Parce qu'il n'y a pas d'exigence de taille minimale pour les microcalcifications (en abrégé Mcs), les plus petites des images de mammographie peuvent facilement être confondues avec du bruit. Des critères tels que la morphologie, l'emplacement et la quantité font partie de la définition des microcalcifications utilisées pour évaluer le niveau de suspicion. [CHE12]

#### III.4.2. Les masses

Les termes « opacité » et « masse » désignent une lésion importante qui occupe une surface et se manifeste sous au moins deux angles différents. Tant que sa tridimensionnalité n'est pas confirmée, une opacité potentielle qui ne peut être vue que comme une occurrence unique est considérée comme une asymétrie. La forme, le contour et la densité de ces agrégats sont autant de caractéristiques essentielles à décrire. [CHE12]

La **Figure III-5** montre quatre formes distinctes que l'on peut distinguer : ronde, ovale, lobulaire et irrégulière.



**Figure III-5:** Les différentes formes d'une masse.

- La forme ronde (a) correspond à une masse qui est sphérique, circulaire ou globuleuse.
- La forme ovale (b) est caractérisée par une masse présentant une forme elliptique ou en forme d'œuf.

- La forme lobulée (c) est une masse présentant une légère ondulation dans sa forme.
- La forme irrégulière (d) est réservée aux masses dont la forme est aléatoire et ne peut être caractérisée par les termes précédemment cités. [CHE 12]

Cinq types de contour des masses : circonscrit (a), micro-lobulé (b), masqué (c), indistinct (d) et spiculé (e).

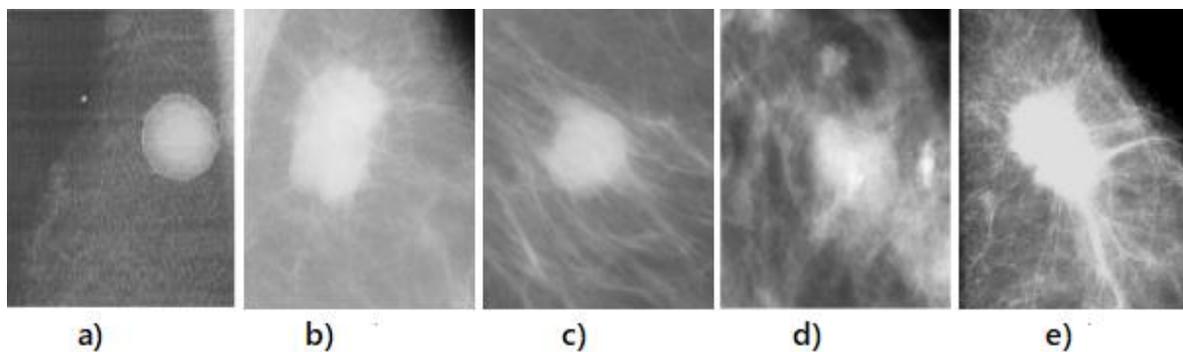
A) **Circonscrit** : Ceci concerne une différence soudaine entre la lésion et le tissu avoisinant. Le contour est donc net et bien défini. Pour qu'une masse soit considérée comme circonscrite, au moins 75% de son contour doit être clairement délimité.

B) **Micro-lobulé** : Dans cette situation, de légères indentations sur le contour créent de petites ondulations.

C) **Masqué** : Un contour masqué est celle qui est cachée par le tissu régulier à proximité. Ce mot est utilisé pour décrire une masse circonscrite où une partie du contour est masquée.

D) **Indistinct** : Dans ce cas, le contour est mal défini. Cette absence de limites nettes (le contraire de circonscrit) peut indiquer une infiltration.

E) **Spiculé** : La masse est caractérisée par des lignes radiales issues du contour de la masse. Ces lignes radiales sont appelées spicules. [CHE12]



**Figure III-6:** Les différents types de contours d'une masse. [CHE12]

### III.5. La densité des mammographies

La densité mammaire est une mesure radiologique de la quantité de tissu mammaire dense ou radio-opaque (tissu fibroglandulaire) par rapport au tissu adipeux radio-transparent. Cette mesure permet d'évaluer la capacité de la mammographie à détecter des lésions : plus le sein est dense, moins la sensibilité de la mammographie est élevée. Une forte densité mammaire est également associée à un risque accru de cancers de l'intervalle et de faux positifs. Toutes les études menées sur la relation entre densité mammaire et cancer du sein ont montré une augmentation significative du risque de cancer chez les femmes ayant une densité

mammaire élevée, qui ne peut pas être expliquée par un manque de détection lors de mammographies antérieures. [TAR08]

Il existe plusieurs classifications proposées pour évaluer la densité mammaire, mais la classification BI-RADS, issue du lexique de l'American College of Radiology (ACR), est celle qui est couramment utilisée en routine. Cette classification se base sur une évaluation visuelle du pourcentage de tissu radio-opaque sur une mammographie, et elle décrit quatre types [TAR08] :

### **III.5.1. Densité de type A**

Le sein est presque entièrement grasseux, avec moins de 25% de tissu glandulaire mammaire présent.

### **III.5.2. Densité de type B**

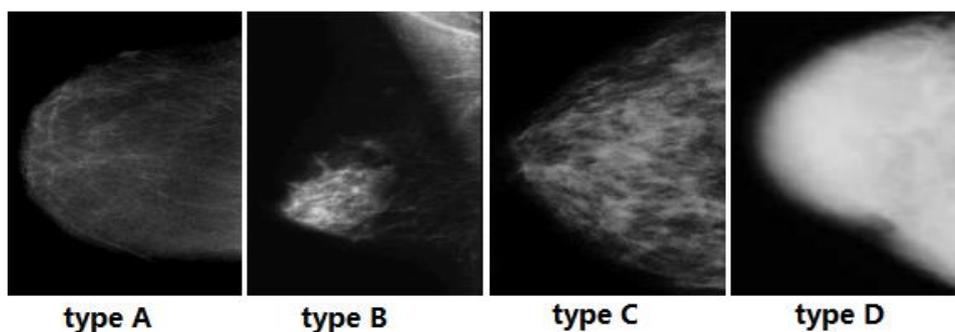
Il y a la présence d'opacités fibroglandulaires éparses, représentant environ 25% à 50% de tissu glandulaire mammaire.

### **III.5.3. Densité de type C**

Le tissu mammaire est dense et hétérogène, ce qui peut compliquer la détection de petites masses. Il représente environ 51% à 75% de tissu glandulaire mammaire.

### **III.5.4. Densité de type D**

Le tissu mammaire est très dense, ce qui peut réduire la sensibilité de la mammographie. Il représente plus de 75% de tissu glandulaire mammaire. Le compte-rendu doit inclure l'appréciation de la densité mammaire, ainsi que l'indication de son homogénéité ou de son hétérogénéité.



**Figure III-7 :** Différents types de densité mammaire.

La fréquence de la densité mammaire importante (types C et D) chez les femmes de 40 ans et plus est d'environ 36-39 % pour le type C et de 8 % pour le type D. En moyenne, les femmes jeunes ont des seins plus denses que les femmes plus âgées, mais la densité mammaire varie considérablement d'une série à l'autre, ce qui montre que la densité mammaire n'est pas faiblement liée à l'âge. Dans une étude portant sur 1 177 patientes suivies pendant 12 ans, 39 % des femmes ont présenté une modification de la densité mammaire, passant des groupes denses aux groupes de seins clairs (classification de Wolfe), même si la majorité de ces femmes sont devenues ménopausées au cours de la période de suivi. Selon l'étude, la densité mammaire varie peu avec l'âge chez 50 % des femmes, et d'autres études ont montré que 10 à 30 % des femmes âgées peuvent conserver des seins très denses. En somme, chez les femmes âgées de 40 à 80 ans, les seins denses et très denses sont présents respectivement dans environ 37 % et 8 % des cas. La densité mammaire diminue généralement avec l'âge et/ou la ménopause, mais cela ne s'applique pas systématiquement à toutes les femmes. [TAR08]

### **III.6. La classification des pathologies mammaires**

Les systèmes de classification les plus répandus et les plus utilisés sont ceux de Le Gal et de BI-RADS.

#### **III.6.1. La classification de Legal**

En 1976, le Service de Radiodiagnostic de l'Institut Curie à Paris, dirigé par Le Gal, a développé la classification de Le Gal [LEG84]. Cette classification décrit cinq types morphologiques qui présentent une valeur prédictive de malignité croissante (voir Tableau 1):

Type 01	Les Mcs de Type 1 sont annulaires, arciformes ou polyédriques et présentent un risque de cancer du sein quasi nul.
Type 02	Les Mcs de Type 2 sont rondes et de tailles variables, et leur risque de carcinome est de 22%.
Type 03	Les Mcs de Type 3 sont poussiéreuses et pulvérulentes, avec un risque de cancer de 36%.
Type 04	Les Mcs de Type 4 sont irrégulières et sont associées à un risque de cancer de 56%.
Type 05	Les Mcs de Type 5 sont vermiculaires ou branchées et ont un risque de carcinome de 90%.

**Tableau III-1** : Classification de Le Gal. [CHE12]

Bien que l'ancienne classification présente l'avantage d'être simple, son principal inconvénient est qu'elle ne repose que sur les microcalcifications et ne prend pas en compte d'autres facteurs tels que [CHE12] :

- L'analyse morphologique des masses.
- L'agencement des microcalcifications.
- L'évaluation des distorsions architecturales.
- L'approche du radiologue pour chaque cas.

### **III.6.2. La classification BI-RADS**

Le système de rapport et de données sur l'imagerie mammaire (BI-RADS) de l'American College of Radiology (ACR) offre une catégorisation normalisée pour les examens mammographiques. Ce mécanisme présente une forte association avec le risque de cancer du sein. La méthode BI-RADS peut apporter aux médecins de famille des résultats importants, les aider à décider du type de suivi et de traitement nécessaires et les inciter à apporter un soutien psychologique et éducatif aux patients. [EBE06]

Le système de classification BIRADS va de 0 à 6, chaque catégorie indiquant un niveau de suspicion différent pour le cancer du sein. Voici une explication de chaque catégorie :

#### **III.6.2.1. BI-RADS 0**

Un diagnostic ne peut être établi avec certitude sans imagerie ou tests supplémentaires, ce que la catégorie 0 désigne comme un examen incomplet. Des problèmes techniques, tels qu'une mauvaise clarté d'image ou un résultat de mammographie original flou, peuvent en être la cause. [EBE06]

#### **III.6.2.2. BI-RADS 1**

La catégorie 1 indique que la mammographie est négative pour toute anomalie significative et est considérée comme normale. [EBE06]

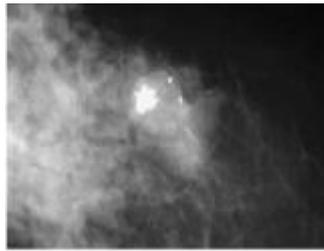
#### **III.6.2.3. BI-RADS 2**

La catégorie 2 est également considérée comme une constatation normale, mais elle peut inclure des caractéristiques bénignes telles que des kystes ou des calcifications. [EBE06]

Les anomalies entrant dans cette catégorie sont les suivantes :

- Masse ronde avec macro-calcifications.
- Ganglion intra-mammaire.

