

الجمهورية الجزائرية الديمقراطية الشعبية
République algérienne démocratique et populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique
جامعة عين تموشنت بلحاج بوشعيب
Université –Aïn Témouchent– Belhadj Bouchaib
Faculté des Sciences et de Technologie
Département Math-Informatique



Projet de Fin d'Études

Pour l'obtention du diplôme de Master en : **CYSIA**
Domaine : **Mathématiques et Informatique**
Filière : **Informatique**
Spécialité : **Cyber sécurité et intelligence artificielle**

Thème

Détection et classification de nodules thyroïdiens bénins et malins sur images échographiques par approches hybrides de Deep Learning

Présenté par :

- 1) Bentahar Chaimaa
- 2) Benbahi Asma

Devant le jury composé de :

Dr Chaoui khadidja	MCB	UAT.B.B (Ain Témouchent)	Présidente
Dr Bendiabdallah Med Hakim	MCA	UAT.B.B (Ain Témouchent)	Examinateur
Dr Beddad Fatima	MCB	UAT.B.B (Ain Témouchent)	Encadrante
Dr Bousahba Nassima	MCA	U.H.B.C (Chlef)	Co-Encadrante

Année Universitaire 2025/2026

Remerciement



Avant toute chose, nous élevons nos louanges à Dieu Tout-Puissant, Maître des destinées, pour Ses innombrables bienfaits et pour nous avoir accordé la santé, la force, la patience et la persévérance tout au long de ce parcours afin de mener à bien ce travail. Sans Sa miséricorde et Sa grâce, rien de tout cela n'aurait été possible. El Hamdulillah !

Nous exprimons nos sincères remerciements à notre encadrante, Dr. BEDDAD Fatima (UAT.B.B), pour son accompagnement, son soutien constant, ses encouragements, ses conseils avisés et ses précieux commentaires tout au long de ce travail.

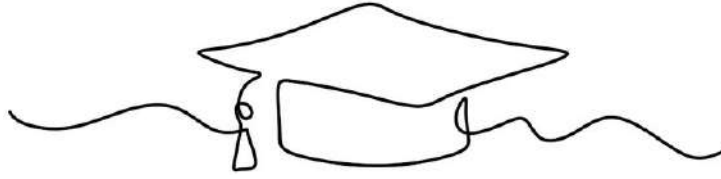
Nous adressons également notre profonde gratitude à notre co-encadrante, Dr. BOUSAHBA Nassima (U.H.B.C), pour sa disponibilité, ses précieux conseils, son soutien et sa contribution essentielle à l'aboutissement de ce projet.

Nous remercions chaleureusement Dr. CHAOUI khadidja, présidente du jury d'examen, ainsi que Dr. BENDIABDALLAH Med Hakim, examinateur, pour le temps qu'ils ont consacré à l'évaluation de ce travail et pour leurs remarques et suggestions constructives.

Nous exprimons notre amour et notre plus profonde gratitude à nos familles, véritables sources de soutien et d'affection. Leur présence, leurs encouragements constants, leur confiance et leurs sacrifices discrets ont constitué une source inépuisable de courage, de sérénité et de paix tout au long de ce parcours.

Nous tenons également à exprimer notre plus profonde gratitude au Dr Lacarne.I, spécialiste en Endocrinologie et Diabétologie, ainsi qu'au Dr Benzeguir Abdelkader pour leur accueil, leur disponibilité, leur précieuse collaboration ainsi que pour les informations et conseils pertinents qu'ils nous ont généreusement prodigués.

Enfin, nous souhaitons exprimer notre profonde gratitude à toutes celles et tous ceux qui, de près ou de loin, ont contribué à la réalisation de ce travail par leur présence, leurs encouragements ou leur soutien. À toutes ces personnes leur contribution, même discrète, a été précieuse et mérite toute notre sincère reconnaissance ainsi que notre profond respect.



Dédicace

Avant tout, je rends grâce à Allah, le Tout-Puissant, qui m'a donné la force, la patience et la foi nécessaires pour mener toutes les épreuves et atteindre cet objectif.

À moi-même

À celle qui n'a jamais abandonné, aux jours où j'ai pleuré sans trouver personne à mes côtés, aux moments où j'ai dû me tenir debout seule face à la vie. Même dans la maladie et les épreuves, j'ai continué à avancer malgré tout. Je dédie ce succès à ma propre persévérance, en hommage à chaque effort fourni pour atteindre ce but.

À ma mère

La femme qui a porté mes rêves avant même que je sache les rêver. Merci pour ton amour infini, tes sacrifices et ton soutien sans limites. Sans toi, rien de tout cela n'aurait été possible.

À mon père

Pilier fondamental de ma vie, symbole de force.
Merci pour ta présence constante, ta confiance absolue quand je doutais de moi, ton travail acharné et ton soutien discret mais constant.

À mes chères sœurs Sarra et Imane

Mes premières amies et les plus beaux soutiens,
Pour leur affection, leurs encouragements dans chaque étape de ma vie et les beaux moments partagés.

À mes chers frères Habib et Islam

Merci pour votre encouragement et votre protection et votre soutien inconditionnel et pour avoir toujours été là pour moi.

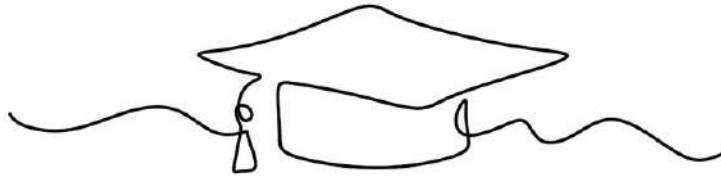
À ma chère nièce Maïssa

Petite lumière qui illumine ma vie.
Merci pour ta douceur, ton sourire et la joie que tu apportes sans le savoir.

À mes amies, Chaimaa, Wissal, Sara, Aya et Amira

Ma famille de cœur. Merci pour vos encouragements, vos visites et ces moments partagés qui ont illuminé mes journées de solitude.

Asma



Dédicace

Je dédie ce modeste travail, le fruit de plusieurs années d'études et d'efforts :

En tout premier lieu, je remercie Allah, le tout puissant, de m'avoir accordé la santé, la volonté, la patience et la force nécessaires pour surmonter les difficultés rencontrées tout au long de ce parcours. Que la paix et les bénédictions soient sur le Prophète Mohammed.

À mon âme

À cette âme qui a patienté lorsque la patience était éprouvante, et qui a persévéré lorsque le chemin semblait long et difficile, et qui a toujours cru en ses capacités malgré les obstacles. À moi-même, pour tous les sacrifices consentis et tous les efforts accomplis. Tu mérites amplement cette réussite.

À l'homme de ma vie, mon père

Mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, mon amour inconditionnel. Que ce travail, soit pour toi un faible témoignage de ma profonde affection, de mon respect et de ma tendresse. Puisse Dieu le tout puissant te préserver et t'accorder santé, longue vie et bonheur éternel.

À ma précieuse Mère

Toi qui as porté les plus lourds fardeaux avec amour et courage, qui as cru en moi jusqu'au bout et m'as toujours encouragé à poursuivre mes rêves. Que ce travail soit le fruit de tes sacrifices. Qu'Allah te protège et te comble de Ses bienfaits.

À mon cher frère Mohamed et à ma petite sœur Nouna

Pour leur affection, leur présence et leurs encouragements tout au long de ce chemin.

À toute ma famille BENTAHAR,

À mes amis,

Qui ont partagé avec moi les joies et les difficultés de ce parcours.

À tous ceux qui m'ont aidée, conseillée et encouragée tout au long de cette aventure académique, je vous exprime ma sincère gratitude et ma profonde reconnaissance.

À vous tous, ce mémoire est dédié avec amour, gratitude et respect.

Chaimaa

Résumé

La détection et la classification automatiques des nodules thyroïdiens ont connu une évolution notable ces dernières années, grâce aux avancées rapides dans les techniques d'imagerie médicale et à l'émergence des méthodes d'apprentissage profond (deep learning), qui se sont révélées particulièrement efficaces dans le domaine de la vision par ordinateur. Ces méthodes se distinguent par leur capacité à extraire automatiquement des caractéristiques discriminantes à partir des images, sans nécessiter d'étapes de prétraitement complexes, permettant ainsi la construction de représentations hiérarchiques riches de l'information visuelle. Ce développement a contribué à améliorer l'efficacité des systèmes de détection, les rendant plus aptes à faire face à la complexité visuelle qui caractérise de nombreuses pathologies médicales, en particulier dans le domaine du diagnostic par imagerie échographique, comme la détection et la classification des nodules thyroïdiens. C'est précisément ce que nous avons examiné dans le cadre de notre mémoire.

Nous proposons un système automatique de détection et de classification des nodules thyroïdiens en deux catégories (Bénin/Malin) à partir d'images échographiques, en nous appuyant sur des architectures d'apprentissage profond combinées selon des stratégies hybrides, afin d'assurer des performances élevées et une robustesse clinique satisfaisante. Nous avons proposé cinq approches hybrides reposant sur les modèles YOLOv8 et YOLOE, ces deux modèles ont d'abord été entraînés individuellement, deux architectures en cascade (YOLOv8→YOLOE et YOLOE→YOLOv8) et deux approches de fusion des prédictions (NMS et WBF). Les expérimentations ont été menées sur le dataset ThyroidXL, composé de 11 545 images échographiques issues de 4 093 patients. L'évaluation individuelle montre que YOLOE obtient le meilleur mAP@50 (90,1 %) et YOLOv8 le meilleur mAP@50-95 (61,1 %), ce qui justifie leur complémentarité. Parmi les approches hybrides, la fusion WBF+NMS est la plus performante globalement, elle affiche le meilleur F1-score (92.9 %) et le meilleur rappel sur le malin (98,50 %), reflétant un équilibre optimal entre fiabilité et exhaustivité des détections. L'approche Fusion WBF présente un F1 malin très proche (89,52 %) et un mAP@50 (97,32 %), tandis que la cascade YOLOv8→YOLOE obtient le meilleur mAP@50 :95 (73.70 %), signe d'une localisation spatiale plus précise. Au regard de l'ensemble des critères, WBF+NMS est retenu comme l'architecture hybride optimale.

Mots-clés : apprentissage profond, détection, classification, échographie thyroïdienne, nodules thyroïdiens, image médicale, approches hybrides, fusion de prédictions.

Abstract

The automatic detection and classification of thyroid nodules has seen significant development in recent years, thanks to rapid advances in medical imaging techniques and the emergence of deep learning methods, which have proven particularly effective in the field of computer vision. These methods are distinguished by their ability to automatically extract discriminating features from images, without requiring complex preprocessing steps, thus enabling the construction of rich hierarchical representations of visual information. This development has contributed to improving the efficiency of detection systems, making them better suited to the visual complexity that characterizes many medical pathologies, particularly in the field of ultrasound imaging diagnosis, such as the detection and classification of thyroid nodules. This is precisely what we examined in our thesis.

We propose an automated system for detecting and classifying thyroid nodules into two categories (benign/malignant) from ultrasound images. This system leverages deep learning architectures combined using hybrid strategies to ensure high performance and satisfactory clinical robustness. We proposed five hybrid approaches based on the YOLOv8 and YOLOE models. These two models were initially trained individually. We also developed two cascade architectures (YOLOv8→YOLOE and YOLOE→YOLOv8) and two prediction fusion approaches (NMS and WBF). Experiments were conducted on the ThyroidXL dataset, which comprises 11,545 ultrasound images from 4,093 patients. Individual evaluation shows that YOLOE achieves the best mAP@50 (90.1%) and YOLOv8 the best mAP@50-95 (61.1%), justifying their complementary nature. Among the hybrid approaches, the WBF+NMS fusion performs best overall, exhibiting the highest F1 score (92.9%) and the best recall for malignant targets (98.50%), reflecting an optimal balance between reliability and comprehensiveness of detections. The WBF Fusion approach stands out with a very close malignant-class F1-score (89.52%) and a mAP@50 (97.32%), while the YOLOv8→YOLOE cascade achieves the best mAP@50 :95 (73.70%), indicating more precise spatial localization. Considering all the criteria, WBF+NMS is selected as the optimal hybrid architecture.

Keywords : deep learning, detection, classification, thyroid ultrasound, thyroid nodules, medical imaging, hybrid approaches, prediction fusion.

الملخص

شهد الكشف والتصنيف الآلي لعقيدات الغدة الدرقية تقدماً ملحوظاً في السنوات الأخيرة، بفضل التطورات السريعة في تقنيات التصوير الطبي وظهور أساليب التعلم العميق، التي أثبتت فعاليتها بشكل خاص في مجال رؤية الحاسوب. تتميز هذه الأساليب بقدرتها على استخلاص السمات المميزة من الصور تلقائياً دون الحاجة إلى خطوات معالجة مسبقة معقدة، مما يتيح بناء تمثيلات هرمية غنية للمعلومات المرئية. وقد ساهم هذا التطور في تحسين كفاءة أنظمة الكشف، وجعلها أكثر ملاءمة للتعقيد البصري الذي يميز العديد من الأمراض، لا سيما في مجال تشخيص التصوير بالموجات فوق الصوتية، مثل الكشف عن عقيدات الغدة الدرقية وتصنيفها. هذا تحديداً ما تناولناه في أطروحتنا.

نقترح نظاماً آلياً للكشف عن عقيدات الغدة الدرقية وتصنيفها إلى فئتين (حميدة/خبيثة) من صور الموجات فوق الصوتية. يستفيد هذا النظام من بنى التعلم العميق المدججة باستخدام استراتيجيات هجينة لضمان الأداء العالي والمتانة السريرية المرضية. اقترحنا خمس مناهج هجينة تعتمد على نموذجي YOLOv8 وYOLOE. تم تدريب هذين النموذجين بشكل منفصل في البداية. قننا أيضاً بتطوير بنيتين متتاليتين (YOLOE→YOLOv8) و(YOLOv8→YOLOE) وثلاث مناهج لدمج التنبؤات (NMS وWBF). أجريت التجارب على مجموعة بيانات ThyroidXL التي تضم 11545 صورة بالموجات فوق الصوتية من 4093 مريضاً. يُظهر التقييم الفردي أن YOLOE يحقق أفضل دقة متوسطة عند 50 (90.1%)، بينما يحقق YOLOv8 أفضل دقة متوسطة عند 50-95 (61.1%)، مما يُبرر طبيعتهما التكاملية. من بين المناهج الهجينة، حقق دمج WBF+NMS أفضل أداء إجمالي، حيث سجل أعلى درجة F1 (92.9%) وأفضل معدل استدعاء للأهداف الخبيثة (98.50%)، مما يعكس توازناً مثالياً بين موثوقية وشمولية الكشف. يتميز نهج WBF بتقارب نتائجه في مقياس F1 للأهداف الخبيثة (89.52%) ونتائجه في مقياس mAP@50 (97.32%)، بينما يحقق نظام المتتالي (YOLOE→YOLOv8) أفضل نتيجة في مقياس mAP@50:95 (73.70%)، مما يدل على دقة تحديد الموقع المكاني. وبناءً على جميع المعايير، تم اختيار WBF+NMS كأفضل بنية هجينة.

الكلمات المفتاحية: التعلم العميق، الكشف، التصنيف، الموجات فوق الصوتية للغدة الدرقية، عقيدات

الغدة الدرقية، التصوير الطبي، الأساليب الهجينة، دمج التنبؤ

Table des matières

Introduction Générale	9
1 Imagerie Médicale	12
1.1 Introduction	12
1.2 Définition d'une image	12
1.3 Image Numérique	12
1.3.1 Définition	12
1.3.2 Image binaire	13
1.3.3 Image en niveau de gris	13
1.3.4 Image en couleur	13
1.4 Les technologies d'imagerie médicale	14
1.4.1 Définition	14
1.4.2 Radiographie	14
1.4.3 Scanner	15
1.4.4 Imagerie par Résonance Magnétique (IRM)	15
1.4.5 Échographie	16
1.4.6 Microscopie	16
1.4.7 Scintigraphie	17
1.4.8 Élastographie	17
1.5 Conclusion	18
2 Pathologies Thyroïdiennes	20
2.1 Introduction	20
2.2 La Glande thyroïdienne	21
2.2.1 Anatomie	21
2.2.2 Structure histologique	21
2.2.3 Hormones thyroïdiennes	22
2.2.4 Rôles physiologiques des hormones thyroïdiennes	23
2.2.5 Régulation de la fonction thyroïdienne	23
2.3 Exploration Clinique et Examens Diagnostiques	23
2.3.1 Examen clinique	24
2.3.2 Analyses biologiques	24
2.3.3 Les examens d'imagerie	25
2.4 Les Pathologies Thyroïdiennes	25
2.4.1 L'hypothyroïdie	25
2.4.2 L'hyperthyroïdie	25
2.4.3 Thyroïdites (maladies inflammatoires)	26
2.4.4 Les nodules thyroïdiens	26

2.4.5	Le Goitre	26
2.4.6	Cancer de la thyroïde	26
2.5	La pathologie nodulaire	26
2.5.1	Définition	26
2.5.2	Les Complications (Malignité)	27
2.5.3	Classifications des nodules (Système TI-RADS)	28
2.6	Le carcinome thyroïdien	29
2.6.1	Définition	29
2.6.2	Les symptômes du cancer thyroïdien	30
2.6.3	Diagnostic et détection précoce	30
2.6.3.1	Interrogatoire et examen clinique	30
2.6.3.2	Examens complémentaires	31
2.7	Conclusion	32
3	Analyse Automatique des images médicales	34
3.1	Introduction	34
3.2	Intelligence Artificielle (IA)	34
3.2.1	Définition	34
3.2.2	Apprentissage Automatique (Machine Learning)	35
3.2.3	Apprentissage Profond (Deep Learning)	36
3.2.4	Le sur-ajustement (Overfitting)	45
3.3	Détection et Classification en Deep Learning	46
3.3.1	Détection de Nodules Thyroïdiens	46
3.3.1.1	Définition	46
3.3.1.2	Les modèles de détection	46
3.3.2	Classification de Nodules Thyroïdiens	50
3.3.2.1	Définition	50
3.3.2.2	Types de classifications	50
3.4	État de l'art	51
3.5	Conclusion	60
4	Résultats et expérimentations	62
4.1	Introduction	62
4.2	Outils utilisés	62
4.2.1	Langage utilisé (Python)	63
4.2.2	Google Colab (Colaboratory)	63
4.2.3	Lightning AI	63
4.2.4	Ultralytics	64
4.2.5	Roboflow	64
4.2.6	Les bibliothèques	64
4.3	Base de données utilisée	65
4.3.1	Analyse comparative et justification du choix de la base de données	65
4.3.2	Base de données retenu et analyse descriptive des données	66
4.4	Méthodes proposées	69
4.4.1	Modèle 1 (RF-DETR)	69
4.4.2	Modèle 2 (yolov8)	70
4.4.3	Modèle 3 (YOLOE)	71
4.4.4	Modèle 4 (yolo26)	72

4.5	Pré-traitement	73
4.5.1	Chargement du Base de données	73
4.5.2	Annotation des données	74
4.5.3	Répartition des données	75
4.5.4	Importation du Dataset	75
4.6	Apprentissage et hyperparamètres des modèles	76
4.7	Résultats et Discussion	79
4.8	Les Approches Hybrides Proposées	87
4.8.1	Approches en cascade	87
4.8.1.1	Cascade YOLOv8 → YOLOE	87
4.8.1.2	Cascade YOLOE → YOLOv8	88
4.8.2	Approches basée sur la fusion des prédictions	89
4.8.2.1	Fusion par Non-Maximum Suppression (NMS)	89
4.8.2.2	Fusion par Weighted Boxes Fusion (WBF)	89
4.8.2.3	Fusion Hybride WBF + NMS	90
4.8.3	Évaluation expérimentale des approches hybrides	90
4.9	Difficultés rencontrées	96
4.10	Conclusion	97
CONCLUSION GÉNÉRALE ET PERSPECTIVES		98

Table des figures

1.1	Image binaire	13
1.2	Image niveau de gris	13
1.3	Palette des niveaux de gris	13
1.4	Images en couleur	14
1.5	Appareil de radiographie	15
1.6	Les étapes d'utilisation du scanner	15
1.7	Les étapes de l'imagerie par l'IRM	16
1.8	Les étapes d'utilisation échographie	16
1.9	Appareil de Microscopie	17
1.10	Les étapes d'utilisation de scintigraphie	17
1.11	Appareil d'élastographie	18
2.1	Anatomie de la glande thyroïdienne	21
2.2	Anatomie de la glande thyroïdienne	22
2.3	Les hormones thyroïdiennes	23
2.4	Illustration d'une thyroïde normale et d'un nodule thyroïdien	27
2.5	Représentation anatomique et microscopique du cancer de la thyroïde	30
2.6	Échographie cervicale dans le bilan du cancer thyroïdien	31
2.7	Ponction à l'aiguille fine d'un nodule thyroïdien sous ultrasons	32
3.1	Relation entre IA et ML et DL	35
3.2	Les types d'apprentissage automatique	35
3.3	Déroulement du machine learning et du deep learning	36
3.4	Structure de modèle de réseau de neurones artificiels	36
3.5	Architecture standard d'un réseau à convolutions	37
3.6	évolution des architectures de CNN	37
3.7	Schéma explicatif d'une convolution	38
3.8	Schéma du "pooling"	38
3.9	Couches Fully Connected (FC) & Fonction Softmax	39
3.10	La structure d'un neurone artificiel	39
3.11	Différentes images de fonctions d'activation	40
3.12	Schéma de l'architecture ViT	41
3.13	Architecture du réseau hybride CNN-Transformer	42
3.14	Structure de l'architecture du EfficientNetV2	43
3.15	Architecture de feature pyramid network (FPN)	43
3.16	Illustration du transfert learning	44
3.17	Illustration du Overfitting et Underfitting	45
3.18	Architecture du modèle Faster R-CNN	47
3.19	L'architecture de l'algorithme YOLO	47

3.20	Évolution des versions de YOLO	48
3.21	Architecture du DETR	49
3.22	Exemple de classification Binaire	50
3.23	Exemple de classification multi-classes	51
3.24	Exemple de classification multi-label	51
4.1	Visualisation des images du dataset ThyroidXL	67
4.2	Distribution des genres dans le jeu de données	67
4.3	Distribution des âges dans le jeu de données	68
4.4	Distribution selon les catégories TI-RADS le jeu de données	68
4.5	RF-DETR Architecture	69
4.6	YOLOv8 Architecture	70
4.7	YOLOE Architecture	71
4.8	Diagramme d'architecture YOLO26	72
4.9	Analyse descriptive du dataset	74
4.10	Annotation avec outils Roboflow	74
4.11	Division des données en données d'apprentissage, de validation et de test	75
4.12	Téléchargement de base de données sur Lightning	75
4.13	data.yaml	76
4.14	Visualisation de la distribution des images des classes d'origine	76
4.15	Visualisation de la distribution des images après l'augmentation	76
4.16	Apprentissage du Modèle 1	77
4.17	Apprentissage du Modèle 2	78
4.18	Apprentissage du Modèle 3	78
4.19	Apprentissage du Modèle 4	78
4.20	Performances de l'apprentissage du Modèle 1	79
4.21	Performances de l'apprentissage du Modèle 2	80
4.22	Performances de l'apprentissage du Modèle 3	80
4.23	Performances de l'apprentissage du Modèle 4	81
4.24	Matrice de confusion du Modèle RF-DETR	83
4.25	Matrice de confusion du Modèle YOLOv8	84
4.26	Matrice de confusion du Modèle YOLOE	84
4.27	Matrice de confusion du Modèle YOLO26	85
4.28	Architecture Hybride en cascade YOLOv8 → YOLOE	88
4.29	Architecture Hybride en cascade YOLOE → YOLOv8	88
4.30	Architecture Hybride basée sur la fusion NMS	89
4.31	Architecture Hybride basée sur la fusion WBF	89
4.32	Architecture Hybride basée sur la fusion WBF+NMS	90
4.33	Matrice de confusion de L'approche hybride cascade	91
4.34	Matrices de confusion des différentes approches de fusion.	91
4.35	Visualisation des prédictions de l'approche en cascade	92
4.36	Visualisation des prédictions de l'approche basée sur la fusion	93

Liste des tableaux

2.1	Classification ACR TI-RADS des nodules thyroïdiens	29
3.1	Récapitulatif de certaines études proposées dans la littérature pour le diagnostic des nodules thyroïdiens	59
4.1	Comparaison des principaux datasets d’images échographiques thyroïdiennes .	66
4.2	Explication de l’architecture du modèle RF-DETR	69
4.3	Explication de l’architecture YOLOv8	70
4.4	Explication de l’architecture YOLOE	71
4.5	Explication de l’architecture YOLO26	73
4.6	Matrice de confusion binaire	81
4.7	Rapports de classification	86
4.8	Analyse du Recall et du FNR pour la classe maligne	86
4.9	Comparaison des performances des modèles de détection	86
4.10	Les Rapports de classification des approches hybrides	92
4.11	Synthèse comparative des indicateurs de performance des cinq approches hybrides évaluées (classe « Toutes catégories » et classe « Nodule malin »)	93

Liste des abréviations

- **AI** : Artificial Intelligence (Intelligence Artificielle)
- **AP** : Average Precision
- **AUC** : Area Under the Curve
- **CNN** : Convolutional Neural Network (Réseau de Neurones Convolutifs)
- **CT** : Computed Tomography (Scanner)
- **DETR** : Detection Transformer
- **DL** : Deep Learning (Apprentissage Profond)
- **EMA** : Exponential Moving Average
- **FN** : False Negative (Faux Négatif)
- **FNA** : Fine-needle Aspiration
- **FP** : False Positive (Faux Positif)
- **FPN** : Feature Pyramid Network (Réseau à Pyramide de Caractéristiques)
- **GPU** : Graphics Processing Unit (Processeur Graphique)
- **Hybrid-CNN** : Hybrid Convolutional Neural Network
- **IoU** : Intersection Over Union
- **IRM** : Imagerie par Résonance Magnétique
- **LR** : Learning Rate (Taux d'Apprentissage)
- **mAP** : Mean Average Precision
- **ML** : Machine Learning (Apprentissage Automatique)
- **NMS** : Non-Maximum Suppression
- **OpenCV** : Open Source Computer Vision Library
- **PAN** : Path Aggregation Network
- **RCNN** : Region-based Convolutional Neural Networks

- **RF-DETR** : Roboflow Detection Transformer
- **RNN** : Recurrent Neural Networks
- **ROI** : Region Of Interest
- **RVB** : Rouge Vert Bleu (Red Green Blue)
- **SPPF** : Spatial Pyramid Pooling Fast
- **TDM** : Tomodensitométrie
- **TEP-FDG** : Tomographie par Émission de Positons au Fluorodésoxyglucose
- **TI-RADS** : Thyroid Imaging Reporting and Data System
- **TN** : True Negative (Vrai Négatif)
- **TPO** : Thyroid Peroxidase
- **TP** : True Positive (Vrai Positif)
- **TRH** : Thyrotropin-Releasing Hormone
- **TSH** : Thyroid-Stimulating Hormone
- **ViT** : Vision Transformer
- **WBF** : Weighted Boxes Fusion
- **XAI** : Explainable Artificial Intelligence
- **YAML** : Yet Another Markup Language
- **YOLO** : You Only Look Once

INTRODUCTION GÉNÉRALE

Les maladies thyroïdiennes figurent parmi les affections endocriniennes les plus fréquentes dans le monde. Parmi les principales pathologies, on distingue les nodules thyroïdiens, les thyroïdites, l'hypothyroïdie, l'hyperthyroïdie ainsi que les cancers thyroïdiens, dont la détection précoce joue un rôle essentiel dans le pronostic et l'efficacité de la prise en charge thérapeutique. La glande thyroïde, en raison de sa localisation anatomique et sa vascularisation particulière, est particulièrement accessible à l'exploration par imagerie médicale, notamment par échographie, qui constitue aujourd'hui l'examen de première intention pour l'évaluation des pathologies thyroïdiennes.

