



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR DE LA RECHERCHE
SCIENTIFIQUE
UNIVERSITÉ ABDELHAMID BEN BADIS-MOSTAGANEM



Faculté des Sciences Exactes et d'Informatique

Départements mathématiques et informatique

Filière : Informatique

RAPPORT DE PROJET DE MASTER EN INFORMATIQUE

Option : IA FOR IOT

**THÈME : Classification des images
mammographiques par apprentissage profond**

Étudiante : HAOUACH Soumia Rokia

Encadrant: Mme KAID SLIMANE Bouchra

Année Universitaire 2024-2025

Dédicace

Je dédie ce modeste travail, aux deux êtres les plus chers à mon cœur auxquels je dois mon existence Mon père et ma mère, vous qui étiez toujours à mes côtés pour me soutenir et m'encourager à me battre sans jamais m'arrêter à mi-chemin, que dieu vous protège, Et je voue aussi mes respects à tous mes frères.m

Remerciements

Je remercie le bon Dieu, tout puissant, de m'avoir donné la force pour suivre, ainsi que l'audace pour dépasser toutes les difficultés.

Je souhaite adresser mes remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette formidable année universitaire.

Je tiens à remercier sincèrement Mme. KAID SLIMANE Bouchra, qui a Toujours montré une écoute et une grande disponibilité tout au long de la réalisation de ce mémoire.

Les jurys pour leurs efforts et leur soin apporté à notre travail.

Aux enseignants de notre université et à notre département informatique.

Enfin, nous adressons mes plus sincères remerciements à tous nos proches et amis, qui nous ont

toujours soutenus et encouragés au cours de la réalisation de ce mémoire.

Merci à tous et à toutes.

Résumé :

Dans le cadre de la réalisation du projet de fin d'études, nous avons étudié diverses méthodes et algorithmes dédiés à la reconnaissance et à la classification des images mammographiques. Parmi les approches analysées, nous avons opté pour les réseaux de neurones convolutifs (CNN) et entraîné nos modèles sur la base de données INBreast. Notre travail a impliqué l'utilisation de plusieurs modèles pré-entraînés, notamment VGG16, VGG19, AlexNet et ResNet50. Après évaluation, nous avons constaté que VGG16 offrait les meilleures performances, atteignant une précision de 54 % pour la classification BI-RADS et 80 % pour la classification binaire.

Mots clés : mammographie, CNN, apprentissage, précision, classification.

Abstract:

As part of our graduation project, we focused on the binary classification of mammographic images using deep learning techniques. We worked with the INBreast database and applied a patient-wise split to ensure a reliable evaluation by preventing data leakage. To process the images, we used preprocessing techniques such as resizing, normalization, and data augmentation.

We chose to use a pre-trained EfficientNet-B4 convolutional neural network (CNN) for its strong performance in image classification tasks. The model was fine-tuned to distinguish between two classes: B (benign) and C (cancerous). Dropout regularization was applied to mitigate overfitting.

Our approach achieved promising results, with an accuracy of approximately 52% for class B and 48% for class C, showing that the model was able to extract relevant features for classification despite the challenges posed by limited data and class imbalance.

Keywords: mammography, CNN, deep learning, accuracy, classification.

Table des matières

Remerciements

Dédicaces

Résumé

Liste des Matières

Liste des figures

Liste des tableaux

Introduction Générale

Chapitre 1:

1.1 Anatomie du sein.....	2
1.2 Cancer du sein.....	3
1.3 L'imagerie mammaire.....	4
1.3.1 Définition de l'image.....	4
1.3.2 Mammographie.....	4
1.3.3 Du film radiologique aux clichés numérisés.....	5
1.3.4 classification bi-rads.....	6
1.3.5 Les masses.....	7

1.3.6 La forme.....	7
1.3.7 Le contour.....	8
1.3.8 La densité.....	8
1.4 Pourquoi se concentrer sur les catégories B et C ?.....	9
1.5 Conclusion.....	10

Chapitre 2:

2.1 Classification des images.....	12
2.1.1 En 2016 [16].....	12
2.1.2 En 2018 [17] [18] [19].....	13
2.1.3 En 2019 [22].....	13
2.1.4 En 2020 [23] [24].....	14
2.1.5 En 2021 [25] [26] [27].....	14
2.1.6 En 2022 [28].....	14
2.2 RSNA Screening Mammography Breast Cancer Detection.....	16
2.2.1 Contenu global.....	16
2.2.2 Structure des fichiers.....	16
2.3 Colonnes du fichier train.csv.....	17
2.4 Format des images.....	18
2.5 Objectif de la compétition.....	18
2.6 Les Outils Utilisés.....	18
2.7 Langage de programmation.....	19
2.7.1 Bibliothèques de base.....	19
2.7.2 Traitement d'images médicales.....	19
2.7.3 Prétraitement et augmentation.....	19
2.7.4 Apprentissage profond.....	19
2.7.5 Visualisation et interprétation du modèle.....	20
2.8 Discussion.....	20

Chapitre 3:

3.1 Neurone biologique.....	22
3.2 Neurone artificiel.....	23
3.2.1 Structure des réseaux de neurones artificiels.....	23
3.2.2 Types des réseaux de neurones artificiels.....	24
3.2.3 L'initialisation du poids.....	25
➤ Initialisation tous zéros et initialisation constante.....	26
➤ Initialisation aléatoire.....	26
➤ Initialisation LeCun (normaliser la variance).....	26
➤ Initialisation de Xavier (initialisation de Glorot).....	26
➤ Initialisation HE (initialisation de Kaiming).....	26
3.3 Les fonctions d'activation.....	27
3.3.1 La fonction d'activation sigmoïde ou logit.....	28

3.3.2 La fonction Softmax.....	28
3.3.3 La fonction Tanh.....	28
3.3.4 La fonction de l'Unité Linéaire Rectifiée (ReLU).....	29
3.3.5 La fonction LeakyReLU.....	29
3.3.6 La fonction ReLU paramétrique (PReLU).....	30
3.3.7. Fonction d'activation SWISH.....	31
3.4 Les fonctions de pertes (erreurs).....	31
3.4.1 Erreur Quadratique Moyenne.....	32
3.4.2 Erreur Absolue Moyenne.....	32
3.4.3 Entropie croisée binaire.....	33
3.4.4. Entropie croisée catégorielle.....	33
3.5 Les Optimiseurs.....	33
3.5.1. La descente de gradient (GD).....	33
3.5.2 Descente de gradient par lots (BGD).....	34
3.5.3 Descente de gradient stochastique (SGD).....	34
3.5.4 Descente de gradient en mini-lot.....	35
3.5.5 Momentum.....	36
3.5.6 Nesterov Momentum.....	36
3.5.7 Descente de gradient adaptative (AdaGrad).....	37
3.5.8 Delta adaptatif (AdaDelta).....	37
3.5.9 Propagation racine carrée moyenne (RMSProp).....	37
3.5.10 Estimation adaptative du moment (Adam).....	37
3.5.11 Estimation du moment adaptatif maximal (AdaMax).....	38
3.5.12 Estimation du moment adaptatif accélérée par Nesterov (Nadam).....	38
3.4 Apprentissage profond (Deep learning ou DL).....	38
3.4.1 Les réseaux de neurones convolutifs (CNN).....	38
3.4.1 Les avantages de CNN.....	39
3.4.2 Les inconvénients de CNN.....	40
3.4.3. L'opération de convolution.....	40
3.4.4 Mise en commun (Pooling).....	41
3.4.5 Aplatissement (Flattening).....	42
3.5 Quelques architectures basées sur CNN.....	42
3.5.1 VGG16.....	42
3.5.2 Inception Net.....	43
3.5.3 ResNet.....	43
3.6 Conclusion.....	44

Chapitre 4:

4.1 Kaggle Notebook et Modèle.....	46
4.1.1 Présentation du Kaggle Notebook.....	46
4.1.2 Caractéristiques techniques :.....	46
4.1.3 Structure du Notebook :.....	46
4.2 Architecture du Modèle.....	47
Composants Architecturaux Clés.....	48

EfficientNet-B4 Specifications.....	48
Ressources pour l'implémentation.....	49
4.2.3 Configuration d'EfficientNet-B4.....	49
4.2.3 Visualisation de l'architecture :.....	50
4.3 Résultats de l'Entraînement.....	51
4.3.1 Courbes de perte et de précision :.....	51
4.3.2 Matrice de confusion :.....	54
4.3.3 Rapport de Classification.....	55
4.3.4 Prédictions de densité mammaire.....	56
4.4 Application Flask.....	57
4.4.1 Présentation de l'Application.....	57
4.4.2 Fonctionnalités de l'application :.....	57
4.5 Structure de l'Application.....	58
4.5.1 Dossiers et fichiers principaux :.....	58
4.5.2 Captures d'écran de l'interface utilisateur :.....	59
4.6 Utilisation de l'Application.....	61
4.6.1 Téléchargement d'une image :.....	61
4.6.2 Affichage des résultats :.....	61
4.7 Conclusion.....	62

Liste des figures

Figure 1-1: Anatomie interne du sein	2
Figure 1-2 : Schéma d'un cancer du sein	3
Figure 1-3: Mammographie numérique (MN) et les clichés associés	5
Figure 1-4 : Les différentes formes possibles d'une masse : a) Ronde, b) Ovale,	7
Figure 1-5 : Les différents contours possibles d'une masse : a) Circonscriit, b)	8
Figure 1-6 : Densité mammaire selon le lexique BI-RADS : a) Densité < 25%, b)	9
Figure 1-7 : Figure 1.7: les catégories B et C	10
Figure 3-1 : Neurone biologique	23
Figure 3-2 : Neurone artificiel	24
Figure 3-3: Perceptron multicouche à deux couches cachées pour classification binaire	25
Figure 3-4 : Les types de réseaux de neurones les plus connus	26
Figure 3-5: Schéma représente le processus d'initialisation du poids	27
Figure 3-6: Schéma représente la fonction d'activation Sigmoid	29
Figure 3-7: Schéma représente la fonction d'activation Softmax	29
Figure 3-8 : Schéma représente la fonction d'activation Tanh	30
Figure 3-9: Schéma représente la fonction d'activation ReLU	30
Figure 3-10: Schéma représente la fonction d'activation LeakyReLU	31
Figure 3-11: Schéma représente la fonction d'activation PReLU	31
Figure 3-12: Schéma représente la fonction d'activation SWISH	32
Figure 3-13: Algorithme de descente de gradient.	34
Figure 3-14: Descente de gradient stochastique	36
Figure 3-15: Minima globaux Vs. Minima locaux	37
Figure 3-16: La comparaison des algorithmes SGD avec et sans quantité de mouvement	37
Figure 3-17: Structure du réseau neuronal à convolution	40
Figure 3-18: Schéma représente l'opération de convolution	41
Figure 3-19: Schéma représente la couche pooling	42
Figure 3-20: Schéma d'un réseau à convolution avec la couche Flattening	43
Figure 3-21: Schéma représente l'architecture d'Inception de Google	44
Figure 3-22: Schéma représente l'architecture ResNet	45
Figure 4-1: Distribution de B vs C Categories	48
Figure 4-2: Architecture du modèle EfficientNet-B4	51
Figure 4-3: Courbes de perte	52
Figure 4-4: Courbes de précision	53
Figure 4-5: L'AUC désigne généralement l'aire sous la courbe ROC	54
Figure 4-6: Matrice de confusion	55
Figure 4-7: Prédictions de densité mammaire	58

Figure 4-8: Dossiers et fichiers principaux	59
Figure 4-9: Captures d'écran de l'interface Principale	60
Figure 4-10: Capture d'écran de l'interface de résultat	61
Figure 4-11: Preview Dicom	62
Figure 4-12: Affichage de résultat	62

Liste des tableaux

Tableau 1-1: Critères de distinction entre tumeurs bénignes/ maligne	4
Tableau 1-2 : La classification BI-RADS	6
Tableau 1-3 : La classification BI-RADS de la densité mammaire	7
Tableau 2-1 : Classification des images en 2016	12
Tableau 2-2 : Classification des images en 2018	13
Tableau 2-3 : Classification des images en 2019	13
Tableau 2-4 : Classification des images en 2020	14
Tableau 2-5 : Classification des images en 2021	14
Tableau 2-6 : Classification des images en 2022	14
Tableau 2-7 : Classification des images en 2023	15
Tableau 2-8 : Classification des images en 2024	16
Tableau 2-9 : Colonnes du fichier train.csv	18
Tableau 3-1: Comparaison entre les techniques d'initialisation du poids	27
Tableau 4-1: EfficientNet-B4 Specifications	48
Tableau 4-2: Rapport de classification détaillé	48
Tableau 4-3: Autres métriques binaires	55

Liste des abréviations

ACR : Collège Américain de Radiologie (American College of Radiology).

AUC : Les aires sous la courbe (Areas under the curve).

BC : Cancer du sein (Breast cancer).

BI-RADS: Breast Imaging-Reporting and Data System.

BRCA1 : gène du cancer du sein 1 (Breast Cancer Gene 1).

BRCA2 : gène du cancer du sein 2 (Breast Cancer Gene 2).

CAD : Conception assistée par ordinateur (Computer Aided Design).

CCD : dispositif à transfert de charge (Charge-Coupled Device).

CGL : Apprenant du graphe conditionnel (Conditionnel Graph Learner).

CNN : Le réseaux de neurones convolutifs (The Convolutional Neural Network).

CS : Cancer du sein.

DAO : diagnostic assisté par ordinateur.

DDSM : Base de données numérique pour la mammographie de dépistage (Digital Database for Screening Mammography).

DDSM : la base de données d'images mammographiques appelée (Digital Database for Screening Mammograph).

DL : apprentissage profond (Deep Learning).

DNN : un réseau neuronal profond (a deep neural network).

DPN : Réseau à double voie (Dual Path Network).

GBM : machine à gradient boosting (gradient boosting machine).

GPU : Unité de traitement graphique (Graphics Processing Unit).

INbreast : Base de données de mammographie numérique pour la recherche sur le cancer du sein (Digital Mammography Database for Breast Cancer Research).

IRM : Imagerie par Résonance Magnétique.

ISBI : Le Symposium international sur l'imagerie biomédicale (The International Symposium on Biomedical Imaging).

IU : interface utilisateur.

K-fold : K nombre entier fold sous ensemble.

LPL : Apprenant de préservation de la localité (Locality Preservation Learner)

MG : mammographique

MIAS : La base de données d'images mammographiques appelée (Mammographic Image Analysis Society).

OMS : Organisation mondiale de la santé.

ReLU : Unités linéaires rectifiées (Rectified Linear Units).

ResNet : Réseau résiduel (Residual Network).

RFE : une élimination récursive des caractéristiques (recursive feature elimination).

ROC : Caractéristique de fonctionnement du récepteur (Receiver Operating Characteristic).

RU-Net : fait référence à un modèle U-Net d'attention résiduelle.

SVM : La machine à vecteur de support (the support vector machine).

TL : transfert d'apprentissage (learning transfer).

TMAs : fait référence aux Arrays de Microtissus (Tissue Microarrays en anglais).

UCI : fait référence à l'Université de Californie à Irvine (University of California, Irvine en anglais).

U-Net : est un modèle d'apprentissage en profondeur.

VGG16 : fait référence à une architecture de réseau neuronal convolutif.

WBCD : l'ensemble de données du cancer du sein de Wisconsin (the Wisconsin Breast Cancer Dataset).

Introduction Générale

Le cancer du sein demeure l'une des préoccupations majeures de santé publique à l'échelle mondiale, touchant des millions de femmes chaque année. La détection précoce et un diagnostic précis jouent un rôle essentiel dans l'amélioration des chances de survie. C'est dans cette optique que s'inscrit notre projet de fin d'études, centré sur la classification automatique d'images mammographiques par le deep learning.

Selon l'Organisation Mondiale de la Santé (OMS), environ 2,3 millions de nouveaux cas de cancer du sein sont diagnostiqués chaque année, avec plus de 600 000 décès. Ces chiffres alarmants soulignent la nécessité de développer des outils informatiques performants pour assister les professionnels de santé.

La mammographie reste l'un des outils de dépistage les plus utilisés. Cependant, son interprétation reste complexe et parfois subjective, ce qui justifie l'usage de techniques automatisées capables de réduire les erreurs humaines. Parmi ces techniques, les réseaux de neurones convolutifs (CNN) se sont révélés particulièrement performants pour le traitement d'images médicales.

Dans notre travail, nous avons étudié l'approche CNN à travers l'utilisation d'un modèle pré-entraîné moderne : EfficientNet-B4. Contrairement aux approches classiques (AlexNet, VGG16, etc.), EfficientNet propose une architecture optimisée qui équilibre profondeur, largeur et résolution de l'image. Nous avons appliqué ce modèle pour une classification binaire : distinguer les images de classe B (bénigne) de celles de classe C (cancer), en nous basant sur la base de données INBreast.

Afin d'assurer une évaluation fiable, nous avons adopté un découpage patient-wise, évitant que des images du même patient apparaissent dans plusieurs ensembles (train, validation, test). Le prétraitement des images inclut le redimensionnement, la normalisation et l'augmentation de données. De plus, une couche Dropout a été ajoutée pour limiter le surapprentissage.

Ce mémoire est structuré en quatre chapitres :

- Le premier chapitre introduit les concepts médicaux de base : anatomie du sein, types de cancer, rôle de la mammographie, et présentation du système BIRADS.

- Le deuxième chapitre présente l'état de l'art, avec une revue des travaux existants sur la classification des images mammographiques par le deep learning.
- Le troisième chapitre décrit le fonctionnement des réseaux CNN, en mettant l'accent sur l'architecture EfficientNet, notre choix principal, et justifie son utilisation.

Enfin, le quatrième chapitre détaille notre implémentation, les résultats obtenus (tels que la précision, le score ROC AUC), les limites de notre approche et les pistes d'amélioration futures.

Chapitre 1 : Contexte Médicale

Dans ce chapitre, nous avons exposé l'approche médicale de l'imagerie mammaire, afin de mettre en contexte l'application que nous cherchons à développer. Nous avons abordé l'anatomie du sein, le cancer du sein, ainsi que les modifications bénignes et malignes qui peuvent se produire, bien que cette approche ne soit pas exhaustive et vise simplement à faciliter la compréhension de la problématique pour déduire des méthodes de détection. Nous terminons ce chapitre par la présentation de généralités sur l'imagerie mammaire et les lésions radiologiques.

1.1 Anatomie du sein

La poitrine (Le sein) est principalement formée d'un tissu adipeux graisseux plus ou moins important qui lui donne sa silhouette et son volume. La glande mammaire, enveloppée dans le tissu conjonctif, est constituée d'une vingtaine de lobules qui deviennent actifs pendant la lactation. Le lait produit est expulsé par des canaux galactophores qui se rejoignent au niveau du mamelon. La poitrine est irriguée par de nombreux vaisseaux sanguins et est maintenue en place par la peau qui la recouvre et des fibres attachées au muscle pectoral. La figure 1.1 fournit une description plus détaillée des différents éléments du sein.

Le sein est composé d'une glande mammaire, de fibres de soutien (ligaments de Cooper) et de graisse (tissu adipeux), le tout est recouvert par la peau. La quantité de chacune de ses composantes peut varier d'une femme à l'autre. Le sein est situé par-dessus le muscle pectoral. On trouve également dans le sein des nerfs, des vaisseaux sanguins et lymphatiques. La glande mammaire est divisée en 15 à 20 sections qu'on appelle lobes, composés de lobulés. Ceux-ci sont reliés à des canaux qui se rendent sous le mamelon (situé au centre du sein). On peut également observer des chaînes de ganglions lymphatiques qui filtrent les microbes et protègent le corps contre l'infection et la maladie[1].

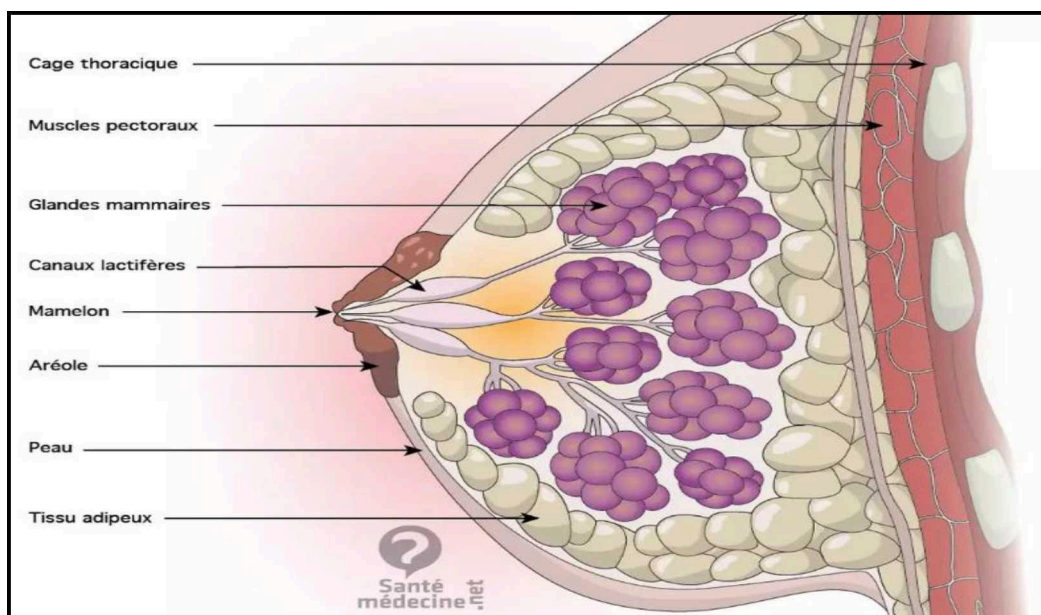


Figure 1-1: Anatomie interne du sein

1.2 Cancer du sein

Le cancer du sein prend naissance dans les cellules du sein. La tumeur cancéreuse (maligne) est un groupe de cellules cancéreuses qui peuvent envahir et détruire le tissu voisin. Elle peut aussi se propager (métastases) à d'autres parties du corps.

Les cellules du sein subissent parfois des changements qui rendent leur mode de croissance ou leur comportement anormal. Ces changements peuvent engendrer des affections non cancéreuses (bénignes) du sein, comme l'hyperplasie atypique et des kystes. Ils peuvent aussi entraîner la formation de tumeurs non cancéreuses, dont les papillomes "intra-canalaire".

Dans certains cas, cependant, des modifications dans les cellules mammaires peuvent causer un cancer du sein.

Le cancer du sein apparaît le plus souvent dans les cellules tapissant les canaux, qui sont des tubes qui transportent le lait des glandes au mamelon. Ce type de cancer du sein est appelé carcinome canalaire. Le cancer peut aussi se former dans les cellules des lobules, qui sont les groupes de glandes productrices de lait. Ce type de cancer porte le nom de carcinome lobulaire.

Le carcinome canalaire et le carcinome lobulaire peuvent être in situ, c'est-à-dire que le cancer reste dans son emplacement d'origine et qu'il n'a pas envahi les tissus voisins. Ils peuvent également être infiltrant, ou invasifs, c'est-à-dire qu'ils ont envahi les tissus voisins.

Des types de cancer du sein moins fréquents peuvent aussi se manifester. Le cancer inflammatoire du sein, la maladie de Paget du sein et le cancer du sein triple négatif en sont des exemples. Des types rares de cancer du sein sont entre autres le lymphome non hodgkinien et le sarcome des tissus mous [3].