- Anomalie de densité grasseuse ou mixte.
- Masses rondes correspondant à un kyste en échographie. (**Figure III-8**)
- Macro-calcifications cutanées.
- Calcifications punctiformes régulières diffuses.
- Calcifications sur matériel de suture.
- Macro-calcifications sans opacité.
- MCS annulaires, semi-lunaires ou sédimentées. [CHE12]



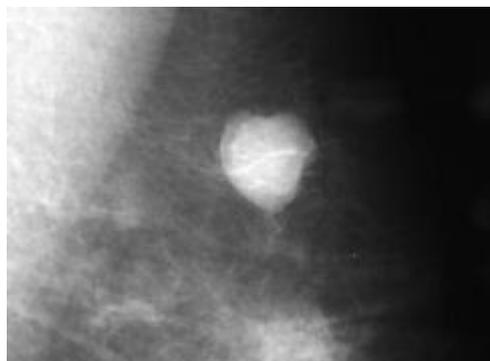
**Figure III-8** : Aspect typique de fibroadénome partiellement calcifié : BI-RADS 2.

#### III.6.2.4. BI-RADS 3

BI-RADS 3 Indique que la mammographie est très probablement normale, mais qu'une mammographie de suivi devrait être effectuée dans six mois. Dans ce groupe, la probabilité de cancer du sein est de 2 %. Il est nécessaire de confirmer que ces analyses de suivi ont été effectuées selon les besoins. [EBE06]

Les anomalies susceptibles d'être classées BIRADS 3 sont :

- Masse circonscrite ronde ou ovale sans micro-lobulations et sans liquide en échographie. (**Figure III-9**)
- Mcs rondes ou punctiformes régulières ou pulvérulentes, peu nombreuses, en petits amas ronds isolés.
- Petit amas rond ou ovale de calcifications amorphes et peu nombreuses. [CHE12]



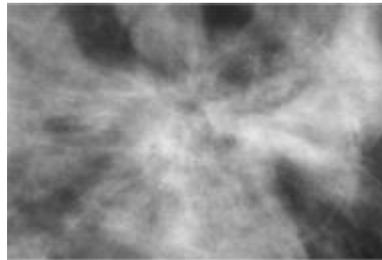
**Figure III-9** : Masses bien circonscrites discrètement lobulées, sans microlobulation, non calcifiées, non typiquement liquidiennes en échographie : BIRADS 3.

### III.6.2.5. BI-RADS 4

Selon BI-RADS 4, il y a une probabilité de 23 % à 34 % que les résultats de la mammographie soient un cancer du sein et que les données soient inquiétantes. Nécessitera une biopsie afin d'obtenir un petit échantillon de tissu pour évaluation. [EBE06]

Les anomalies entrant dans cette catégorie sont les suivantes :

- Présence d'une masse avec des projections en forme de spicules et sans centre dense.
- Masse de forme ronde ou ovale avec des contours lobulés ou masqués.
- Abondance de Mcs pulvérulentes, groupées en petits amas.
- Présence de Mcs irrégulières, polymorphes ou granulaires, mais en petit nombre.
- Abondance de Mcs punctiformes régulières, groupées en petits amas avec des contours ni ronds ni ovales. [CHE12]



**Figure III-10 :** Images spiculée sans centre dense.



**Figure III-11 :** Microcalcifications granulaires irrégulières peu nombreuses : BIRADS 4.

### III.6.2.6. BI-RADS 5

Un résultat de catégorie 5 est considéré comme extrêmement suspect et ayant une probabilité de 95 % d'être un cancer du sein. Il est conseillé de prendre des mesures immédiates, comme passer un test ou consulter un chirurgien. [EBE06]

Il s'agit de [CHE12] :

- Masse avec des projections en forme de spicules et un centre dense.
- Masse avec un contour flou et irrégulier.
- Mcs associées à une anomalie architecturale ou à une opacité.

- Groupement de Mcs , quelle que soit leur morphologie, dont la topographie suit les canaux galactophoriques.
- Mcs vermiculaires, arborescentes ou irrégulières polymorphes ou granulaires, nombreuses et groupées.
- Groupement de Mcs ayant augmenté en nombre ou Mcs dont la morphologie et la distribution sont devenues plus suspectes.

### **III.6.2.7. BI-RADS 6**

La catégorie 6 est un cancer confirmé avec une biopsie positive. Lorsqu'un cancer a déjà été identifié et que des soins supplémentaires, comme la planification d'un traitement, sont nécessaires, il appartient à ce groupe. [EBE06]

Pour savoir s'il y a une tumeur maligne, une biopsie mammaire est effectuée pour acquérir un échantillon de tissu mammaire. Une "biopsie chirurgicale" peut être nécessaire, ou la biopsie peut être effectuée en utilisant une méthode de seringue (une "biopsie à l'aiguille"). Si une biopsie à l'aiguille est une possibilité, c'est généralement ce qui est choisi. [EBE06]

## **III.7. Les systèmes de Détection/Diagnostic Assistés par Ordinateur (DAO)**

Les premières recherches sur les systèmes de traitement automatique des images de mammographie visaient à donner aux radiologues une seconde interprétation pour identifier et diagnostiquer plus rapidement les lésions malignes, qu'il s'agisse de tumeurs ou de microcalcifications. Ces systèmes sont appelés systèmes d'aide à la détection/diagnostic par opérateur (DAO). Afin de trouver et de localiser des lésions sur des clichés mammographiques, des systèmes dits d'assistance à la détection automatique (DAOe) sont utilisés. Le terme " Diagnostic Assistée par Ordinateur " (DAOx) fait référence à un système complet de traitement des images de mammographie, du prétraitement à la catégorisation et à la prise de décision. Le succès de ces systèmes est le résultat de leur efficacité, de leur cohérence et de leur capacité à proposer des méthodes fiables d'aide à la détection ou à la reconnaissance des lésions mammaires. [CHE12]

### **III.7.1. Détection Assistée par Ordinateur (DAOe)**

Les applications actuelles des systèmes de détection assistée par ordinateur pour la mammographie concernent la mammographie diagnostique ; cependant, leur utilisation dans la mammographie de dépistage n'a pas encore été déterminée. Ciato et al, [CIA03] ont

comparé les résultats de la lecture des images mammographiques par les radiologues aux résultats obtenus par le système DAOe dans le cadre d'un examen national de compétence pour la mammographie de dépistage en Italie. Les auteurs sont arrivés à la conclusion qu'une lecture unique utilisant un système DAOe produit des résultats de lecture comparables à la lecture par deux radiologues. [CHE12]

### **III.7.2. Diagnostic Assistée par Ordinateur (DAOx)**

La catégorisation biaisée des lésions mammaires par un radiologue peut différer d'un radiologue à l'autre pour une même lésion. Cependant, un système automatique d'Aide au Diagnostic par Ordonnateur (DAOx) qui base ses résultats sur les caractéristiques ou les descriptions de la lésion produit systématiquement les mêmes résultats pour une même image et peut donc servir de référence. Un système automatique d'Aide au Diagnostic par Ordonnateur (DAOx) est destiné à donner un second avis au radiologue. Ce système commence généralement par un prétraitement pour améliorer la qualité de l'image et réduire le bruit. Cette étape vise à supprimer l'effet de la diffusion qui est fréquemment visible sur les images radiologiques et échographiques. La tâche de détection d'anomalies de toutes sortes (image normale, présence de masse(s), présence de microcalcification(s), présence des deux catégories d'anomalies) est facilitée par ce prétraitement. L'étape suivante consiste à dériver des descripteurs mathématiques qui capturent le mieux les caractéristiques de ces lésions. Enfin, une étape de catégorisation et de prise de décision basée sur des algorithmes appropriés est adoptée. [CHE12]

### **III.7.3. Métriques d'évaluation des CAD**

L'implémentation de métriques d'évaluation dans les tâches DL est essentielle pour créer un classifieur optimal. Ils sont utilisés au cours de deux étapes clés d'un processus typique de catégorisation des données : la formation et les tests. Pendant la phase d'apprentissage, il est utilisé pour améliorer l'algorithme de classification. Cela indique que la mesure d'évaluation est utilisée pour faire une distinction et choisir la meilleure solution. Actuellement, l'efficacité du classificateur construit est évaluée à l'aide de la métrique d'évaluation, par exemple, en tant qu'évaluateur lors de l'étape de test du modèle utilisant des données masquées. Le nombre d'instances négatives et positives effectivement catégorisées, respectivement, est défini comme TN et TP. La quantité d'occurrences positives et négatives incorrectement identifiées est également définie comme FN et FP, respectivement. [ALZ21]

### III.7.3.1. Matrice de confusion

Une matrice de confusion contient des détails sur les classifications réelles et prévues qu'un système de classification a établies. Les données de la matrice sont régulièrement analysées pour évaluer l'efficacité de ces modèles. Le tableau ci-dessous affiche la matrice de confusion pour un classificateur à deux classes. [SAN12]

	Prédiction positive	Prédiction négative
Classe positive	Vrai positif	Faux positif
Classe négative	Faux négatif	Vrai négatif

**Tableau III-2** : matrice de confusion. [MON03]

**Vrai positif (VP)** : est le pourcentage de cas positifs correctement classés comme appartenant à la classe positive.

**Faux positif (FP)** : est le pourcentage de cas négatifs mal classés comme appartenant à la classe positive.

**Faux négatif (FN)** : est le pourcentage de cas positifs mal classés comme appartenant à la classe négative.

**Vrai négatif (VN)** : est le pourcentage de cas négatifs correctement classés comme appartenant à la classe négative. [MON03]

Maintenant, nous allons passer à discuter les critères de performances des systèmes d'aide au diagnostic comme le taux de reconnaissance (accuracy), la sensibilité (également connu sous le nom rappel), de la spécificité et de F1-score, ces métriques sont importantes pour évaluer la performance des modèles de classification :

### III.7.3.2. Le taux de reconnaissance

Calcule le pourcentage de classes correctement prédites pour tous les échantillons évalués.

