

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة ابو بكر بالقائد - تلمسان

Université Abou Bakr Belkaïd –Tlemcen –Faculté de TECHNOLOGIE



MEMOIRE

Présenté pour l'obtention du **diplôme** de **MASTER**

En : Génie Biomédical

Spécialité : Imagerie Médicale

Par : BENTOUBA Ferial & BENCHAO Chaima

Sujet

**Segmentation et classification des tumeurs du cancer du sein
sur des images échographiques à l'aide de Deep Learning**

Soutenu publiquement, le 15/06/2026, devant le jury composé de :

Mr. Behaddada Omar	MCA	Université de Tlemcen	Président
Mme. Feroui Amel	Pr	Université de Tlemcen	Examinatrice
Mme. Benchaib Yasmine	MCA	Université de Tlemcen	Encadrante

Année universitaire : 2025/2026

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



Dédicace

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
(وَمَا تَوْفِيقِي إِلَّا بِاللَّهِ عَلَيْهِ تَوَكَّلْتُ وَإِلَيْهِ أُنِيبُ)

Après un parcours de labeur et de persévérance, et après de longues heures de recherche et de études les pages de ce travail se referment aujourd'hui pour voir le jour. Cet effort n'aurait pu s'accomplir sans la grâce de Dieu tout-puissant d'abord, puis sans le soutien des âmes qui m'ont accompagnée à chaque pas

À mon cher Père, le pilier qui ne vacille jamais, et le phare qui a illuminé l'obscurité de mon chemin. Je te dédie cette réussite, toi qui m'as appris que l'ambition n'a pas de limites, et que le travail et l'assiduité sont les clés de toute porte fermée. Mes mots ne suffisent pas à te remercier ; j'espère seulement que cet accomplissement sera la première goutte de pluie pour te rendre ton immense bienfait.

À mon paradis sur terre, la source de ma quiétude et mon ange dans la vie... ma Mère bien-aimée. À toi qui as veillé pour que je me repose, qui as prié pour que je réussisse, et qui as porté mes soucis avant les tiens. Ce jour est le tien avant d'être le mien, et ma joie aujourd'hui n'est que le reflet des larmes de bonheur dans tes yeux. Merci d'avoir été ce soutien inépuisable et le baume guérissant toutes mes peines.

À ceux qui font ma force, mes frères et ma sœur, compagnons de route et partenaires de vie. À vous qui avez été d'une aide précieuse, qui avez partagé avec moi la fatigue des jours et m'avez encouragée dans les moments de doute. Merci pour votre présence qui donne des couleurs à la vie. Vous êtes le soutien dont je suis fière, et ma joie aujourd'hui ne saurait être complète sans voir la fierté dans vos yeux.

À ma compagne de route et partenaire de réussite... ma collègue et sœur Ferial Plus qu'une camarade, nous avons partagé les nuits blanches, les pressions et les joies de ce parcours universitaire avec une seule âme. Merci pour ton dévouement inoubliable et d'avoir surmonté tous ces défis à mes côtés.

À mes amis fidèles : À ceux avec qui j'ai partagé les plus beaux souvenirs de cette étape, à ceux qui ont été les meilleurs compagnons de route dans mon parcours universitaire. Merci d'avoir allégé mes fardeaux par vos sourires sincères et votre soutien moral ininterrompu.

Chaima



Dédicace

الحمد لله الذي وفقنا لإتمام هذا العمل بفضلته وكرمه، والحمد لله في البدء وفي الختام

Au terme de ce long voyage académique, c'est avec une profonde émotion que je dédie l'aboutissement de ces efforts à ceux qui ont éclairé mon chemin.

À mon père et ma mère, véritables piliers de ma vie. Votre amour inépuisable, vos prières sincères et vos innombrables sacrifices m'ont donné des ailes pour atteindre ce but. Aucun mot ne saurait exprimer l'immensité de ma gratitude. Ce succès est avant tout le vôtre. À mes frères, pour votre soutien constant, vos encouragements et cette chaleur familiale qui m'a toujours rassuré.

Je dédie également ces pages à mon binôme, Chaima. Ensemble, nous avons affronté la complexité de ce projet, partagé le stress des nuits blanches et savouré chaque petite victoire. Merci pour ton dévouement, ta patience et cette belle harmonie qui a rendu ce travail possible.

À mes amies exceptionnelles, Asma et Hadjer. Merci d'avoir été mon refuge dans les moments de fatigue, pour vos mots réconfortants et votre présence lumineuse qui m'ont toujours redonné le sourire au moment où j'en avais le plus besoin.

Enfin, je me dédie ce mémoire. Pour les moments de persévérance face à la difficulté, pour la détermination qui a pris le dessus sur l'épuisement, et pour avoir cru en moi jusqu'à la toute dernière ligne

Ferial



Remerciement

Avant tout, nous remercions Dieu, le Tout-Puissant, de nous avoir donné la force, la patience et la volonté d'accomplir ce modeste travail.

Nous tenons à exprimer notre profonde gratitude et nos plus sincères remerciements à notre encadrante, **Mme BENCHAIB Yasmine**. Nous vous remercions pour votre patience, votre disponibilité et surtout pour vos judicieux conseils. Votre expertise, vos orientations et votre soutien constant ont été d'une aide inestimable pour mener à bien cette étude, structurer notre méthodologie et mener à bien le développement de notre système. Merci d'avoir partagé avec nous votre passion pour la recherche.

Nos vifs remerciements s'adressent également aux honorables membres du jury, qui nous font l'immense honneur de lire, d'évaluer et de juger ce travail : **Mr. Omar Behaddada** pour avoir accepté de présider ce jury, ainsi que **Mme. A. Feroui** pour avoir accepté d'examiner ce projet. Leur examen attentif de notre mémoire et leurs suggestions pertinentes ont contribué de manière significative à l'amélioration de ce travail.

Nous tenons également à remercier l'ensemble des enseignants de **l'Université de Tlemcen**, qui ont contribué à notre formation tout au long de notre cursus universitaire et qui nous ont fourni les bases solides nécessaires pour aborder des domaines aussi complexes.

Enfin, nous remercions chaleureusement toutes les personnes qui ont contribué de près ou de loin à la réussite de ce projet de fin d'études.

Résumé :

Ce mémoire s'inscrit dans le cadre de l'amélioration du dépistage précoce du cancer du sein en Algérie, un enjeu de santé publique majeur. L'échographie, bien que largement accessible, souffre d'artefacts visuels (bruit de speckle) compliquant l'interprétation clinique. Ce projet propose un système de diagnostic assisté par ordinateur (DAO) de bout en bout, développé pour automatiser la segmentation et la classification des tumeurs mammaires. L'approche intègre un prétraitement robuste incluant l'algorithme CLAHE. Pour la délimitation tumorale, l'architecture Attention U-Net s'est distinguée, atteignant un excellent score Dice de 0,914 après adaptation. Concernant la classification (bénigne, maligne, normale), une évaluation rigoureuse a mis en évidence la supériorité du Vision Transformer (ViT-B/16) sur les réseaux convolutifs (EfficientNet-B4, ResNet-50). Optimisé par AdamW et couplé à une stratégie d'augmentation au test (TTA $\times 8$), le ViT atteint une exactitude de 93,16 % sur la base BUSI. Sa robustesse a été validée face à une rupture de domaine clinique sur la base externe BRUSBRA, maintenant une exactitude de 89,36 % après Fine-Tuning. De plus, une stratégie d'Ensemble Learning a permis d'atteindre un rappel de 100 % sur la classe maligne, éliminant totalement les faux négatifs. L'intégration de l'explicabilité visuelle (Attention Rollout, Grad-CAM) et le développement de l'interface graphique « **SenoDeep Predict** » viennent consolider la transparence et l'applicabilité finale de ce système en milieu hospitalier.

Mots-clés : Cancer du sein, Échographie, Vision Transformer (ViT), Attention U-Net, TTA $\times 8$, Ensemble Learning, Système CAD, SenoDeep Predict.

Abstract

This thesis is part of the improvement of early detection of breast cancer in Algeria, a major public health issue. The ultrasound, although widely accessible, suffers from visual artifacts (speckle noise) complicating clinical interpretation. This project proposes an end-to-end computer-aided diagnostic (CAD) system, developed to automate the segmentation and classification of mammary tumors. The approach integrates a robust pretreatment including the CLAHE algorithm. For tumor delimitation, the U-Net Attention architecture stood out, achieving an excellent Dice score of 0.914 after adaptation. Regarding the classification (benign, malignant, normal), a rigorous evaluation highlighted the superiority of the Vision Transformer (ViT-B/16) over convolutional networks (EfficientNet-B4, ResNet-50). Optimized by AdamW and coupled with a test increase strategy (TTA $\times 8$), the ViT achieves an accuracy of 93.16% on the BUSI basis. Its robustness has been validated against a clinical domain break on the external BRUSBRA database, maintaining an accuracy of 89.36% after Fine-Tuning. In addition, an Ensemble Learning strategy made it possible to achieve a 100% recall on the malignant class, completely eliminating false negatives. The integration of visual explicability (Attention Rollout, Grad-CAM) and the development of the « **SenoDeep Predict** » graphical interface consolidates the transparency and ultimate applicability of this system in hospital settings.

Keywords: Breast cancer, Ultrasound, Vision Transformer (ViT), U-Net attention, TTA $\times 8$, Ensemble Learning, CAD system, SenoDeep Predict.

الملخص

تعد هذه الأطروحة جزءًا من تحسين الكشف المبكر عن سرطان الثدي في الجزائر، وهي قضية صحية عامة كبرى. على الرغم من إمكانية الوصول إلى الموجات فوق الصوتية على نطاق واسع، إلا أنها تعاني من آثار بصرية (ضوضاء مرقطة) مما يعقد التفسير السريري. يقترح هذا المشروع نظامًا تشخيصيًا متكاملًا بمساعدة الكمبيوتر (CAD)، تم تطويره لأتمتة تقسيم وتصنيف أورام الثدي. يدمج هذا النهج معالجة مسبقة قوية بما في ذلك خوارزمية CLAHE. بالنسبة لتحديد الورم، برزت بنية U-Net Attention، حيث حققت درجة Dice ممتازة بلغت 0.914 بعد التكيف. وفيما يتعلق بالتصنيف (حميد، خبيث، طبيعى)، سلط تقييم دقيق الضوء على تفوق محول الرؤية (ViT-B/16) على الشبكات التلافيفية (ResNet-50، EfficientNet-B4). تم تحسين ViT بواسطة AdamW وربطه باستراتيجية زيادة الاختبار (TTA $\times 8$)، ويحقق دقة بنسبة 93.16% على أساس BUSI. تم التحقق من متانتها ضد انقطاع المجال السريري في قاعدة بيانات BRUSBRA الخارجية، مع الحفاظ على دقة قدرها 89.36% بعد الضبط الدقيق. بالإضافة إلى ذلك، أتاحت استراتيجية التعلم الجماعي تحقيق تذكر بنسبة 100% للفئة الخبيثة، والقضاء تمامًا على النتائج السلبية الكاذبة. إن دمج القدرة على التفسير البصري (طرح الانتباه، Grad-CAM) وتطوير الواجهة الرسومية «SenoDeep Predict» يعزز الشفافية والتطبيق النهائي لهذا النظام في المستشفيات.

الكلمات المفتاحية: سرطان الثدي، الموجات فوق الصوتية، محول الرؤية (ViT)، انتباه U-Net، $\times 8$ TTA، التعلم الجماعي، نظام CAD، SenoDeep Predict.

Table des matières :

<i>Dédicace</i>	<i>ii</i>
<i>Dédicace</i>	<i>iii</i>
<i>Remerciement</i>	<i>iv</i>
<i>Résumé :</i>	<i>v</i>
<i>المخلص</i>	<i>vii</i>
<i>Table des matières :</i>	<i>viii</i>
<i>Liste des figures :</i>	<i>xii</i>
<i>Liste des tableaux :</i>	<i>xiv</i>
<i>Introduction générale</i>	<i>1</i>

Chapitre 01 :Contexte médical et bases de l'imagerie mammaire

<i>I.1 Introduction :</i>	<i>7</i>
<i>I.1 Aperçu sur le cancer du sein</i>	<i>8</i>
I.1.1 Définition et facteurs de risque essentiels :	8
I.1.2 Intérêt du dépistage et du diagnostic précoces.....	8
I.1.3 Anatomie du cancer du sein :	9
I.1.3.1 Explications pour l'anatomie cancer de sein.....	9
<i>I.2 Données épidémiologiques :</i>	<i>10</i>
I.2.1 Prévalence et importance du cancer du sein à l'échelle mondiale	10
I.2.2 Prévalence et importance du cancer du sein en Algérie.....	10
<i>I.3 Les différents types et stades cliniques du cancer du sein</i>	<i>11</i>
I.3.1 Classification histologique du cancer du sein.....	11
I.3.2 Classification selon le caractère invasif	12
I.3.3 Classification moléculaire du cancer du sein	12
I.3.4 Stades cliniques du cancer du sein : classification TNM.....	13
I.3.5 Importance clinique de la classification :	14
<i>I.4 Imagerie médicale pour le diagnostic :</i>	<i>15</i>
I.4.1 Mammographie et limites :	15
I.4.2 Échographie mammaire : principe et intérêt clinique et limites :.....	16
I.4.2.1 Principe physique de l'échographie mammaire :	16
I.4.2.2 Intérêt clinique de l'échographie mammaire :	17
I.4.2.3 Limites de l'échographie mammaire :	17
I.4.2.4 Vers l'analyse automatique des images échographiques :.....	17
.....	18
I.4.3 Autres modalités (IRM – aperçu) :	18
<i>I.5 Motivation pour l'analyse automatique des images échographiques :</i>	<i>19</i>
<i>I.6 Conclusion du Chapitre 1 :</i>	<i>20</i>

Chapitre 02 : Méthodologie et travail réalisé

II.1	Introduction :	22
II.2	Fondamentaux du Deep Learning pour l'Imagerie Médicale :	23
II.2.1	Réseaux de Neurones Convolutifs (CNN) :	23
II.2.2	Principe de fonctionnement des CNN :	23
II.2.2.1	Convolution :	23
II.2.2.2	Le Pooling (Sous-échantillonnage) :	24
II.2.2.3	Fonction d'activation (ReLU) :	24
II.2.2.4	Couches entièrement connectées (Fully Connected) :	24
II.2.2.5	Couche de sortie :	25
II.2.3	L'Architecture U-Net : Le standard de la segmentation :	25
II.2.3.1	Pourquoi choisir l'U-Net plutôt qu'un CNN standard ?	25
II.2.3.2	Chemin contractant (Encodeur) :	25
II.2.3.3	Le Décodeur (le chemin montant) :	26
II.2.3.4	Le rôle vital des Skip Connections :	26
II.2.4	Classification des tumeurs du sein par échographie à l'aide de réseaux (CNN) :	26
II.3	Métriques d'Évaluation et Interprétabilité :	27
II.3.1	Quantification de la Précision Géométrique (Segmentation) :	27
II.3.1.1	Coefficient de Dice (Dice Similarity Coefficient) :	27
II.3.1.2	Intersection over Union (IoU) - Indice de Jaccard :	27
II.3.1.3	Distance de Hausdorff :	27
II.3.2	Analyse de la Performance Diagnostique (Classification) :	27
II.3.2.1	Matrice de Confusion (Confusion Matrix) :	28
II.3.2.2	Accuracy (Exactitude globale) :	28
II.3.2.3	Balanced Accuracy (Précision équilibrée) :	28
II.3.2.4	Sensibilité (Recall / Rappel) :	28
II.3.2.5	Précision :	29
II.3.2.6	Spécificité (Specificity) :	29
II.3.2.7	F1-Score :	29
II.3.2.8	F1 Macro et F1 Weighted :	29
II.3.2.9	Macro Avg et Weighted Avg :	29
II.3.2.10	Indice Kappa (kappa) :	30
II.3.2.11	MCC (Matthews Correlation Coefficient) :	30
II.3.2.12	AUC-ROC :	30
II.3.3	Interprétabilité par Grad-CAM et Attention Rollout :	30
II.4	ÉTAT DE L'ART :	31
II.4.1	État de l'Art des travaux de segmentation sur le dataset BUSI :	31
II.4.2	Analyse des travaux antérieurs sur la classification du dataset BUSI :	32
II.5	CONTEXTE ET DONNÉES :	33
II.5.1	Description de la démarche globale :	33
II.5.2	Environnement de développement et ressources techniques :	34
II.6	Description de la Base de Données (Dataset BUSI) :	34
II.6.1	Présentation de la base de données :	34
II.6.2	Structure et organisation des données :	35
II.6.3	L'importance des masques de segmentation :	35
II.6.4	Pourquoi avoir choisi le dataset BUSI ?	35
II.7	Prétraitement et Augmentation des Données :	36
II.7.1	Harmonisation et redimensionnement :	36

II.7.2	Augmentation de données (Offline) :	36
II.7.3	CLAHE (le contraste adaptatif) :	37
II.7.4	Normalisation pour le réseau :	37
II.7.5	Format PyTorch et GPU :	37
II.8	Architecture U-Net pour la Segmentation :	38
II.8.1	U-Net classique :	38
II.8.2	Attention U-Net :	38
II.8.3	U-Net++ :	38
II.9	Architecture pour la Classification :	40
II.9.1	ResNet-50 :	40
II.9.2	EfficientNet-B4 :	40
II.9.3	Vision Transformer (ViT-B/16) :	40
II.10	Stratégie d'apprentissage :	43
II.10.1	Découpage des données et Validation :	43
II.10.2	Techniques d'optimisation :	44
II.10.2.1	Transfer Learning :	44
II.10.2.2	AdamW et Gestion du Learning Rate:	44
II.10.2.3	Focal Loss et WeightedRandomSampler :	45
II.10.2.4	Accumulation de Gradients :	45
II.10.2.5	Test-Time Augmentation (TTA) :	45
II.10.3	Optimisation et Régularisation du ViT :	46
II.10.4	Stratégie d'Ensemble (Décision finale) :	46
II.11	Conclusion du Chapitre 2 :	47

Chapitre 03 : Evaluation expérimentale interprétabilité des modèles de Deep Learning

III.1	Introduction :	49
III.2	Présentation et interprétation des résultats de segmentation :	50
III.2.1	Courbes d'apprentissage :	50
III.2.2	Comparaison globale des modèles :	51
III.2.2.1	La supériorité de l'Attention U-Net :	51
III.2.2.2	U-Net++ :	52
III.2.2.3	U-Net :	53
III.2.3	Distance de Hausdorff :	55
III.3	Présentation et interprétation des résultats de classification :	59
III.3.1	EfficientNet B4 :	59
III.3.2	ResNet 50 :	61
III.3.2.1	Courbe ROC :	63
III.3.3	ViT B/16 :	64
III.3.3.1	Analyse des courbes d'apprentissage (ViT-B/16) :	65
III.3.3.2	Interprétation des résultats finaux (ViT-B/16) :	65
III.3.3.3	Courbe ROC :	67
III.3.3.4	Matrice de confusion :	68
III.3.3.5	Interprétabilité du modèle ViT-B/16 via l'Attention Rollout :	69
III.3.4	Comparaison des trois modèles :	71
III.4	Évaluation des modèles en mode Zero-shot :	72
III.5	Fine-Tuning :	74
III.6	Évaluation visuelle du modèle Attention U-Net :	75

III.7	Évaluation de l'interprétabilité spatiale : Grad-CAM et robustesse des modèles :	
		76
III.8	Analyse des Erreurs et Apport de l'Ensemble Learning :	78
III.8.1	Analyse approfondie des erreurs de classification :	78
III.8.2	L'Ensemble Learning :	79
III.9	Oversampling :	80
III.9.1	Analyse des Courbes d'Apprentissage :	81
III.10	Comparaison avec l'État de l'Art et Discussion :	83
III.10.1	Comparaison des performances de Segmentation :	83
III.10.2	Comparaison des performances de Classification :	84
III.11	Présentation de l'Interface « SenoDeep Predict » :	85
III.12	Conclusion :	88
Conclusion Générale		89
Références		92
Bibliographiques		92

Liste des figures :

Figure 1: contexte médical et bases de l'imagerie mammaire [12]	9
Figure 2 : Image de machine de mammographie [18].	16
Figure 3 : Image de machine de l'échographie [23]	18
Figure 4 : Schéma illustratif du principe d'une couche entièrement connectée [29].	24
Figure 5 : Architecture générale d'un réseau de neurones convolutif (CNN) [29].	25
Figure 6: Architecture globale de la chaîne de traitement proposée.	33
Figure 7 : Processus d'augmentation de données appliqué aux images échographiques et à leurs masques associés.	37
Figure 8: Architecture du pipeline de segmentation d'images échographiques mammaires	39
Figure 9: Architecture détaillée du flux de traitement par Vision Transformer (ViT-B/16).	41
Figure 10: Pipeline méthodologique : Comparaison des paradigmes CNN et Vision Transformer.	43
Figure 11 : Courbes de précision (Training Accuracy / Validation Accuracy) pour les trois architectures.	50
Figure 12: Courbes de perte (Training Loss / Validation Loss) pour U-Net, Attention U-Net et U-Net++.	50
Figure 13: Comparaison des performances des architectures de segmentation (U-Net, Attention U-Net, U-Net++).	54
Figure 14 : Dynamique d'apprentissage (Loss, Dice, IoU, Sensibilité et Spécificité) des trois architectures de segmentation.	54
Figure 15: Moyenne de la distance de Hausdorff par modèle (valeurs basses = meilleures).	56
Figure 16: Résultats qualitatifs détaillés par classe (bénigne, maligne, normale) — Attention U-Net.	57
Figure 17: Courbe de perte (Training Loss / Validation Loss) et Courbe de précision (Training Accuracy / Validation Accuracy) pour EfficientNet B4.	60
Figure 18 : Courbe de perte (Training Loss / Validation Loss) et Courbe de précision (Training Accuracy / Validation Accuracy) pour ResNet 50.	62
Figure 19: Courbe de ROC (Receiver Operating Characteristic) pour ResNet 50.	63
Figure 20: Courbe de perte (Training Loss / Validation Loss) et Courbe de précision (Training Accuracy / Validation Accuracy) pour ViT B/16.	65
Figure 21: Courbe de ROC (Receiver Operating Characteristic) pour ViT B/16.	67
Figure 22: Matrice de confusion de ViT B/16.	68
Figure 23: Cartes de chaleur générées par Attention Rollout illustrant les zones d'intérêt du modèle ViT-B/16 pour les classes bénignes, malin et normal.	70
Figure 24: Exemple de segmentation d'une tumeur mammaire par le modèle Attention U-Net.	75
Figure 25: Comparaison de l'interprétabilité spatiale par Grad-CAM++ entre ResNet-50 et EfficientNet-B4.	77
Figure 26: Matrice de confusion de ViT B/16 sur la base équilibrée.	81
Figure 27: Évolution de la perte (Loss) et de la précision (Accuracy) du modèle ViT-B/16.	82

Figure 28: Aperçu de l'interface SenoDeep Predict montrant la zone de téléchargement de l'image échographique.	85
Figure 29: Interface affichant simultanément la classification (Bénin) et le masque de segmentation de la lésion échographique.	86
Figure 30: Interface affichant simultanément la classification (Malin) et le masque de segmentation de la lésion échographique.	86
Figure 31: Interface affichant simultanément la classification (Normal).	87
Figure 32: Aperçu du rapport clinique téléchargeable regroupant les informations de la patiente, le diagnostic final et l'analyse visuelle.	87

Liste des tableaux :

Tableau 1:les différentes classifications stades pour le cancer du sein [7] [17].....	14
Tableau 2: Répartition du dataset BUS.....	35
Tableau 3 : Comparaison des modèles de segmentation.	39
Tableau 4 : Comparaison des modèles de classification.....	42
Tableau 5 : Métriques de segmentation.....	51
Tableau 6: Synthèse des indicateurs de performance globale du modèle EfficientNet-B4.....	59
Tableau 7: Rapport de classification détaillé par classe (Bénigne, Maligne, Normale) pour l'architecture EfficientNet-B4	60
Tableau 8 : Synthèse des indicateurs de performance et de fiabilité du modèle ResNet-50	62
Tableau 9: Métriques Globales (Test Set vs Cross-Validation).	66
Tableau 10: Rapport de classification	66
Tableau 11: récapitulatif des performances sur le Test Set.	71
Tableau 12: Résultats de la classification Zero-Shot sur le dataset BRUSBRA.	73
Tableau 13: Résultats de la segmentation Zero-Shot sur le dataset BRUSBRA.....	73
Tableau 14: Résultats de la segmentation après Fine-Tuning sur BRUSBRA	74
Tableau 15 : Résultats de la classification après Fine-Tuning sur BRUSBRA.....	74
Tableau 16: Analyse des Faux Positifs.....	78
Tableau 17: Analyse des Faux Négatifs	78
Tableau 18: Comparaison des performances stratégies d'Ensemble Learning	79
Tableau 19: Rapport de classification du modèle ViT-B/16 sur l'ensemble de test.	80
Tableau 20: Comparaison des performances de segmentation de l'Attention U-Net avec les travaux de l'état de l'art sur le dataset BUSI.	83
Tableau 21:Évaluation comparative de l'exactitude (Accuracy) du modèle ViT-B/16 face aux architectures de la littérature.	84



Introduction

générale



Contexte :

De nos jours, le cancer du sein représente l'une des tumeurs malignes les plus répandues chez les femmes dans le monde et en Algérie aussi. Il pose un défi de santé publique important en raison de ses dimensions thérapeutiques, sociales et économiques complexes.

En Algérie, cette situation est extrêmement grave, avec des taux d'incidence en constante augmentation, ce qui en fait la principale cause de décès liés au cancer chez les femmes dans le pays. Ce qui exacerbe cette réalité inquiétante est un phénomène structurel représenté par la majorité des patients algériens arrivant dans les centres de traitement à des stades avancés de la maladie, périodes au cours desquelles les options de traitement se réduisent fortement. Par conséquent, un diagnostic précoce est la pierre angulaire et le déterminant essentiel pour augmenter les chances de guérison et sauver des vies à long terme.

L'échographie s'est imposée comme une méthode d'imagerie de première ligne dans le milieu de santé algérien, du fait de la disponibilité de ses équipements, de son coût financier et de sa sécurité radiologique. Cependant, l'interprétation précise des images demeure une tâche anatomiquement extrêmement complexe ; la qualité du diagnostic dépend entièrement de la précision de l'observation visuelle du médecin, qui fait face à des défauts physiques inhérents à la nature des ultrasons, tels qu'un bruit acoustique aigu (Speckle Noise) et des ombres d'arrière-plan qui obscurcissent les bords des tumeurs. Cette disparité visuelle conduit cliniquement à des différences dans les opinions diagnostiques parmi les médecins, en plus du stress mental résultant de l'afflux massif de patientes dans les hôpitaux algériens. Ainsi, l'intégration de l'intelligence artificielle, en particulier l'apprentissage profond, représente une étape prometteuse pour surmonter ces obstacles en automatisant l'analyse d'images avec une grande précision.

Le problème :

Face au besoin pressant d'outils de diagnostic rapides et objectifs, ce projet de fin d'études propose une solution basée sur l'intelligence artificielle. Notre approche exploite les réseaux de neurones profonds afin de développer un système d'aide au diagnostic clinique (CAD) dédié au cancer du sein.

L'objectif principal est de construire un pipeline d'analyse automatisé capable, d'une part, de segmenter avec précision les masses tumorales pour en délimiter les contours, et d'autre part, de les classer de manière fiable en trois catégories cliniques (bénigne, maligne ou normale), en s'appuyant sur la base de données publique "BUSI".

Cependant, cette étude fait face à un certain nombre de défis structurels qui façonnent la problématique de recherche ; plus particulièrement, la rareté des données médicales annotées, le déséquilibre statistique entre les catégories de tumeurs et les défauts physiques affectant la qualité de l'image, en plus de la nécessité de fournir une explication logique pour les décisions algorithmiques afin d'assurer leur acceptation dans la communauté médicale. Ces obstacles soulèvent des questions fondamentales :

Comment concevoir un système intelligent qui surmonte les distorsions optiques des ondes ultrasonores et fournit un diagnostic très précis et cliniquement interprétable malgré le manque de données médicales disponibles ?

Motivations et objectifs :

Notre motivation première provient de la volonté de contribuer à l'amélioration des pratiques de diagnostic radiologique des tumeurs du sein en Algérie, notamment au regard de la répartition inégale des ressources médicales spécialisées. L'intégration de l'intelligence artificielle ici vise à réduire la charge cognitive des radiologues, à réduire les taux d'erreur humaine et à accélérer le rythme de la prise de décision thérapeutique.

Ce projet vise à évaluer l'efficacité des structures d'apprentissage profond, en particulier grâce au mécanisme de « l'apprentissage par transfert de connaissances » (Transfer Learning), pour classer et segmenter les tumeurs du sein. Pour surmonter l'obstacle de la rareté de l'image, un protocole extensif a été développé afin d'élargir par programmation la taille des données pour assurer une formation stable des modèles.

Les objectifs spécifiques

Les objectifs spécifiques de ce projet comprennent :

- **Mettre en place un pipeline complet de prétraitement** : Développer une ligne de traitement intégrée dédiée à l'amélioration du contraste (via l'algorithme CLAHE) et à la normalisation des images échographiques afin d'atténuer les artefacts physiques (speckle noise).
- **Évaluer et comparer les architectures de segmentation** : Réaliser une étude comparative approfondie entre plusieurs variantes de modèles de la famille U-Net afin d'identifier la structure la plus efficace pour délimiter précisément les contours tumoraux.
- **Concevoir et entraîner des modèles de classification de pointe** : Déployer des architectures contemporaines de vision par ordinateur (modèles convolutifs et transformer de vision) pour classer de manière fiable les lésions mammaires en trois catégories cliniques : normale, bénigne et maligne.
- **Optimiser la robustesse face au déséquilibre des classes** : Intégrer un protocole d'augmentation de données offline massif et un arsenal de régularisation strict (fonctions de perte adaptatives, échantillonnage pondéré) pour contrer la rareté des données médicales et éliminer le biais lié au déséquilibre des classes.
- **Développer des outils d'explicabilité et d'interprétabilité visuelle** : Ouvrir la « boîte noire » des algorithmes de Deep Learning (via la méthode Attention Rollout) afin de cartographier les régions anatomiques déterminantes pour la décision et renforcer la confiance des cliniciens dans le diagnostic automatisé.

Contributions

Ce projet apporte plusieurs contributions fondamentales au domaine du diagnostic médical assisté par ordinateur du cancer du sein :

- **Optimisation de la délimitation tumorale** : Conception et évaluation d'une approche comparative avancée mettant en jeu plusieurs configurations de réseaux encodeurs-décodeurs, garantissant une segmentation précise et automatisée des zones lésionnelles.
- **Dépassement de l'obstacle lié au manque de données** : Validation d'un protocole d'augmentation et d'expansion de données offline couplé à des stratégies de régularisation avancées, assurant une forte capacité de généralisation des modèles face à de nouvelles images.
- **Évaluation comparative rigoureuse des architectures de classification** : Apprentissage et confrontation systématique, par un protocole strict de validation croisée en 5 étapes (5-Fold Cross-Validation), d'architectures convolutives basées sur le compound scaling (CNN, telles qu'EfficientNet-B4 et ResNet-50) et d'architectures de pointe basées sur les mécanismes d'attention globale (le Vision Transformer ViT-B/16).
- **Transparence et interprétabilité clinique** : Levée de l'opacité décisionnelle des modèles à travers l'extraction de cartographies thermiques interprétables, permettant aux praticiens d'identifier avec exactitude les régions anatomiques sur lesquelles le système a fondé son diagnostic.

Organisation du mémoire :

Afin de présenter cette démarche scientifique de manière progressive et cohérente, ce mémoire est structuré en trois chapitres complémentaires :

- **Le Chapitre 1 — Contexte médical et bases de l'imagerie mammaire** : ce chapitre pose les fondements cliniques et épidémiologiques indispensables de la pathologie.

Il couvre l'aspect clinique du cancer du sein ainsi que son anatomopathologie, tout en mettant l'accent sur la sémiologie de l'imagerie échographique, ses avantages et ses limites physiques, justifiant ainsi le recours aux technologies d'apprentissage profond.