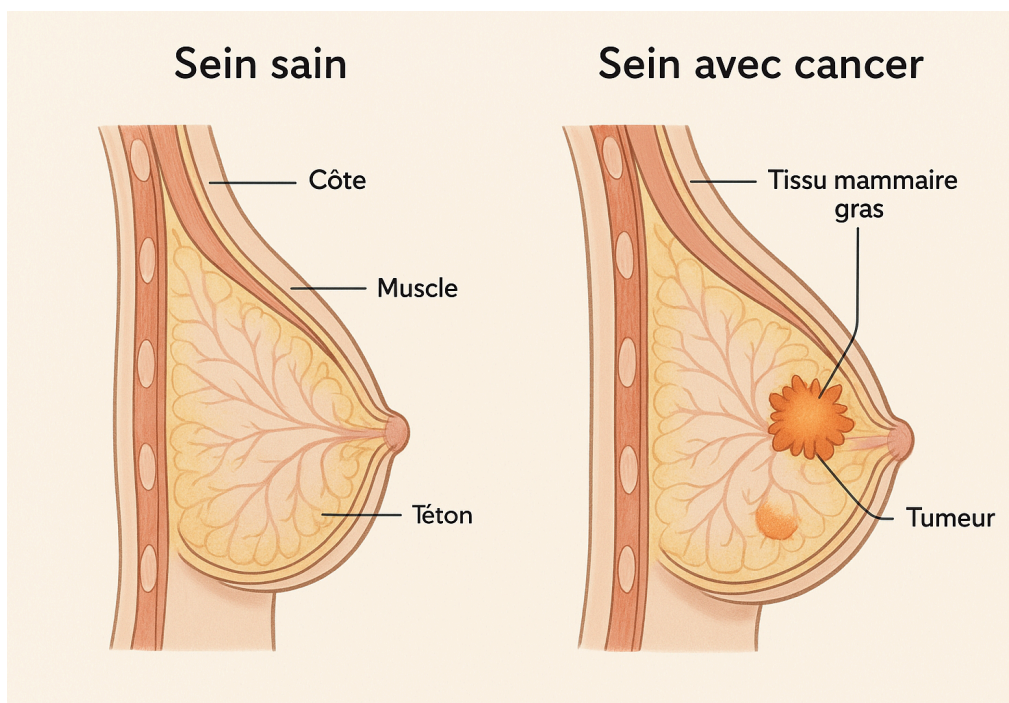


Figure 1-2 : Schéma d'un cancer du sein.

Il y a deux grandes catégories de tumeurs sont connues :

- ❖ Les tumeurs bénignes.
- ❖ Les tumeurs malignes ou cancers.

Le tableau 1.1 décrit les points par lesquels on peut différencier une tumeur bénigne d'une autre maligne.

Tumeurs bénignes	Tumeurs malignes
Bien limitée	Mal limitée
Encapsulée	Non encapsulée
Histologiquement semblable au tissu d'origine	Plus ou moins semblable au tissu d'origine (Dédifférenciation, différenciation aberrante)
Cellules régulières	Cellules irrégulières (cellules cancéreuses)
Croissance lente	Croissance rapide
Refoulement sans destruction des tissus voisins	Envahissement des tissus voisins
Pas de récurrence locale après exérèse Complète	Récurrence possible après exérèse supposée Totale
Pas de métastases	Métastases

Tableau 1-1 : Critères de distinction entre tumeurs bénignes/ maligne [5].

1.3 L'imagerie mammaire

L'imagerie mammaire est l'utilisation de différentes techniques d'imagerie pour évaluer et diagnostiquer les affections mammaires, en particulier le cancer du sein.

1.3.1 Définition de l'image

L'image est une représentation visuelle de quelque chose, telle qu'elle est perçue ou conçue dans l'esprit humain. Elle peut être créée à partir d'une variété de médias tels que la peinture, la photographie, le cinéma, l'art numérique, etc.

1.3.2 Mammographie

La mammographie est une radiographie du sein. C'est l'exploration du sein la plus ancienne et la plus souvent réalisée. C'est également la seule technique reconnue pour le dépistage du cancer du sein. Elle utilise de faibles doses de rayons X. c'est pourquoi tout risque de grossesse doit être signalé au praticien ou radiologue [6].

Une mammographie de dépistage permet de découvrir des cancers du sein à un stade précoce. Cet examen est recommandé tous les 2 ans chez la femme [6].

Selon Bermond, près de 40% des cancer « ratés » à la mammographie sont dus à une mauvaise qualité des clichés. Il s'agit le plus souvent d'un défaut de positionnement du sein [7].

1.3.3 Du film radiologique aux clichés numérisés

Il y a deux méthodes de reproduction de l'image mammographique à partir du film radiologique :

- a) **La numérisation indirecte** : s'effectue à partir d'un film et une caméra CCD ou d'un numériseur de film à balayage laser (scanner optique) [8]. Cependant, La numérisation d'une radiographie ne peut jamais révéler plus de détails que l'image originale. Elle conserve en outre toutes les imperfections et les limitations de l'image initiale (bruit, dynamique, artefacts de développement, etc.).
- b) **La mammographie numérique** : la technique de la mammographie numérique remplace le système classique d'imagerie par un détecteur qui transforme les rayons X en signaux électroniques. Cette méthode permet une plus grande plage dynamique que le couple écran-film utilisé auparavant. Cette technologie présente de multiples avantages :

➤ Avantages de la mammographie numérique

Cette technologie présente de multiples avantages :

- a) La mammographie numérique permet la manipulation des images à l'aide d'ordinateurs haute résolution pour répondre aux besoins spécifiques de l'utilisateur, tels que le zoom, l'inversion de contraste et le fenêtrage. De plus, cette méthode utilise une dose d'irradiation plus faible que les méthodes d'imagerie traditionnelle.
- b) Cette technologie numérique accélère considérablement l'acquisition des images et élimine le temps consacré au développement de la pellicule, ce qui permet d'accroître l'offre d'examen des patients [9].

Comme elle représente dans la figure 1-2



Figure 1-3 : Mammographie numérique (MN) et les clichés associés [2].

1.3.4 classification BI-RADS

La classification BI-RADS (Breast Imaging-Reporting And Data System) est utilisée par les radiologues lors de la mammographie, de l'échographie et de l'IRM pour définir les anomalies vues et permettre de savoir ce qui doit être fait par la suite : retour au dépistage, suivi rapproché ou biopsie.

Le tableau .2 décrit La classification BI-RAD

Catégorie	Imagerie	Risque de cancer	Conduite
0	Investigation incomplète		Investigation à compléter ou comparer avec films antérieurs
1	Normale	0 %	Retour au dépistage
2	Anomalie bénigne	0 %	Retour au dépistage
3	Anomalie probablement bénigne	> 0 % mais ≤ 2%	Suivi 6 mois
4 4A 4B 4C	Anomalie demandant une biopsie faiblement suspecte modérément suspecte très suspecte	> 2 % mais < 95 % > 2 % à ≤ 10 % > 10 % à ≤ 50 % > 50 % à < 95 %	Biopsie
5	Anomalie fortement suspecte d'un cancer	≥ 95 %	Biopsie
6	Cancer prouvé à la biopsie	100 %	Chirurgie

Tableau 1-2 : La classification BI-RADS [10].

Il existe aussi une classification BI-RADS spécifique pour exprimer **la densité mammaire**. La densité est exprimée par une note qualitative et un pourcentage qui correspond aux catégories BIRADS :[10]

Le tableau 1-3 donne la classification BI-RADS exprimée en densité mammaire ou tout simplement mettre.

Catégorie BI-RADS	Description	% tissu glandulaire
A	Seins presque entièrement graisseux	< 25 %
B	Seins composés de zones de densités fibroglandulaires éparses	25 à 50 % approximativement
C	Seins denses de façon hétérogène, pouvant masquer des petites masses	51 à 75 % approximativement
D	Seins extrêmement denses, diminuant la sensibilité de la mammographie	> 75 %

Tableau 1-3 : La classification bi-rads de la densité mammaire [4].

1.3.5 Les masses

Une opacité ou une masse est une zone occupant un espace important et vue sous deux angles différents. Si une opacité potentielle est observée sur une seule incidence, elle est appelée asymétrique jusqu'à ce que son caractère tridimensionnel soit confirmé.

Différentes caractéristiques de ces masses sont à écrire à savoir la forme, le contour et la densité [11].

1.3.6 La forme

Selon la description du BI-RADS (D'Orsi et al., 2003 [12]), les masses mammaires peuvent avoir la forme ronde (figure 1-4 a), ovale (figure 1-4 b), lobulée (figure 1-4 c) ou irrégulière (figure 1-4 d) [11].

- (a) Ronde : Il s'agit de masse sphérique, circulaire ou globuleuse.
- (b) Ovale : Elle présente une forme elliptique (ou en forme d'œuf).
- (c) Lobulée : La forme de la masse présente une légère ondulation.
- (d) Irrégulière : Cette appellation est réservée aux masses dont la forme est aléatoire et ne peut être caractérisée par les termes cités ci-dessus.

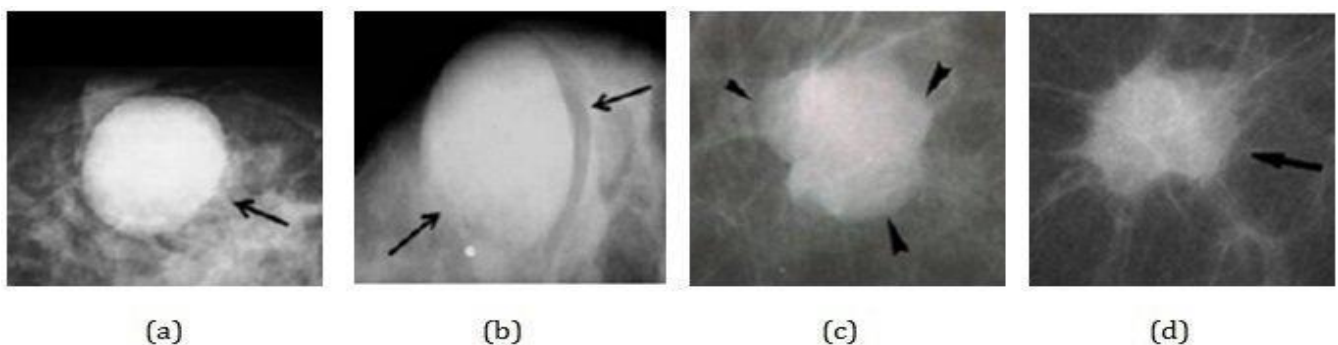


Figure 1-4 : Les formes d'une masse : a) Ronde, b) Ovale, c) Lobulée et d) Irrégulière [11].

1.3.7 Le contour

Le contour des masses mammaires est soit circonscrit (figure 1-5 a), soit microlobulé (figure 1-5 b), soit masqué (figure 1-5 c) soit indistinct (figure 1-5 d), soit spiculé (figure 1-5 e). On détaille dans ce qui suit ces différentes notions : [11].

- (a) Circonscrit : Il s'agit d'une transition brusque entre la lésion et le tissu environnant. Le contour est alors net et bien défini. Pour qu'une masse soit qualifiée de circonscrite, il faut qu'au moins 75% de son contour soit nettement délimité.
- (b) Microlobulé : Dans ce cas, de courtes dentelures du contour créent de petites ondulations.
- (c) Masqué : Un contour masqué est un contour qui est caché par le tissu normal adjacent. Ce terme est employé pour caractériser une masse circonscrite dont une partie du contour est cachée.
- (d) Indistinct : Dans ce cas, le contour est mal défini. Ce caractère indistinct (le contraire de circonscrit) peut correspondre à une infiltration.
- (e) Spiculé : La masse est caractérisée par des lignes radiaires prenant naissance sur le contour de la masse. Ces lignes radiaires sont appelées les spicules.

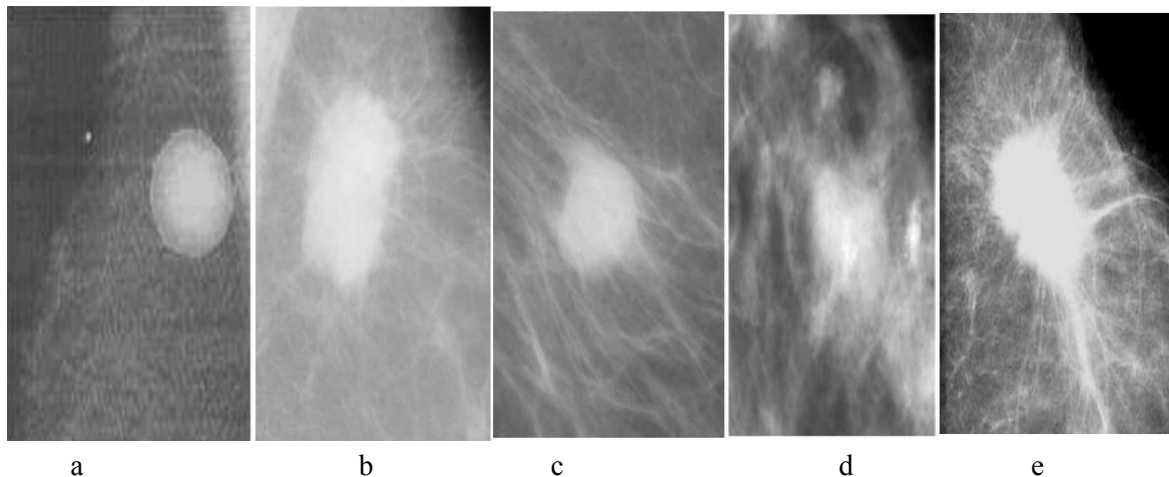


Figure 1-5 : Les différents contours possibles d'une masse : a) Circonscrit, b) Microlobulé, c) Masqué, d) Indistinct et e) Spiculé [11].

1.3.8 La densité

L'apparence du sein normal varie fortement d'une femme à l'autre (figure 1-6). Le facteur le plus remarquable est la grande variabilité de la densité radiologique de la zone mammaire. Wolfe a été le premier à établir une relation entre la densité du tissu mammaire et le risque de développer un cancer (Wolfe, 1976 [13]). D'autres études remettent en question ce lien (LeTreut et Dilhuydy, 1988 [14]). Néanmoins, cette classification tissulaire est utilisée dans d'autres travaux sans faire de lien avec le facteur de risque (Suckling et al., 1995 [15]).

Afin de standardiser les rapports mammographiques, la classification BI-RADS de l'ACR définit 4 classes de la composition du sein [11].

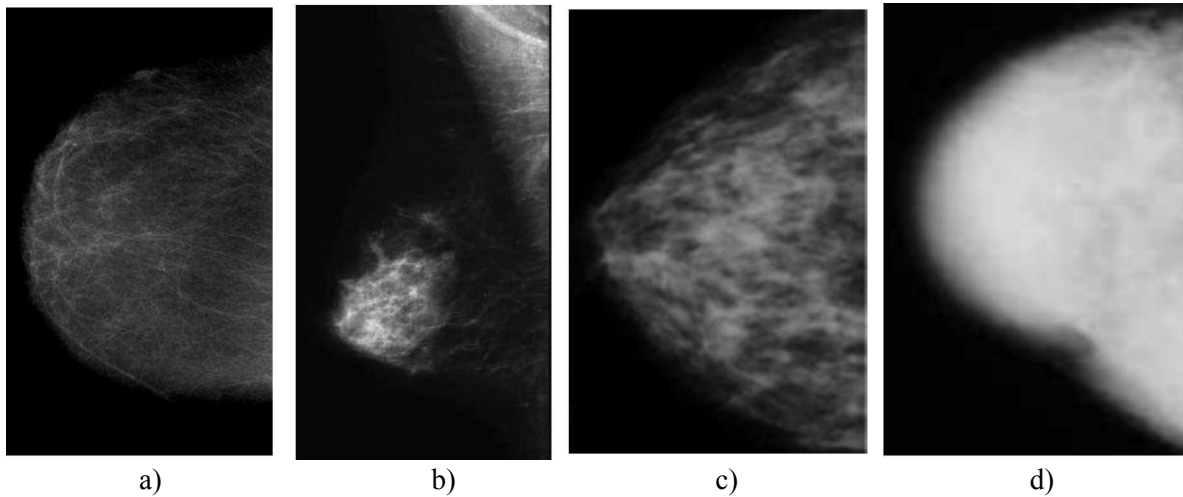


Figure 1-6 : Densité mammaire selon le lexique BI-RADS : a) Densité < 25%, b) Densité entre 25 et 50%, c) Densité entre 50 et 75% et d) Densité > 75% [11].

- (a) **Stade 1** : Le sein est presque entièrement graisseux et homogène, radio transparent et facile à lire (moins de 25% de la glande mammaire).
- (b) **Stade 2** : Il y a des opacités fibro-glandulaires dispersées. Le sein est graisseux et hétérogène (approximativement 25 à 50% de la glande mammaire).
- (c) **Stade 3** : Le tissu mammaire est dense et hétérogène (approximativement 51 à 75% de la glande mammaire)
- (d) **Stade 4** : Le tissu mammaire est extrêmement dense et homogène. La mammographie est alors difficile à interpréter puisque la densité peut masquer une lésion (plus de 75% de la glande mammaire) (figure 1-5)..

1.4 Pourquoi se concentrer sur les catégories B et C ?

Dans le cadre de notre projet, nous avons choisi de nous focaliser exclusivement sur les classes de densité mammaire B et C car ces catégories représentent des profils diagnostiques à la fois fréquents et particulièrement ambigus. La densité de type B, avec des zones éparses de tissu glandulaire au sein d'une matrice essentiellement graisseuse, et la densité de type C, caractérisée par un tissu mammaire dense de façon hétérogène, posent des défis spécifiques en termes d'interprétation radiologique. Ces deux classes ne constituent pas uniquement une part importante des examens de mammographie, mais elles correspondent également à des zones critiques où l'apparition de petites masses peut facilement être masquée par la densité du tissu.

En se concentrant sur les catégories B et C, notre approche par apprentissage profond vise à affiner la précision du dépistage dans des contextes particulièrement sensibles. Ce choix permet non seulement d'optimiser l'utilisation des ressources pour traiter les cas les plus complexes, mais également d'améliorer la détection précoce des anomalies, ce qui est crucial pour une meilleure prise en charge du cancer du sein.

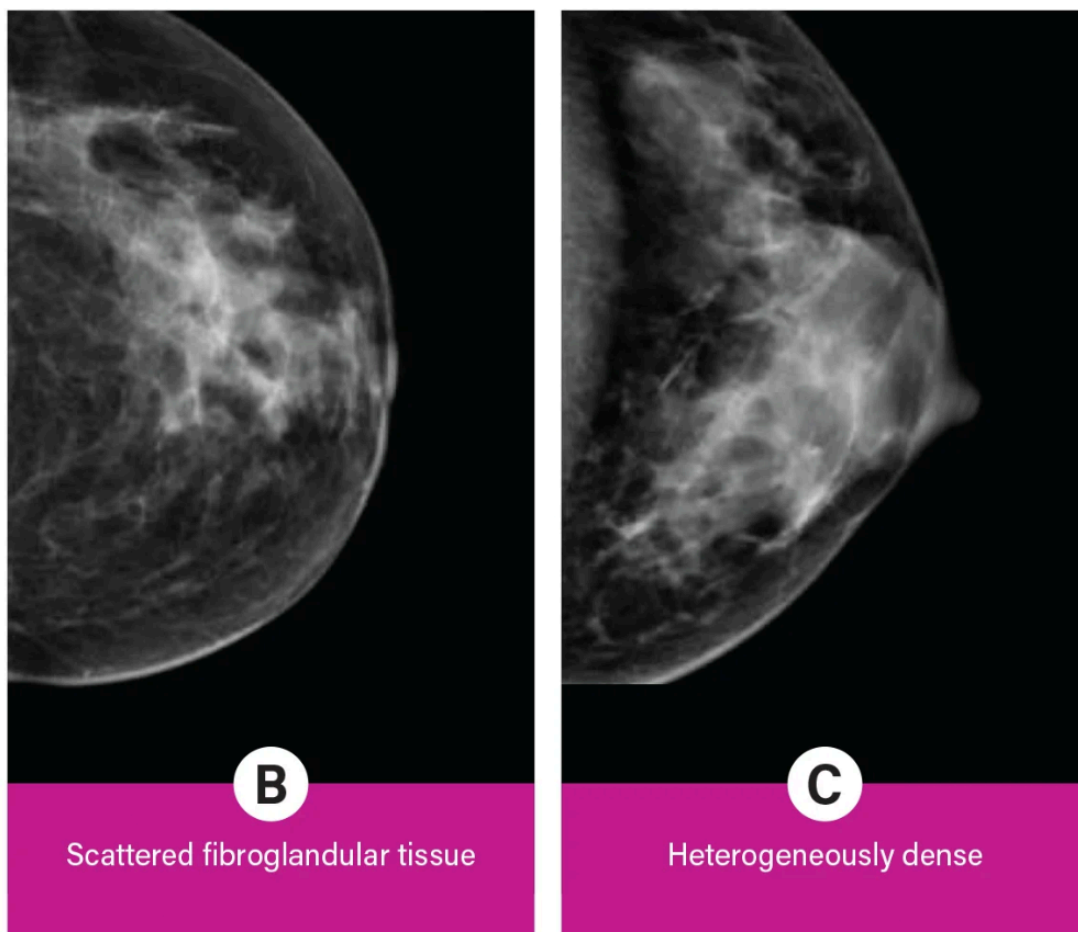


Figure 1-7: les catégories B et C.

1.5 Conclusion

Dans ce chapitre, Nous avons présenté quelques concepts généraux relatifs à l'anatomie du sein, à l'importance de la mammographie et aux différentes techniques d'imagerie mammaire.

En raison du faible contraste de nombreuses lésions cancéreuses et de leur grande ressemblance avec le tissu mammaire, l'analyse des images mammographiques est une tâche extrêmement difficile qui requiert une grande expertise et une concentration soutenue.

Chapitre 2 : L'état de l'art

La classification des images mammographiques constitue un enjeu majeur dans la détection précoce du cancer du sein, l'une des principales causes de mortalité chez les femmes dans le monde. Afin d'améliorer le diagnostic et d'assister les radiologues, de nombreuses recherches ont été menées dans le but de développer des méthodes d'analyse automatique, reposant principalement sur des techniques d'intelligence artificielle.

Ce chapitre présente un état de l'art centré sur les travaux scientifiques majeurs ayant contribué à l'avancement des méthodes de classification d'images mammographiques. Il met en lumière les différentes approches méthodologiques, les architectures de réseaux utilisées, les jeux de données mobilisés, ainsi que les résultats obtenus. L'objectif est de fournir un panorama des techniques les plus performantes, en vue d'identifier les approches les plus pertinentes pour notre étude.

2.1 Classification des images

La classification des images mammographiques repose souvent sur l'extraction de caractéristiques spécifiques, telles que la texture, la forme ou l'intensité des tissus présents dans les images. Plusieurs algorithmes de classification supervisée et non supervisée ont été proposés pour résoudre ce problème, notamment les réseaux de neurones convolutifs (CNN), les méthodes basées sur les arbres de décision, les machines à vecteurs de support (SVM) et les réseaux bayésiens. Pour cela dans ce chapitre, nous examinerons les différentes approches et résultats présentés dans chaque article, en mettant en évidence les résultats obtenus par chaque méthode.

2.1.1 En 2016 [16]

Auteurs	Approche / Méthodologie	Données / Architectures	Résultats clés
Dayong Wang et al.	Détection du cancer métastatique dans les images de biopsies	VGG16, AlexNet, GoogleNet, DeepFace + Pathologiste	Score classification : 0,925 (DL), 0,966 (pathologiste), 0,995 (combiné). Localisation : 0,7051 (DL), 0,733 (pathologiste).

Tableau 2-1 : Classification des images en 2016.

2.1.2 En 2018 [17] [18] [19]

Auteurs	Approche / Méthodologie	Données / Architectures	Résultats clés
Dalal Bardou et al.	Comparaison CNN vs SVM	Codage BoW/LLC vs CNN	CNN : 96-98 % (binaire), 83-88 % (multiclasse)
Yeeleng S. Vang et al.	InceptionV3 + DPN + ensemble learning	Challenge BACH	Amélioration de 12,5 % par rapport à l'état de l'art
Aditya Golatkart et al.	Patches par densité nucléaire + CNN	Images 4 catégories (normal, bénin, in situ, invasif)	85 % (4 classes), 93 % (bénin vs malin)
S. Karthik et al.	DNN + RFE sur WBCD	Données hyperspectrales	Exactitude : 98,62 %
Mehdi H. Motlagh et al.	ResNet50/152 pré-entraînés	BreakHis & TMAs	99,8 % (4 cancers), 98,7 % (sein bénin/malin), AUC : ~0,996

Tableau 2-2 : Classification des images en 2018.

2.1.3 En 2019 [22]

Auteurs	Approche / Méthodologie	Données / Architectures	Résultats clés
Dina Abdelhafiz et al.	RU-Net + ResNet pour segmentation & classification	Mammographies (MG)	Meilleure précision avec images binaires que niveaux de gris

Tableau 2-3 : Classification des images en 2019.

2.1.4 En 2020 [23] [24]

Auteurs	Approche / Méthodologie	Données / Architectures	Résultats clés
Heyi Li et al.	DualCoreNet : LPL + CGL	DDSM, INbreast	Segmentation & classification supérieures aux modèles existants
Ghulam Murtaza et al.	Étude de multiples modalités & techniques DL	Revue de littérature	Guide complet pour classification du cancer du sein

Tableau 2-4 : Classification des images en 2020.