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN} \quad \text{Équation III-1}$$

### III.7.3.3. Sensibilité ou rappel

Utilisé pour calculer la fraction de modèles positifs qui sont correctement classés.

$$Sensibilité = \frac{VP}{VP + FN} \quad \text{Équation III-2}$$

### III.7.3.4. Spécificité

Utilisé pour déterminer le pourcentage de modèles négatifs correctement catégorisés.

$$\text{Spécificité} = \frac{VN}{FN + VN} \quad \text{Équation III-3}$$

### III.7.3.5. La Précision

Utilisée pour déterminer les modèles positifs qui peuvent être correctement prédits par tous les autres modèles d'une classe positive.

$$\text{Précision} = \frac{VP}{VP + FP} \quad \text{Équation III-4}$$

### III.7.3.6. F1-Score

Fait une moyenne harmonique des taux de rappel et de précision.

$$F1 - \text{Score} = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \quad \text{Équation III-5}$$

## III.8. Revue de littérature

Dans ce contexte, nous avons entrepris des travaux portant sur la classification des images mammographiques selon leur densité. Pour ce faire, nous exposons cinq articles récents issus de la littérature scientifique afin d'en améliorer la compréhension.

### III.8.1. Une méthode d'apprentissage profond pour classer les catégories de densité mammographique du sein

L'étude de Mohamed et al. [MOH18] utilise un modèle basé sur l'apprentissage en profondeur (un réseau de neurones convolutifs (CNN)) pour la classification de la densité mammaire en deux catégories : "densité dispersée" et "densité hétérogène". Les chercheurs ont recueilli un vaste ensemble de données d'imagerie mammographique. Toutes les images ont été collectées à partir d'une cohorte de 1 427 femmes ayant subi un dépistage par mammographie numérique standard entre 2005 et 2016. Lors de la formation du CNN, ils ont obtenu une précision de 94%.

### III.8.2. Segmentation et classification entièrement automatisées de la densité mammaire à l'aide de l'apprentissage profond

L'étude de Saffari et al. [SAF20] vise à développer une méthode de segmentation et de classification des tissus mammaires entièrement automatisée et numérisée à l'aide de

techniques avancées d'apprentissage en profondeur. La méthode proposée utilise des réseaux antagonistes génératifs conditionnels (cGAN) pour la segmentation des tissus denses et un réseau neuronal convolutif (CNN) pour la classification de la densité mammaire basée sur la normalisation du système de rapport et de données d'imagerie mammaire (BI-RADS). L'étude a utilisé 410 images de 115 patients de l'ensemble de données INbreast pour le dépistage par mammographie. Le cadre proposé a atteint une précision, un coefficient de dés et un indice de Jaccard de 98 %, 88 % et 78 %, respectivement, pour la segmentation de la densité mammaire, et une précision, une sensibilité et une spécificité de 97,85 %, 97,85 % et 99,28 %, respectivement, pour la classification de la densité mammaire.

### **III.8.3. Évaluation de la densité mammographique du sein à l'aide de l'apprentissage profond : mise en œuvre clinique**

Dans une étude menée par Lehman et al. [LEH18], des méthodes de réseau de neurones à convolution profonde (DCNN) ont été utilisées pour la classification de la densité mammaire. L'étude a porté sur 41 479 mammographies numériques de dépistage obtenues auprès de 27 684 femmes entre janvier 2009 et mai 2011. Les images ont été classées en deux catégories : denses et non denses. L'étude a atteint une précision de 87 % dans la classification de la densité mammaire.

### **III.8.4. Estimation automatisée de la densité mammographique à l'aide d'un réseau entièrement convolutif**

De plus, Lee et Nishikawa [LEE18] ont développé un algorithme utilisant un réseau entièrement convolutif, pour segmenter à la fois le sein et les zones fibroglandulaires denses sur les images mammographiques. Ils ont utilisé 455 images, puis ils ont affiné un réseau d'apprentissage en profondeur pré-formé (VGG 16) pour entraîner le réseau à segmenter à la fois le sein et les zones fibroglandulaires. Leur algorithme a montré une forte corrélation avec les cotes de densité BI-RADS données par les radiologues. Plus précisément, 0,81 pour la vue CC, 0,79 pour la vue MLO et 0,85 pour la moyenne CC-MLO.

### **III.8.5. Détermination de la densité mammographique du sein à l'aide d'un réseau neuronal convolutionnel profond**

Ciritsis et al. [CRI19] Ont mené une étude dans laquelle ils ont utilisé un réseau de neurones à convolution profonde avec 11 couches convolutives et 3 couches entièrement connectées pour la classification de la densité mammaire. Ils ont formé et validé le modèle sur

un jeu de données augmenté et l'ont testé sur deux jeux de données différents. Le premier jeu de test comprend 850 MLO et 880 CC, le deuxième jeu de test comprend 100 MLO, 100 CC. L'algorithme a montré une conformité avec les rapports radiologiques correspondants de 71,7 % pour les projections MLO et de 71,0 % pour les projections CC lorsqu'il a été testé sur le premier ensemble de données de test. Cependant, l'accord s'est amélioré pour différencier les seins denses et gras pour les deux projections (MLO = 88,6 % et CC = 89,9 %). Dans le deuxième ensemble de données de test, l'accord a atteint 99 % pour les projections MLO et 96 % pour les projections CC pour différencier les seins gras et denses, respectivement.

### **III.8.6. Classification de la densité mammographique à l'aide d'informations d'histogramme multirésolution**

Dans cet article, Muhimmah et al. [MUH06] proposent une approche d'histogramme multi-résolution pour classer la densité mammaire à l'aide d'informations sur la texture et les intensités., ils ont utilisé un classificateur DAG-SVM, cette technique a produit un accord de 77,57 %.

## **III.9. Conclusion**

Dans ce chapitre, nous avons introduit quelques notions générales concernant l'anatomie et le cancer de sein, ainsi que les méthodes d'imagerie médicale utilisées pour un dépistage précoce, en mettant l'accent sur la mammographie. Les pathologies mammaires ont été présentées en examinant leur forme, leur contour et leur densité, et les différentes techniques de classification utilisées par les radiologues. Nous avons présenté aussi quelques systèmes d'aide au diagnostic pour la classification de la densité mammaire. Ces systèmes d'aide au diagnostic permettent d'améliorer l'efficacité et la précision de l'interprétation des anomalies mammaires.

Dans le prochain chapitre nous présenterons notre travail qui ne permettra de définir un système d'aide au diagnostic pour la classification des densités des mammographies en se basant sur la norme BI-RADS.

# Chapitre

---

# 4

# CLASSIFICATION DE LA DENSITE MAMMOGRAPHIQUE

## **IV.1. Introduction**

Avec la difficulté de réaliser des systèmes d'aide au diagnostic permettant de classer les densités des mammographies en se basant sur les bases de données disponibles dans le domaine médical en raison de leur texture complexe, de leur taille limitée, de leur variabilité et le manque des étiquetage dans plusieurs bases mammographiques. Il existe un fort besoin de concevoir des architectures simples et puissantes capables de gérer avec succès le problème de la classification de la densité des mammographies selon la norme BI-RADS. Pour atteindre cet objectif, nous présentons dans ce chapitre l'architecture du système basé sur l'apprentissage profond. Ce système permet de réaliser des prédictions en se basant sur la base de données mammographique de RSNA (the Radiological Society of North America) afin d'automatiser la tâche de classification de la densité mammographique.

## **IV.2. Environnement de développement de système**

Ce projet a été développé dans les conditions suivantes

### **IV.2.1. Environnement matériel**

Notre travail est réalisé sur la plateforme de Kaagle [KAG23], cette plateforme web héberge la plus grande communauté scientifique au monde. Il reçoit près de 150 000 soumissions chaque mois et offre à la communauté scientifique des outils et des ressources puissants pour les aider à réaliser toutes les avancées de la science des données. Semblable à Datascientest, Kaggle fournit un environnement Jupyter Notebooks personnalisable et non configuré. Un accès GPU gratuit est disponible, ainsi qu'une quantité substantielle de données et de code publiés par la communauté. Vous pouvez réaliser tous vos projets de science des données à travers Kaggle, qui contient tous les outils et données nécessaires. Plus de 500 000 bases de données publiques et 400 000 revues accessibles publiquement. Dans cette plateforme les utilisateurs peuvent :

- ✓ Trouver et publier des bases de données
- ✓ Explorer et créer des modèles dans un environnement en ligne adapté.
- ✓ Travailler avec d'autres professionnels et passionnés.
- ✓ Faire des concours pour se tester sur des sujets innovants.

### **IV.2.2. Langage de programmation et bibliothèque utilisée**

Le langage de programmation et les bibliothèques utilisés sont les suivants :

### **IV.2.2.1. Langage Python**

Les professionnels de l'informatique utilisent le langage de programmation open source le plus couramment utilisé, Python [PYT20]. Ce langage a été créé en pensant à la gestion de l'infrastructure, à l'analyse de données ou au développement de logiciels. L'un des avantages de Python est qu'il permet aux développeurs de se concentrer davantage sur ce qu'ils font que sur la façon dont ils le font. Il a libéré les développeurs des contraintes des formulaires des systèmes linguistiques plus anciens. En conséquence, écrire du code en Python est plus rapide que d'écrire du code dans d'autres langages.

### **IV.2.2.2. Tensorflow**

Tensorflow [TEN20] est un framework open source complet pour la création d'applications d'apprentissage automatique. Il s'agit d'un outil mathématique symbolique qui effectue diverses opérations ciblées sur la formation et l'inférence de réseaux de neurones profonds à l'aide de flux de données et de programmation différentiable. Il permet aux programmeurs de créer des applications d'apprentissage automatique en utilisant une gamme d'instruments, de cadres et d'actifs communautaires.

Tensor Flow de Google est actuellement le framework d'apprentissage en profondeur le plus connu au monde. Tous les produits de Google intègrent l'apprentissage automatique pour améliorer le moteur de recherche, la traduction, les légendes d'images ou les suggestions.

### **IV.2.2.3. Keras**

Keras [KER20] est une bibliothèque Python open-source basée sur les travaux du développeur Google François Chollet dans le cadre du projet Oneiros, qui est un système d'exploitation de robot intelligent neuro-électronique ouvert. Le logiciel multiplateforme a été publié pour la première fois le 28 mars 2015. Cette bibliothèque a pour but de faciliter la formation rapide de réseaux neuronaux. Dans ce contexte, Keras fonctionne comme une interface de programmation d'application (API) pour accéder et programmer divers frameworks d'apprentissage automatique plutôt qu'un framework autonome. Les outils de Keras incluent Theano, Tensor Flow et Microsoft Cognitive Toolkit (anciennement CNTK).