Traditionnellement, le diagnostic et l'interprétation des images échographiques thyroïdiennes reposent sur l'expertise de radiologues et des endocrinologues spécialisés. Bien qu'efficace, ce processus reste long et coûteux. Ces dernières années, les progrès de l'imagerie médicale et des techniques d'apprentissage automatique et profond ont permis de développer des systèmes intelligents plus performants pour la détection précoce et la classification des anomalies thyroïdiennes à partir d'images échographiques.

Malgré ces progrès, plusieurs défis persistent dans les méthodes de classification des maladies thyroïdiennes. L'interprétation manuelle des images échographiques demeure lente et chronophage, ce qui ralentit le dépistage et la prise en charge des patients, notamment en raison du manque de spécialistes expérimentés. À cela s'ajoutent les risques de divergences d'interprétation ainsi que la difficulté à identifier certaines anomalies atypiques. Pour répondre à ces défis, les systèmes de diagnostic assisté par ordinateur (CAD) se sont développés afin d'aider les praticiens dans leur travail quotidien. Avec l'essor de l'intelligence artificielle (IA) et du deep learning (DL), des avancées significatives ont été réalisées dans le domaine du diagnostic médical, permettant d'en améliorer les performances. Le deep learning a démontré un fort potentiel en imagerie médicale, notamment grâce aux réseaux de neurones convolutifs (CNN), qui ont démontré une grande efficacité dans diverses tâches d'analyse d'images médicales, notamment la classification automatique et la détection précoce de maladies, avec des performances comparables, voire supérieures, à celles des experts humains. Plus récemment, les architectures basées sur les Transformers ont montré leur efficacité dans l'analyse des images médicales grâce à leur capacité à modéliser les

relations globales dans les images. Par ailleurs, certaines approches permettent également la localisation automatique des régions suspectes, ce qui constitue un complément important pour assister le diagnostic clinique.

Dans le cadre de notre projet de mémoire de fin d'études, l'objectif principal est de concevoir un système basé sur l'apprentissage profond pour le diagnostic automatique et la détection précoce des maladies thyroïdiennes à partir d'images échographiques. L'approche proposée repose sur l'entraînement et la comparaison de plusieurs architectures avancées telles que RF-DETR, YOLOv8, YOLOE et YOLO26 combinées au sein de pipelines hybrides. Deux stratégies de fusion sont explorées, une architecture en cascade et une architecture basée sur la fusion par NMS (Non-Maximum Suppression) et WBF (Weighted Boxes Fusion), visant à améliorer la robustesse et les performances globales du système. Dans ce but, nous avons utilisé les plateformes Google Colab, Lightning AI et Roboflow, en exploitant le langage Python et ses bibliothèques afin d'améliorer l'efficacité de l'entraînement grâce à leurs GPU gratuits et à un accès facile aux datasets.

Des expérimentations sont menées sur un ensemble de données (dataset), disponibles en ligne, sur HuggingFace. Il s'agit du jeu de données ThyroidXL, qui contient 11 545 images d'échographie de la thyroïde, chacune étant associée à une catégorie précise.

Ce mémoire est organisé en quatre chapitres. Le premier chapitre présente des concepts de base de l'imagerie médicale en général et plus particulièrement de l'imagerie thyroïdienne. Le deuxième chapitre est consacré aux maladies thyroïdiennes, nous exposons ces notions fondamentales ainsi que les caractéristiques de ces maladies. Dans le troisième chapitre, nous abordons les concepts fondamentaux de l'apprentissage profond, nous présentons également quelques architectures de réseaux neuronaux convolutifs (CNN) ainsi que quelques travaux récents sur la détection et la classification automatiques des maladies thyroïdiennes. Finalement, le quatrième chapitre est consacré à la partie expérimentale et aux résultats obtenus, où nous détaillons les méthodes proposées, la base de données utilisée, les étapes d'entraînement et d'évaluation des modèles, ainsi que l'approche hybride proposée dans ce travail, et nous terminons avec une discussion et une analyse des performances des modèles réalisés. Une conclusion générale est présentée à la fin du mémoire, rappelant le contexte du travail ainsi que les perspectives de recherches futures.

Chapitre 1

Imagerie médicale

Chapitre 1

Imagerie Médicale

1.1 Introduction

L'imagerie médicale regroupe un ensemble de techniques permettant d'obtenir des images du corps humain à des fins diagnostiques, thérapeutiques ou de recherche médicale. Ces examens, souvent réalisés dans des centres spécialisés, utilisent des équipements technologiques avancés afin de visualiser les organes, les os, les tissus ainsi que les processus biologiques [1]. Elle joue un rôle fondamental dans l'évolution de la médecine et de la recherche en utilisant diverses techniques permettant l'examen et la visualisation des organes internes sans intervention chirurgicale.

1.2 Définition d'une image

Une image numérique est une représentation d'une image sous forme de données informatiques. Contrairement aux photos argentiques traditionnelles, elle est composée d'éléments discrets et peut être stockée, transmise et modifiée par un ordinateur[2].

1.3 Image Numérique

1.3.1 Définition

Une image numérique est la représentation d'une image bidimensionnelle (dessin, icône, photographie, etc.) sous la forme d'un ensemble fini de valeurs numériques, appelées éléments d'image ou pixels. Chaque pixel possède une valeur unique qui détermine sa couleur et sa luminosité et tous forment l'image complète. Il existe trois types de couleur dans l'images numériques : image binaire, image niveau de gris, image en couleur [3].

1.3.2 Image binaire

C'est une image qui ne prend que des pixels avec uniquement deux valeurs 0 (noir) et 1 (blanc). Le niveau de gris est donc codé sur un seul bit, ce qui permet une représentation simple (voir Figure 1.1).



Figure 1.1 – Image binaire [4]

1.3.3 Image en niveau de gris

Une image en niveaux de gris est une image dont les couleurs varient du blanc au noir. Chaque pixel est codé par le niveau de l'intensité lumineuse, qui est souvent un nombre entier compris entre 0 et 255. La valeur 0 indique la luminosité nulle, ou noir, tandis que la valeur 255 indique la luminosité maximale, ou blanc [5] (voir Figure 1.2).



Figure 1.2 – Image niveau de gris [4]

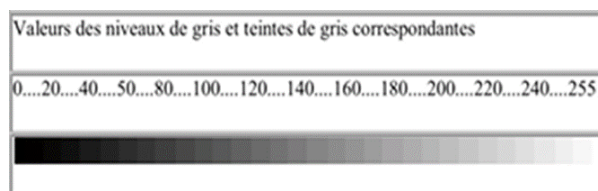


Figure 1.3 – Palette des niveaux de gris [6]

1.3.4 Image en couleur

Elle est obtenue par la combinaison de trois couleurs dites primaires : rouge, vert et bleu (RVB) (voir Figure 1.4). Chaque couleur est codée comme une

image à niveaux de gris, avec des valeurs allant de 0 à 255. Pour $R=V=B=0$ nous aurons un noir pur, et pour $R=V=B=255$ nous aurons un blanc pur. La représentation des images couleurs se fait donc soit par une image dont la valeur du pixel est une combinaison linéaire des valeurs des trois composantes couleurs, soit par trois images distinctes représentant chacune une composante couleur, on distingue généralement deux types d'images : [4]

- Images 24 bits ;
- Images à palettes (256 couleurs) ;



Figure 1.4 – Images couleurs [7]

1.4 Les technologies d'imagerie médicale

1.4.1 Définition

L'imagerie médicale est devenue essentielle dans le domaine de la santé. Regroupe l'ensemble des moyens physiques ou des techniques utilisées par la médecine pour le diagnostic mais aussi pour le traitement d'un grand nombre de pathologies pour visualiser les cellules d'un organisme (corps humain). Grâce aux images collectées, ils peuvent détecter les pathologies, suivre leur évolution et adapter le traitement à chaque patient. Ces technologies telles que la radiographie, la CT, l'IRM et l'échographie qui fournissent des informations importantes pour établir un diagnostic fiable, à assurer un suivi médical efficace et à éviter des interventions inutiles [8].

1.4.2 Radiographie

La radiographie repose sur l'utilisation des rayons X, capables de traverser les tissus de manière différente selon leur densité. En plaçant un détecteur relié à un ordinateur derrière la zone à examiner (autrefois un film photographique), on obtient une image sur laquelle les organes sont plus ou moins clairs. Ainsi les os, plus denses, apparaissent en blanc, tandis que les tissus mous, moins

denses, apparaissent en nuances de gris. Cette image étant bidimensionnelle, le radiologue doit déduire l'anatomie en 3D grâce à ses connaissances anatomiques et de son expérience [9].

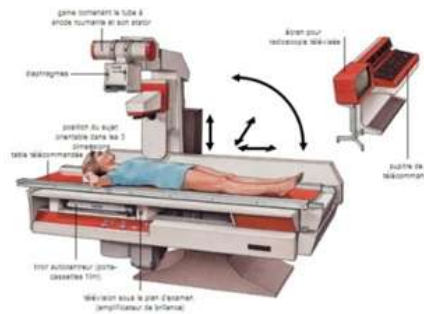


Figure 1.5 – Appareil de radiographie [10]

1.4.3 Scanner

Le scanner, également appelé tomodensitométrie, est une technique qui d'obtenir des images en coupe d'un organe. Il vise à apporter plus de précisions sur les résultats d'une radiographie ou d'une échographie, notamment la localisation et l'étendue d'une lésion sur un organe ou un tissu [11].

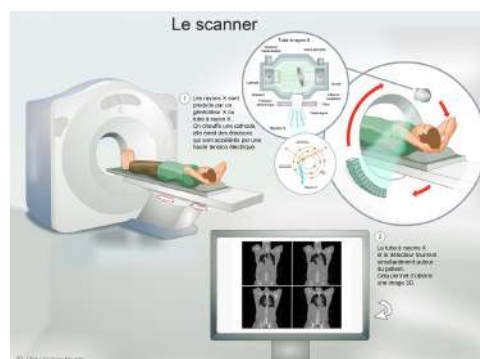


Figure 1.6 – Les étapes d'utilisation du scanner [12]

1.4.4 Imagerie par Résonance Magnétique (IRM)

L'IRM (imagerie par résonance magnétique) est une technique d'imagerie médicale récente fondée sur la résonance magnétique nucléaire. Elle permet d'obtenir des images en deux ou trois dimensions de l'intérieur du corps, de manière non invasive et sans irradiation [13]. Cet examen repose sur l'excitation des noyaux d'hydrogène du corps pour produire des images. Elle est préférée à la Tomodensitométrie (TDM) lorsqu'une analyse plus de détails pour les tissus mous est nécessaire, notamment pour l'étude des anomalies du cerveau, de la colonne vertébrale, des muscles et du foie. L'IRM est particulièrement efficace pour identifier les tumeurs dans ces tissus [14].



Figure 1.7 – Les étapes de l'imagerie par l'IRM [15]

1.4.5 Échographie

L'imagerie par ultrasons, ou échographie, est une technique qui permet de visualiser l'intérieur du corps grâce à des sons à haute fréquence. Ces ondes sonores sont émises par un transducteur, qui capte ensuite les échos réfléchis pour les convertir en images dynamiques. Les échographes modernes offrent des détails sur les tissus mous sans recourir aux radiations ionisantes. L'un des principaux avantages de cette méthode est sa non invasivité. Contrairement à d'autres techniques comme les rayons X, l'échographie n'expose pas le patient aux effets potentiellement nocifs des radiations [16].



Figure 1.8 – Les étapes d'utilisation échographie [17]

1.4.6 Microscopie

La microscopie regroupe un ensemble de techniques permettant d'observer des structures biologiques et visualiser des éléments invisibles à l'œil nu. L'appareil utilisé pour cela est le microscope. La qualité de l'image obtenue dépend principalement du pouvoir séparateur, c'est-à-dire de la capacité à distinguer deux points voisins, plus que du simple le grossissement [18].



Figure 1.9 – Appareil de Microscopie [19]

1.4.7 Scintigraphie

La scintigraphie, également connue sous le nom de tomographie par émission monophotonique, est une technique qui repose sur l'injection d'un traceur faiblement radioactif dans l'organisme. Ce traceur se fixe de manière ciblée sur l'organe à visualiser. Une caméra spéciale appelée gamma-caméra capte les photons émis, et ces données sont traitées pour former une image de l'organe étudié. Cette méthode permet non seulement d'obtenir des images, mais aussi d'évaluer le fonctionnement des organes, notamment le cerveau ou le cœur [20].

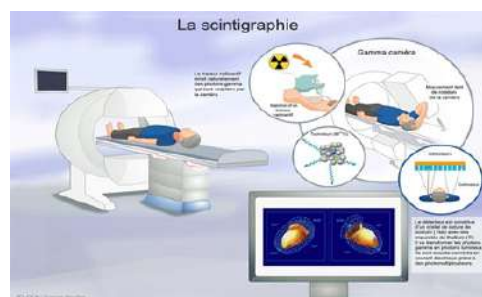


Figure 1.10 – Les étapes d'utilisation de scintigraphie [17]

1.4.8 Élastographie

L'élastographie est un précieux outil de diagnostic non invasif qui permet d'obtenir des informations essentielles sur la rigidité des tissus. Elle fonctionne de manière analogue de palpation (toucher) d'un organe, mais de façon non-invasive grâce, en utilisant des ondes associées à l'imagerie. Cette méthode mesure combien un tissu est dur ou mou, ce qui facilite la détection d'anomalies, telles que des tumeurs sans biopsie immédiate. Grâce à sa capacité à fournir des résultats rapides, précis et fiables sans nécessiter de procédures invasives telles que des biopsies, l'élastographie est une technique importante en pratique médicale moderne [21].



Figure 1.11 – Appareil d'élastographie [22]

1.5 Conclusion

Dans ce chapitre, nous avons présenté une brève introduction aux concepts liés au domaine du traitement d'images. Nous avons détaillé le principe l'imagerie médicale, définie comme l'ensemble des technologies permettant de visualiser des parties du corps humain ou des organes, son but est de diagnostiquer et guider les gestes de traitement. Par suit, nous avons répertorié quelques méthodes utilisées dans le domaine de l'imagerie, telles que : Radiologie, Scanner et IRM.

Chapitre 2

Pathologies thyroïdiennes

Chapitre 2

Pathologies Thyroïdiennes

2.1 Introduction

Les maladies thyroïdiennes représentent un enjeu majeur de santé publique en raison de leur forte prévalence et de leurs conséquences potentielles sur la santé générale. Parmi ces affections, les nodules thyroïdiens sont particulièrement fréquents. On estime que 2 à 6 % de la population adulte présente des nodules palpables, en moyenne 4 %, tandis que de 40 à 60% peuvent être détectés par échographie. Ces nodules sont plus fréquents chez les femmes et leur incidence augmente avec l'âge [23].

Bien que la majorité des nodules soient bénins, une faible proportion peut être malins, ce qui rend indispensable une évaluation précise et précoce. Le cancer de la thyroïde est relativement rare, représentant environ 1 % de l'ensemble des cancers diagnostiqués et près de 500 à 1 500 nouveaux cas par an à l'échelle nationale [24]. Son incidence est en augmentation constante depuis plusieurs décennies, principalement en raison d'une meilleure détection des formes papillaires et des microcarcinomes grâce aux examens d'imagerie et à la cytoponction. Cette augmentation est largement attribuée au surdiagnostic plutôt qu'à une réelle augmentation de l'incidence de la maladie [23].

L'échographie est la méthode principale pour le diagnostic précoce et la caractérisation des nodules thyroïdiens. Cependant, l'interprétation des images échographiques reste dépendante de l'expérience du radiologue, entraînant une variabilité inter-observateur et parfois des biopsies inutiles. Les facteurs de risque incluent l'exposition aux radiations, les antécédents familiaux de cancer thyroïdien, certaines mutations génétiques et la carence en iode dans certaines régions.

Dans ce contexte, l'introduction des modèles d'apprentissage profond offre une opportunité prometteuse d'automatiser l'analyse d'images échographiques, d'améliorer la précision et de détecter précocement les nodules malins, tout en réduisant la variabilité entre praticiens et les erreurs diagnostiques.

2.2 La Glande thyroïdienne

2.2.1 Anatomie

La glande thyroïde est une glande endocrine de couleur brun-rouge, vascularisée et en forme de papillon, située au milieu de la partie antérieure du cou. Dans des conditions normales, elle s'étend de la cinquième vertèbre cervicale (C5) à la première vertèbre thoracique (T1). Elle pèse en moyenne de 15 à 25 g et constitue la plus grande des glandes endocrines. Elle se compose de deux lobes, gauche et droit (voir Figure 2.1), qui sont connectés par un isthme [25].

La thyroïde fabrique deux hormones essentielles qui régulent le métabolisme et agissent sur le fonctionnement de divers organes tels que le cœur, les muscles, le système nerveux ou encore le tube digestif [26].

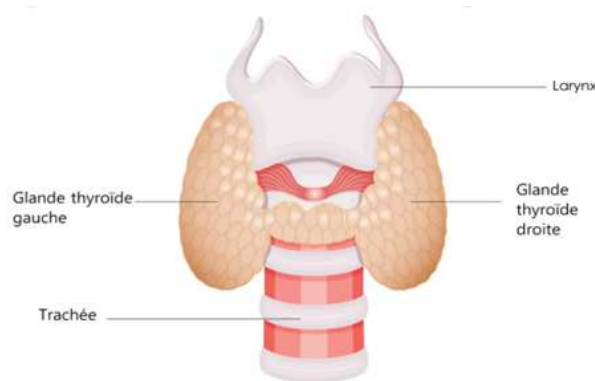


Figure 2.1 – Anatomie de la glande thyroïdienne [27]

2.2.2 Structure histologique

La thyroïde humaine regroupe à l'intérieure d'une structure morphofonctionnelle unique, le follicule thyroïdien "la vésicule thyroïdienne", deux types de cellules :

a. Cellules folliculaires

Encore appelées cellules vésiculaires ou thyrocytes, sont les cellules de la paroi d'un follicule thyroïdien et leur taille varie selon l'activité. Quand les follicules sont au stade de repos (inactif), les cellules sont plates et simples squameuses avec une quantité abondante de colloïde. Lorsque les follicules sont très actifs, les cellules folliculaires sont simples en colonne avec une colloïde peu abondante. Dans un état d'activité moyenne, les cellules sont cuboïdes simples, et la cavité contient une quantité modérée de colloïde (voir Figure 2.2). Ils sécrètent deux hormones qui régulent le rythme du métabolisme : T3 et T4 [28].

b. Cellules parafolliculaires (cellules C)

Sont également appelées cellules claires ou cellules lumineuses. Ce sont de grandes cellules polyédriques, à coloration pâle avec des noyaux ovales et excentriques. Ils sont largement répartis entre les cellules folliculaires et leur membrane du socle, et peuvent aussi se trouver entre des cellules folliculaires adjacentes sans atteindre la lumière (voir Figure 2.2). Ils sécrètent une hormone appelée calcitonine, dont la sécrétion est indépendante de l'hypophyse cérébrale et dépend principalement du taux de calcium sérique [28].

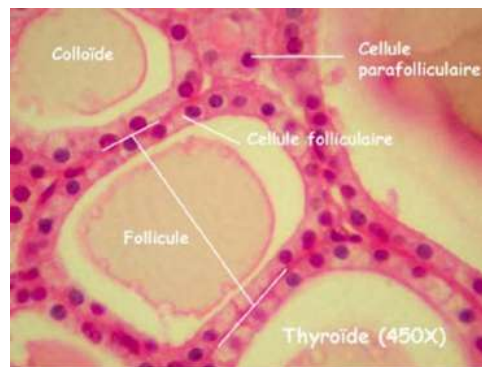


Figure 2.2 – Anatomie de la glande thyroïdienne [29]

2.2.3 Hormones thyroïdiennes

La thyroïde produit deux majeures hormones :

- T3 (Triiodothyronine)
- T4 (Thyroxine)

La T4, principale hormone produite par la thyroïde, a un effet minime sur le métabolisme de base et est convertie en T3, l'hormone la plus active, dans le foie et d'autres tissus et dépend de plusieurs facteurs selon les besoins de l'organisme. La majeure partie de T4 et T3 dans le sang est transportée par la globuline liant la thyroxine, tandis qu'une petite partie circule librement et constitue la forme active. Pour fabriquer les hormones thyroïdiennes, la thyroïde capte l'iode présent dans l'eau et les aliments et l'utilise dans la synthèse hormonale [30].