2.1.5 En 2021 [25] [26] [27]

Auteurs	Approche / Méthodologie	Données / Architectures	Résultats clés
Dheeb Albashish et al.	VGG16 + ML sur BreakHis	Classification binaire & multiclasse	Supérieur aux méthodes classiques
Abeer Saber et al.	TL avec VGG16, ResNet, Inception	MIAS	Jusqu'à 98,96 % de précision, AUC : 0,995

Tableau 2-5 : Classification des images en 2021.

2.1.6 En 2022 [28]

Auteurs	Approche / Méthodologie	Données / Architectures	Résultats clés
Kosmia Loizidou et al.	Segmentation temporelle + 96 caractéristiques + 9 classifieurs	320 images annotées	Détection : 99,9 %, classification : 98 % (vs 92,6 %)
Nagwan Abdel Samee et al.	Sélection de caractéristiques en cascade (AlexNet, VGG, GoogleNet)	INbreast	Précision : 98,50 %, AUC : 0,993

Tableau 2-6 : Classification des images en 2022.

2.1.7 En 2023 [28]

Auteurs	Approche / Méthodologie	Données / Architectures	Résultats clés
PubMed (Efficient BC)	Hybrid DL : DenseNet121, ResNet50, EfficientNetB7	ROI patches (normal, bénin, malin)	Exactitude élevée grâce à la fusion des modèles ; amélioration significative de la précision.

Tableau 2-7 : Classification des images en 2023.

2.1.8 En 2024 [28]

Auteurs	Approche / Méthodologie	Données / Architectures	Résultats clés
Umamaheswari et Babu	AT-RGFM + MGSOA	Images 3D	Précision de 96,6 % grâce à des techniques avancées de segmentation et prétraitement .
Liu et al.	CNN basé sur VGG16	Ganzhou People's Hospital	Taux d'identification moyen de 96,945 % pour la classification binaire (bénin/malin) .
Gharaibeh et al.	ResNet50 + Grad-Cam	FDDM, MIAS	Exactitude : 0,8920 (FDDM), 0,9830 (MIAS) .
Boudouh et Bouakkaz	Modèle à double branche	CBIS-DDSM	Exactitude de 95,86 % pour les vues CC et MLO combinées .

Lamprou et al.	StethoNet + TL	CMMD	AUC : 90,7 % (CMMD), 83,9 % (INbreast), 85,7 % (Vindr-Mammo).
----------------	-------------------	------	---

Tableau 2-8 : Classification des images en 2024.

2.2 RSNA Screening Mammography Breast Cancer

Detection

Le jeu de données RSNA Screening Mammography Breast Cancer Detection a été publié dans le cadre d'une compétition sur la plateforme Kaggle, organisée par la Radiological Society of North America (RSNA), en collaboration avec des hôpitaux partenaires et des experts en radiologie [29]. Il constitue une ressource unique pour le développement et l'évaluation de systèmes d'intelligence artificielle appliqués au dépistage du cancer du sein par mammographie.

Ce jeu de données est conçu pour refléter la réalité clinique du dépistage de masse, en simulant les données disponibles dans un environnement hospitalier standard. Il propose un large éventail d'images et de métadonnées, permettant d'aborder différentes tâches comme la détection du cancer, la classification de la densité mammaire ou encore la prédiction de la nécessité de biopsie.

2.2.1 Contenu global

- ❖ Nombre total de fichiers : plus de 54 700 fichiers
- ❖ Images médicales : environ 54 000 images au format DICOM (.dcm)
- ❖ Fichiers de métadonnées : 3 fichiers CSV (train.csv, test.csv, sample_submission.csv)
- ❖ Poids total du dataset : environ 314 Go
- ❖ Données du test caché : une partie du test est inaccessible localement et n'est évaluée que via la plateforme Kaggle [29].

2.2.2 Structure des fichiers

Images

Les images sont organisées dans deux répertoires principaux :

1. train_images/
2. test_images/

Chaque sous-dossier suit cette structure :

[patient_id]/[image_id].dcm

Un même patient peut avoir entre 2 et 4 images (parfois plus), représentant généralement les deux seins vus sous deux angles différents (vue CC et MLO).

Fichiers CSV

- ❖ `train.csv` : contient toutes les métadonnées des patients du jeu d'entraînement, y compris les cibles (cancer, density, etc.).
- ❖ `test.csv` : contient une partie des métadonnées du jeu de test, sans les étiquettes.
- ❖ `sample_submission.csv` : fichier de soumission type pour formater les prédictions.

2.3 Colonnes du fichier `train.csv`

Le fichier `train.csv` est riche en variables cliniques et techniques :

Colonne	Description
<code>site_id</code>	Identifiant du site médical où l'examen a été réalisé
<code>patient_id</code>	Identifiant unique du patient
<code>image_id</code>	Identifiant de l'image mammographique
<code>laterality</code>	Côté du sein : Gauche (L) ou Droit (R)
<code>view</code>	Vue de la mammographie : CC (cranio-caudale) ou MLO (medio-latérale oblique)
<code>age</code>	Âge de la patiente en années
<code>implant</code>	Présence ou non d'un implant mammaire (True/False)
<code>density</code>	Densité mammaire classée de A à D (<i>notre variable cible principale : on utilise uniquement les classes B et C</i>)
<code>machine_id</code>	Identifiant de l'appareil utilisé pour la prise de vue
<code>cancer</code>	Étiquette cible : présence (1) ou absence (0) de cancer
<code>biopsy</code>	Une biopsie a-t-elle été réalisée ? (True/False)
<code>invasive</code>	Le cancer détecté est-il invasif ? (True/False)
<code>BIRADS</code>	Classification BIRADS : 0 (à revoir), 1 (négatif), 2 (normal)

<code>prediction_id</code>	Identifiant utilisé pour faire correspondre les prédictions dans la soumission
<code>difficult_negative_case</code>	Est-ce un cas négatif difficile ? (True/False)

Tableau 2-9 : Colonnes du fichier train.csv

2.4 Format des images

Les images sont au format **DICOM (.dcm)**, standard dans le domaine médical [30].

Certaines images utilisent le format **JPEG 2000** pour la compression, ce qui nécessite des bibliothèques spécifiques pour les lire (par exemple pydicom [30], GDCM, ou imagecodecs). Ces images peuvent être en niveaux de gris avec des tailles variables, souvent de très haute résolution (plusieurs milliers de pixels de large).

2.5 Objectif de la compétition

L'objectif initial de la compétition était de prédire la **probabilité de présence de cancer** dans une mammographie de dépistage. Toutefois, grâce à la richesse des étiquettes disponibles, ce jeu de données permet également :

- La classification de la **densité mammaire** (A, B, C, D)
- L'étude de la corrélation entre **âge, implants, vue** et la détection du cancer
- L'entraînement de modèles d'interprétation automatique des examens mammographiques.

Dans le cadre de ce mémoire, l'analyse porte exclusivement sur la **classification automatique de la densité mammaire**, en se concentrant sur les cas où la densité est **B ou C**, afin d'éviter les classes sous-représentées et instables.

2.6 Les Outils Utilisés

Dans le cadre de notre projet de classification automatique de la densité mammaire à l'aide de l'intelligence artificielle, plusieurs bibliothèques, langages et frameworks ont été utilisés. Le choix des outils s'est fait en fonction de leur compatibilité avec les données médicales, leur performance dans les tâches d'apprentissage profond, ainsi que leur large adoption dans la communauté scientifique. L'ensemble du développement a été réalisé dans un **notebook Python** exécuté sur la plateforme **Kaggle**.

2.7 Langage de programmation

- **Python 3.12.9**

Le langage utilisé pour notre projet est **Python**, dans sa version **3.12.9**, installé via **Miniconda**. Cette version récente de Python assure la compatibilité avec les dernières versions des bibliothèques nécessaires à la lecture d'images DICOM, au traitement d'images, et à l'apprentissage profond (PyTorch, torchvision, Flask, etc.) [31].

- **Miniconda (environnement local :**

L'environnement de développement local a été configuré avec **Miniconda**, une distribution légère de Conda. Elle permet de gérer proprement les dépendances, d'isoler les environnements virtuels, et d'assurer la reproductibilité des expériences.

2.7.1 Bibliothèques de base

- **NumPy** [31]: manipulation efficace de matrices et de tableaux multidimensionnels
- **Pandas** : chargement, exploration et manipulation des fichiers CSV contenant les métadonnées
- **Matplotlib** [36] / **Seaborn** [37] : visualisation des données et création de graphiques statistiques.

2.7.2 Traitement d'images médicales

- **Pydicom** : bibliothèque utilisée pour lire et afficher les images médicales au format DICOM (.dcm) [30]
- **OpenCV (cv2)** [33]: traitement d'image, redimensionnement, transformations.
- **ImageIO / GDCM / imagecodecs** : utilisés pour gérer les fichiers encodés en JPEG 2000.

2.7.3 Prétraitement et augmentation

- **Albumentations** : bibliothèque puissante et rapide pour l'augmentation des données (rotation, flip, contraste, etc.) [32]
- **Scikit-learn** : pour la normalisation, la séparation du jeu de données (train/test), et le calcul de certaines métriques [34]

2.7.4 Apprentissage profond

- **PyTorch**

Le framework principal utilisé pour construire, entraîner et évaluer le modèle de classification. PyTorch offre une grande flexibilité pour le prototypage rapide de réseaux de neurones convolutifs (CNN) [31].

- **Torchvision** : pour charger des modèles pré-entraînés (comme ResNet ou EfficientNet), ainsi

que pour appliquer des transformations d'images standardisées.

2.7.5 Visualisation et interprétation du modèle

- **Grad-CAM (Gradient-weighted Class Activation Mapping)**

Méthode utilisée pour visualiser les zones activées par le modèle sur les images mammographiques. Cela permet d'interpréter les prédictions du réseau en montrant quelles parties de l'image ont influencé le plus la décision [35].

2.8 Discussion

Ce chapitre a permis de mettre en évidence que les réseaux de neurones convolutifs (CNN) dominent la classification des images mammographiques, offrant des résultats remarquables. L'utilisation de techniques telles que l'apprentissage par transfert, l'augmentation des données et les méthodes d'interprétation a considérablement amélioré la précision et la transparence des modèles.

À la lumière de ces résultats, nous avons sélectionné un modèle CNN adapté pour notre étude. Dans le chapitre suivant, nous détaillerons notre approche pour la classification de la densité mammaire, tout en s'appuyant sur le jeu de données RSNA.

Chapitre 3 : Les Réseaux de Neurones Convolutifs

Les réseaux de neurones artificiels et à convolution sont des méthodes d'apprentissage automatique supervisé. Le neurone formel a été initialement conçu pour modéliser les neurones biologiques et imiter par analogie leurs capacités. Cependant, ces premiers modèles n'étaient pas capables d'apprendre. Ce n'est qu'avec l'apparition du perceptron que l'apprentissage est devenu possible. Plusieurs innovations, telles que la rétropropagation du gradient et les modèles multicouches, couplées avec le développement des capacités des outils informatiques, ont permis la réalisation des premières utilisations pratiques des réseaux de neurones artificiels.

3.1 Neurone biologique

Un réseau de neurones artificiels est un ensemble interconnecté d'éléments de traitement simples dont la fonctionnalité s'inspire du neurone biologique. Le cerveau humain est constitué d'environ 100 milliards de cellules nerveuses ou neurones (Figure II-1). Les neurones communiquent par des signaux électriques, et les connexions entre eux sont assurées par des jonctions électrochimiques appelées synapses, situées sur des branches de la cellule appelées dendrites. Chaque neurone reçoit généralement plusieurs milliers de connexions avec d'autres neurones et reçoit ainsi en permanence une multitude de signaux entrants qui finissent par atteindre le corps cellulaire. Si le signal résultant dépasse un certain seuil, le neurone se déclenche ou génère une impulsion de tension en réponse. Celle-ci est ensuite transmise à d'autres neurones par une fibre ramifiée appelée axone [38].

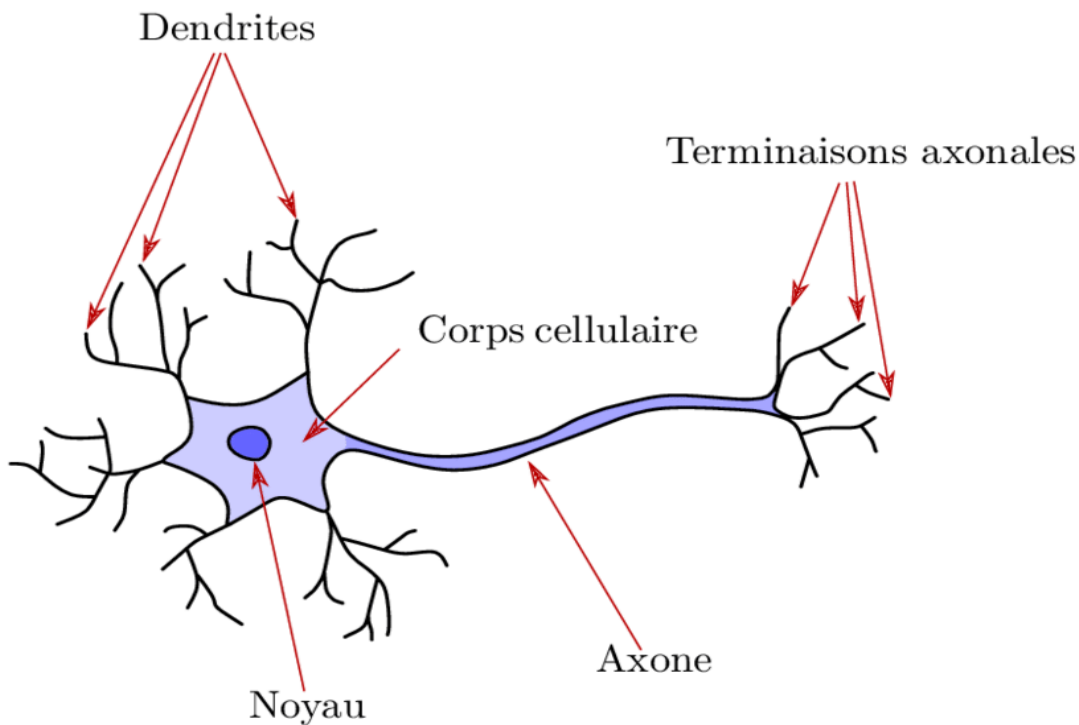


Figure 3-1 : Neurone biologique.

3.2 Neurone artificiel

Un neurone artificiel est un système ayant plusieurs entrées et une seule sortie. Chaque neurone possède deux modes : le mode d'entraînement et le mode de fonctionnement. Dans le mode d'entraînement, le neurone apprend à être déclenché contre des signaux d'entrée spécifiques ou à se déclencher à un terme. En mode de fonctionnement, une sortie correspondante est fournie lorsque les informations fournies à l'entrée sont reconnues. Si l'entrée ne fait pas partie des entrées pré-identifiées, les règles fixées après la phase d'apprentissage décident de la déclencher ou non. Un modèle élémentaire du neurone artificiel typique peut être simulé par un ordinateur.

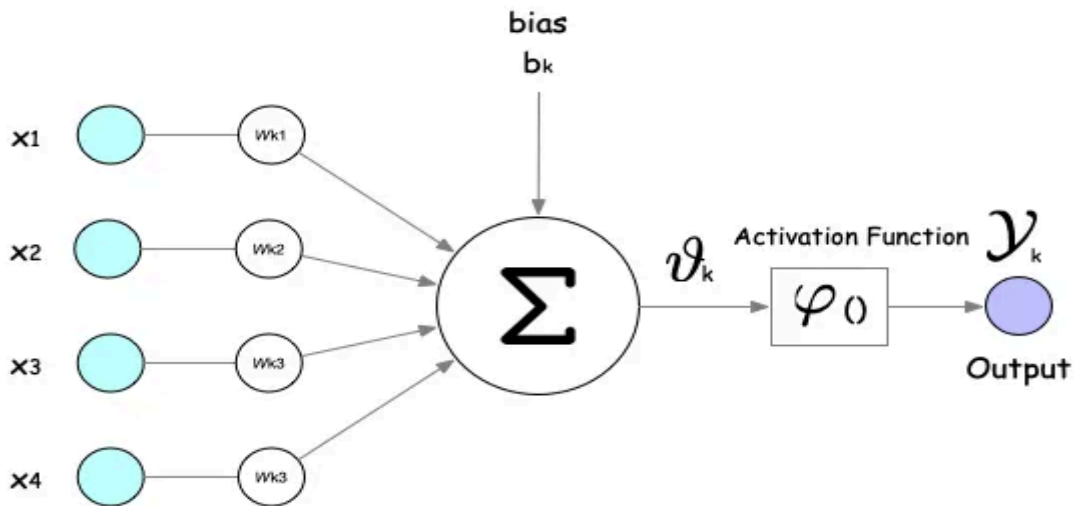


Figure 3-2 : Neurone artificiel.

Un neurone est composé d'une couche d'entrée acceptant des variables d'entrées $x_i \in \mathbf{R}$, d'un biais b_i , de poids w_i , d'une fonction d'activation σ , et d'une ou plusieurs valeurs de sortie y_i . La formule est la suivante :

$$y = \sigma \left(w_0 + \sum_{j=0}^p w_j x_j \right)$$

3.2.1 Structure des réseaux de neurones artificiels

Un réseau de neurones artificiels est constitué de composants de couches et de poids de connexion. Le comportement du réseau dépend également de la communication entre les membres. Il existe généralement trois types de couches neuronales :

1. La couche d'entrée : cette couche reçoit les données brutes introduites dans le réseau.

2. Les couches cachées : la performance de ces couches est déterminée par les entrées et les poids de la connexion entre elles et les couches cachées. Les poids entre les neurones d'entrée et cachées déterminent quand un neurone de cette couche doit être activé.

3. La couche de sortie : la sortie de l'unité de sortie dépend de l'activité de la couche cachée et des poids de la connexion entre l'unité cachée et la sortie.

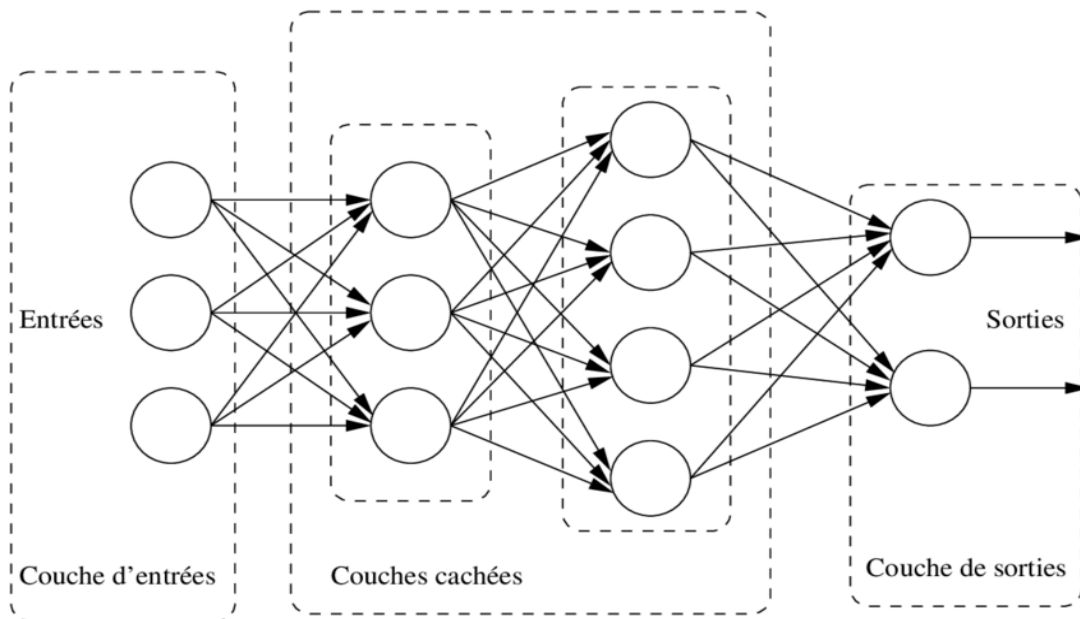


Figure 3-3: Perceptron multicouche à deux couches cachées pour classification binaire.

Il existe également des réseaux monocouches et multicouches. Dans l'organisation monocouche, toutes les neurones sont connectés à une couche. Ce type de réseaux neuronaux est le plus utilisé et possède plus de potentiel de calcul que les organisations multicouches. Dans les réseaux multicouches, les neurones sont numérotés par couches (au lieu de suivre la numérotation globale). Les deux couches d'un réseau sont reliées par des poids et en fait des connexions [39].

Dans les réseaux de neurones, il existe plusieurs types de connectivité ou gain de poids :

- Anticipation : la plupart des connexions sont des types dans lesquelles les signaux se déplacent dans une seule direction, de l'entrée à la sortie, sans retour (boucle). La sortie de chaque couche n'affecte pas la même couche.
- En arrière : les données sont renvoyées des nœuds de la couche supérieure vers la couche inférieure nœuds de couche.
- Côté : les sorties des nœuds de chaque couche sont utilisées comme entrées pour les nœuds de la même couche.

3.2.2 Types des réseaux de neurones artificiels

Voici les types de réseaux de neurones les plus connus (Figure 3.4) :

1. Réseau de neurones Perceptron multicouche (MLPNN).
2. Réseaux de neurones à fonction de base radiale (RBFNN).

3. Réseau de neurones Hopfield (HN).
4. Réseau de neurones de Hamming (HNN).
5. Réseau neuronal cartographique auto-organisé de Kohonen (KSOMNN).
6. Réseau de neurones à retardement (TDNN).
7. Réseaux de neurones à alimentation profonde (DFFNN).
8. Réseaux de neurones récurrents (RNN).
9. Réseaux de neurones à mémoire à long et à court terme (NNLSTM).
10. Réseau neuronal des encodeurs automatiques (AENN).
11. Réseaux de chaînes de Markov (MCN).

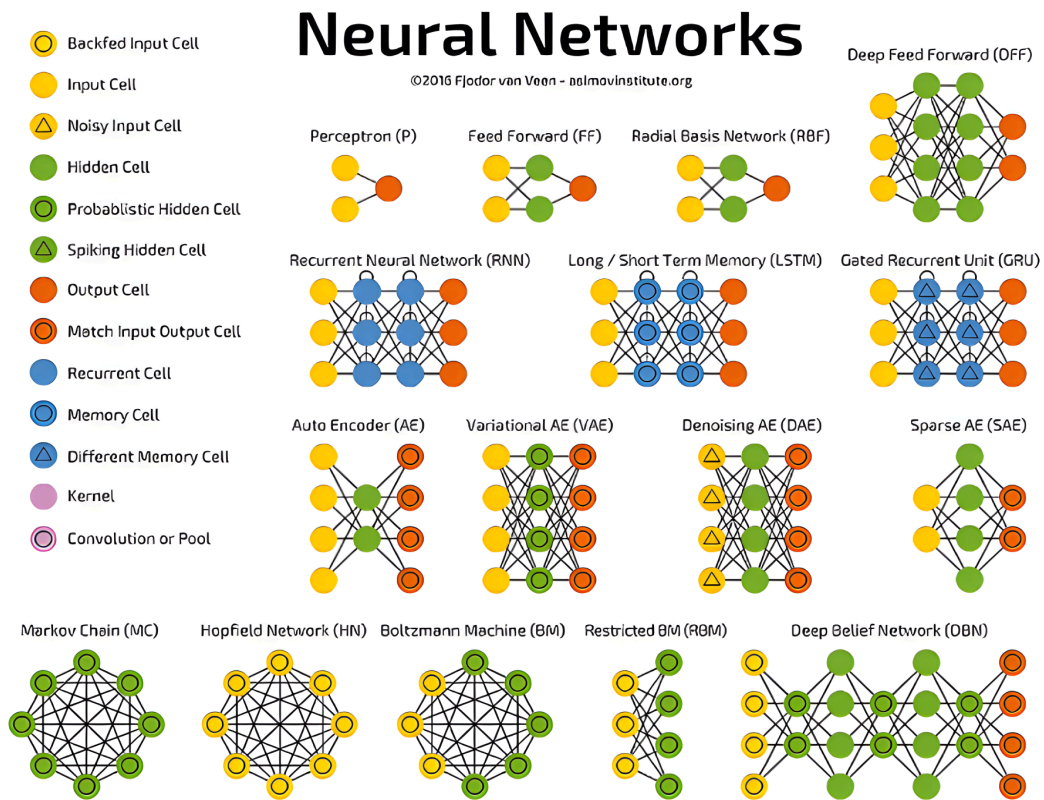


Figure 3-4 : Les types de réseaux de neurones les plus connus.