### **IV.2.2.4. Numpy**

Le nom NumPy [NUM23] signifie en fait "Python numérique". Une bibliothèque open source Python est en cours de discussion. On utilise cet outil pour la programmation

scientifique Python, en particulier pour la programmation en science des données, en ingénierie, en mathématiques ou en sciences.

La science des données repose sur des calculs scientifiques extrêmement compliqués. Les Data Scientists ont besoin d'outils puissants pour effectuer ces calculs. Cette bibliothèque est très utile pour effectuer des opérations statistiques et mathématiques en Python. Lors de la multiplication de matrices ou de tableaux multidimensionnels, elle fonctionne à merveille. L'intégration avec C/C++ et Fortran est très simple.

### **IV.2.2.5. Matplotlib**

Pour Python et son extension numérique NumPy, Matplotlib [MAT23] est un outil d'affichage de données et de traçage graphique multiplateforme. En conséquence, il présente un solide substitut open source pour MATLAB. Les API (Application Programming Interfaces) pour matplotlib permettent aux programmeurs d'incorporer des graphiques dans les applications GUI.

### **IV.2.2.6. PyDicom**

PyDicom [OMA23] est un package Python pur pour traiter les fichiers DICOM, y compris les images médicales, les rapports et les objets de radiothérapie. PyDicom simplifie la lecture et l'écriture de ces fichiers complexes dans des structures pythoniques organiques pour une manipulation simple.

### **IV.2.2.7. Pandas**

L'outil Python open source le plus fréquemment utilisé pour les activités de science des données, d'analyse de données et d'apprentissage automatique s'appelle Pandas [PAN23]. Il est construit au-dessus de Numpy, un programme différent qui prend en charge les tableaux multidimensionnels. Pandas, l'un des packages de gestion de données les plus connus, est fréquemment inclus dans toutes les distributions Python et fonctionne bien avec de nombreux autres modules de science des données au sein de l'écosystème Python. Pandas simplifie la réalisation de nombreuses tâches chronophages et répétitives associées à l'utilisation des données, telles que:

- Purge et saisie des données.
- Normalisation des données.
- Illustration de données.
- Évaluation statistique et analyse des données.

### IV.2.2.8. Sklearn

La bibliothèque d'apprentissage automatique Python la plus puissante et la plus fiable s'appelle Scikit-learn [JVC22], parfois connue sous le nom de Sklearn. Elle propose une variété d'outils efficaces pour l'apprentissage et la modélisation statistiques automatiques, y compris la classification, la régression et le clustering, via une interface utilisateur Python cohérente. Cette bibliothèque, qui est en grande partie écrite en Python, est construite sur NumPy, SciPy et Matplotlib.

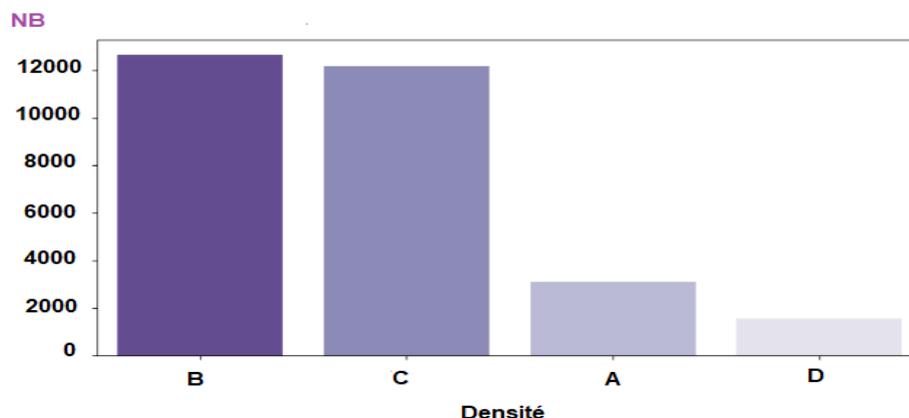
### IV.2.2.9. Tkinter

Tkinter (interface Tk) [TIK17] est un module Python inclus dans la bibliothèque standard du langage et vous permet de créer des interfaces utilisateur graphiques avec des fenêtres, des widgets et des événements tels que les pressions sur le clavier et les mouvements du curseur.

Parce que Tkinter fonctionne sous Windows et la majorité des systèmes Unix, les interfaces créées avec lui sont portables.

## IV.3. La base de données utilisée

Dans cette étude, nous visons à expérimenter des techniques basées sur l'apprentissage en profondeur pour la classification des densités dans les mammographies. Pour cela, nous utilisons la base de données fournie par la Radiological Society of North America (RSNA) disponible en ligne sur la plateforme Kaggle [RSN23], cette base contient 54 706 images provenant de 11 913 patients, où chaque patient peut avoir au moins 4 images mammographiques au format DICOM. Cependant, pour notre étude, nous avons utilisé uniquement 24350 images en raison du manque d'étiquetage de la densité des mammographies.



**Figure IV-1** : distribution des classes de densité dans la base de données RSNA.

### IV.4. Architectures générale de travail

L'architecture générale de notre travail englobe deux phases principales ; dans la première étape nous effectuons quelques prétraitements sur les images mammographiques, tels que la normalisation, suppression des annotations et le redimensionnement. Tandis que la deuxième étape consiste à utiliser l'apprentissage par transfert à travers quatre modèles connus, le meilleur modèle selon les critères d'évaluation sera élu pour la classification des densités des mammographies.

L'interface graphique du modèle finale a été développée en utilisant la bibliothèque Tkinter.

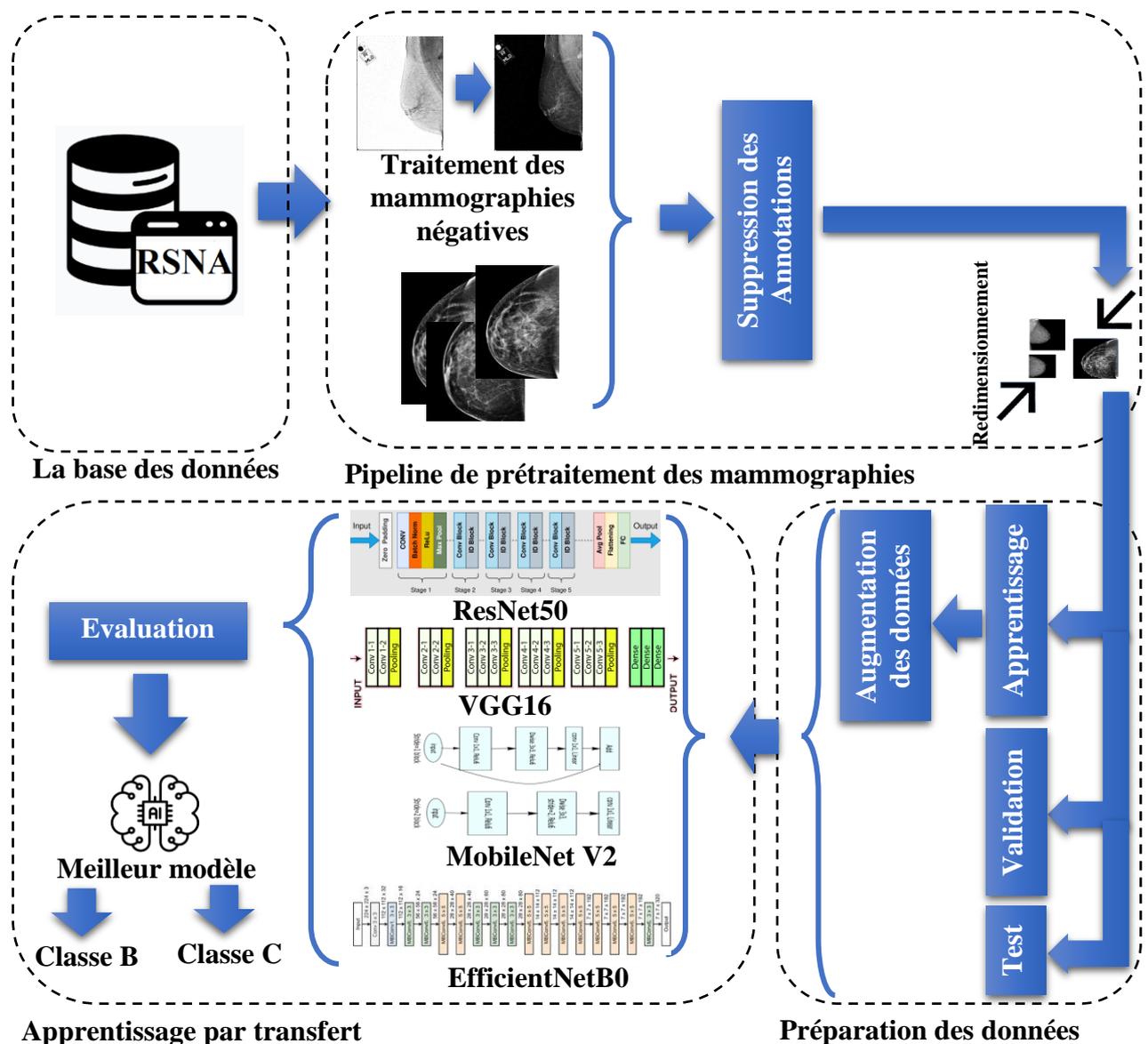


Figure IV-2 : L'Architecture générale de travail.

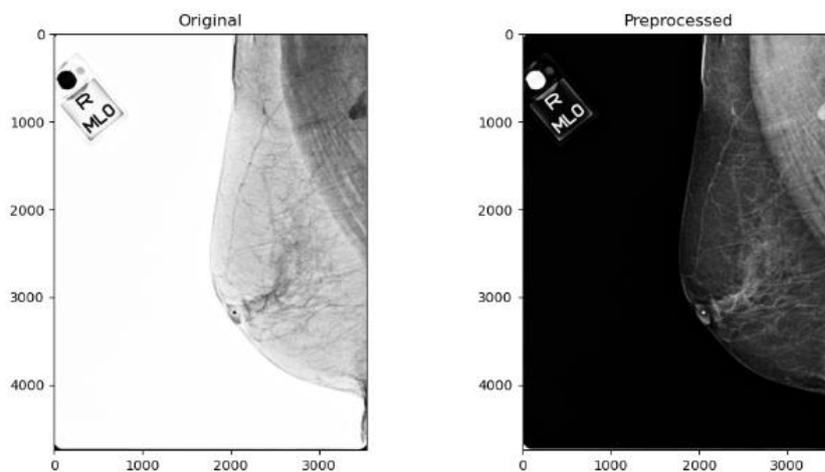
### IV.4.1. Prétraitement des mammographies

Dans cette étape nous avons effectué certaines étapes de prétraitement sur les images qui ont une densité B et C uniquement car le nombre de mammographies ayant la densité A ou D est très faible dans les bases de données disponible actuellement. Les prétraitements effectués sont: la normalisation, la suppression des annotations et le redimensionnement.