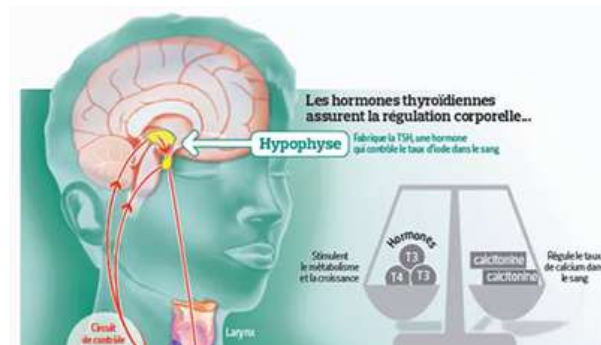


Figure 2.3 – Les hormones thyroïdiennes [31]

2.2.4 Rôles physiologiques des hormones thyroïdiennes

Les hormones thyroïdiennes T3L et T4L, surtout la T3L, agissent en se liant au récepteur TR, qui active ou réprime la transcription de nombreux gènes. Elles stimulent l'ensemble des métabolismes, augmentent le métabolisme de base et participent à la thermorégulation. Elles favorisent la lipolyse, la protéolyse, augmentent la consommation d'oxygène du cœur et renforcent ses effets chronotrope et inotrope. Elles contribuent aussi à la croissance des os, au développement du système nerveux et des tissus, tandis qu'un déficit en iode entraîne retard de croissance et baisse du QI [32].

2.2.5 Régulation de la fonction thyroïdienne

La TSH, sécrétée par l'hypophyse sous stimulation de la TRH, stimule la production d'hormones thyroïdiennes et la prolifération des thyrocytes. Les hormones thyroïdiennes exercent un rétrocontrôle négatif sur TRH et TSH via le TR β 2. L'iodure module la synthèse hormonale par l'effet Wolff-Chaikoff et sa carence augmente la sensibilité à la TSH. D'autres facteurs comme la somatostatine, dopamine, glucocorticoïdes, stress et certaines cytokines régulent également la TSH. Ainsi, la fonction thyroïdienne est contrôlée par un réseau complexe d'hormones et de signaux extracellulaires [33].

2.3 Exploration Clinique et Examens Diagnostiques

Les médecins examinent tout d'abord la personne et palpent le cou pour évaluer une éventuelle hypertrophie de la thyroïde. Selon les résultats de cet examen, d'autres analyses ou examens complémentaires peuvent être nécessaires, notamment lorsque les médecins ne peuvent déterminer si le trouble provient de la thyroïde ou de l'hypophyse.

2.3.1 Examen clinique

Palpation de la thyroïde

La palpation est le premier examen du médecin pour explorer la thyroïde et permet de retrouver la présence d'un goitre ou d'un nodule. Des examens complémentaires sont effectués selon les manifestations constatées [27].

Signes généraux

Un dysfonctionnement peut entraîner des symptômes subtils et facilement négligés. Connaître les signes des problèmes de thyroïde permet un diagnostic précoce et un traitement efficace. Voici quelques signes à surveiller [34] :

- Modifications de la fréquence cardiaque
- Perte de cheveux et santé de la thyroïde
- Faiblesse musculaire et douleurs articulaires
- Gonflement du cou
- Enrouement et changements de voix

2.3.2 Analyses biologiques

Dosage TSH

La TSH, ou thyroïdostimuline, est le marqueur principal du bilan thyroïdien. Produite par l'hypophyse, stimule la thyroïde pour qu'elle produise les hormones T3 et T4, généralement c'est le premier examen prescrit pour évaluer la fonction thyroïdienne [35].

Dosage des hormones T3 et T4

La T3 libre est la forme la plus active des hormones thyroïdiennes, dont environ 80 % provient de la conversion périphérique de la T4 dans le foie et les reins. La T4 libre (FT4) représente la fraction non liée aux protéines de transport, donc biologiquement active, et constitue environ 0,03 % de la T4 totale [35].

Dosages sanguins

Un dosage sanguin permet de rechercher des anticorps anti-thyroïde, notamment les anticorps anti-TPO (ThyroPérOxydase), afin de détecter une maladie de la thyroïde d'origine auto-immune ou un dysfonctionnement thyroïdien [27].

2.3.3 Les examens d'imagerie

Échographie cervicale

L'échographie thyroïdienne, effectuée à l'aide d'une sonde à ultrasons sur le cou, permet de visualiser la thyroïde, et de rechercher la présence de nodules. C'est un examen simple et non invasif réalisé surtout pour l'évaluation diagnostique des nodules de la thyroïde [27].

Scintigraphie thyroïdienne

Cet examen indolore consiste à injecter un produit radioactif, comme de l'iode ou du technétium, qui se fixe sur la thyroïde et permet de différencier un nodule « chaud » sécrétant des hormones thyroïdiennes et le plus souvent bénin, d'un nodule froid [27].

Biopsie à l'aiguille fine (FNA)

L'aspiration à l'aiguille fine (FNA) est une procédure utilisée par les prestataires de soins de santé pour prélever un échantillon de cellules sur une masse suspecte ou une zone anormale de corps. C'est aussi appelée une biopsie à l'aiguille fine. La FNA consiste à utiliser une fine aiguille et une seringue pour extraire des cellules, des tissus et des liquides. Le professionnel de la santé envoie ensuite l'échantillon à un laboratoire de cytologie. Là, un cytotechnologue examine les cellules et un spécialiste appelé pathologiste les analyse. La FNA est le type de biopsie le moins invasif, mais il a des limites [36].

2.4 Les Pathologies Thyroïdiennes

2.4.1 L'hypothyroïdie

Cette maladie thyroïdienne survient lorsque la thyroïde ne produit pas suffisamment d'hormones thyroïdiennes. Les causes principales de ce trouble thyroïdien sont la maladie de Hashimoto (une maladie auto-immune), des antécédents de chirurgie thyroïdienne, des traitements radioactifs ou certains médicaments comme l'amiodarone, l'immunothérapie et le lithium [37].

2.4.2 L'hyperthyroïdie

Cette maladie thyroïdienne survient lorsque la glande thyroïde produit les hormones thyroïdiennes en quantité excessive. Ses principales causes sont les

maladies auto-immunes, notamment la maladie de Basedow, ainsi que les nodules thyroïdiens. Elle peut aussi être liée à certains médicaments, des infections virales ou, plus rarement, un dysfonctionnement hypophysaire [37].

2.4.3 Thyroïdites (maladies inflammatoires)

Les thyroïdites sont des affections causées par un processus infectieux, inflammatoire ou immunitaire de la thyroïde. Elles comprennent plusieurs types : thyroïdite aiguë suppurée, thyroïdite subaiguë, thyroïdite indolore ou silencieuse, thyroïdite fibreuse et Thyroïdite du post-partum [38].

2.4.4 Les nodules thyroïdiens

Un nodule thyroïdien est une masse focale développée dans la glande thyroïde. Dans la plupart des cas, il ne provoque aucun symptôme et bénin. Sa présence est plus fréquente chez les femmes et sa prévalence à augmente avec l'âge, ou dans certains particuliers [39].

2.4.5 Le Goitre

Le goitre c'est la pathologie de la thyroïde la plus fréquente. Le goitre simple correspond aux hypertrophies thyroïdiennes diffuses normo fonctionnelles (absence d'hyper ou d'hypothyroïdie), non inflammatoires (sont exclues les thyroïdites), et non cancéreuses. Les formes multinodulaires peuvent devenir symptomatiques, et s'accompagner d'un risque de complications [40].

2.4.6 Cancer de la thyroïde

Le cancer de la thyroïde (CT) est le cancer endocrinien le plus fréquent. Il se développe à partir des tissus de la glande thyroïde dans, lorsque les cellules prolifèrent de manière anormale, qui peuvent parfois se propager à d'autres parties du corps. Dans la majorité des cas, Il s'agit de tumeurs bien différenciées, principalement représentées par les carcinomes papillaires et vésiculaires [41].

2.5 La pathologie nodulaire

2.5.1 Définition

Un nodule thyroïdien est une croissance qui se forme au sein de la glande thyroïde se située à la base de votre cou. Ces nodules peuvent présenter des tailles et des textures ; certains peuvent sont très petits passent inaperçus, tandis

que d'autres peuvent devenir volumineux et visibles. Ils peuvent être solides ou contenir du liquide, dans ce cas appelée nodule kystiques [42].

Dans la majorité des situations, le nodule thyroïdien est indolore et ne provoque aucun symptôme. Il est souvent découvert lors de la palpation du cou par le patient ou le médecin ou soit l'occasion d'une imagerie réalisée pour une autre indication, comme écho-doppler des carotides ou scanner thoracique. Les nodules peuvent être uniques ou multiples sans que cela ne modifie la façon de les explorer. Leur évolution est généralement lente et la prise en charge est rarement urgente. Les nodules sont très fréquents dans la population générale, particulièrement chez la femme, et leur fréquence augmente avec l'âge [43].

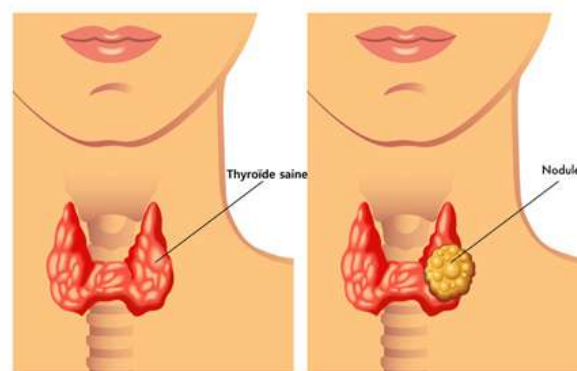


Figure 2.4 – Illustration d'une thyroïde normale et d'un nodule thyroïdien [44]

2.5.2 Les Complications (Malignité)

Les nodules thyroïdiens sont le plus souvent asymptomatiques et de nature bénigne. Toutefois, dans certains cas, ils peuvent se manifester par des signes cliniques qui orientent le diagnostic.

1. Signes locaux :

Un nodule peut parfois entraîner :

- Une douleur au niveau cervical, généralement en faveur d'une thyroïdite ou d'une hémorragie intrakystique.
- Un nodule de grande taille peut provoquer des signes de compression des structures voisines, tels qu'une dysphagie, une dyspnée ou une dysphonie. La présence de ces symptômes compressifs doit faire évoquer un risque accru de malignité.

2. Signes fonctionnels :

Selon l'activité hormonale de la thyroïde, différents symptômes peuvent apparaître :

— **En cas d’hyperthyroïdie :**

Des symptômes tels que des palpitations, une perte de poids, des tremblements ou une intolérance à la chaleur sont en faveur d’une hyperthyroïdie, souvent liée à un nodule hyperfonctionnel ou à une thyroïdite.

— **En cas d’hypothyroïdie :**

À l’inverse, une fatigue, une prise de poids et une intolérance au froid peuvent évoquer une hypothyroïdie, fréquemment associée à une thyroïdite chronique, notamment de type Hashimoto.

3. Facteurs de risque de malignité :

Certains éléments de l’anamnèse constituent des facteurs de risque de cancer thyroïdien :

- Antécédents d’irradiation cervicale, surtout durant l’enfance
- Antécédents familiaux de cancer thyroïdien
- Croissance rapide du nodule
- Nodule isolé ou goitre associé
- Élévation du taux de TSH

Bien que les nodules soient plus fréquents chez les femmes, les formes les plus agressives sont plus souvent observées chez les hommes, en particulier les sujets âgés [45].

2.5.3 Classifications des nodules (Système TI-RADS)

Le système ACR TI-RADS est un système de compte rendu des nodules thyroïdiens à l’échographie, proposé par l’American College of Radiology (ACR).

Il utilise une notation standardisée pour les comptes rendus, fournissant aux utilisateurs des recommandations sur le moment d’effectuer une ponction à l’aiguille fine (FNA) ou un suivi échographique des nodules suspects, et sur la possibilité de ne pas intervenir sur les nodules bénins ou non suspects [46].

Cinq critères échographiques du système ACR TI-RADS

1. **Composition :** Kystique, Spongiforme, Mixte, Solide
2. **Échogénicité :** Anéchogène, Hyper- ou isoéchogène, Hypoéchogène, Très hypoéchogène
3. **Forme :** Plus large que haute, Plus haute que large
4. **Contours :** Réguliers, Mal définis, Lobulé/irrégulier, Extension extrathyroïdienne

5. **Foyers échogènes** : Aucun, Artefact en queue de comète étendu, Microcalcifications, Calcifications périphériques/en anneau, Foyers échogènes ponctués [46].

Catégories TI-RADS :

	Score	Catégories TI-RADS	Malignité	Recommandation
TR1	0 points	Bénin	0.3%	Pas de FNA nécessaire
TR2	2 points	Non suspect	1.5%	Pas de FNA nécessaire
TR3	3 points	Faible suspicion	4.8%	Suivi ≥ 1.5 cm, FNA ≥ 2.5 cm
TR4	4–6 points	Suspicion modérée	9.1%	Suivi ≥ 1 cm, FNA ≥ 1.5 cm
TR5	≥ 7 points	Haute suspicion	35%	Suivi ≥ 0.5 cm, FNA ≥ 1 cm

Table 2.1 – Classification ACR TI-RADS des nodules thyroïdiens [46]

2.6 Le carcinome thyroïdien

2.6.1 Définition

Le cancer de la thyroïde est une tumeur maligne qui se développe au sein de la glande thyroïde. Bien qu'il soit relativement rare, son incidence a augmenté au cours des dernières décennies. Il touche principalement les femmes, en particulier entre 30 et 50 ans. Ce cancer peut prendre soit des cellules folliculaires, soit des cellules C parafolliculaires. Il se manifeste généralement par des nodules indolores, ce qui rend son dépistage parfois complexe [47].

Selon le type de cellules atteintes, l'évolution du cancer thyroïdien varie :

- **Le carcinome papillaire** : il s'agit du type le plus fréquent et le moins agressif. Il évolue lentement et reste le plus souvent limité à la thyroïde.
- **Le carcinome folliculaire** : également à croissance lente, il peut toutefois se propager aux ganglions lymphatiques et à d'autres organes.
- **Le carcinome médullaire** : plus agressif, il peut se disséminer rapidement aux ganglions lymphatiques ainsi qu'aux organes voisins.
- **Le carcinome anaplasique** : forme rare mais très agressive, caractérisée par une évolution rapide et un diagnostic souvent posé à un stade avancé [48].

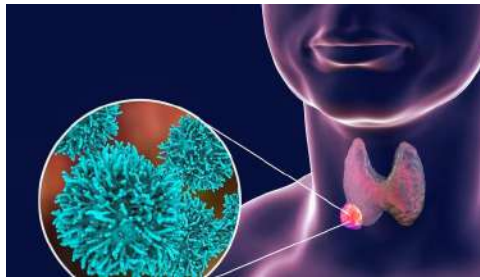


Figure 2.5 – Représentation anatomique et microscopique du cancer de la thyroïde [49]

2.6.2 Les symptômes du cancer thyroïdien

- **Bosse au niveau du cou :**

L'apparition d'une masse dure dans la partie antérieure du cou, généralement indolore, correspondant à la localisation de la glande thyroïde.

- **Ganglions cervicaux :**

Dans certains cas, le cancer peut s'étendre aux ganglions lymphatiques cervicaux, provoquant leur inflammation ou une augmentation de volume. Cela peut entraîner une gêne ou une douleur au niveau du cou, voire de l'oreille.

- **Difficulté à avaler :**

Une gêne ou une difficulté à avaler des aliments solides ou liquides.

- **Enrouement ou modifications de la voix :**

Si la tumeur irrite ou atteint le nerf laryngé récurrent, elle peut entraîner des modifications de la voix, telles qu'un enrouement persistant ou un affaiblissement du ton vocal.

2.6.3 Diagnostic et détection précoce

2.6.3.1 Interrogatoire et examen clinique

Aucun signe n'est spécifique, et le plus souvent, il s'agit d'un nodule d'allure clinique banale. Quelques données peuvent cependant être utiles à préciser :

- Antécédents de maladie thyroïdienne ou d'irradiation cervicale.
- Antécédents familiaux de cancer thyroïdien ou de néoplasie endocrinienne multiple de type 2.
- Âge inférieur à 15 ans ou supérieur à 60 ans.
- Sexe masculin (les nodules chez l'homme sont plus souvent malins).
- Nodule palpable, dur, adhérent, ou qui grossit rapidement.
- Symptômes de compression ou paralysie nerveuse : voix rauque, difficulté à avaler, essoufflement[50].

2.6.3.2 Examens complémentaires

Imagerie : Échographie cervicale et thyroïdienne est l'examen de référence pour tout nodule palpable. Elle permet de :

- Décrire la taille, la localisation et le nombre de nodules.
- Identifier les signes échographiques de malignité.
- Examiner les ganglions cervicaux.
- Servir de référence pour le suivi futur[50].



Figure 2.6 – Échographie cervicale dans le bilan du cancer thyroïdien [51]

Autres examens

- Scintigraphie : seulement si la TSH est basse, pour détecter un nodule hyperfonctionnel.
- IRM ou TDM cervicales : rarement nécessaires, sauf cas particuliers (nodules plongeants, goitres volumineux).
- TEP-FDG : pas utilisée pour le diagnostic initial.

Cytologie (ponction à l'aiguille fine) : Indiquée en fonction de la taille du nodule, de ses caractéristiques échographiques et du contexte de risque, selon un système de classification et des recommandations.

- Non indiquée pour les kystes purs, sauf si volumineux ou douloureux.
- Pour les nodules supérieurs à 2 cm, la ponction est généralement réalisée même sans facteur de risque.
- Pour les nodules de 0,7 à 2 cm, la ponction dépend des facteurs de risque ou des caractéristiques suspectes[50].

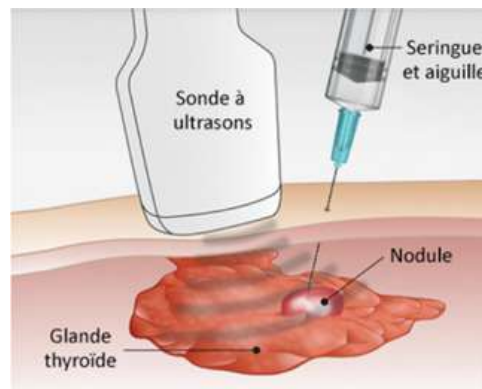


Figure 2.7 – Ponction à l’aiguille fine d’un nodule thyroïdien sous ultrasons [52]

Biologie : Devant un nodule thyroïdien, les dosages suivants sont réalisés :

- TSH sérique pour évaluer la fonction thyroïdienne.
- Calcémie en préopératoire pour éliminer une pathologie parathyroïdienne associée.
- Calcitonine sérique systématiquement proposée en préopératoire.

D’autres examens biologiques peuvent être réalisés selon le contexte, notamment dans le cadre d’un bilan préopératoire[50].

2.7 Conclusion

Les maladies thyroïdiennes constituent un enjeu majeur de santé publique, pouvant entraîner des complications graves en raison de l’absence d’un diagnostic précoce. Parmi elles, le cancer de la thyroïde, bien que moins fréquent, reste préoccupant, notamment chez les adultes jeunes et d’âge moyen. Grâce aux avancées en imagerie médicale et en intelligence artificielle, l’analyse des anomalies thyroïdiennes peut être améliorée en tant qu’outil d’aide au dépistage et au diagnostic. La prévention, le suivi régulier sont essentiels pour réduire les risques et améliorer la qualité de vie des patients. Dès lors, dans le prochain chapitre, nous abordons les concepts généraux de la classification d’images particulièrement la machine learning, le deep learning, quelques approches de classification et certains travaux récents portant sur la détection et l’analyse des nodules thyroïdiens.

Chapitre 3

Analyse automatique des images médicales

Chapitre 3

Analyse Automatique des images médicales

3.1 Introduction

L'analyse automatique des images médicales est une branche spécialisée de la vision par ordinateur et de l'intelligence artificielle (IA), axée sur l'interprétation et l'extraction d'informations pertinentes à partir des examens médicaux. Grâce à des algorithmes avancés, ce domaine automatise la détection des structures biologiques et des anomalies dans des données d'imagerie complexes, telles que les radiographies, les échographies ou les IRM. L'objectif principal consiste à soutenir les radiologues et les cliniciens en leur fournissant des données quantitatives précises pour guider le diagnostic, planifier les traitements et assurer le suivi à long terme des patients.

3.2 Intelligence Artificielle (IA)

3.2.1 Définition

L'intelligence artificielle (IA) est un domaine de l'informatique fondé sur des bases mathématiques et des concepts issus des sciences cognitives. Elle vise à concevoir des systèmes capables de résoudre des problèmes à forte complexité logique ou algorithmique [53]. L'intelligence artificielle désigne la capacité d'une machine à reproduire des comportements associés à l'être humain, tels que le raisonnement, la planification et la créativité. Elle permet aux systèmes techniques de percevoir leur environnement, de contrôler ces perceptions, de résoudre des problèmes et d'entreprendre des actions pour atteindre un objectif précis. L'ordinateur reçoit, analyse et réutilise des données préalablement préparées ou collectées via des capteurs. Ces systèmes peuvent également ajuster leur comportement en analysant les effets de leurs actions précédentes, tout en fonctionnant de manière autonome [54].

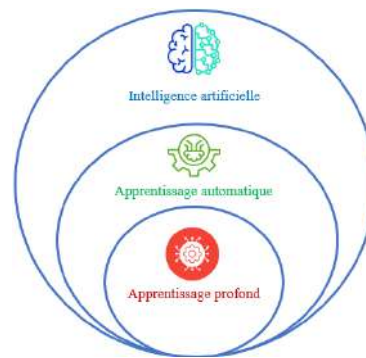


Figure 3.1 – Relation entre IA et ML et DL

3.2.2 Apprentissage Automatique (Machine Learning)

3.2.2.1 Définition

L'apprentissage automatique est un sous-domaine de l'intelligence artificielle qui utilise des algorithmes formés sur des ensembles de données pour créer des modèles qui permettent aux machines d'effectuer des tâches associées aux capacités humaines, telles que la catégorisation d'images, l'analyse de données ou la prédiction [55]. Il intègre des techniques telles que : l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage semi-supervisé et l'apprentissage par renforcement (voir Figure 3.2).

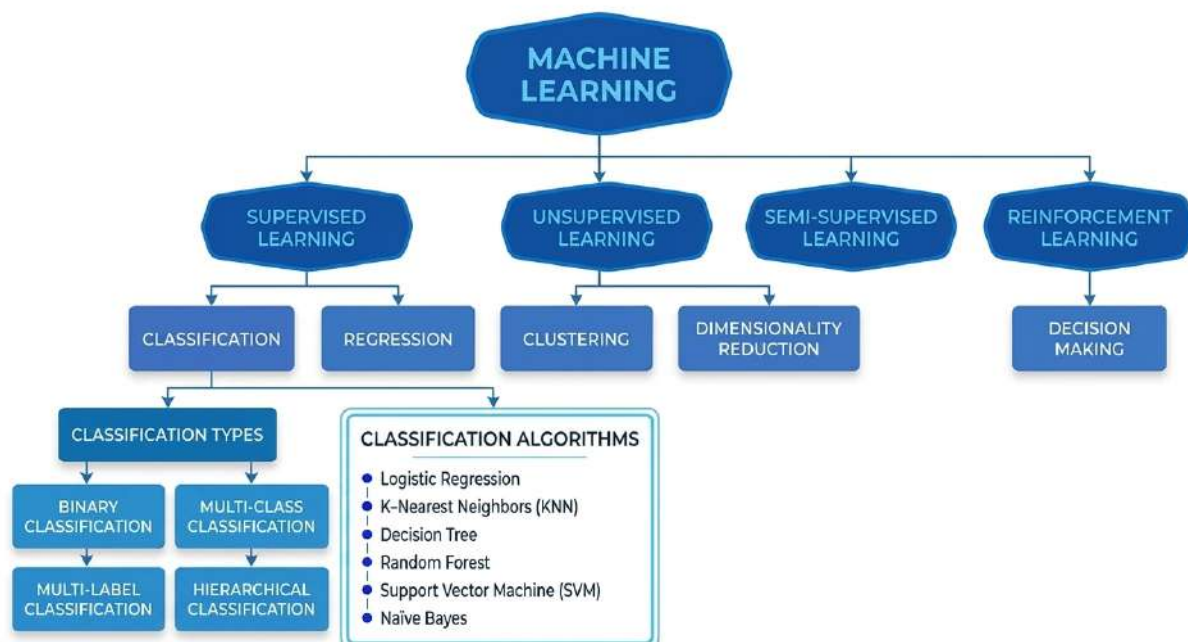


Figure 3.2 – les types d'apprentissage automatique

3.2.3 Apprentissage Profond (Deep Learning)

L'apprentissage profond est une technique plus avancée dans le domaine de l'apprentissage automatique, qui utilise des réseaux de neurones artificiels profonds pour traiter les données. Alors que l'apprentissage automatique peut utiliser des algorithmes simples pour identifier des motifs dans les données, l'apprentissage profond utilise plusieurs couches de neurones (similaires à la structure du cerveau humain) pour créer des modèles plus complexes et précis. Cela rend le deep learning plus adapté pour gérer des tâches plus complexes telles que la reconnaissance d'images, la compréhension de la parole ou la traduction linguistique. L'apprentissage profond nécessite souvent de grands ensembles de données et une capacité de calcul considérable [56].