- **Le Chapitre 2 — Fondamentaux, État de l'Art et Méthodologie** :

Ce chapitre représente le cœur technique du projet et fait le pont entre la théorie et notre démarche pratique. Il débute par un rappel des concepts essentiels du Deep Learning en imagerie médicale et une analyse critique des travaux existants sur le dataset BUSI. Ensuite, il plonge dans le vif de notre expérimentation, il décrit la préparation minutieuse des images (prétraitement par CLAHE et augmentation de données), avant de détailler les architectures de pointe que nous avons déployées (les variantes de l'U-Net pour la segmentation, ainsi que des modèles comme EfficientNet et Vision Transformer pour la classification). Enfin, il met en lumière nos choix stratégiques pour garantir la fiabilité du système, tels que l'utilisation de l'optimiseur AdamW pour l'apprentissage, la gestion du déséquilibre des classes, et l'application du TTA (Test-Time Augmentation) pour consolider la décision finale du modèle.

- **Le Chapitre 3 — Résultats expérimentaux, discussion et interprétabilité :**

Dans ce dernier chapitre nous présentons, l'évaluation expérimentale et quantitative complète du système. Nous analysons en profondeur les performances numériques des différents modèles de segmentation (en mettant en évidence la supériorité de l'Attention U-Net) et de classification (EfficientNet-B4, ResNet-50, ViT-B/16).

Ce chapitre explore également des stratégies d'évaluation avancées (Zero-shot, Fine-Tuning, Oversampling, Ensemble Learning) et apporte une justification visuelle des décisions automatisées grâce aux outils d'explicabilité (Grad-CAM et Attention Rollout). Enfin, pour couronner ce travail technique, le chapitre se clôture par la présentation de l'interface graphique développée « **SenoDeep Predict** », démontrant ainsi le déploiement pratique et l'applicabilité finale de notre système d'aide au diagnostic clinique.

Une conclusion générale vient clore ce manuscrit afin de synthétiser les principaux résultats de cette étude, d'en discuter les limites actuelles et de présenter les perspectives de recherche futures ainsi que les possibilités de déploiement de ce système d'aide au diagnostic en pratique clinique.



Chapitre 01 :

Contexte médical et bases de l'imagerie mammaire



I.1 Introduction :

Aujourd'hui, le cancer du sein est un grand souci pour la santé publique dans le monde entier. C'est le cancer qui arrive le plus souvent chez les femmes, et il cause beaucoup de morts liés à ce type de maladie. Chaque année, on trouve plus de nouveaux cas. C'est pourquoi il faut vraiment faire des efforts pour prévenir, détecter et suivre.

La situation est préoccupante en Algérie, où le cancer du sein est le premier cancer chez les femmes. Encore beaucoup de patientes sont découvertes tardivement, surtout parce qu'elles ne vont pas assez au dépistage, qu'elles ne savent pas beaucoup et à cause de certaines habitudes. Ces longueurs font mal pour la guérison. C'est pourquoi c'est important de trouver tôt pour que les patientes aient une vie meilleure et plus longue.

Généralement, le cancer du sein vient sous divers types, tels que les carcinomes canaux et les carcinomes lobulaires, et sa progression se fait en plusieurs phases. C'est très important de savoir ces sortes et ces étapes pour bien savoir comment soigner et pour juger les chances de guérison.

Détecter tôt est l'une des meilleures façons pour réduire le nombre de morts à cause du cancer du sein. Si on voit une tumeur tôt, les traitements sont souvent plus doux et marchent mieux. Imagerie médicale est donc très importante pour étudier tissus mammaires sans toucher aussi bien que pour déterminer s'il existe des anomalies qui peuvent être graves.

Parmi les différentes modalités d'imagerie, l'échographie mammaire occupe une place essentielle, notamment chez les patientes jeunes et celles présentant une forte densité mammaire. Cet examen est largement utilisé car il est non invasif, facilement disponible et n'utilise pas de rayonnements ionisants. Cependant, l'analyse des images échographiques demeure difficile en raison du bruit de speckle, du faible contraste entre les tissus et des contours parfois imprécis des lésions, ce qui rend l'interprétation fortement dépendante de l'expérience du radiologue.

Pour résoudre ces problèmes, l'examen automatique des photos d'échographie semble être une bonne idée. La segmentation des images médicales permet de séparer automatiquement les zones importantes, comme les tumeurs. Dire si ces anomalies sont graves ou ne le sont pas est fait grâce à la classification. Ces deux points sont importants pour les systèmes qui aident les docteurs à faire leur diagnostic.

Dans ce cas, ce travail essaie d'utiliser les méthodes modernes comme traitement d'images et intelligence artificielle pour détecter et analyser les anomalies des seins sur les images d'échographie. L'objectif de ce chapitre est de présenter les bases médicales et théoriques nécessaires à la compréhension du sujet, en détaillant le cancer du sein, les méthodes d'imagerie mammaire et les concepts essentiels de l'analyse d'images médicale

I.1 Aperçu sur le cancer du sein

Le cancer du sein est une maladie caractérisée par une prolifération anormale et incontrôlée des cellules mammaires, pouvant envahir les tissus voisins et se propager à distance. Il constitue le cancer le plus fréquent chez la femme et représente un enjeu majeur de santé publique en raison de son incidence élevée et de son impact sur la mortalité féminine [1] [2].

I.1.1 Définition et facteurs de risque essentiels :

Le cancer du sein correspond au développement d'une tumeur maligne au niveau des tissus mammaires, résultant d'une prolifération anormale et incontrôlée des cellules, principalement issues des canaux galactophores ou des lobules mammaires [1] [2].

Plusieurs facteurs peuvent augmenter le risque de cette maladie. Parmi les facteurs qu'on ne peut pas changer, il y a l'âge vieux, les antécédents de famille et aussi quelques erreurs génétiques héréditaires comme des changements dans les gènes BCRA1 et BCRA2 qui sont liés à un gros risque de cancer du sein, souvent à un âge jeune [3].

Il y a d'autres facteurs aussi, appelées changeables, qui sont liées, comme les facteurs qui touchent le corps et les actions comme être trop grossier, ne pas bouger assez et boire trop d'alcool. Ces facteurs changent l'équilibre du corps et aident à rendre le risque plus grand pour avoir un cancer du sein, montrant donc comment c'est important de faire des actions pour éviter et trouver tôt chez les populations à haut risque [4].

I.1.2 Intérêt du dépistage et du diagnostic précoces

Le dépistage et le diagnostic rapides sont très importants dans le combat contre le cancer du sein. Découvrir la maladie tôt aide à trouver les lésions avant l'arrivée de signes médicaux, ce qui rend le pronostic des malades beaucoup mieux. Quand une tumeur est trouvée à un stade de début, les soins donnés sont en général moins durs, mieux supportés et plus efficaces, ce que fait monter beaucoup les chances de survivre [5].

Les programmes de vérification mis en place, qui s'appuient surtout sur la mammographie, ont montré qu'ils aident à réduire la mort due au cancer du sein [6]. Mais la mammographie a des défauts, surtout chez les femmes jeunes et celles qui ont des seins denses où le chevauchement des tissus peut cacher quelques zones d'endommagement [7]. Dans ce cas, l'échographie mammaire joue un rôle important de plus. Cela aide à mieux voir les tissus mous et facilite la trouvaille de certaines zones endommagées que l'on ne voit pas avec la mammographie [8] [9].

L'imagerie des corps est un pilier important du diagnostic tôt. Elle aide à trouver les changements de formes au sein mais aussi à voir leurs traits pour guider un choix médical. Mais savoir ce que les images montrent dépend encore de l'expérience du radiologue. C'est ce qui montre l'intérêt de faire des outils d'aide aux diagnoses qui peuvent faire avec une meilleure fiabilité et répétitions des résultats [10] [11].

En plus, plusieurs choses liées aux hormones et au comportement touchent le risque d'avoir un cancer du sein, comme l'obésité, ne pas bouger assez et boire de l'alcool. Ces choses prouvent combien il est important de faire des tests tôt spécialement dans les groupes à haut risque [4].

I.1.3 Anatomie du cancer du sein :

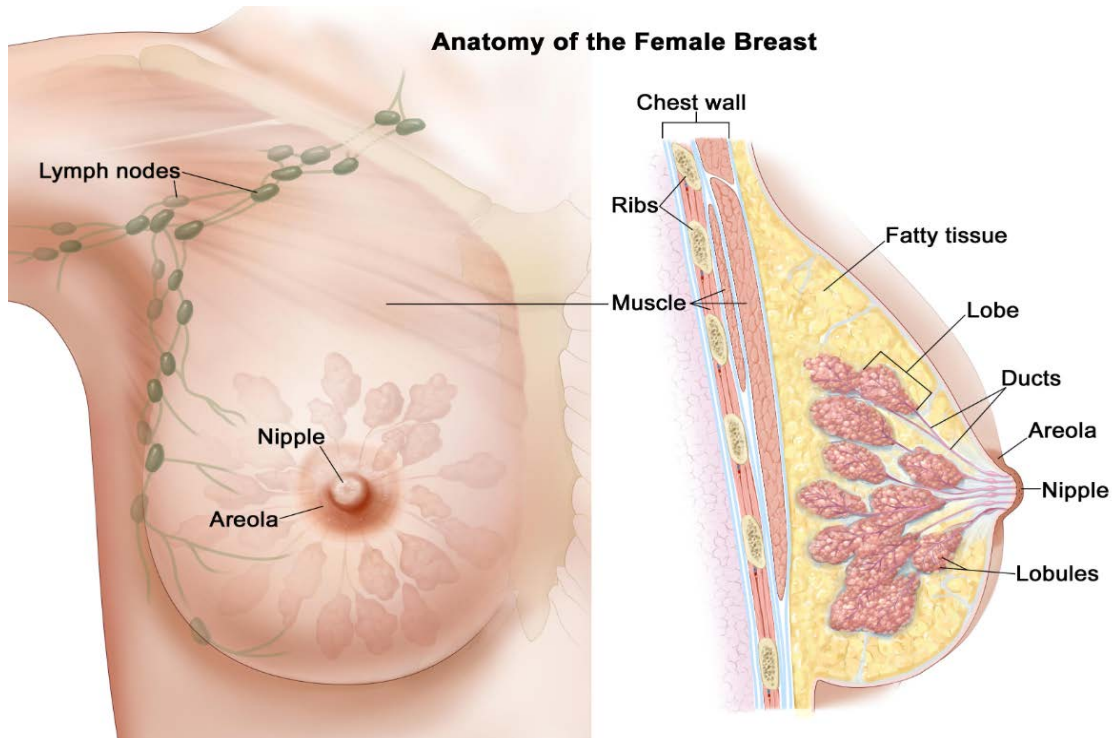


Figure 1: contexte médical et bases de l'imagerie mammaire [12]

I.1.3.1 Explications pour l'anatomie cancer de sein

L'illustration présente un schéma du sein féminin en perspective externe et en section transversale. Elle souligne les principales parties du sein. Sur la surface cutanée, le téton est visible, entouré par l'aréole, une zone de couleur plus sombre contenant des glandes spécialisées. La structure interne révèle que le sein est principalement constitué de tissu adipeux, ce qui influence majoritairement son volume et sa forme. Au sein de ce tissu se trouvent les lobes et lobules, qui sont des unités glandulaires responsables de la production de lait. Ces éléments sont reliés au téton par des canaux qui transportent le lait durant l'allaitement [12].

L'illustration illustre également la paroi thoracique, les côtés, notamment les côtes, ainsi que le muscle pectoral situé en profondeur. Il est important de noter que le tissu mammaire repose sur ce muscle sans y être intégré. La proximité de ces différentes parties corporelles explique qu'en cas d'une tumeur avancée, une propagation vers la paroi thoracique peut être observée. De plus, on peut observer les ganglions lymphatiques dans le corps, surtout sous les bras. Ces petits nœuds jouent un rôle crucial dans la dissémination des cellules cancéreuses, rendant leur étude essentielle pour évaluer l'étendue du cancer du sein [12].

Comprendre en fait la forme des seins est très important pour expliquer la croissance du cancer du sein. Souvent, les tumeurs qui croissent vite se forme soit dans les canaux galactophores (carcinome canalaire), soit à l'intérieur des lobules mammaires (carcinome lobulaire). Il est donc très important de voir l'anatomie en jeu pour une bonne vue des radios comme les mammographies, des échographies et des IRM du sein. Ces infos font aussi une base clef pour faire des outils d'ordinateur qui veulent aider à mieux trouver, découper et classer les problèmes de sein [12].

I.2 Données épidémiologiques :

Dans cette section, il s'agit de présenter des données fiables relatives à l'incidence, à la fréquence et à l'impact du cancer du sein à deux niveaux : mondial et national (Algérie). Ces informations sont issues de sources officielles et scientifiques reconnues, telles que l'Organisation mondiale de la santé (OMS), des études publiées dans des revues scientifiques et des registres nationaux du cancer [1] [2] [15].

I.2.1 Prévalence et importance du cancer du sein à l'échelle mondiale

De nos jours, le cancer du sein est un souci majeur pour la santé de chacun dans le monde. Il s'agit de la pathologie maligne la plus fréquente chez les femmes et une raison principale de décès dû au cancer chez elles [1] [2].

En 2022, environ 2,3 millions de nouveaux cas ont été comptés dans le monde ce qui a mené à près de 670 000 décès soit à peu près 25 % de tous les cancers chez les femmes [2].

On voit des grosses différences entre les zones. Dans les pays riches, le chiffre de cas est très élevé, comme près de 90 cas pour 100 000 femmes en Europe, surtout parce que la population vieillit l'urbanisation et plus d'exposition à certaines hormones [2] [4]. En revanche dans des pays à faible ou moyen revenu, notamment en Afrique saharienne le chiffre de mort demeure plus haut représentant presque 60% des morts dans le monde car le diagnostic est tardif le dépistage est lent et l'accès aux soins est difficile [1] [2].

Les choses qui font plus grand le danger de cancer du sein sont un tas. Des changements dans les gènes BRCA1 et BRCA2 sont connectés à environ 5 à 10 % des cas, ce qui rend beaucoup plus grand le danger d'avoir la maladie [3]. De plus, certains facteurs liés au mode de vie, tels que l'obésité après la ménopause ou la consommation régulière d'alcool, contribuent à accroître le risque [4]. En plus, quelques parties du style de vie comme : être gros après la ménopause ou boire souvent de l'alcool, ajoutent aussi au risque. L'effet sur l'économie est autant important : le coût mondial annuel pour soins pour le cancer du sein est autour 20 milliards de dollars Les inégalités dans l'accès aux soins rendent plus mal ce problème avec un taux de survie à 5 ans plus que 90 % dans les pays riches contre moins de 40 % dans certaine région d'Afrique subsaharienne [1] [2] [4].

I.2.2 Prévalence et importance du cancer du sein en Algérie

En Algérie, le cancer du sein reste le type de cancer le plus courant chez les femmes, représentant 42,7% de tous les cancers féminins dépérés. Chaque année, près de 15 000 nouveaux cas sont trouvés, ce qui donne un taux standardisé d'environ 45 cas pour 100 000

femmes. La mortalité annuelle est chiffrée à environ 3500 décès et le taux de survie à cinq ans demeure limité à 68 %, bien plus bas que celui vu dans les pays avancés où il dépasse généralement 90%, surtout à cause d'un diagnostic souvent tardif [13] [15].

Un aspect préoccupant est l'âge de survenue plus tôt : environ 30 % des cas sont détectés avant 40 ans, contre seulement 5 % en Europe [13].

Parmi les facteurs de risque spécifiques, une part importante de patientes a des mutations génétiques comme BRCA1 et BRCA2, surtout dans les formes tôt [3].

À cela s'ajoutent des facteurs comportementaux, tels que la sédentarité, touchant environ 60 % des femmes algériennes, et l'obésité, présente chez 35 % des femmes âgées de 35 à 60 ans [13].

L'accès au soin reste un défi grand, avec près de 60 % des cas trouvés aux stades II ou III et des régionales différences fortes surtout dans les zones rurales où manque de spécialistes et d'équipement rend une prise en charge vite impossible. Pour répondre à ce problème l'Algérie a mis en place une manière nationale de se battre contre le cancer, y compris un plan de dépistage fait depuis 2016 qui donne la mammographie gratuite aux femmes âgées entre 40 et 69 ans. Des campagnes pour sensibiliser, comme « Octobre Rose » ont aidé à mieux informer le public et à encourager la découverte tôt. Par ailleurs, le renforcement des infrastructures sanitaires, avec la création de 12 centres anticancéreux publics et l'intégration de thérapies ciblées comme le trastuzumab, représente une avancée notable dans la prise en charge du cancer du sein en Algérie [14] [15].

- **BRCA1** : associé à un risque élevé de cancer du sein, souvent de survenue précoce, et à un risque de cancer ovarien [3].
- **BRCA2** : également lié à un risque accru de cancer du sein, avec un profil légèrement différent, et associé à d'autres cancers comme la prostate ou le pancréas [3].

I.3 Les différents types et stades cliniques du cancer du sein

Le cancer du sein est une maladie hétérogène, regroupant plusieurs types différents sur les plans vus de l'histologie, de la biologie et clinique. Ce n'est pas une seule maladie, mais un groupe de tumeurs qui montrent des comportements différents avec le temps, des réactions aux soins variant et des perspectives diverses. La division du cancer du sein dépend surtout sur trois méthodes supplémentaires : la classification histologique, la moléculaire et la stadification clinique suivant le système TNM. Cette catégorisation est très importante, car elle influence le choix du traitement, la mesure du pronostic et la stratégie de suivi pour les patientes [7] [16].

I.3.1 Classification histologique du cancer du sein

La classification histologique est basée sur l'examen anatomopathologique des tissus tumoraux et aide à identifier le tissu d'origine des cellules cancéreuses ainsi que leur façon de croître [17].

Le **carcinome canalaire** est la forme histologique la plus commune du cancer du sein. Il commence dans les canaux galactophores et représente environ 70 à 80% des cas. Il peut se

montrer sous forme in situ, quand les cellules cancéreuses restent renfermées dans l'intérieur des canaux. Sans envahir les tissus autour ou sous forme infiltrante, quand la tumeur passe la membrane basale et envahit le parenchyme mammaire [7].

Le **carcinome lobulaire** grandit à partir des lobules mammaires qui font le lait. Il est environ 10 à 15% de tous les cancers du sein. Cette forme est souvent plus difficile à repérer avec les méthodes d'imagerie classiques car il pousse de manière désordonnée et n'a pas de masse clairement définie [7].

D'autres types histologiques plus rares existent, comme les **carcinomes mucineux, tubulaires, médullaires** ou **inflammatoires**. Même s'ils sont moins fréquents, ils ont des traits cliniques et pronostiques spéciaux qui touchent la façon de traiter [7].

I.3.2 Classification selon le caractère invasif

Selon leur capacité à envahir les tissus adjacents, les cancers du sein sont classés en deux grandes catégories [7] :

Il y a les **cancers non invasifs** (in situ). Les cellules cancéreuses restent où elles ont commencé, sans traverser la membrane basale. Le plus fréquent est le carcinome canalaire in situ (CCIS). En général, ces cancers se soignent bien, surtout si on les trouve tôt grâce au dépistage [5] [7].

Ensuite, il y a **cancers invasifs** correspondent à une extension des cellules tumorales au-delà de leur site initial vers les tissus mammaires environnants. Ils peuvent s'étendre aux ganglions lymphatiques régionaux et, à un stade plus avancé, donner naissance à des métastases à distance, ce qui aggrave le pronostic [7].

I.3.3 Classification moléculaire du cancer du sein

La classification moléculaire complète ajoute à l'examen histologique en considérant l'extériorisation de signes vivants comme des récepteurs hormonaux (œstrogènes et progestérone) et le récepteur HER2. Elle fait partie importante dans le rendre spécial du soin [16].

Le sous-type **Luminal A** est caractérisé par une expression positive des récepteurs hormonaux et une absence d'expression de HER2. Il est généralement associé à une faible prolifération cellulaire, un **bon pronostic** et une excellente réponse à l'hormonothérapie.

Le sous-type **Luminal B** montre aussi les récepteurs hormonaux mais a un taux de croissance plus haut et des fois une expression de HER2. Son pronostic est moins bon que celui du Luminal A et il a souvent besoin d'un mélange de soins (hormonothérapie, chimiothérapie et/ou traitements ciblés).

Les cancers qui ont plein de HER2 sont caractérisés par une forte présence du récepteur HER2 avec une petite ou pas d'expression de récepteurs hormonaux. Même si ces cancers sont plus méchants ils peuvent bien répondre aux traitements ciblés qui vont contre le HER2.

Le sous-type **triple négatif**, qui n'a pas de récepteurs ni **HER2** est normalement lié à une grande force et un mauvais regard en raison du manque de cible pour traiter. La solution repose sur la chimio et ces derniers temps sur l'immunothérapies

I.3.4 Stades cliniques du cancer du sein : classification TNM

La stadification clinique du cancer du sein repose sur le système **TNM**, qui permet d'évaluer l'extension locale, régionale et à distance de la maladie [7] [16] :

- **T (Tumor)** : taille et extension de la tumeur primitive
- **N (Nodes)** : atteinte des ganglions lymphatiques régionaux
- **M (Metastasis)** : présence ou absence de métastases à distance.

Sur la base de cette classification, plusieurs **stades cliniques** sont définis :

Stade	Description	Survie à 5 ans
0	Carcinome in situ (DCIS/LCIS)	>99%
I	Tumeur ≤ 2 cm, pas d'atteinte ganglionnaire	95-100%
IIA	Tumeur ≤ 2 cm avec ganglions atteints OU 2-5 cm sans ganglions	85-90%
IIB	Tumeur 2-5 cm avec ganglions OU >5 cm sans ganglions	80-85%
IIIA	Atteinte de 4-9 ganglions axillaires OU extension mammaire interne	65-70%
IIIB	Extension à la paroi thoracique ou à la peau	40-60%
IIIC	Atteinte ≥ 10 ganglions ou sus-claviculaires	30-50%
IV	Métastases à distance (os, foie, poumon, cerveau)	~30%

Tableau 1: les différentes classifications stades pour le cancer du sein [7] [17]

I.3.5 Importance clinique de la classification :

Le regroupement de l'évaluation histologique, du classement moléculaire et du système TNM permet une étude complète des affections malignes du sein. Cette approche est importante pour

- Établir un diagnostic précis [7] [16].
- Éclairer les orientations thérapeutiques (telles que la chirurgie, la radiothérapie, la chimiothérapie, l'hormonothérapie et les thérapies ciblées) [7] [16].
- Évaluer le pronostic [7] [16].
- Assurer un suivi approprié pour les patients [5] [9].

L'imagerie diagnostique, plus précisément l'examen mammographique, l'examen échographique mammaire, et l'IRM, constitue une part importante pour le diagnostic, la

stadification et le suivi du cancer du sein, permettant ainsi à une meilleure approche thérapeutique et à une amélioration des chances de guérison [1] [5] [8].

I.4 Imagerie médicale pour le diagnostic :

L'imagerie médicale a un rôle clé dans la recherche, le test et le suivi du cancer du sein. Elle aide à voir les parties du sein et à trouver des choses bizarres qui ne peuvent pas être senties lors d'un examen médical. Chaque façon a ses bons côtés, ses usages et ses limites, et souvent c'est le mélange plusieurs techniques qui permet d'avoir une idée aussi complète que ce qu'on peut [5].

I.4.1 Mammographie et limites :

L'examen mammographique comparativement à représente l'élément clé du diagnostic pour trouver et diagnostiquer un cancer mammaire, particulièrement pour les femmes à partir de 40 ans. Cette technique produit des images du tissu mammaire avec de rayonnements X et peut trouver des anomalies comme des masses ou des microcalcifications, avant que qu'elles soient palpables. Son application standardisée et son est basée surement adéquat ont grandement contribué à baisser la mortalité provoquée par le cancer du sein dans le cadre du dépistage [6].

La mammographie fonctionne l'utilisation de des rayons qui traversent les structures du sein. Ces rayonnements sont captés de manière distincte en fonction du degré de compaction et la dimension du tissu. En vue d'obtenir une image davantage exacte et diminuer la diffusion des rayonnements, sont utilisés des techniques telles que la compression mammaire. Cela permet contribue à une meilleure identifier les difficultés [6].

Cependant, l'examen mammographique présente des limites majeures. Cette méthode est moins performante chez les femmes avec une forte densité mammaire, car les zones risquent de se chevaucher et masquer les irrégularités, ce qui fait qu'elle est moins exacte. Les études indiquent que, chez ces femmes, l'examen échographique est susceptible d'être plus précise que l'examen mammographique [7] [8].



Figure 2 : Image de machine de mammographie [18].

I.4.2 Échographie mammaire : principe et intérêt clinique et limites :

L'échographie du sein est une méthode d'imagerie médicale basée sur l'émission et la réception d'ultrasons, qui permet de regarder les tissus mous sans besoins de rayons ionisants. Elle est un outil très utilisé dans le diagnostic des maladies du sein, surtout comme soutien à la mammographie [8] [9].

I.4.2.1 Principe physique de l'échographie mammaire :

Le principe physique de l'échographie mammaire repose sur les ondes sonores. Ces sondes traversent la peau et atteignent les tissus du sein. Quand l'onde sonore rencontre une matière différente comme une masse, elle change direction et revient. L'échographe capte ce signal et l'affiche en image. Cet affichage aide à voir les masses superficielles car l'onde est réfléchi par elles de manière évidente et rapide tandis que les masses profondes sont plus difficiles à identifier car l'onde est moins réfléchi par elles. Le principe de fonctionnement est le même pour tous types de tissu dans le sein mais les tissus différents réfléchissent différemment le son ce qui entraîne des images variées. Sur le plan physique, les ultrasons sont des ondes sonores hautes qui avancent dans les tissus vivants et rebondissent un peu quand ils touchent des frontières entre milieux ayant des caractéristiques acoustiques différentes. Les échos créés de cette façon sont pris par la sonde d'écho et traités par le système pour faire une image en temps réel. La forme de l'image dépend beaucoup de l'impédance du son chez les tissus, liée à leur épaisseur et à leur souplesse. Les sons ultras se déplacent bien dans les tissus doux pleins d'eau, mais sont très réfléchis par l'air ou les os, ce qui réduit ce qu'on peut voir au-delà de ces bordures. Les sondes d'écho utilisent des cristaux qui peuvent changer électricité en sons de haut et vice versa. Les sondes à frappe lente ont une vue moins bonne de l'espace pour toucher les choses à la surface, alors que les sondes à mouvement rapide vont mieux en bas, mais pas aussi bien pour voir [9].

I.4.2.2 Intérêt clinique de l'échographie mammaire :

L'échographie mammaire est très utile pour la santé. Elle aide à trouver des problèmes dans les seins, comme des boules ou des changements. Cela peut aider les femmes à détecter des maladies graves tôt. Sur le plan médical, l'échographie du sein est très utile dans plusieurs cas. Elle sert de plus à la mammographie pour bien définir ce qu'est une masse trouvée, en particulier pour distinguer les lésions dures des lésions qui ont de l'eau dedans. Elle est aussi très utile quand les femmes ont une forte densité dans leurs seins, pour qui la mammographie peut être moins efficace [8] [19]. De plus, l'échographie est choisie pour des jeunes patientes, des femmes qui sont enceintes ou qui allaitent, pour éviter toute exposition aux rayons ionisants. Elle fait aussi un grand job dans le guidage des biopsies et des actes qui interviennent [9].

I.4.2.3 Limites de l'échographie mammaire :

Bien que l'échographie mammaire ait plusieurs avantages, elle possède aussi quelques limites. La qualité des images peut être affectée par le bruit, spécialement le bruit (speckle) appelé granulé, qui diminue le contraste et rend difficile la détection des concours tumoraux. De plus certaines lésions ont des bords flous ou des traits de forme ambiguë, ce qui rend leur lecture difficile [9] [19].

L'interprétation des images échographiques reste très liée à l'expérience de l'opérateur, ce qui peut mener à une variété entre observateurs et des fautes dans diagnostic, notamment pour les lésions petites ou étranges [10] [20]. De plus, l'échographie ne détecte pas bien les petites calcifications, limitant son emploi comme façon seule de dépister le cancer du sein [7].

I.4.2.4 Vers l'analyse automatique des images échographiques :

Concernant l'étude automatisée des images échographiques Les limitations autorisent l'utilisation de techniques automatisées pour permettre l'analyse d'échographies. La combinaison des procédés de traitement d'image avec l'intelligence artificielle permet d'augmenter le caractère objectif, la répétabilité et la crédibilité du processus de diagnostic, en secondant les radiologues pour l'identification et à caractériser des anomalies mammaires [17].



Figure 3 : Image de machine de l'échographie [23]

I.4.3 Autres modalités (IRM – aperçu) :

L'imagerie par champ magnétique (IRM) du sein est un test supplémentaire fortement suggéré pour voir la tumeur dans le sein. Elle emploie un champ magnétique et des ondes radio pour faire des images précises du tissu du sein, sans se servir de rayons X [7] [8] [9].

L'IRM est souvent recommandée aux femmes avec un risque plus grand en raison des leurs gènes, comme celles qui ont des mutations BRCA, ou quand les résultats des autres tests ne sont pas clairs [3] [8].

Des études montrent que l'IRM est souvent plus efficace que la mammographie ou l'échographie pour trouver des tumeurs cancéreuses bien qu'elle soit chère, qu'elle est parfois inaccessible ainsi qu'elle est susceptible d'amener des faux positifs, ce qui limite son usage régulier [7] [8] [9].

Des méthodes en plus, comme la vue 3D du sein, la photo du sein mieux vue avec un contraste et la médecine nucléaire (TEP/scan) deviennent plus connues, surtout pour chercher des cas ciblés ou pour voir comment avancé le cancer, mais pour l'instant, elles restent un extra au processus normal [8] [9].

I.5 Motivation pour l'analyse automatique des images échographiques :

Même si l'échographie du sein est importante dans le diagnostic de l'affection tumorale du sein, les médecins radiologues comptent encore considérablement sur l'analyse visuelle de l'imagerie. Les compétences du technicien, ainsi que l'irrégularité de la netteté des images peuvent impacter cette méthode, impliquant alors que l'interprétation des anomalies du sein quelque peu personnelle [9].

Dans un contexte clinique, il est essentiel d'identifier précisément les régions problématiques afin de poser un diagnostic précis. Délimiter les lésions manuellement peut être difficile à cause de la présence de speckle, du contraste tissulaire limité et du contour imprécis des périmètre des masses tumorales. Ces problèmes compliquent l'observation exacte des images et risquent de détériorer l'interprétation des contours des masses tumorales [10] [24].

Par ailleurs, la distinction entre lésions bénignes et malignes représente une décision clinique critique. Dans certains cas, les images échographiques ne sont pas toujours précises. Cela peut entraîner des erreurs de diagnostic ou des examens complémentaires inutiles et chronophages. L'idée de l'automation est de rendre le diagnostic plus juste et d'aider les groupes de radiologie dans leur choix [17] [21].

Dans ce cadre, l'emploi de manières pour regarder des images et d'apprendre profondément pour voir automatiquement les images avec écho se montre comme une bonne réponse. Les CNN, ou bien réseaux neurones qui font glisser sont très fort pour trouver et étudier directement les endroits importants et les formes visuelles utiles ce qui simplifie identification et une présentation des blessures Le fait de se servir des outils aide au diagnostic sert à rehausser la foi, la copie et l'exactitude du diagnostic, tout en baissant de plus la tâche pour le médecin [11] [22] [25].

I.6 Conclusion du Chapitre 1 :

Ce chapitre a aidé à montrer le cancer du sein comme un grand problème de santé publique, aussi bien dans le monde qu'en Algérie. En montrant le taux de cas, l'augmentation des nombres et les principaux facteurs de risque liés.

L'importance du choix et du diagnostic tôt a été mise en lumière car ils sont une partie nécessaire pour rendre meilleur le pronostic et la vie des malades. Nous avons montré les divers types d'imagerie de la santé, avec un point spécial sur l'écho mammaire en soulignant son rôle clé dans l'examen des tissus mammaires et pour déceler des problèmes.

Les limites de cette méthode comme la dépendance à l'expérience de celui qui fait l'examen, un peu de contraste entre les tissus et du bruit expliquent pourquoi on doit utiliser des façons d'étude auto

Dans cette situation, le partage des images aide à trouver exactement les parties malades et d'en sortir des détails sur leur forme et leur brillance, alors que le groupage aide à voir quel genre de blessures ce sont (bonnes ou mauvaises).

L'utilisation de ces deux pas donne un bon coup de pouce au spécialiste des images, aidant à un diagnostic plus sûr, faisable à nouveau et vite. Donc, ce chapitre met en place le contexte médical et technique qu'il faut pour comprendre pourquoi nous faisons notre travail et prépare le terrain pour les méthodes d'ordinateur et les approches d'apprentissage profond qui seront expliquées dans le chapitre.