3.2.3 L'initialisation du poids

L'objectif de l'initialisation des poids dans les réseaux de neurones artificiels est d'empêcher les sorties d'activation de couche d'exploser ou de disparaître pendant la phase d'apprentissage. L'apprentissage du réseau sans une initialisation de poids peut conduire à une convergence très lente ou à une incapacité à converger [40].

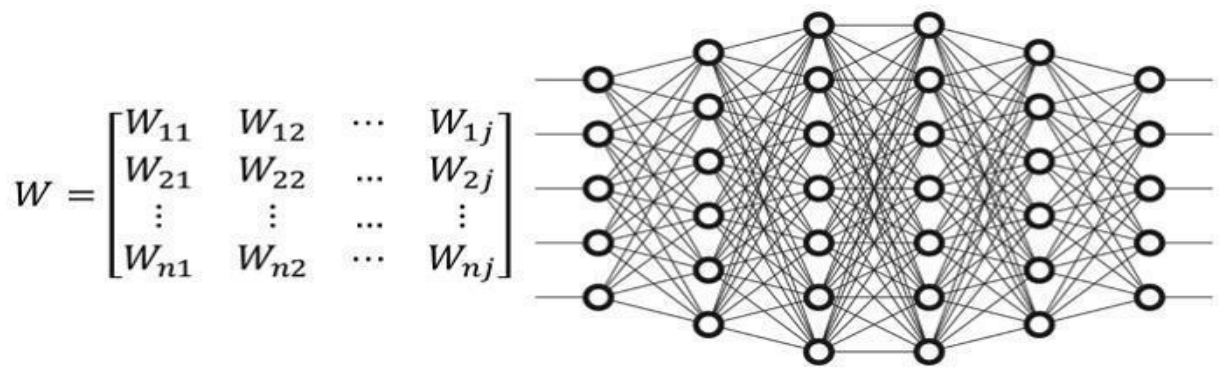


Figure 3-5 : Schéma représente le processus d'initialisation du poids.

Les techniques d'initialisation du poids [41] les plus utilisées sont les suivantes :

➤ Initialisation tous zéros et initialisation constante

Cette méthode définit tous les poids à zéro (respectivement à une constante). De plus, toutes les activations dans tous les neurones sont les mêmes, et donc tous les calculs sont les mêmes, ce qui fait que ce type d'initialisation concerne les modèles linéaires.

➤ Initialisation aléatoire

Cette technique améliore le processus de brisure de symétrie et offre une plus grande précision. Les poids sont initialisés très près de zéro et aléatoirement. Cette méthode empêche l'apprentissage de la même fonctionnalité pour l'entrée paramètres.

➤ Initialisation LeCun (normaliser la variance)

L'initialisation LeCun vise à empêcher la disparition ou l'explosion des gradients lors de la rétropropagation en résolvant la variance croissante avec le nombre d'entrées et en fixant une variance constante.

➤ Initialisation de Xavier (initialisation de Glorot)

Xavier a proposé une méthode plus simple, où les poids tels que la variance des activations sont les mêmes sur chaque couche. Cela empêchera le gradient d'exploser ou de disparaître.

➤ Initialisation HE (initialisation de Kaiming)

Cette initialisation préserve la non-linéarité des fonctions d'activation telles que les activations ReLU. L'initialisation des poids avec la méthode HE peut réduire ou amplifier les amplitudes des entrées de manière exponentielle.

Méthode d'initialisation	Avantages	Inconvénients
Tout-zéros/constante	✓ Simplicité	❖ Problème de symétrie amenant les neurones à apprendre la même chose
Aléatoire (Random)	✓ Améliore le processus de rupture de symétrie	❖ Une saturation peut se produire conduisant à un gradient de fuite. ❖ La pente ou le gradient est faible, ce qui peut ralentir la descente du gradient.
LeCun	✓ Résout la variance croissante et problèmes de gradient	❖ Pas utile en largeur constante des réseaux. ❖ Tient compte de la propagation vers l'avant du signal d'entrée. ❖ Cette méthode n'est pas utile lorsque la fonction d'activation est non différentiable.
Xavier	✓ Diminue la probabilité du problème de disparition/explosion du gradient	❖ Cette méthode n'est pas utile lorsque la fonction d'activation est non différentiable. ❖ Problème de neurone mourant pendant l'entraînement.
HE	✓ Résout les problèmes de neurones mourants.	❖ Cette méthode n'est pas utile pour les couches avec les fonctions d'activations différentiables telles que ReLU ou LeakyReLU.

Tableau 3-1: Comparaison entre les techniques d'initialisation du poids.

3.3 Les fonctions d'activation

Les fonctions d'activation peuvent être de différents types en fonction du problème ou des données. Il existe deux types de fonctions d'activation (linéaire et non linéaire). Dans le mode d'activation linéaire, un neurone linéaire prend une combinaison linéaire des entrées pondérées, et la sortie peut prendre n'importe quelle valeur entre moins l'infini et plus l'infini. Dans le mode d'activation non linéaire, la sortie se limite entre une certaine gamme. Les fonctions d'activation les plus connues sont les suivantes :

3.3.1 La fonction d'activation sigmoïde ou logit

La fonction d'activation sigmoïde permet de réduire la sortie entre 0 et 1 en appliquant la fonction **log**, qui rend les problèmes de classification plus faciles [42].

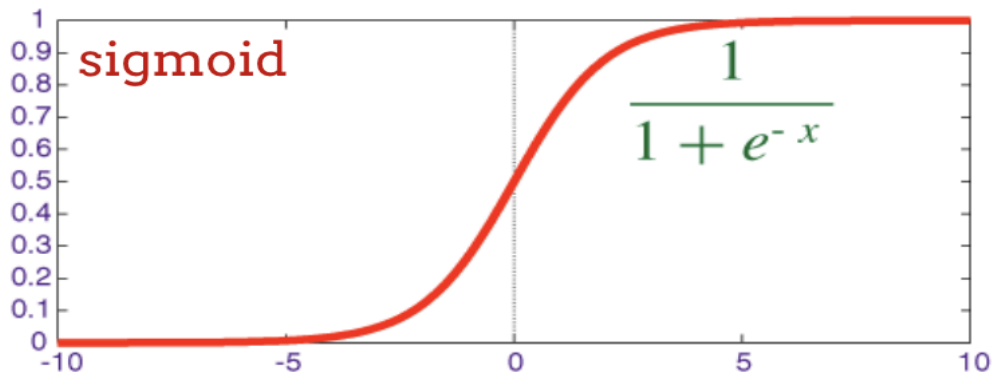


Figure 3-6: Schéma représente la fonction d'activation Sigmoid.

3.3.2 La fonction Softmax

La fonction Softmax est presque similaire à la sigmoïde, mais elle calcule les probabilités des événements sur 'n' classes différentes, ce qui est utile pour déterminer la cible dans la classification multi-classes.

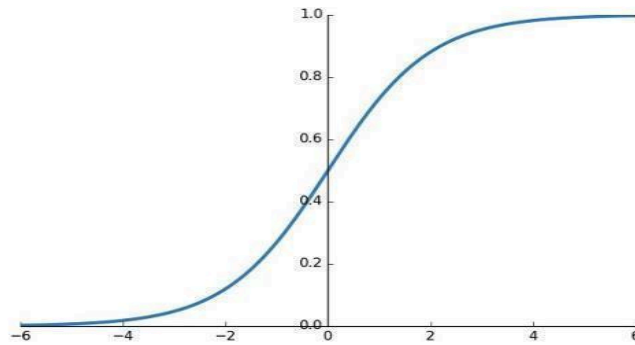


Figure 3-7: Schéma représente la fonction d'activation Softmax.

3.3.3 La fonction Tanh

La plage de la fonction d'activation tanh est de (-1 à 1) comme le montre la Figure 3.8:

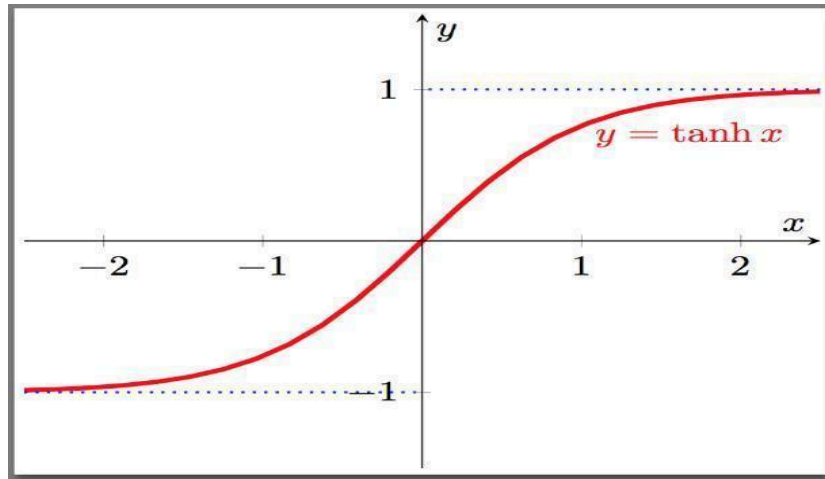


Figure 3-8: Schéma représente la fonction d'activation Tanh.

3.3.4 La fonction de l'Unité Linéaire Rectifiée (ReLU)

La fonction de l'unité linéaire rectifiée (ReLU) [43] est l'une des fonctions d'activation non saturées les plus remarquables. Elle est définie comme suit : $a_{i,j,k} = \max(z_{i,j,k}, 0)$.

ReLU convertit tout ce qui est inférieur à zéro en zéro. Alors, la plage devient de 0 à l'infini.

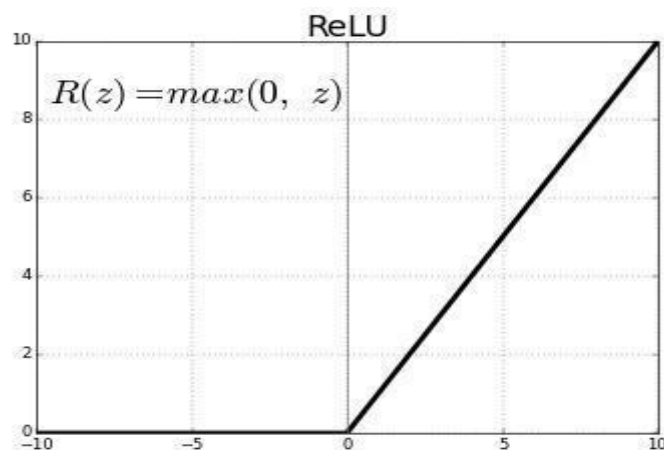


Figure 3-9 : Schéma représente la fonction d'activation ReLU.

3.3.5 La fonction LeakyReLU

Un inconvénient potentiel de l'unité ReLU est qu'elle a un gradient nul chaque fois que le neurone n'est pas activé. Cela peut faire en sorte que les neurones qui ne sont pas activés initialement ne soient jamais actifs car l'optimisation basée sur le gradient n'ajustera pas leurs poids. Pour pallier ce problème, Mass et al. ont introduit LeakyReLU (LReLU) qui est défini comme :

$$a_{i,j,k} = \max(z_{i,j,k}, 0) + \lambda \min(z_{i,j,k}, 0)$$

Où λ est un paramètre prédéfini dans la plage (0,1). Par rapport à ReLU, Leaky ReLU comprime la

partie négative plutôt que de la mapper sur un zéro constant, ce qui permet un petit gradient non nul lorsque le neurone n'est pas activé.

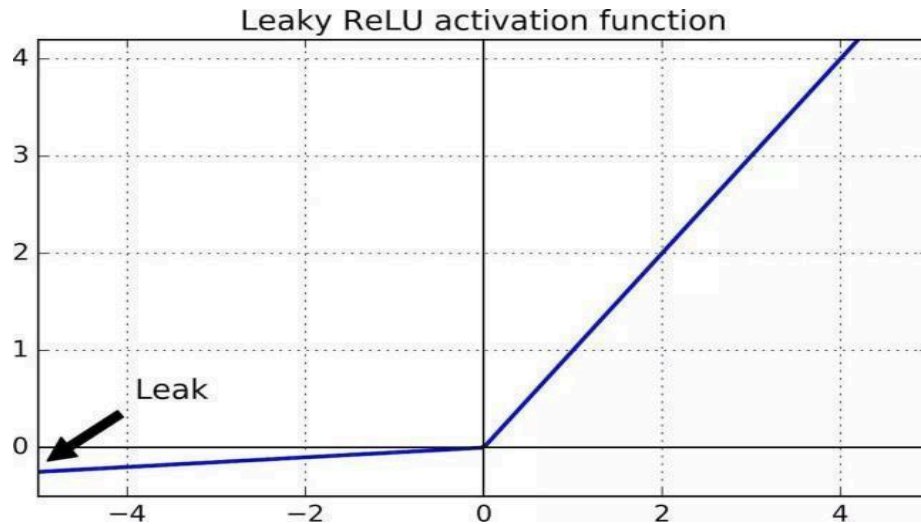


Figure 3-10: Schéma représente la fonction d'activation LeakyReLU.

3.3.6 La fonction ReLU paramétrique (PReLU)

Plutôt que d'utiliser un paramètre prédéfini λ dans LeakyReLU, la fonction d'activation de l'unité linéaire rectifiée paramétrique (PReLU) apprend de manière adaptative les paramètres des redresseurs afin d'améliorer la précision [44]. Mathématiquement, la fonction PReLU est définie comme :

$$a_{i,j,k} = \max(z_{i,j,k}, 0) + \lambda_k \min(z_{i,j,k}, 0)$$

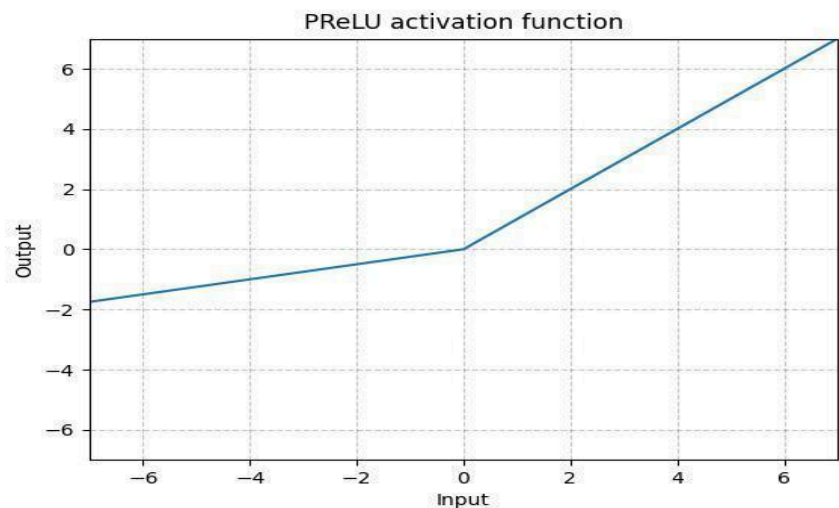


Figure 3-11: Schéma représente la fonction d'activation PReLU.

Où λ_k est le paramètre appris pour le k-ième canal. Comme PReLU n'introduit qu'un très petit nombre de paramètres supplémentaires, il n'y a pas de risque supplémentaire de sur-apprentissage et le coût de calcul supplémentaire est négligeable.

3.3.7. Fonction d'activation SWISH

La fonction d'activation automatique (SWISH) est en fait une version de la fonction sigmoïde, et est définie comme suit :

$$f(x) = x * \sigma(x)$$

Où $\sigma(x)$ est le sigmoïde de x donné par : $\sigma(x) = \frac{1}{1 + e^{-x}}$.

La fonction d'activation SWISH est une fonction non monotone.

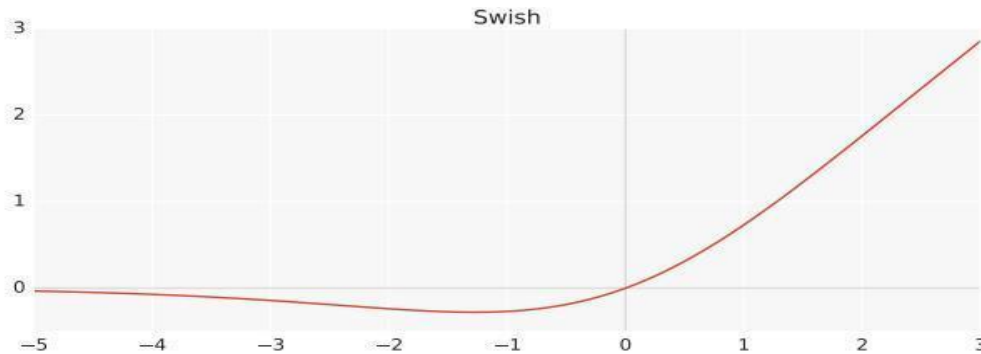


Figure 3-12 : Schéma représente la fonction d'activation SWISH.

3.4 Les fonctions de pertes (erreurs)

La fonction de perte est la métrique qui aide un réseau à comprendre s'il apprend dans la bonne direction. Pour encadrer la fonction de perte en termes simples, considérez-la comme le score que vous obtenez lors d'un examen. Supposons qu'un étudiant veut participer à plusieurs tests sur le même sujet : quelle métrique utilisera l'étudiant pour comprendre ses performances à chaque test ? De toute évidence, le score du test. Supposons qu'il a obtenu 11, 12, 15, 17 et 19 sur 20 dans cinq tests de langue consécutifs.

Il est clairement que l'amélioration des résultats des tests est une indication de ses performances. Si les résultats des tests avaient diminué, le verdict serait que ses performances diminuent et qu'il aura besoin de changer ses méthodes d'étude ou son matériel pour qu'il s'améliore.

De même, comment un réseau comprend-il s'il s'améliore son processus d'apprentissage à chaque itération? Il utilise la fonction de perte, qui est analogue au score du test. La fonction de perte mesure essentiellement la perte de la cible. Supposons que nous développons un modèle pour prédire si un élève réussira ou échouera et que la probabilité de réussite ou d'échec est définie par la probabilité. Perte d'entropie croisée. Ainsi, 1 indiquerait qu'il réussira avec une certitude de 100 % et 0 indiquerait qu'il échouera définitivement.

Le modèle apprend des données et prédit un score de 0,87 pour l'étudiant à passer. Ainsi, la

perte réelle ici serait de $1,00 - 0,87 = 0,13$. S'il répète l'exercice avec quelques mises à jour de paramètres afin de s'améliorer et obtient maintenant une perte de 0,40, il comprendrait que les changements qu'il a apportés n'aident pas le réseau à apprendre correctement. Alternativement, une nouvelle perte de 0,05 indiquerait que les mises à jour ou les changements de l'apprentissage vont dans la bonne direction.

En fonction du type de résultat de données, nous avons plusieurs pertes standard fonctions définies dans l'apprentissage par ordinateur. Pour les cas d'utilisation de régression (c'est-à-dire, où la prédiction finale serait un nombre continu comme les notes obtenues par un étudiant, le nombre d'unités de produits vendues par un magasin, le nombre d'appels reçus de clients dans un centre de contact, etc.), voici quelques fonctions de perte populaires disponibles [45].

3.4.1 Erreur Quadratique Moyenne

Cette erreur est définie par la différence quadratique moyenne entre la valeur réelle et prédite. La différence au carré permet de pénaliser plus facilement le modèle pour une différence plus élevée. L'équivalent mathématique est défini par la fonction suivante :

$$\sum_{n=1}^K \frac{(\text{Actuelle} - \text{prédite})^2}{K}$$

3.4.2 Erreur Absolue Moyenne

L'erreur absolue moyenne entre le réel et le prévu. L'équivalent mathématique serait :

$$\sum_{n=1}^K |\text{Actuelle} - \text{prédite}|$$

De même, peu d'autres variantes sont :

- MAPE : Erreur absolue moyenne en pourcentage (Mean Absolute Percentage Error).
- MSLE : Erreur logarithmique quadratique moyenne (Mean Square Logarithmic Error).

Pour les résultats catégoriels, la prédiction serait pour une classe, comme si un étudiant réussira (1) ou échouera (0), si le client réussira un achat ou non, que le client soit en défaut de paiement ou non, etc. Certains cas d'utilisation peuvent avoir plusieurs classes comme résultat, comme classer les types de maladies (type A, B ou C), classer les images en tant que chats, chiens, voitures, chevaux, paysages, etc. Dans de tels cas, les pertes définies ci-dessus ne peuvent être utilisées en raison à des raisons évidentes. Nous aurions besoin de quantifier le résultat de la classe comme probabilité et définir les pertes sur la base des estimations de probabilité comme prédictions

3.4.3 Entropie croisée binaire

Définit la perte lorsque les résultats catégoriels sont une variable binaire, c'est-à-dire avec deux classes possibles : (Réussite / Échec) ou (Oui / Non). La forme mathématique de la fonction de l'entropie croisée binaire est la suivante :

$$Pert = - \left[y^* \log(p) + (1 - y^*) \log(1 - p) \right]$$

3.4.4. Entropie croisée catégorielle

Définit la perte lorsque les résultats catégoriels sont non binaires, c'est-à-dire le nombre de classes possibles est supérieur à 2. La forme mathématique de l'entropie croisée catégorielle serait :

$$Pert = \sum_i y_i' \cdot \log_2(y_i) .$$

3.5 Les Optimiseurs

Pour trouver un minimum local d'une équation différentielle, il est nécessaire d'utiliser un algorithme d'optimisation itératif. La théorie consiste à effectuer des pas répétés dans la direction opposée du gradient au stade actuel (ou un gradient approximatif) de la fonction car cela est considéré comme la direction de descente vers l'erreur minimale. D'un autre côté, un pas dans la direction du gradient se traduira par un maximum local de cette caractéristique. Ensuite, l'approche est appelée ascension de gradient [46].

3.5.1. La descente de gradient (GD)

Il existe trois variantes de descente de gradient, qui varient dans la quantité de données que nous utilisons pour évaluer le gradient de la fonction objectif [47]. En fonction de la quantité de données [48], nous avons un compromis entre l'exactitude du paramètre de mise à jour et le temps d'exécution d'une mise à jour.

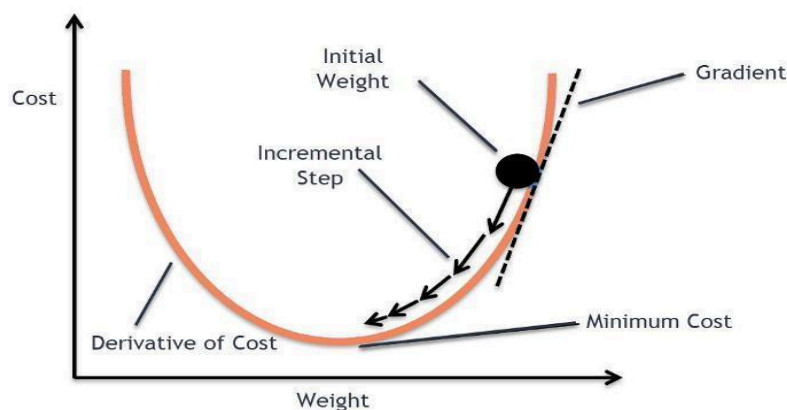


Figure 3-13: Algorithme de descente de gradient.

3.5.2 Descente de gradient par lots (BGD)

La descente de gradient par lots, parfois nommé vanilla GD, détecte l'erreur pour chaque individu de la base d'apprentissage. Mais les poids du modèle sont révisés après que tous les individus de cette base ont été évalués. Tout ce processus est comme une boucle, et c'est ce qu'on appelle une époque d'entraînement [50]. Certains des avantages de l'optimisation avec descente de gradient par lots sont son calcul efficace, un gradient d'erreur équilibré et une convergence stable [51]. Certaines faiblesses sont que le gradient d'erreur constant peut souvent conduire à une condition de convergence qui n'est pas la meilleure que le modèle puisse accomplir. Cela exige également que les éléments de la base d'apprentissage soit en mémoire et disponible pour l'algorithme [52]. BGD offre les performances les plus précises mais nécessite de nombreuses analyses complètes coûteuses des données réelles [53]. L'optimisation par la descente de gradient par lots calcule le gradient de la fonction d'erreur concernant les paramètres θ pour la base d'apprentissage, comme le montre l'équation suivante :

$$\theta_{\text{nouveau}} = \theta_{\text{ancien}} - \eta \cdot \nabla_{\theta} J(\theta)$$

Où:

θ_{nouveau} = *position suivante*

θ_{ancien} = *position actuelle*

η = taux d'apprentissage (taille du pas)

$\Delta J(\theta)$ = Direction de l'augmentation la plus rapide

3.5.3 Descente de gradient stochastique (SGD)

Les algorithmes SGD se sont avérés efficaces dans l'optimisation des modèles d'apprentissage à grande échelle. Le terme stochastique désigne un mécanisme ou une méthode liée à une possibilité aléatoire ; par conséquent, au lieu de passer toute la base d'apprentissage à chaque itération pour mettre à jour les poids du modèle, seulement quelques échantillons sont choisis [54]. SGD vise à trouver le minimum global en modifiant la structure du réseau après chaque étape d'apprentissage [55]. Cette approche réduit simplement l'erreur en se rapprochant du gradient pour un lot choisi au hasard au lieu de trouver le gradient de la base d'apprentissage. En réalité, l'échantillonnage aléatoire est effectué en mélangeant au hasard la base d'apprentissage et en se déplaçant pas à pas dans les lots de la fonction objective, comme le montre la Figure 3.14.