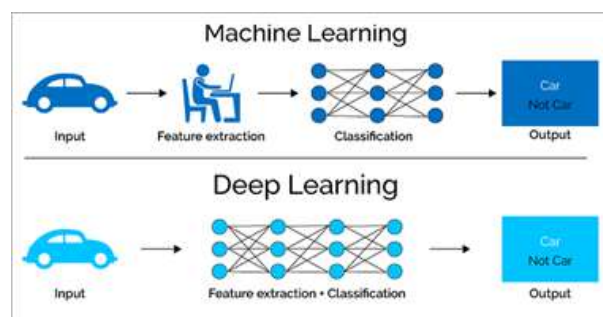


Figure 3.3 – Déroulement du machine learning et du deep learning [57]

Dans l'apprentissage profond, on utilise des réseaux de neurones artificiels pour concevoir des algorithmes capables de reproduire les activités du cerveau humain. Ils sont constitués de plusieurs dizaines de couches de neurones, où chaque couche est chargée de recevoir et d'interpréter les données de la couche précédente (voir Figure 3.4).

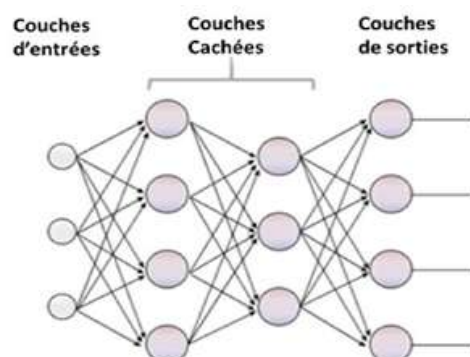


Figure 3.4 – Structure de modèle de réseau de neurones artificiels [58]

Parmi les modèles de Deep Learning utilisés, on trouve :

3.2.3.1 Convolutional Neural Networks (CNN)

Un réseau neuronal convolutionnel (CNN), également connu sous le nom de ConvNet, est un type spécialisé d'algorithme d'apprentissage profond principalement conçu pour des tâches nécessitant la reconnaissance d'objets, notamment la classification, la détection et la segmentation d'images. Les CNN sont utilisés dans divers scénarios pratiques, tels que les véhicules autonomes, les systèmes de caméras de sécurité, et d'autres. Les CNN se trouvent dans de nombreux types d'architectures, conviennent à différentes tâches telles que la classification d'images, la segmentation et même la détection d'objets. Parmi celles-ci, les architectures CNN les plus largement reconnues LeNet-5, AlexNet, VGG16 et ResNet (voir Figure 3.6).

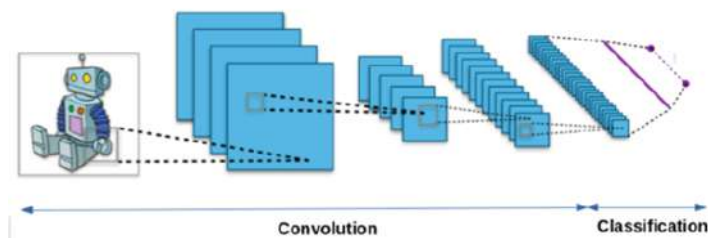


Figure 3.5 – Architecture standard d'un réseau à convolutions [59]

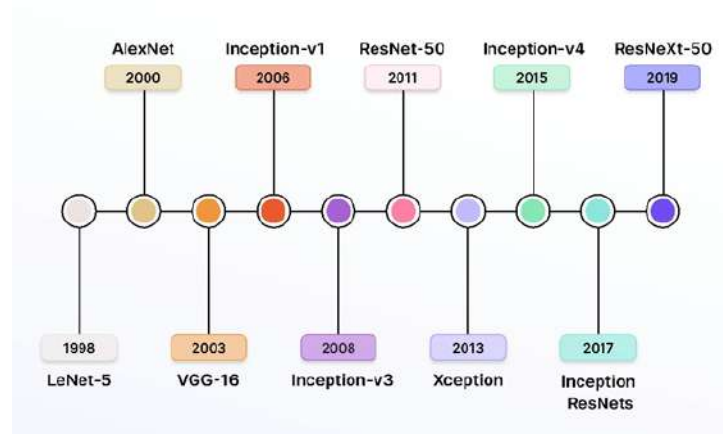


Figure 3.6 – évolution des architectures de CNN [60]

Une architecture CNN est formée de couches de traitement indépendantes :

1. Couche de convolution

La couche de convolution est un élément essentiel des réseaux de neurones convolutifs, et constitue généralement leur première couche. Elle est composée de filtres qui sont des matrices de valeurs appelées feature maps. Chaque couche de convolution reçoit une image en entrée et génère une feature map en

appliquant un filtre à cette image. Par exemple, une image de 5x5 pixels et un filtre de 3x3. La couche applique ainsi chaque filtre à toutes les images en entrée afin d'extraire les caractéristiques recherchées [61].

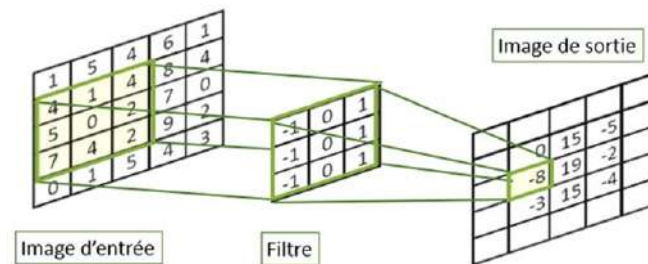


Figure 3.7 – Schéma explicatif d'une convolution [62]

2. Couche pooling

C'est la deuxième étape essentielle dans l'élaboration d'un réseau de neurones convolutifs et se place souvent entre deux couches de convolution. Elle reçoit en entrée plusieurs feature maps et applique à chacune l'opération de pooling. Une couche de pooling agit comme une couche de réduction en divisant l'image en blocs et ne conserve que le maximum de chaque bloc. Ainsi, elle réduit la dimension de l'image tout en gardant les caractéristiques importantes. En sortie, on obtient le même nombre de feature maps, mais plus petites [61].

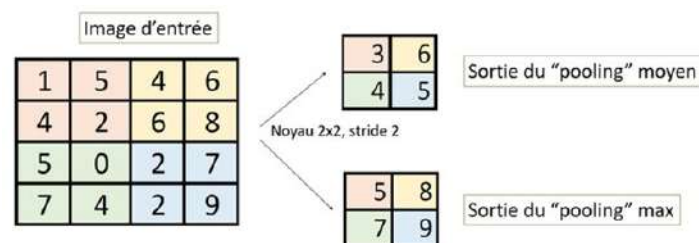


Figure 3.8 – Schéma du "pooling" [62]

3. Couches entièrement connectées

Il s'agit de l'une des dernières étapes du réseau de neurones. Leur entrée correspond à la dernière couche de pooling. Elles sont complètement connectées, c'est-à-dire que tous les neurones sont reliés entre eux, permettant l'apprentissage de combinaisons non linéaires entre les feature maps de la dernière couche du réseau de neurones. Comme les couches de convolution, elles sont associées à une fonction d'activation généralement la fonction softmax qui permet d'associer à chaque sortie du réseau de neurones convolutifs un score, ensuite transformé en probabilité d'appartenir à un label d'image [63].

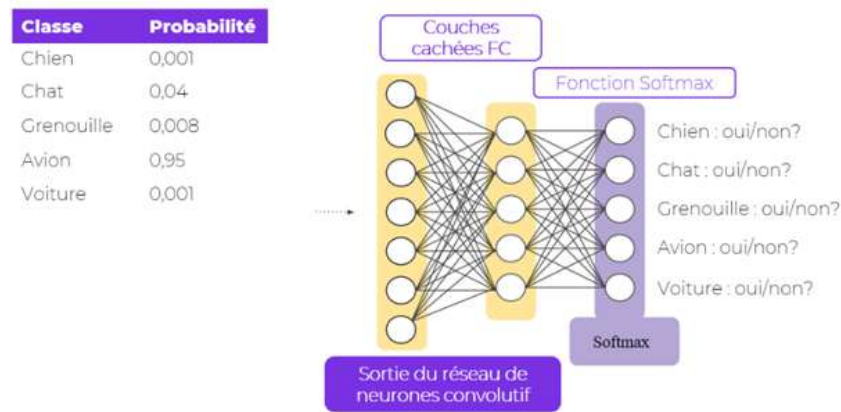


Figure 3.9 – Couches Fully Connected (FC) & Fonction Softmax [63]

4. Fonction d'activation

Une fonction d'activation est une fonction mathématique appliquée à la sortie linéaire d'un neurone artificiel. Elle va reproduire le potentiel d'activation que l'on retrouve dans le domaine de la biologie du cerveau humain. Elle va permettre le passage d'information ou non de l'information si le seuil de stimulation est atteint. Concrètement, elle va avoir pour rôle de décider si une réponse du neurone doit être activée ou pas [64]. Un neurone (voir Figure 3.10) ne va faire qu'appliquer la fonction suivante :

$$X = \sum(\text{entrée} \times \text{poids}) + \text{biais} \quad (3.1)$$

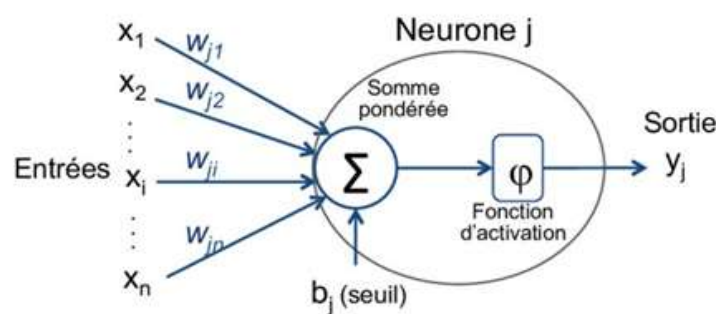


Figure 3.10 – La structure d'un neurone artificiel [65]

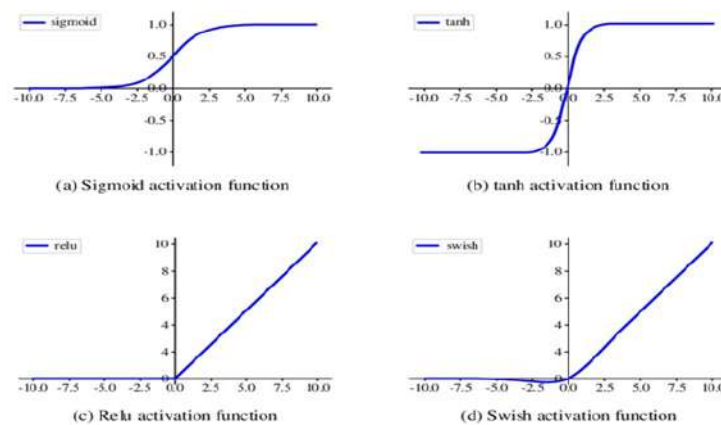


Figure 3.11 – Différentes images de fonctions d'activation [66]

3.2.3.2 Architectures avancées de Deep Learning

1 - Vision Transformer (ViT)

Le modèle Vision Transformer (ViT) a été largement utilisé dans le domaine de la vision par ordinateur, en particulier pour les tâches de classification d'images. En 2017, l'architecture Transformer proposée par l'équipe de Google pour résoudre la tâche de traduction automatique a été largement utilisée dans l'apprentissage profond, abandonnant les architectures traditionnelles CNN et RNN, l'ensemble de la structure du réseau étant entièrement composé du mécanisme d'Attention. En 2020, l'équipe de Google a appliqué la structure Transformer aux tâches de classification d'images et a proposé le modèle ViT basé sur cette structure, qui est devenu un travail marquant dans l'application des Transformers dans le domaine de la vision par ordinateur en raison de son modèle « simple », efficace et de sa forte capacité de mise à l'échelle, ce qui a également suscité des recherches ultérieures[67].

Le ViT capture mieux le contexte global que les CNN, et est particulièrement efficace pour :

- L'imagerie à grand champ de vision où le contexte des nodules est important,
- Les cas multi-nodulaires nécessitant une analyse complète de la thyroïde,
- L'apprentissage des relations spatiales entre les structures anatomiques.

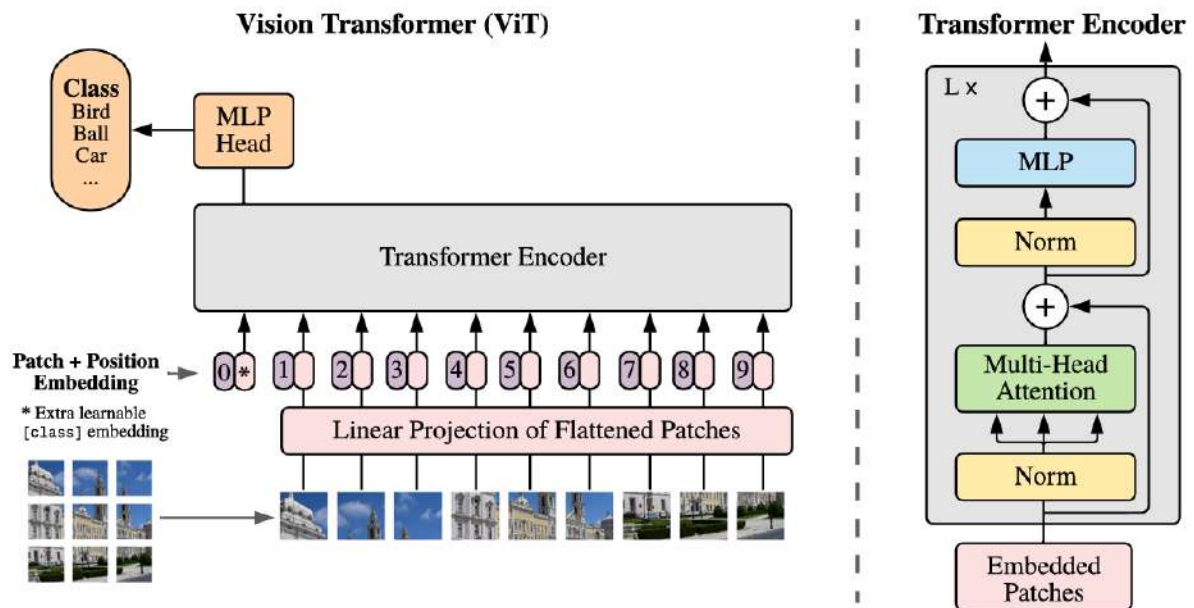


Figure 3.12 – Schéma de l'architecture ViT [68]

2 - Modèles hybrides CNN-Transformer

Les architectures hybrides CNN-Transformer ont été proposées comme une solution prometteuse, combinant les biais inductifs locaux des CNN avec les capacités de modélisation globale des Transformers, permettant de surmonter les limites et les faiblesses de chacun, et améliorant ainsi les performances. En général, dans un modèle hybride, le CNN joue principalement le rôle d'extraction d'informations locales et d'introduction d'un biais inductif, tout en réduisant la charge computationnelle liée au traitement d'images haute résolution, améliorant ainsi l'efficacité globale. Le rôle principal du ViT réside dans la modélisation de l'information contextuelle globale, capturant les dépendances à longue portée entre différentes régions de l'image. Cette combinaison permet au modèle d'extraire des représentations de caractéristiques plus riches, renforçant sa compréhension de l'image. D'autre part, grâce à cette intégration structurelle, la complexité computationnelle peut être efficacement réduite, améliorant encore l'efficacité [69].

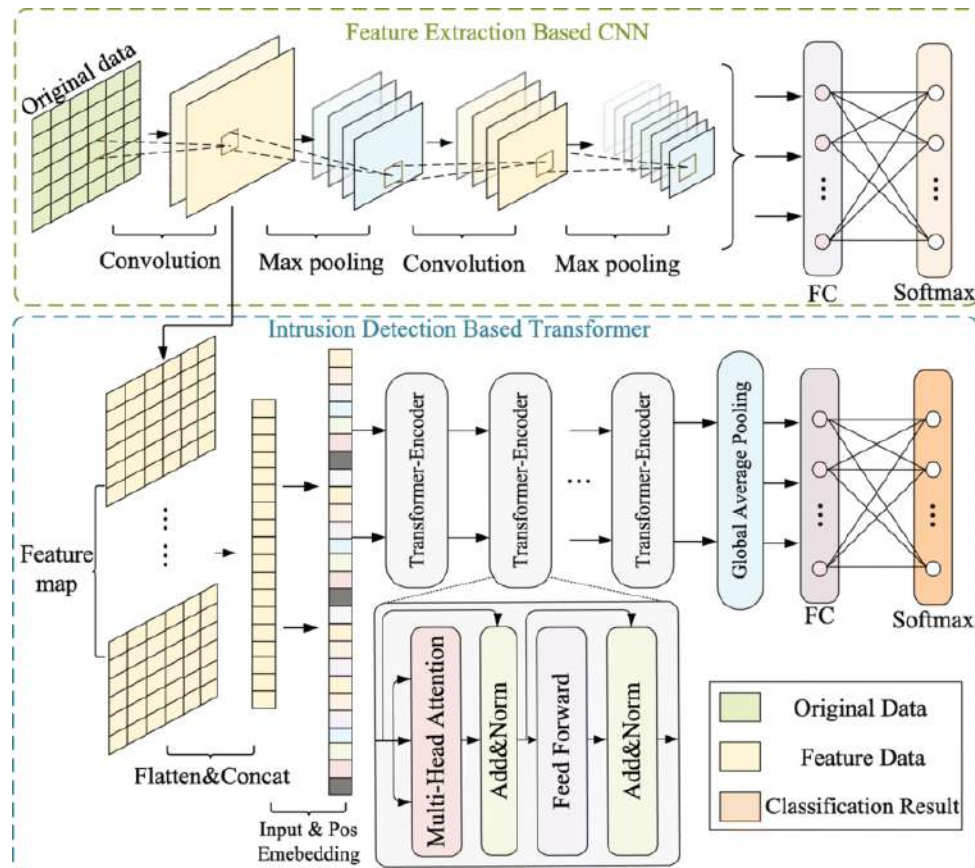


Figure 3.13 – Architecture du réseau hybride CNN-Transformer [70]

3 - EfficientNetV2

EfficientNetV2 est une architecture révolutionnaire de réseau de neurones convolutionnel proposée par l'équipe Google Research dirigée par Tan et Le (2021). EfficientNetV2 atteint un équilibre entre la taille des paramètres, l'efficacité computationnelle et la précision, ce qui le rend adapté à divers scénarios de calcul. Il a été largement utilisé pour des tâches telles que la classification d'images, la détection d'objets et la segmentation sémantique.[71]

EfficientNetV2 offre :

- Une réduction de 90 % du nombre de paramètres par rapport à ResNet, tout en conservant une précision similaire ;
- Un entraînement et une inférence plus rapides ;
- Une meilleure adaptation au déploiement sur des dispositifs périphériques (échographie mobile).

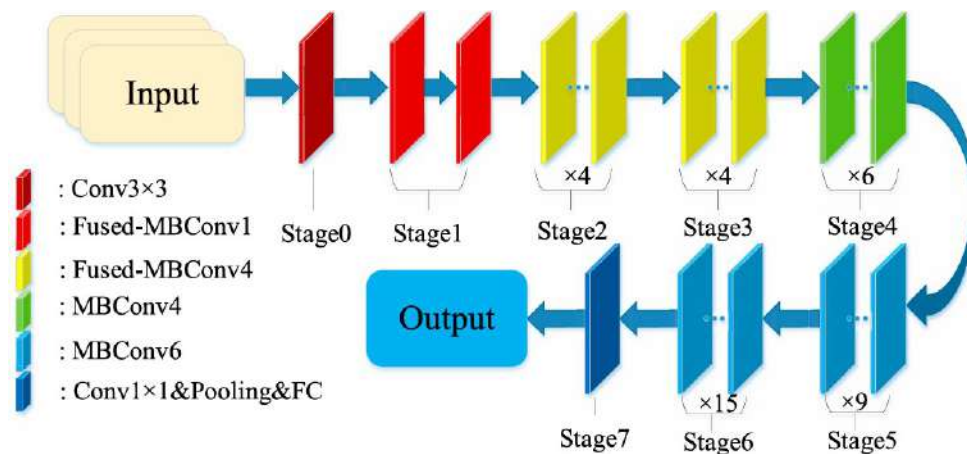


Figure 3.14 – Structure de l'architecture du EfficientNetV2 [72]

4 - Multi-Scale Feature Pyramids

Les pyramides de caractéristiques sont un composant fondamental des systèmes de reconnaissance pour la détection d'objets à différentes échelles. Cependant, les détecteurs d'objets récents basés sur l'apprentissage profond ont évité les représentations pyramidales, en partie parce qu'elles sont coûteuses en calcul et en mémoire. Une architecture descendante avec des connexions latérales a été développée afin de générer des cartes de caractéristiques sémantiques de haut niveau à toutes les échelles. Cette architecture, appelée Feature Pyramid Network (FPN), permet d'améliorer l'extraction de caractéristiques et constitue une solution efficace pour la détection d'objets multi-échelles[73].

Les réseaux de pyramides de caractéristiques (FPN) ou les réseaux d'agrégation de chemins (PA-Net) prennent en charge :

- Des nodules de tailles variables (de 2 mm à 40 mm) ;
- La détection simultanée à plusieurs échelles ;
- Une meilleure détection des petits nodules (essentielle pour un dépistage précoce).

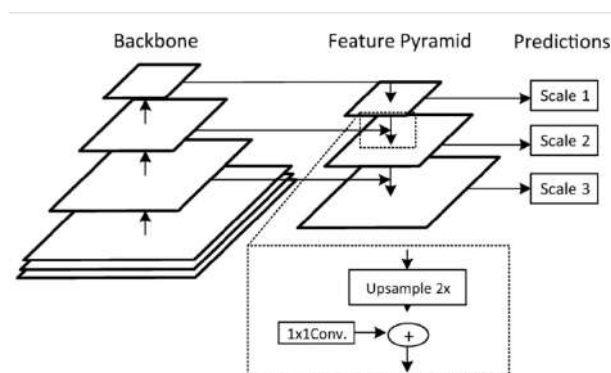


Figure 3.15 – Architecture de feature pyramid network (FPN) [74]

3.2.3.3 Transfert learning

L'apprentissage par transfert est une technique d'apprentissage automatique regroupe l'ensemble des méthodes qui aident au transfert des acquisitions à partir de la résolution d'un problème donné, pour en résoudre un autre. Il s'agit d'une technique en un modèle d'intelligence artificielle, déjà entraîné sur une mission bien définie, est adapté pour une autre mission similaire, mais plus ciblée. Cette approche utilise les facultés déjà acquises afin d'améliorer l'efficacité et la performance du modèle sur la nouvelle mission. L'objectif est d'utiliser moins de temps de travail, mais aussi moins de données [75] (voir Figure 3.16).