Les images de la base de données finale seront enregistrées sous forme de fichiers PNG et organisées dans les deux catégories B et C.

#### IV.4.1.1. Normalisation des mammographies

Dans cette phase deux méthode de prétraitement ont été effectuées, la première consiste à calculer l'image négative pour certain mammographies ayant un fond blanc pour rendre toutes les mammographies uniforme la Figure IV-3 présente un exemple d'une mammographie négative.



**Figure IV-3 :** traitement des mammographies négatives.

A cause de la variabilité des niveaux de gris dans les fichiers DICOM, la deuxième méthode consiste à normaliser les pixels des images mammographiques dans une plage spécifique. La fonction utiliser pour effectuer cette opération est la suivante :

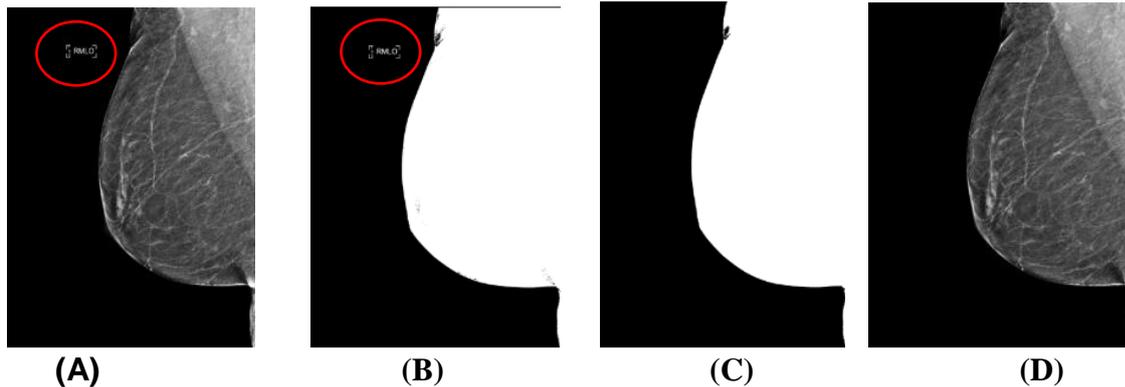
$$\text{Image} = (\text{image} - \text{np.min}(\text{image})) / (\text{np.max}(\text{image}) - \text{np.min}(\text{image})) \quad \text{Équation IV-1}$$

Le but de cette phase est de faciliter le processus d'entraînement des modèles utilisées.

#### IV.4.1.2. Suppression des annotations

Il a été remarqué que les annotations et les marqueurs de texte dans les images mammographiques sont mal placés et gêne le processus de classification des densités des mammographies. Nous suivons donc un processus en plusieurs étapes pour résoudre ce

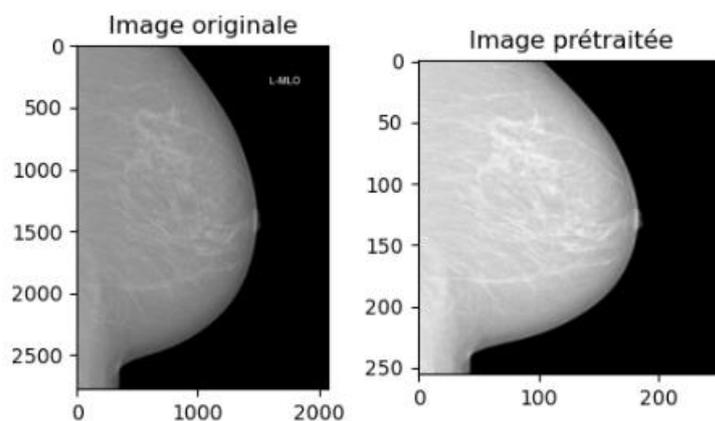
problème : D'abord, nous convertissons l'image en niveaux de gris et nous appliquons une binarisation avec un seuil = 20, le seuil a été choisi après une série d'expérimentation pour séparer l'organe, l'étiquette et le fond de la mammographie. Ensuite, nous choisissons le contour le plus grand parmi tous les contours extraits ce contour représente les bords de la glande mammaire. Finalement, nous utilisons ce contour pour créer un masque qui efface les étiquettes lorsqu'il est combiné avec l'image originale. La figure suivante présente les différentes étapes de ce processus.



**Figure IV-4 :** exemple de suppression des annotations. (A : image original avec étiquette, B : image binaire avec étiquette, C : image avec mask , D : image prétraité).

#### IV.4.1.3. Redimensionnement des mammographies

La variabilité des tailles des mammographies peuvent rendre leur traitement automatique difficile. Pour résoudre ce problème, nous avons redimensionné les images prétraitées à une taille de 256x256 pixels, la taille fixe permet de réduire la complexité du modèle.



**Figure IV-5 :** Exemple de redimensionnement.

Nous avons enregistré les images prétraitées au format PNG après les avoir redimensionnées car elles étaient à l'origine au format DICOM. Maintenant, nous pouvons gérer plus facilement notre base de données.

## IV.4.2. Préparation des données

La préparation des données dans notre travail comprend deux phases :

### IV.4.2.1. Partition des données

Notre base de données a été partagée en trois sections : entraînement, validation et test. L'allocation était de 80%, 10% et 10% respectivement. Le tableau suivant présente le nombre des images dans chaque section :

base de données	Nombre d'images
Entraînement	19 480
Validation	2 434
Test	2436

**Tableau IV-1:** Récapitulatif des données d'entraînement, validation et de test.

#### IV.4.2.1.1. La base d'apprentissage

Nous avons utilisé un total de 19 480 images prétraitées, 9740 images de la classe B et 9740 de la classe C. avec l'utilisation des techniques d'augmentation de données simples, pour développer plus rapidement des modèles plus précis.

#### IV.4.2.1.2. La base de validation

Nous avons utilisé 1217 images dans chaque classe B et C pour évaluer les performances du modèle, mais notre utilisation réelle de la base de validation est pour ajuster les hyperparamètres.

#### IV.4.2.1.3. La base de test

Dans cette base, nous avons utilisé un total de 2 436 images. Une fois le modèle entraîné et validé, nous avons évalué les performances de ce dernier avec l'utilisation des métriques comme le taux de reconnaissance, la précision et le score F1.

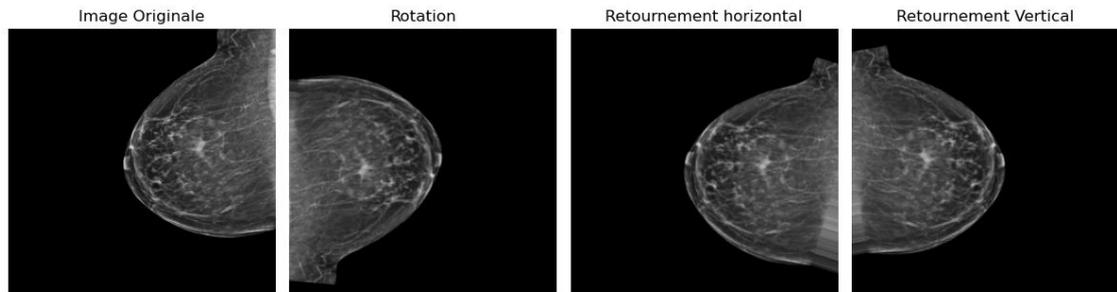
### IV.4.2.2. Augmentation de la base des données

Nous avons également utilisé l'augmentation des données seulement pour la base d'apprentissage pour générer de nouvelles images :

**Rotation** : Les images sont légèrement tournées dans une plage de  $\pm 20$  degrés. Cela permet au modèle d'apprendre à reconnaître les lésions dans des orientations différentes.

**Retournement horizontal** : Les images peuvent être retournées horizontalement de manière aléatoire. Cela augmente la variabilité des données et permet au modèle de mieux généraliser.

**Retournement vertical** : Les images peuvent également être retournées verticalement de manière aléatoire. Cela ajoute une autre forme de variation aux données d'entraînement.



**Figure IV-6** : Exemple des 3 techniques d'augmentations appliquées sur une mammographie de classe B.

### IV.4.3. Apprentissage par transfert

Nous avons utilisé 4 modèles du réseau de neurones convolutifs profonds VGG 16, MobileNet\_V3, Resnet50 et EfficientNet\_B0 avec des modifications simples sur l'architecture et les mêmes paramètres.

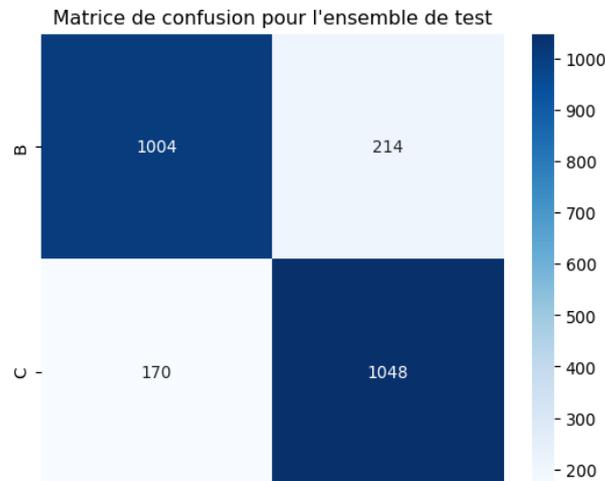
Nous avons ajouté une couche Global Average Pooling 2D pour réduire les dimensions des caractéristiques extraites et capturer les informations les plus importantes. Une couche dense avec 256 neurones avec la fonction d'activation relu, et pour éviter le sur apprentissage, nous avons inclus une couche Dropout avec un taux de 0,5. Enfin, la couche de sortie utilise une fonction d'activation sigmoïde pour la classification binaire "B" ou "C".

Pour assurer une exécution efficace du modèle, nous avons pris soin de suivre les bonnes pratiques requises pour mener notre étude. Nous avons ainsi utilisé Adam comme un optimiseur, réglant le taux d'apprentissage à 0,001 pour atteindre des performances optimales.