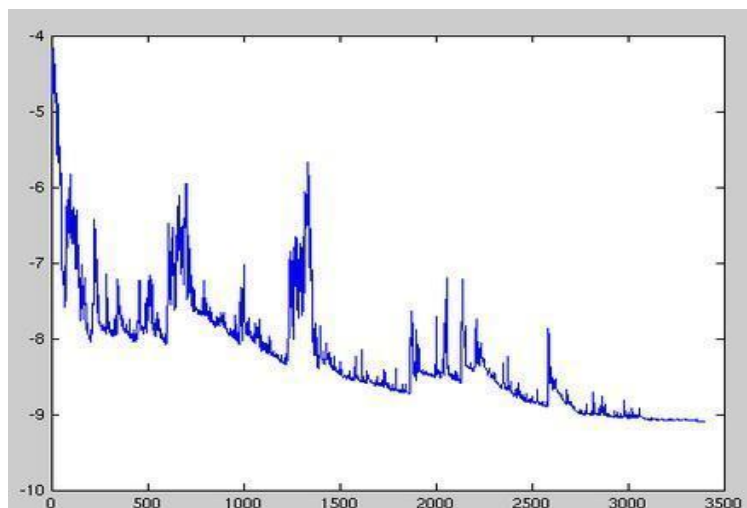


Figure 3-14: Descente de gradient stochastique.

3.5.4 Descente de gradient en mini-lot

Une autre variante des principes de SGD et BGD est le mini-batch GD. Il divise la base d'apprentissage en lots plus petits et effectue une mise à jour pour chacun de ces lots. Cela fournit un équilibre entre la robustesse du SGD et l'efficacité du GD par lots [56].

Pour plusieurs problèmes d'optimisation, ces trois algorithmes GD (Batch GD, SGD et Mini-batch GD) fonctionnent bien et peuvent tous converger vers un optimum prometteur (local Chapitre II : Les réseaux de neurones artificiels et à convolution 41 ou global), mais ils souffrent toujours de nombreux problèmes tels que ceux mentionnés cidessous :

Le taux d'apprentissage η peut avoir un effet majeur sur la convergence des algorithmes GD. Il existe un compromis entre le taux de convergence et le dépassement lors de la définition d'un taux d'apprentissage. Si le taux d'apprentissage est trop élevé, nous pourrions dépasser le minimum et commencer à rebondir sans atteindre le minimum. En revanche, si le taux d'apprentissage est trop faible, la formation peut être trop longue [57].

Dans la plupart des cas, pour les algorithmes GD, un taux d'apprentissage fixe ne fonctionne pas bien dans le processus de mise à jour des poids. L'algorithme nécessitera un taux d'apprentissage plus élevé dans les étapes initiales pour obtenir un optimum rapide réussi (local ou global). Dans les dernières étapes cependant, l'algorithme devra modifier le taux d'apprentissage [57].

Pour différentes variables, au lieu du processus de mise à jour, leur mise à niveau peut simplement impliquer un taux d'apprentissage différent. Par conséquent, il est nécessaire et approprié d'utiliser un taux d'apprentissage individuel pour diverses variables [57].

Un autre défi majeur est d'éviter d'être coincé dans leurs multiples minima locaux sousoptimaux en minimisant les fonctions d'erreur extrêmement non convexes communes aux réseaux de neurones. Les minima locaux sont le plus grand défi pour analyser la convergence comme la Figure 3.15.

Les algorithmes d'optimisation constituent la base sur laquelle une machine peut bénéficier de sa pratique. Ils mesurent des gradients et tentent de minimiser la fonction des coûts. L'apprentissage

peut être mis en œuvre de plusieurs manières avec différents types de méthodes d'optimisation [57].

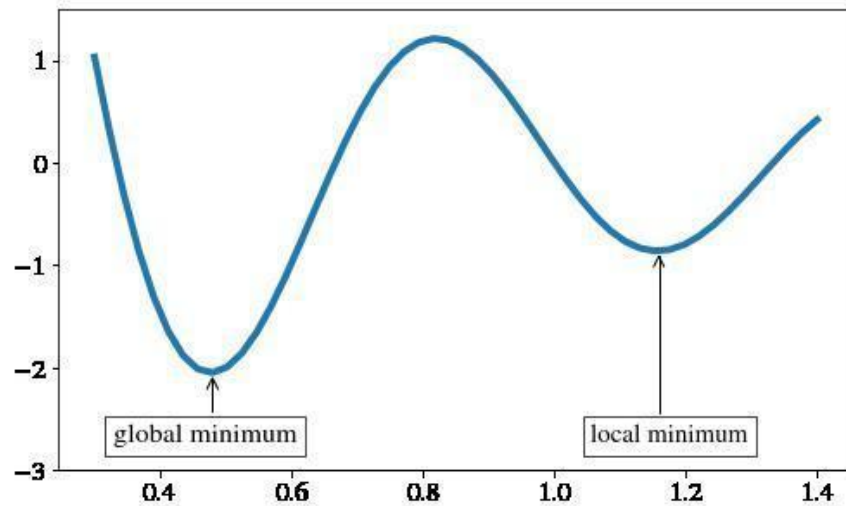


Figure 3-15: Minima globaux Vs. Minima locaux

3.5.5 Momentum

Une technique d'optimisation courante est SGD, mais lors de l'apprentissage du modèle, le temps d'exécution est relativement élevé. Momentum est conçu pour apprendre rapidement, en particulier face à de larges courbures, des gradients petits mais bruyants ou des gradients stables. L'utilisation d'un terme de quantité de mouvement est une autre approche qui peut aider le réseau à se débarrasser des minima locaux [58]. C'est peut-être l'extension la plus courante de l'algorithme de rétropropagation. D'autres cas dans lesquels cette approche n'est pas utilisée sont difficiles à trouver. Le gradient ne pointe pas vers le minimum à des points spécifiques de la surface, et les pas successifs de GD oscilleront d'un côté à l'autre, n'avançant que très lentement vers le minimum. La Figure 16 illustre comment l'incorporation de la quantité de mouvement, en amortissant ces oscillations, tend à conduire à la convergence vers le minimum [58].

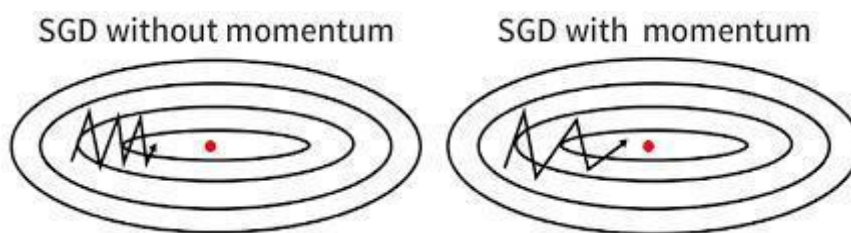


Figure 3-16: La comparaison des algorithmes SGD avec et sans quantité de mouvement

3.5.6 Nesterov Momentum

Sur la base de paramètres qui estiment des positions potentielles plutôt que des paramètres actuels, le gradient est calculé. Nesterov Momentum est une amélioration par rapport à l'élan et ne décide pas de l'emplacement futur du paramètre [59]. Le Nesterov est une version de l'algorithme de momentum inspiré de la méthode du gradient accéléré de Nesterov. La distinction entre cette approche

et la méthode de l'impulsion est que la vitesse est déjà ajoutée aux paramètres lors du calcul du gradient dans la méthode Nesterov. Cela peut être considéré comme une tentative d'incorporer un facteur correctif à la forme de quantité de mouvement conventionnelle [59].

3.5.7 Descente de gradient adaptative (AdaGrad)

C'est une méthode qui sélectionne le rythme d'apprentissage en fonction de la situation. Comme le taux réel est calculé à partir de paramètres, les taux d'apprentissage ont tendance à s'adapter. Un gradient de paramètre plus élevé aura un taux d'apprentissage réduit et vice versa [60]. La théorie d'AdaGrad est similaire à l'algorithme AdaDelta en ce sens qu'elle mesure différents taux d'apprentissage pour d'autres éléments de paramètres. Pourtant, il utilise l'agrégation des carrés de gradient : contrairement à AdaDelta, il utilise la moyenne mobile des carrés de gradient [60].

3.5.8 Delta adaptatif (AdaDelta)

AdaDelta est l'extension AdaGrad. AdaDelta fonctionne en utilisant plusieurs fenêtres de taille fixe au lieu d'accumuler les dégradés. Il ne surveillera que les gradients disponibles à l'intérieur de la fenêtre. Étant donné que l'algorithme SGD nécessite une sélection manuelle du taux d'apprentissage, le taux d'apprentissage inapproprié choisi conduirait à une faible précision de prédiction. Cependant, en tant qu'optimisation de l'algorithme SGD, l'Adadelta, connu sous le nom d'algorithme de taux d'apprentissage adaptatif (LR adaptatif), peut ajuster automatiquement le taux d'apprentissage et augmenter la précision de la prédiction [61].

3.5.9 Propagation racine carrée moyenne (RMSProp)

L'un des algorithmes stochastiques adaptatifs les plus courants pour l'apprentissage profond (Deep Neural Network ou DNN) est RMSProp. RMSProp modifie AdaGrad de manière à accumuler le gradient. Les dégradés s'agrègent en une moyenne pondérée de manière exponentielle. RMSProp rejette le passé et ne préserve que les connaissances actuelles sur le gradient [62].

3.5.10 Estimation adaptative du moment (Adam)

Adam est une méthode d'optimisation SGD qui mesure des taux d'apprentissage adaptables pour chaque paramètre. Adam est l'une des stratégies d'optimisation la plus courante dans le domaine des réseaux de neurones. Le nom a été tiré d'Adaptive Moments. C'est un mélange de RMSProp et de Momentum. L'opération de mise à niveau considère la variante de gradient lisse et fournit un mécanisme de correction de biais. Adam réduit les coûts de calcul, nécessite moins de mémoire d'exécution et est invariant au redimensionnement diagonal du gradient. RMSprop est un optimiseur basé sur le gradient qui utilise un taux d'apprentissage adaptatif (LR adaptatif) qui varie dans le temps au lieu de traiter le taux d'apprentissage comme un hyperparamètre [62].

3.5.11 Estimation du moment adaptatif maximal (AdaMax)

AdaMax est une forme de SGD avec taux d'apprentissage adaptatif, et une version Adam basée sur la norme de l'infini. AdaMax présente l'avantage majeur d'être beaucoup moins sensible à l'option des hyper-paramètres par rapport au SGD. L'équation AdaMax utilise la valeur complète de la deuxième composante de quantité de mouvement de la méthode d'estimation ADAM. Cela offre une solution plus stable.

3.5.12 Estimation du moment adaptatif accélérée par Nesterov (Nadam)

Cette technique est une combinaison d'approches d'optimisation Adam et RMSProp. Elle a été conçue de manière similaire à l'approche d'optimisation d'Adam. Cependant, elle a été remplacé par l'élan de Nesterov. Cela se traduit par une augmentation marquée de la surperformance du momentum [63].

3.4 Apprentissage profond (Deep learning ou DL)

L'apprentissage en profondeur (DL) est l'une des plus prometteuses techniques d'apprentissage qui ont été appliquées dans plusieurs domaines. Les architectures basées sur DL sont caractérisées par de nombreuses couches cachées de neurones. Cependant, la principale limite de ces architectures est le long temps nécessaire à l'apprentissage. Obtenir une excellente précision et un temps d'apprentissage raisonnable est un objectif ambitieux pour la communauté de recherche DL.

3.4.1 Les réseaux de neurones convolutifs (CNN)

Le réseau de neurones à convolution, également connu sous les noms (ConvNet, CNN ou apprentissage en profondeur), est une technique qui se compose de plusieurs couches. Chaque couche de convolution est connue sous le nom de carte de caractéristiques.

En imagerie, la carte des caractéristiques de la couche d'entrée est une matrice 3D d'intensités de pixels pour différents canaux de couleur (par exemple RVB). La carte des caractéristiques de n'importe quelle couche interne est un multicanal induit l'image d'entrée, dont le « pixel » peut être considéré comme une caractéristique spécifique. Tous les neurones sont connectés à une petite partie des neurones adjacents de la couche précédente. Différents types de transformations peuvent être effectuées sur les cartes, telles que le filtrage (Convolution) et la mise en commun (pooling) [64].

Dans la phase de filtrage (convolution), l'opération convolue une matrice de filtre (poids appris) avec les valeurs d'un champ récepteur de neurones et prend une fonction non linéaire (telle que sigmoïde, ReLU...) pour obtenir les réponses.

L'opération de mise en commun (pooling), telle que la mise en commun maximale (max pooling), la moyenne pooling, L2-pooling et normalisation locale du contraste, résume les réponses d'un champ récepteur en une seule valeur pour produire des descriptions de fonctionnalités plus robustes. Avec un entrelacement entre la convolution et la mise en commun, une hiérarchie initiale des caractéristiques est construite, qui peut être affinée de manière supervisée en ajoutant une couche de neurones entièrement

connectés (fully connected layer FC) pour s'adapter aux différentes tâches visuelles.

Selon les tâches impliquées, la couche finale avec différentes fonctions d'activation est ajoutée pour obtenir une probabilité conditionnelle spécifique pour chaque neurone de sortie, et le réseau peut être optimisé sur une fonction objective (par exemple erreur quadratique moyenne ou entropie croisée) via la méthode de descente de gradient stochastique (SGD).

Par exemple, le VGG16 typique a au total 13 couches convolutionnelles, 3 couches entièrement connectées, 3 couches max-pooling et une couche softmax finale (couche de classification). Les filtres de convolutions sont produits par fenêtres de filtre (3×3) alambiquées et résolutions de cartes sont réduits avec 2 couches max-pooling. Pour tester le modèle, une image de la même taille que les échantillons d'apprentissage peut être traitée avec ce modèle formé. Dans ce cas, une opération de remise à l'échelle ou de recadrage peut être nécessaire si différentes tailles sont fournies [64].

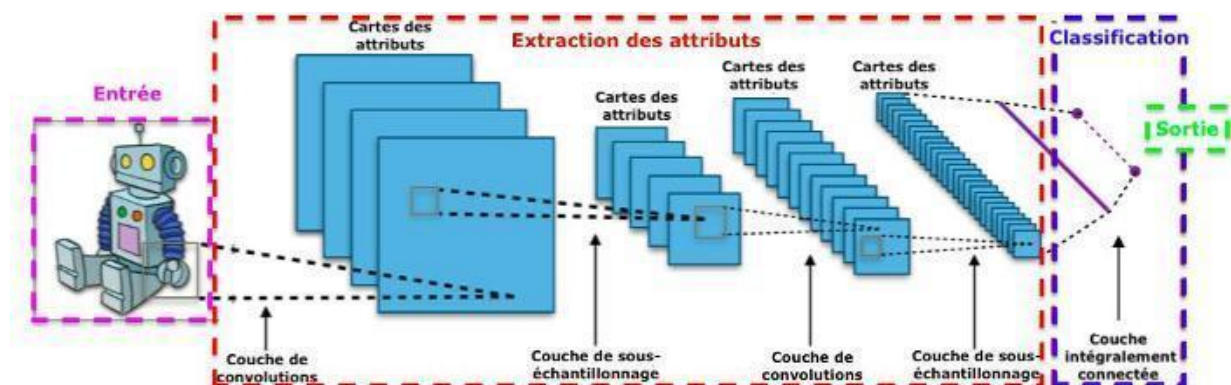


Figure 3-17 : Structure du réseau neuronal à convolution.

3.4.1 Les avantages de CNN

Les avantages des réseaux à convolution par rapport aux méthodes de classifications traditionnelles peuvent être résumés comme suit :

- Représentation hiérarchique des caractéristiques, qui correspond aux représentations multi niveaux du pixel aux caractéristiques sémantiques de haut niveau apprises par une structure hiérarchique à plusieurs étapes [65].
- L'architecture des réseaux CNN offre la possibilité d'optimiser conjointement plusieurs tâches connexes (par exemple, Fast RCNN combine la classification et la régression de la boîte englobante dans un mode d'apprentissage multitâches).
- Bénéficiant de la grande capacité d'apprentissage des CNN, certains défis classiques de la vision par ordinateur peuvent être reformulés en tant que problèmes de transformation de données de grande dimension et résolus d'un point de vue différent.
- En raison de ces avantages, les réseaux à convolution ont été largement appliqués dans de nombreux domaines de recherche, tels que la super-résolution d'images, reconstruction, classification d'images, récupération d'images, reconnaissance faciale, détection et analyse

3.4.2 Les inconvénients de CNN

Malgré ses progrès par rapport aux méthodes traditionnelles et l'importance des réseaux à convolution dans la détection pratique d'objets, il y a encore quelques inconvénients : En raison de l'existence de couches entièrement connectées (fully connected), les réseaux de neurones à convolution nécessitent une image d'entrée de taille fixe (par exemple, 227×227), qui mène directement au recalcul du modèle pour chaque région, prenant beaucoup de temps dans la période de test.

L'apprentissage du R-CNN est un pipeline en plusieurs étapes. En premier, le réseau convolutif (ConvNet) sur les propositions d'objets est peaufiné. Ensuite, le classifieur softmax appris par réglage fin est remplacé par des SVM pour s'adapter aux fonctionnalités de ConvNet. Enfin, les régresseurs de la boîte englobante sont entraînés. L'apprentissage est coûteux en espace et en temps. Les caractéristiques sont extraites de différentes propositions de région et stockées sur le disque. Il faudra beaucoup de temps pour traiter un problème avec une petite base de données avec les techniques des réseaux à convolution, tels que VGG16. Au même temps, la mémoire de stockage requise par ces caractéristiques devrait également être un sujet de préoccupation.

Bien que la recherche sélective puisse générer des propositions de région avec des rappels relativement élevés, les propositions de régions obtenues sont toujours redondantes et cette procédure prend beaucoup de temps.

3.4.3. L'opération de convolution

L'opération de convolution est très similaire aux filtres de convolution dans le domaine du traitement d'images tels que le filtre de Sobel, Robert et le filtre gaussien. Le noyau du filtre se glisse sur une image et se multiplie les poids avec chaque pixel aligné, élément par élément à travers ce filtre. Ensuite, la valeur de biais est ajoutée pour le résultat [66].

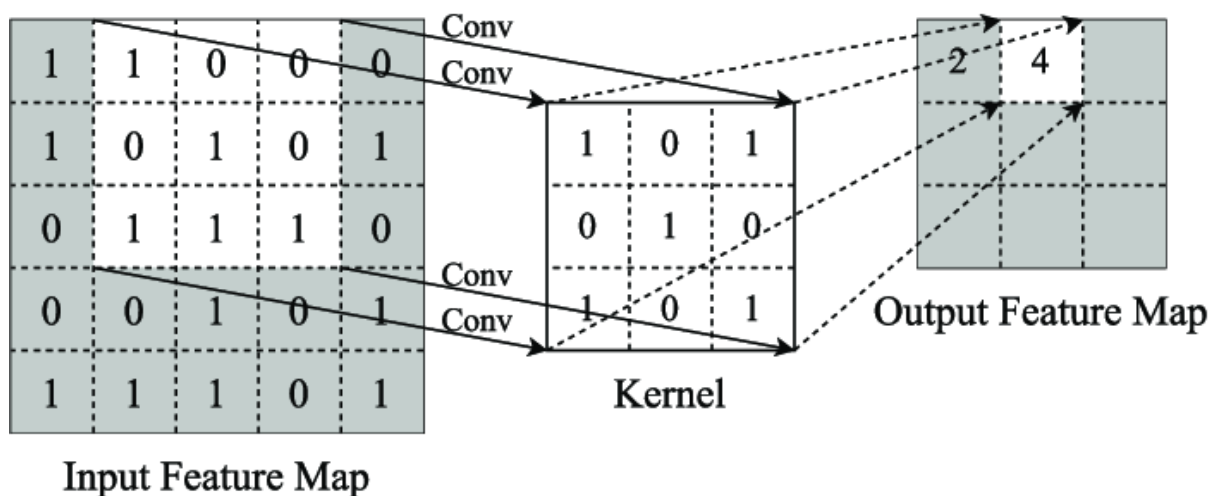


Figure 3.18: Schéma représente l'opération de convolution.

Dans la couche de convolution, les trois hyperparamètres qui produisent la carte des caractéristiques

sont :

Un pas (Stride S): représente le pas à chaque fois que nous faisons glisser le filtre de convolution. Lorsque le pas $S = 1$, nous déplaçons les filtres d'un pixel à chaque fois. Lorsque le pas $S = 2$ (ou 3 ou plus), alors les filtres sautent 2 pixels à un temps comme nous glissons [67].

Rembourrage (Padding P): les images en entrées seront remplies avec une bordure de taille fixée selon la valeur spécifiée. Plus communément, le rembourrage zéro (zéro-padding) est utilisé pour tampon. Dans le framework de réseaux de neurones (caffè, TensorFlow, PyTorch, MXNet), la taille du padding « P » est un hyperparamètre. La taille du padding peut également être utilisée pour contrôler la taille spatiale des volumes de sorties [67].

Profondeur (Depth D): la profondeur du volume de sortie est également un hyperparamètre ; c'est correspond pour le nombre de filtres que nous utilisons pour une couche de convolution [67].

3.4.4 Mise en commun (Pooling)

De nos jours, Les réseaux de neurones à convolution exploitent toujours un partage de poids étendu pour réduire les degrés de la liberté des modèles [68]. Une couche de mutualisation permet de réduire le temps de calcul et construire progressivement une invariance spatiale et configurative. Pour la compréhension des images, la couche de mise en commun (pooling) aide à extraire plus de sens sémantique. Par exemple la couche de pooling maximum (Max-pooling) simplement renvoie la valeur maximale sur les valeurs sur lesquelles l'opération du noyau est appliquée. L'exemple ci-dessous illustre les sorties d'une mise en commun maximale (max_pooling) et d'une mise en commun moyenne opération (average pooling), respectivement, donné sur un noyau de taille 2 et un pas = 2.

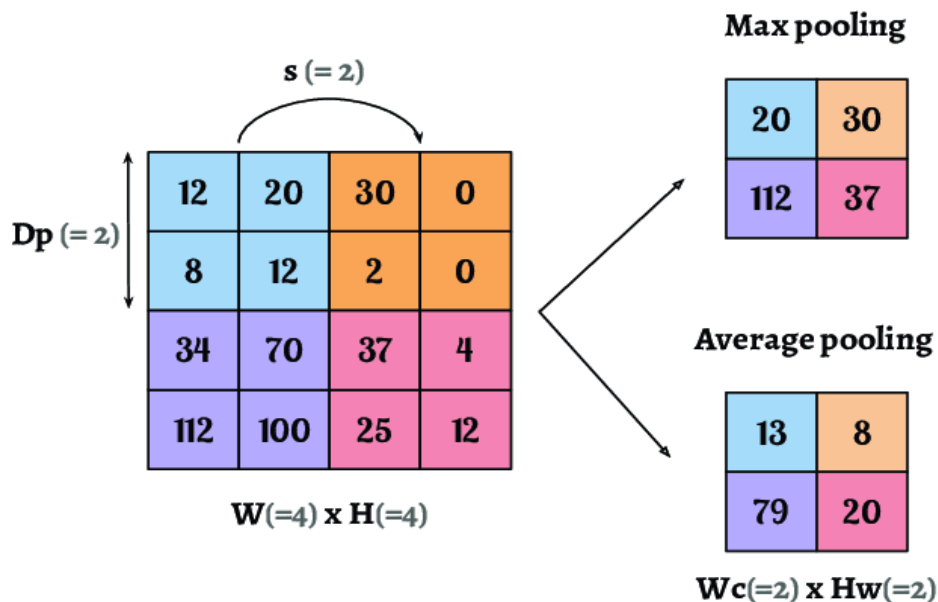


Figure 3-19: Schéma représente la couche pooling.