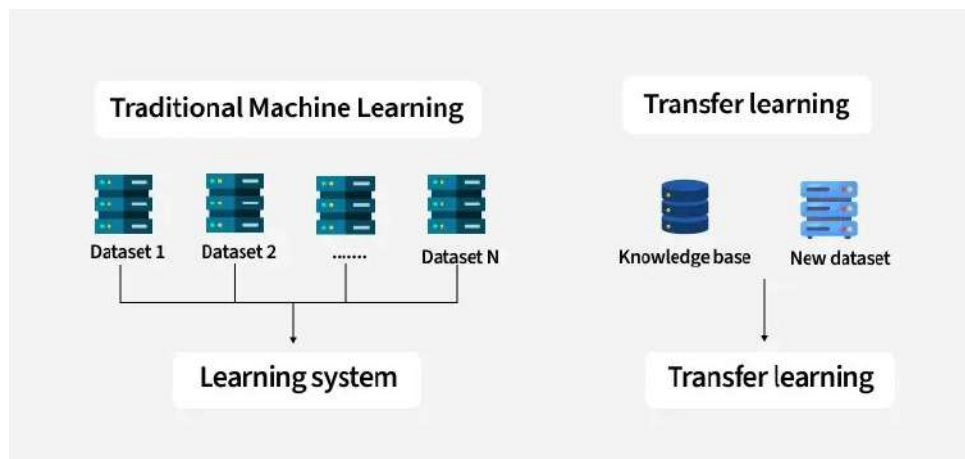


Figure 3.16 – Illustration du transfert learning [76]

L'apprentissage par transfert vous permet de réutiliser des modèles pré-entraînés pour de nouvelles tâches. Les principales techniques sont :

1. Feature Extraction : consiste à prendre un modèle pré-entraîné sur un large ensemble de données connexes et à l'adapter à une nouvelle tâche cible, souvent avec moins de données.
2. Fine-Tuning : est une forme spécifique d'apprentissage par transfert dans laquelle tout ou partie des couches d'un modèle pré-entraîné sont réentraînées sur de nouvelles données pour accomplir la tâche cible.
3. Apprentissage Multitâche : entraîne un modèle unique à accomplir simultanément plusieurs tâches connexes, en partageant des représentations communes entre ces différentes tâches.
4. Apprentissage Fédéré : est une approche d'entraînement décentralisée dans laquelle le modèle est entraîné sur plusieurs appareils ou serveurs détenant des échantillons de données locales, sans qu'aucun échange de données brutes ne soit nécessaire.[76]

3.2.4 Le sur-ajustement (Overfitting)

L'overfitting se produit lorsqu'un modèle apprend excessivement les détails et le bruit des données d'entraînement, perdant ainsi sa capacité de généralisation sur de nouvelles données (voir Figure 3.17). Cela se manifeste par une erreur d'entraînement plus faible accompagnée d'une erreur de validation/test plus élevée. Ce phénomène peut être causé par plusieurs facteurs, notamment la complexité du modèle (nombre élevé de couches, de neurones et de paramètres), l'absence de mécanismes de régularisation (Dropout), Les données insuffisantes ou non représentatives, ainsi que par la mémorisation au lieu de la généralisation.

Le contraire est le sous-apprentissage (underfitting), se produit lorsqu'il existe encore une marge d'amélioration dans l'apprentissage des données d'entraînement. Lorsque les erreurs de prédiction sur les ensembles de données d'entraînement et de validation sont élevées et la différence entre les deux est très faible, le modèle est considéré comme sous-ajusté [77].

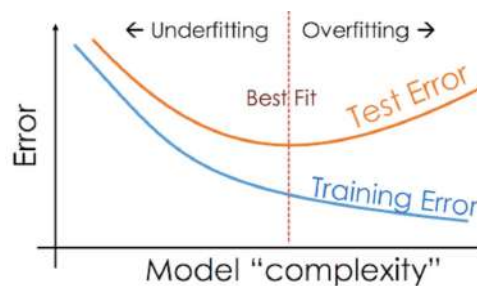


Figure 3.17 – Illustration du Overfitting et Underfitting [78]

Dans le cas où il y aurait peu d'exemples d'apprentissage, le modèle peut parfois apprendre des bruits ou voir ces performances se dégrader en raison de la présence de détails indésirables dans les exemples d'apprentissage du modèle sur les nouveaux exemples (test/validation) [77]. Il existe plusieurs techniques pour lutter contre ce phénomène dans le processus d'apprentissage, parmi lesquelles :

- Augmentation des données ;
- Utilisation de Dropout dans le modèle afin de désactiver aléatoirement certains neurones pendant le processus d'apprentissage ;
- Ajustement des hyperparamètres (nombre de couche, type de fonction d'activation, learning rate, optimiseur [SGD, Adam,...], batch-size, nombre d'époque,...) ;
- Validation croisée K-Fold ;
- Arrêt précoce (EarlyStopping) ;
- Normalisation des données ;
- ... etc

3.3 Détection et Classification en Deep Learning

3.3.1 Détection de Nodules Thyroïdiens

3.3.1.1 Définition

La détection d'objets est une tâche de vision par ordinateur qui consiste à identifier et localiser les objets présents dans des images numériques. Elle fait partie application de l'intelligence artificielle visant à doter aux ordinateurs d'une capacité de compréhension visuel de manière similaire à l'être humain, notamment en identifiant et en classant les objets selon des catégories sémantiques.

La détection d'objets combine ces deux sous-tâches :

- La localisation d'objet : est une méthode permettant de déterminer la position d'un objet dans une image en le délimitant au moyen d'une boîte englobante.
- La classification d'objet : elle consiste à attribuer à l'objet détecté une catégorie correspondante. Ainsi la détection combine ces deux approches, la détection d'objets permet afin d'estimer simultanément l'emplacement et la nature des différentes instances d'objets dans une ou plusieurs images [79].

3.3.1.2 Les modèles de détection

3.3.1.2.1. Faster RCNN

Un Faster R-CNN est une méthode de détection d'objets basée sur un réseau neuronal, permettant d'identifier et localiser des objets dans une image à l'aide de boîtes englobantes. Elle s'appuie sur un réseau neuronal convolutionnel est une technique pour extrait les caractéristiques visuelles puis utilise un réseau de propositions de régions afin de générer susceptibles de contenir des objets. Ces régions sont ensuite analysées et classées par un ensemble de couches entièrement connectées pour produire les résultats finaux de reconnaissance d'objets. L'algorithme est appelé « Faster » R-CNN reflète son amélioration en termes plus rapide et plus précise par rapport les algorithmes précédents de détection d'objets tels que R-CNN. Il est largement utilisé dans des applications, notamment les véhicules autonomes, la détection faciale et la surveillance d'objets [80].

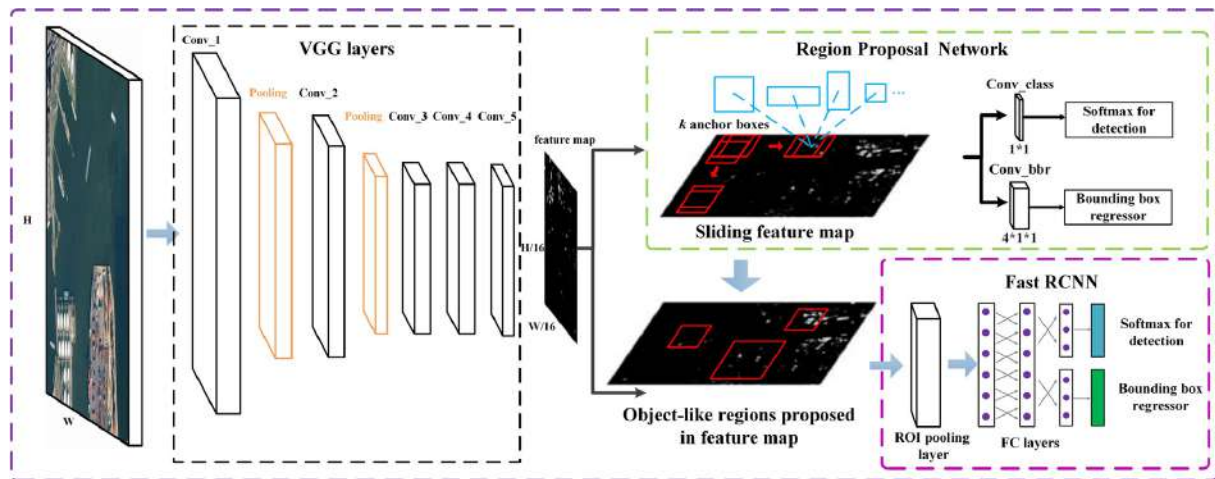


Figure 3.18 – Architecture du modèle Faster R-CNN [81]

3.3.1.2.2. YOLO

You Only Look Once (YOLO) est un algorithme d'apprentissage profond largement utilisé pour la détection d'objets. Il repose sur une approche combinant classification et régression, ce qui lui confère à plusieurs avantages notamment l'architecture est simple, une taille de modèle réduite et une vitesse de calcul rapide. Le mécanisme de YOLO consiste à redimensionner l'image d'entrée à une taille fixe, puis à la diviser l'image en grille de $S \times S$ cellules. Chaque cellule est chargée de détecter les objets dont le centre se trouve à l'intérieur. Pour cela, chaque cellule prédit plusieurs boîtes de détection, chacune calculant non seulement sa propre position, mais aussi un score de prédiction. Ce score représente la probabilité qu'un objet détecté soit présent dans la cellule de réseau prédite. Comme il peut y avoir plusieurs boîtes dans une cellule, YOLO sélectionne automatiquement celles ayant le score le plus élevé pour la prédiction afin d'identifier la catégorie d'objet et sa localisation [82].

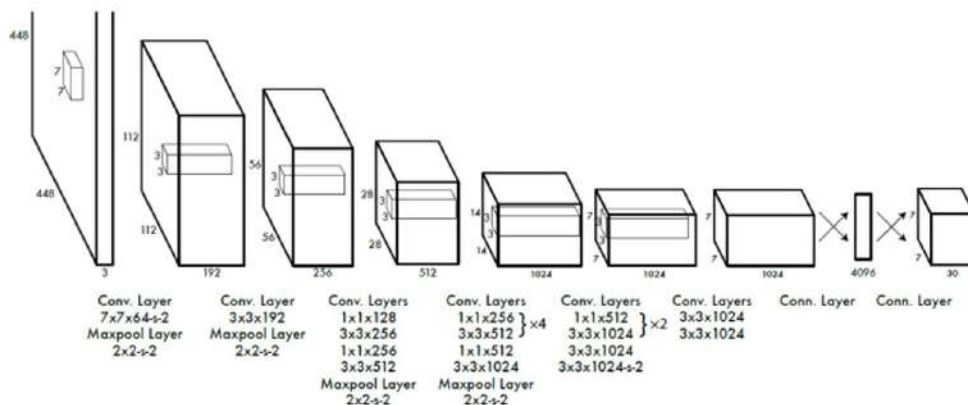


Figure 3.19 – L'architecture de l'algorithme YOLO [83]

Évolution des modèles YOLO

Le modèle YOLO a considérablement évolué au fil du temps, et chaque nouvelle version a apporté des améliorations à son architecture par rapport à la précédente en termes de vitesse, de précision et de fonctionnalités. Les principales versions de YOLO sont les suivantes : YOLOv1 (2016), YOLOv2 (2016), YOLOv3 (2018), YOLOv4 (2020), il existe d'autres versions proposées par Ultralytics comme YOLOv5 (2020) et YOLOv8 (2023), YOLOv11 (2024), YOLO26 (2026) [Hidayatullah et al., 2026 [84]]. De plus, pour chaque version de YOLO, il existe différentes variantes telles que Nano (n), Small (s), Medium (m) et Large (l), qui varient en taille, en complexité et en fonction du nombre de paramètres du modèle.

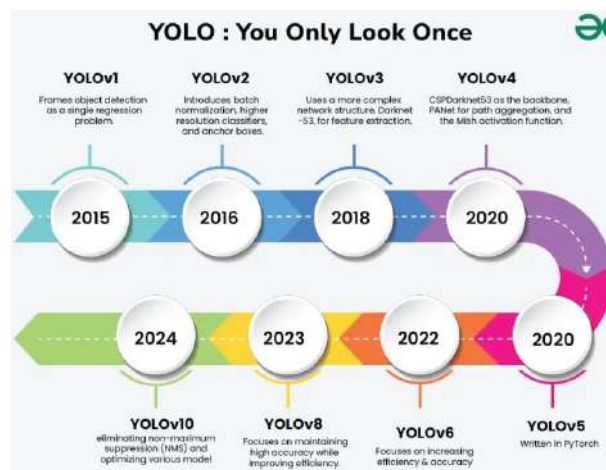


Figure 3.20 – Évolution des versions de YOLO [85]

- YOLOv8

YOLOv8 est un modèle de vision par ordinateur conçu pour la détection d'objets, la segmentation d'images et la classification. Sa architecture est la dernière version des modèles de détection d'objets YOLO, connue pour sa rapidité et sa précision. Développé par Ultralytics, il introduit des améliorations pour la détection d'objets en temps réel. Contrairement aux méthodes traditionnelles basées sur des fenêtres glissantes, YOLO traite la détection comme un problème de régression unique et prédit directement les encadrements et les probabilités de classe en un seul passage sur l'image, ce qui le rend plus rapide [86].

- YOLO26

Ultralytics YOLO26 représente la plus récente version de la série YOLO de détecteurs d'objets en temps réel, élaborée intégralement pour les appareils embarqués et à consommation réduite. Il propose une approche simplifiée qui sup-

prime la complexité superflue tout en incorporant des innovations spécifiques pour fournir un déploiement plus rapide, léger et accessible [87].

- YOLOE

Un modèle puissant et efficace qui fonctionne comme un œil humain, reconnaissant n'importe quel objet à travers divers types d'invites : invites basées sur du texte, indices visuels, ou même sans aucune invite. Il s'appuie sur les forces des modèles YOLO mais est conçu pour une utilisation plus flexible dans le monde réel, tout en conservant la vitesse et la légèreté qui ont rendu YOLO célèbre [88].

3.3.1.2.3. DETR

DETR (Détection Transformer) est un modèle avancé en vision par ordinateur qui exploite l'architecture Transformer, initialement développée pour le traitement du langage naturel, afin de réaliser des tâches telle que la détection et la classification d'objets dans les images. Contrairement aux modèles classiques basées sur les réseaux convolutionnels ou des méthodes (comme YOLO), DETR est un modèle adopte une stratégie de prédiction directe par ensembles qui utilise une architecture encodeur-décodeur de type Transformer pour détecter simultanément les objets présenter dans une image. Cette approche permet de simplifier le processus de détection tout en offrant une solution efficace et élégante par rapport aux méthodes traditionnelles [89].

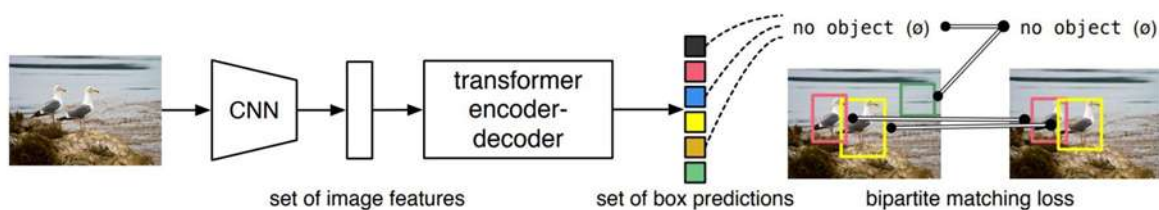


Figure 3.21 – Architecture du DETR [90]

3.3.1.2.4. RF-DETR

RF-DETR (Roboflow Detection Transformer) est une famille de modèle de détection d'objets basée sur l'architecture DETR et les transformeurs. Il intègre un backbone DINOv2 pré-entraîné, des mécanismes d'attention déformable, ainsi que par un pipeline modulaire sans ancres. Parmi les principales contributions

incluent l'utilisation de la recherche d'architecture neuronale (NAS) avec partage de poids pour recevoir des détecteurs optimisés en temps réel, une généralisation dans des environnements encombrés ou ambigus, ainsi que des stratégies d'adaptation au changement de domaine, incluant des techniques telles que le hard négative mining [91]. RF-DETR repose sur une architecture basée sur les transformeurs inspirée DETR, enrichie par des améliorations du champ réceptif afin d'augmenter à la fois la précision et la rapidité. Contrairement aux modèles YOLO traditionnel, qui s'appuient sur des ancres et de la suppression non maximale (NMS)[92].

3.3.2 Classification de Nodules Thyroïdiens

3.3.2.1 Définition

La classification d'images consiste à attribuer des étiquettes aux images selon des catégories prédéfinies en analysant l'image d'entrée afin d'identifier la classe correspondante. Dans le domaine médical, notamment pour les images échographiques thyroïdiennes, elle vise à identifier et classer les nodules thyroïdiens. Étant donné que l'analyse manuelle de ces images est complexe et dépend fortement de l'expertise des médecins, plusieurs approches automatiques ont été développées pour localiser, segmenter et classer les nodules thyroïdiens, afin d'améliorer la précision et la fiabilité du diagnostic grâce aux avancées d'apprentissage automatique et à la vision par ordinateur.

Il existe différents types de classification, parmi eux on peut citer :

3.3.2.2 Types de classifications

• Classification Binaire

Aussi appelée classification binomiale, est la forme la plus simple de classification supervisée, où chaque donnée est assignée à l'une des deux classes possibles. Le modèle analyse les caractéristiques d'entrée afin de déterminer à quelle classe elle appartient [93] (voir Figure 3.22).

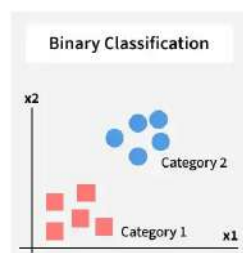


Figure 3.22 – Exemple de classification Binaire [93]

• Classification Multi-classes

Elle est utilisée lorsque les données doivent être réparties entre plus de deux catégories. Le modèle évalue les caractéristiques de chaque entrée afin de lui attribuer une seule classe parmi plusieurs classes possibles [93] (voir Figure 3.23).

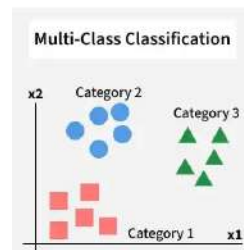


Figure 3.23 – Exemple de classification multi-classes [93]

• Classification multi-label

Permet à une même donnée d'appartenir simultanément à plusieurs catégories. Contrairement à la classification multiclasse, les classes ne sont pas mutuellement exclusives, et le modèle peut assigner plusieurs étiquettes à une même instance [93] (voir Figure 3.24).

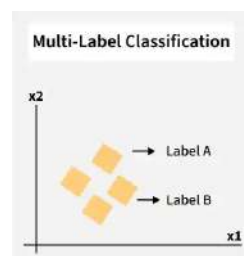


Figure 3.24 – Exemple de classification multi-label [93]

3.4 État de l'art

Cette section est dédiée à la présentation de quelques travaux récents dans le domaine du diagnostic automatique des maladies thyroïdiennes. Nous avons étudié des approches basées sur les modèles de apprentissage profond.

Il existe de nombreux ouvrages et études différentes dans la littérature consacrée à l'exploitation des images échographiques, ainsi qu'à la détection et à la segmentation des nodules thyroïdiens. Dans le cadre de notre recherche, nous avons sélectionné un ensemble d'articles pertinents.

Dans ce cadre [Tao et al., 2022 [94]] ont proposé un modèle basé sur l'architecture ResNet-50, enrichie par l'ajout de modules d'attention, afin de réaliser

un diagnostic différentiel des nodules thyroïdiens classés entre les catégories 3 et 5 du système TI-RADS. Ce modèle vise à exploiter et à fusionner des informations issues de différentes modalités d'imagerie, notamment l'échographie mode B (GSU), le Doppler couleur (CDFI), l'élastographie de déformation (SE), ainsi que les masques de région d'intérêt (ROI). L'étude s'appuie sur un ensemble de données comprenant 1138 nodules thyroïdiens, tous confirmés par examen histopathologique et acquis en coupes transversales et longitudinales. Les résultats obtenus montrent des performances prometteuses pour la détection du cancer thyroïdien, avec une précision de 86,1%, une AUC de 0,928, une sensibilité de 84,7% et une spécificité de 87,5%, mettant ainsi en évidence l'intérêt de l'intégration multimodale.

De leur côté [Wang et al., 2022 [95]] ont proposé un modèle basé sur un réseau de neurones convolutif entièrement convolutionnel utilisant le principe du soft label learning (SL-FCN), visant à améliorer à la fois le traitement ciblé du cancer du sein et le diagnostic du cancer de la thyroïde à partir de plusieurs jeux de données. Pour le cancer du sein, leur approche permet la segmentation automatique de l'amplification du récepteur HER2 dans des images FISH et DISH. Pour faciliter le diagnostic du cancer de la thyroïde, la méthode proposée segmente automatiquement le carcinome papillaire de la thyroïde (CPT) à partir d'images issues de prélèvements par aspiration à l'aiguille fine et de lames entières colorées selon la méthode de Papanicolaou. Le modèle a été évalué et comparé à treize méthodes d'apprentissage profond récentes ou populaires. Les résultats expérimentaux montrent des performances élevées, sur l'ensemble de données DISH, la méthode proposée atteint une précision comprise entre 87,77% à 94,64%, avec des scores F1 compris entre 81,67% et 85,14%. Tandis que l'ensemble de données FISH, obtient une précision de 93,54% et un score F1 de 86,98%. Par ailleurs, les résultats sont statistiquement supérieurs à la plupart des méthodes de référence ($p < 0,05$). En ce qui concerne la segmentation du carcinome papillaire de la thyroïde, la modèle atteint une précision élevée de 99,99% et un score F1 de 89,82%.

Dans une autre étude, [Zhao et al., 2023 [96]] ont évalué quatre modèles de réseaux de neurones convolutifs profonds (DCNN) pré-entraînés, notamment AlexNet, ShuffleNetV2, MobileNetV3 et ResNet-34, pour la classification de différentes pathologies thyroïdiennes à partir de 3194 images SPECT. Les classes considérées incluent la maladie de Basedow, la thyroïdite subaiguë, les tumeurs thyroïdiennes et les cas normaux. Le modèle ResNet-34 amélioré a obtenu les meilleurs résultats avec une précision globale de 94,4% et de 93,1% en validation interne. Pour garantir la fiabilité de ces résultats, le modèle a été

soumis à une validation croisée à cinq plis afin d'évaluer plus précisément ses performances, et ses performances diagnostiques ont été comparées à celles de médecins nucléaires juniors et seniors. De plus, les régions d'attention spécifiques à chaque classe ont été visualisées à l'aide de cartes de chaleur basées sur la méthode Grad-CAM, mettant en évidence la capacité du modèle à localiser les zones pertinentes pour la classification.

Pour apporter plus de précision, [Nguyen et al., 2023 [97]] ont étudié l'apport des techniques d'intelligence artificielle explicable (XAI) dans le diagnostic des nodules thyroïdiens, en particulier leur impact sur la confiance des médecins et des patients. Leur approche repose sur l'utilisation de modèles CNN pour la détection des nodules, combinés à des méthodes d'explicabilité telles que Grad-CAM, l'estimation de densité par noyau (Kernel Density Estimation) et les cartes de densité. L'étude a été menée sur un ensemble de 14171 images échographiques provenant de 970 patients vietnamiens. Contrairement aux travaux axés principalement sur les performances quantitatives, cette étude met l'accent sur l'évaluation qualitative à travers des enquêtes auprès des utilisateurs, sans rapporter de métriques classiques telles que la précision. Les résultats montrent que l'intégration de méthodes XAI, notamment Grad-CAM, contribue significativement à renforcer la confiance des médecins dans les systèmes automatisés de diagnostic.

Dans ce cadre [Amgad et al., 2024 [98]] ont proposé une approche basée sur des techniques d'ensemble, en exploitant différentes versions partitionnées de modèles d'apprentissage profond afin d'améliorer la classification du cancer de la thyroïde. Leur méthode a été évaluée sur plusieurs architectures, notamment ResNet50, MobileNetV2, VGG19, DenseNet121 et ViT, dans le but d'obtenir une classification précise et efficace du cancer de la thyroïde. Les expériences ont été réalisées à partir d'examens cytologiques collectés au sein de l'Unité de Cytopathologie du NCI, Université du Caire, entre 2015 et 2021. Cet ensemble comprend 2008 images RVB, réparties en deux classes principales : adénomes (859 images) et carcinomes (1149 images). Toutes les images ont été redimensionnées à une résolution de 224 x 224 pixels. En particulier, les résultats montrent que le modèle ResNet50 est le plus performant parmi les approches testées, avec une précision de 92%, une spécificité de 95,5%, une sensibilité de 90,6%, un score F1 de 92,9% ainsi qu'une AUC de 97,4%.

En faisant référence à d'autres techniques basées sur l'architecture CNN, [Archana et al., 2024 [99]] examinent l'intégration des techniques d'apprentissage profond pour le diagnostic des pathologies thyroïdiennes. Face au be-

soin croissant d'un diagnostic rapide et précis des différentes affections thyroïdiennes, notamment l'hypothyroïdie, l'hyperthyroïdie, la thyroïdite, les nodules, le cancer et l'euthyroïdie. Les auteurs explorent ainsi le potentiel des CNN comme alternative aux méthodes de diagnostic conventionnelles. Les données utilisées proviennent de diverses sources médicales et incluent différents types d'images, telles que les images échographiques, les IRM et les analyses histopathologiques. L'image a d'abord été redimensionnée à 96×96 pixels avec une profondeur de 3 (correspondant au canal RGB) afin d'accélérer le traitement. Par suit, Toutes les images du dossier sont redimensionnées au format 220×220 pixels avant d'être traitées par un modèle d'apprentissage automatique utilisant différentes méthodes. Ce redimensionnement garantit l'homogénéité des images dans l'ensemble du document. Les résultats expérimentaux montrent des performances élevées, avec une précision de 97% sur l'ensemble de test.