Dans la classification binaire, il est habituel d'utiliser la fonction de perte entropie croisée binaire, c'est celle que nous avons employée ici également. Afin d'affiner notre apprentissage avec intuition tout en garantissant les meilleures pratiques, nous avons pris soin du processus étape par étape, notamment en employant des techniques telles que l'arrêt précoce où le temps minimum était fixé à 5 itérations consécutives avant arrêt précoce. Grâce à cette méthodologie rigoureuse, nous avons réussi à réaliser notre objectif initial qui était celui d'identifier le meilleur modèle possible.

## IV.5. Les résultats

En outre, nous avons utilisé une matrice de confusion pour évaluer les performances de nos modèles. Voici les résultats de la matrice de confusion :



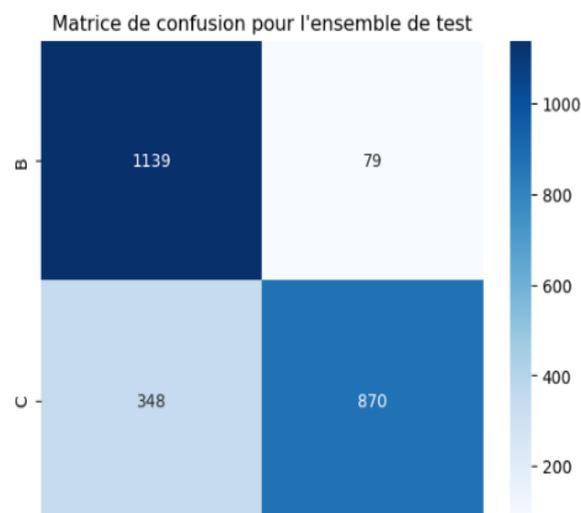
**Figure IV-7 :** Matrice de confusion du transfert d'apprentissage du modèle VGG 16.

Vrai Positif (VP): 1004 (le modèle a correctement prédit la classe de densité B).

Faux Positif (FP): 214 (le modèle a prédit la classe de densité C, mais la classe réelle était B).

Faux Négatif (FN): 170 (le modèle a prédit la classe de densité B, mais la classe réelle est C).

Vrai Négatif (VN) : 1048 (le modèle a correctement prédit la classe de densité C).



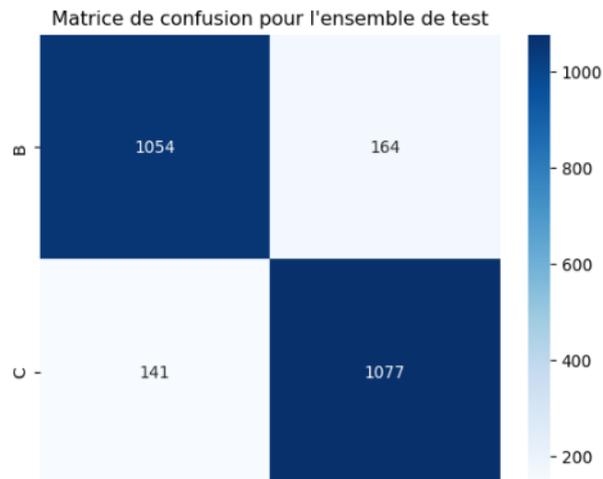
**Figure IV-8 :** Matrice de confusion du transfert d'apprentissage du modèle ResNet 50.

Vrai Positif (VP) : 1139 (le modèle a correctement prédit la classe de densité B).

Faux Positif (FP) : 79 (le modèle a prédit la classe de densité C, mais la classe réelle était B).

Faux Négatif (FN) : 348 (le modèle a prédit la classe de densité B, mais la classe réelle est C).

Vrai Négatif (VN) : 870 (le modèle a correctement prédit la classe de densité C).



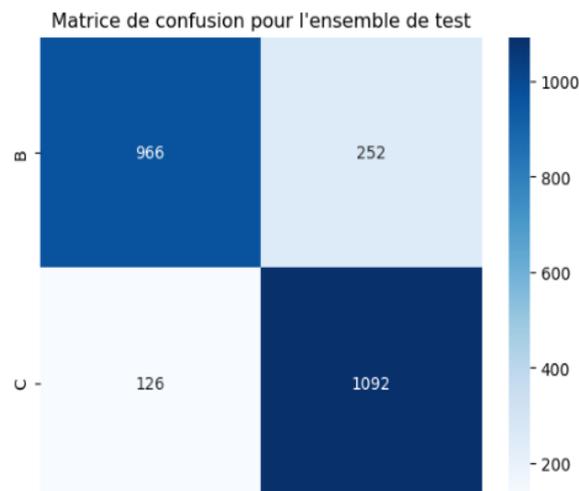
**Figure IV-9 :** Matrice de confusion du transfert d'apprentissage du modèle EfficientNetB0.

Vrai Positif (VP) : 1054 (le modèle a correctement prédit la classe de densité B).

Faux Positif (FP) : 164 (le modèle a prédit la classe de densité C, mais la classe réelle était B).

Faux Négatif (FN) : 141 (le modèle a prédit la classe de densité B, mais la classe est était C).

Vrai Négatif (VN) : 1077 (le modèle a correctement prédit la classe de densité C).



**Figure IV-10 :** Matrice de confusion du transfert d'apprentissage du modèle MobileNet V2.

Vrai Positif (VP) : 966 (le modèle a correctement prédit la classe de densité B).

Faux Positif (FP) : 252 (le modèle a prédit la classe de densité C, mais la classe réelle était B).

Faux Négatif (FN) : 126 (le modèle a prédit la classe de densité B, mais la classe réelle est C).

Vrai Négatif (VN) : 1092 (le modèle a correctement prédit la classe de densité C).

Il était remarqué que la plupart des matrices de confusion montrent un plus grand nombre de vrais positifs et de vrais négatifs par rapport aux faux positifs et aux faux négatifs. Cela suggère que les modèles réussissent souvent à identifier avec précision les images de mammographie selon leur densité.

En termes de métriques d'évaluation supplémentaires, voici les scores de précision, de rappel et de f1-score pour chaque model :

<b>Métriques</b> <b>Modèle</b>	<b>Précision</b>	<b>Rappel</b>	<b>Score F1</b>
<b>VGG 16</b>	82%	85%	82%
<b>Resnet 50</b>	93%	76%	82%
<b>EfficientNetB0</b>	86%	88%	86%
<b>MobileNet V2</b>	79%	88%	82%

**Tableau IV-2:** métriques d'évaluation supplémentaires.

Les métriques fournissent des données utiles sur la performance des modèles. Le score de précision impressionnant de 93 % de ResNet 50 démontre sa capacité à réduire les fausses prédictions positives. Cependant, son score de rappel est légèrement inférieur de 76 %, ce qui suggère qu'il peut avoir manqué certains cas vraiment positifs. Cependant, EfficientNetB0 montre un bon équilibre des performances avec une précision de 86 %, un rappel élevé de 88 % et un score F1 de 86 %. Cela met en évidence son efficacité dans la prédiction précise des cas positifs et la collecte de nombreux cas positifs réels. En termes de performances, VGG 16 et MobileNet V2 sont comparables, atteignant une précision et des scores F1 de 82 %. Notamment, MobileNet V2 excelle dans le rappel avec un score de 88 %, démontrant sa capacité à reconnaître les véritables cas positifs. Dans l'ensemble, EfficientNetB0 semble être le modèle le plus efficace, affichant une précision, un rappel et un score F1 équilibrés pour toutes les mesures.

Lors de notre recherche visant à classer les images de mammographie selon leur densité B ou C, nous avons testé quatre modèles différents. Après avoir analysé de nombreux résultats, nous avons constaté que le modèle EfficientNetB0 s'est révélé être le plus performant parmi tous.

Ce modèle a vraiment fait preuve d'une efficacité remarquable, affichant une précision exceptionnelle ainsi que de bonnes mesures d'évaluation. Sa capacité à distinguer avec précision les différentes densités de tissus mammaires est impressionnante. Ces résultats sont particulièrement prometteurs dans le domaine médical, où une telle précision peut jouer un rôle crucial dans le diagnostic des anomalies mammaires.

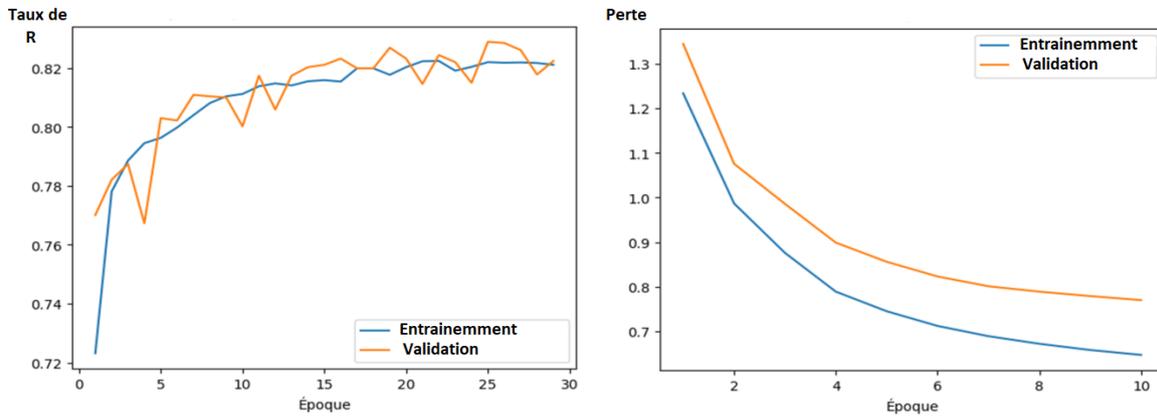
En optant pour le modèle EfficientNetB0, nous avons pu obtenir des performances supérieures dans notre tâche de classification des images de mammographie en fonction de leur densité. Ces résultats positifs renforcent notre confiance dans l'utilisation de ce modèle pour des applications similaires, et soulignent l'importance de poursuivre la recherche et le développement de techniques de pointe dans le domaine médical.