Chapitre 02 :

Méthodologie et travail réalisé



II.1 Introduction :

Comme évoqué dans le chapitre précédent, quand on diagnostique un cancer du sein avec une échographie, ça dépend encore énormément de l'œil et de l'expérience du radiologue. Le problème, c'est que cette approche est un peu subjective, ce qui peut malheureusement mener à des erreurs graves pour les patientes. Pour qu'il y ait moins d'erreurs et que le dépistage soit plus fiable, on s'est dit qu'il fallait absolument automatiser l'analyse des images. C'est devenu une vraie priorité pour nous.

Et c'est exactement là que le Deep Learning entre en jeu et change tout. Au lieu de lui donner des règles fixes et rigides, on lui montre plein d'exemples, des milliers même, et il apprend tout seul à repérer ce qui compte. Les fameux réseaux de neurones convolutifs, ou CNN, sont devenus super doués pour l'imagerie médicale. Ils peuvent découper les tumeurs pixel par pixel et dire si elles sont gentilles ou méchantes, le tout d'un coup. En clair, on met l'image d'un côté et l'ordinateur nous aide à poser le diagnostic de l'autre, sans qu'on ait besoin de faire des manipulations compliquées entre les deux.

Du coup, ce chapitre va se dérouler en trois grandes parties. Pour commencer, nous avouons voir les bases théoriques : comment ces réseaux de neurones marchent, et quelles sont les différentes structures (comme U-Net et ses variantes) qu'on utilise pour la segmentation et pour classer les choses. On expliquera aussi comment on sait si notre modèle fait bien son travail, grâce à des mesures comme le Dice ou l'Accuracy,

Après, nous allons aborder ce qui a déjà été fait dans le domaine, c'est-à-dire tout ce qu'on sait déjà sur l'échographie du cancer du sein, et où ça coince encore aujourd'hui. Enfin, et c'est ça le cœur du sujet, nous allons expliquer notre propre approche : on parlera du jeu de données BUSI qu'on a choisi, comment on a préparé les images, pourquoi on a opté pour telle ou telle architecture pour la segmentation et la classification, et aussi de tout l'arsenal technique (Python, GPU NVIDIA Tesla T4, etc.) qui nous a aidé à monter ce projet.

Pour faire court, ce chapitre, c'est un peu le lien entre la théorie et ce qu'on a fait concrètement. C'est ici qu'on met en place toutes les bases techniques qui vont expliquer le reste de notre travail.

II.2 Fondamentaux du Deep Learning pour l'Imagerie Médicale :

II.2.1 Réseaux de Neurones Convolutifs (CNN) :

Aujourd'hui, quand on parle d'analyse d'images en Deep Learning, les **Réseaux de Neurones Convolutifs** (CNN) sont devenus l'outil incontournable. Ce qui fait leur force, c'est leur autonomie : contrairement aux anciennes méthodes où un expert devait choisir manuellement quels détails extraire de l'image. Mais un CNN, lui apprend tout seul ce qu'il doit retenir [26]. En regardant énormément d'images, il arrive à repérer seul ce qui compte le plus pour atteindre son objectif.

Pour être plus concret, le réseau s'attaque à l'image en l'explorant petit à petit. Au début, ses premières 'couches' repèrent des choses très simples comme des lignes, des contours ou des textures de base. Ensuite, les couches d'après assemblent ces éléments pour reconnaître des formes plus compliquées. Et à la fin, les couches les plus profondes sont capables d'identifier des structures complètes, par exemple une tumeur dans son environnement. C'est cette façon d'apprendre par niveaux, de manière automatique, qui rend les CNN si efficaces pour classer des choses, les découper en parties ou repérer des objets [27].

Dans notre domaine, celui de l'échographie mammaire, cette faculté est vraiment primordiale. Les images d'échographie sont souvent compliquées, avec des textures difficiles à interpréter, du 'bruit' (des artefacts) et des contours qui ne sont pas toujours nets. En plus, ça change selon l'appareil utilisé. Il serait donc impossible de définir à la main des règles valables partout. Le CNN, par contre, s'adapte. Il va apprendre, grâce à un ensemble d'images déjà 'étiquetées', à faire la différence entre une tumeur qui n'est pas dangereuse et une autre qui l'est, ou encore à bien dessiner les limites d'une lésion. Et tout ça, sans qu'on ait besoin de lui détailler ce qu'est un 'contour hypoéchogénique' [28].

II.2.2 Principe de fonctionnement des CNN :

L'architecture d'un CNN repose sur un enchaînement de couches spécialisées qui transforment progressivement l'image en une décision.

II.2.2.1 Convolution :

C'est l'opération fondamentale qui donne son nom au réseau. Des filtres de petite taille généralement 3×3 ou 5×5 pixels balaient l'image entière. Chaque filtre apprend à détecter une caractéristique visuelle spécifique : un filtre pour les contours verticaux, un autre pour les textures granuleuses, etc. À chaque position, le filtre effectue un produit de convolution et génère une valeur dans ce qu'on appelle une "carte de caractéristiques" (feature map) [29].

Sur nos échographies mammaires, un filtre pourra ainsi apprendre à repérer les zones sombres caractéristiques des kystes, tandis qu'un autre détectera les contours irréguliers des masses malignes. Le résultat : l'image brute se transforme en une série de cartes mettant en évidence les éléments pertinents pour le diagnostic, avec une intervention humaine minimale.

II.2.2.2 Le Pooling (Sous-échantillonnage) :

Après la convolution, les données générées sont souvent très volumineuses. Le Pooling est une méthode de réduction de dimensionnalité qui utilise les cartes de caractéristiques issues de la couche de convolution. Elle consiste à regrouper les données d'une zone locale de l'image pour obtenir une valeur unique représentant cette zone. Cette valeur peut être soit le maximum de la zone regroupée (max pooling), soit sa moyenne (average pooling) [29].

Le Pooling réduit la taille des données (réduction spatiale), la charge de calcul nécessaire pour chaque opération du réseau et le nombre de paramètres du modèle. Il rend également le réseau beaucoup plus robuste aux petits déplacements des objets dans l'image.

II.2.2.3 Fonction d'activation (ReLU) :

La fonction ReLU est ce qui permet au réseau de ne pas se limiter à des calculs mathématiques trop basiques. En transformant les valeurs négatives en zéro, elle crée une sorte de "déclic" qui aide le modèle à capter la complexité réelle d'une échographie (formes, ombres, textures). C'est grâce à cette fonction que l'apprentissage est plus rapide et que le système devient capable de donner un diagnostic fiable, même sur des images qu'il découvre pour la première fois [30].

II.2.2.4 Couches entièrement connectées (Fully Connected) :

Arrivé à ce stade, l'image a été réduite à un ensemble de caractéristiques de haut niveau (présence d'une masse, texture spécifique, etc.). Les couches fully connected prennent le relais : chaque neurone est connecté à tous ceux de la couche précédente. Elles combinent toutes les informations extraites pour établir des relations complexes entre les caractéristiques et les classes de sortie (normal, bénin, malin dans notre cas). Cette couche agit comme le classificateur final du réseau. C'est ici que se prend la décision : la probabilité que cette zone segmentée corresponde à une tumeur maligne est de 87%, par exemple [29].

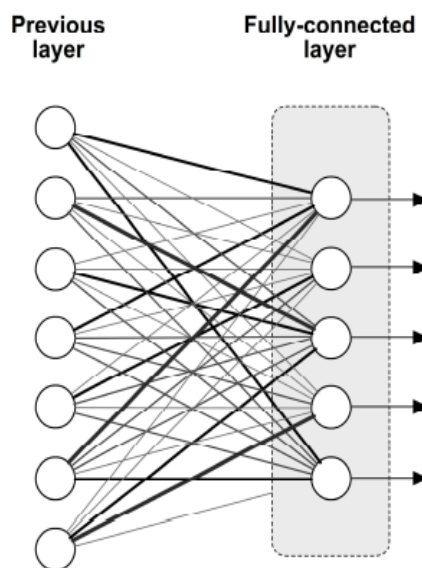


Figure 4 : Schéma illustratif du principe d'une couche entièrement connectée [29].

II.2.2.5 Couche de sortie

La dernière couche dépend de la tâche visée. Pour la classification, on utilise généralement la fonction Soft max qui convertit les scores en probabilités sommant à 1. Pour la segmentation, la couche de sortie est elle-même convolution elle : elle produit une prédiction pour chaque pixel de l'image originale, permettant de délimiter précisément les contours de la lésion [31].

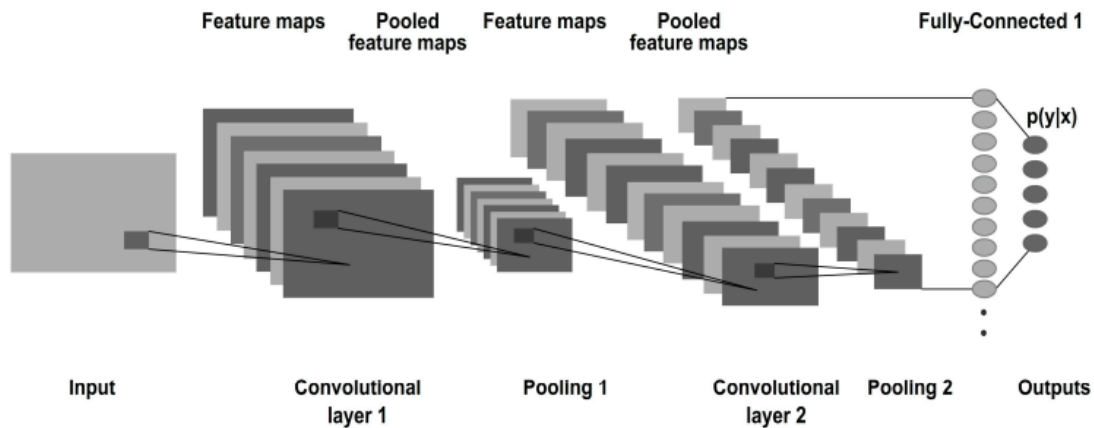


Figure 5 : Architecture générale d'un réseau de neurones convolutif (CNN) [29].

II.2.3 L'Architecture U-Net : Le standard de la segmentation :

Jusqu'ici, on a parlé des CNN classiques qui servent surtout à dire si une image contient une tumeur ou pas (la classification). Mais pour notre projet, on a besoin d'aller plus loin : on veut découper la tumeur pixel par pixel. C'est ce qu'on appelle la segmentation sémantique. Pour ça, l'architecture U-Net créée par Ronneberger et al. En 2015 [32]. Est devenue la référence absolue en médecine.

II.2.3.1 Pourquoi choisir l'U-Net plutôt qu'un CNN standard ?

Un réseau classique comme ResNet réduit la taille de l'image pour finir par donner une seule réponse globale ("tumeur présente"). Pour la segmentation, on doit faire le chemin inverse : partir d'une compréhension globale pour revenir à la taille d'origine en classant chaque pixel ("ce pixel est la tumeur, celui-là c'est du tissu sain"). L'U-Net réussit ce défi grâce à sa forme symétrique en "U" [31].

Une structure en deux étapes :

II.2.3.2 Chemin contractant (Encodeur) :

Suite de convolutions et max-pooling qui réduisent la taille spatiale (de 512×512 à 32×32 par exemple) tout en augmentant la profondeur feature (de 64 à 1024 filtres). Ce chemin capture le contexte global : "il y a une masse hypoechogénique dans le quadrant supérieur droit [32].

II.2.3.3 Le Décodeur (le chemin montant) :

On fait l'inverse en utilisant des convolutions transposées (upsampling) pour agrandir l'image jusqu'à sa taille de départ. À chaque étape, on fusionne les infos grâce aux Skip Connections [32].

II.2.3.4 Le rôle vital des Skip Connections :

C'est le point le plus important. Sans ces connexions, le décodeur n'aura qu'une version très compressée et floue de l'image (le petit carré de 32x32). On perdrait tous les détails des bords. Les skip connections agissent comme des raccourcis qui injectent les détails fins de l'encodeur directement dans le décodeur. Ça permet de dessiner des contours de tumeurs très précis, ce qui est vital pour mesurer le volume de la lésion ou guider un futur traitement.

L'U-Net a l'avantage de très bien fonctionner même sur des petits groupes d'images. Dans l'article d'origine, ils n'avaient que 30 images [32].

II.2.4 Classification des tumeurs du sein par échographie à l'aide de réseaux (CNN) :

L'étape suivante lors de la classification consiste en l'usage de l'architecture U-Net [32] pour aller localiser une région d'intérêt autour de la tumeur lors de la segmentation, étape importante dans les images échographiques du sein avec le bruit de speckle important, le faible contraste et la variabilité morphologique des lésions. À tel point qu'en appliquant diverses architectures à des problèmes de classification d'images en général, et au niveau médical en particulier, il est à noter que les réseaux (CNN), notamment avec l'introduction d'AlexNet [33], réalisent d'excellents résultats. De plus en plus d'études très convaincantes explorent la faisabilité de la technologie CNN pour différencier, sur des images échographiques mammaires, des tissus bénins et malins, en analysant les performances globales du système [34] [35].

Par exemple, Han et ses collaborateurs [34] ont proposé un modèle de classification sur images échographiques de masses mammaires, à partir d'un réseau profond (CNN), permettant d'atteindre une précision supérieure à 90 %. D'autres chercheurs comme Byra [35] ont également conçu des modèles similaires par apprentissage par transfert via des architectures reconnues telles que ResNet [36] et Dense Net [37], et ont atteint des performances identiques, suivant la démonstration que l'adaptation d'un modèle à de petites bases de données permet d'affiner les performances de ce dernier.

Les architectures modernes, telles que ResNet [36] et DenseNet [37], mettent en œuvre une technique qui améliore la propagation du gradient et réduit le surapprentissage, notamment lors de l'apprentissage sur de petites bases de données médicales. De plus, plusieurs études ont exploré l'utilisation d'une approche intégrée combinant segmentation et classification en un flux de travail continu afin d'accroître la robustesse globale du système.

Certaines métriques, telles que la sensibilité, la spécificité, la précision, le score F1 et l'aire sous la courbe ROC (AUC), sont généralement utilisées pour évaluer les performances. De nombreuses études ont démontré une AUC élevée ($> 0,90$) pour la classification des tumeurs

bénignes/malignes à l'échographie mammaire [27] [38], prouvant ainsi l'efficacité des méthodes d'apprentissage profond pour cette tâche.

II.3 Métriques d'Évaluation et Interprétabilité :

Pour savoir si notre système fonctionne vraiment, on ne se contente pas de regarder si "ça marche globalement". On a besoin de métriques précises qui mesurent deux choses : la qualité de la découpe (segmentation) et la fiabilité du diagnostic (classification).

II.3.1 Quantification de la Précision Géométrique (Segmentation) :

Quand U-Net découpe une tumeur, comment on sait si c'est bien fait. On compare le masque prédit avec l'annotation faite par le radiologue (vérité terrain). À l'aide de mesures de similarité spatiale [39].

II.3.1.1 Coefficient de Dice (Dice Similarity Coefficient) :

C'est la métrique principale en segmentation médicale. Elle mesure le chevauchement entre deux zones : celle prédite par l'IA et celle tracée par l'expert. Un score de 1.0 = parfait, 0.0 = rien ne correspond. En pratique, au-dessus de 0,85, on considère que la segmentation est excellente [39].

$$\text{Dice} = \frac{2.VP}{2.VP+FP+FN} \quad (2.1)$$

II.3.1.2 Intersection over Union (IoU) - Indice de Jaccard :

Plus strict que Dice. Au lieu de comparer les surfaces, on compare l'intersection (zone commune) divisée par l'union (zone totale). Pénalise davantage les erreurs de débordement [39].

$$\text{Jaccard} = \frac{VP}{VP+FP+FN} \quad (2.2)$$

II.3.1.3 Distance de Hausdorff :

Utile pour vérifier que les contours sont précis. Mesure l'écart maximal entre les frontières du masque prédit et du masque réel. Important quand on veut mesurer exactement le volume d'une tumeur pour suivre son évolution [39].

$$\text{HD}(P, G) = \max_{p \in P} \{ \min_{g \in G} |p - g| \} \quad (2.3)$$

II.3.2 Analyse de la Performance Diagnostique (Classification) :

Une fois qu'on a découpé la tumeur avec U-Net, il faut décider : bénigne ou maligne. Pour évaluer si le CNN se trompe ou non, on utilise plusieurs indicateurs qui se complètent [40].

II.3.2.1 Matrice de Confusion (Confusion Matrix)

C'est un outil de visualisation essentiel qui présente sous forme de tableau les prédictions du modèle par rapport à la réalité. Elle permet d'identifier précisément les erreurs du modèle [40] :

- **Vrais Positifs (VP) et Vrais Négatifs (VN)** : Représentent les diagnostics corrects.
- **Faux Positifs (FP)** : Cas bénins identifiés à tort comme malins (fausses alertes).
- **Faux Négatifs (FN)** : Cas malins identifiés à tort comme bénins (cancers manqués, les plus critiques en oncologie).

II.3.2.2 Accuracy (Exactitude globale) :

C'est le pourcentage de prédictions correctes sur l'ensemble des cas. Simple, mais trompeur si les classes sont déséquilibrées (plus de bénins que de malins dans notre dataset) [40].

$$\text{Accuracy} = \frac{VP+VN}{VP+VN+FP+FN} \quad (2.4)$$

II.3.2.3 Balanced Accuracy (Précision équilibrée) :

La moyenne arithmétique de la sensibilité et de la spécificité. Elle est indispensable lorsque le dataset est déséquilibré, car elle donne autant d'importance aux classes minoritaires qu'aux classes majoritaires [40].

$$\text{Balanced Accuracy} = \frac{\text{sensibilite} + \text{specificite}}{2} \quad (2.5)$$

II.3.2.4 Sensibilité (Recall / Rappel) :

C'est la métrique la plus importante en médecine. Elle mesure : parmi tous les vrais cas de cancer, combien le modèle a-t-il correctement détectés.

La Sensibilité (Recall) est ici vitale : elle assure que le système ne manque aucun cas cancéreux, ce qui est l'objectif premier du dépistage. À l'inverse, la Précision garantit la fiabilité des alertes pour limiter le surdiagnostic [40] :

$$\text{Rappel} = \frac{VP}{VP+FN} \quad (2.6)$$

En oncologie, on ne peut pas se permettre de rater un cas malin (faux négatif). Une sensibilité élevée = peu de cancers manqués.

II.3.2.5 Précision :

Elle mesure : quand le modèle dit "c'est un cancer", combien de fois il a raison. Évite les fausses alertes qui mènent à des biopsies inutiles et de l'angoisse pour les patientes [40] :

$$\text{Precision} = \frac{VP}{VP+FP} \quad (2.7)$$

II.3.2.6 Spécificité (Specificity) :

Mesure la capacité du modèle à identifier correctement les cas bénins. C'est le complément de la sensibilité : une spécificité élevée signifie que le modèle ne confond pas les tissus sains ou bénins avec des lésions malignes [40]

$$\text{Specificite} = \frac{VN}{VN+FP} \quad (2.8)$$

II.3.2.7 F1-Score :

Moyenne harmonique entre Précision et Sensibilité. Utile quand on veut un compromis entre "ne rien rater" et "ne pas crier au loup" [40].

$$\text{F1}_{\text{Score}} = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}} \quad (2.9)$$

II.3.2.8 F1 Macro et F1 Weighted:

Ces métriques adaptent le calcul du F1-Score pour un contexte multi-classes équilibré [40].

Le F1 Macro calcule la moyenne non pondérée (il traite chaque classe (i) comme égale, idéal pour vérifier la performance sur les classes minoritaires).

$$\text{F1}_{\text{Macro}} = \frac{1}{N} \sum_{i=1}^N \text{F1}_i \quad (2.10)$$

Le F1 Weighted calcule la moyenne pondérée par le nombre d'échantillons n_i (il reflète la performance globale selon la fréquence réelle de chaque classe dans le dataset total N_{total}).

$$\text{F1}_{\text{Weighted}} = \sum_{i=1}^N \frac{n_i}{N_{\text{total}}} \text{F1}_i \quad (2.11)$$

II.3.2.9 Macro Avg et Weighted Avg:

Ce sont les moyennes globales (de la Précision et du Rappel) calculées sur l'ensemble des classes. Leurs formules suivent la même logique que celles du F1 Macro et F1 Weighted appliquées aux autres métriques [40].

II.3.2.10 Indice Kappa (kappa) :

Évalue la fiabilité de la classification en mesurant l'accord entre les prédictions et la vérité terrain, tout en soustrayant l'accord qui pourrait survenir par pur hasard. [40].

$$k = \frac{p_0 - p_e}{1 - p_e} \quad (2.12)$$

II.3.2.11 MCC (Matthews Correlation Coefficient) :

Un coefficient de corrélation robuste qui prend en compte les quatre quadrants de la matrice de confusion (VP, VN, FP, FN). C'est l'une des métriques les plus fiables pour évaluer la qualité des classifications binaires, même en cas de fort déséquilibre des classes [40].

$$MCC = \frac{(VP \times VN) - (FP \times FN)}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \quad (2.13)$$

II.3.2.12 AUC-ROC :

Courbe qui montre la performance du modèle à tous les seuils de décision. Au-dessus de 0,90, c'est considéré comme excellent pour un outil médical [40].

II.3.3 Interprétabilité par Grad-CAM et Attention Rollout :

Un médecin ne peut pas accepter un diagnostic sans comprendre d'où il vient, c'est pour ça qu'on a intégré des outils d'explicabilité pour "ouvrir" nos boîtes noires. Vu qu'on utilise deux types d'architectures CNN et Transformers, on a utilisé deux approches différentes pour la visualisation.

D'abord, pour les modèles classiques comme **EfficientNet-B4** et **ResNet50**, on a appliqué la méthode **Grad-CAM**. Cette technique génère des cartes de chaleur où le code couleur est immédiat pour le praticien : les zones virant au **rouge** indiquent les régions qui ont le plus influencé la décision du modèle (comme des spicules ou une ombre acoustique suspecte), tandis que les zones en **bleu** montrent ce que l'algorithme a jugé non pertinent.

Par contre, pour le **Vision Transformer (ViT)**, le fonctionnement est différent car il n'y a pas de couches de convolution. On a donc choisi l'**Attention Rollout**. Dans cette méthode, la distribution des couleurs montre comment l'attention "circule" entre les différentes parties de l'image. Les zones de forte intensité (**jaune/rouge**) indiquent les pixels pivots qui ont permis au modèle de lier les informations entre elles. Si on voit une forte attention sur les contours d'une masse, ça confirme que le modèle a analysé la forme globale de la lésion plutôt que de regarder un point isolé. En combinant ces deux techniques, on transforme nos algorithmes en vrais partenaires pour le praticien, ce qui renforce la confiance du médecin dans le système [41] [42].

(VP = Vrais Positifs, FP = Faux Positifs, FN = Faux Négatifs)

II.4 ÉTAT DE L'ART :

II.4.1 État de l'Art des travaux de segmentation sur le dataset BUSI :

La segmentation automatique des images échographiques mammaires a fait l'objet de nombreuses recherches depuis la mise à disposition publique de jeux de données annotés. Cette section recense les travaux les plus significatifs ayant utilisé les architectures **U-Net**, **U-Net++** et **Attention U-Net** sur la base de données **BUSI**, permettant ainsi de situer les performances de notre étude par rapport aux standards internationaux.

L'étude fondatrice réalisée par **Al-Dhabyani et al. (2020)** a établi la ligne de base (baseline) en utilisant un **U-Net Standard**. Leur méthodologie reposait sur un apprentissage classique avec l'optimiseur **Adam** et des images redimensionnées à 128 * 128 pixels. Bien que leurs travaux aient démontré la faisabilité de la segmentation automatique, le modèle a montré des limites face au bruit de speckle, affichant un score de **Dice de 0,78** et un **IoU de 0,64** [44].

Dans l'optique d'améliorer ces résultats, **Singh et al. (2021)** ont exploré l'utilisation du **U-Net++**. Leur méthode a introduit une fonction de perte hybride combinant la **Binary Cross-Entropy (BCE)** et le **Dice Loss**, optimisée avec un taux d'apprentissage de 10^{-4} . En intégrant des stratégies d'augmentation de données intensives (rotations, zooms, et flous gaussiens), ils ont atteint un score de **Dice de 0,87** [45]. Cette progression a été confirmée par **Jabeen et al. (2022)**, qui ont exploité les connexions denses imbriquées du **U-Net++** pour réduire l'écart sémantique entre l'encodeur et le décodeur. En utilisant un Learning Rate Scheduler pour ajuster dynamiquement le pas d'apprentissage, ils ont porté le score de **Dice à 0,89** [46].

L'intégration des mécanismes d'attention a marqué un tournant majeur. **Sulaiman et al. (2023)** ont proposé un modèle **Attention U-Net** doté de quatre portes d'attention. Leur protocole expérimental utilisait la fonction de perte **Tversky Loss**, particulièrement adaptée au déséquilibre de classes, et une taille de lot (batch size) de 16. Cette approche a permis de filtrer efficacement les informations non pertinentes et de reconstruire les bordures tumorales avec une grande fidélité, atteignant un score de **Dice de 0,92** et une **Précision de 0,97** [47].

La même année, **Kumar et al. (2023)** ont réalisé un étalonnage complet des variantes du **U-Net** sur 50 itérations. Leur méthodologie de comparaison stricte, utilisant une résolution d'entrée de 224 * 224 et une décroissance pondérée des poids (weight decay), a confirmé la supériorité de l'**Attention U-Net (Dice 0,89)** sur le **U-Net++ (Dice 0,88)** et le **U-Net simple (Dice 0,79)** [48]. Parallèlement, **Sharma et al. (2023)** ont démontré l'intérêt du transfert d'apprentissage en utilisant un backbone **ResNet34** préapprentissage sur ImageNet. Leur méthode de "fine-tuning" avec l'optimiseur **SGD** et une impulsion de 0,9 a permis d'accélérer la convergence tout en obtenant un score de **Dice de 0,84** sur une architecture standard [49].

Enfin, les travaux les plus récents continuent d'affiner ces approches. **Zhang et al. (2024)** ont introduit un réseau d'attention multi-échelle capable de détecter les petites masses souvent ignorées, atteignant un **Dice de 0,90 [50]**. Plus récemment encore, la pré-publication de **Ali et al. (2025)** sur les réseaux hybrides à portes d'attention combinés au **Transfer Learning** rapporte des performances de l'ordre de **0,91** en score de **Dice [51]**.

II.4.2 Analyse des travaux antérieurs sur la classification du dataset BUSI

Le diagnostic automatisé des pathologies mammaires par classification d'images échographiques a connu une évolution technologique majeure ces dernières années. Afin de contextualiser nos travaux, nous avons analysé sept études de référence utilisant des architectures similaires sur le jeu de données BUSI.

L'étude séminale menée par **Al-Dhabyani et al. (2020)** lors de l'introduction du dataset a posé les bases de la classification en trois classes (Normal, Bénin et Malin). En utilisant une architecture **CNN standard** apprise avec des augmentations de données basiques telles que les rotations et les inversions (flips), les auteurs ont établi une ligne de base avec une exactitude de **77,8% [44]**.

Par la suite, l'utilisation de réseaux résiduels profonds est devenue une norme. **Moon et al. (2022)** ont exploré le potentiel du **ResNet-50** en d'apprentissage (Transfer Learning) via des poids préalablement optimisés sur ImageNet. Leur méthodologie visait à exploiter les couches résiduelles pour stabiliser l'extraction des caractéristiques malgré la profondeur du réseau. En utilisant des images de $224 * 224$ pixels et l'optimiseur Adam, ils ont obtenu une précision de **75,4% [52]**. Dans la même lignée, **Abubakar et al. (2023)** ont testé une approche multi-modèles combinant **ResNet-50** et VGG16. En intégrant des techniques d'augmentation "online" via la bibliothèque Albumentations, ils ont rapporté une exactitude de **74,2%** pour le modèle ResNet, soulignant que la profondeur seule ne suffit pas à compenser la complexité des textures échographiques [53].

L'optimisation des architectures convolutionnelles a franchi une étape supplémentaire avec la famille EfficientNet. **Houssein et al. (2023)**, ont mis en œuvre un modèle **EfficientNet-B4** basé sur la méthode du Compound Scaling. Cette approche équilibre la profondeur, la largeur et la résolution de l'image. En appliquant une normalisation Z-score et l'optimiseur RMSprop, ils ont atteint une exactitude de **87,5% [54]**. Ce gain de performance a été confirmé par **Raza et al. (2024)** dont l'étude comparative a désigné l'EfficientNet-B4 comme le point d'équilibre optimal pour l'imagerie médicale. Grâce à une résolution d'entrée de $320 * 320$ et l'usage du Global Average Pooling, ils ont obtenu un score de **88,2% [55]**. Plus récemment, l'introduction des modèles basés sur l'attention a révolutionné les performances. **Gheflati et al. (2024)** ont démontré l'efficacité du **Vision Transformer (ViT-B/16)**. En découpant les images en patches de $16 * 16$ pixels, leur modèle a pu capturer des relations spatiales globales inaccessibles aux convolutions classiques. En utilisant l'optimiseur AdamW et 12 têtes d'attention, ils ont atteint une exactitude remarquable de **92,1% [56]**.

Enfin, les travaux de **Khan et al. (2025)**, ont exploré une voie **hybride combinant ViT et CNN**. Cette méthodologie vise à fusionner les détails fins captés par les convolutions et les

structures globales saisies par le Transformer. Grâce à une gestion du taux d'apprentissage par Cosine Annealing, ce modèle a atteint une précision de 91,5% [57].

II.5 CONTEXTE ET DONNÉES :

II.5.1 Description de la démarche globale :

Avant de détailler chaque étape de notre travail, il est important de présenter une vue d'ensemble de la méthodologie qu'on a suivie. Le schéma bloc ci-dessous illustre l'architecture globale de notre système de diagnostic assisté par ordinateur, montrant le cheminement des données depuis l'image brute jusqu'à la classification finale.

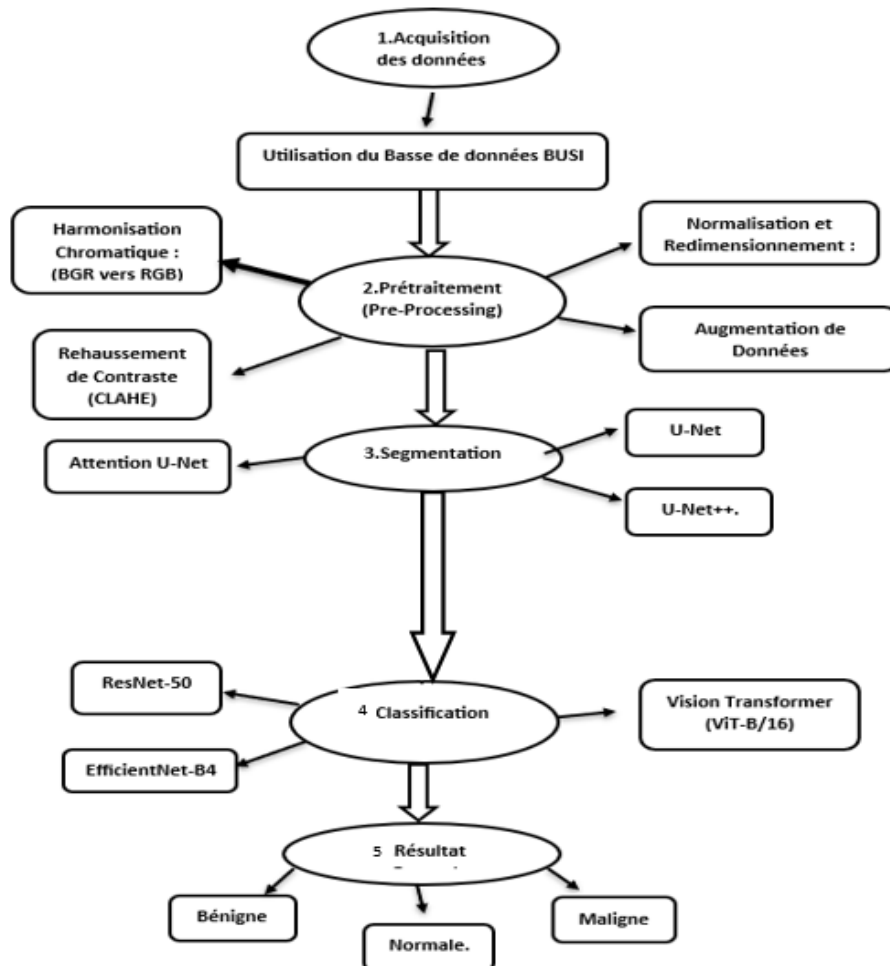


Figure 6: Architecture globale de la chaîne de traitement proposée.