Dans un autre travail sur les images médicales thyroïdiennes, [Wen et al., 2025 [100]] ont développé des modèles d'apprentissage automatique multimodaux afin d'améliorer la précision diagnostique des nodules thyroïdiens hypervascularisés (grades d'Adler 2 et 3), souvent difficiles à caractériser en échographie standard. Leur méthode repose sur l'intégration de caractéristiques issues de l'imagerie en mode B et du Doppler de puissance (PDI), à partir d'une cohorte rétrospective de 315 patients. Les images échographiques multimodales ont été traitées à l'aide d'un modèle de segmentation basé sur l'apprentissage profond, combiné à une technique de seuillage du canal rouge, suivie de l'extraction de caractéristiques radiomiques (1910 caractéristiques via PyRadiomics) et d'une dérivation de caractéristiques par apprentissage profond (1000 caractéristiques extraites à partir de ResNet). Cinq algorithmes d'apprentissage automatique ont été évalués, notamment la forêt aléatoire, la régression logistique, SVM, XGBoost et le réseau TABPFN. Parmi eux, le SVM a obtenu les meilleures performances dans les analyses unimodales : en mode B, il a atteint une AUC de 0,89 (précision : 0,84 ; rappel : 0,94), tandis qu'en mode PDI, il a atteint une AUC de 0,86 (précision : 0,82 ; rappel : 0,97).

C'est dans ce cadre [Hanani et al., 2025 [101]] ont exploré l'utilisation d'un classificateur supervisé CatBoost pour prédire les récurrences chez les patients atteints de cancers thyroïdiens bien différenciés, en s'appuyant sur un jeu de données comprenant 383 patients et intégrant des variables démographiques, cliniques et pathologiques. Les performances du modèle ont été comparées à celles d'autres méthodes d'ensemble, telles que Extra Trees, LightGBM et XGBoost. Par ailleurs, une analyse SHAP a été utilisée afin d'évaluer l'importance des variables et d'améliorer l'interprétabilité du modèle. Les résultats obtenus

montrent que le modèle CatBoost surpasse les approches concurrentes, avec une précision de 97% et une aire sous la courbe ROC de 0,99. De plus, les analyses SHAP locales ont permis de mieux interpréter les prédictions individuelles, révélant que certaines erreurs de classification étaient liées à une surévaluation d'un seul facteur au détriment d'autres indicateurs cliniquement pertinents. Ces résultats confirment à la fois la performance prédictive élevée et l'intérêt interprétable de l'approche proposée.

[Rashed et al., 2025 [102]] ont évalué l'efficacité des réseaux de neurones convolutifs (CNN) avec apprentissage par transfert pour la classification des nodules thyroïdiens à partir d'un jeu de données échographiques public, validé par biopsie, composé de 483 images (197 bénignes et 286 malignes). Neuf CNN pré-entraînés ont été étudiés, notamment ResNet50, ResNet101, VGG16, VGG19, DenseNet121, EfficientNetB0, InceptionV3, InceptionResNetV2 et Xception. Les modèles ont été entraînés en utilisant des techniques d'augmentation des données, d'équilibrage des classes et une validation croisée à dix plis. Les résultats montrent que le modèle ResNet50 a obtenu les meilleures performances, avec une précision de 96,90% et une aire sous la courbe ROC de 0,97, suivi par ResNet101 et d'EfficientNetB0. Les autres modèles ont enregistré des précisions comprises entre 87% et 90%, avec des valeurs d'AUC variant de 0,89 à 0,93. Ces résultats soulignent la supériorité de ResNet50 et confirment le potentiel de l'apprentissage par transfert basé sur les CNN comme outil fiable d'aide à la décision pour la classification des nodules thyroïdiens.

Aussi dans ce cadre que [Tathagat et al., 2025 [103]] ont proposé une architecture d'apprentissage profond innovante, nommée TATHA, dédiée à l'amélioration de la segmentation des images échographiques de la thyroïde. Le modèle a été évalué sur une base de données comprenant 99 cas. Il intègre des procédures de prétraitement visant à réduire le bruit de speckle et à améliorer le contraste, tandis que des techniques de détection des contours sont utilisées pour fournir des entrées de meilleure qualité au processus de segmentation. Les résultats expérimentaux montrent des performances élevées sur l'ensemble des plis de validation, avec une précision comprise entre 94,1% et 95,3%, des scores de Dice variant de 0,9065 à 0,9214, ainsi qu'une AUC comprise entre 0,9312 et 0,9442. Ces résultats confirment l'efficacité du modèle TATHA et soulignent son potentiel en tant qu'outil d'aide à la décision pour les cliniciens et les chercheurs en imagerie thyroïdienne.

[Wang et al., 2025 [104]] ont proposé un modèle d'apprentissage profond innovant, nommé ThyroNet-X4 Genesis, pour classifier automatiquement les no-

dules thyroïdiens en bénins ou malins. Basé sur l'architecture ResNet, ce modèle améliore l'extraction des caractéristiques grâce à l'intégration de convolutions groupées et de noyaux de convolution de plus grande taille, optimisant ainsi l'analyse des images échographiques. Le modèle a été entraîné et validé sur des jeux de données public, puis évalué sur un jeu de validation externe provenant de l'hôpital central de HanZhong. Les résultats obtenus montrent une précision de 85,55% en phase d'entraînement, 71,70% en validation interne et 67,02% en validation externe. Malgré une baisse de performance sur les données externes, le modèle surpasse plusieurs approches classiques, ce qui souligne son potentiel pour une application clinique dans la classification des nodules thyroïdiens.

Dans un autre travail, [Vahdati et al., 2024 [105]] ont proposé un pipeline basé sur l'apprentissage profond pour la détection et la classification des nodules thyroïdiens en bénins ou malins, à partir d'images échographiques transversales et longitudinales issues de 983 patients. Deux modèles YOLO v5 ont été entraînés séparément pour chaque type d'incidence, tandis que cinq modèles ont été développés dans le cadre d'une validation croisée, puis combinés à l'aide de la méthode de suppression non maximale (NMS) afin d'améliorer la généralisation. Les sorties de ces modèles ont ensuite été exploitées par un classificateur XG-Boost pour produire la prédiction finale de malignité. Les modèles d'ensemble ont atteint un $mAP@0,5$ de 0,797 pour les images transversales et de 0,716 pour les images longitudinales. Le pipeline complet a obtenu une AUROC de 0,84 (IC à 95% : 0,75-0,91), avec une sensibilité de 84% et une spécificité de 63%. En comparaison, la méthode ACR-TIRADS appliquée au même ensemble de données a présenté une sensibilité de 76% et une spécificité de 34% ($p = 0,003$). Ces résultats démontrent la supériorité du modèle proposé pour l'évaluation diagnostique des nodules thyroïdiens.

Dans cette partie nous avons présenté différents travaux de la littérature relatifs au domaine de notre projet. Afin de simplifier et résumer ces différentes méthodes, le tableau 3.1 représente le récapitulatif sur l'ensemble des méthodes résumées ci-dessus.

Auteurs	Dataset	Méthodologies	Résultats
[Tao et al., 2022] [94]	Thyroid Ultra-sound Dataset (1138 nodules thyroïdiens, 781 patients)	Modèle Deep Learning basé sur ResNet-50 + modules d'attention pour fusionner caractéristiques multimodales (GSU, Doppler, elastographie, ROI masks) et classifieur bénin vs malin	Accuracy=86.1%, AUC=0.928, Sensitivity=84.7%, Specificity=87.5%
[Wang et al., 2022] [95]	4 datasets cliniques : 2 DISH breast (270), 1 FISH breast (200), 131 PTC thyroid WSIs	Soft Label FCN (fully convolutional network) pour segmentation automatique des images; soft-label learning (soft weight soft-max loss) Comparé à 13 modèles (U-Net, SegNet, DeepLabv3+, YOLOv5, etc.)	DISH1 : Acc=87,77% DISH2 : Acc=94,64% FISH : Acc=93,54% Thyroïde (PTC) : Acc=99,99%
[Zhao et al., 2023] [96]	3194 SPECT scintigraphy thyroid images multicentriques (training n=2067, internal val n=514, external val n=613)	Comparaison de 4 modèles CNN pré-entraînés (AlexNet, ShuffleNetV2, MobileNetV3, ResNet-34) Amélioration du modèle; modèle retenu évalué par 5-fold CV; interprétation via Grad-CAM	Meilleur modèle : ResNet-34 amélioré Accuracy=94,4% (validation interne) Accuracy=93,1% (validation externe)
[Nguyen et al., 2023] [97]	14171 échographies thyroïdiennes (970 patients vietnamiens)	CNN pour détection de nodules + XAI methods (Grad-CAM, Kernel Density Estimation, Density map)	Étude de la confiance des médecins et patients via sondage; pas de métrique de précision

[Amgad et al., 2024] [98]	Images cytologiques FNA du National Cancer Institute – Cairo University (2008 images)	Plusieurs modèles Deep Learning : ResNet50, DenseNet121, VGG19, MobileNetV2, Vision Transformer + techniques Ensemble Learning	ResNet50 : Accuracy=92% specificity=95.5%, sensitivity=90.6%, F1-score=92.9% AUC=97.4% Meta-learning : Accuracy=95.1%
[Archana et al., 2024] [99]	Images médicales de la thyroïde (radiographie et échographie) contenant plusieurs types de maladies thyroïdiennes	Utilisation d'un modèle Convolutional Neural Network (CNN) aux images X-ray et Ultrasound avec prétraitement, optimisation et élimination du bruit	Accuracy=97%
[Wen et al., 2025] [100]	315 patients, échographie B-mode + PDI	Fusion multimodale combinant radiomiques + deep learning features (SVM, RF, XGBoost, etc.)	Modèle multimodal montre 89% (AUC test) 0.86 (F1-score)
[Hanani et al., 2025] [101]	Données cliniques de patients avec cancer thyroïdien	CatBoost supervisé + explicabilité SHAP sur données cliniques pour prédiction de récurrence	Meilleure précision de prédiction de récurrence qu'avec d'autres classifieurs, et interprétabilité des facteurs de risque via SHAP

[Rashed et al., 2025] [102]	Dataset public d'images échographiques de la thyroïde (483 images, confirmé par biopsie)	Transfer Learning avec CNNs pré-entraînés (ResNet50, ResNet101, VGG16/19, DenseNet121, EfficientNetB0, InceptionV3, InceptionResNetV2, Xception), augmentation des données, équilibrage des classes, validation croisée 10-fold	ResNet50 : Accuracy=96.90% AUC=0.97 Precision=96.93% Recall=96.90% F1-score=96.90%
[Tathagat et al., 2025] [103]	Images échographiques thyroïdiennes	Approche hybride de segmentation thyroïdienne combinant un réseau de deep learning avec mécanisme d'attention sur les caractéristiques, extraction multi-échelle et validation statistique pour renforcer la robustesse et la fiabilité clinique	Accuracy=94.1% Sensitivity=93.4%
[Wang et al., 2025] [104]	DDTI Thyroid Ultrasound Images+ dataset externe (HanZhong Central Hospital)	ThyroNet-X4 Genesis, Utilisation des (CNN) basé sur ResNet pour classifier automatiquement les nodules thyroïdiens bénins et malins à partir d'images échographiques.	Accuracy=71.70% (validation interne) Accuracy=67.02% (validation externe)
[Vahdati et al., 2024] [105]	Images échographiques de la thyroïde de 983 patients avec des vues transversale et longitudinale	Utilisation de modèles YOLOv5 pour détection et classification des nodules thyroïdiens à partir de vues transversale et longitudinale, avec fusion des prédictions via XGBoost pour une décision finale bénin/malin.	AUROC=0.84 Sensitivity=84% Specificity=63% F1 score=0.76

Table 3.1 – Récapitulatif de certaines études proposées dans la littérature pour le diagnostic des nodules thyroïdiens

3.5 Conclusion

Ce chapitre a été consacré à la présentation des notions de base de l'intelligence artificielle, ainsi que des principaux types d'apprentissage. Nous avons également abordé les méthodes de détection et de classification d'images, qui jouent un rôle essentiel dans l'analyse et l'interprétation des images. Le dernier chapitre est consacré à la présentation des outils utilisés pour l'implémentation des modèles proposés, et les résultats obtenus ainsi que leur discussion.

Chapitre 4

Résultats et Expérimentations

Chapitre 4

Résultats et expérimentations

4.1 Introduction

Le diagnostic des maladies thyroïdiennes repose principalement sur l'analyse des images échographiques de la glande thyroïde, notamment pour la détection et la caractérisation des nodules thyroïdiens, essentiels dans l'identification de pathologies graves comme le cancer de la thyroïde. Cependant, la distinction entre les nodules bénins et malins reste une tâche complexe nécessitant une expertise médicale avancée. Ainsi, dans le cadre de notre travail, nous proposons un système automatique basé sur des approches hybrides de deep learning, combinant plusieurs architectures complémentaires afin de pallier les limites de chaque modèle et d'améliorer la robustesse ainsi que les performances globales du système.

4.2 Outils utilisés

Pour la réalisation de notre système, nous avons utilisé plusieurs plateformes différentes :

- Google Colab intégré à Google Drive, qui permet de stocker, partager et accéder facilement aux notebooks.
- Lightning ai qui donne un accès immédiat aux GPU gratuitement, sans avoir besoin de passer par des demandes d'autorisation comme sur Colab Pro et une mémoire plus grande par rapport à la RAM de colab.
- Roboflow, une plateforme qui facilite l'annotation, la gestion et l'exportation des jeux de données destinés à l'entraînement des modèles de vision par ordinateur.

4.2.1 Langage utilisé (Python)

Python est un langage de programmation informatique largement utilisé dans les applications Web, le développement de logiciels, la science des données et le machine learning. De plus, python est un langage open source, multiplateforme et orienté objet, qui fonctionne sur différents systèmes d'exploitation tels que Windows, macOS et Linux. Les développeurs l'utilisent parce qu'il est efficace, facile à apprendre et peut s'exécuter sur diverses plateformes. Il dispose d'une immense bibliothèque de modules et d'une communauté très active, ce qui facilite l'apprentissage et l'exécution de projets complexes [106].

4.2.2 Google Colab (Colaboratory)

Google Colab est un environnement de développement Python hébergé dans le cloud qui permet d'écrire et d'exécuter du code directement depuis un navigateur web. Elle offre un accès gratuit aux ressources puissantes mais limitées en matière de temps, notamment des GPU et des TPU sans nécessiter d'installation locale. Basé sur Jupyter Notebook, il combine texte, code exécutable et visualisations dans un document interactif. L'intégration avec Google Drive facilite le stockage, le partage et la collaboration en temps réel. Il est particulièrement adapté aux projets de data science et d'intelligence artificielle [107].

4.2.3 Lightning AI

Lightning AI est une plateforme open source qui réunit la construction de modèles et la création d'applications IA au même endroit. Elle fournit des GPU cloud, des DevBoxes prêts à l'emploi, l'entraînement, le déploiement et l'orchestration avec zéro configuration. Elle permet de coder ensemble depuis le navigateur, de prototyper rapidement, d'entraîner sur des clusters GPU et de servir des modèles en production [108].

- Tous les modèles ont été entraînés à l'aide d'un GPU NVIDIA L4, ce qui a permis d'accélérer les calculs et de réduire considérablement le temps d'entraînement grâce à ses capacités de traitement parallèle élevées.
- Dans le cadre de notre projet, nous avons utilisé deux machines différentes. Le premier système était un LAPTOP-CIETJ12C, équipé d'un processeur Intel Pentium i3-4415U et de 4 Go de RAM. Le second appareil était un DESKTOP-FF8J8VQ, doté d'un processeur Intel Core i5-8350U et de 16 Go de RAM.

4.2.4 Ultralytics

Ultralytics est une plateforme de vision par ordinateur qui permet d'annoter des jeux de données, d'entraîner des modèles YOLO sur des GPU cloud et de les déployer à l'échelle mondiale depuis un espace de travail unique. Elle simplifie l'ensemble du processus, des données brutes aux applications d'IA prêtes pour la production, en remplaçant les outils fragmentés. Elle est également une entreprise spécialisée dans le développement de solutions d'intelligence artificielle, connue pour la création et la maintenance de la série de modèles YOLO [109].

4.2.5 Roboflow

Roboflow est une plateforme de vision par ordinateur développée pour aider à créer, entraîner et déployer facilement des modèles d'intelligence artificielle à partir d'images et de vidéos. Connue pour son innovation, elle transforme la manière dont les entreprises utilisent les technologies visuelles. Elle est largement utilisée pour simplifier la création de modèles de détection d'objets (comme YOLO, Faster R-CNN, etc.) sans avoir à construire toute l'infrastructure soi-même [110].

4.2.6 Les bibliothèques

- **TensorFlow**

Est une bibliothèque développée par l'équipe Google Brain en 2011 sous le nom DistBelief, pour les réseaux de neurones de Deep Learning. En 2015, elle est devenue open source sous le nom TensorFlow. Elle permet de résoudre des problèmes mathématiques complexes et aide les chercheurs de développer des architectures d'apprentissage expérimentales et à les transformer en logiciels [111].

- **PyTorch**

Est un framework open source d'apprentissage automatique qui accélère le passage du prototypage de recherche au déploiement en production, offrant une flexibilité et rapidité grâce aux graphes de calcul dynamiques. Il prend en charge un large éventail de cas d'utilisation, du traitement automatique du langage naturel, de la vision par ordinateur à l'apprentissage par renforcement et à l'IA générative, grâce à un écosystème robuste de bibliothèques, d'outils et d'intégrations [112].

- **OpenCV (cv2)**

Est une bibliothèque open source de vision par ordinateur et d'apprentissage automatique, conçue pour accélérer l'intégration de la perception ma-

chine dans les produits commerciaux. Sous licence Apache 2, elle est facile à utiliser et à modifier pour les entreprises et comprend plus de 2 500 algorithmes optimisés. [113].

- **Matplotlib**

Est une bibliothèque Python open source créée en 2002 pour visualiser les signaux électriques cérébraux de patients épileptiques, en reproduisant les capacités graphiques de MATLAB au sein de Python. [114]. Elle offre une flexibilité élevée permettant de produire une large gamme de visualisations, avec une prise en main accessible aux débutants et des fonctionnalités avancées pour les utilisateurs expérimentés. [115].

- **Numerical Python (Numpy)**

Est une bibliothèque Python open source largement utilisée en sciences et en ingénierie. Elle contient des structures de données multidimensionnelles, telles que les tableaux homogènes et les tableaux N-dimensionnels, ainsi qu'une vaste bibliothèque de fonctions permettant de manipuler efficacement ces structures [116].

- **Scikit-learn**

Est une bibliothèque open source d'apprentissage automatique en Python, créée en 2007 par David Cournapeau dans le cadre du Google Summer of Code. Aussi connue sous son nom abrégé sklearn, elle fournit une vaste gamme d'outils et d'algorithmes permettant de faciliter le développement de modèles de machine learning, notamment pour la classification, la régression, le clustering et le gradient boosting [117].

- **Seaborn**

Est une bibliothèque Python de visualisation de données basée sur matplotlib. Elle propose une interface de haut niveau pour la création facilement des graphiques statistiques attrayants et informatifs. [118].

4.3 Base de données utilisée

4.3.1 Analyse comparative et justification du choix de la base de données

La base de données ThyroidXL a été choisie en raison de ses avantages par rapport aux autres datasets existants. Contrairement à des jeux de données tels que Thyroid Ultrasound Images (900 images) ou TN3K (environ 3493 images), il se distingue par un volume beaucoup plus important avec plus de 11000 images, ce qui le rend mieux adapté à l'entraînement de modèles de deep learning et

permet d'éviter le surapprentissage. Il est également polyvalent car il permet de réaliser différentes tâches telles que la classification, la détection et la segmentation des nodules thyroïdiens, offrant ainsi une grande flexibilité dans le développement du projet. Par ailleurs, les images étant issues d'échographies médicales, elles sont parfaitement adaptées au diagnostic assisté par ordinateur. Enfin, l'accès au dataset via Hugging Face est soumis à une demande préalable en raison de la sensibilité des données médicales. Après acceptation des conditions d'utilisation et validation de la demande, le dataset a été mis à disposition dans un cadre strictement académique et expérimental après validation.

Dataset	Nombre d'images	Type de tâche
Thyroid_Ultrasound_Images	900	Classification
Thyroid_phantom_dataset	1278	Classification, Détection, Segmentation
TN3K Dataset	3493	Segmentation
TN5000-2025	5000	Détection, Classification
ThyroidXL	11545	Classification, Détection, Segmentation

Table 4.1 – Comparaison des principaux datasets d'images échographiques thyroïdiennes

4.3.2 Base de données retenu et analyse descriptive des données

4.3.2.1 Présentation du base de données ThyroidXL

L'ensemble de données que nous allons utiliser dans ce travail est le dataset ThyroidXL [119], disponible en ligne sur la plateforme HuggingFace [120]. Ce jeu de données (dataset) est constitué d'images d'échographie de la glande thyroïde (voir Figure 4.1), utilisées pour l'analyse et la détection des nodules thyroïdiens. Ce dataset a été conçu pour soutenir la recherche et le développement de modèles d'intelligence artificielle dans le domaine de l'imagerie médicale, notamment pour la mise en place de systèmes d'aide au diagnostic permettant la détection et la classification des nodules thyroïdiens à partir d'images échographiques. Il est composé de 11 545 images d'échographie de la thyroïde provenant de 4 093 patients, collectées et annotées par des radiologues experts à l'Hôpital national d'endocrinologie du Vietnam. L'ensemble de données est organisé en deux sous-ensembles principaux : un ensemble d'entraînement (train) contenant 9 541 images issues de 3 275 patients et un ensemble de test (test) contenant 2 094 images provenant de 739 patients. Les images sont associées à des annotations permettant d'effectuer plusieurs tâches d'apprentissage profond telles que la classification, la détection et la segmentation des nodules thyroïdiens.

Les données comprennent des cas bénins et malins, permettant ainsi de développer et d'évaluer des modèles de diagnostic assisté par ordinateur pour la détection automatique des nodules thyroïdiens dans les images échographiques. Grâce à sa taille importante et à ses annotations validées par des experts médicaux, le dataset ThyroidXL constitue l'un des plus grandes ressources publiques disponibles pour l'étude et l'analyse automatique des nodules thyroïdiens à l'aide des techniques d'apprentissage profond, ce qui en fait une ressource précieuse pour la recherche en imagerie médicale et en apprentissage automatique.

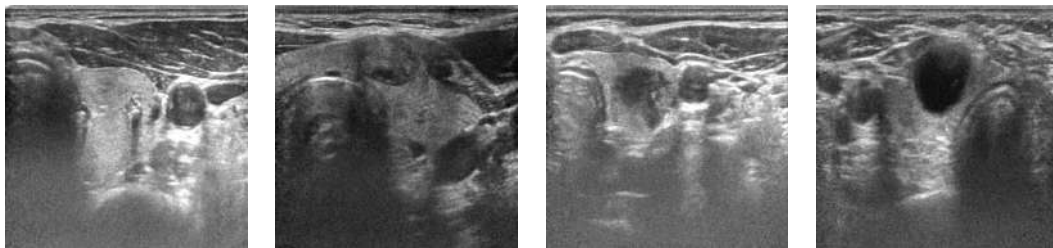


Figure 4.1 – Visualisation des images du dataset ThyroidXL

4.3.2.2 Analyse descriptive des données

Afin de mieux caractériser cette base de données, une analyse descriptive a été réalisée en étudiant la répartition démographique et clinique des patients, notamment selon le sexe, l'âge, la localité ainsi que les catégories TI-RADS.

• Distribution selon le genre

L'analyse démographique met en évidence une très forte prédominance féminine au sein de jeux de donnée. Comme l'illustre la figure sur la distribution des genres, pour l'ensemble d'entraînement 3000 comprend environ femmes de sexe féminin contre 350 pour le sexe masculin seulement. De même, l'ensemble de tests avec environ 650 femmes contre une centaine d'hommes.