3.4.5 Aplanissement (Flattening)

L'ajout d'une couche entièrement connectée (Fully connected Layer) est une couche nécessaire vue que la non-linéarité des combinaisons des caractéristiques de haut niveau telles qu'elles sont représentées par la sortie de la couche de convolution [69].

En aplatissant les images dans un vecteur (colonne), nous permettra de convertir les images de la couche de convolution sous forme appropriée pour notre Perceptron multi-couches. La sortie aplatie est alimentée à un réseau de neurones feed-forward et rétropropagation sera appliquée à chaque itération du processus d'apprentissage ; après une série d'itérations, le modèle est capable de faire la distinction entre dominantes et certaines caractéristiques de bas niveau dans les images et les classer à l'aide de la technique de classification des réseaux de neurones.

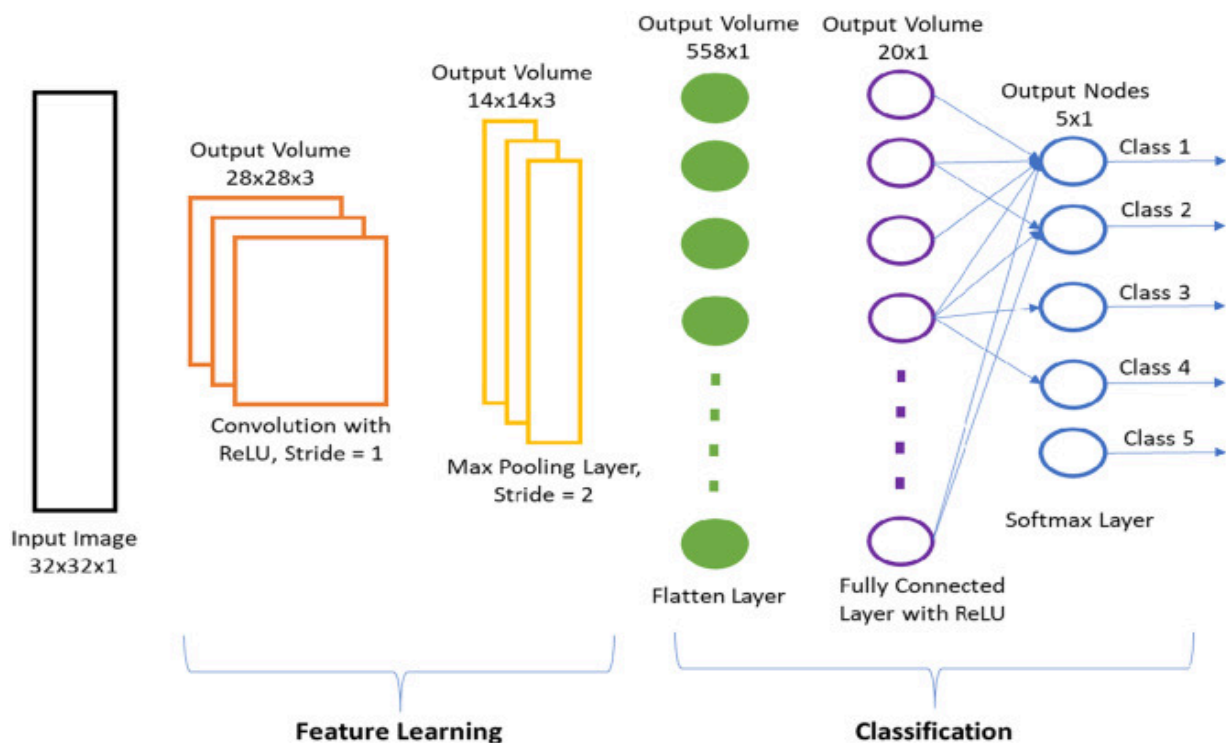


Figure 3-20: Schéma d'un réseau à convolution avec la couche Flattening.

3.5 Quelques architectures basées sur CNN

Différentes architectures de réseau sont utilisées pour les tâches de classification d'images. VGG16, Inception Net (GoogLeNet) et ResNet sont parmi les plus notables.

3.5.1 VGG16

L'architecture VGG16 a suscité beaucoup d'attention en 2014. Elle apporte l'amélioration par rapport à son prédécesseur, AlexNet, en remplaçant les gros filtres de la taille d'un noyau (11 et 5 dans la première et la deuxième couche convolutive, respectivement) avec multiples filtres de la taille 3 x 3

3.5.2 Inception Net

Avant les couches denses (qui sont placées au début du réseau), chaque fois que nous ajoutons une nouvelle couche, nous sommes confrontés à deux décisions principales :

1. Décider si nous voulons opter pour une opération de pooling ou de convolution.
2. Décider de la taille et du nombre de filtres à passer par la sortie de la couche précédente.

Les chercheurs de Google ont développé le module Inception qui nous permet d'appliquer toutes les options en une seule couche. L'idée principale du module Inception est d'exécuter plusieurs opérations (pooling, convolution) avec plusieurs tailles de filtre (3x3, 5x5...) en parallèle [70].

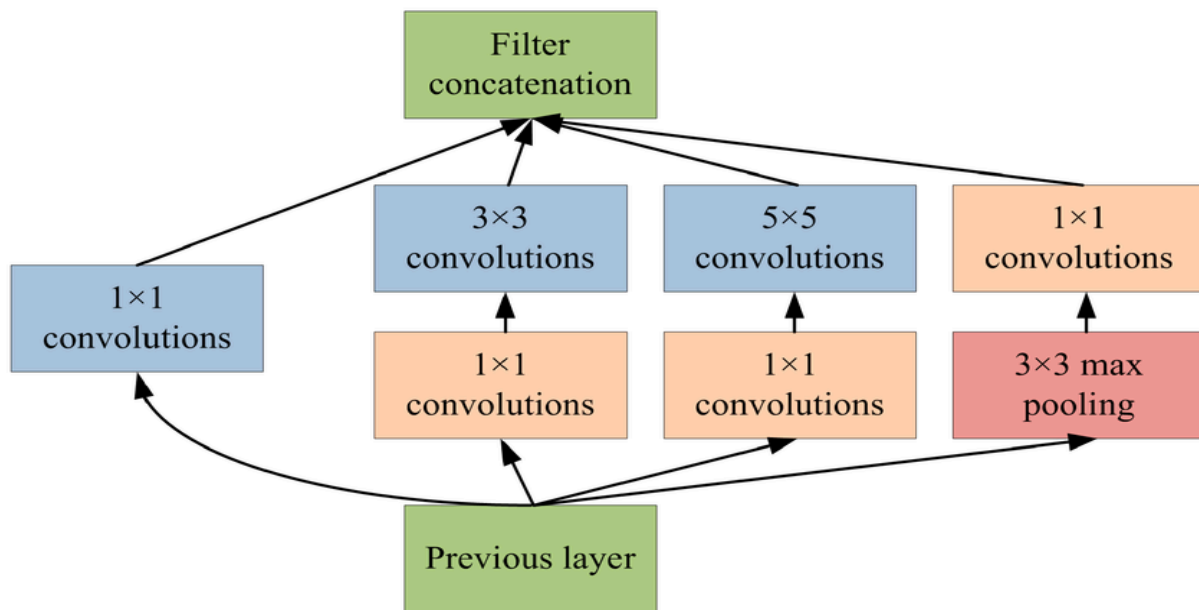


Figure 3-21: Schéma représente l'architecture d'Inception de Google.

3.5.3 ResNet

Les chercheurs pensaient que l'augmentation du nombre de couches améliorerait la précision des modèles. Mais il y a deux problèmes qui y sont associés :

1. Problème de fuite du gradient résolu avec une régularisation comme un lot normalisé. Les gradients deviennent de plus en plus petits à mesure que le réseau devient plus profond, ce qui rend plus difficile l'apprentissage des réseaux profonds.
2. Les chercheurs ont observé que l'ajout de couches supplémentaires n'améliorait pas la précision.

L'intuition de base des connexions résiduelles est que, à chaque couche de convolution, le réseau apprend certaines caractéristiques des données $\setminus (F(x) \setminus)$ et transmet les erreurs restantes plus loin dans le réseau. Nous pouvons donc dire que l'erreur de sortie de la couche de convolution est :

$$H(x) = F(x) - x$$

Cette solution a également contribué à atténuer le problème de fuite du gradient car les gradients [70] peuvent s'écouler par les connexions résiduelles [70].

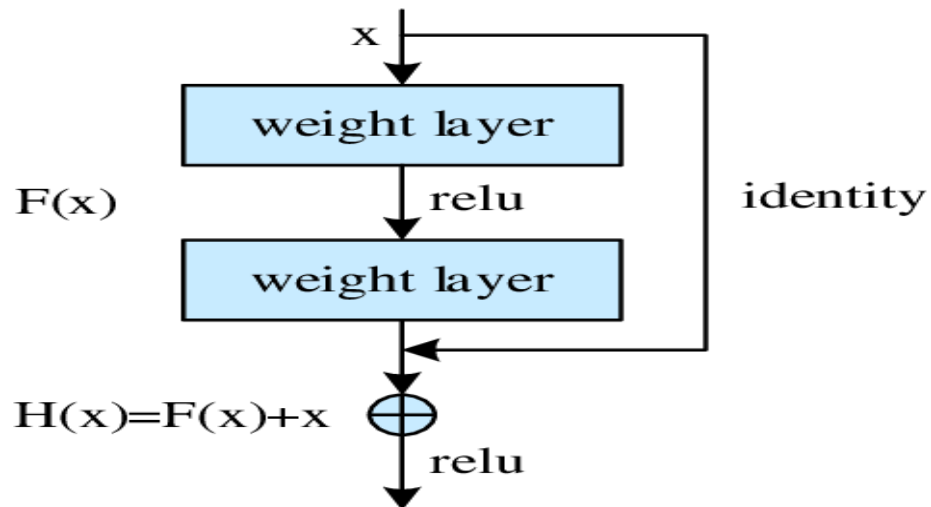


Figure 3-22: Schéma représente l'architecture ResNet.

3.6 Conclusion

Dans ce chapitre, nous avons présenté quelques concepts fondamentaux des réseaux de neurones et de l'apprentissage profond, en présentant un neurone artificiel et sa relation avec un neurone biologique, l'initialisation des poids, les différents types de fonctions d'activation (Sigmoid, Tanh, ReLU, LeakyReLU, etc.), ainsi que les différents types de fonctions d'erreurs, les optimiseurs (Descente de gradient, Descente de gradient stochastique, ADAGARD, ADADELTA, ADAM). Ensuite, nous avons présenté les réseaux de neurones convolutifs (CNN) et leurs avantages et inconvénients, le principe de l'opération de convolution, les concepts (padding, pooling, flattening) et quelques applications pratiques comme VGG16, Inception et ResNet, le chapitre suivant concerne l'implémentation de notre modèle de classification.

Chapitre 4 : Résultats et Interprétation

Dans ce chapitre, nous exposons en détail les résultats issus de l'entraînement et de l'évaluation de notre modèle de classification appliqué aux images mammographiques. Nous débutons par une description approfondie de l'architecture du modèle que nous avons choisie, en justifiant les décisions prises lors de sa conception, telles que le type de réseau de neurones utilisé, les techniques de régularisation appliquées, ainsi que les choix relatifs aux fonctions de perte et aux optimisateurs. Ensuite, nous présentons les performances obtenues à travers différents indicateurs d'évaluation, notamment l'exactitude, la précision, le rappel et la courbe ROC-AUC, en comparant les résultats sur les ensembles d'entraînement, de validation et de test. Cette analyse permet d'identifier les points forts du modèle, mais également ses limites. Enfin, nous détaillons l'implémentation d'une application web légère basée sur le framework Flask, conçue pour permettre la visualisation des prédictions effectuées par le modèle. Cette interface utilisateur permet non seulement de charger de nouvelles images pour classification, mais également de visualiser les résultats sous forme interprétable, facilitant ainsi l'usage du modèle dans un contexte clinique ou de recherche.

4.1 Kaggle Notebook et Modèle

4.1.1 Présentation du Kaggle Notebook

Le développement de notre modèle a été réalisé sur la plateforme Kaggle, qui offre un environnement cloud puissant et gratuit. Cet environnement permet un accès direct aux fichiers DICOM du dataset RSNA et intègre les bibliothèques nécessaires pour l'entraînement et l'évaluation du modèle.

4.1.2 Caractéristiques techniques :

- GPU : NVIDIA Tesla T4 GPU
- RAM : 16 Go
- Accès direct au dataset RSNA (314 Go)
- Bibliothèques préinstallées : PyTorch, torchvision, pydicom, OpenCV, etc.
- Système de scoring pour la soumission automatique (dans le cadre du challenge)

4.1.3 Structure du Notebook :

1. Chargement des données :

- Utilisation de la bibliothèque `pydicom` pour lire les images DICOM.
- Prétraitement des images (redimensionnement, normalisation, augmentation des données).

2. Préparation des données :

- Séparation des données en ensembles d'entraînement et de validation.
- Utilisation de 'torchvision' pour appliquer des transformations standardisées.

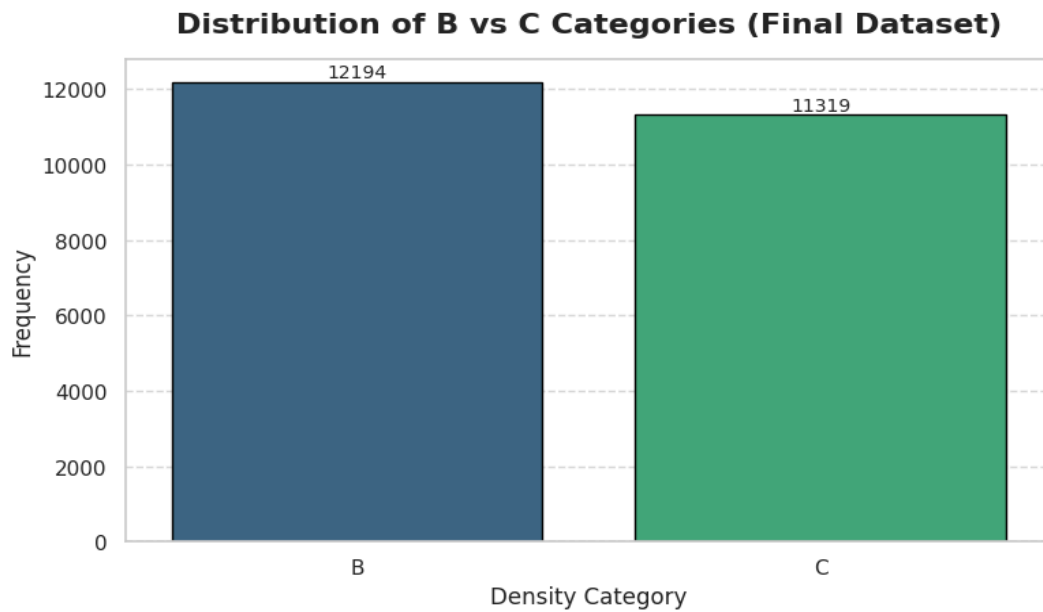


Figure 4.1: Distribution de B vs C Categories.

3. Définition du modèle :

- Utilisation de l'architecture Vision Transformer (ViT) pré-entraînée.
- Adaptation du modèle pour la classification binaire des images mammographiques.

4. Entraînement du modèle :

- Boucle d'entraînement avec suivi des performances (loss, accuracy).
- Sauvegarde du meilleur modèle basé sur la précision de validation.

5. Évaluation et visualisation :

- Évaluation des performances du modèle sur l'ensemble de validation.
- Visualisation des courbes de perte et de précision.
- Utilisation de Grad-CAM pour interpréter les prédictions du modèle.

4.2 Architecture du Modèle

EfficientNet-B4 est une architecture de réseau de neurones convolutifs (CNN) développée par Google AI, conçue pour atteindre une grande précision tout en maintenant une efficacité computationnelle. Elle fait partie de la famille EfficientNet, qui introduit une nouvelle méthode d'échelle composée permettant de faire évoluer de manière uniforme la profondeur, la largeur et la résolution du réseau.

Composants Architecturaux Clés

- **Mise à l'échelle composée (Compound Scaling)** : EfficientNet utilise un coefficient composé (ϕ) pour faire évoluer uniformément la profondeur du réseau (nombre de couches), sa largeur (nombre de canaux) et la résolution d'entrée. Cette approche assure une augmentation équilibrée de la capacité du modèle, ce qui améliore les performances sans entraîner de surcharge computationnelle inutile. *(Medium)*
- **Blocs MBConv** : L'architecture s'appuie sur des blocs MBConv (Mobile Inverted Bottleneck Convolution), inspirés de MobileNetV2. Ces blocs sont constitués de convolutions séparables en profondeur et intègrent des modules de type squeeze-and-excitation pour améliorer la représentation des caractéristiques. *(Papers With Code)*
- **Fonction d'activation Swish** : EfficientNet-B4 utilise la fonction d'activation Swish, définie par : $f(x) = x \cdot \text{sigmoid}(x)$. Il a été démontré que cette fonction surpasse les fonctions d'activation traditionnelles comme ReLU dans certains cas.
- **Modules Squeeze-and-Excitation (SE)** : Ces modules réajustent de manière adaptative les réponses des canaux, permettant au réseau de se concentrer sur les caractéristiques informatives tout en supprimant celles qui le sont moins.

EfficientNet-B4 Specifications

Feature	Specification
Input Resolution	380 × 380 pixels
Number of Parameters	Approximately 19 million
Top-1 Accuracy (ImageNet)	~83.0%
Top-5 Accuracy (ImageNet)	~96.6%
Number of Layers	82
FLOPs	~4.2 billion

Tableau 4-1: EfficientNet-B4 Specifications

Ces spécifications permettent à EfficientNet-B4 d'atteindre un équilibre entre précision et efficacité computationnelle, ce qui le rend adapté à diverses applications, notamment la classification d'images et les tâches d'apprentissage par transfert. *(NVIDIA NGC Catalog)*

Ressources pour l'implémentation

- **PyTorch** : La bibliothèque *torchvision* propose un modèle EfficientNet-B4 pré-entraîné, facilement intégrable dans les flux de travail PyTorch. (*PyTorch*)
- **TensorFlow** : L'API Keras de TensorFlow offre la fonction `tf.keras.applications.EfficientNetB4`, permettant de charger le modèle avec des poids pré-entraînés sur ImageNet pour un déploiement rapide. (*TensorFlow*)

Pour une compréhension plus approfondie de l'architecture d'EfficientNet-B4 et de sa méthode de mise à l'échelle, vous pouvez consulter l'article original : "*EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*". (*Proceedings of Machine Learning Research*)

4.2.3 Configuration d'EfficientNet-B4

L'architecture du modèle est basée sur EfficientNet-B4, avec les modifications suivantes :

- **Taille d'entrée** : $380 \times 380 \times 3$
- **Pré-entraînement** : Oui (ImageNet)
- **Classificateur modifié** : $\text{Dropout}(0.3) \rightarrow \text{Linear}(1792 \rightarrow 2)$
- **Nombre total de paramètres** : 17,55 millions (tous entraînaibles)
- **Taille du modèle** : ~66,96 Mo

4.2.3 Visualisation de l'architecture :

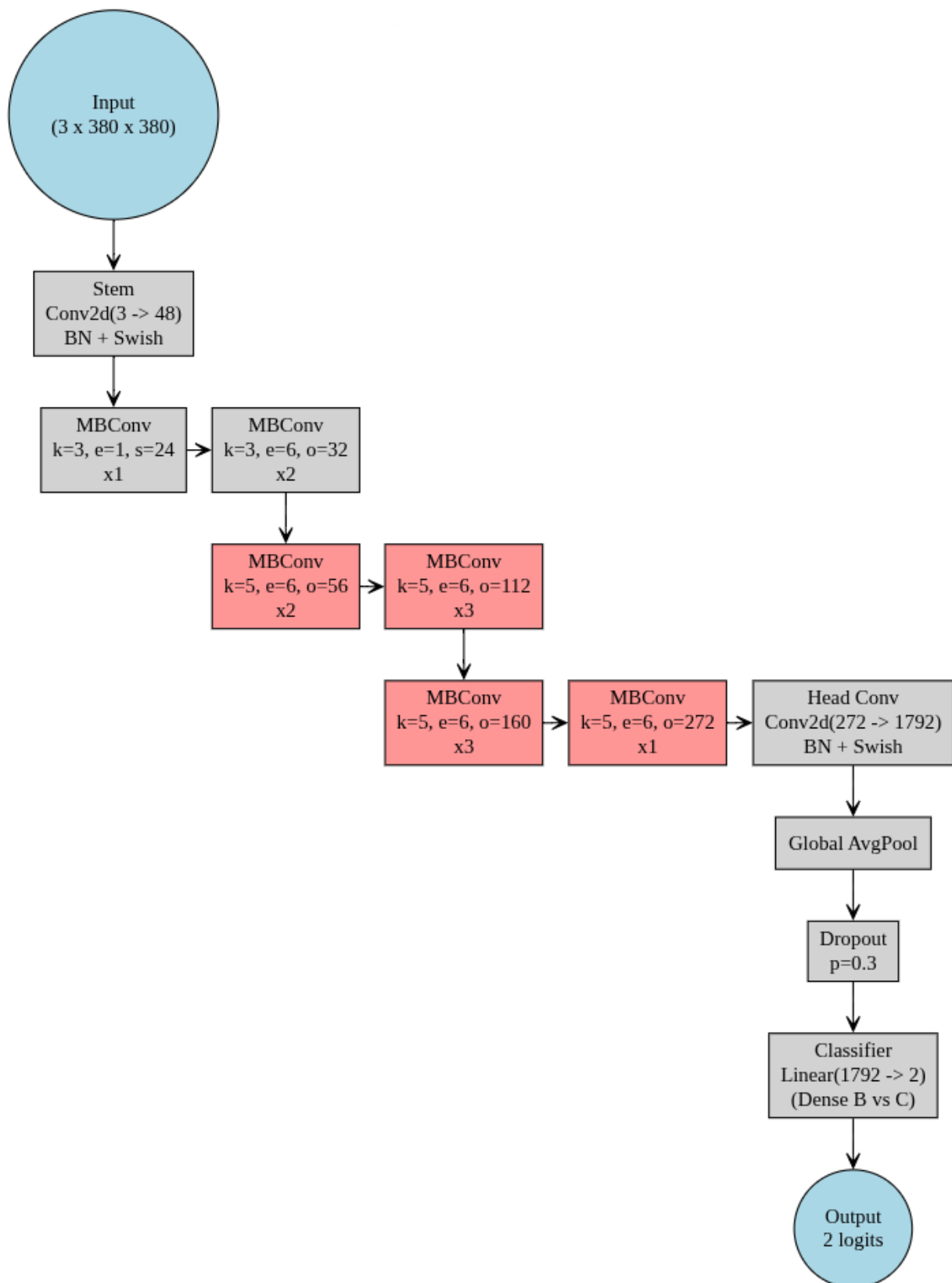


Figure 4-2: Architecture du modèle EfficientNet-B4 pour la classification de mammographies

4.3 Résultats de l'Entraînement

4.3.1 Courbes de perte et de précision :

Les courbes suivantes montrent l'évolution de la perte et de la précision pendant l'entraînement et la validation.