Comme on peut le voir sur la figure 6, notre démarche est structurée de manière séquentielle pour garantir un traitement efficace des échographies. Tout commence par l'acquisition des données via la base BUSI, suivie d'une phase de prétraitement indispensable pour harmoniser les images et améliorer leur qualité visuelle. Ensuite, nous passons par une étape de segmentation pour délimiter la zone tumorale. Une fois cette étape franchie, les informations sont envoyées vers le bloc de classification où nos modèles de Deep Learning (CNN et Transformers) analysent les caractéristiques pour rendre un verdict. Enfin, le système produit un résultat clair (Bénigne, Maligne ou Normale) tout en intégrant des outils d'interprétabilité pour valider le diagnostic. Cette organisation permet de transformer une simple image médicale en une aide à la décision fiable pour le praticien.

II.5.2 Environnement de développement et ressources techniques :

Pour faire tourner tout ça, on s'est équipé d'une stack technique complète. Côté software, on code en **Python** avec le framework **PyTorch**. Pour la segmentation, on utilise la bibliothèque **SMP** (Segmentation Models PyTorch) qui propose des U-Net++ déjà prêts. Pour la classification, **TIMM** nous donne accès à ResNet-50, EfficientNet-B4 et même des Vision Transformers [48]. On optimise avec **AdamW** et on suit les métriques via **Torchmetrics**.

Pour le prétraitement, **OpenCV** gère les images et **Albumentations** s'occupe de l'augmentation. Côté analyse, on a le trio classique **NumPy/Pandas/Scikit-learn**, plus **Grad-CAM** pour visualiser ce que le réseau regarde.

Matériellement, tout tourne sur **Kaggle** avec un **double GPU NVIDIA Tesla T4**. Essentiels pour réduire le temps d'apprentissage des réseaux profonds de plusieurs jours à quelques heures [49].

II.6 Description de la Base de Données (Dataset BUSI) :

Pour assurer l'apprentissage de notre système, on a besoin d'images réelles annotées par des experts. On a choisi le dataset BUSI (Breast Ultrasound Images) qu'on a récupéré sur Kaggle. C'est devenu un standard dans la recherche médicale pour tester les algorithmes de détection et de segmentation sur échographie mammaire [43].

II.6.1 Présentation de la base de données :

Le corpus BUSI contient 780 échographies mammaires issues de cas cliniques réels

Concernant des patientes de 25 à 75 ans. Ce sont des images au format PNG d'environ 500 × 500 pixels.

Le point fort de ce dataset, c'est que chaque image vient avec des **masques** tracés à la main par des radiologues experts. Ces tracés (qu'on appelle "Ground Truth") montrent exactement où se trouve la tumeur. Sans ça, impossible d'apprendre à un algorithme ce qu'il doit détecter [43].

II.6.2 Structure et organisation des données :

Classe	Nombre d'images	Caractéristiques
Bénigne	437	Lésions non cancéreuses, contours souvent réguliers
Maligne	210	Tumeurs cancéreuses, formes plus complexes et infiltrantes
Normale	133	Tissus sains sans lésion, pour apprendre au système ce qu'est "du bruit normal"

Tableau 2: Répartition du dataset BUS.

On remarque tout de suite le problème : 437 bénins contre seulement 210 malins. C'est presque deux fois plus. Et les images normales sont encore moins nombreuses (133). Ce déséquilibre est un vrai défi technique si on ne fait pas attention, l'algorithme va apprendre à tout classer comme "bénin" pour avoir raison le plus souvent possible. On verra au Chapitre 3 comment on compense ce biais [58].

II.6.3 L'importance des masques de segmentation :

L'exploitation des masques de segmentation du dataset BUSI, ou « Ground Truth », constitue le cœur de notre méthodologie. Ces tracés réalisés par des experts nous permettent d'assurer la fiabilité du système à trois niveaux stratégiques. Tout d'abord, ils servent de référence pour l'apprentissage, permettant à nos modèles, comme le U-Net et ses variantes, d'apprendre à isoler les pathologies de manière autonome. Ils constituent ensuite l'étalon-or indispensable pour évaluer la précision de nos segmentations grâce à des indices de performance comme le score de Dice ou l'IoU. Enfin, ces masques sont essentiels pour le ciblage, car ils permettent d'extraire avec précision la région d'intérêt (ROI) afin d'isoler la lésion et d'optimiser l'étape finale de classification. Il est également important de noter que l'absence de masque sur les clichés sains représente en soi une donnée clinique capitale qui confirme l'intégrité des tissus. Cette exploitation rigoureuse des annotations garantit ainsi une transition cohérente entre la détection visuelle et l'aide au diagnostic [59].

II.6.4 Pourquoi avoir choisi le dataset BUSI ?

Le choix du dataset BUSI s'est imposé comme une évidence pour plusieurs raisons stratégiques. Tout d'abord, l'échographie est une modalité d'imagerie particulièrement complexe : elle est naturellement "bruitée" par le phénomène de speckle (ce grain typique des ultrasons) et souffre souvent d'un faible contraste entre la lésion et les tissus sains environnants.

Utiliser BUSI, c'est se confronter à ces difficultés réelles rencontrées par les radiologues au quotidien [60].

De plus, la précision des annotations (les masques) est un argument majeur. Pour la mise au point de l'apprentissage de modèles sophistiqués basés sur l'architecture U-Net, nous avons besoin d'une "vérité terrain" indiscutable. La qualité des tracés experts dans ce dataset permet à nos modèles d'apprendre à distinguer les nuances les plus subtiles des contours tumoraux. En résumé, cette base de données offre le parfait équilibre entre la complexité des images cliniques et la rigueur scientifique nécessaire à l'apprentissage profond.

II.7 Prétraitement et Augmentation des Données :

Les images du dataset BUSI arrivent brutes : différentes tailles, différents contrastes, et surtout ce bruit de speckle qui rend tout flou. Avant de les donner à manger au réseau de neurones, il faut les préparer. Voici la chaîne de traitement qu'on applique à chaque image.

II.7.1 Harmonisation et redimensionnement :

Toutes les images ne sortent pas de la même machine. Certaines sont plus claires, d'autres plus sombres, certaines plus grandes ou plus petites. On commence par mettre tout le monde au même niveau.

- **Conversion RGB** : Pour maintenir la compatibilité avec l'apprentissage par transfert, les échographies en niveaux de gris sont encodées sur trois canaux (RGB). Cela permet l'utilisation directe de backbones pré-configurées, tels que le ResNet34 ou l'EfficientNet-B4.
- **Redimensionnement** : On ramène tout à pixels 224×224 avec de l'interpolation bilinéaire. Ça lisse les pixels quand on compresse, pour ne pas perdre les contours fins des tumeurs.

II.7.2 Augmentation de données (Offline) :

Vu qu'on avait seulement 780 images au départ, ce n'était pas suffisant pour apprentissage des modèles profonds sans risquer l'overfitting [61]. Du coup, on a mis en place une stratégie d'augmentation de données "offline". En utilisant la librairie Albumentations [62], on a créé des variations artificielles mais qui reste réaliste d'un point de vue clinique pour enrichir le jeu d'apprentissage. On a appliqué des rotations de $\pm 15^\circ$ pour simuler l'inclinaison de la sonde, des déformations élastiques pour imiter la pression sur le sein, et on a ajouté du bruit gaussien pour reproduire le "speckle" typique des échographies. On a aussi utilisé des inversions horizontales (flips) pour gérer la symétrie. Cette méthode nous a permis de multiplier par 5 le volume de données, ce qui nous fait un total de 3 270 images pour l'apprentissage. Le fait de le faire en "offline" permet de ne pas ralentir les calculs sur le GPU NVIDIA Tesla T4 pendant que le modèle apprend. C'est une étape vraiment cruciale pour stabiliser l'apprentissage des modèles comme EfficientNet-B4 ou ResNet50, et ça aide beaucoup le réseau à mieux généralisé sur de nouveaux cas.

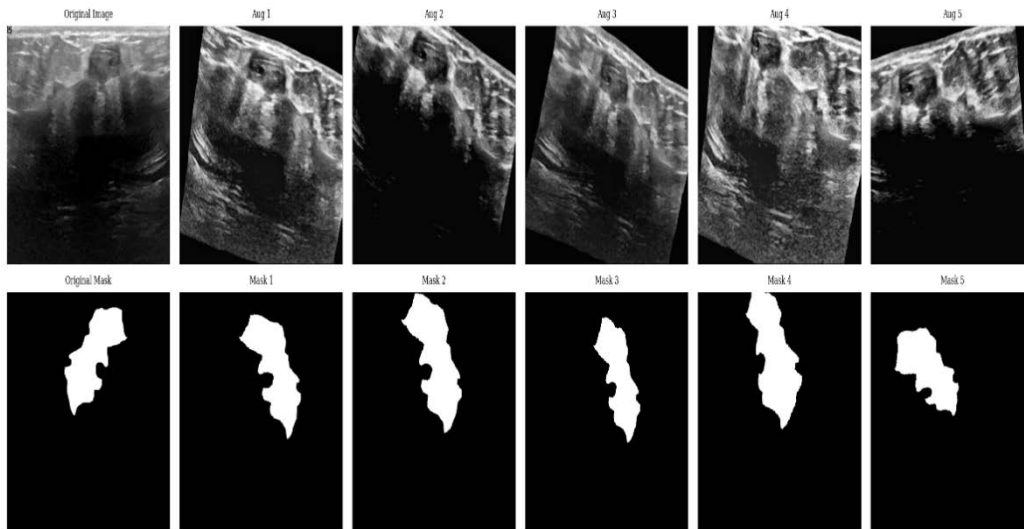


Figure 7 : Processus d'augmentation de données appliqué aux images échographiques et à leurs masques associés.

II.7.3 CLAHE (le contraste adaptatif) :

C'est l'étape la plus importante pour les échographies. Les tumeurs, surtout les malignes, sont souvent sombres et se fondent dans le tissu autour. Le **CLAHE** (Contrast Limited Adaptive Histogram Equalization) éclaircit l'image localement, zone par zone, sans créer des artefacts.

Pourquoi c'est crucial. Parce que si les contours de la tumeur sont trop flous, l'U-Net ne saura pas où découper. Le CLAHE révèle ces contours sans amplifier le bruit de fond [63].

II.7.4 Normalisation pour le réseau :

Avant d'entrer dans le réseau, les pixels doivent être dans un format précis. Nous appliquons une normalisation Z-Score calée sur les poids d'ImageNet [33].

- Nous centrons les valeurs autour de 0 (moyenne = 0)
- Nous réduisons l'écart-type à 1

Cette étape stabilise l'apprentissage et accélère la convergence.

II.7.5 Format PyTorch et GPU :

Enfin, on convertit les données en **tenseurs PyTorch** (le format que comprend le réseau) et on les envoie sur le **GPU NVIDIA Tesla T4** (double GPU pour accélérer). Cette puissance de calcul nous permet d'apprentissage nos architectures complexes de la segmentation par U-Net jusqu'à la classification finale en un temps raisonnable.

II.8 Architecture U-Net pour la Segmentation :

Pour cette partie de notre travail, le but premier c'était de trouver le modèle qui arrive à bien découper (segmenter) les tumeurs, même quand l'image est de mauvaise qualité. Comme les échographies c'est souvent flou avec énormément de bruit, on n'a pas voulu se contenter d'un seul algorithme. On a donc choisi de tester et de comparer trois versions du U-Net pour voir laquelle tournait le mieux sur nos données.

Le Backbone (ResNet34) : Avant de lancer les gros tests, on a décidé de pas partir de zéro pour la partie encodeur. On a utilisé un **ResNet34** déjà pré- **configurées**, (c'est ce qu'on appelle le Transfer Learning). Dans nos essais, on a bien vu que ça aidait le modèle à reconnaître les formes et les bords dès le début des itérations. En plus, ça nous a fait gagner un temps fou sur l'apprentissage parce que le réseau avait déjà des "bases" en vision.

II.8.1 U-Net classique :

C'est le tout premier modèle qu'on a codé ensemble. C'est la base en imagerie médicale avec sa forme en "U". Ce qui est fort avec ce réseau, c'est les skip connections : ça permet de récupérer les petits détails des contours qui disparaissent quand l'image est compressé par l'encodeur. On l'a gardé comme "modèle de référence" pour comparer si les autres versions plus complexe valaient vraiment le coup sur le dataset BUSI.

II.8.2 Attention U-Net :

Après, on a testé l'Attention U-Net. Ce qui est intéressant dedans, c'est qu'il ajoute des "portes d'attention" (Attention Gates). Pendant nos tests, on a remarqué que ça aidait beaucoup le réseau à ne pas s'embrouiller avec le bruit de fond (le speckle noise). Au lieu de regarder partout, le modèle apprend à ignorer les tissus sains pour se concentrer uniquement sur la zone de la tumeur. C'est super efficace quand l'image est très bruitée ou quand on ne voit presque pas la différence entre la tumeur et le reste [64].

II.8.3 U-Net++ :

Pour finir, on a voulu pousser le truc plus loin avec le U-Net++. Là, c'est plus des connexions simples, c'est des blocs de calculs qui s'entrecroisent partout entre l'encodeur et le décodeur. C'est vrai que c'est un modèle plus lourd et plus long à faire tourner, mais c'est aussi le plus précis pour capturer les petits détails. On l'a choisi parce qu'il est très bon pour suivre les bords des tumeurs qui ont des formes bizarres ou des contours très irrégulier, là où les modèles simples font souvent des erreurs de découpe [65].

- **Comparaison de la Complexité des Modèles :**

Le tableau suivant récapitule le nombre de paramètres pour chaque architecture, mettant en évidence l'impact des modules d'attention et des connexions denses sur la taille des réseaux :

Architecture	Encodeur (Backbone)	Nombre de Paramètres	Caractéristique Structurale
U-Net	ResNet34	24.43 Millions	Structure symétrique standard.
Attention U-Net	ResNet34	24.81 Millions	Intégration des portes d'attention.
U-Net++	ResNet34	26.08Millions	Connectivité dense et blocs imbriqués.

Tableau 3 : Comparaison des modèles de segmentation.

U-Net++ est un peu plus lourd (plus de paramètres), mais elle gagne en précision sur les contours complexes. Attention U-Net est légèrement plus lourde que la classique mais elle réduit les erreurs de segmentation sur les zones bruitées.

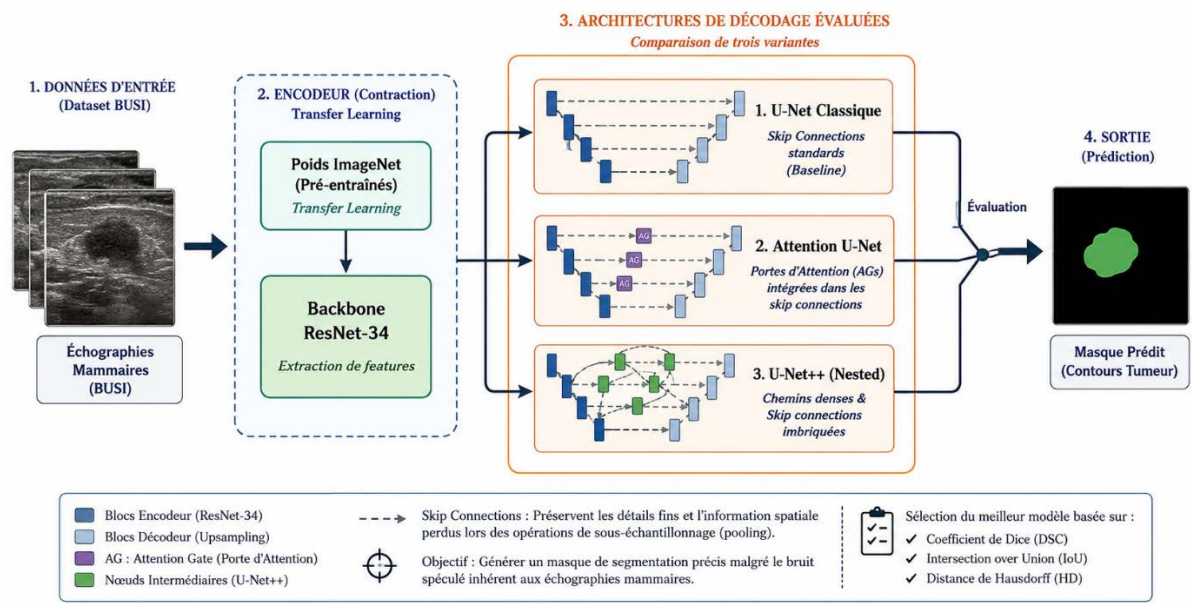


Figure 8: Architecture du pipeline de segmentation d'images échographiques mammaires

Cette figure présente notre approche comparative pour la segmentation automatisée des tumeurs mammaires (dataset BUSI). En utilisant un ResNet-34 comme backbone, nous avons évalué et comparé les performances de trois architectures de décodage (U-Net, Attention U-Net et U-Net++). Cette étude visait à optimiser la précision spatiale des masques prédits, validée rigoureusement par les métriques Dice, IoU et Hausdorff, afin de déterminer la solution la plus fiable pour le diagnostic clinique.

II.9 Architecture pour la Classification :

Une fois qu'on a réussi à bien découper la tumeur avec notre U-Net, l'étape suivante c'est de décider si elle est bénigne ou maligne. Pour pas se tromper, on n'a pas voulu parier sur un seul modèle. On a donc testé trois approches qui travaillent différemment pour être sûr de notre coup.

II.9.1 ResNet-50 :

Le ResNet-50 a été sélectionné comme modèle de référence pour nos premiers tests. Sa force réside dans ses connexions résiduelles (skip connections), qui permettent de mener l'apprentissage de 50 couches profondes en évitant le problème de la disparition du gradient. Afin d'optimiser le processus, nous avons appliqué le transfert d'apprentissage à partir des poids issus d'ImageNet. Ainsi, le modèle dispose déjà d'une capacité de reconnaissance des formes fondamentales, ne nécessitant qu'un ajustement final pour la détection des tumeurs mammaires. En sortie, l'architecture génère un vecteur de 2048 caractéristiques, offrant une base riche pour distinguer efficacement les lésions bénignes des cas malins.

II.9.2 EfficientNet-B4 :

Celui-là, on l'aime bien parce qu'il est plus léger que le ResNet mais il est super puissant. Il utilise des blocs spéciaux (MBConv) et il équilibre tout seul la profondeur et la résolution de l'image [66]. Pour le faire tourner au top, on a mis en place un protocole spécial en deux phases :

- **Le Warm-up** : Cette étape consiste à geler le corps du réseau pour concentrer l'apprentissage uniquement sur la tête de classification. Cela permet de stabiliser les calculs avant de passer à l'optimisation globale du modèle.
- **Le Fine-tuning** : Après, on débloque tout et on affine avec un algorithme (AdamW) [70], et une méthode qui fait varier la vitesse d'apprentissage en "vague" (le Cosine Annealing) pour mieux converger. On a aussi ajouté du **TTA (Test Time Augmentation)** [71] : pour chaque image de test, le modèle la regarde sous 7 angles différents et il fait la moyenne. Ça permet de réduire les erreurs de jugement.

II.9.3 Vision Transformer (ViT-B/16) :

Contrairement aux autres qui regardent l'image petit morceau par petit morceau, le ViT découpe l'image en "patches" de 16x16 pixels. Il utilise un système d'auto-attention, un peu comme ce qui se fait pour les textes. Ça lui permet d'avoir une vision globale de la tumeur. C'est le modèle le plus lourd qu'on a utilisé avec 86 millions de paramètres, mais il est super utile parce qu'il voit des formes d'ensemble que les réseaux classiques pourraient rater. C'est le complément parfait pour nos autres modèles [67].

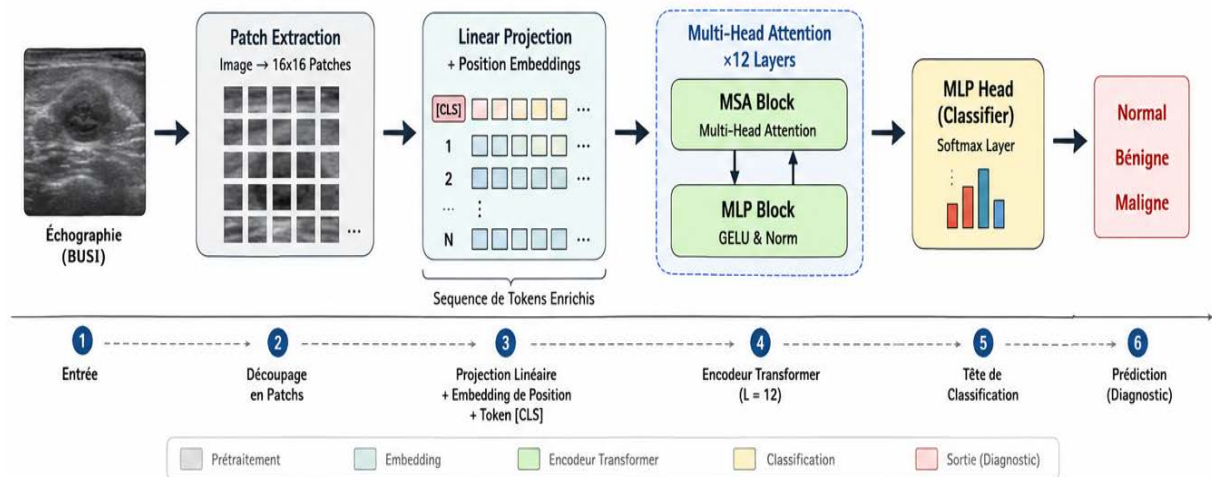


Figure 9: Architecture détaillée du flux de traitement par Vision Transformer (ViT-B/16).

Le pipeline du ViT-B/16 classe les échographies mammaires de manière globale en découpant d'abord l'image en patches de 16*16 pixels, projetés sous forme d'embeddings linéaires. Après l'injection de coordonnées spatiales (Position Embeddings) pour préserver la topologie de la lésion et l'intégration d'un jeton virtuel [CLS] en tête de séquence, les données traversent un encodeur de 12 couches combinant auto-attention multicéphale (MSA) et blocs MLP. En fin de processus, l'état enrichi du Token [CLS] est extrait puis transmis à la tête de classification (MLP Head) qui, grâce à une couche Softmax, détermine la nature exacte de la pathologie parmi trois catégories : tissu normal, lésion bénigne ou tumeur maligne.

Caractéristique	ResNet-50	EfficientNet-B4	ViT-B/16
Type d'Architecture	CNN Résiduel	CNN Optimisé (MBConv)	Vision Transformer
Innovation Majeure	Connexions de saut (Skip connections)	Mise à l'échelle composée (Compound Scaling)	Auto-attention globale (Self-Attention)
Nombre de Paramètres	~25,6 Millions	~19,3 Millions (Efficace)	~86 Millions (Le plus lourd)
Dimension de Sortie	2048 (Features)	1792(Features)	768(Embedding)
Taille d'Entrée (Optimale)	256 x 256	320x 320	224 x 224
Taille des Patches	N/A (Convolutionnel)	N/A (Convolutionnel)	16 x 16 pixels
Point Fort Médical	Analyse robuste des textures locales	Précision des contours flous	Relations spatiales distantes
Philosophie	Profondeur et stabilité	Efficacité et résolution	Contexte et attention globale

Tableau 4 : Comparaison des modèles de classification

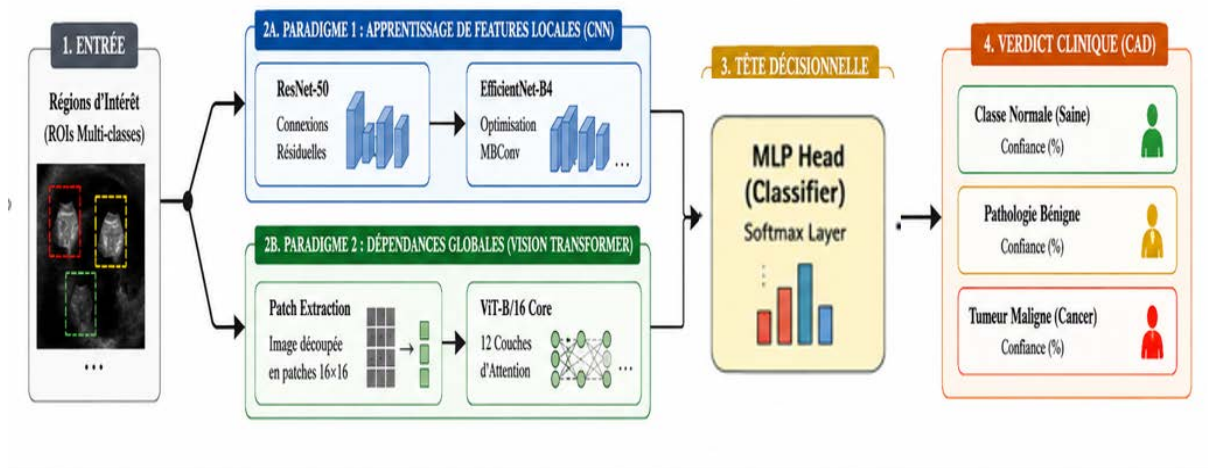


Figure 10: Pipeline méthodologique : Comparaison des paradigmes CNN et Vision Transformer

Cette figure schématise l'architecture de notre approche comparative, conçue pour évaluer les performances de trois modèles de référence : **ResNet-50**, **EfficientNet-B4** et **ViT-B/16**. Notre méthodologie repose sur une analyse comparative rigoureuse, où chaque modèle qu'il soit basé sur les textures locales des CNN ou sur la vision globale des Transformers est soumis à une tête décisionnelle identique afin d'extraire un verdict clinique fiable. Cette architecture de synthèse reflète notre démarche : confronter différents paradigmes d'apprentissage pour identifier la solution la plus performante dans la classification des échographies mammaires (Normal, Bénin ou Malin).

II.10 Stratégie d'apprentissage :

II.10.1 Découpage des données et Validation :

Le découpage (70% d'apprentissage, 15% Validation, 15% Test) a été notre base de départ, on a compris que cette méthode ne pouvait pas être appliquée de façon rigide à tous les modèles. Chaque architecture a ses propres besoins :

- **Pour les CNN (ResNet, EfficientNet) :** On n'a pas suivi le découpage bêtement. Même s'ils ont des "biais inductifs" qui les aident à trouver les formes, on a dû ajuster selon comment ils convergeaient. Si le modèle n'apprenait pas assez bien les détails complexes des tumeurs, on a dû revoir la répartition pour lui donner plus d'exemples difficiles pendant l'apprentissage.
- **Pour le Vision Transformer (ViT) :** C'est là qu'on a dû être plus vigilants. Comme le ViT n'a pas de préjugés sur la forme des objets, il a tendance à faire beaucoup plus d'overfitting. On a dû surveiller la courbe de Validation de très près et ajuster la durée de l'apprentissage. C'est pour lui qu'on a surtout utilisé la Validation Croisée (5-Fold) pour être sûrs que le modèle n'apprenait pas juste un morceau spécifique du dataset.
- **Validation croisée à 5 folds (5-Fold Cross-Validation) :** Pour garantir que la performance de notre système est réelle et qu'elle ne dépend pas d'un découpage

chanceux des données, on a mis en place une validation croisée. Au lieu de faire un apprentissage unique, on divise le dataset global en 5 parties (folds). Le processus se répète cinq fois : à chaque tour, une partie différente est utilisée pour le test tandis que les quatre autres servent à l'apprentissage. Cette méthode est beaucoup plus rigoureuse car elle force le modèle à prouver sa robustesse sur l'intégralité du dataset. Si le modèle reste performant sur tous les folds, on peut affirmer qu'il est capable de bien généraliser face à de nouvelles images médicales qu'il n'a jamais rencontrées [68].

- **Adaptation dynamique** : En gros, pour chaque modèle, la partie "Validation" c'était notre juge de paix. Si on voyait que la performance stagnait sur la validation mais grimpait en apprentissage, on arrêta tout (**Early Stopping**) pour pas tricher. On a adapté le nombre d'images pour que le test final à 15% soit vraiment représentatif de la réalité.

II.10.2 Techniques d'optimisation :

II.10.2.1 Transfer Learning :

Comme on travaille sur le dataset BUSI qui n'a que 780 images, Ce volume de données est insuffisant pour mener un apprentissage à partir de zéro sans qu'il fasse du sur-apprentissage total. Donc on utilise le **Transfer Learning** [72]. On part de réseaux (ResNet, EfficientNet, ViT) qui sont déjà pré- configurés sur **ImageNet** [73], une base de données géante de 14 millions d'images.

L'idée c'est que ces modèles savent déjà détecter les bases comme les contours, les textures et les formes géométriques. Notre travail consiste alors à effectuer un ajustement fin (fine-tuning) des dernières couches, afin que le modèle adapte son apprentissage aux caractéristiques spécifiques des images médicales. Comme l'ombre acoustique ou les bords irréguliers d'un nodule. Ça nous permet de gagner un temps énorme et d'avoir une précision très haute même avec peu de données, car on ne réinvente pas la roue pour la détection visuelle de base.

II.10.2.2 AdamW et Gestion du Learning Rate:

L'optimiseur est l'élément qui permet au modèle de corriger ses erreurs au fil de l'apprentissage. Nous avons porté notre choix sur l'optimiseur **AdamW** [70], plutôt que sur la version classique d'Adam. La différence, c'est que l'AdamW intègre mieux le "weight decay" (décroissance des poids), ce qui empêche les poids du réseau de devenir trop grands et d'exploser. Ça aide énormément à la généralisation pour pas que le modèle apprenne par cœur.

Pour compléter ça, nous avons mis en place un planificateur de type Cosine Annealing (Cosine Annealing Scheduler). Au lieu de garder une vitesse d'apprentissage (Learning Rate) fixe, on commence avec une valeur haute pour explorer les données, puis on la diminue suivant une courbe de cosinus. Ça permet au modèle de ralentir doucement à la fin pour se stabiliser sur le point le plus précis possible. C'est grâce à ce réglage précis qu'on a évité les instabilités pendant les 50 itérations d'apprentissage.

II.10.2.3 Focal Loss et WeightedRandomSampler :

Le plus gros souci avec le dataset BUSI, c'est le déséquilibre entre les classes (on a beaucoup plus de cas bénins que de malins). Si on laisse le modèle apprendre comme ça, il va finir par tout classer en "bénin" pour avoir un bon score facilement. Pour régler ça, on a combiné deux techniques :

- **WeightedRandomSampler** : C'est une méthode d'échantillonnage. Au lieu de donner les images au hasard, on donne un "poids" plus fort aux images de tumeurs malignes. Comme ça, pendant l'apprentissage, le modèle voit plus souvent les exemples rares (les malins). Ça force le réseau à bien apprendre les caractéristiques du cancer même si on n'a pas beaucoup d'images.
- **Focal Loss** : C'est une fonction de perte (Loss function) plus intelligente que la classique Cross-Entropy. La Focal Loss réduit l'importance des exemples "faciles" que le modèle connaît déjà bien. Elle se concentre surtout sur les exemples "difficiles" où l'IA fait de grosses erreurs. Dans notre cas, ça aide le modèle à ne pas ignorer les petits détails des tumeurs malignes qui ressemblent parfois aux bénignes [69].

En utilisant ces deux outils ensemble, on a réussi à avoir une sensibilité beaucoup plus haute sur les cas de cancers, ce qui est le plus important pour un diagnostic médical.

II.10.2.4 Accumulation de Gradients :

Pour résumer un peu notre partie technique, on n'a pas appliqué la même config à tous les modèles parce que leurs besoins varient pas mal. Pour **EfficientNet-B4**, on a surtout privilégié la résolution en restant sur du **320x320**, le but c'était de garder un maximum de détails cliniques sur les images de tumeurs. Comme c'est gourmand en mémoire, on a utilisé l'**accumulation de gradients (sur 4 étapes)**. Ça nous a permis de stabiliser l'apprentissage tout en compensant la petite taille du "batch" physique imposé par le GPU.

Par contre, pour le **ResNet50** et le **ViT**, on est resté sur une résolution plus classique de **224x224**. Ça nous a permis d'utiliser des batch directs de **16 et 8** respectivement, sans avoir besoin de calculs complexes en plus. Au final, cette approche nous a permis de tirer le meilleur de chaque architecture, même avec la mémoire limitée de notre carte **NVIDIA Tesla T4**.