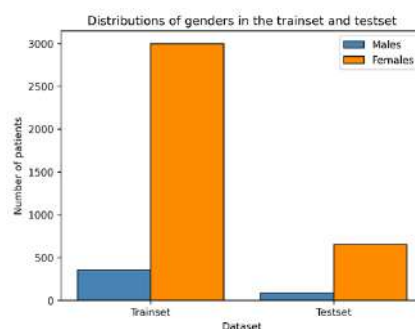


Figure 4.2 – Distribution des genres dans le jeu de données [119]

• Distribution selon l'âge

L'analyse de la distribution des âges montre une répartition globalement similaire entre les ensembles d'entraînement et de test. Nos résultats montrent une corrélation positive entre l'âge et la fréquence des pathologies thyroïdiennes dans notre population. La fréquence selon les tranches d'âge continue son ascension entre la tranche d'âge 30 et 65 ans avec une concentration plus marquée autour des âges moyens, notamment entre 40 et 60 ans. Elle décline après 75 ans et plus. L'âge reste un facteur de risque quant à la survenue des affections thyroïdiennes.

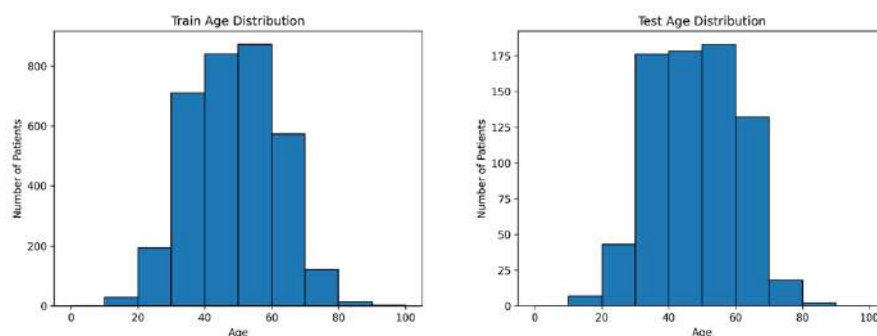


Figure 4.3 – Distribution des âges dans le jeu de données [119]

• Distribution selon les catégories TI-RADS

Comme l'illustre la figure 4.4, dans le train set, on observe une très forte prédominance des catégories élevées, qui constituent la majorité des cas, notamment les classes TI-RADS 3 et TI-RADS 4, avec également une bonne représentation de la catégorie TI-RADS 5. De même, dans l'ensemble de test, on observa une dominance de la catégorie TI-RADS 5, suivie des catégories TI-RADS 4 et TI-RADS 3. Cependant, dans les deux ensembles (train et test), on constate que les catégories TI-RADS 1 et TI-RADS 2 sont très faiblement représentées, voire quasiment absentes.

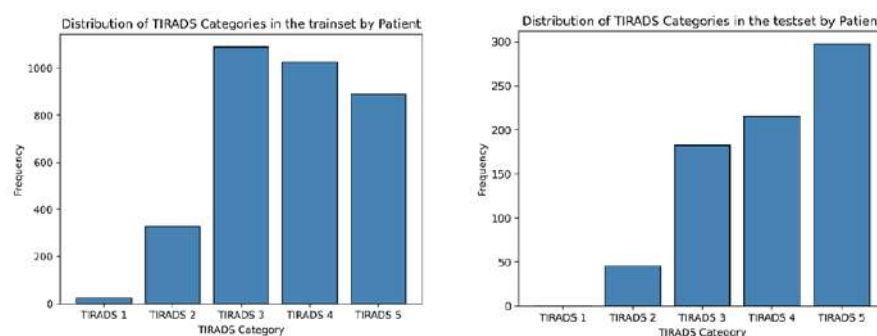


Figure 4.4 – Distribution selon les catégories TI-RADS le jeu de données [119]

4.4 Méthodes proposées

4.4.1 Modèle 1 (RF-DETR)

RF-DETR adopte une architecture séquentielle où le Backbone extrait les caractéristiques visuelles, le Projecteur et l'Encoder les enrichissent, le Décodeur structuré en Decoder Groups raffine les détections, et le Detection Head produit les prédictions finales (voir Figure 4.5).

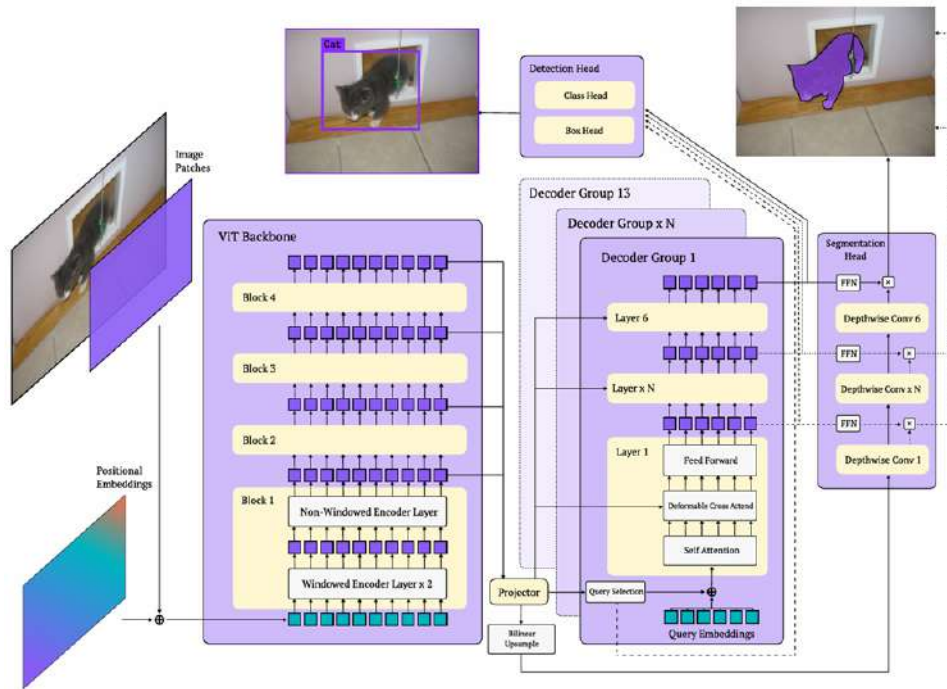


Figure 4.5 – RF-DETR Architecture [121]

Section	Couche	Méthode utilisée	Rôle principal
Input	Patch Embedding	ViT Patch Embedding, Positional Encoding	Découper l'image en tokens et injecter l'information de position spatiale pour préparer l'image.
Backbone	Blocs 1 → 4 (sorties : 3, 6, 9, 12)	Windowed Self-Attention ×2, Non-Windowed Self-Attention, SSL Pretraining (DINOv2), Multi-layer Feature Taps	Extraire des caractéristiques visuelles riches et robustes grâce à l'attention locale et globale.
Projecteur	Bloc 5	Linear Projection (Conv 1 × 1), Bilinear Upsampling, Sinusoidal Positional Encoding	Faire le pont entre le backbone et le décodeur DETR.
Encoder	Bloc 6	Deformable Self-Attention, Two-Stage Initialization, Query Selection	Raffiner les caractéristiques et générer les premières régions candidates.
Décodeur	Blocs 7 → 10	Self-Attention (8 têtes), Deformable Cross-Attention (16 têtes, 2 points), FFN, Lite Refpoint Refine, BBox Reparametrization, LayerNorm	Affiner les points de référence et stabiliser la régression des boîtes.
Decoder Groups	13 groupes	Contrastive Denoising, Shared Decoder Weights, Hungarian Matching	Accélérer la convergence par débruitage contrastif avec 13 groupes parallèles.
Detection Head	Bloc 11	MLP Class Head (90 classes), MLP Box Head (4 coordonnées), IA BCE Loss, Deep Supervision	Produire les prédictions finales des classes et des boîtes englobantes.

Table 4.2 – Explication de l'architecture du modèle RF-DETR

4.4.2 Modèle 2 (yolov8)

L'architecture de YOLOv8, issue de l'évolution des versions précédentes de l'algorithme YOLO basées sur les réseaux de neurones convolutifs (CNN), repose sur une structure modulaire composée de trois blocs principaux : le **Backbone**, le **Neck** et le **Head** (voir Figure 4.6).

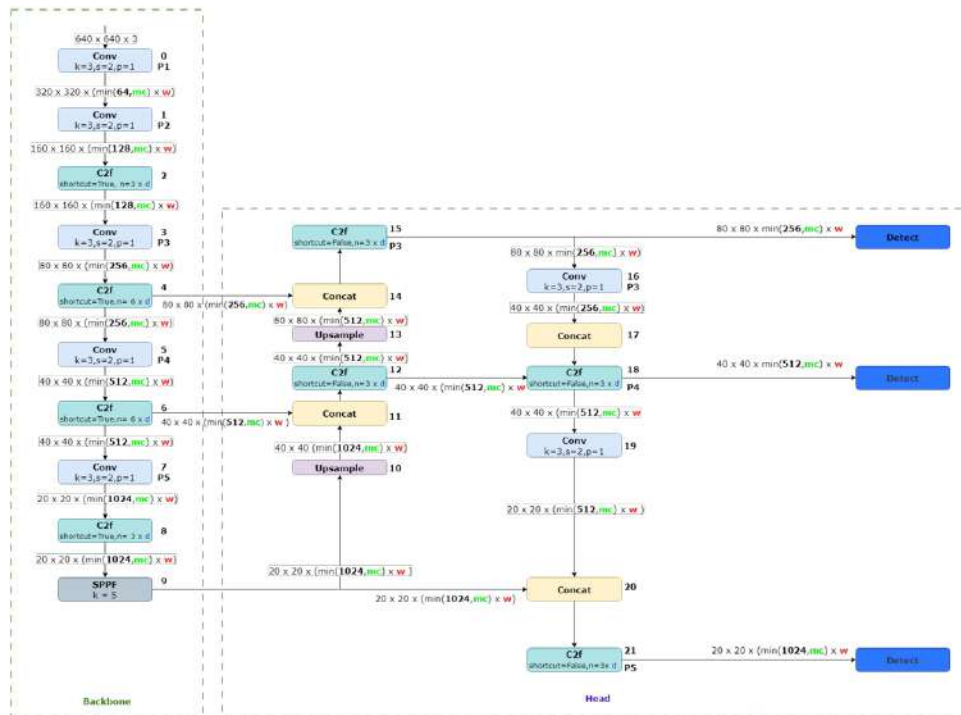


Figure 4.6 – YOLOv8 Architecture [122]

Section	Couches	Modules utilisés	Résolution	Rôle principal
Backbone	0 – 9	Conv, C2f, SPPF	640 × 640 → 20 × 20	Extraire les features à plusieurs échelles
Neck FPN	10 – 15	Upsample, Concat, C2f	20 × 20 → 80 × 80	Fusionner features profondes et superficielles (top-down)
Head PAN	16 – 21	Conv, Concat, C2f	80 × 80 → 20 × 20	Recombinaison des niveaux (bottom-up) + préparer la détection
Detect P3	15	Detect head	80 × 80 × 256	Détecter les petits objets
Detect P4	18	Detect head	40 × 40 × 512	Détecter les objets moyens
Detect P5	21	Detect head	20 × 20 × 1024	Détecter les grands objets

Table 4.3 – Explication de l'architecture YOLOv8

4.4.3 Modèle 3 (YOLOE)

L'architecture de YOLOE s'organise en trois blocs successifs : un Backbone assurant l'extraction hiérarchique des caractéristiques, un Neck réalisant leur agrégation et leur enrichissement à plusieurs échelles, et un Head intégrant le module YOLOEDetect, propre à cette architecture, qui produit les sorties finales du réseau (voir Figure 4.7).

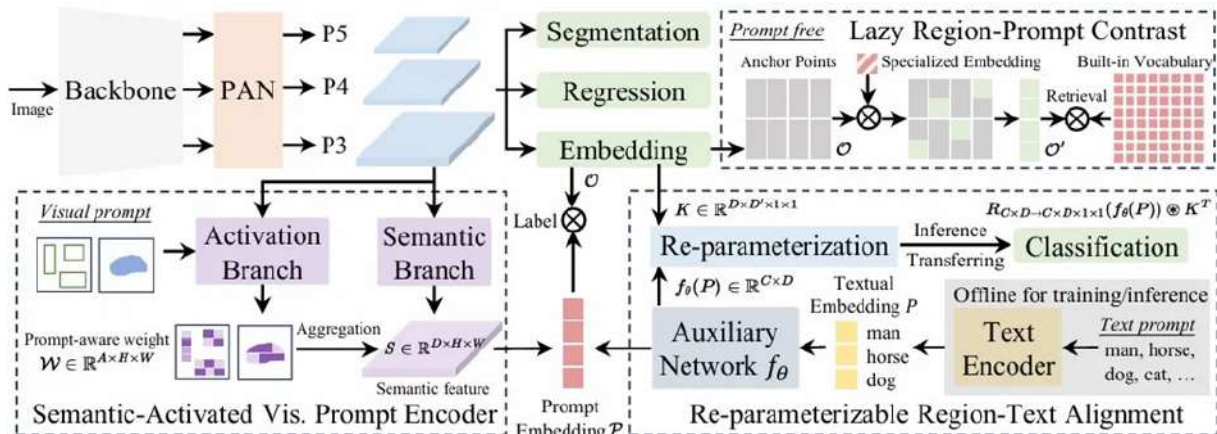


Figure 4.7 – YOLOE Architecture [123]

Section	Couches	Modules utilisés	Résolution	Rôle principal
Backbone	0–9	Conv, C2f, SPPF	$640 \times 640 \rightarrow 20 \times 20$	Extraire les caractéristiques à plusieurs échelles.
Neck FPN	10–15	Upsample, Concat, C2f	$20 \times 20 \rightarrow 80 \times 80$	Fusionner les caractéristiques selon une stratégie top-down.
Head PAN	16–21	Conv stride-2, Concat, C2f	$80 \times 80 \rightarrow 20 \times 20$	Recombinaison des informations multi-échelles selon une stratégie bottom-up.
YOLOEDetect	22	Cv2, Cv3, Cv4, DFL, RepRTA (SwiGLUFFN), SAVPE	P3 : $80 \times 80 \times 256$ P4 : $40 \times 40 \times 512$ P5 : $20 \times 20 \times 512$	Prédire les boîtes englobantes, les embeddings, la classification contrastive texte-région et l'encodage Visual Prompt.

Table 4.4 – Explication de l'architecture YOLOE

4.4.4 Modèle 4 (yolo26)

Le modèle YOLO26 est composé de trois blocs principaux : le **Backbone** pour l'extraction des caractéristiques, le **Neck** pour la fusion multi-échelle des informations, et le **Head** pour la détection et la classification des objets (voir la Figure 4.8).

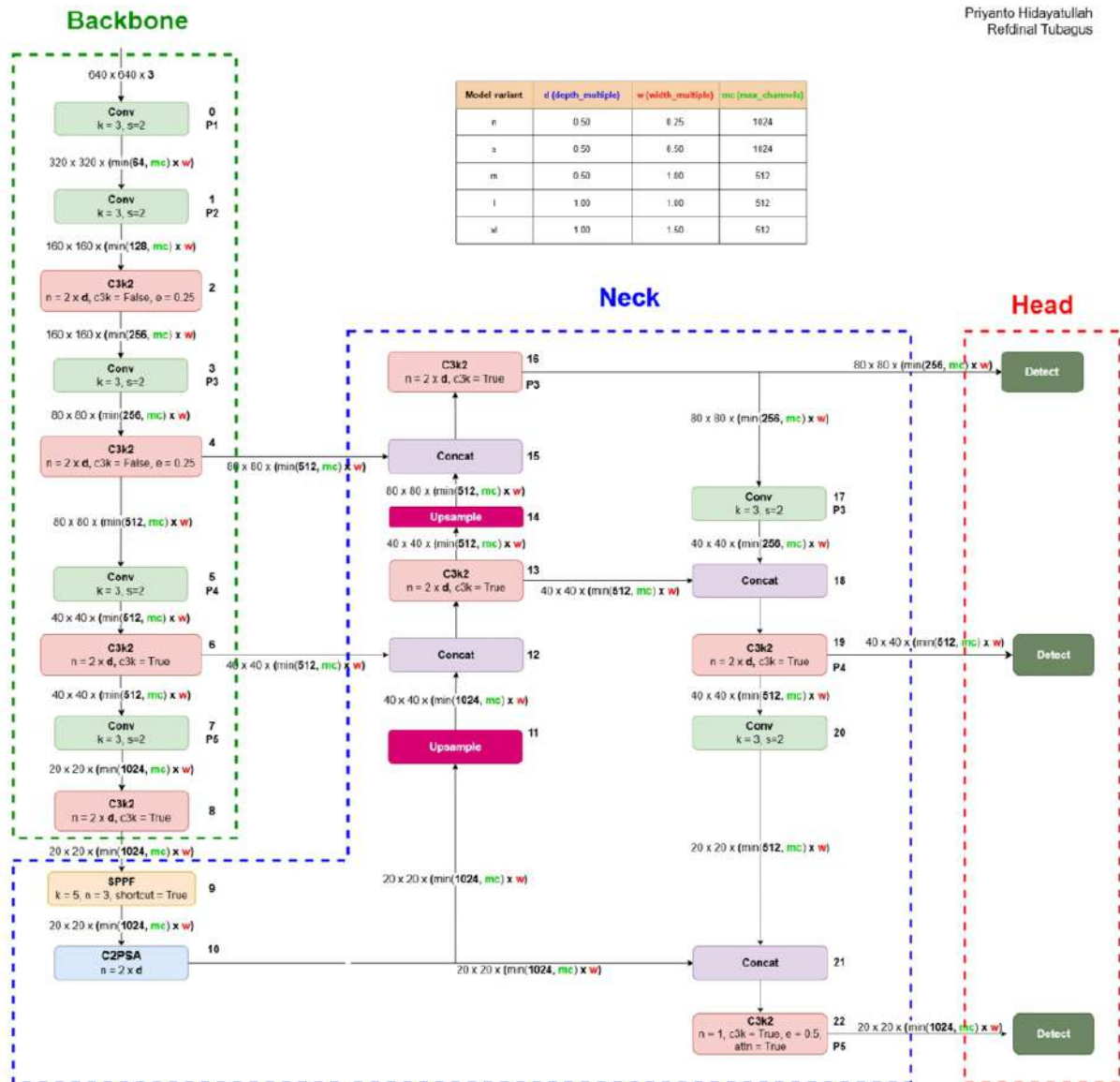


Figure 4.8 – Diagramme d'architecture YOLO26 [84]

Section	Couches	Blocs principaux	Résolution	Rôle
Backbone	0–8	Conv (k=3, s=2), C3k2	80×80, 40×40, 20×20	Extraction hiérarchique des caractéristiques visuelles à différentes échelles (P3, P4, P5).
	9–10	SPPF, C2PSA	20×20	Élargissement du champ réceptif et mise en évidence des régions importantes grâce au mécanisme d'attention.
Neck (FPN+PAN)	11–16	Upsample, Concat, C3k2	20×20 → 80×80	Fusion ascendante des caractéristiques afin de propager les informations sémantiques vers les hautes résolutions.
	17–22	Conv (s=2), Concat, C3k2	80×80 → 20×20	Fusion descendante permettant de réintégrer le contexte global aux basses résolutions.
Head	—	Detect × 3	80×80 / 40×40 / 20×20	Prédiction des boîtes englobantes, des scores de confiance et des classes sur trois échelles.

Table 4.5 – Explication de l'architecture YOLO26

4.5 Pré-traitement

4.5.1 Chargement du Base de données

Dans cette partie, nous avons chargé le base de données ThyroidXL à partir de la plateforme Roboflow, une étape essentielle dans la création d'un pipeline d'apprentissage automatique dédié à l'imagerie médicale. Le projet intitulé thyroid-project qui compose un total de 11621 images échographiques annotée pour la détection binaire. Par ailleurs, à partir de Roboflow Health Check (voir Figure 4.9) pour obtenir des informations sur votre jeu de données notamment la taille et le nombre des images, l'équilibre des classes pour chaque image, telle que ce dernier indique la classe 0 pour les cas bénins et le 1 pour les cas malins.

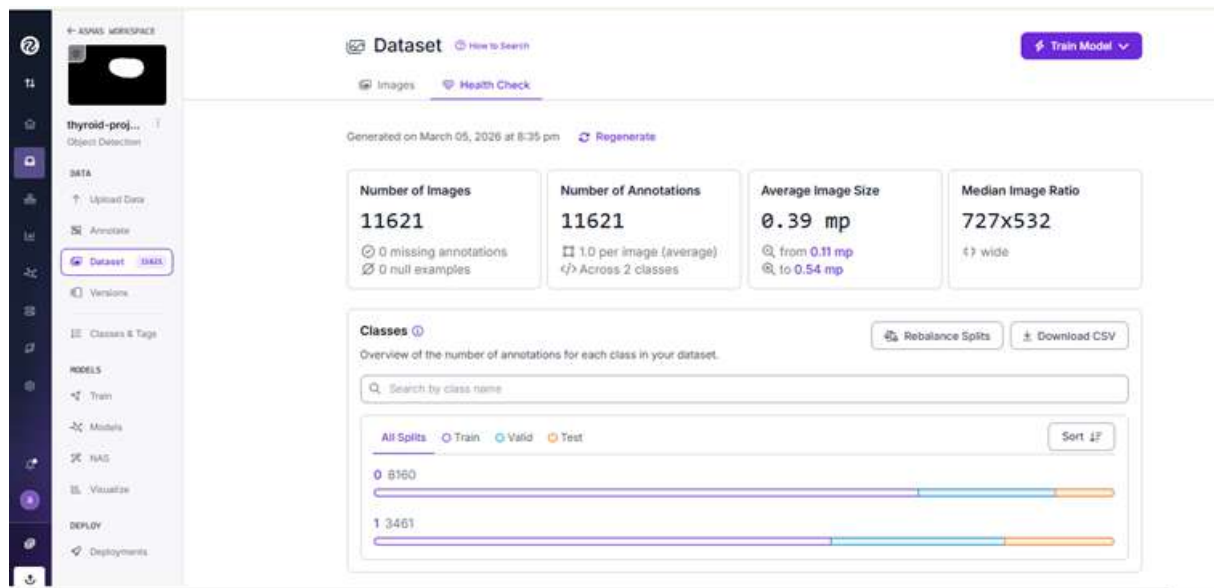


Figure 4.9 – Analyse descriptive du dataset

4.5.2 Annotation des données

Dans notre projet, une fois téléchargé les images la plate-forme Roboflow propose une annotation manuelle des images (voir Figure 4.10) dans le but d'étiqueter les données pour la tâche de détection, à l'aide des boîtes englobante dans les régions d'intérêt correspondant aux nodules thyroïdiens sur les images échographiques. Les annotations obtenues sont ensuite utilisées pour entraîner les modèles de détection ce qui permet de garantir une meilleure précision et d'améliorer leur capacité à identifier automatiquement les zones pathologiques.

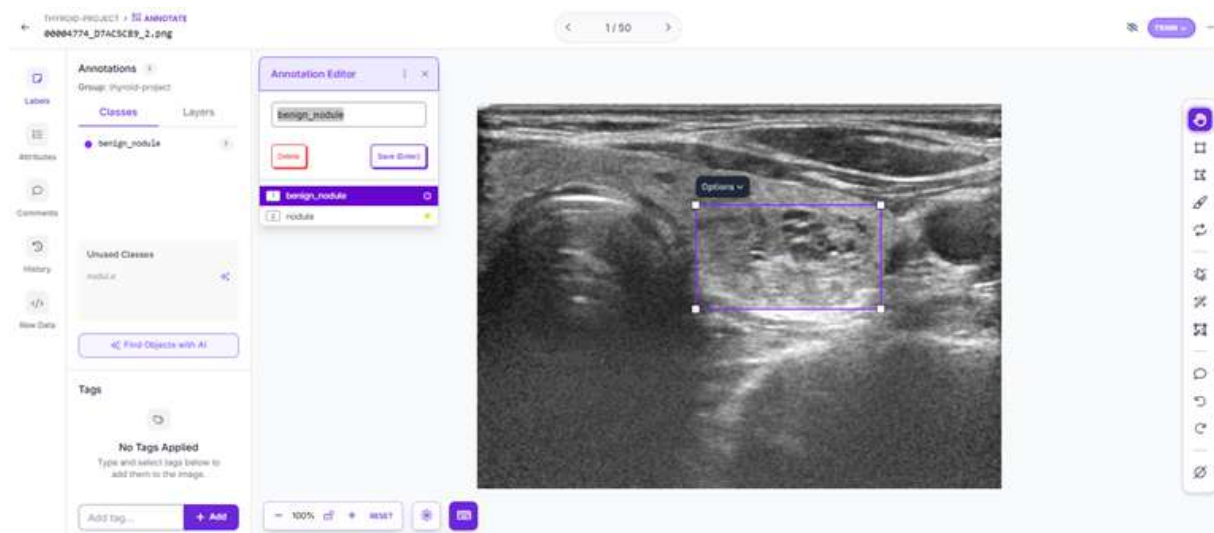


Figure 4.10 – Annotation avec outils Roboflow

4.5.3 Répartition des données

La base de données a été répartie en trois sous-ensembles essentiels (voir Figure 4.11).