Nous pouvons résumer les résultats dans le tableau suivant:

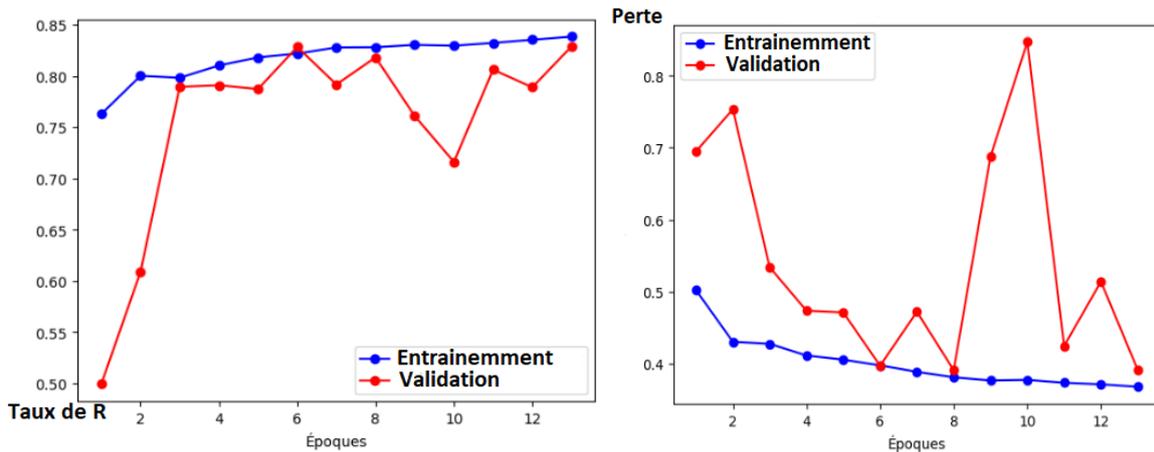
Modèle	Taux d'apprentissage		
	Entraînement	Validation	Test
VGG 16	82,18%	82,85%	84,24%
ResNet 50	82,82%	81,83%	82,47%
EfficientNetB0	85,64%	86,35%	87,48%
MobileNe V2	83,64%	82,57%	84,48%

**Tableau IV-3:** Les résultats obtenus.

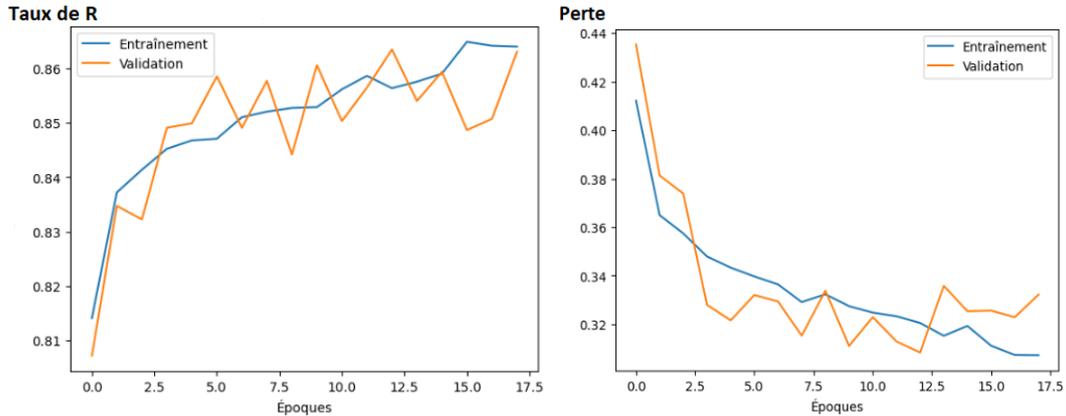
Les figures suivantes représentent le taux de reconnaissance (Taux de R) et la perte (erreur) pour les bases d'apprentissage (Entraînement) et de validation :



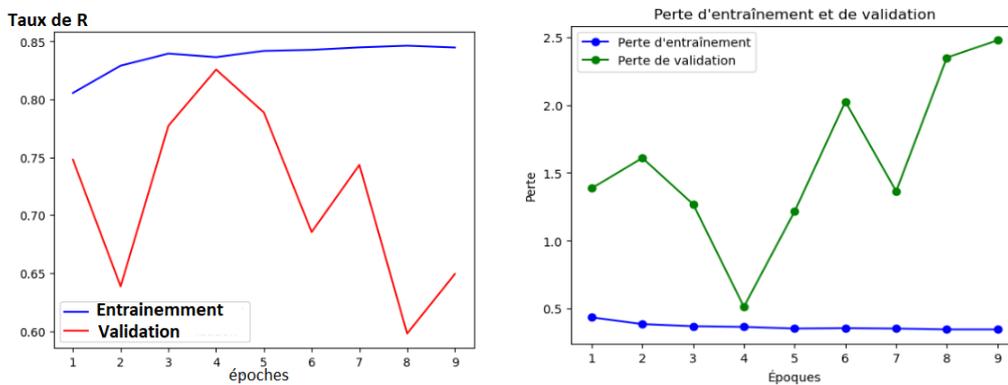
**Figure IV-11 :** Evolution des taux d'apprentissage et erreurs de VGG 16.



**Figure IV-12 :** Evolution des taux d'apprentissage et erreurs de RESNET 50.



**Figure IV-13 :** Evolution des taux d'apprentissage et erreurs d'EfficientNetB0.



**Figure IV-14 :** Evolution des taux d'apprentissage et erreurs de MobileNet V2.

Après avoir vu les chiffres et le tableau des résultats des quatre modèles, il était observé que le taux de reconnaissance du test dans tous les cas est légèrement supérieure à le taux de reconnaissance de l'entraînement et de la validation. Il y a peu de différences entre ces valeurs.

Nous avons également vu que le modèle EfficientNetB0 a obtenu le plus grand taux de reconnaissance parmi les autres dans les trois bases de données (entraînement avec 85,64 %, validation avec 86,35 % et test avec 87,48 %).

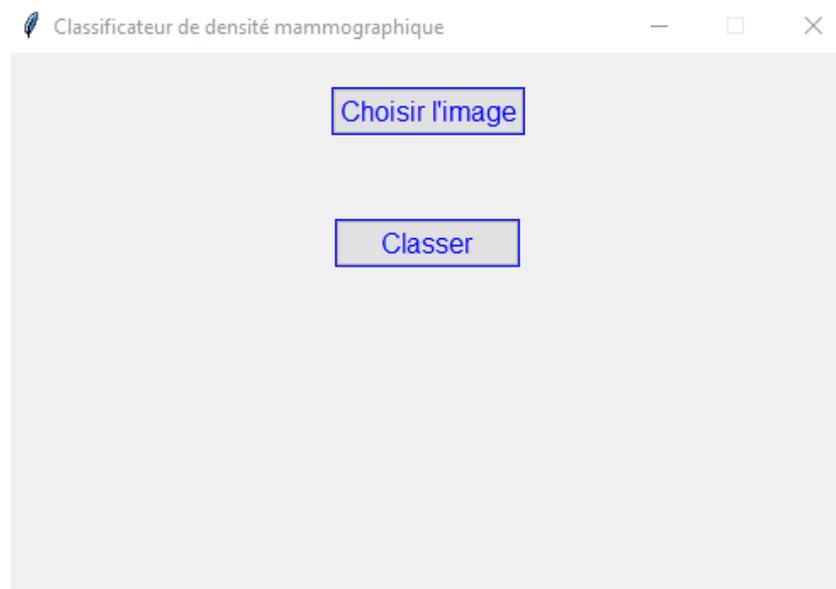
Le modèle ResNet 50 a obtenu le taux de reconnaissance le plus faible par rapport aux autres modèles dans les trois bases de données (entraînement, validation et test). Les taux étaient de 82,82%, 81,83% et 82,47%, respectivement.

MobileNetV2 et VGG 16 obtiennent également des performances similaires dans les trois ensembles de données, mais le modèle MobileNetV2 offre un taux de reconnaissance légèrement supérieure à celle du modèle VGG16.

## IV.6. Déploiement de l'application

Nous avons utilisé la bibliothèque Tkinter de Python pour créer l'interface graphique de notre application. Tkinter nous a permis de concevoir une interface utilisateur graphique

attractive pour notre projet afin de catégoriser les images mammographiques selon leurs densités. L'interface comprend des fonctionnalités telles qu'un bouton de téléchargement pour choisir une image mammographique, un bouton de classification pour exécuter notre modèle de classification et afficher les résultats, et une zone d'affichage pour l'image choisie.



**Figure IV-15 :** Interface graphique.



**Figure IV-16 :** Interface graphique (résultats de la classification d'une image de la classe B).

Grâce à Tkinter, notre application offre une méthode simple et conviviale pour classer les images mammographiques en fonction de leur densité.

## **IV.7. Conclusion**

Après avoir expérimenté avec les modèles EfficientNet, VGG16, ResNet50 et MobileNet V2, nous pouvons tirer certaines conclusions , chaque modèle a ses avantages et il est important de choisir en fonction des besoins spécifiques du projet, tels que la précision requise, les ressources disponibles et les contraintes de temps de calcul. Les expérimentations effectuées avec ces quatre modèles nous a permis de mieux comprendre leurs performances respectives et de faire un choix informé pour la tâche de classification des densités des mammographies.

## CONCLUSION GENERALE

Le cancer du sein est une condition sérieuse qui affecte les femmes, et la détection précoce joue un rôle crucial pour améliorer les chances de rétablissement. La mammographie est une technique de dépistage essentielle, mais son interprétation peut être complexe en raison de la densité mammaire variable.

L'intelligence artificielle, et en particulier l'apprentissage automatique, joue un rôle essentiel dans la lutte contre le cancer du sein. Ces techniques permettent de développer des modèles prédictifs en exploitant des bases de données, ce qui permet d'automatiser certaines tâches et de réaliser des prédictions précises. L'un des atouts majeurs de l'apprentissage automatique et de l'apprentissage profond est leur adaptabilité à une large gamme de problèmes, ce qui explique leur popularité croissante dans ce domaine.

Dans ce projet, notre objectif principal était de créer un système de classification des images de mammographie selon leurs densités en se basant sur la norme BI-RADS de ACR, en les plaçant spécifiquement dans les catégories B ou C. Nous avons utilisé la base de données RSNA, qui se compose de 24350 mammographies. Des étapes de prétraitement ont été appliquées pour améliorer la qualité des données, afin d'arriver à une catégorisation précise, nous avons effectuées des expérimentations de transfert d'apprentissage à travers quatre modèles de points d'apprentissage profond : VGG16, ResNet50, MobileNetV2 et EfficientNetB0. L'un des modèles testés, a affiché les meilleurs résultats, avec un taux de reconnaissance de 85,64 % dans l'apprentissage et 87,48 % sur le test, d'une précision de 86%, un rappel de 88% et un score F1 de 86%. Ces résultats mettent en évidence la pertinence d'EfficientNetB0 pour classer les densités des images mammographiques.

Cependant, il est vivement recommandé d'utiliser d'autres bases de données afin d'élargir le nombre de classes de densités à quatre (A, B, C et D). L'utilisation de bases de données bien traitées garantit une représentation adéquate des différentes catégories de densité mammaire, ce qui améliore la précision de la classification des images. De plus, l'application de prétraitements supplémentaires permet de préparer les données de manière plus efficace, améliorant ainsi leur qualité et réduisant les variations indésirables. En combinant ces approches, il est possible d'obtenir des résultats plus précis et fiables dans la classification des densités des images mammographiques.