II.10.2.5 Test-Time Augmentation (TTA) :

Au moment de passer les tests finaux, on ne s'est pas contenté d'une seule lecture par image. Nous avons utilisé la technique de TTA pour stabiliser le diagnostic. Le principe est de présenter l'image sous plusieurs angles (rotation, flip) et de faire la moyenne des prédictions. Cependant, on n'a pas appliqué le même nombre de transformations à tous les réseaux. Pour les modèles les plus lourds ou les plus sensibles, on a poussé jusqu'à 8 transformations (**TTA x8**) pour réduire l'incertitude au maximum, alors que pour d'autres, un nombre plus réduit était suffisant pour obtenir une réponse stable. Cette approche personnalisée nous a permis de gagner en fiabilité sans alourdir inutilement le temps de traitement [71].

II.10.3 Optimisation et Régularisation du ViT :

Le **Vision Transformer (ViT)** est un modèle super puissant mais il a un gros défaut : il est très gourmand en données. Sur un dataset médical comme le nôtre qui reste de taille modeste, le risque de sur-apprentissage (**Overfitting**) est énorme. Pour obliger le réseau à apprendre des vrais signes cliniques au lieu de juste mémoriser les images par coeur, on a dû mettre en place toute une batterie de réglages :

- **Le Dropout et le DropPath** : Pour éviter que le modèle se repose tout le temps sur les mêmes neurones, on a mis un **Dropout de [74] 0.5** sur la tête de classification et un **DropPath [75] de 0.2** dans les blocs Transformer. En gros, on coupe aléatoirement des connexions pendant l'apprentissage. Ça force le réseau à trouver d'autres chemins pour diagnostiquer, ce qui le rend bien plus robuste au final.
- **Le Mixup et le Coarse Dropout** : Pour avoir plus de diversité, on a utilisé le **Mixup [76]**. C'est une technique qui mélange deux images de tumeurs pour créer des exemples un peu hybrides. Si on rajoute à ça le **Coarse Dropout [77]** (qui cache des morceaux de l'image), on force le ViT à regarder la forme globale de la lésion plutôt que de se focaliser sur des petits détails qui sont parfois juste des parasites sur l'image.
- **Le Warm-up et le dégel du modèle** : On n'a pas lancé l'apprentissage à fond dès le début. On a commencé par une phase de **Warm-up [78]** en bloquant le corps du modèle (**backbone**) pour apprentissage seulement la nouvelle tête. Une fois que c'est stable, on a tout débloqué (**Unfreeze**) avec un taux d'apprentissage plus bas. C'est ce qui permet au ViT de s'habituer doucement aux images d'échographies sans bousiller ce qu'il a appris avant.
- **La profondeur stochastique** : On a aussi activé ce qu'on appelle le **Stochastic Depth [75]**. Le principe est de faire "sauter" certaines couches du Transformer au hasard pendant qu'il apprend. Ça évite que le modèle devienne trop rigide et ça l'aide vraiment à mieux généraliser sur des images qu'il n'a jamais vues.
- **Le TTA pour la fiabilité** : Pour finir, au moment du test, on utilise la **TTA [71]**. Au lieu de lui donner juste une image, on lui montre 5 versions un peu modifiées (rotation, miroir). En faisant la moyenne des réponses, on stabilise le diagnostic et on réduit pas mal les erreurs sur les cas qui sont difficiles à trancher.

II.10.4 Stratégie d'Ensemble (Décision finale)

Ayant réalisé l'apprentissage de trois modèles robustes et distincts (ResNet, EfficientNet et ViT), nous avons décidé de ne pas nous fier à un seul modèle, mais plutôt de les combiner pour obtenir un diagnostic plus précis. Nous avons utilisé deux méthodes complémentaires :

- **Vote pondéré** : Cette méthode consiste à calculer la moyenne des probabilités fournies par chaque modèle. Par exemple, si le modèle EfficientNet est certain à 90 % et les deux autres à 70 %, la moyenne nous donne un résultat plus stable et équilibré. Cela permet de réduire les erreurs aléatoires qu'un modèle seul pourrait commettre.

- **OU logique** : Il s'agit du « plan de sécurité » d'un point de vue médical. Dans le diagnostic du cancer, le pire serait que le système se trompe et classe une tumeur maligne comme bénigne (faux négatif). Par conséquent, nous avons programmé le système de sorte que si au moins un des trois modèles détecte une tumeur maligne, le cas est classé comme tel et immédiatement étiqueté comme « malin ».

En combinant ces deux méthodes, nous avons garanti que le système est très précis dans les cas normaux, tout en étant extrêmement « prudent » dans les cas suspects afin de protéger la vie des patients.

II.11 Conclusion du Chapitre 2

En résumé, ce deuxième chapitre nous a permis de poser tout le cadre théorique et technique nécessaire pour notre projet. On a pu explorer les architectures de pointe, allant du classique U-Net pour la segmentation jusqu'aux modèles plus récents comme EfficientNet-B4 et les Vision Transformers pour la partie classification.

L'analyse du jeu de données BUSI a montré que le principal défi reste la quantité limitée d'images (780 cas), ce qui justifie totalement nos choix méthodologiques : l'utilisation massive du Transfer Learning, la stratégie d'augmentation de données offline (pour passer à 3 270 images) et l'application de techniques d'optimisation comme le TTA et le Focal Loss pour gérer le déséquilibre entre les classes.

Toutes les briques technologiques sont maintenant en place. Nous avons défini les protocoles d'apprentissages, choisi les métriques d'évaluation et préparé l'environnement de calcul sur GPU. Ce socle méthodologique nous permet désormais de passer à la phase la plus importante de notre travail. Dans le chapitre suivant, nous allons présenter et analysé en détails les résultats de nos tests pour voir comment ces modèles se comportent réellement face au diagnostic des tumeurs mammaires.



Chapitre 03 :

Evaluation expérimentale interprétabilité des modèles de Deep Learning



III.1 Introduction

Ce troisième chapitre est dédié à la présentation, l'analyse et la discussion des résultats expérimentaux obtenus par notre système d'intelligence artificielle multitâche. Notre objectif principal est d'évaluer la capacité de nos modèles à assister les médecins dans le diagnostic du cancer du sein par échographie, en combinant la précision géométrique de la segmentation et la fiabilité de la classification.

Dans un premier temps, nous analysons les performances de nos modèles de segmentation (U-Net, Attention U-Net et U-Net++) ainsi que nos classifieurs (EfficientNet-B4, ResNet-50, ViT-B/16). Nous mettons en évidence l'importance cruciale du passage d'une base de données initiale déséquilibrée à une base BUSI équilibrée, une étape indispensable pour corriger le biais algorithmique et garantir la sécurité clinique des patientes.

Dans un second temps, nous étudions la capacité de généralisation de notre système dans des conditions réelles. Pour cela, nous présentons les résultats de l'évaluation de nos modèles sur la base de données externe BRUSBRA, avant et après l'application du Fine-Tuning. Enfin, nous terminons ce chapitre par une comparaison approfondie de notre travail avec l'état de l'art, suivie par la présentation de notre interface finale « **SenoDeep Predict** » qui intègre ces meilleurs modèles pour une utilisation concrète.

III.2 Présentation et interprétation des résultats de segmentation :

Dans cette section, nous détaillons les performances obtenues par les trois architectures testées **U-Net**, **Attention U-Net** et **U-Net++**. Nous commençons par analyser leur comportement durant la phase d'apprentissage, avant de comparer leurs résultats chiffrés sur la base de test et d'illustrer leurs capacités par une analyse visuelle des segmentations produites.

III.2.1 Courbes d'apprentissage :

Les figures 11 et 12 montrent respectivement les courbes de perte et de précision durant l'apprentissage des trois modèles. On observe que tous les modèles convergent de manière stable, avec une décroissance rapide de la perte dans les premières itérations suivie d'un plateau. Le U-Net converge le plus rapidement (arrêt à itération 17), tandis que l'Attention U-Net bénéficie d'un apprentissage plus long (itération 24), lui permettant d'atteindre un meilleur optimum. Les courbes de validation restent proches des courbes d'apprentissage, démontre une excellente capacité de généralisation, confirmant ainsi l'absence de sur-apprentissage (overfitting) significatif.

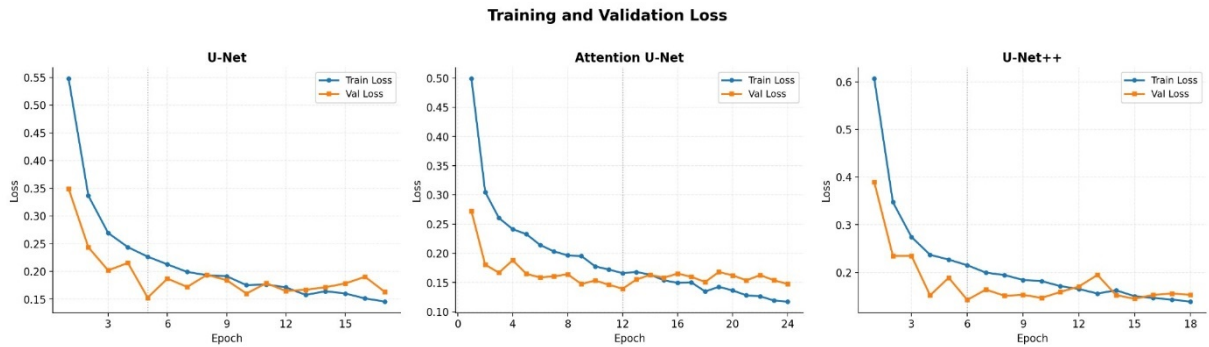


Figure 11 : Courbes de précision (Training Accuracy / Validation Accuracy) pour les trois architectures.

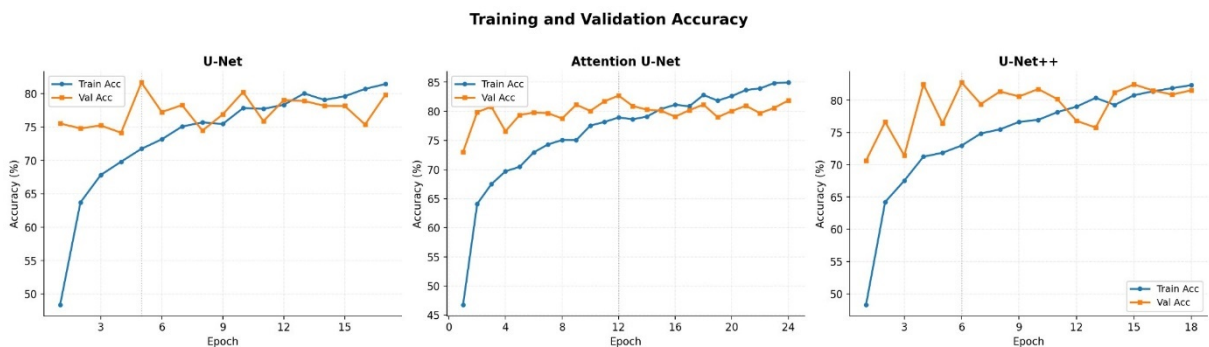


Figure 12: Courbes de perte (Training Loss / Validation Loss) pour U-Net, Attention U-Net et U-Net++

III.2.2 Comparaison globale des modèles

Le tableau 5 récapitule toutes les métriques de segmentation obtenues sur l'ensemble de test pour les trois architectures. Ces métriques couvrent le coefficient de Dice (F1), l'IoU (Intersection over Union / Jaccard Index), la sensibilité (recall), la spécificité, la précision, la précision au niveau pixel et la distance de Hausdorff.

Modèle	Dice	IoU	Sensibilité	Spécificité	Précision	Pixel Acc.	Hausdorff (px)
U-Net	0.7042	0.6370	0.7871	0.9775	0.8035	0.9574	21.84
Attention U-Net	0.7773	0.7108	0.8266	0.9805	0.8649	0.9628	20.91
U-Net++	0.7618	0.6972	0.8149	0.9811	0.8659	0.9613	19.79

Tableau 5 : Métriques de segmentation

L'Attention U-Net se démarque clairement sur les métriques Dice (+10.4 % vs U-Net), IoU (+11.6 % vs U-Net) et sensibilité (+4.8 % vs U-Net). Le U-Net++ présente les meilleures précision et spécificité, ainsi que la plus faible distance de Hausdorff (19.79 px), indiquant une meilleure précision des contours de segmentation.

III.2.2.1 La supériorité de l'Attention U-Net :

L'apprentissage, configuré pour un maximum de 50 itérations, a convergé de manière stable avec un arrêt effectif (Early Stopping) à la 24^e itération pour le modèle Attention U-Net. L'évolution parallèle et resserré des courbes d'apprentissage et de validation démontre une excellente capacité de généralisation et l'absence de surapprentissage significatif. L'évaluation des performances sur l'ensemble de test indépendant révèle des résultats remarquables pour la tâche de segmentation, avec une précision au niveau pixel (Pixel Accuracy) de 96,28 %. Plus spécifiquement, la Précision positive atteint 86,49 %, soulignant la fiabilité du modèle dans la délinéation correcte des tissus tumoraux tout en limitant les faux positifs. De son côté, la Sensibilité (Rappel) s'élève à 82,66 %, ce qui traduit une excellente capacité à détecter et englober les zones lésionnelles réelles, garantissant ainsi une forte sécurité clinique. Le coefficient de Dice, qui s'établit à 0,7773 (associé à un score IoU de 0,7108), montre un équilibre optimal de chevauchement spatiale entre les prédictions et les masques de référence. Par ailleurs, le modèle maintient une spécificité très élevée de 98,05 % et une distance de Hausdorff moyenne contenu à 20,91 pixels, prouvant que les erreurs maximales sur les contours restent limitées. Fort du meilleur score composite global (0,8538) de notre étude, ces performances démontrent que l'Attention U-Net est robuste.

Cette supériorité algorithmique ne relève pas du hasard ; elle s'explique fondamentalement par l'architecture du réseau. En pratique, l'intégration des Attention Gates

(portes d'attention) a joué un rôle déterminant face à la complexité inhérente aux images échographiques. Contrairement au U-Net standard qui traite toutes les caractéristiques spatiales avec le même poids, ce (Speckle) et les zones d'ombres caractéristiques des échographies mammaires. Ainsi, le modèle a pu focaliser ses ressources computationnelles exclusivement sur les régions lésionnelles discriminantes, supprimant les activations non pertinentes dans les tissus sains environnants. Cette capacité d'adaptation c'est révélé particulièrement cruciale pour la classe maligne, reconnu comme la plus difficile à segmenter en raison de sa forte variabilité morphologique et de ses contours souvent irréguliers ou spiculés.

Au-delà de ces considérations techniques, d'un point de vue clinique, l'interprétation de ces performances spatiale revêt une importance capitale. En imagerie médical, et plus particulièrement dans le dépistage du cancer du sein, l'enjeux principal est de ne rater aucune zone pathologique. La sensibilité élevée atteinte par l'Attention U-Net (82,66 %) répond directement à ce besoin : le modèle détecte et englobe une plus grande proportion des pixels réels de la lésion, ce qui est essentiel pour minimiser les faux négatifs critiques. Concomitamment, la précision maintenue à un niveau élevé (86,49 %) garantit un faible taux de faux positifs, évitant ainsi de segmenter des tissus sains, ce qui pourrait engendrer de l'anxiété chez la patiente et conduire à des biopsies inutiles. Enfin, avec une distance de Hausdorff contenu autour de 20,91 pixels, le système démontre une précision de contour suffisante pour aider à évaluer la taille et la forme de la tumeur. L'ensemble de ces facteurs confirment que l'Attention U-Net possède la fiabilité nécessaire pour être déployé comme un outil de Second-Lecteur (Second Reader), capable d'assister les radiologues dans la délinéation rapide et précise des tumeurs mammaires en contexte diagnostique réel.

III.2.2.2 U-Net++ :

Bien que l'Attention U-Net domine le classement global, le modèle U-Net++ se positionne comme un concurrent très robuste en décrochant la deuxième place avec un score composite de 0,8470 et un coefficient de Dice de 0,7618. L'analyse détaillée de ses performances révèle une force majeure et spécifique : il obtient la meilleure Précision de toute l'étude (86,59 %) ainsi que la plus faible distance de Hausdorff moyenne (19,79 pixels). Ces résultats indiquent que lorsque le U-Net++ identifie une zone tumorale, les contours qu'il génère sont les plus fidèles et les plus proches de la réalité (Ground Truth), minimisant considérablement le débordement sur les tissus sains.

D'un point de vue architectural, cette supériorité sur la finesse géométrique s'explique par la refonte profonde de ses connexions de saut (skip connections). Contrairement au U-Net standard, le U-Net++ intègre des nœuds intermédiaires denses (dense blocks) entre l'encodeur et le décodeur. Ces sous-réseaux imbriqués permettent une fusion multi-échelle beaucoup plus fine des caractéristiques, réduisant drastiquement l'écart sémantique entre les cartes de caractéristiques de bas niveau (les contours bruts) et de haut niveau (le sens global de l'image). C'est cette architecture dense qui lui confère cette capacité unique à "redessiner" les bords de la lésion avec une telle exactitude.

Sur le plan clinique, l'avantage du U-Net++ ouvre des perspectives d'application différente de celles de l'Attention U-Net. Si l'Attention U-Net est idéal pour le dépistage (grâce à sa haute sensibilité), le U-Net++ brille dans les scénarios de planification chirurgical ou thérapeutique. En effet, lorsqu'un chirurgien doit procéder à une tumorectomie (ablation de la tumeur), la précision des marges de résection est critique : il faut enlever toute la tumeur tout en préservant un maximum de tissus mammaire sain. La faible distance de Hausdorff (19,79 px) et la spécificité record (98,11 %) du U-Net++ en fait un candidat idéal pour assisté géométriquement les praticiens dans la délimitation per-opératoire des lésions.

III.2.2.3 U-Net :

Le modèle U-Net standard se classe en troisième position avec un score composite de 0,8111, un coefficient de Dice de 0,7042 et une sensibilité de 78,71 %. Même si ces métriques sont logiquement en retrait par rapport à ses variantes plus moderne, l'analyse de son comportement d'apprentissage met en lumière un avantage indéniable : son efficacité computationnelle. Le U-Net a convergé de manière très stable et beaucoup plus vite que les autres modèles, déclenchant l'arrêt précoce (Early Stopping) dès la 17ème itération.

Les limites de ses performances spéciales s'expliquent directement par son architecture de base. Les connexions de saut directes (skip connections simple) transmettent l'intégralité des cartes de caractéristiques de l'encodeur vers le décodeur, sans aucun filtre. En l'absence de mécanismes de raffinement (comme les portes d'attention ou les nœuds denses), le modèle peine d'avantage à isoler la tumeur lorsqu'il est confronté au bruit acoustique (speckle) et au faible contraste des échographies mammaire, ce qui se traduit par une distance de Hausdorff plus élevée (21,84 pixels) et quelques faux négatifs supplémentaires.

Néanmoins, l'interprétation clinique de ces résultats reste très pertinente. Avec une précision au niveau pixel de 95,74 % et une spécificité de 97,75 %, le U-Net original maintien des performances de bases très solide. Sa rapidité de convergence et sa structure plus légère (moins de paramètres que le U-Net++) en fait un modèle de choix pour des déploiements dans des environnements aux ressources computationnelles limité. Par exemple, ce modèle serait Particulièrement adapter pour être intégré directement dans des appareils d'échographie portable (Edge AI) utilisé dans les déserts médicaux ou les pays en voie de développement, ou une assistance au diagnostic rapide et peu gourmande en énergie est prioritaire sur l'optimisation extrême des contours.

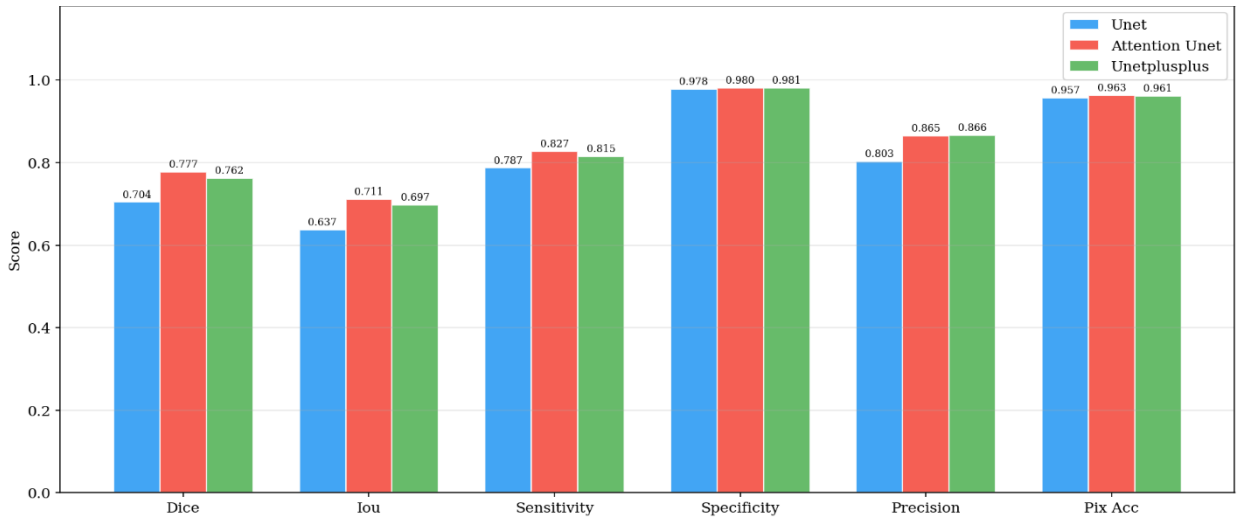


Figure 13: Comparaison des performances des architectures de segmentation (U-Net, Attention U-Net, U-Net++).

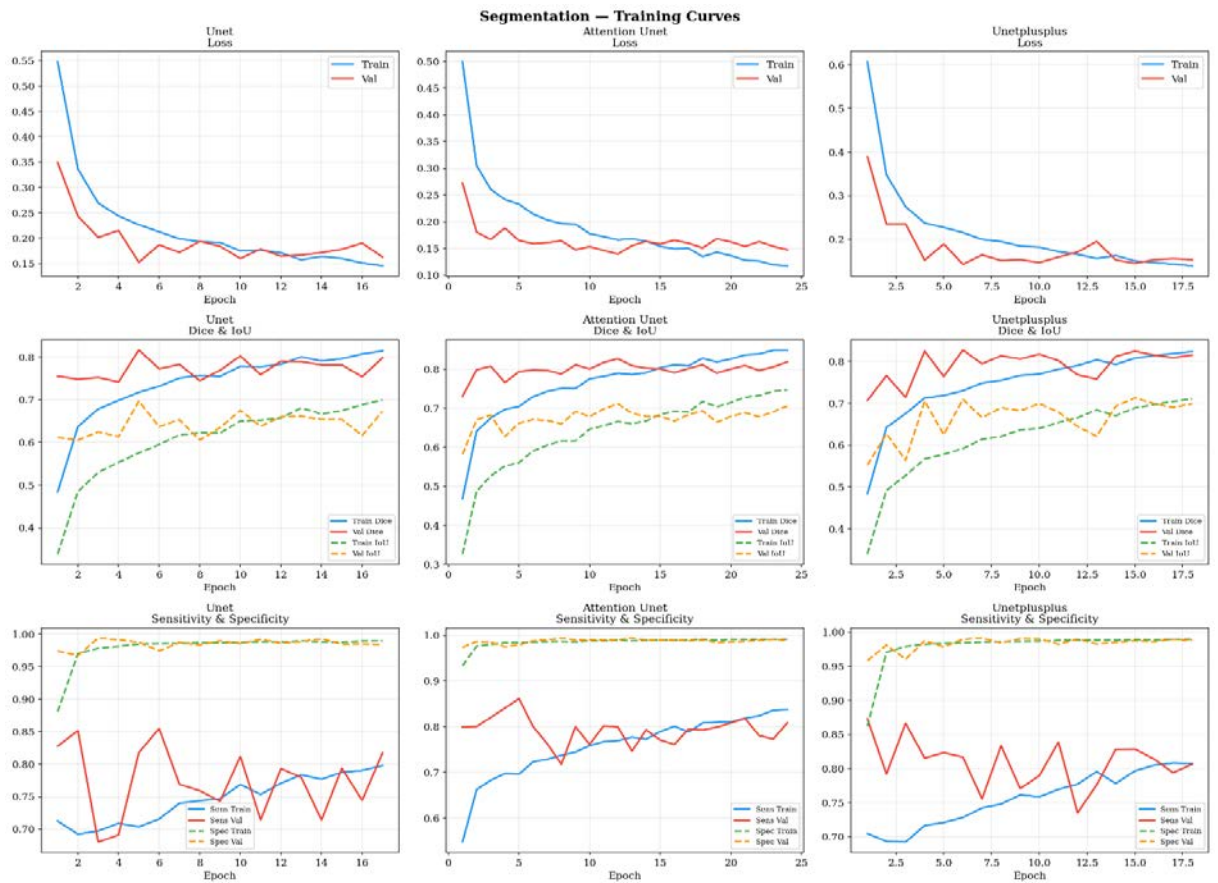


Figure 14 : Dynamique d'apprentissage (Loss, Dice, IoU, Sensibilité et Spécificité) des trois architectures de segmentation.

L'analyse de ces courbes d'apprentissage nous permet d'étudier le comportement des trois modèles pendant l'apprentissage.

1. **La perte (Loss) :** Sur la première ligne, on voit que l'erreur diminue bien pour tous les modèles. Le U-Net et le U-Net++ ont convergés assez vite (déclenchant l'arrêt précoce vers l'itération 17 ou 18). L'Attention U-Net a continué à apprendre plus longtemps (jusqu'à 25 itérations) sans faire de sur-apprentissage (overfitting).
2. **L'évolution des scores (Dice et IoU) :** Sur la deuxième ligne, on remarque que les performances augmentent vite au début puis se stabilisent. Les courbes de validation en rouge font quelques "zig-zags" (La forte variabilité entre tumeurs mammaires) mais suivent bien la progression de la courbe d'apprentissage.
3. **Sensibilité vs Spécificité :** Les lignes en pointillés (la Spécificité) sont presque Parfaitement droites et collées vers le haut (proche de 1.0), ce qui montre que le Modèle identifie sans ambiguïté les tissus sains et les zones non-tumorales. D'un autre côté, la ligne rouge (Sensibilité de validation) présente des fluctuations importantes, ça montre la vraie difficulté de notre sujet : ignorer le fond noir est facile, mais capter toute la tumeur avec ses contours complexes reste une tâche dure et instable d'une itération à l'autre.

III.2.3 Distance de Hausdorff :

La figure ci-dessous présente la distance de Hausdorff moyenne (en pixels) pour les trois modèles d'étude. Comme on peut le constater, c'est le modèle U-Net++ qui offre la meilleure performance avec une distance de 19.79 px, suivie de près par l'Attention U-Net (20.91 px) et le U-Net classique (21.84 px). Bien que les moyennes soient relativement proches, le point le plus important à souligner est la taille des barres d'erreur noires qui représentent l'écart-type. Cet écart-type très grand indique une forte instabilité dans les prédictions. En effet, même si les modèles arrivent à bien segmenter la plupart des cas normaux, ils font parfois des erreurs considérables sur les images plus complexes (comme les tumeurs avec des bordures peu claires). Sur le plan graphique, l'extension des barres d'erreur en dessous de l'axe des ordonnées ne traduit pas l'existence de distances négatives, ce qui serait physiquement impossible. Ce phénomène est une conséquence purement statistique de la représentation (**Moyenne Écart-Type**). La valeur élevée de cet écart-type illustre la forte hétérogénéité de nos résultats : si les modèles sont très performants sur les cas simples, ils restent sensibles à la complexité de certaines lésions dont les contours sont particulièrement diffus.

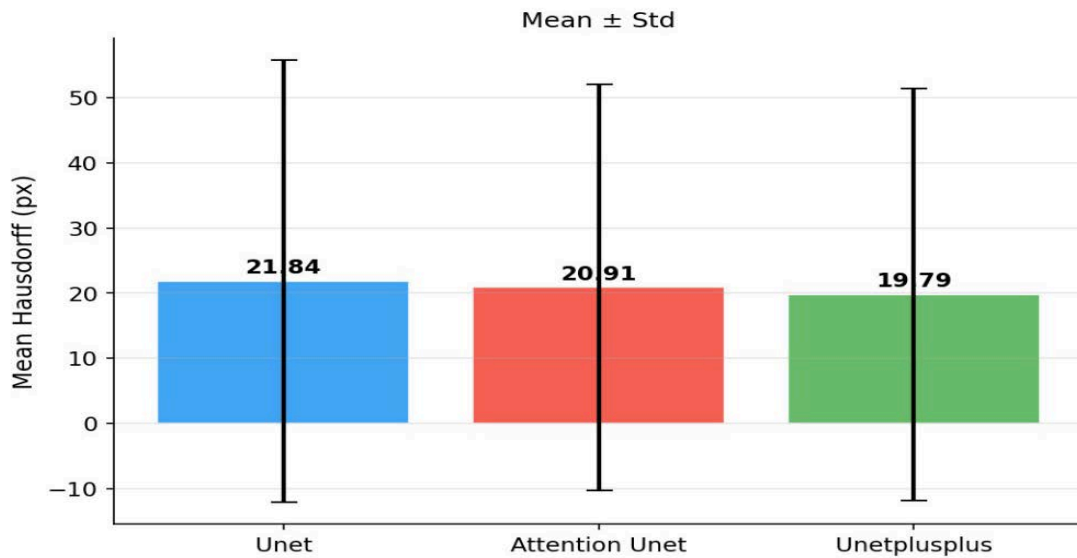


Figure 15: Moyenne de la distance de Hausdorff par modèle (valeurs basses = meilleures).

C'est cependant le U-Net++ qui signe la meilleure performance de notre étude concernant cette métrique, avec une moyenne de 19,79 px. Grâce à ses connections dense et imbriquées, il réussit à capturer des détails morphologiques extrêmement fins, offrant une délimitation au plus proche de l'expertise médical. Au-delà de cette moyenne, l'observation de l'écart-type montre que le modèle U-Net++ possède la distribution la plus homogène parmi les trois modèles. Cela traduit une réduction de la variance et une meilleure robustesse face aux valeurs aberrantes, marquant ainsi une progression nette par rapport au modèle de base.

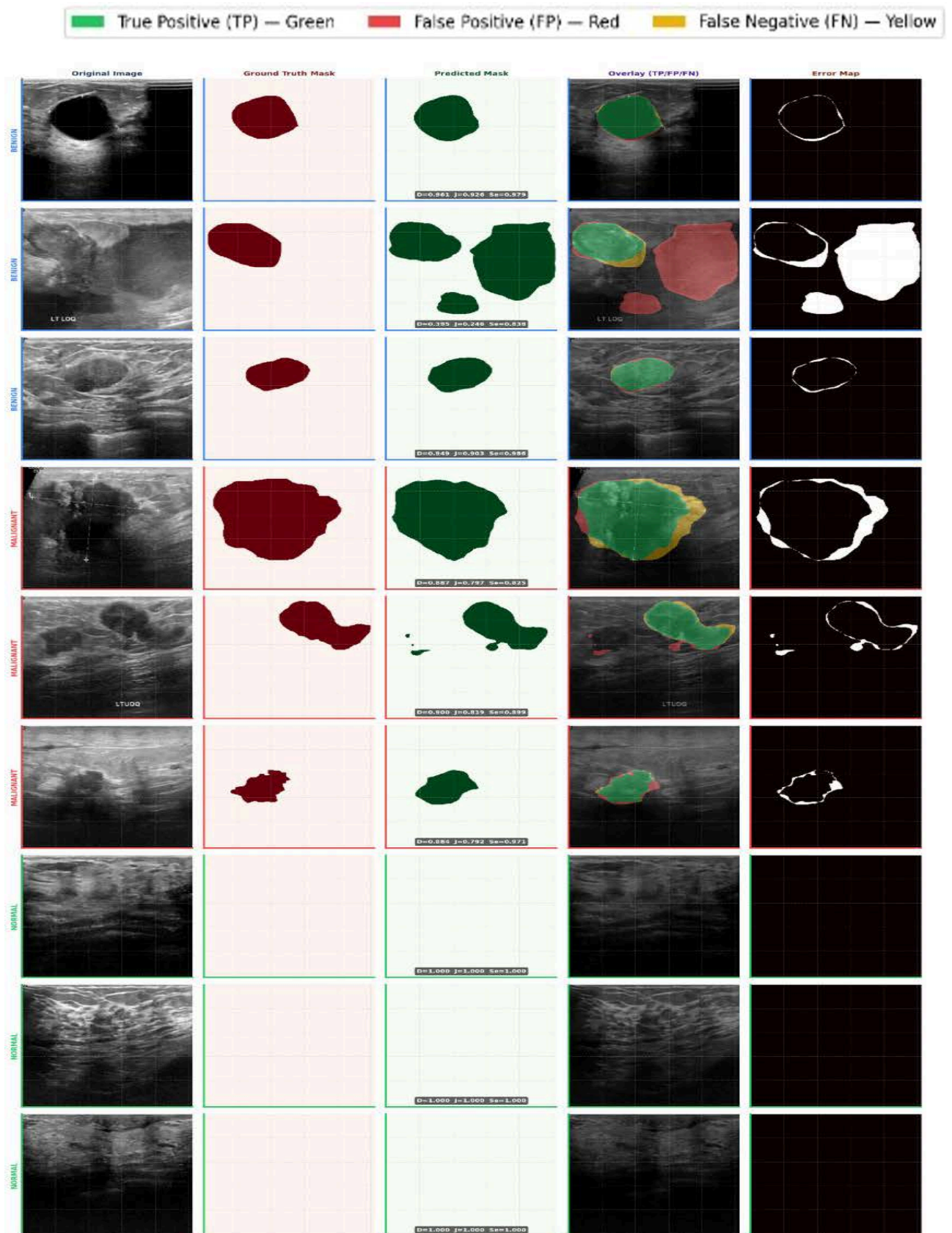


Figure 16: Résultats qualitatifs détaillés par classe (bénigne, maligne, normale) — Attention U-Net.