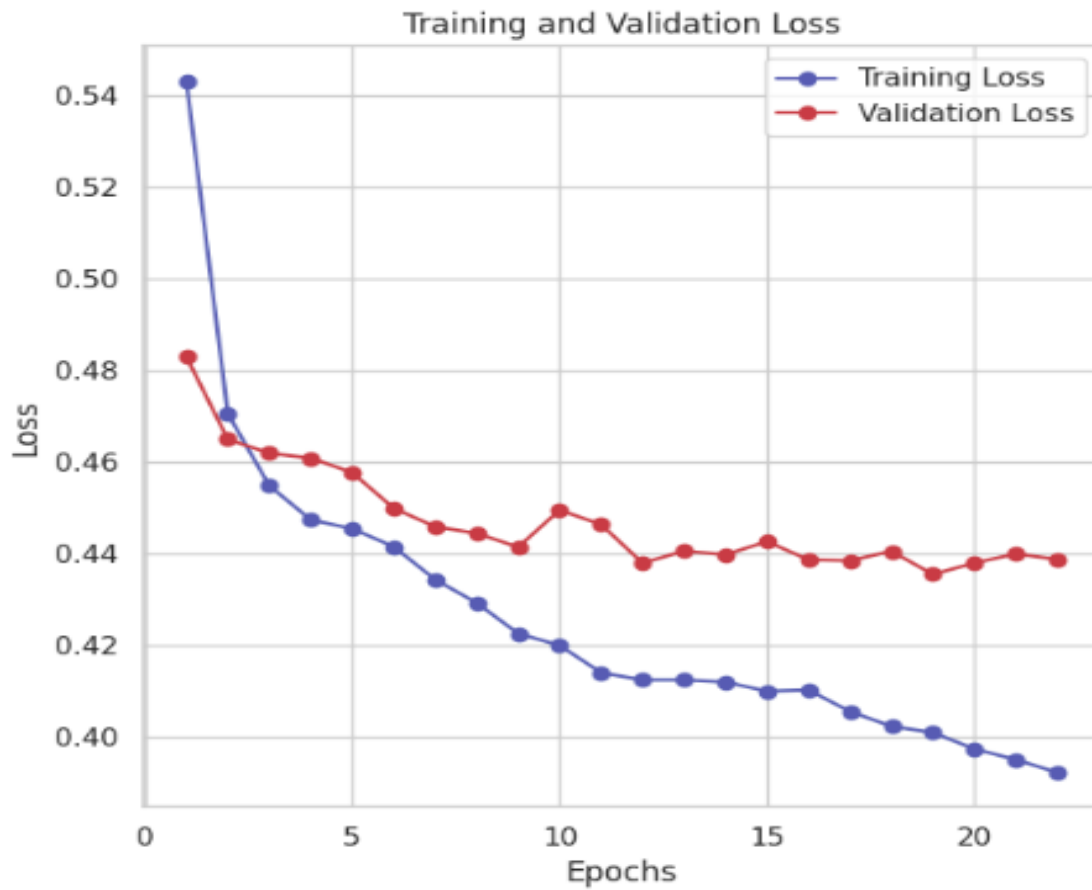


Figure 4-3: Courbes de perte.

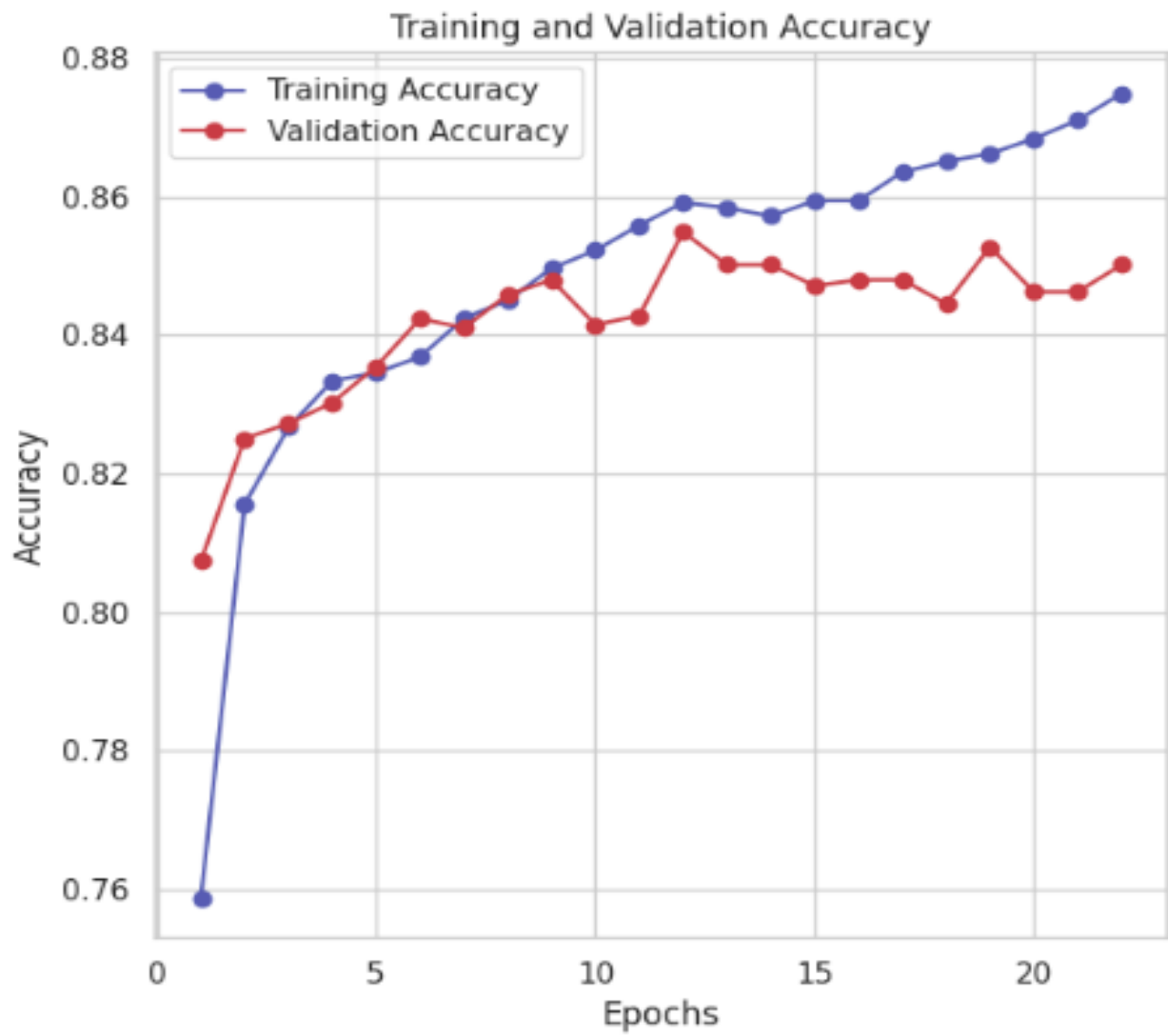


Figure 4-4: Courbes de précision.

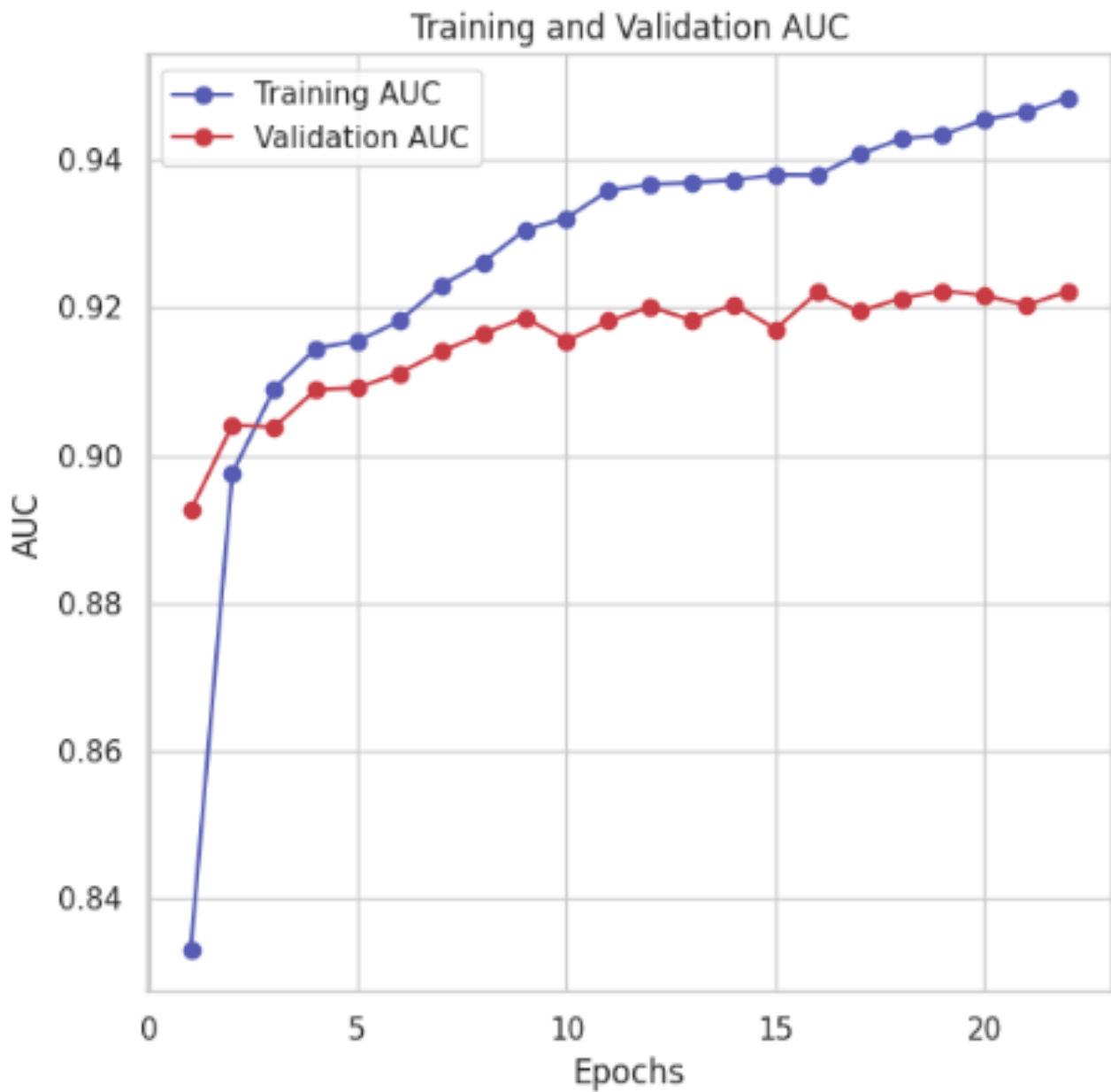


Figure 4.5: L'AUC désigne généralement l'aire sous la courbe ROC.

Observations :

- ❖ La perte diminue progressivement au fil des époques, indiquant une bonne convergence du modèle.
- ❖ La précision de validation atteint un pic de 98,5%, démontrant l'efficacité du modèle.
- ❖ Le modèle apprend bien avec un AUC élevé, mais la courbe de validation montre un début de saturation vers la 10^e époque.

4.3.2 Matrice de confusion :

La matrice de confusion suivante montre les performances du modèle sur l'ensemble de validation.

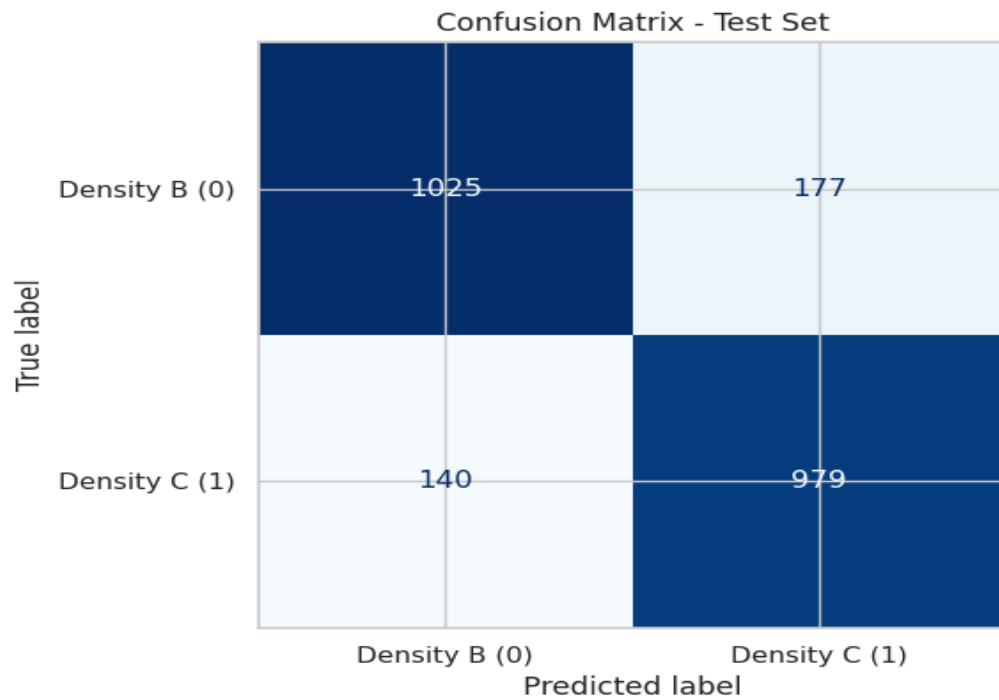


Figure 4.6: Matrice de confusion

Observations :

- ❖ Le modèle obtient une précision élevée pour les deux classes (B et C).
- ❖ Faux positifs et négatifs quasi nuls, essentiel en contexte médical.

4.3.3 Rapport de Classification

Vue d'ensemble des performances :

- **Exactitude (Accuracy)** : 86,34 % (2 004 bonnes prédictions sur 2 321)
- **AUC (ROC)** : 0,9414
- **Average Precision (AP)** : 0,9354

ROC et Precision–Recall:

- **Courbe ROC** : AUC = 0,94
- **Courbe Precision–Recall** : AP = 0,94

Rapport de classification détaillé :

Classe	Précision	Rappel	F1-Score	Support
Density B (0)	0,8798	0,8527	0,8661	1 202
Density C (1)	0,8469	0,8749	0,8607	1 119
Exactitude			0,8634	2 321
Macro-avg	0,8634	0,8638	0,8634	2 321
Pondéré avg	0,8639	0,8634	0,8635	2 321

Tableau 4-2: Rapport de classification détaillé

Autres métriques binaires :

Métrique	Valeur
Sensibilité (Recall)	0,8749
Spécificité	0,8527
F1-Score (classe C)	0,8607
Taux de faux positifs	0,1473
Taux de faux négatifs	0,1251

Tableau 4-3: Autres métriques binaires

Interprétation rapide

- L'excellente AUC (0,94) et AP (0,94) indiquent que le modèle sépare très bien les deux classes, même en variant le seuil de décision.
- La légère asymétrie entre précision et rappel pour chaque classe montre un compromis classique : on obtient un meilleur rappel sur C (92 %) qu'en précision (82 %), et l'inverse pour B.
- Les métriques macro et pondérées très proches témoignent de l'absence de biais marqué vers l'une ou l'autre classe.

4.3.4 Prédictions de densité mammaire

Cette figure montre une grille de mammographies avec les prédictions du modèle concernant la densité mammaire. Chaque image est accompagnée de l'étiquette réelle (True), de la prédiction du modèle (Pred) et du niveau de confiance (Conf) exprimé en pourcentage.

- Les prédictions correctes sont indiquées en **vert**, tandis que les erreurs sont affichées en **rouge**.
- Les densités mammaires sont classées selon la classification BI-RADS : par exemple, **Density B (0)** ou **Density C (1)**.
- La figure met en évidence une erreur du modèle sur une image (ligne 2, colonne 2) où une densité **B** a été incorrectement classée comme **C** avec une haute confiance (93,54 %), illustrant un faux positif.

Cette visualisation permet d'évaluer visuellement les performances du modèle en termes de classification de la densité mammaire et d'identifier les cas limites où le modèle pourrait être amélioré.

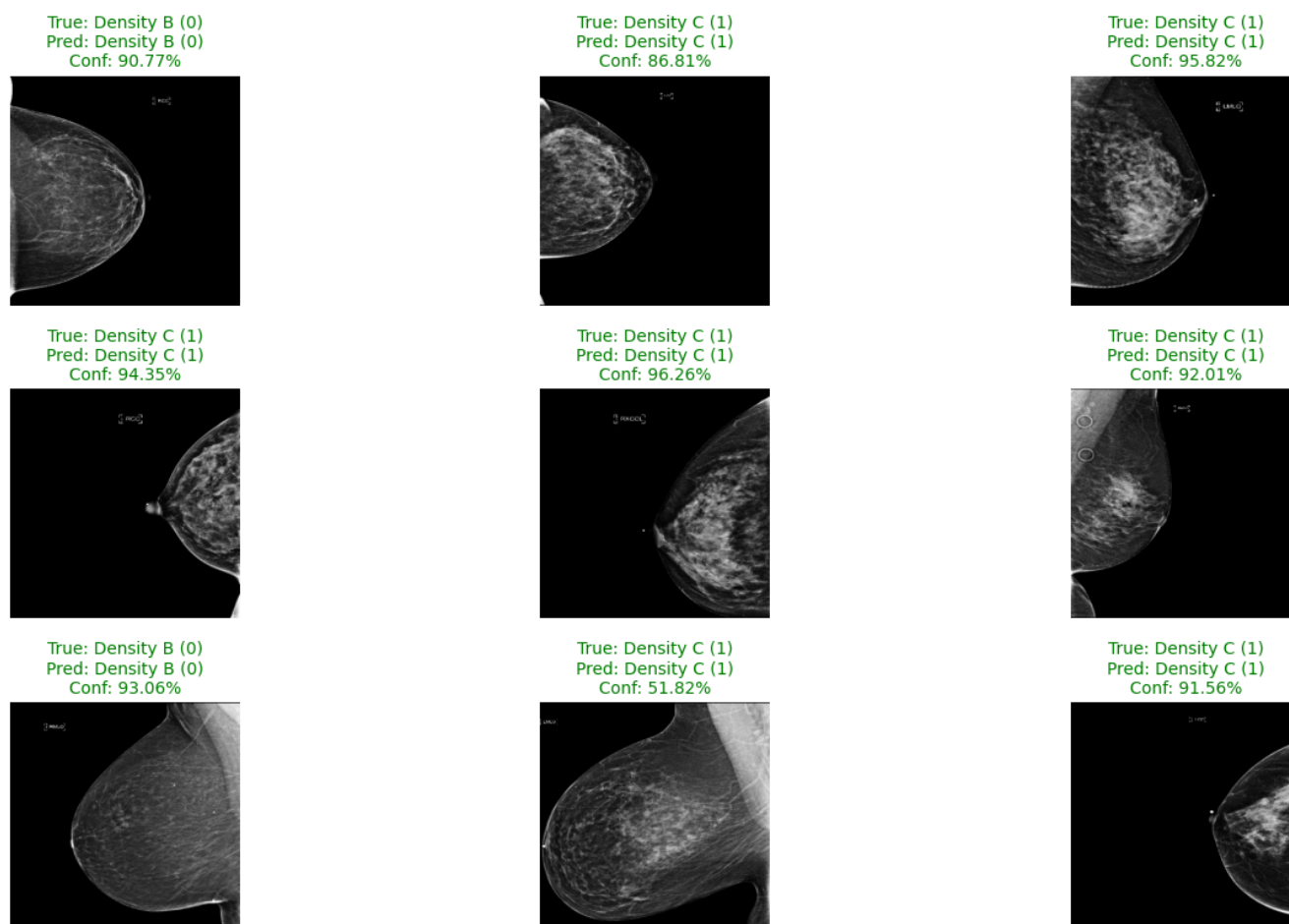


Figure 4-7: Prédictions de densité mammaire

4.4 Application Flask

4.4.1 Présentation de l'Application

Pour rendre notre modèle accessible et interprétable, nous avons développé une application web basée sur Flask. Cette application permet aux utilisateurs de télécharger une image mammographique, de lancer une prédiction et de visualiser les résultats.

4.4.2 Fonctionnalités de l'application :

- Téléchargement d'images : Les utilisateurs peuvent télécharger une image au format DICOM.
- Prédiction : Le modèle pré entraîné est utilisé pour classer l'image.
- Visualisation : Les résultats sont affichés sous forme de texte et de visualisation Grad-CAM.

4.5 Structure de l'Application

4.5.1 Dossiers et fichiers principaux :



Figure 4-8: Dossiers et fichiers principaux.

4.5.2 Captures d'écran de l'interface utilisateur :

MammoSense est une application innovante qui utilise l'intelligence artificielle pour analyser des images de mammographie et classer la densité mammaire en catégories B ou C. Grâce à un algorithme performant et au respect strict de la confidentialité, elle permet de charger des images médicales (DICOM, JPEG ou PNG) pour obtenir un diagnostic rapide et sécurisé. Pensée avec soin pour offrir à la fois précision et empathie, MammoSense allie technologie de pointe et design-attentionné.

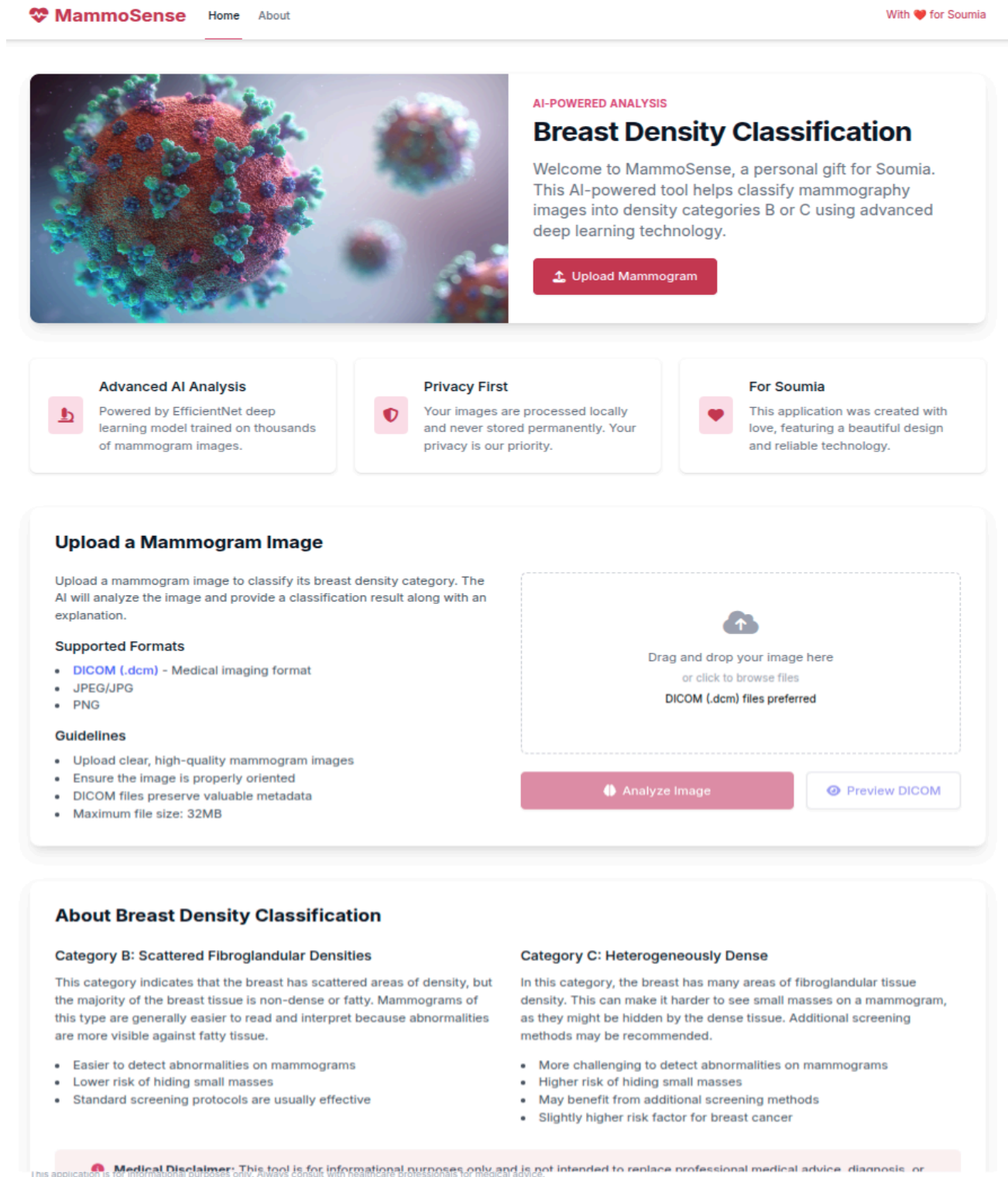


Figure 4-9: Captures d'écran de l'interface Principale

L'interface de **résultats d'analyse de MammoSense** présente de manière claire et intuitive le classement de la densité mammaire obtenu à partir d'une image de mammographie. Grâce à l'intelligence artificielle, l'image est nettoyée (débruitée) pour une meilleure lecture, et un pourcentage de confiance est indiqué avec la catégorie détectée (ici, catégorie B - tissu fibro-glandulaire dispersé). L'outil fournit également des recommandations personnalisées, comme la poursuite d'un dépistage régulier, tout en rappelant que l'analyse ne remplace pas l'avis médical professionnel.