## Références bibliographiques

---

- [SAM59] Samuel AL. Some studies in machine learning using the game of checkers. IBM: J Res Dev. 1959;3:210–29.
- [MIT97] Mitchell TM. Machine learning. New York: McGraw-Hill; 1997.
- [ALP14] Alpaydin E. Introduction to machine learning. 3rd ed. Cambridge, MA: The MIT Press; 2014.
- [NAQ15] El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In machine learning in radiation oncology (pp. 3-11). Springer, Cham.
- [SOM16] Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016, August). A review of machine learning techniques using decision tree and support vector machine. In 2016 international conference on computing communication control and automation (ICCUBEA) (pp. 1-7). IEEE.
- [OSI17] Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., & Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. International Journal of Computer Trends and Technology (IJCTT), 48(3), 128-138.
- [SIN15] Sinha, P., & Sinha, P. (2015). Comparative study of chronic kidney disease prediction using KNN and SVM. International Journal of Engineering Research and Technology, 4(12), 608-12.
- [AHU20] Ahuja, R., Chug, A., Gupta, S., Ahuja, P., & Kohli, S. (2020). Classification and clustering algorithms of machine learning with their applications. *Nature-inspired computation in data mining and machine learning*, 225-248.
- [PIC94] Picton, P. (1994). What is a neural network?. In Introduction to Neural Networks (pp. 1-12). Palgrave, London.
- [SOU20] Souissi, L. (2020). Entraînement d'un réseau de neurones par algorithme génétique.
- [NWA18] Nwankpa, C., Ijomah, W., Gachagan, A. et Marshall, S. (2018). Fonctions d'activation : Comparaison des tendances de la pratique et de la recherche pour l'apprentissage en profondeur. préimpression arXiv arXiv:1811.03378 .
- [GLO10] Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256). JMLR Workshop and Conference Proceedings.
- [NAI10] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807-814).
- [ZHE15] Zheng, H., Yang, Z., Liu, W., Liang, J., & Li, Y. (2015, July). Improving deep neural networks using softplus units. In *2015 International joint conference on neural networks (IJCNN)* (pp. 1-4). IEEE.

- [CLE15] Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*.
- [DUA21] Duarte, Gabriel & Pereira, Tamara & Jhones, Erik & Mesquita, Diego & Junior, Amauri. (2021). How do loss functions impact the performance of graph neural networks?.1-7. 10.21528/CBIC2021-161.
- [PIJ17] Li, P. (2017). Optimization algorithms for deep learning. Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong.
- [BAS19] Bashar, A. (2019). Survey on evolving deep learning neural network architectures. *Journal of Artificial Intelligence*, 1(02), 73-82.
- [BAS18] Basirat, M., & Roth, P. M. (2018). The quest for the golden activation function. *arXiv preprint arXiv:1808.00783*.
- [ALZ21] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8(1), 1-74.
- [GOO16] Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning, vol. 1. Cambridge: MIT press; 2016.
- [SHE20] Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404,132306.
- [ALB21] Albashish, D., Al-Sayyed, R., Abdullah, A., Ryalat, M. H., & Almansour, N. A. (2021,July). Deep CNN model based on VGG16 for breast cancer classification. In 2021 International conference on information technology (ICIT) (pp. 805-810). IEEE.
- [ZHA16] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p.770–8.
- [KIM21] Kim, B., Yuvaraj, N., Sri Preethaa, K. R., & Arun Pandian, R. (2021). Surface crack detection using deep learning with shallow CNN architecture for enhanced computation. *Neural Computing and Applications*, 33, 9289-9305.
- [ZEN18] W. Liu and K. Zeng, “SparseNet: A Sparse DenseNet for Image Classification,” *ArXiv*, vol. abs/1804.05340, pp. 1- 17, 2018.
- [SAN18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018.
- [ZAK20] Zaki, S. Z. M., Zulkifley, M. A., Stofa, M. M., Kamari, N. A. M., & Mohamed, N. A.(2020). Classification of tomato leaf diseases using MobileNet v2. *IAES International Journal of Artificial Intelligence*, 9(2), 290.
- [TAN19] Tan, M., Le, Q. V., 2019. EfficientNet: Rethinking model scaling for convolutionalneural networks. 36th Int. Conf. Mach. Learn. ICML 2019 2019-June, 10691–10700.
- [ATI21] Atila, Ü., Uçar, M., Akyol, K., & Uçar, E. (2021). Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, 61, 101182.

- [MAR18] Mark, S., Andrew, H., Menglong, Z., Andrey, Z., Liang-Chieh, C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4510–4520.
- [FAT17] Fatma, M., & Feriel, I. (2017). Classification des images de mammographie (Doctoral dissertation, Université Mouloud Mammeri).
- [CHE12] Cheikhrouhou, I. (2012). Description et classification des masses mammaires pour le diagnostic du cancer du sein (Doctoral dissertation, Université d'Evry-Val d'Essonne).
- [TAR08] Tardivon, A. (2008, November). Densité mammaire et cancer du sein. In 30<sup>e</sup> Journées de la Société Française de Sénologie et de Pathologie Mammaire (SFSPM), LaBaule, 2008. Prévention du cancer du sein (mythe ou réalité?) (pp. 93-101). Datebe SAS.
- [LEG84] LeGal, M., Chavanne, G., and Pellier, D. (1984). Valeur diagnostique des microcalcifications groupées découvertes par mammographie. a propos de 227 cas. *Bull Cancer*, 71:57–64.
- [EBE 08] Eberl, M. M., Fox, C. H., Edge, S. B., Carter, C. A., & Mahoney, M. C. (2006). BI-RADS classification for management of abnormal mammograms. *The Journal of the American Board of Family Medicine*, 19(2), 161-164.
- [SAN12] Santra, A. K., & Christy, C. J. (2012). Genetic algorithm and confusion matrix for document clustering. *International Journal of Computer Science Issues (IJCSI)*, 9(1), 322.
- [MON03] Monard, M. C., & Batista, G. E. A. P. A. (2003). Graphical methods for classifier performance evaluation. *Proc. Advances in Logic, Artificial Intelligence and Robotics (LAPTEC'2003)*, 59-67.
- [MOH18] Mohamed, A. A., Berg, W. A., Peng, H., Luo, Y., Jankowitz, R. C., & Wu, S. (2018). A deep learning method for classifying mammographic breast density categories. *Medical physics*, 45(1), 314-321.
- [SAF20] Saffari, N., Rashwan, H. A., Abdel-Nasser, M., Kumar Singh, V., Arenas, M., Mangina, E., ... & Puig, D. (2020). Fully automated breast density segmentation and classification using deep learning. *Diagnostics*, 10(11), 988.
- [LEH18] Lehman, C. D., Yala, A., Schuster, T., Dontchos, B., Bahl, M., Swanson, K., & Barzilay, R. (2019). Mammographic breast density assessment using deep learning: clinical implementation. *Radiology*, 290(1), 52-58.
- [LEE18] Lee, J., & Nishikawa, R. M. (2018). Automated mammographic breast density estimation using a fully convolutional network. *Medical physics*, 45(3), 1178-1190.
- [CRI19] Ciritsis, A., Rossi, C., Vittoria De Martini, I., Eberhard, M., Marcon, M., Becker, A. S., ... & Boss, A. (2019). Determination of mammographic breast density using a deep convolutional neural network. *The British journal of radiology*, 92(1093), 20180691.
- [MUH06] Muhimmah, I., & Zwiggelaar, R. (2006, October). Mammographic density classification using multiresolution histogram information. In Proceedings of the International Special Topic Conference on Information Technology in Biomedicine, ITAB (pp.26-28).

## Webographie

- [KAG23] Kaggle : Tout ce qu'il faut savoir sur cette plateforme. (s. d.). Consulté 12 juin 2023, à l'adresse <https://datascientest.com/kaggle-tout-ce-quil-a-savoir-sur-cette-plateforme>
- [PYT20] Python : Définition et utilisation de ce langage informatique. (2020, mars 31). Consulté 12 juin 2023, à l'adresse <https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1445304-python-definition-et-utilisation-de-ce-langage-informatique/>
- [TEN23] What is TensorFlow ? How it Works ? Introduction & Architecture. (s. d.). Consulté 12 juin 2023, à l'adresse <https://www.guru99.com/what-is-tensorflow.html>
- [KER20] Keras : Une bibliothèque open source pour la constitution de réseaux neuronaux.(2020, octobre 8). IONOS Digital Guide. Consulté 12 juin 2023, à l'adresse <https://www.ionos.fr/digitalguide/web-marketing/search-engine-marketing/quest-ce-que-keras/>
- [NUM23] NumPy : La bibliothèque Python la plus utilisée en Data Science. (s. d.). Consulté 12 juin 2023, à l'adresse <https://datascientest.com/numpy>
- [MAT23] What Is Matplotlib In Python? How to use it for plotting? (s. d.). ActiveState. Consulté 12 juin 2023, à l'adresse <https://www.activestate.com/resources/quick-reads/what-is-matplotlib-in-python-how-to-use-it-for-plotting/>
- [OMA23] Introducing PyDicom, Its Classes, Methods, and Attributes | by Omar Alkousa | Towards Data Science. (s. d.). Consulté 12 juin 2023, à l'adresse <https://towardsdatascience.com/introducing-pydicom-its-classes-methods-and-attributes-518c1d71162>
- [PAN23] What Is Pandas in Python ? Everything You Need to Know—ActiveState. (s. d.). (2022 , 9 août) Consulté 12 juin 2023, à l'adresse <https://www.activestate.com/resources/quick-reads/what-is-pandas-in-python-everything-you-need-to-know/>
- [JVC22] JVC, J. (2022, juin 23). Scikit-Learn : Guide de démarrage rapide en Machine Learning avec Python. Data Transition Numérique. Consulté 12 juin 2023, à l'adresse <https://www.data-transitionnumerique.com/scikit-learn-python/>
- [TIK17] Tkinter – l'Informatique, c'est fantastique ! (2017, janvier 30). Consulté 12 juin 2023, à l'adresse <https://info.blaise-pascal.fr/tkinter/>
- [RSN23] Carr, C.; Kitamura, F.; Kalpathy-Cramer, J.; Mongan, J.; Andriole, K.; Vazirabad, M.; Riopel, M.; Ball, R.; Dane, S. RSNA Screening Mammography Breast Cancer Detection. 2022. disponible en ligne: <https://kaggle.com/competitions/rsna-breast-cancer-detection>. Consulté 28 janvier 2023.