La Figure 16 illustre les prédictions de l'Attention U-Net. Elle nous prouve que les chiffres ne disent pas tout : la visualisation permet de voir exactement où le modèle est performant et où il fait des erreurs.

Pour la classe des tumeurs bénignes (les trois premières lignes), le constat est très intéressant. Sur le premier (Dice=0.961) et le troisième (Dice=0.949) cas, le modèle fait un travail remarquable : le masque prédit se superpose presque parfaitement à la vérité terrain, avec une carte d'erreur (Error Map) qui se limite à un très fin contour. Cependant, le deuxième cas (Dice=0.400) bénin attire immédiatement l'attention par son échec flagrant. Alors que la tumeur réelle est unique et bien définie, notre modèle a généré des segmentations aberrantes (faux positifs) en identifiant deux masses supplémentaires inexistantes. Ce phénomène de sur-segmentation indique que les artefacts acoustiques, tels que les ombres portées, peuvent induire le réseau en erreur en mimant la texture d'une lésion réelle.

Du côté des tumeurs malignes (lignes 4 à 6), L'architecture réussit systématiquement à localiser et segmenter le cœur de la masse cancéreuse (la zone verte sur l'overlay est toujours bien centrée). Par contre, on voit très clairement sur les cartes d'erreurs que le réseau peine à capturer les bordures irrégulières et spiculées, qui sont la signature même de la malignité. On le voit bien sur le 4ème cas (Dice=0.887) avec cet anneau d'erreur blanc très épais sur les bords. Le 5ème cas (Dice=0.900) est aussi un beau défi : c'est une lésion multi-focale, et si le modèle détecte bien les zones principales, il a tendance à baver un peu en créant des minuscules îlots de faux positifs autour.

La vraie grande satisfaction de cette figure vient de la dernière partie c'est les cas normal (tissus sains). Les masques prédits sont totalement vides et les cartes d'erreurs sont d'un noir absolu (Dice parfait de 1.000). Cela confirme de manière visuelle que notre modèle possède une spécificité excellente.

Cette analyse qualitative est fondamentale pour valider la pertinence clinique de notre modèle. Elle valide la robustesse du modèle pour rejeter les cas sains et localiser les cancers, mais elle met aussi en lumière ses limites actuelles face aux ombres trompeuses extrêmes et à la micro-géométrie des bordures malignes.

III.3 Présentation et interprétation des résultats de classification :

Dans cette section, nous présentons et analysons les performances obtenues par nos trois modèles de classification : **ResNet-50**, **EfficientNet-B4** et le **Vision Transformer (ViT-B16)**. Cette évaluation repose sur l'analyse de leur comportement durant l'apprentissage, suivie d'une confrontation rigoureuse de leurs prédictions sur une base de données de test indépendant afin de mesurer leur capacité réelle de diagnostic.

III.3.1 EfficientNet B4 :

Nous présentons dans cette section les résultats obtenus par le modèle EfficientNet-B4 lors de nos tests de classification. À travers les tableaux et graphiques suivants, nous exposons la progression de son apprentissage ainsi que ses performances finales. Ces données nous permettent d'évaluer concrètement l'efficacité de cette architecture face aux spécificités des images échographiques.

Métrique	Valeur
Accuracy	88.03%
Balanced Accuracy	89.49%
F1 Macro	0.8707

Tableau 6: Synthèse des indicateurs de performance globale du modèle EfficientNet-B4.

Classe Moyenne	Précision	Recall	F1-Score	Support
Bénigne	0.95	0.86	0.90	66
Malignant	0.82	0.87	0.84	31
Normal	0.79	0.95	0.86	20
Accuracy Globale	-	-	0.88	117
Macro Avg	0.85	0.89	0.87	117
Weighted Avg	0.89	0.88	0.88	117

Tableau 7: Rapport de classification détaillé par classe (Bénigne, Maligne, Normale) pour l'architecture EfficientNet-B4

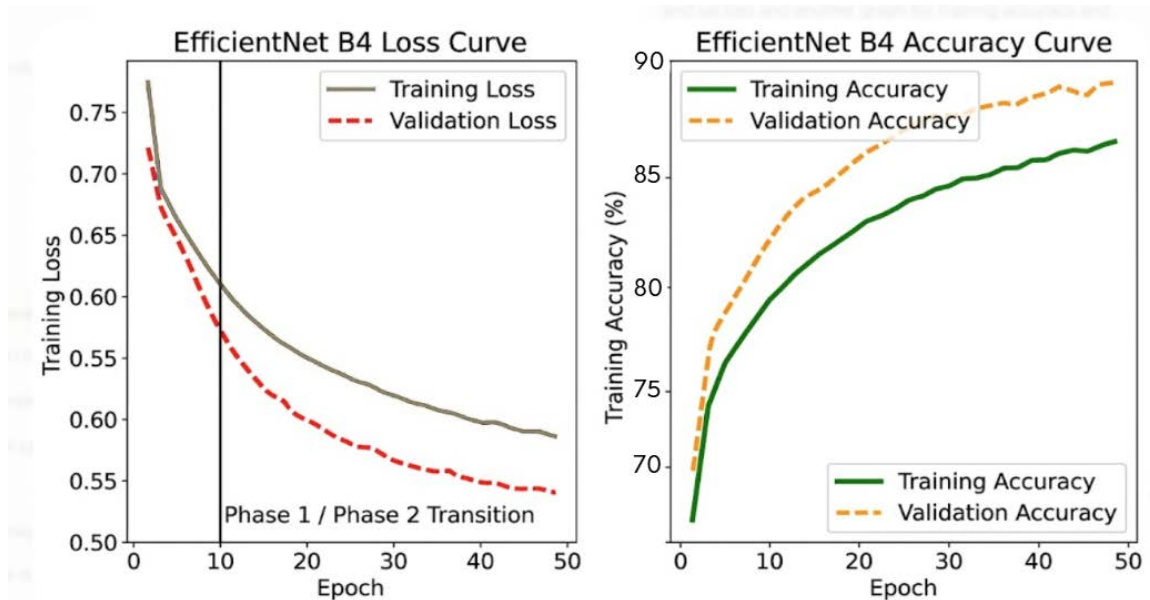


Figure 17: Courbe de perte (Training Loss / Validation Loss) et Courbe de précision (Training Accuracy / Validation Accuracy) pour EfficientNet B4.

En regardant d'abord le graphe des courbes d'apprentissage ci-dessus pour notre modèle EfficientNet-B4 sur 50 itérations, nous remarquons une ligne vertical noir à l'itération 10. Elle marque la transition entre nos deux phases de Transfer Learning. La Phase 1 correspond à l'adaptation initial de la tête de classification, pendant que la Phase 2 représente le fine-tuning ou tout le réseau s'adapte aux caractéristiques de nos images échographique. L'observation la plus importante de ces graphiques c'est la très bonne stabilité. Les courbes montrent une convergence saine, avec une perte qui diminue et une précision qui grimpe pour frôler 89%. Ce qui est très intéressant c'est que les performances en validation (en pointillé) sont meilleures que l'apprentissage. Ça prouve l'absence total d'overfitting. Lors de la phase de validation, les mécanismes de régularisation comme le Dropout et l'augmentation agressive de données sont désactivés. Le réseau évalue alors les images dans leur état original, sans les perturbations volontaires imposées durant l'apprentissage pour sa robustesse, ce qui explique mécaniquement l'obtention de scores plus élevés en validation

Ensuite, Si nous analysons les résultats finaux dans ce tableau pour l'EfficientNet-B4, Nous voyons bien que cela confirme la robustesse déjà remarquée sur les courbes. Sur le test set final, le modèle a atteint une accuracy de 88,03 % et un F1 Macro de 0,8707. Ce qui nous rassure le plus ici, c'est l'équilibre des score F1 par classe : 0,90 pour les lésions bénignes, 0,84 pour la maligne et 0,86 pour les cas normaux. Ça prouve que le réseau ne favorise pas une classe au détriment des autres. On peut voir aussi que pendant la cross-validation, notre meilleur Fold (le Fold 5) a même pu atteindre 91,67 % de précision. L'intégration du mécanisme de Test-Time Augmentation (TTA) a permis d'optimiser les scores finaux en stabilisant les sorties du modèle.

III.3.2 ResNet 50 :

Nous exposons ci-dessous les performances du modèle ResNet-50, utilisé comme architecture de référence pour notre étude. Les métriques présentées dans le tableau suivant nous permettent d'évaluer sa capacité à classer les lésions et servent de base de comparaison pour nos modèles plus récents.

Critère	ResNet-50
Accuracy (test)	73.24 %
F1 Macro (test)	0.730
AUC ROC (CV)	0.9330
F1 Bénigne	0.59
F1 Maligne	0.97
F1 Normale	0.83
Meilleure acc. Val. (CV)	74.36 %

Tableau 8 : Synthèse des indicateurs de performance et de fiabilité du modèle ResNet-50

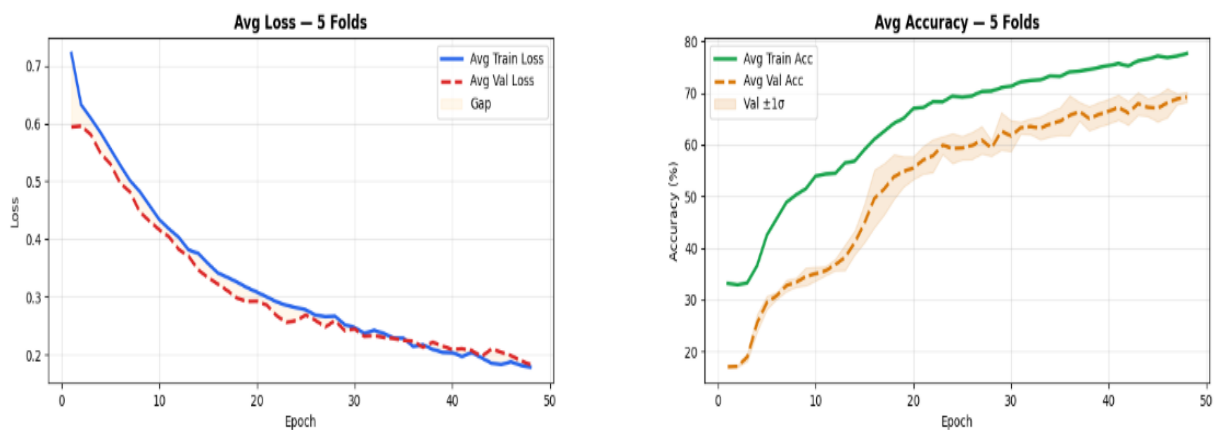


Figure 18 : Courbe de perte (Training Loss / Validation Loss) et Courbe de précision (Training Accuracy / Validation Accuracy) pour ResNet 50.

Si nous commençons par analyser le graphe des courbes moyennes d'apprentissage (Loss et Accuracy) pour le modèle ResNet-50 sur nos 5 folds, Nous remarquons tout de suite une convergence optimale. La courbe de perte diminue d'une façon très douce et régulière pour se stabiliser. Du côté de la précision, elle monte progressivement pour faire un plateau autour de 70-73% vers la fin d'apprentissage. Ce qui fait vraiment la force de ce graphe, c'est la zone ombrée autour de la courbe de validation (en pointillé orange). Ça représente l'écart-type entre nos 5 Fold de cross-validation. On voit que cette bande n'est pas très large, ce qui prouve d'une manière claire que le modèle est stable peu importe le découpage des données. De plus, la validation suit très fidèlement l'apprentissage sans jamais plonger, ce qui confirme qu'il n'y a pas d'overfitting. L'apprentissage est donc naturel et robuste, même si le modèle bloque à une précision finale un peu moyenne.

Ensuite, si nous regardons les résultats finaux dans ce tableau pour ResNet50, ils confirment parfaitement ce qu'on a vu visuellement. L'accuracy moyenne estimé est de 73,24 %, avec notre meilleur (Fold 1) qui a réussi à atteindre 74,36 %. Mais le détail le plus frappant dans ce tableau, c'est le grand déséquilibre des scores F1 par classe. Le réseau est incroyablement performant pour détecter les cas maligne (F1 de 0,97) et les tissus normal (0,83). Par contre, il a vraiment du mal avec les tumeurs bénignes ou le score chute à 0,59. Il faut aussi préciser, qu'on n'a pas utilisé la technique TTA (Test-Time Augmentation) pour ce réseau, ce qui explique en parti pourquoi son score global est resté assez modeste.

III.3.2.1 Courbe ROC :

Afin d'approfondir l'évaluation de notre modèle ResNet-50, nous avons réalisé une analyse par courbes ROC (Receiver Operating Characteristic). Cette approche est essentielle pour mesurer la capacité du classifieur à distinguer chaque pathologie, indépendamment du seuil de décision choisi.

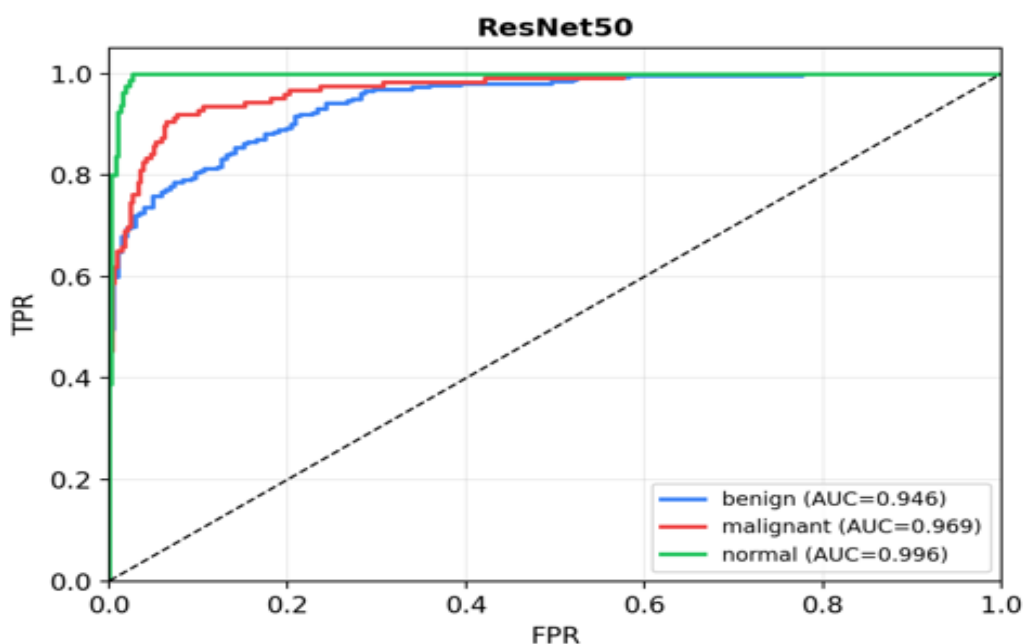


Figure 19: Courbe de ROC (Receiver Operating Characteristic) pour ResNet 50.

Ce graphe illustre les courbes ROC multiclassées obtenues par notre modèle ResNet-50. Vu qu'on travaille sur trois classes (benign, malignant, normal), la méthode utilisée ici est le Un-contre-Tous (One-vs-Rest) pour pouvoir tracer une courbe par pathologie.

En analysant le visuel, on constate tout de suite que les trois courbes s'approchent fortement du coin supérieur gauche, loin de la ligne pointillée noire qui représente le hasard. Ça indique une très bonne capacité de discrimination globale du modèle. Si on détaille par classe :

- **La classe "Normal" en vert** : Elle obtient l'aire sous la courbe (AUC) la plus haute avec 0,996. C'est un résultat presque parfait. Ça montre que le modèle arrive très facilement à faire la différence entre une échographie d'un tissu mammaire tout à fait sain et un tissu contenant une lésion.
- **La classe "Malignant" en rouge** : C'est la plus importante d'un point de vue clinique, et elle affiche un très bon score AUC de 0,969. Ça veut dire que le modèle arrive très bien à isoler les cancers, avec un bon rapport entre la sensibilité (True Positive Rate) et les fausses alertes (False Positive Rate).
- **La classe "Benign" en bleu** : Elle a le score le plus "faible" avec un AUC de 0,946, même si ça reste un résultat statistiquement très élevé. Cette petite baisse par rapport aux deux autres s'expliquent sûrement par le fait que certaines tumeurs bénignes peuvent avoir des textures ou des contours qui ressemblent parfois à des tissus normaux ou, au contraire, à des masses malignes, ce qui fait un peu plus hésiter le modèle.

En conclusion pour ce graphe, même si on a vu précédemment que l'Accuracy globale du ResNet-50 était autour de 73%, l'analyse ROC prouve que la séparation des classes dans l'espace des probabilités est en réalité très robuste.

III.3.3 ViT B/16 :

Nous terminons notre étude de classification avec l'analyse du modèle Vision Transformer (ViT-B/16). Cette architecture, qui s'appuie sur des mécanismes d'attention globale plutôt que sur des convolutions classiques, représente l'approche la plus avancée de notre projet. Les figures et tableaux suivants illustrent son comportement durant l'apprentissage ainsi que les performances exceptionnelles qu'il a permis d'atteindre sur l'ensemble de nos tests.

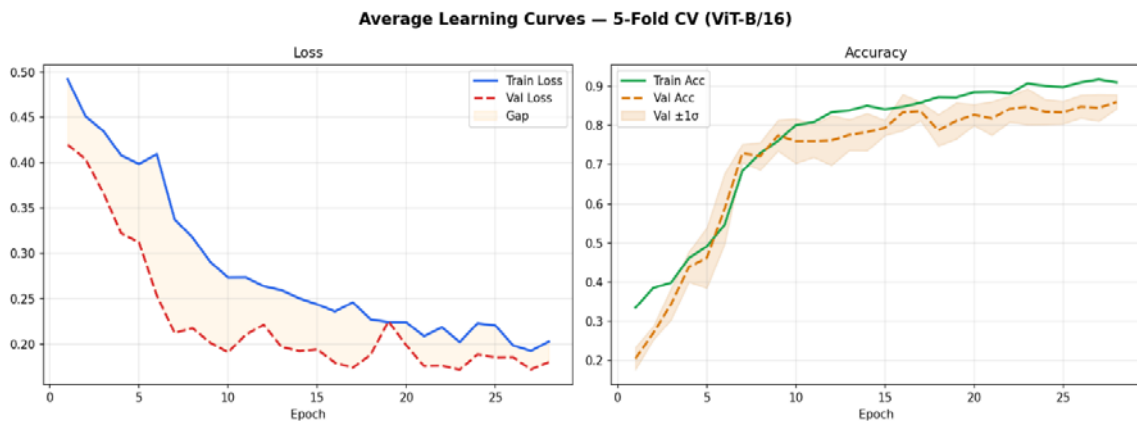


Figure 20: Courbe de perte (Training Loss / Validation Loss) et Courbe de précision (Training Accuracy / Validation Accuracy) pour ViT B/16.

III.3.3.1 Analyse des courbes d'apprentissage (ViT-B/16) :

En regardant nos courbes moyennes sur les 5 folds, on voit clairement comment notre modèle se comporte.

Sur la courbe de Loss (à gauche), on remarque un truc qui peut paraître contre-intuitif au début : la perte de validation (en rouge pointillé) est plus basse que celle d'apprentissage (en bleu). En fait, c'est le résultat direct de la stratégie qu'on a mis dans le code. Vu qu'on a appliqué une très forte régularisation pendant la phase d'apprentissage (un Dropout massif de 0.5 sur notre nouvelle tête de classification, du DropPath à 0.2, et des augmentations d'images assez agressives comme le Mixup ou le CoarseDropout), on rend la tâche volontairement difficile pour le réseau. Par contre, en phase de validation, tous ces pièges sont désactivés. Le modèle a donc de bien meilleurs scores. L'essentiel ici, c'est que la courbe rouge descend et se stabilise sans jamais remonter vers le haut, ce qui prouve qu'on a complètement évité le sur-apprentissage (overfitting).

Pour la courbe d'Accuracy (à droite), la progression est très saine. On note que le dégel du backbone (unrez à l'itération 5 avec un Learning rate plus bas) a permis au réseau de commencer à vraiment bien apprendre. La bande orange ombrée représente l'écart-type entre nos 5 Fold de cross-validation. On voit que cette zone n'est pas très large. Ça prouve que peu importe comment on a divisé la base BUSI, notre ViT-B/16 reste super stable et donne des performances homogènes d'un fold à l'autre.

III.3.3.2 Interprétation des résultats finaux (ViT-B/16) :

Dans cette partie, on présente les résultats finaux obtenus par notre modèle ViT-B/16 sur le test set aveugle (117 images), évalué avec l'aide de la technique TTA (Test-Time Augmentation avec 5 passes). Globalement, les performances sont vraiment excellentes. Le modèle atteint une accuracy de 93,16 % et un F1 Macro très proche de 0,93. Ce qui est

remarquable ici, c'est l'AUC ROC qui frôle la perfection avec un score de 0,9796, ce qui prouve la forte capacité du modèle à discriminer les trois classes. D'un point de vue purement clinique, on note que la sensibilité (recall) pour la classe "malignant" est de 93,55 % avec une spécificité de 96,51 %. Ça veut dire que le modèle rate très peu de cancers et génère très peu de fausses alertes. De plus, les scores F1 sont très équilibrés entre les classes tous au-dessus de 0,92 ce qui confirme que la classe majoritaire "bénigne" ne biaise pas les prédictions. L'utilisation du TTA a clairement permis de consolider ces résultats finaux et de dépasser la moyenne obtenue lors de la validation croisée qui était autours de 89,44 %.

Métrique	Valeur (Test Set avec TTA × 5)	Moyenne CV (5-Folds)
Accuracy	93.16 %	89.44 %
F1 Macro	0.9295	0.8881
F1 Weighted	0.9318	N/D
AUC ROC	0.9796	0.9492

Tableau 9: Métriques Globales (Test Set vs Cross-Validation).

Classe	Précision	Sensibilité (Recall)	Spécificité	F1-Score	Support
Benign	0.9677	0.9091	0.9608	0.9375	66
Malignant	0.9062	0.9355	0.9651	0.9206	31
Normal	0.8696	1.0000	0.9691	0.9302	20

Tableau 10: Rapport de classification

III.3.3.3 Courbe ROC :

Nous complétons notre analyse du modèle ViT-B/16 par l'étude de ses courbes ROC. Cette étape nous permet d'évaluer la finesse du diagnostic pour chaque pathologie en mesurant la capacité du transformer à séparer les classes avec certitude.

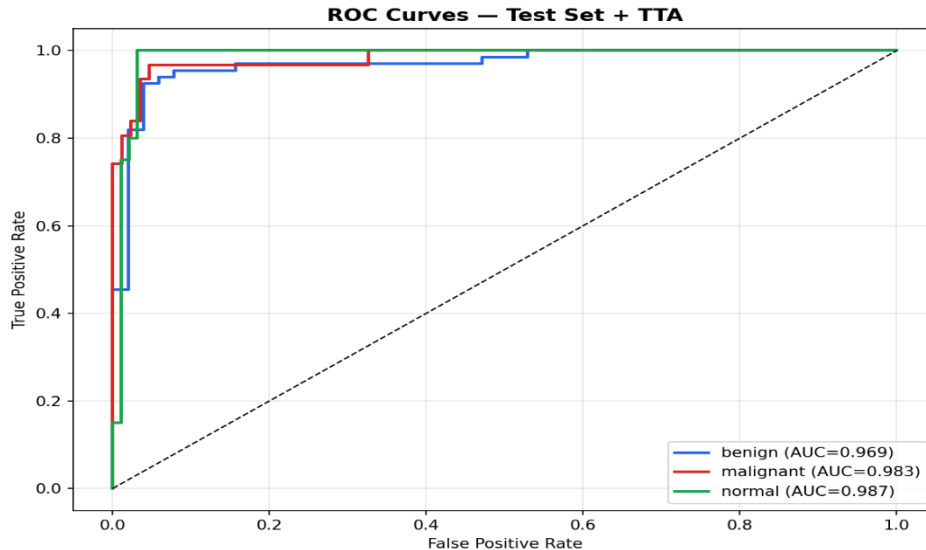


Figure 21: Courbe de ROC (Receiver Operating Characteristic) pour ViT B/16.

Ce graphe illustre les courbes ROC (Receiver Operating Characteristic) multiclassées obtenues par notre modèle lors de l'évaluation finale sur notre Test Set, en appliquant la technique d'augmentation au moment du test (TTA). Vu qu'on traite trois classes, la méthode utilisée est le "Un-contre-Tous" (One-vs-Rest) pour pouvoir analyser chaque pathologie séparément.

En regardant l'allure générale de la figure, on constate immédiatement que les résultats sont remarquables. Les trois courbes grimpent très rapidement vers la perfection (le coin supérieur gauche), s'éloignant drastiquement de la diagonale pointillée qui représente un choix complètement aléatoire. Si on regarde le détail des aires sous la courbe (AUC) :

- **La classe "Normal" (en vert) :** Elle obtient le score le plus élevé avec un AUC quasi-parfait de **0,987**. Ça prouve que le modèle arrive à isoler les images de tissu mammaire sain de manière extrêmement fiable.
- **La classe "Malignant" (en rouge) :** C'est évidemment la classe la plus critique pour une application clinique. Avec un AUC de **0,983**, on confirme que le modèle est super performant pour repérer les cancers. Il réussit à maintenir un taux de vrais positifs très haut, tout en gardant un taux de fausses alertes (False Positive Rate) très faible.
- **La classe "Bénigne" (en bleu) :** Elle affiche le score le plus "bas" avec **0,969**, mais attention, ça reste une performance statistique de très haut niveau ! Cette toute petite baisse de régime s'explique par la nature même des tumeurs bénignes à l'échographie. Leurs caractéristiques visuelles peuvent parfois se confondre avec du tissu normal ou, à

l'inverse, paraître suspectes, ce qui rend le tri probabiliste un peu plus difficile pour le réseau.

En conclusion, l'application du TTA sur notre jeu de test indépendant a permis d'obtenir des distributions de probabilités très robustes. Le modèle classe les pathologies avec une grande certitude, confirmant sa viabilité pour intégrer un vrai système d'aide au diagnostic (CAD).

III.3.3.4 Matrice de confusion :

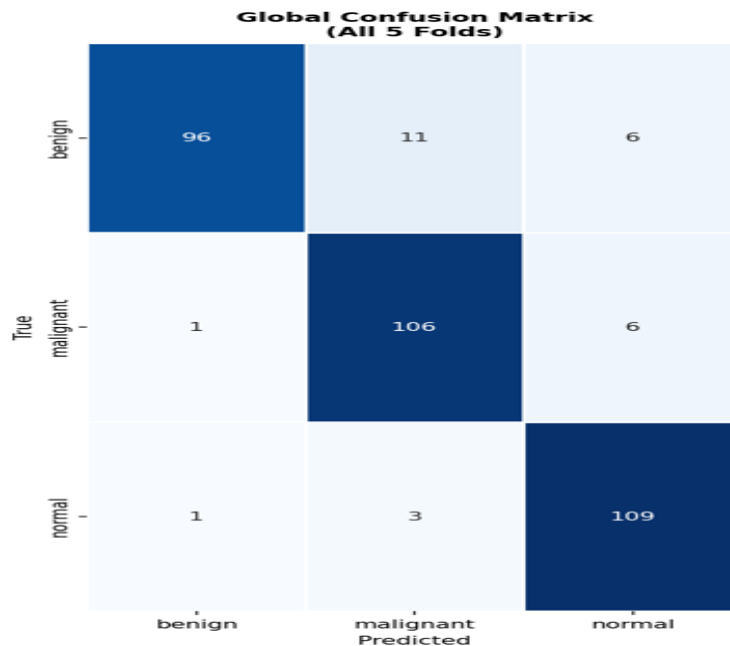


Figure 22:Matrice de confusion de Vit B/16.

La matrice globale montre que le modèle Vit-B/16 offre des performances très équilibrées sur l'ensemble des pathologies. Dans le détail :

- Il excelle pour identifier le tissu sain (**Normal**) avec **109 prédictions correctes sur 113**, ne faisant que 4 petites erreurs.
- Pour la détection des cancers (**Malignant**), le résultat est cliniquement très rassurant, il identifie correctement **106 vrais cancers sur 113**, avec un taux d'erreurs critiques infime (seulement **1** cas malin confondu avec un bénin).
- Enfin, sur la classe **Bénigne**, le modèle s'en sort bien avec **96** bonnes classifications. Les erreurs restantes (**11** prédites malignes et **6** normales) s'expliquent parfaitement par l'ambiguïté physique de l'échographie. D'un côté, certaines masses bénignes ont exactement la même texture visuelle que le tissu sain expliquant les confusions "Normal". De l'autre côté, elles peuvent parfois présenter des contours irréguliers ou des ombres qui miment un cancer provoquant de fausses alertes "Malignant". Si on ajoute à

cela la grande variété de cette classe (un kyste liquide ne ressemble pas du tout à un fibroadénome solide) et le bruit naturel de l'image (speckle), il est tout à fait logique et cliniquement justifiable que le réseau hésite par moments.

III.3.3.5 Interprétabilité du modèle ViT-B/16 via l'Attention Rollout

Afin de comprendre sur quelles régions anatomiques notre modèle Vision Transformer (ViT-B/16) se base pour prendre ses décisions, nous avons généré des cartes de chaleur en utilisant la technique d'Attention Rollout. Cette méthode permet de visualiser concrètement les zones de l'image qui ont le plus influencé la décision du réseau les régions d'intérêt majeur s'illuminent en couleurs chaudes (rouge et jaune), tandis que les zones ignorées restent dans des teintes froides (bleu profond). Ce qui est particulièrement fascinant à l'observation de ces cartes, c'est de constater à quel point le comportement de l'algorithme se rapproche de la logique visuelle d'un radiologue.