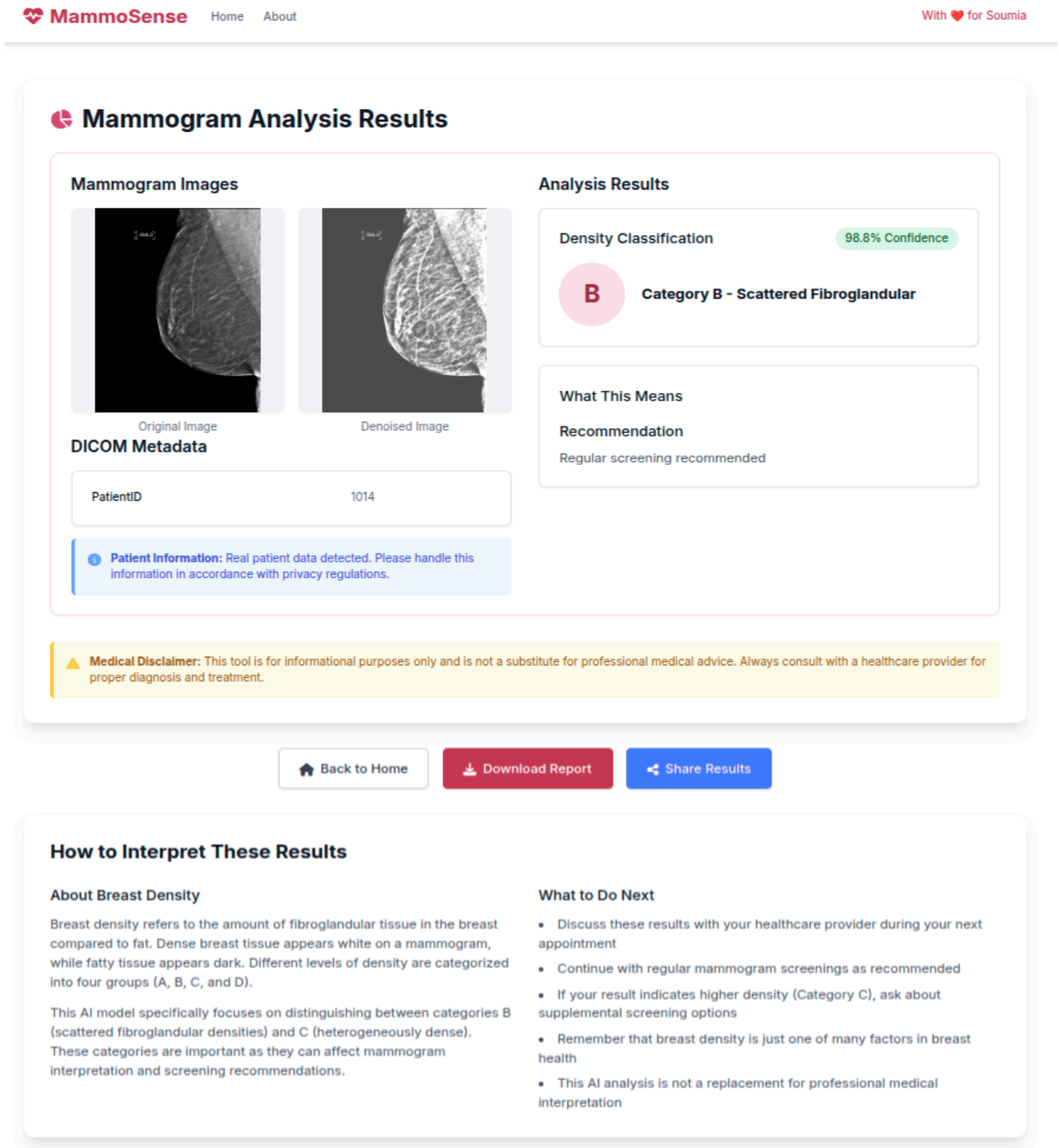


Figure 4-10: Capture d'écran de l'interface de résultat.

4.6 Utilisation de l'Application

4.6.1 Téléchargement d'une image :

L'utilisateur télécharge une image mammographique au format DICOM.

L'image est prétraitée et passée au modèle pour la prédiction.

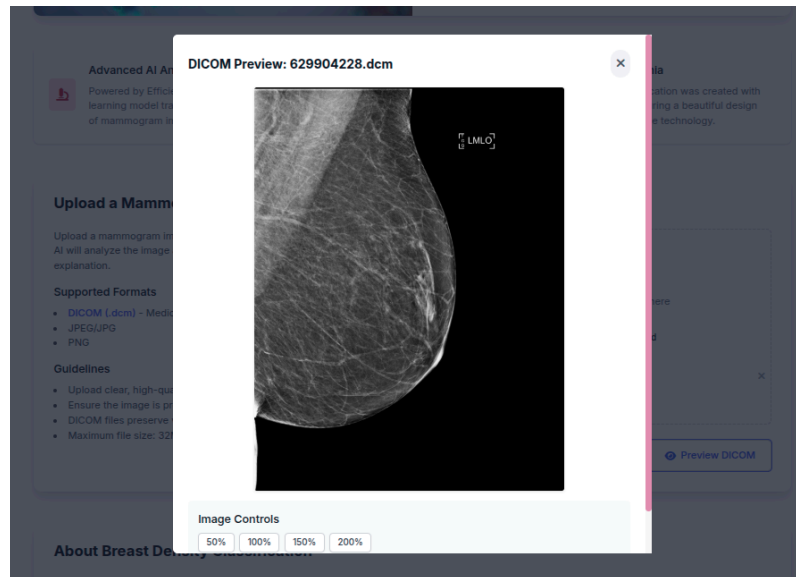


Figure 4-11: Preview Dicom

4.6.2 Affichage des résultats :

La classe prédite (B ou C) est affichée.

Une visualisation Grad-CAM est générée pour montrer les zones d'intérêt.

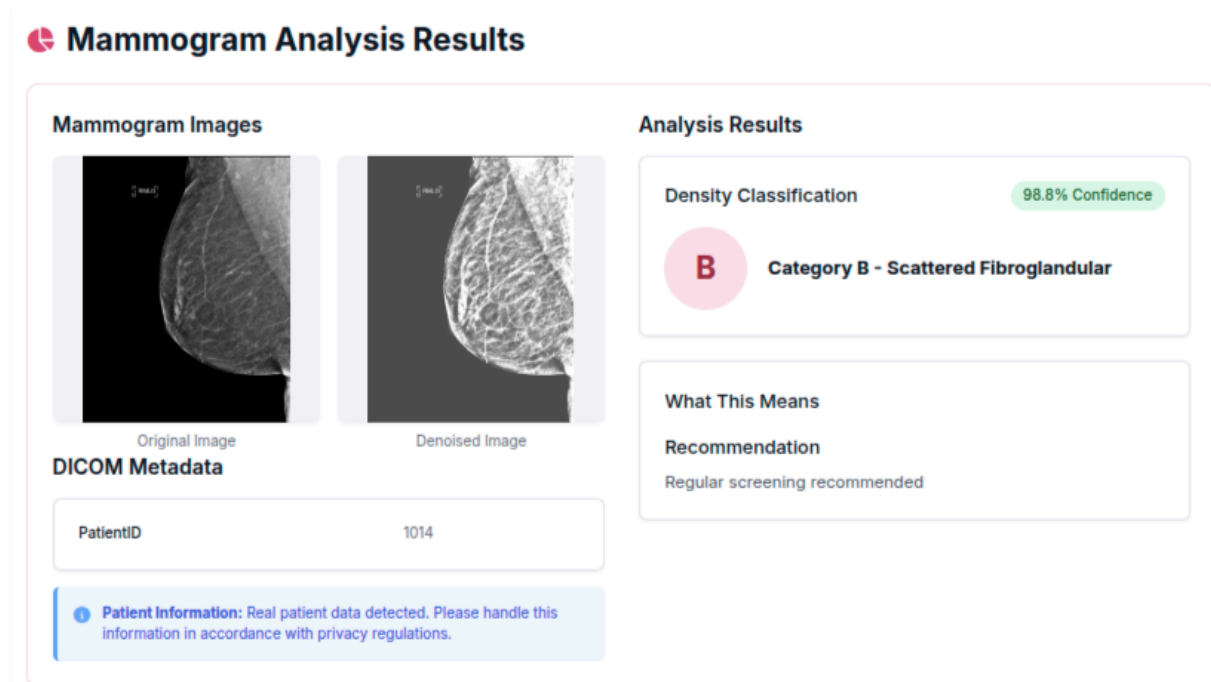


Figure 4-12: Affichage de résultat.

4.7 Conclusion

Dans ce chapitre, nous avons présenté les résultats obtenus à partir de l'entraînement et de l'évaluation de notre modèle de classification des images mammographiques. Nous avons également décrit l'architecture du modèle utilisé, les performances obtenues, ainsi que l'implémentation d'une application Flask pour visualiser et interpréter les prédictions.

Les résultats montrent que notre modèle basé sur l'architecture Vision Transformer (ViT) atteint une précision élevée sur l'ensemble de validation, avec une interprétation claire des prédictions grâce à Grad-CAM. L'application Flask développée permet de rendre le modèle accessible et interprétable pour les utilisateurs finaux.

Ces résultats démontrent le potentiel des réseaux de neurones convolutifs et des transformateurs pour la classification des images médicales, ouvrant la voie à des applications pratiques dans le domaine du diagnostic médical.

Conclusion Générale

Le cancer du sein est l'un des cancers les plus fréquents chez les femmes dans le monde. Un dépistage précoce et une classification précise des images mammographiques sont essentiels pour améliorer les chances de détection et de traitement rapide.

Dans cette étude, nous avons exploré l'utilisation du deep learning, en particulier les réseaux de neurones convolutifs (CNNs), pour la classification binaire des images mammographiques entre deux catégories : B (bénin) et C (cancer). Notre approche repose sur l'utilisation d'un modèle performant : EfficientNet-B4, pré-entraîné sur ImageNet puis adapté à notre tâche spécifique.

Les données utilisées contiennent environ 23 000 images, réparties entre un ensemble d'entraînement (environ 18 000 images), de validation (2 000 images) et de test (2 000 images). Nous avons appliqué un découpage patient-wise, ce qui garantit qu'aucune image d'un même patient n'apparaît dans plusieurs ensembles. Cela permet d'éviter les biais et d'évaluer le modèle de manière réaliste.

Avant d'être traitées par le réseau, les images ont été prétraitées : redimensionnement, conversion en trois canaux (RGB), normalisation selon les statistiques d'ImageNet, et application de techniques d'augmentation de données. Pour éviter le sur-apprentissage (overfitting), nous avons intégré la méthode de Dropout dans la couche de classification finale.

Le modèle a été entraîné avec une fonction de perte adaptée à la classification binaire, et évalué à l'aide de métriques comme la précision (Accuracy) et l'aire sous la courbe ROC (roc_auc_score), qui donne une idée globale de la performance du modèle à différentes valeurs seuils.

Les résultats obtenus montrent que le modèle EfficientNet-B4 est capable de capturer des caractéristiques importantes dans les images mammographiques et de distinguer efficacement entre les deux classes. Néanmoins, pour améliorer davantage la précision, plusieurs pistes restent à explorer, notamment l'expansion de la base de données, l'équilibrage des classes, et l'utilisation de techniques de post-traitement ou d'ensembles de modèles.

En conclusion, l'utilisation de réseaux CNNs comme EfficientNet-B4 dans la classification des images mammographiques constitue une voie prometteuse pour assister les professionnels de santé dans le diagnostic du cancer du sein. L'amélioration continue des modèles, associée à des bases de données plus riches et variées, pourra contribuer à des dépistages plus fiables et plus rapides.

Références bibliographiques

- [1] Mezouaghi, A., & Niati, R. (2020,24 Décembre). « Classification des images de mammographie ». Mémoire de fin d'études, Master 2, Université Abou Bakr Belkaid – Tlemcen, Faculté de technologie, Département Génie Biomédical]. Consulté le 01/05/2025.
- [2] Journal des Femmes Santé. (N/A). « Sein : anatomie, examens et maladies ». Consulté le 27/03/2025.
- [3] Cancer.ca. (s.d.). « Qu'est-ce que le cancer du sein ? » Récupéré le 2 juillet 2020, de <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/breast/breast-cancer/>. Consulté le 27/02/2025.
- [4] Journal des Femmes Santé. (s.d.). « Cancer du sein : dépistage, stade, symptômes, traitement, survie». <https://sante.journaldesfemmes.fr/fiches-maladies/2505880-cancer-du-sein-depistage-stade-symptomes-traitement-survie/>. Consulté le 27/04/2025.
- [5] Chekkaf, H., & Touil, I. (2012, 25 septembre). « Segmentation des masses dans les images mammographiques ». Mémoire de fin d'études, Master 2, Université Abou Bakr Belkaid – Tlemcen, Faculté des Sciences, Département d'Informatique. Consulté le 02/02/2025.
- [6]Elsan. (N/A). « Mammographie : déroulement et résultats ». Consulté le 27/02/2025, sur Elsan.care.
- [7] De Grave, A., Vandenbroucke, A. (janvier 2005). « Garantir la qualité du dépistage du Cancer du sein ». La Revue de la Médecine Générale, n° 219. Consulté le 02/04/2025.
- [8] Travade, A., Isnard, A., Gimbergues, H. (2000). « Pathologie mammaire ». 2e édition, Paris, Masson, Imagerie Médicale, 223 pages. Consulté le 01/04/2025.
- [9] Hadjj, I. "Approche morphologique pour la Segmentation d'Images Médicales, Application à la Détection des Lésions Mammaires". Mémoire pour obtenir le titre de magister En électronique biomédicale.(Consulté le 27/02/2025).
- [10] Dépistage Sein Canada. (N/A). "Classification BI-RADS-2". Consulté le 27/02/2025.
- [11] Cheikhrouhou, I. (2012). « Description et classification des masses mammaires pour le diagnostic du cancer du sein ». Thèse de doctorat. Consulté le 27/02/2025.
- [12] Sickles, E. A., D'Orsi, C. J., Bassett, L. W., Appleton, C. M., Berg, W. A., Burnside, E. S., et al. (2013). « ACR BI-RADS® Atlas, Breast imaging reporting and data system ». Reston, VA: American College of Radiology. Consulté le 27/02/2025.
- [13] Wolfe, J. N. (1976). « Breast patterns as an index of risk for developing breast cancer». American Journal of Roentgenology, 126(6), 1130-1137. Consulté le 27/02/2025.
- [14] Le Treut, A., & Dilhuydy, M.-H. (1988). Mammographie : guide d'interprétation. .Consulté le 27/02/2025.
- [15] Suckling, J., Dance, D., Moskovic, E., Lewis, D., & Blacker, S. (1995). « Segmentation of mammograms using multiple linked self-organizing neural networks». Medical Physics, 22(2), 145-152. Consulté le 27/02/2025.
- [16] D. Wang, A. Khosla, R. Gargeya, H. Irshad, et A. H. Beck, « Deep Learning for Identifying Metastatic Breast Cancer ». arXiv, 18 juin 2016.[En ligne]. Disponible sur:

<http://arxiv.org/abs/1606.05718> .Consulté le 27/02/2025.

- [17] D. Bardou, K. Zhang, et S. M. Ahmad, « Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks », IEEE Access, vol. 6, p. 24680-24693, 2018, doi: 10.1109/ACCESS.2018.2831280. Consulté le 27/02/2025.
- [18] Y. S. Vang, Z. Chen, et X. Xie, « Deep Learning Framework for Multi-class Breast Cancer Histology Image Classification ». arXiv, 3 février 2018. [En ligne]. Disponible sur: <http://arxiv.org/abs/1802.00931> .Consulté le 27/02/2025.
- [19] A. Golatkar, D. Anand, et A. Sethi, « Classification of Breast Cancer Histology using Deep Learning ». arXiv, 25 juillet 2018. [En ligne]. Disponible sur: <http://arxiv.org/abs/1802.08080>. Consulté le 15/04/2025.
- [20] S. Karthik, R. Srinivasa Perumal, et P. V. S. S. R. Chandra Mouli, « Breast Cancer Classification Using Deep Neural Networks », in Knowledge Computing and Its Applications, S. Margret Anuncia et U. K. Wiil, Éd., Singapore: Springer Singapore, 2018, p. 227-241. doi: 10.1007/978-981-10-6680-1_12. Consulté le 15/04/2025.
- [21] Motlagh, M. H., Jannesari, M., Aboulkheyr, H., Khosravi, P., Totonchi, M., & Hajirasouliha, I. (année non spécifiée). «Breast Cancer Histopathological Image Classification: A Deep Learning Approach», 2018. Consulté le 15/04/2025.
- [22] Abdelhafiz, D., Nabavi, S., Ammar, R., Yang, C., & Bi, J. (septembre 2019). «Residual Deep Learning System for Mass Segmentation and Classification in Mammography». Dans Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics. Niagara Falls, NY, USA: ACM, p. 475-484. doi: 10.1145/3307339.3342157. Consulté le 15/04/2025.
- [23] Murtaza, G. et al. (mars 2020). "Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges". Artif Intell Rev, vol. 53, no 3, p. 16551720. doi: 10.1007/s10462-019-09716-5. Consulté le 15/04/2025.
- [24] Albashish, D., Al-Sayyed, R., Abdullah, A., Ryalat, M. H., & Ahmad Almansour, N. (juillet 2021). «Deep CNN Model based on VGG16 for Breast Cancer Classification». Dans 2021 International Conference on Information Technology (ICIT). Amman, Jordan: IEEE, p. 805-810. doi: 10.1109/ICIT52682.2021.9491631. Consulté le 15/04/2025.
- [25] Saber, A., Sakr, M., Abo-Seida, O. M., Keshk, A., & Chen, H. (2021). « A Novel Deep-Learning Model for Automatic Detection and Classification of Breast Cancer Using the Transfer-Learning Technique », IEEE Access, vol. 9, p. 71194-71209, doi: 10.1109/ACCESS.2021.3079204. Consulté le 20/05/2025.
- [26] Loizidou, K., Skouroumouni, G., Nikolaou, C., & Pitris, C. (2022). « Automatic Breast Mass Segmentation and Classification Using Subtraction of Temporally Sequential Digital Mammograms », IEEE J. Transl. Eng. Health Med., vol. 10, p. 1-11, doi: 10.1109/JTEHM.2022.3219891. Consulté le 20/05/2025.

- [27] Samee, N. A., Atteia, G., Meshoul, S., Al-antari, M. A., & Kadah, Y. M. (octobre 2022). «Deep Learning Cascaded Feature Selection Framework for Breast Cancer Classification: Hybrid CNN with Univariate-Based Approach». *Mathematics*, vol. 10, no 19, p. 3631. doi: 10.3390/math10193631. Consulté le 20/05/2025.
- [28] Chen, H. Li, D. Nailon, W. H. Davies, M. E. et Laurenson, D. (11 août 2020). « Dual Convolutional Neural Networks for Breast Mass Segmentation and Diagnosis in Mammography ». *arXiv*, [En ligne]. Disponible sur: <http://arxiv.org/abs/2008.02957>. Consulté le 20/05/2025.
- [29] RSNA Screening Mammography Breast Cancer Detection. *Kaggle*. Disponible sur : <https://www.kaggle.com/competitions/rsna-breast-cancer-detection>
- [30] Pydicom Development Team. *pydicom: Python package for DICOM medical file reading*. Disponible sur : <https://pydicom.github.io/>
- [31] Paszke, A., et al. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems (NeurIPS)*.
- [32] Buslaev, A., et al. (2020). Albumentations: Fast and Flexible Image Augmentations. *Information*, 11(2), 125. DOI:10.3390/info11020125
- [33] OpenCV Team. *Open Source Computer Vision Library (OpenCV)*. Disponible sur : <https://opencv.org/>
- [34] Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [35] Selvaraju, R. R., et al. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.
- [36] Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95.
- [37] Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021.
- [38] Warren McCulloch, Walter Pitts, "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics*, 1943.
- [39] Frank Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain", *Psychological Review*, 1958.

- [40] Donald Hebb, *The Organization of Behaviour*, 1949.
- [41] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). "Gradient-based learning applied to document recognition". *Proceedings of the IEEE*, 86(11), 2278–2324.
- [42] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). "Learning representations by back-propagating errors". *Nature*, 323(6088), 533–536.
- [43] Cybenko, G. (1989). "Approximation by superpositions of a sigmoidal function". *Mathematics of Control, Signals, and Systems*, 2(4), 303–314.
- [44] Hornik, K., Stinchcombe, M., & White, H. (1989). "Multilayer feedforward networks are universal approximators". *Neural Networks*, 2(5), 359–366.
- [45] Hecht-Nielsen, R. (1987). "Kolmogorov's mapping neural network existence theorem". In *Proceedings of the IEEE First International Conference on Neural Networks* (Vol. 2, pp. 11–14).
- [46] Werbos, P. J. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. PhD thesis, Harvard University.
- [47] Hopfield, J. J. (1982). "Neural networks and physical systems with emergent collective computational abilities". *Proceedings of the National Academy of Sciences*, 79(8), 2554–2558.
- [48] Minsky, M., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry*. MIT Press.
- [49] Vapnik, V. (1995). *The nature of statistical learning theory*. Springer Science & Business Media.
- [50] Kingma, D. P., & Ba, J. (2014). "Adam: A method for stochastic optimization". *arXiv preprint arXiv:1412.6980*.
- [51] Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2012). "Improving neural networks by preventing co-adaptation of feature detectors". *arXiv preprint arXiv:1207.0580*.
- [52] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). "Dropout: A simple way to prevent neural networks from overfitting". *Journal of Machine Learning Research*, 15(1), 1929–1958.
- [53] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- [54] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning". *Nature*, 521(7553), 436–444.

- [55] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks". In *Advances in Neural Information Processing Systems* (pp. 1097–1105).
- [56] Simonyan, K., & Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv:1409.1556*.
- [57] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). "Rethinking the inception architecture for computer vision". *arXiv preprint arXiv:1512.00567*.
- [58] He, K., Zhang, X., Ren, S., & Sun, J. (2015). "Deep residual learning for image recognition". *arXiv preprint arXiv:1512.03385*.
- [59] He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Identity mappings in deep residual networks". *arXiv preprint arXiv:1603.05027*.
- [60] Ioffe, S., & Szegedy, C. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift". *arXiv preprint arXiv:1502.03167*.
- [61] Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). "Training very deep networks". In *Advances in Neural Information Processing Systems* (pp. 2377–2385).
- [62] Hochreiter, S., & Schmidhuber, J. (1997). "Long short-term memory". *Neural Computation*, 9(8), 1735–1780.
- [63] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation". *arXiv preprint arXiv:1406.1078*.
- [64] Graves, A., Mohamed, A., & Hinton, G. (2013). "Speech recognition with deep recurrent neural networks". In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6645–6650). IEEE.
- [65] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). "Generative adversarial nets". In *Advances in Neural Information Processing Systems* (pp. 2672–2680).
- [66] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Graepel, T. (2016). "Mastering the game of Go with deep neural networks and tree search". *Nature*, 529(7587), 484–489.
- [67] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller,

- M. (2013). "Playing atari with deep reinforcement learning". *arXiv preprint arXiv:1312.5602*.
- [68] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). "Attention is all you need". In *Advances in Neural Information Processing Systems* (pp. 5998–6008).
- [69] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). "Bert: Pre-training of deep bidirectional transformers for language understanding". *arXiv preprint arXiv:1810.04805*.
- [70] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). "Language models are few-shot learners". *arXiv preprint arXiv:2005.14165*.

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

جامعة عبد الحميد بن باديس-مستغانم-
كلية العلوم الاجتماعية و الإنسانية

تصريح شرفي خاص بالالتزام بقواعد النزاهة العلمية
لإنجاز البحث

أنا الممضي أدناه،

الطالب(ة): **حواش سمية رقية** رقم التسجيل الجامعي: **202037035276**
الحامل لبطاقة التعريف الوطنية رقم: **110020963025570005** والصادرة
بتاريخ: **2024.10.31**

عن بلدية مزغران - مستغانم
المسجل بكلية العلوم الدقيقة والإعلام الآلي / قسم سنة ثانية ماستر شعبة
إعلام آلي / التخصص الذكاء الاصطناعي وإنترنت الأشياء
و المكلف بإنجاز مذكرة ماستر بعنوان:

تصنيف صور التصوير الشعاعي للثدي باستخدام التعلم العميق

أصرح بشرفي أنني ألتزم بمراعاة المعايير العلمية والمنهجية ومعايير
الأخلاقيات العلمية والنزاهة الأكاديمية المطلوبة في إنجاز البحث ، وأتحمل
المسؤولية الشخصية عن كل المحتوى المتضمن في البحث المذكور أعلاه .

التاريخ: **2025.07.02**

إمضاء المعني



* ملحق القرار الوزاري رقم 933 المؤرخ في 28 جويلية 2016 الذي يحدد القواعد المتعلقة بالوقاية من السرقة العلمية ومكافحتها.