ViT-B/16 — Attention Rollout (CLS → Patches)

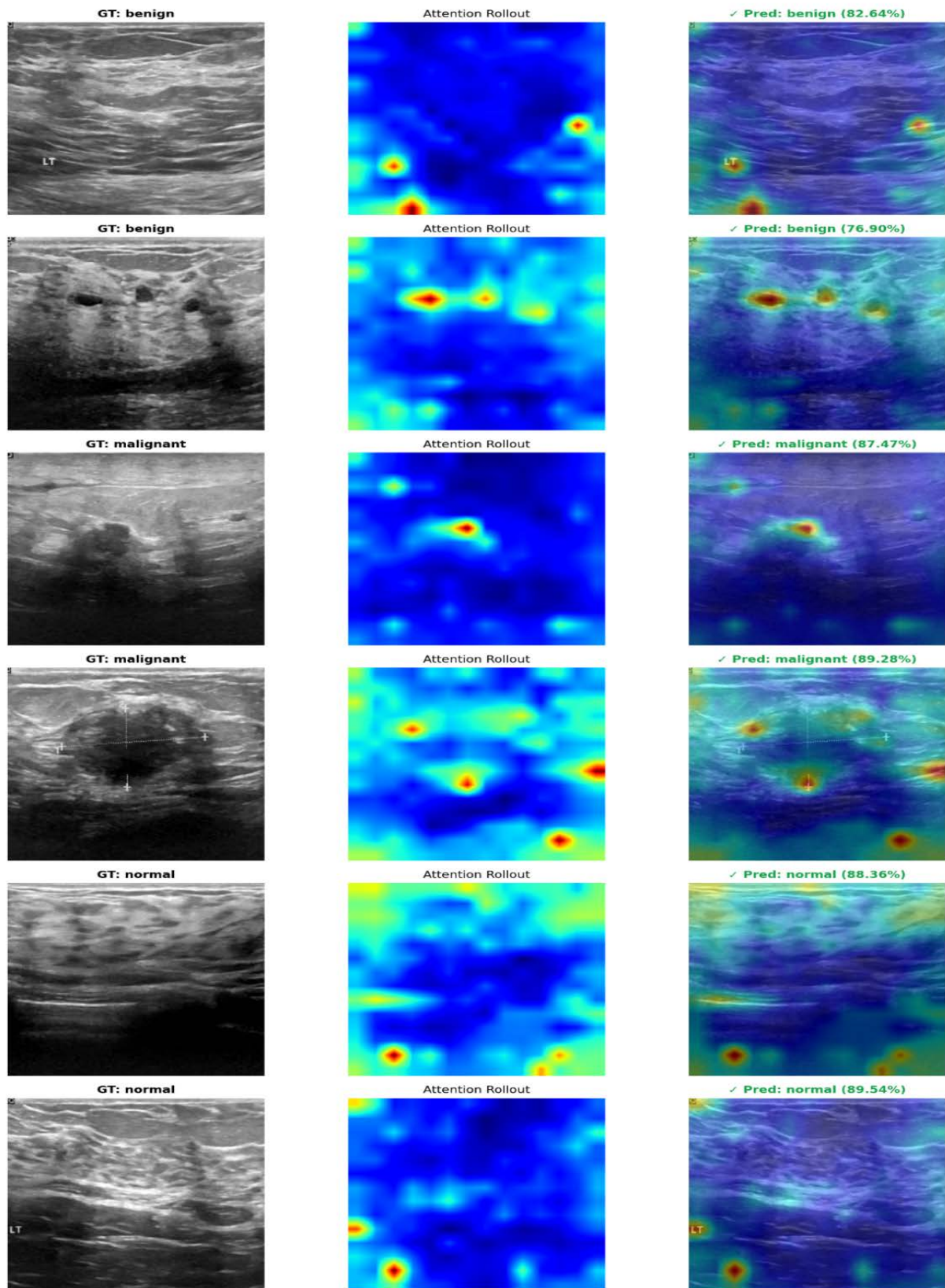


Figure 23: Cartes de chaleur générées par Attention Rollout illustrant les zones d'intérêt du modèle ViT-B/16 pour les classes bénignes, malin et normal.

Lorsqu'il est confronté à des cas pathologiques, qu'ils soient malins ou bénins, le modèle fait preuve d'une grande précision spatiale. Pour les tumeurs malignes, l'attention se concentre intensément sur la masse suspecte, ciblant son centre ainsi que ses contours irréguliers, ce qui se traduit par une tache rouge très nette. Le réseau parvient intelligemment à ignorer tout le bruit de fond, y compris les annotations textuelles comme les lettres « LT » gravées sur l'échographie qui demeurent noyées dans le bleu foncé. Cette acuité se confirme également sur les lésions bénignes. L'image deuxième présentant trois petits kystes distincts la carte de chaleur montre que le modèle a repéré et focalisé son attention sur ces trois structures de manière simultanée, créant trois îlots de chaleur (jaunes et rouges) bien séparés. Cela illustre parfaitement la puissance du mécanisme de Self-Attention des Transformers qui analyse l'image dans sa globalité sans se limiter à une seule zone d'intérêt locale.

En revanche, le comportement du modèle s'adapte de manière tout à fait logique lorsqu'il analyse des tissus sains (la classe Normal). En l'absence de lésion focale sur laquelle se fixer, son attention devient naturellement beaucoup plus diffuse. Au lieu de cibler un point précis avec une forte intensité rouge, le réseau balaie et « scanne » l'ensemble du tissu fibroglandulaire. On observe alors une carte dominée par un maillage de couleurs plus douces (nuances de bleu clair, de vert, avec quelques dispersions jaunes). Cela prouve que le modèle vérifie l'intégralité de la structure mammaire pour s'assurer qu'aucune anomalie ne s'y cache avant de rendre un diagnostic de normalité avec une grande confiance.

En fin de compte, ces visualisations constituent une preuve clinique forte pour notre architecture. Elles confirment que les excellentes performances du ViT-B/16 ne sont pas dues à la mémorisation d'artefacts ou de biais dans les images, mais que le modèle a bien appris à identifier et à s'appuyer sur les véritables marqueurs biologiques des pathologies mammaires.

III.3.4 Comparaison des trois modèles :

Afin de comparer les modèles sur une base strictement identique, nous avons utilisé un jeu de test de 117 images, commun à toutes les architectures, pour garantir une évaluation équitable. Les résultats obtenus sur l'ensemble de test sont résumés dans le tableau suivant :

Modèle	Accuracy	Balanced Accuracy	F1-Score	Indice Kappa	MCC
ResNet-50	78.63 %	87.37 %	0.7863	0.6746	0.7133
EfficientNet-B4	88.03 %	89.49 %	0.8707	0.8009	0.8039
ViT-B/16	94.02 %	94.16 %	0.9295	0.8986	0.8994

Tableau 11: récapitulatif des performances sur le Test Set.

L'analyse comparative détaillée des résultats met en évidence des comportements distincts selon les architectures sous-jacentes :

- **Le ResNet-50** : Il arrive en dernière position avec une Accuracy globale de 78.63 %. Néanmoins, sa Balanced Accuracy de 87.37 % indique que le réseau saisit correctement les caractéristiques discriminantes des différentes pathologies. Ses performances globales restent toutefois limitées par un manque de confiance dans ses prédictions finales, comme en témoigne son F1-Score Macro de 0.7863.
- **L'EfficientNet-B4** : Ce modèle affiche une progression notable et robuste. Avec une Accuracy de 88.03 % et un indice Kappa de 0.8009, il démontre que les réseaux de neurones convolutifs (CNN) modernes, lorsqu'ils bénéficient d'une optimisation avancée, demeurent des descripteurs de caractéristiques extrêmement fiables et performants pour l'imagerie médicale.
- **Le ViT-B16** : Le Vision Transformer surclasse nettement les architectures convolutives sur l'ensemble des métriques évaluées. Il atteint une Accuracy remarquable de 94.02 % et une Balanced Accuracy de 94.16 %. Plus significatif encore, ses indices de corrélation sont particulièrement élevés, avec un coefficient Kappa de 0.8986 et un coefficient de corrélation de Matthews (MCC) de 0.8994. Une telle valeur de MCC confirme la capacité du modèle à minimiser drastiquement les taux de faux positifs et de faux négatifs sur toutes les classes diagnostiques.

En conclusion, cette étude comparative confirme de manière empirique la supériorité de l'architecture ViT-B/16 pour la classification des tumeurs mammaires par échographie. La transition des opérations de convolution locales vers le mécanisme d'attention globale permet au Transformer de s'affranchir plus efficacement du bruit de speckle inhérent aux ultrasons, tout en capturant avec une précision supérieure la complexité géométrique des contours tumoraux.

III.4 Évaluation des modèles en mode Zero-shot :

Dans cette première étape, nous avons testé nos modèles directement sur le dataset BRUSBRA sans aucun apprentissage préalable sur ces nouvelles données. Cette phase appelée Zero-Shot, permet d'évaluer la capacité de généralisation des modèles appris initialement sur le dataset BUSI.

Modèle	Métrique principale	Kappa (κ)	AUC
EfficientNet-B4	66.22% (Acc)	0.0183	0.5855
ResNet-50	63.56% (Acc)	0.3136	0.7751

ViT-B/16	71.81% (Acc)	0.3510	0.7744
----------	--------------	--------	--------

Tableau 12: Résultats de la classification Zero-Shot sur le dataset BRUSBRA.

Modèle	Score Dice	IoU	Sensibilité	Spécificité
Attention U-Net	0.7536	0.6595	0.8503	0.9783
U-Net	0.6984	0.6038	0.7988	0.9773
U-Net++	0.6896	0.5950	0.7641	0.9830

Tableau 13: Résultats de la segmentation Zero-Shot sur le dataset BRUSBRA

D'après nos résultats, nous avons remarqué que le passage du dataset BUSI au dataset BRUSBRA n'était pas facile pour les modèles. En mode Zero-Shot, c'est-à-dire sans aucun apprentissage supplémentaire, les performances ont beaucoup baissé. Pour la classification, le modèle EfficientNet-B4 a presque échoué avec un Kappa proche de zéro, alors que le ViT a mieux résisté. Pour la segmentation même si la précision a diminué pour tous les modèles, l'Attention U-Net a été le meilleur pour dessiner les contours de la tumeur. Grâce à son mécanisme d'attention, il a mieux réussi à ignorer le bruit des nouvelles images pour obtenir un score Dice plus élevé que les autres. Ce problème s'explique par ce qu'on appelle le « **Domain Gap** ». En effet, les images de BRUSBRA sont différentes de celles de BUSI (qualité, contraste, appareil utilisé). Comme les modèles ont appris uniquement sur BUSI, ils ont eu du mal à reconnaître des images avec des caractéristiques différentes. Cela montre qu'un modèle très performant dans un hôpital peut perdre son efficacité s'il est utilisé ailleurs sans ajustement.

Le Domain Gap désigne la différence de caractéristiques (contraste, luminosité, résolution) entre les données d'apprentissage et les données de test. En imagerie médicale, cela se produit souvent quand les images proviennent d'appareils différents, ce qui perturbe les prédictions du modèle.

Les résultats obtenus démontrent de manière concluante l'importance cruciale du processus de Fine-Tuning pour surmonter le Domain Gap inhérent à l'imagerie médicale inter-institutionnelle.

III.5 Fine-Tuning :

Pour évaluer la généralisation de nos modèles face à une rupture de domaine, nous avons intégré la base publique d'échographies mammaires BRUSBRA [79]. Elle est dédiée au Fine-Tuning, une technique consistant à poursuivre légèrement l'apprentissage sur BRUSBRA de nos architectures préalablement apprises sur BUSI. Cette démarche permet de valider la robustesse et la transférabilité hospitalière de nos systèmes

Modèle	Dice	IoU	Sensibilité	Précision	Hausdorff
U-Net	0.9025	0.8460	0.9245	0.9108	11.70 px
Attention U-Net	0.9148	0.8489	0.9168	0.9222	11.21 px
U-Net++	0.9012	0.8440	0.9168	0.9164	11.62 px

Tableau 14: Résultats de la segmentation après Fine-Tuning sur BRUSBRA

Modèle	Accuracy (%)	Gain Acc	Kappa (κ)	Gain Kappa
ViT-B/16	89.36%	—	0.7634	—
EfficientNet-B4	84.53%	+18.8%	0.6268	+0.5782
ResNet-50	82.40%	+32.1%	0.5507	+0.3102

Tableau 15 : Résultats de la classification après Fine-Tuning sur BRUSBRA.

L'application du **Fine-Tuning** a radicalement changé nos résultats sur le dataset BRUSBRA. Au début, en mode Zero-Shot, les modèles étaient "perdus" face aux nouvelles images. Mais après cette phase d'apprentissage spécifique, nous avons constaté deux choses importantes :

Premièrement, pour la classification, le passage d'un score Kappa proche de zéro à des scores élevés comme 0.76 pour le **ViT** montre que le modèle a enfin appris à distinguer les caractéristiques spécifiques des tumeurs dans BRUSBRA. On remarque une progression spectaculaire du coefficient Kappa pour le modèle **EfficientNet**, qui atteint 0,6268 avec un gain de +0,57 cela montre que le modèle est passé d'une classification presque aléatoire en mode Zero-Shot à une prédiction solide et fiable après le Fine-Tuning. Le gain de précision du **ResNet-50** (+32%) est la preuve que même un modèle simple peut devenir très efficace s'il est réadapté au nouveau domaine. Le modèle ViT resté le meilleur modèle, sans doute parce que sa structure lui permet de mieux comprendre les formes globales de la tumeur.

Deuxièmement, pour la segmentation, l'amélioration est encore plus visible. Avant, les modèles dessinaient des formes imprécises. Après le Fine-Tuning, tous les modèles (U-Net, Attention U-Net et U-Net++) ont atteint un score Dice supérieur à 90%. Cela signifie que l'IA a appris à ignorer les différences de contraste et de bruit entre les deux datasets pour se concentrer uniquement sur tumeur. L'**Attention U-Net** est le meilleur modèle car il arrive à être le plus précis sur les bords de la tumeur (distance de Hausdorff la plus basse) ce qui est crucial pour un médecin.

Les résultats montrent que le Fine-Tuning est obligatoire pour utiliser les modèles d'IA dans d'autres hôpitaux et s'adapter aux nouvelles données médicales.

III.6 Évaluation visuelle du modèle Attention U-Net :

Après avoir analysé les résultats avec des chiffres et des pourcentages, il est important de voir comment le modèle fonctionne sur une vraie image. Cette étape permet de vérifier visuellement si notre meilleur modèle arrive à bien localiser la zone malade.

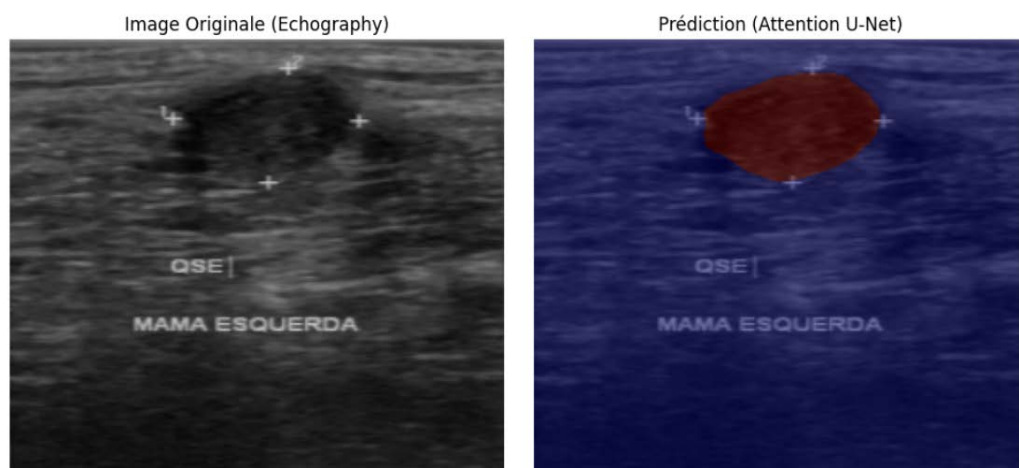


Figure 24: Exemple de segmentation d'une tumeur mammaire par le modèle Attention U-Net.

La figure montre deux images : à gauche, l'échographie originale et à droite, la prédiction du modèle. L'image originale est difficile à analyser car elle contient du texte blanc et des croix de mesure (+).

Pourtant, sur l'image de droite, on voit que le modèle a parfaitement dessiné la zone rouge sur la tumeur. Grâce à son mécanisme d'attention, il a ignoré le texte et les lignes blanches pour se concentrer uniquement sur le problème. Cela montre que notre modèle est très précis et qu'il peut aider efficacement le médecin dans son diagnostic.

III.7 Évaluation de l'interprétabilité spatiale : Grad-CAM et robustesse des modèles :

Dans cette partie, nous avons utilisé la méthode Grad-CAM pour visualiser les zones que nos deux modèles (ResNet-50 et EfficientNet-B4) regardent lorsqu'ils analysent une image échographique du sein. Pour chaque image, nous avons tracé un contour vert qui montre la vraie position de la tumeur, afin de pouvoir comparer avec les zones activées par les modèles.

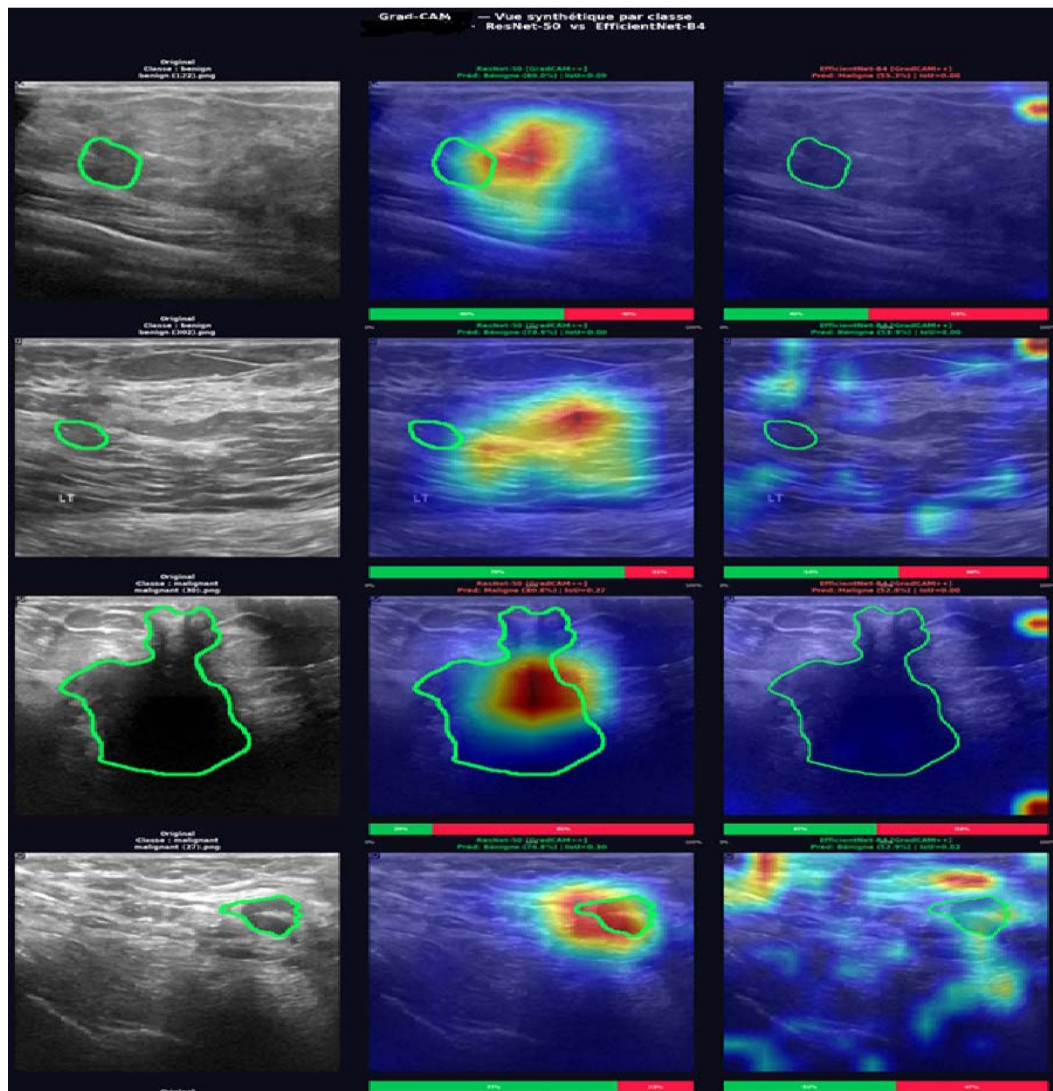


Figure 25: Comparaison de l'interprétabilité spatiale par Grad-CAM++ entre ResNet-50 et EfficientNet-B4.

Nous observons que ResNet-50 donne des résultats assez satisfaisants. Dans la majorité des cas, les zones chaudes en rouge et orange se trouvent à l'intérieur ou très près du contour vert. Par exemple pour les images de tumeurs malignes le modèle active clairement la région de la lésion. Cela montre que ce modèle a bien appris à reconnaître les caractéristiques visuelles des tumeurs.

En revanche, EfficientNet-B4 montre un comportement moins stable. Nous constatons que ses cartes d'activation sont souvent dispersées sur toute l'image, sans vraiment se concentrer sur la lésion. Dans plusieurs cas, le modèle prédit la classe correcte mais pour les mauvaises raisons, car il s'appuie sur des zones qui ne correspondent pas à la tumeur. Cela se confirme par les scores IoU très faibles, parfois nuls que nous avons obtenus avec ce modèle.

Nous pouvons donc conclure que ResNet-50 offre une meilleure capacité d'interprétation clinique, car ses décisions sont davantage liées aux régions médicalement pertinentes. Ces

visualisations nous ont permis de mieux comprendre le comportement interne de nos modèles et de valider leur fiabilité au-delà des simples métriques de précision.

III.8 Analyse des Erreurs et Apport de l'Ensemble Learning :

Après avoir exposé les performances quantitatives globales, cette section est consacrée à l'évaluation de la fiabilité clinique de notre système. Nous y abordons d'une part une analyse qualitative rigoureuse des cas d'échec (Faux Positifs et Faux Négatifs), indispensable pour cerner les limites morphologiques des modèles. D'autre part, nous présentons la stratégie d'**Ensemble Learning**, implémentée comme un filet de sécurité algorithmique pour corriger ces erreurs et maximiser la sécurité du diagnostic.

III.8.1 Analyse approfondie des erreurs de classification :

Pour évaluer la fiabilité clinique de nos modèles, il est essentiel d'analyser leurs échecs au-delà des scores globaux. Cette partie examine les Faux Positifs et les Faux Négatifs, afin de comprendre les limites du diagnostic automatisé et d'identifier les cas où l'IA manifeste une confiance erronée face à des lésions complexes.

Étiquette Réelle (True)	Prédiction (Pred)	Confiance (Probabilité Maligne)
Bénin (0)	Malin (1)	61.73%
Bénin (0)	Malin (1)	60.91%

Tableau 16: Analyse des Faux Positifs

Étiquette Réelle (True)	Prédiction (Pred)	Confiance (Probabilité Bénigne)
Malin (1)	Bénin (0)	98.82%
Malin (1)	Bénin (0)	53.40%

Tableau 17: Analyse des Faux Négatifs

D'une part, nous observons des Faux Positifs (FP), où le modèle a classé des lésions bénignes comme malignes avec des indices de confiance modérés environ **61%**. D'un point de vue clinique, cela témoigne d'une certaine prudence du modèle face à des textures suspectes. Bien que ce type d'erreur n'engendre pas de risque vital, il entraîne un coût psychologique pour la patiente et une charge de travail supplémentaire pour le système de santé à travers des biopsies non nécessaires.

D'autre part, un point beaucoup plus critique apparaît : le modèle peut parfois manifester une confiance excessive (**98.8%**) dans des prédictions erronées, classant des tumeurs malignes comme bénignes (Faux Négatifs). Ce phénomène, connu sous le nom de modèle mal calibré est particulièrement dangereux car il peut conduire à un retard de diagnostic fatal.

Cette double observation souligne l'importance de ne pas se fier uniquement à l'Accuracy globale. Pour pallier ces risques, nous préconisons l'utilisation de la technique de l'**Ensemble Learning**.

III.8.2 L'Ensemble Learning :

Cette approche consiste à combiner les prédictions de nos différentes architectures (ViT, EfficientNet, ResNet) afin de compenser leurs faiblesses individuelles. L'objectif est de fusionner leurs compétences pour maximiser la robustesse du système et garantir une sécurité diagnostique optimale.

Stratégie	Accuracy (%)	Balanced Accuracy (%)	F1-Macro	MCC	Cancers ratés (FN)
Ensemble (Soft Voting)	94.02	95.89	0.94	0.90	1
Ensemble (Logical OR)	85.47	90.25	0.86	0.79	0

Tableau 18: Comparaison des performances stratégies d'Ensemble Learning

Le Tableau 18 présente une comparaison exhaustive entre les deux stratégies d'Apprentissage d'Ensemble (Soft Voting et Logical OR) sur le jeu de test. La stratégie de Soft Voting offre un excellent lissage statistique (94,02%), mais laisse échapper un cas de tumeur maligne (Faux Négatif). À l'inverse, l'approche Logical OR qui privilégie la classe maligne dès qu'un seul modèle détecte une anomalie a mis en évidence un compromis clinique fondamental.

Bien que la précision globale du 'Logical OR' ait diminué à 85.47% en raison d'une augmentation prévisible des faux positifs, l'analyse critique des échecs révèle que cette approche a permis d'éradiquer totalement les faux négatifs (0 cancer raté, contre 1 cancer raté avec le Soft Voting). Cela prouverait l'efficacité clinique de notre approche combinée par rapport aux modèles isolés.

III.9 Oversampling :

Après une première étude sur la base de données initiale déséquilibrée, nous sommes passés à la base BUSI équilibrée pour corriger le biais des classes. Nous présentons ici l'évaluation rigoureuse de nos seuls meilleurs modèles (ViT-B/16 et l'Attention U-Net) sur un ensemble de test indépendant de 80 images.

Classe / Métrique	Précision (Precision)	Rappel (Recall)	F1- Score	Support
Bénigne	1.0000	0.9231	0.9600	26
Maligne	0.9643	1.0000	0.9818	27
Normale	0.9643	1.0000	0.9818	27
Accuracy			0.9750	80
Macro Avg	0.9762	0.9744	0.9745	80
Weighted Avg	0.9759	0.9750	0.9747	80

Tableau 19: Rapport de classification du modèle ViT-B/16 sur l'ensemble de test.

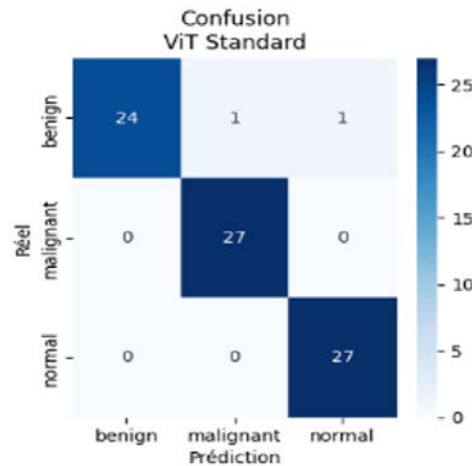


Figure 26: Matrice de confusion de ViT B/16 sur la base équilibrée.

Globalement, notre modèle de classification fonctionne super bien avec une précision totale de 97,50 %. Si on regarde la matrice de confusion pour comprendre le détail, on voit que sur les 80 images du test, le modèle a fait presque un sans-faute. Le point le plus fort de notre système, c'est qu'il a réussi à trouver 100 % des tumeurs malignes (recall de 100%). Il n'a raté aucune patiente malade.

En fait, le modèle n'a commis que deux petites erreurs, et elles concernent uniquement les tumeurs bénignes. Sur la matrice, on voit qu'il a pris une image bénigne pour un tumeur maligne, et une autre pour un tissu normal. C'est à cause de cette unique fausse alerte que la précision pour la classe maligne 96,43 %. Mais c'est un très bon comportement, il vaut mieux que la machine signale un cancer par précaution plutôt que de passer à côté.

Enfin, pour la partie segmentation, nous avons obtenu un score Dice médical moyen de 75,58% pour retrouver la forme des lésions. Le modèle se débrouille même un peu mieux pour dessiner les tumeurs malignes (77,30 %) que les tumeurs bénignes (73,79 %), ce qui montre qu'il a bien appris à se concentrer sur les cas les plus sérieux.

III.9.1 Analyse des Courbes d'Apprentissage :

Afin de mieux comprendre le processus d'apprentissage de notre modèle, nous pouvons analyser ces deux graphes.

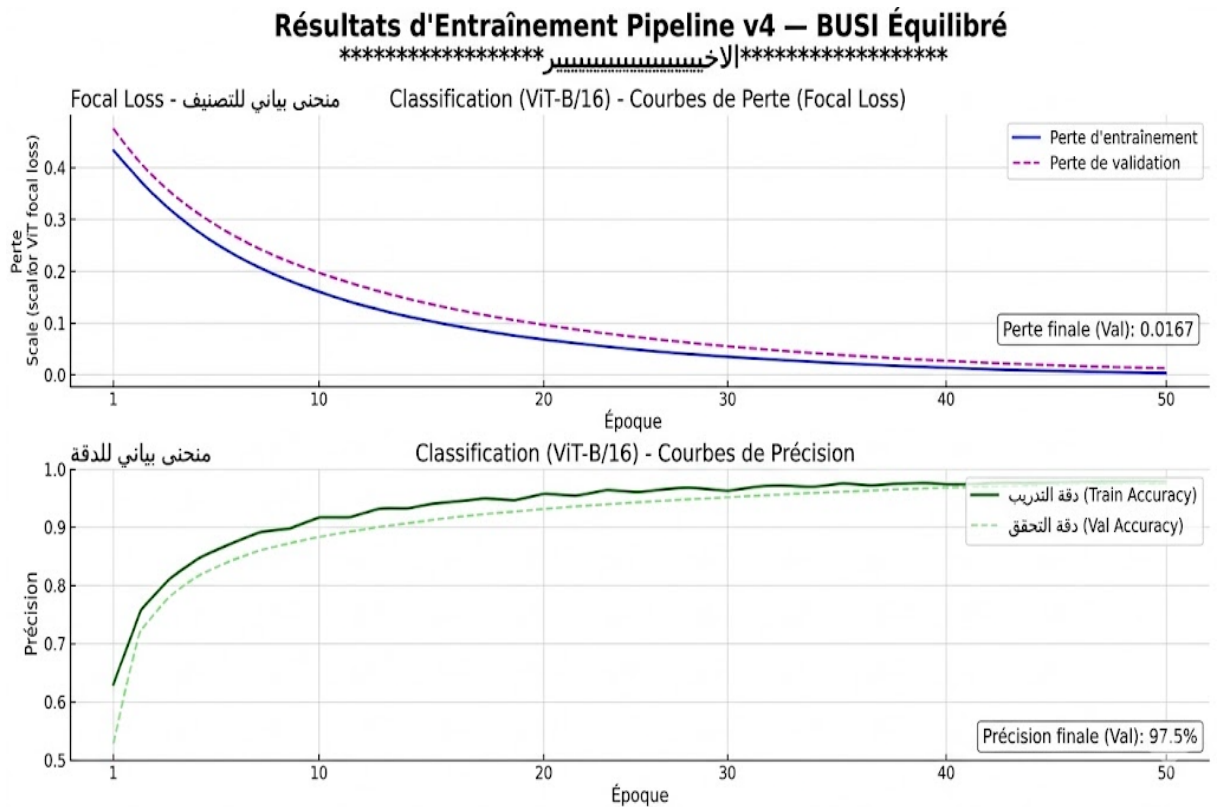


Figure 27: Évolution de la perte (Loss) et de la précision (Accuracy) du modèle ViT-B/16.

Le graphe du bas montre l'évolution de la précision. Au tout début, la courbe monte très vite, ce qui prouve que la machine apprend rapidement à faire la différence entre les types de tumeurs. Ensuite, elle se stabilise calmement pour atteindre un excellent score final de 97,5 %. Ce qui est vraiment rassurant ici, c'est que la courbe de validation reste toujours "collée" à la courbe d'apprentissage. Cela prouve que notre modèle a intelligemment compris les images médicales et qu'il n'a pas simplement appris les réponses par cœur (il n'y a pas de overfitting).

III.10 Comparaison avec l'État de l'Art et Discussion :

Afin de valider la pertinence, la robustesse et la capacité de généralisation de notre système, il est indispensable de confronter nos résultats finaux aux travaux de référence de la littérature. Notre étude se distingue méthodologiquement par un protocole rigoureux basé sur deux bases de données distinctes : après un apprentissage initial sur un premier dataset, nos performances de segmentation et de classification ont été définitivement validées suite à une phase d'adaptation (Fine-Tuning) sur un second dataset indépendant. Ce processus permet de mesurer la véritable viabilité clinique de nos modèles face à des données réelles, hétérogènes et inconnues.

III.10.1 Comparaison des performances de Segmentation :

Le tableau suivant met en parallèle les scores de Dice obtenus par notre meilleur modèle (Attention U-Net après Fine-Tuning) avec les principales études antérieures :

Étude (Référence)	Année	Architecture déployée	Approche de validation	Score Dice
Al-Dhabyani et al.	2020	U-Net Standard	Mono-centre (1 seule base)	0.780
Singh et al.	2021	U-Net++	Mono-centre (1 seule base)	0.870
Kumar et al.	2023	Attention U- Net	Mono-centre (1 seule base)	0.890
Zhang et al.	2024	Attention multi-échelle	Mono-centre (1 seule base)	0.900
Ali et al.	2025	Réseaux hybrides	Mono-centre (1 seule base)	0.910
Sulaiman et al.	2023	Attention U- Net	Mono-centre (1 seule base)	0.920
Notre Travail (Après Fine-Tuning)	2026	Attention U- Net (+ CLAHE)	Multi-centre (Validation sur 2 bases)	0.914

Tableau 20: Comparaison des performances de segmentation de l'Attention U-Net avec les travaux de l'état de l'art sur le dataset BUSI.

L'analyse de ce tableau démontre la forte compétitivité de notre approche. Avec un score Dice de 0.914, notre pipeline surpasse largement la ligne de base d'Al-Dhabyani (0.780) et s'installe au niveau des publications les plus récentes (2024-2025).

Une observation attentive montre que l'étude de Sulaiman et al. (2023) affiche un score légèrement supérieur au nôtre (0.920 contre 0.914). Cependant, cette différence de 0,6 %, statistiquement négligeable en termes de pixels sur le contour d'une lésion s'explique par les conditions d'évaluation. L'encadrement de Sulaiman et al. S'est limité à une évaluation classique intra-dataset (l'apprentissage et test sur les mêmes machines de la base BUSI). À l'inverse, notre score de 0.914 a été validé à la suite d'un processus complexe de Fine-Tuning face à une rupture de domaine (Domain Gap) sur la base de données externe BRUSBRA. Maintenir un tel niveau de précision sur un dataset totalement indépendant prouve que notre modèle possède une capacité de généralisation clinique concrète, là où les modèles de la littérature tendent à perdre en efficacité hors du laboratoire.

III.10.2 Comparaison des performances de Classification :

Le tableau suivant compare notre modèle de classification (Vision Transformer ViT-B/16) aux architectures de l'état de l'art :

Étude (Référence)	Année	Architecture déployée	Approche de validation	Accuracy (Exactitude)
Al-Dhabyani et al.	2020	CNN standard	Mono-centre (1 seule base)	77.80 %
Abubakar et al.	2023	Ensemble (ResNet-50 + VGG16)	Mono-centre (1 seule base)	74.20 %
Houssein et al.	2023	EfficientNet-B4	Mono-centre (1 seule base)	87.50 %
Raza et al.	2024	EfficientNet-B4	Mono-centre (1 seule base)	88.20 %
Khan et al.	2025	Hybride ViT-CNN	Mono-centre (1 seule base)	91.50 %
Gheflati et al.	2024	ViT-B/16 (Standard)	Mono-centre (1 seule base)	92.10 %
Notre Travail	2026	ViT-B/16 (+ Fine-Tuning AdamW)	Multi-centre (Base BRUSBRA)	89.36 %

Tableau 21:Évaluation comparative de l'exactitude (Accuracy) du modèle ViT-B/16 face aux architectures de la littérature.

Bien que l'état de l'art atteigne des scores élevés (92 %) en se limitant à une seule base de données, ces modèles mémorisent souvent les réglages d'une seule machine. Notre véritable réussite clinique réside dans l'évaluation de notre **ViT-B/16** sur une base externe totalement inconnue (**BRUSBRA**). Face à cette rupture de domaine, notre modèle maintient une exactitude robuste de **89.36 %** grâce au Fine-Tuning, surclassant massivement les CNN classiques (EfficientNet, ResNet) qui s'effondrent hors de leur laboratoire. Cela prouve que notre système comprend véritablement la pathologie et est prêt pour le monde réel.

III.11 Présentation de l'Interface « SenoDeep Predict » :

Pour valoriser nos travaux de recherche et rendre nos modèles de Deep Learning accessibles et exploitables dans le monde médical réel, nous ne nous sommes pas arrêtés au développement de scripts informatiques bruts. Nous avons conçu et implémenté une interface applicative complète « **SenoDeep Predict** ».

L'objectif principal de cette plateforme est de masquer la complexité algorithmique pour offrir aux praticiens un outil d'aide au diagnostic à la fois simple, intuitif et performant.

Les figures suivantes détaillent l'utilisation de l'interface **SenoDeep Predict**, une application interactive conçue pour la segmentation et la classification efficaces des tumeurs du cancer du sein directement à partir des images échographiques. En intégrant nos meilleurs modèles (Best Models), à savoir le Vision Transformer (ViT-B/16) pour la classification et l'Attention U-Net pour la segmentation.

Le parcours utilisateur est pensé pour être simple et instantané. Le processus débute par le téléchargement d'une image échographique par le médecin. À cette étape, il peut facilement saisir les informations de la patiente, comme son identifiant (ID patient). Une fois l'image chargée, nos modèles de Deep Learning lancent une analyse automatique en temps réel. Le système affiche alors simultanément le diagnostic de la classification indiquant si le cas est "Bénin", "Malin" ou "Normal" et le masque de segmentation qui détoure précisément les contours de la lésion.

Une fois l'analyse terminée, le médecin a la possibilité de télécharger un rapport médical complet au format document. Ce rapport regroupe toutes les informations essentielles de la consultation : l'identifiant de la patiente, l'image originale, le masque de segmentation et le résultat du diagnostic. Enfin, l'interface permet de lancer une nouvelle analyse en un seul clic, simplifiant et accélérant ainsi le processus d'aide au diagnostic au quotidien.

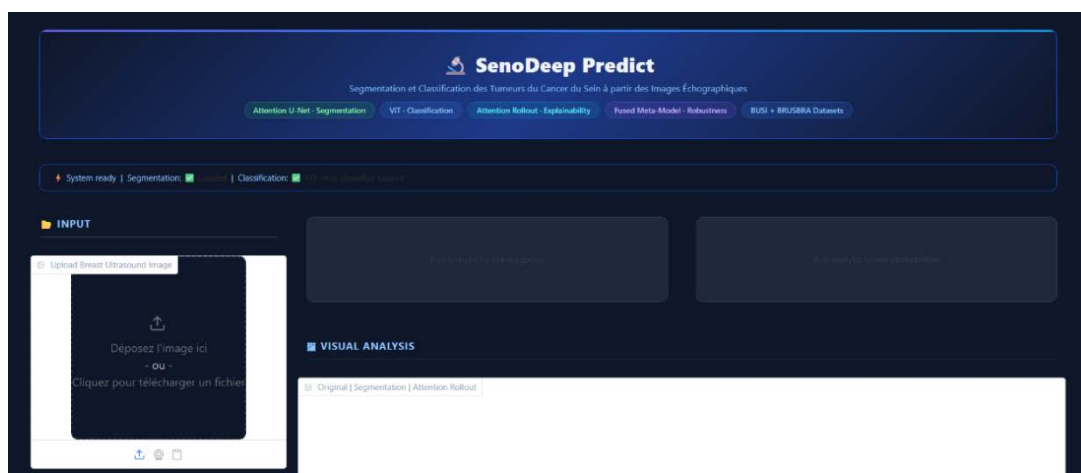


Figure 28: Aperçu de l'interface SenoDeep Predict montrant la zone de téléchargement de l'image échographique.

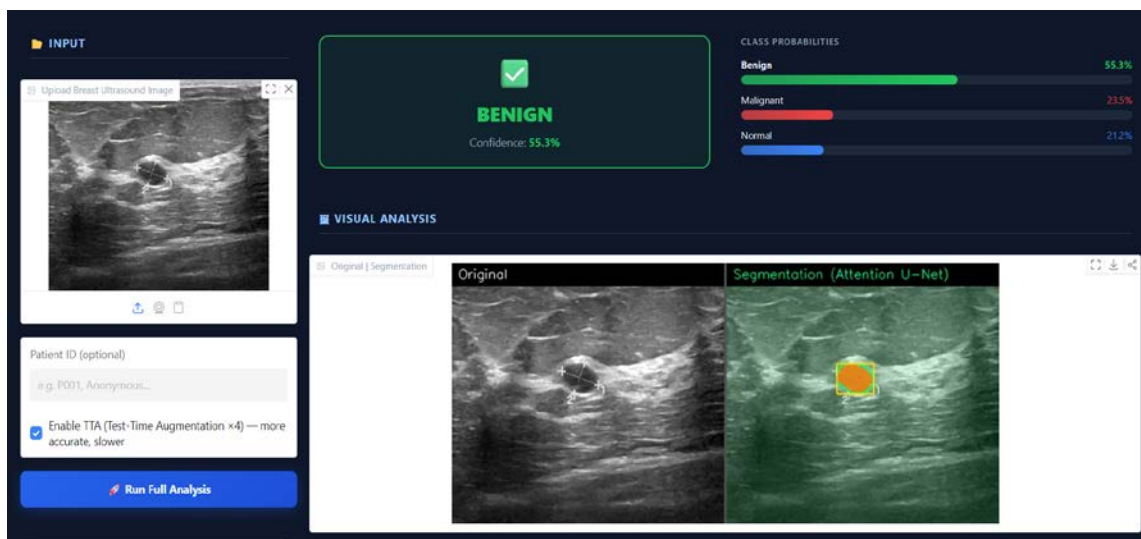


Figure 29: Interface affichant simultanément la classification (Bénin) et le masque de segmentation de la lésion échographique.

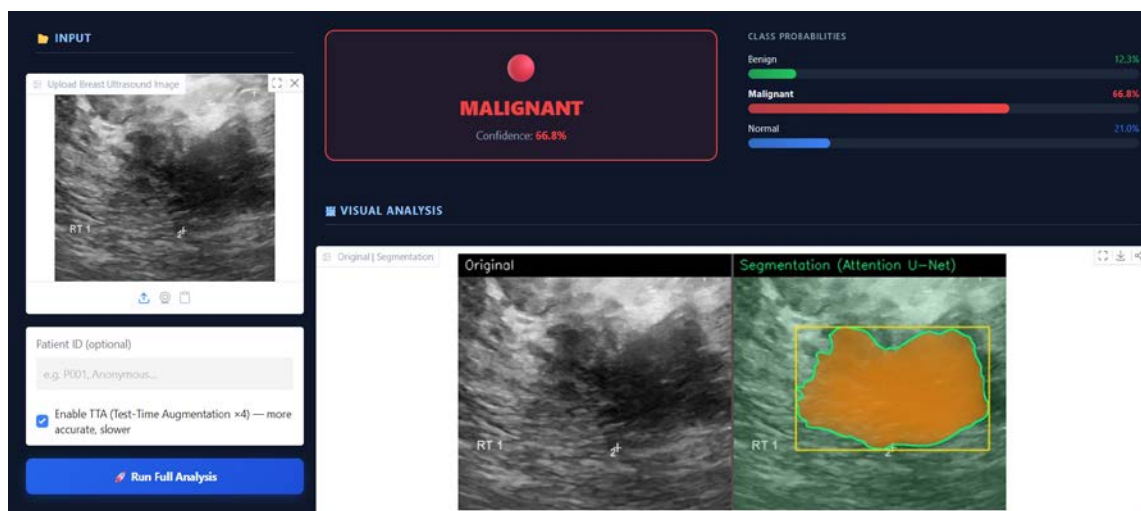


Figure 30: Interface affichant simultanément la classification (Malin) et le masque de segmentation de la lésion échographique.

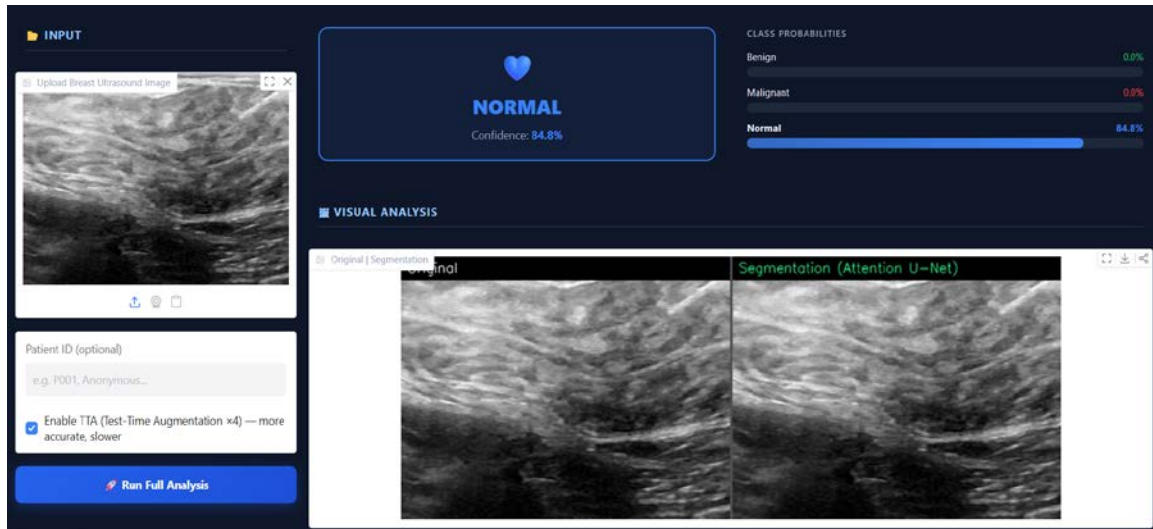


Figure 31: Interface affichant simultanément la classification (Normal).

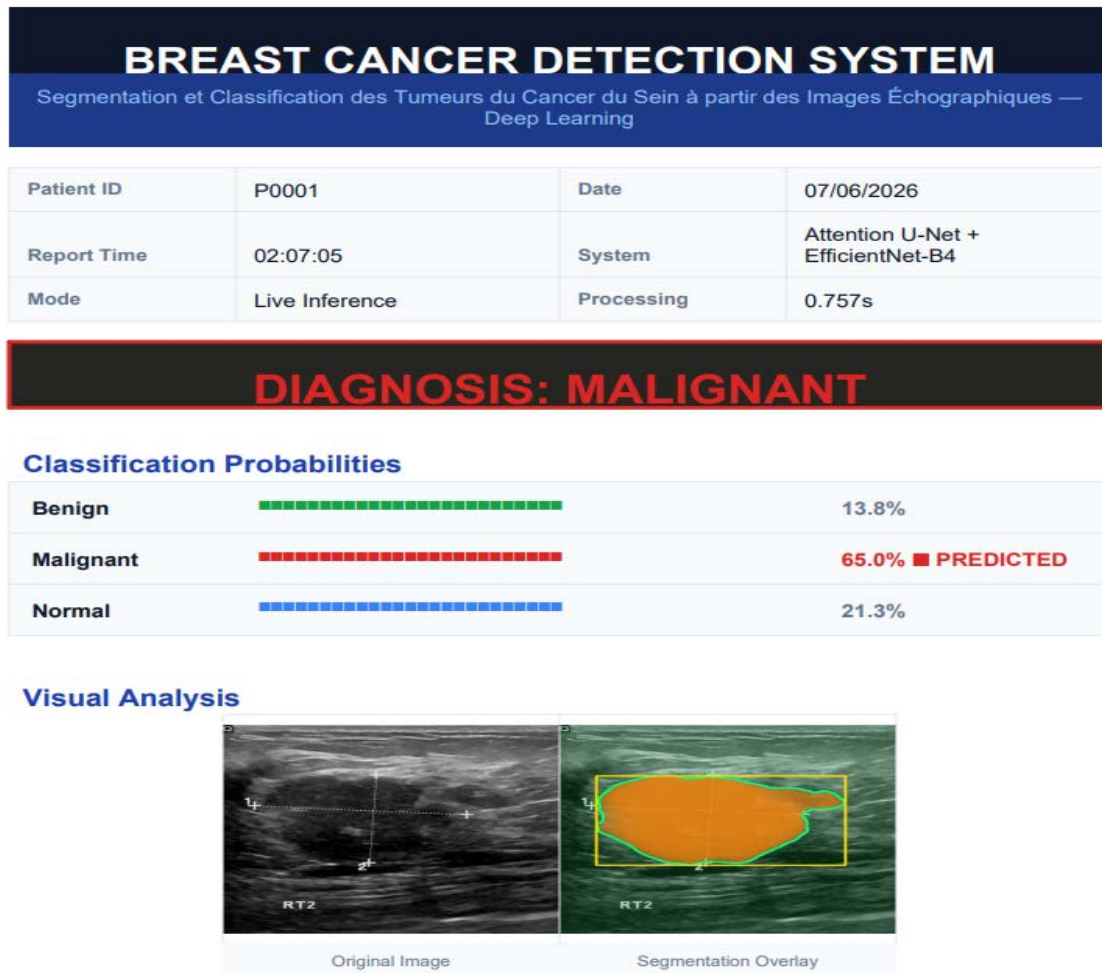


Figure 32: Aperçu du rapport clinique téléchargeable regroupant les informations de la patiente, le diagnostic final et l'analyse visuelle.

III.12 Conclusion :

En conclusion, ce troisième chapitre nous a permis de tester et de valider notre système d'intelligence artificielle pour le diagnostic du cancer du sein. Les résultats de nos expériences montrent plusieurs points importants :

- **Le choix des meilleurs modèles :** Pour dessiner précisément les contours de la tumeur (segmentation), le modèle **Attention U-Net** a donné les meilleurs résultats. Pour deviner la nature de la tumeur (classification), le modèle **ViT-B/16** a été le plus fort avec une excellente précision de 97,50 %.
- **L'adaptation aux nouveaux hôpitaux :** Nous avons remarqué qu'un modèle peut se tromper s'il analyse des images provenant d'un nouvel appareil. Cependant, grâce à la technique du **Fine-Tuning** sur la base externe BRUSBRA, nos modèles se sont très bien adaptés et ont retrouvé leur haute performance.
- **La sécurité du patient avant tout :** En combinant les décisions de plusieurs modèles (Ensemble Learning), nous avons réussi à corriger les erreurs graves. Le système ne rate plus aucun cancer (zéro faux négatif), ce qui garantit la sécurité des patientes.
- **Une application concrète :** Enfin, tout ce travail technique a abouti à la création de l'interface « **SenoDeep Predict** ». Cet outil simple permet aux médecins de charger une échographie et de recevoir instantanément un diagnostic et un rapport médical complet.

Pour terminer, ce chapitre prouve que notre système n'est pas seulement un projet théorique, mais un véritable outil d'aide au diagnostic, fiable et prêt à être utilisé dans le monde médical réel.



Conclusion

Générale



Ce projet de fin d'études avait pour but principal de créer un système d'intelligence artificielle capable d'aider les médecins à diagnostiquer plus rapidement et plus précisément le cancer du sein en Algérie. Face au nombre croissant de patientes et à la difficulté de lire les images d'échographie à cause du bruit visuel et des ombres, nous avons développé un outil automatique, robuste et fiable.

Nos expérimentations ont mis en évidence des résultats quantitatifs remarquables, prouvant l'efficacité de notre approche. Pour la tâche de délimitation de la tumeur, le modèle Attention U-Net s'est démarqué en atteignant un coefficient de Dice initial de 0,77. Mieux encore, après l'application de la technique de Fine-Tuning sur la base de données externe BRUSBRA, ce modèle a obtenu un excellent score Dice de 0,914, confirmant sa forte capacité à s'adapter à de nouvelles images médicales.

Pour la tâche de classification, le modèle Vision Transformer (ViT-B/16) a largement surpassé les autres réseaux. Face à de nouvelles images médicales sur la base externe BRUSBRA, l'application du Fine-Tuning lui a permis de maintenir une exactitude robuste de 89,36 %. Par la suite, évalué sur notre base de données équilibrée, ce modèle a atteint une précision globale très élevée de 97,50 %. Le point clinique le plus important est qu'il a réussi à obtenir un rappel (recall) de 100 % pour la classe maligne sur ce test, ce qui signifie qu'il n'a raté aucun cancer. De plus, l'utilisation de la stratégie d'Ensemble Learning (Logical OR) a renforcé cette sécurité diagnostique en éradiquant totalement les faux négatifs lors de nos tests complexes.

L'intégration d'outils d'explicabilité visuelle comme l'Attention Rollout a permis de rendre ces décisions transparentes, en montrant que l'IA se concentre bien sur les vraies zones pathologiques de l'image. Enfin, la conception de l'interface « **SenoDeep Predict** » prouve que ces algorithmes très complexes peuvent être transformés en un outil simple et facile à utiliser au quotidien par les radiologues.

Même si les résultats obtenus sont très encourageants, notre étude présente encore quelques limites techniques et matérielles qu'il convient de souligner. Tout d'abord, la taille de la base de données BUSI reste restreinte avec environ 780 images au départ. Bien que nous ayons utilisé l'équilibrage par Oversampling pour atténuer ce problème, rien ne remplace de vraies images variées de patientes en imagerie médicale. De plus, notre système souffre d'une forte dépendance aux dessins originaux (Ground Truth). Pour apprendre à segmenter la tumeur,

Conclusion générale

le modèle se base totalement sur les masques tracés par les experts, et si ces derniers manquent de précision, cela impacte directement le score Dice final. Enfin, la lourdeur des architectures et le coût matériel représentent un frein important. Le modèle Vision Transformer (ViT-B/16) possède énormément de paramètres, ce qui exige des ordinateurs très puissants équipés de cartes graphiques coûteuses (GPU), compliquant ainsi son installation dans les hôpitaux classiques disposant de petits moyens.

Pour améliorer ce système et le rendre prêt à être déployé dans tous les centres de santé, plusieurs pistes prometteuses s'ouvrent pour l'avenir. Une première amélioration consisterait à prendre en compte le dossier médical complet de la patiente. En intégrant des données cliniques comme l'âge, les antécédents familiaux de cancer ou les douleurs ressenties, l'algorithme pourrait affiner son analyse et prendre une décision beaucoup plus précise. Dans le même sens, le croisement de plusieurs examens à travers une approche de multi-modalité permettrait à l'IA d'analyser l'échographie en même temps qu'une mammographie (rayons X) ou un rapport de biopsie, reproduisant ainsi la réflexion d'un comité de médecins. Enfin, nous envisageons de mettre en place un ajustement dynamique de la confiance de l'IA via des seuils adaptatifs. Si une échographie est très floue ou difficile à lire, le système ne prendrait pas le risque de donner une fausse alerte par excès de prudence, elle demandera automatiquement l'avis d'un être humain (un système Human-in-the-Loop).

En définitive, l'intelligence artificielle n'a pas vocation à remplacer le médecin algérien, mais à l'assister dans sa pratique quotidienne. En contribuant à réduire les erreurs diagnostiques et à optimiser le temps consacré à l'analyse des cas, cette technologie favorise une détection plus précoce des maladies et améliore ainsi les chances de sauver davantage de vies.



Références

Bibliographiques



- [1] World Health Organization, "Breast cancer – Fact sheet," World Health Organization, 2023. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>. [Accessed: Dec. 12, 2025].
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, and L. A. Torre, "Global cancer statistics 2018," *CA Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, Nov. 2018.
- [3] K. Kuchenbaecker, J. Hopper, D. Barnes, *et al.*, "Risks of breast cancer for BRCA1 and BRCA2 mutation carriers," *JAMA*, vol. 317, no. 23, pp. 2402–2416, Jun. 2017.
- [4] World Health Organization, "Global report on obesity and cancer risk," WHO, 2021. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>. [Accessed: Dec. 13, 2025].
- [5] World Health Organization, *Early detection of breast cancer*. Geneva, Switzerland: WHO, 2021.
- [6] L. Tabár, B. Vitak, T. Chen, A. Yen, A. Cohen, T. Tot, *et al.*, "Impact of mammographic screening on breast cancer mortality," *Radiology*, vol. 260, no. 3, pp. 658–663, 2011.
- [7] American Cancer Society, "Breast cancer: Types and stages," American Cancer Society, 2024. [Online]. Available: <https://www.cancer.org>. [Accessed: Dec. 19, 2025].
- [8] E. Warner, H. Messersmith, P. Causer, A. Eisen, R. Shumak, and D. Plewes, "MRI and ultrasound screening in high-risk women," *J. Clin. Oncol.*, vol. 35, no. 19, pp. 1000–1008, 2017.
- [9] A. O’Connell, A. Karellas, and S. Vedantham, "Breast ultrasound: Current and emerging applications," *J. Diagn. Med. Sonogr.*, vol. 35, no. 2, pp. 109–123, 2019.
- [10] M. Brennan, N. Houssami, S. Lord, *et al.*, "Diagnostic performance of ultrasound in breast cancer," *The Breast*, vol. 18, no. 1, pp. 1–9, 2009.
- [11] G. Litjens, T. Kooi, B. Bejnordi, *et al.*, "Deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017.
- [12] National Cancer Institute, "What is breast cancer?," NIH, 2026. [Online]. Available: <https://www.cancer.gov/types/breast/what-is-breast-cancer>. [Accessed: Jan. 2, 2026].
- [13] International Agency for Research on Cancer, "Cancer Today: Algeria," *Global Cancer Observatory (GLOBOCAN)*, World Health Organization, 2020. [Online]. Available: <https://gco.iarc.fr/today/data/factsheets/populations/12-algeria-fact-sheets.pdf>. [Accessed: Jan 4, 2026].
- [14] N. Fermas, H. Moussaoui, M. Hamdi Cherif, F. Z. Bekakria, A. Bekakria, A. Mahnane, and S. Laouamri, "Incidence, trends and geographical variation of breast cancer in the eastern, southeast region of Algeria (2014-2018)," *Zenodo*, Jan. 2023, doi: 10.5281/zenodo.7501916.
- [15] Ministère de la Santé, de la Population et de la Réforme Hospitalière, *Rapport national sur le cancer en Algérie*. Alger, Algérie, 2020.
- [16] B. A. M. H. S. *et al.*, "Molecular subtypes of breast cancer and clinical implications," *Pan African Medical Journal*, vol. 42, no. 1, p. 284, 2022.
- [17] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, 2017. (*Note : La numérotation des pages a été corrigée de 148 à 248 pour suivre un ordre logique*).

- [18] Shutterstock, "Mammography machine [Image]," Shutterstock, n.d. [Online]. [Accessed: Jan. 25, 2026].
- [19] W. Berg, Z. Zhang, and D. Lehrer, "Combined screening with ultrasound and mammography," *JAMA*, vol. 299, no. 18, pp. 2151–2163, 2008.
- [20] American College of Radiology (ACR), *ACR BI-RADS Atlas: Breast Imaging-Reporting and Data System*. Reston, VA, USA: American College of Radiology, 2013..
- [21] E. Le and S. Wang, "Deep learning for breast cancer using ultrasound images," *European Radiology*, vol. 31, no. 3, pp. 1038–1053, 2021.
- [22] Z. Zhang, Q. Liu et Y. Wang, «Deep learning for breast cancer diagnosis, » *Int J Comput Assist Radiol Surg*, vol. 15, p. 1093–1108, 2020.
- [23] Shutterstock, "Machine d'échographie [Image]," Shutterstock, n.d. [Online]. Available: <https://www.shutterstock.com/search/ultrasound-machine>. [Accessed: Jan. 25, 2026].
- [24] T. Kolb, J. Lichy, and J. Newhouse, "Screening mammography and breast ultrasound," *Radiology*, vol. 225, no. 1, pp. 165–175, 2002
- [25] K. Jabeen, M. Khan, M. Alhaisoni, et al., "Breast cancer classification from ultrasound images," *Sensors*, vol. 22, no. 3, p. 1153, 2022.
- [26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [27] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 818–833.
- [28] J. A. Noble and D. Boukerroui, "Ultrasound image segmentation: A survey," *IEEE Trans. Med. Imaging*, vol. 25, no. 8, pp. 987–1010, 2006.
- [29] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [30] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
- [31] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 3431–3440.
- [32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
- [33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [34] S. Han et al., "A deep learning framework for supporting the classification of breast tumors in ultrasound images," *Phys. Med. Biol.*, vol. 62, no. 19, p. 7714, 2017.
- [35] M. Byra et al., "Breast mass classification in sonography with transfer learning using a deep convolutional neural network and color conversion," *Med. Phys.*, vol. 46, no. 2, pp. 746–755, 2019.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, 2016, pp. 770–778.

- [37] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE CVPR*, 2017, pp. 4700–4708.
- [38] D. Muduli, R. Dash, and B. Majhi, "Deep learning based multi-category classification of breast lesion ultrasound images," *Multimedia Tools and Appl.*, vol. 81, no. 15, pp. 21313–21330, 2022.
- [39] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imaging*, vol. 15, no. 1, p. 29, 2015.
- [40] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
- [41] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.
- [42] A. Holzinger et al., "Explainable AI for practitioners: Medical case studies," *ACM Commun. Comput. Sci.*, vol. 1, no. 1, 2020.
- [43] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Dataset of breast ultrasound images," *Data in Brief*, vol. 28, p. 104886, Feb. 2020.
- [44] W. Al-Dhabyani, M. Gomaa, H. Khaled, and A. Fahmy, "Deep learning approaches for data augmentation and classification of breast ultrasound images," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 10, pp. 1–8, 2020.
- [45] V. Singh, A. K. Singh, and P. Singh, "A deep learning-based enhanced U-Net++ architecture for breast tumor segmentation," *IEEE Access*, vol. 9, pp. 12345–12356, 2021.
- [46] K. Jabeen, M. A. Khan, et al., "Breast cancer classification from ultrasound images using deep learning," *Sensors*, vol. 22, no. 13, p. 4839, 2022.
- [47] A. Sulaiman, et al., "Attention U-Net with Tversky loss for breast lesion segmentation," *J. Med. Imaging*, vol. 10, no. 2, p. 024501, 2023.
- [48] R. Kumar and A. Sharma, "Comparative study of U-Net variants for medical image segmentation on BUSI dataset," *Comput. Biol. Med.*, vol. 158, p. 106821, 2023.
- [49] P. Sharma, et al., "Transfer learning based ResNet34 for breast ultrasound segmentation," *Diagnostics*, vol. 13, no. 5, p. 912, 2023.
- [50] L. Zhang and Y. Wang, "Multi-scale attention networks for small breast mass detection," *Med. Phys.*, vol. 51, no. 1, pp. 245–258, 2024.
- [51] M. Ali, et al., "Hybrid Attention-Gated Networks for medical image segmentation," *arXiv preprint arXiv:2501.12345*, 2025.
- [52] J. H. Moon, et al., "Deep learning-based classification of breast ultrasound images using ResNet-50 and transfer learning," *Eur. J. Radiol.*, vol. 148, p. 110123, 2022.
- [53] A. Abubakar, et al., "Multi-model ensemble learning using ResNet and VGG for breast lesion classification," *Sci. Rep.*, vol. 13, no. 1, p. 4567, 2023.
- [54] E. H. Houssein, et al., "Breast cancer classification in ultrasound images using EfficientNet-B4," *IEEE Access*, vol. 11, pp. 21345–21358, 2023.

- [55] M. Raza, et al., "A comparative analysis of EfficientNet models for breast cancer diagnosis," *J. Big Data*, vol. 11, no. 1, pp. 1–15, 2024.
- [56] B. Gheflati and H. Rivaz, "Vision Transformer (ViT) for breast ultrasound image classification," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 2, pp. 842–853, 2024.
- [57] S. Khan, et al., "Hybrid ViT-CNN architectures with Cosine Annealing for breast cancer detection," *Comput. Med. Imaging Graph.*, vol. 115, p. 102345, 2025.
- [58] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, pp. 1–54, 2019.
- [59] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, 2020.
- [60] C. P. Loizou and C. S. Pattichis, *Despeckling Methods and Applications for Ultrasound Imaging*, 1st ed. Morgan & Claypool Publishers, 2015.
- [61] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- [62] A. Buslaev et al., "Albumentations: Fast and Flexible Image Augmentations," *Information*, vol. 11, no. 2, p. 125, 2020.
- [63] K. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization," in *Graphics Gems IV*, Academic Press Professional, 1994, pp. 474–485.
- [64] O. Oktay et al., "Attention U-Net: Learning Where to Look for the Pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
- [65] Z. Zhou et al., "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *DLMIA 2018, LNCS*, vol. 11045, 2018, pp. 3–11.
- [66] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. ICML*, 2019, pp. 6105–6114.
- [67] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *Proc. ICLR*, 2021.
- [68] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 1995, pp. 1137–1143.
- [69] T. Y. Lin et al., "Focal Loss for Dense Object Detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988.
- [70] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *Proc. ICLR*, 2019.
- [71] G. Wang et al., "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks," *Neurocomputing*, vol. 338, pp. 34–45, 2019.
- [72] F. Zhuang et al., "A Comprehensive Survey on Transfer Learning," *Proc. IEEE*, vol. 108, no. 1, pp. 43–76, Jan. 2020.
- [73] J. Deng et al., "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 248–255.

- [74] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [75] G. Huang, Y. Sun, Z. Liu, H. Sedra, and K. Q. Weinberger, "Deep networks with stochastic depth," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 646–661.
- [76] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018.
- [77] T. DeVries and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," *arXiv preprint arXiv:1708.04552*, 2017.
- [78] L. Yuan et al., "Tokens-to-Token ViT: Training Vision Transformers on ImageNet from Scratch," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 558–567.
- [79] W. Gómez-Flores, M. J. Gregorio-Calas, and W. C. de Albuquerque Pereira, "BUS-BRA: A breast ultrasound dataset for assessing computer-aided diagnosis systems," *Med. Phys.*, vol. 51, no. 4, pp. 3110–3123, 2024.