- 70% (8135 images) pour l'entraînement
- 20% (2324 images) pour la validation
- 10% (1162 images) pour le test

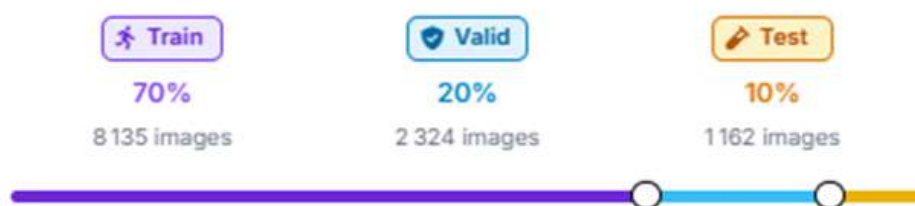


Figure 4.11 – Division des données en données d'apprentissage, de validation et de test

4.5.4 Importation du Dataset

Après avoir préparé et annoté les données, nous avons exporté le Dataset dans différents formats adaptés à l'entraînement des modèles, notamment le format YOLO pour YOLOE, YOLOv8 et YOLO26, ainsi que le format COCO pour RF-DETR. Dans notre projet ces deux formats ont été explorés pour l'entraînement. Le chargement des données a été effectué depuis Roboflow à l'aide de script Python permettant d'accéder aux images ainsi qu'à leurs annotations.

La figure présente deux captures d'écran de terminaux Linux montrant l'exécution de scripts Python pour télécharger des données depuis Roboflow.

Le terminal de gauche exécute un script qui configure Roboflow, définit un projet et télécharge le dataset 'yolo8'.

```
from roboflow import Roboflow
rf = Roboflow(api_key="9U5Dm4wXMO42nc")
project = rf.workspace("asmas-workspace-bekfy").project("thyroid-project")
dataset = project.version(1).download("yolo8")
print("Download finished")
```

Le terminal de droite exécute un script similaire mais télécharge le dataset 'coco'.

```
from roboflow import Roboflow
rf = Roboflow(api_key="Dc7SDM4d4qK9NC6Z9GrP")
project = rf.workspace("asmas-workspace-bekfy").project("thyroid-project")
dataset = project.version(5).download("coco")
```

Figure 4.12 – Téléchargement de base de données sur Lightning

Ce processus génère automatiquement un fichier de configuration au format YAML, nommé data.yaml (voir Figure 4.13). Ce fichier contient les chemins d'accès aux dossiers des images et des labels, organisés selon les ensembles d'entraînement, de validation et de test, ainsi que les informations relatives aux classes cibles.

```

thyroid-project-2 > ! data.yaml
1  names:
2    - Benign
3    - Malignant
4  nc: 2
5  roboflow:
6    license: MIT
7    project: thyroid-project
8    url: https://universe.roboflow.com/asmas-workspace-bekfy/thyroid-project/dataset/2
9    version: 2
10 workspace: asmas-workspace-bekfy
11 test: ../test/images
12 train: ../train/images
13 val: ../valid/images

```

Figure 4.13 – data.yaml

Après le chargement du dataset, nous avons constaté un déséquilibre dans la distribution du nombre d'images par classe, comme illustré dans la Figure 4.14, ce qui peut entraîner des problèmes d'apprentissage tels que le sur-apprentissage. Pour résoudre ce problème, nous avons appliqué une technique d'oversampling consistant à dupliquer les images de la classe sous-représentée (Figure 4.15). L'objectif est d'équilibrer les classes du dataset sans introduire de nouvelles images artificielles par transformation.

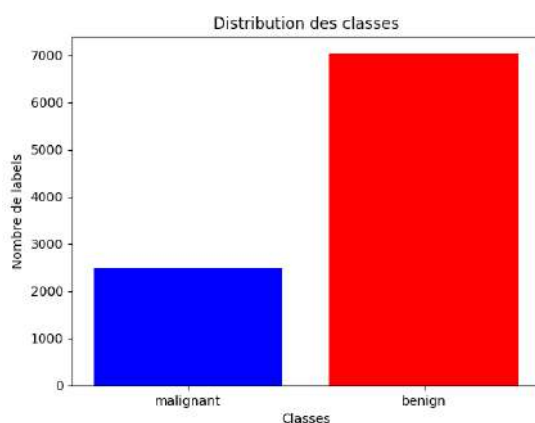


Figure 4.14 – Visualisation de la distribution des images des classes d'origine

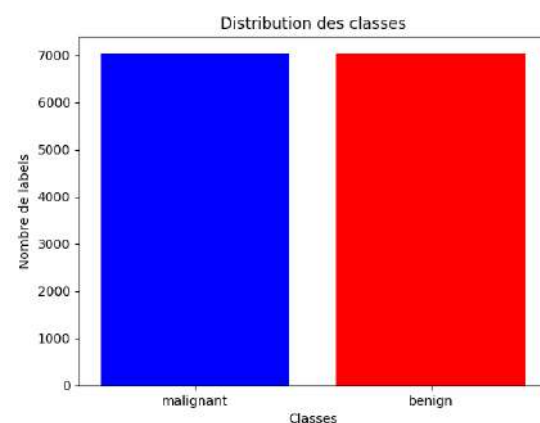


Figure 4.15 – Visualisation de la distribution des images après l'augmentation

4.6 Apprentissage et hyperparamètres des modèles

Pour entraîner nos différents modèles (4 architectures), le choix des valeurs des hyperparamètres est une étape importante afin de garantir des résultats optimaux. Dans notre travail, ces paramètres ont été choisis expérimentalement de la manière suivante :

- **Optimiseur** : nous avons utilisé l'optimiseur AdamW, une variante de Adam, intégrant une régularisation par décroissance des poids pour améliorer la généralisation, avec un taux d'apprentissage initial fixé à 10^{-4} ou 10^{-5} , car ils permettent une convergence plus rapide et efficace.

- **Epochs / batch size** : nous avons fixé le nombre d'époques à 100 et la taille des lots (batches) de 8 à 16 images par itération.
- **Taille de l'image** : 640×640 pixels, la taille à laquelle chaque image d'entrée est redimensionnée avant d'être traitée au réseau de neurones.
- **weight_decay** : un coefficient de décroissance des poids fixé à 5×10^{-4} , est une technique qui empêche les poids de devenir trop grands pendant l'entraînement, en les réduisant un peu à chaque mise à jour afin d'éviter le surapprentissage et améliorer la généralisation du modèle.

Afin d'améliorer la performance du modèle tout en évitant le surapprentissage, plusieurs techniques de régularisation et d'optimisation ont été mises en œuvre telles que :

1. **EarlyStopping** : cette technique permet d'interrompre automatiquement l'entraînement du modèle si la performance sur les données de validation ne s'améliore plus (après un certain nombre d'époques), tout en rétablissant les meilleurs poids obtenus et de réduire le temps de formation inutile.
2. **Data augmentation (augment = True)** : permet d'appliquer des transformations aléatoires aux images d'entraînement telles que la rotation, le zoom, le décalage...etc, afin d'augmenter artificiellement la diversité du jeu de données et de renforcer la robustesse du modèle face aux variations des images lors de la phase de test.

Nous présentons le processus d'entraînement des modèles afin de mettre en évidence son comportement d'apprentissage au fil du temps. Durant l'entraînement, nous avons utilisé mAP comme une métrique d'évaluation pour apprécier les performances, les Figures 4.16, 4.17, 4.18 et 4.19 montrent quelques performances à certains epochs de l'apprentissage.

```
[2026-05-13 09:46:38] [INFO] rf-detr - Best EMA mAP improved to 0.0081 (epoch 0)
Metric __rfdetr_effective_map__ improved, New best score: 0.448
[2026-05-13 10:37:00] [INFO] rf-detr - Best regular mAP saved to /teamspace/studios/this_studio/output/checkpoint_best_regular.pth (e
[2026-05-13 10:37:01] [INFO] rf-detr - Best EMA mAP improved to 0.4481 (epoch 0)
Metric __rfdetr_effective_map__ improved by 0.036 >= min_delta = 0.001, New best score: 0.484
[2026-05-13 11:27:51] [INFO] rf-detr - Best EMA mAP improved to 0.4844 (epoch 1)
Metric __rfdetr_effective_map__ improved by 0.014 >= min_delta = 0.001, New best score: 0.498
[2026-05-13 12:18:17] [INFO] rf-detr - Best regular mAP saved to /teamspace/studios/this_studio/output/checkpoint_best_regular.pth (e
[2026-05-13 12:18:17] [INFO] rf-detr - Best EMA mAP improved to 0.4983 (epoch 2)
Metric __rfdetr_effective_map__ improved by 0.008 >= min_delta = 0.001, New best score: 0.506
[2026-05-13 13:08:47] [INFO] rf-detr - Best regular mAP saved to /teamspace/studios/this_studio/output/checkpoint_best_regular.pth (e
[2026-05-13 13:08:47] [INFO] rf-detr - Best EMA mAP improved to 0.5062 (epoch 3)
Metric __rfdetr_effective_map__ improved by 0.011 >= min_delta = 0.001, New best score: 0.517
[2026-05-13 13:59:46] [INFO] rf-detr - Best regular mAP saved to /teamspace/studios/this_studio/output/checkpoint_best_regular.pth (e
[2026-05-13 13:59:46] [INFO] rf-detr - Best EMA mAP improved to 0.5174 (epoch 4)
Metric __rfdetr_effective_map__ improved by 0.004 >= min_delta = 0.001, New best score: 0.521
[2026-05-13 14:51:02] [INFO] rf-detr - Best regular mAP saved to /teamspace/studios/this_studio/output/checkpoint_best_regular.pth (e
[2026-05-13 14:51:02] [INFO] rf-detr - Best EMA mAP improved to 0.5211 (epoch 5)
Metric __rfdetr_effective_map__ improved by 0.005 >= min_delta = 0.001, New best score: 0.526
[2026-05-13 15:42:35] [INFO] rf-detr - Best regular mAP saved to /teamspace/studios/this_studio/output/checkpoint_best_regular.pth (e
[2026-05-13 15:42:35] [INFO] rf-detr - Best EMA mAP improved to 0.5261 (epoch 6)
[2026-05-13 16:33:49] [INFO] rf-detr - Best regular mAP saved to /teamspace/studios/this_studio/output/checkpoint_best_regular.pth (e
```

Figure 4.16 – Apprentissage du Modèle 1

Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
1/100	2.16G	1.469	1.696	1.506	13	640: 100% ██████████ 750/750 [02:42<00:00, 4.611t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 73/73 [00:14<00:00, 5.171t/s]
	all	2324	2324	0.401	0.437	0.373 0.177
2/100	2.22G	1.42	1.354	1.477	21	640: 100% ██████████ 750/750 [02:13<00:00, 5.631t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 73/73 [00:13<00:00, 5.441t/s]
	all	2324	2324	0.551	0.595	0.599 0.337
3/100	2.14G	1.396	1.283	1.457	14	640: 100% ██████████ 750/750 [02:00<00:00, 6.231t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 73/73 [00:12<00:00, 5.761t/s]
	all	2324	2324	0.658	0.726	0.732 0.422
4/100	2.14G	1.366	1.206	1.431	16	640: 100% ██████████ 750/750 [01:54<00:00, 6.531t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 73/73 [00:12<00:00, 5.861t/s]
	all	2324	2324	0.655	0.757	0.749 0.454
5/100	2.22G	1.326	1.149	1.405	24	640: 100% ██████████ 750/750 [01:54<00:00, 6.551t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 73/73 [00:12<00:00, 5.901t/s]
	all	2324	2324	0.693	0.771	0.769 0.473
6/100	2.14G	1.306	1.108	1.391	24	640: 100% ██████████ 750/750 [01:54<00:00, 6.521t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 73/73 [00:12<00:00, 5.881t/s]
	all	2324	2324	0.677	0.776	0.77 0.488

Figure 4.17 – Apprentissage du Modèle 2

Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
1/100	8.80G	1.592	1.745	1.843	3	640: 100% ██████████ 1017/1017 [00:29<00:00, 2.011t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 291/291 [00:46<00:00, 6.281t/s]
	all	2324	2324	0.147	0.285	0.162 0.0661
2/100	6.51G	1.521	1.684	1.776	5	640: 100% ██████████ 1017/1017 [00:18<00:00, 2.601t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 291/291 [00:46<00:00, 6.301t/s]
	all	2324	2324	0.57	0.682	0.659 0.308
3/100	6.66G	1.499	1.527	1.764	3	640: 100% ██████████ 1017/1017 [00:15<00:00, 2.711t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 291/291 [00:46<00:00, 6.311t/s]
	all	2324	2324	0.556	0.702	0.675 0.357
4/100	6.66G	1.462	1.406	1.731	5	640: 100% ██████████ 1017/1017 [00:14<00:00, 2.711t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 291/291 [00:46<00:00, 6.321t/s]
	all	2324	2324	0.585	0.7	0.687 0.322
5/100	6.66G	1.423	1.377	1.699	3	640: 100% ██████████ 1017/1017 [00:14<00:00, 2.721t/s]
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 291/291 [00:46<00:00, 6.311t/s]

Figure 4.18 – Apprentissage du Modèle 3

Epoch	GPU_mem	box_loss	cls_loss	dfl_loss	Instances	Size
1/100	2.45G	1.524	2.187	0.01802	6	640: 100% ██████████ 881/881 4.41t/s 3:19<0.2s
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 66/66 3.61t/s 18.4s0.2s
	all	2092	2092	0.318	0.262	0.174 0.0684
2/100	2.84G	1.475	1.497	0.01748	8	640: 100% ██████████ 881/881 4.91t/s 2:40<0.2s
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 66/66 6.11t/s 10.8s0.2s
	all	2092	2092	0.641	0.716	0.725 0.423
3/100	2.85G	1.43	1.32	0.01679	6	640: 100% ██████████ 881/881 5.01t/s 2:55<0.2s
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 66/66 6.01t/s 11.0s0.2s
	all	2092	2092	0.628	0.728	0.734 0.423
4/100	2.85G	1.403	1.22	0.0166	3	640: 100% ██████████ 881/881 5.01t/s 2:54<0.2s
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 66/66 6.21t/s 10.6s0.2s
	all	2092	2092	0.668	0.715	0.738 0.424
5/100	2.85G	1.376	1.166	0.01621	3	640: 100% ██████████ 881/881 5.11t/s 2:54<0.2s
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 66/66 6.21t/s 10.6s0.2s
	all	2092	2092	0.712	0.705	0.600 0.490
6/100	2.85G	1.367	1.119	0.01601	5	640: 100% ██████████ 881/881 5.01t/s 2:55<0.2s
	Class	Images	Instances	Box(P	R	mAP50 mAP50-95): 100% ██████████ 66/66 6.21t/s 10.6s0.2s
	all	2092	2092	0.715	0.763	0.605 0.471

Figure 4.19 – Apprentissage du Modèle 4

4.7 Résultats et Discussion

Une attention particulière a été portée à la classe nodule, qui correspond exclusivement aux cas malins.

L'évaluation des performances des modèles a été effectuée sur un jeu de données unique présenté dans la section 4.3. Les expérimentations ont été réalisées sur des sous-ensembles séparés pour l'apprentissage, la validation et le test. Le nombre de paramètres pour chaque modèle évalué est de 33.6 M pour le Modèle 1 (RF-DETR), 51 165 430 pour le Modèle 3 (YOLOE), 3 006 038 pour le Modèle 2 (YOLOv8) et 2 504 580 pour le Modèle 4 (YOLO26).

Lors de la phase d'entraînement du modèle RF-DETR appliqué à la détection de nodules nous avons configuré le paramètre EarlyStopping avec une patience de 15 itérations. Ainsi, l'apprentissage a été arrêté à la 29ème époque sur 100, avec un taux de $\text{map}@0,50$ de 83% pour l'ensemble de validation et de 84% pour l'EMA $\text{map}@0,50$, ce qui correspond à une valeur de $\text{mAP}@0,50 : 0,95$ de 51%, comme illustré dans la Figure 4.20. L'apprentissage a duré 7 heures, 24 minutes et 38 secondes.

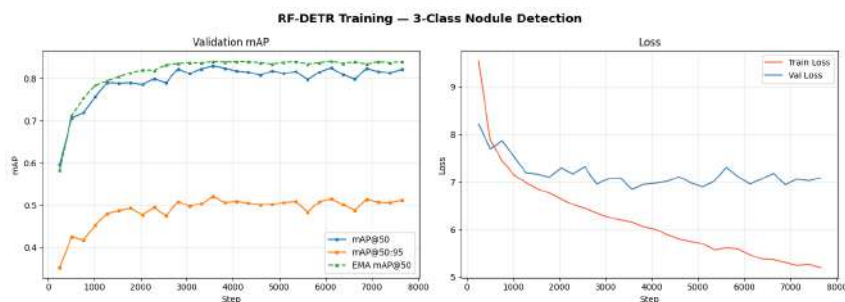


Figure 4.20 – Performances de l'apprentissage du Modèle 1

Durant la phase d'apprentissage, nous avons configuré le paramètre EarlyStopping à 15 pour le Modèle YOLO8n. Ainsi, l'apprentissage s'est arrêté au bout de la 68ème époque sur 100, avec un $\text{map}@0,50$ de 89.5% et un $\text{mAP}@0,50 : 0,95$ de 61.1%. Les pertes finales d'entraînement pour box_loss , cls_loss et dfl_loss étaient approximativement de 0.92, 0.59 et 1.13, tandis que les pertes de validation étaient d'environ 1.18, 0.65 et 1.33. L'apprentissage a duré 2 heures 26 minutes.

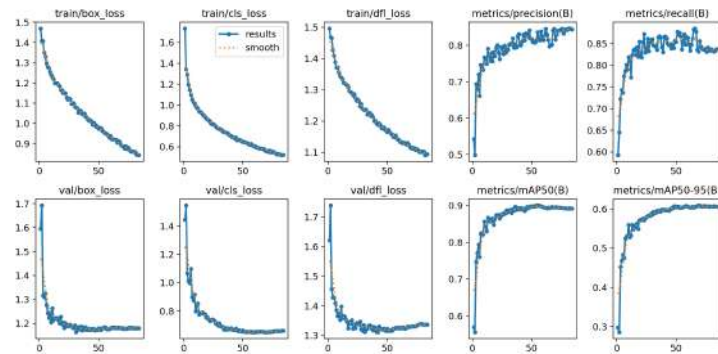


Figure 4.21 – Performances de l'apprentissage du Modèle 2

Concernant le modèle YOLOE, l'EarlyStopping est défini avec une patience de 10 itérations. L'apprentissage est interrompu à la 48 ème époque sur un maximum de 100. Les résultats obtenus sur l'ensemble de validation indiquent un mAP50 global de 90.1% et un mAP@0,50 :0,95 de 61.0% (voir Figure 4.22). L'entraînement a duré approximativement 6 heures et 50 minutes.

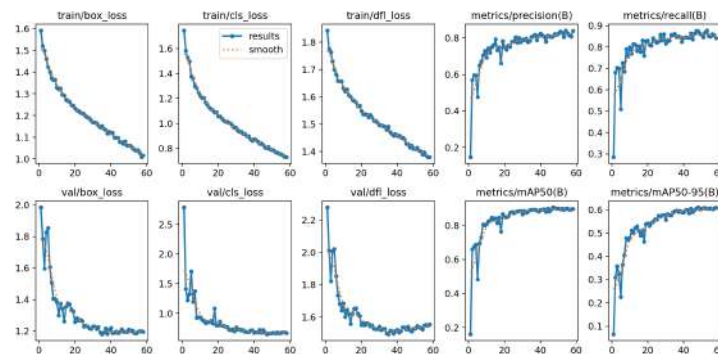


Figure 4.22 – Performances de l'apprentissage du Modèle 3

Enfin, pour la phase d'apprentissage du modèle YOLO26, nous avons également utilisé une valeur de 15 pour l'EarlyStopping. L'entraînement s'est déroulé sur 68 époques sur 100, avec un map@0,50 de 89.5% et un mAP@0,50 :0,95 de 61.1%, Les pertes d'entraînement (box_loss, cls_loss, dfl_loss) ont affiché une décroissance régulière et stable, convergeant respectivement vers environ 1.07, 0.72 et 0.012 (voir la Figure 4.23), la durée totale était de 2 heures et 49 minutes.

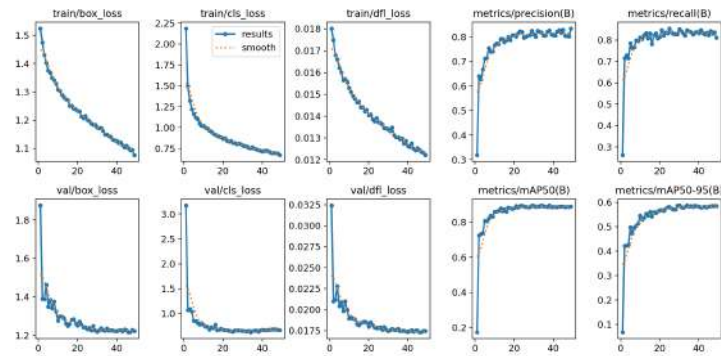


Figure 4.23 – Performances de l'apprentissage du Modèle 4

Par ailleurs, nous avons affiché les matrices de confusion pour chaque modèle afin de faciliter l'analyse et la visualisation des erreurs de classification. Cet outil permet de confronter les prédictions du modèle aux valeurs réelles, en offrant une vue structurée des correspondances et des divergences entre classes prédites et classes effectives. Chaque ligne de la matrice représente la classe prédite, tandis que chaque colonne correspond à la classe réelle (ou vice versa). Le tableau 4.6 illustre la matrice de confusion pour un problème de classification binaire.

	Classe réelle positive	Classe réelle négative
Classe prédite positive	True Positif (TP)	False Positif (FP)
Classe prédite négative	False Négatif (FN)	True Négatif (TN)

Table 4.6 – Matrice de confusion binaire

Telle que :

- Vrai Positif (True Positif "TP") : lorsque la classe prédite est positive et la classe réelle est également positive ;
- Faux Positif (False Positif "FP") : lorsque la classe prédite est positive alors que la classe réelle est négative ;
- Vrai Négatif (True Negatif "TN") : lorsque la classe prédite est négative mais la classe réelle est négative ;
- Faux Négatif (False Negatif "FN") : lorsque la classe prédite est négative et la classe réelle est également positive ;

Dans le but d'évaluer rigoureusement la performance des modèles, cette étude s'appuie sur des métriques classiques telles que la précision, le rappel, le score F1, l'accuracy et la mAP. Les équations (1) à (4) permettent de calculer ces

différentes métriques. Les performances globales s'améliorent lorsque ces indicateurs augmentent.

- Précision : Il s'agit du nombre total de vrais positifs divisé par le nombre total d'entrées marquées comme appartenant à la classe positive ;

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \quad (4.1)$$

- Recall : Il s'agit du nombre total de vrais positifs divisé par le nombre total d'éléments qui appartiennent réellement à la classe positive ;

$$\text{Recall} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (4.2)$$

- F1_score : Calculé en combinant la précision et le rappel du modèle. C'est particulièrement utile lorsque les classes sont déséquilibrées.

$$\text{F1Score} = 2 * \text{precision} * \frac{\text{recall}}{(\text{precision} + \text{recall})} \quad (4.3)$$

- Intersection over Union (IoU) : mesure le chevauchement entre une boîte englobante prédite et la boîte englobante réelle.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (4.4)$$

- mAP (mean Average Precision) : mesure globale utilisée notamment en détection et segmentation d'objets, représentant la moyenne des précisions moyennes sur toutes les classes et différents seuils d'IoU.

$$\text{mAP} = \sum_{k=1}^n \frac{\text{AP}(k)}{n} \quad (4.5)$$

Où n est le nombre de catégories et $\text{AP}(k)$ la précision moyenne de la classe k . Un AP plus élevé reflète de meilleures performances de détection.

On distingue notamment le mAP50, calculé à un seuil d'IoU fixe de 0,5, et le mAP50-95, moyenné sur plusieurs seuils d'IoU allant de 0,5 à 0,95 par pas de 0,05, offrant ainsi une évaluation plus stricte et plus complète de la précision de localisation.

Les Figures 4.25, 4.26 et 4.27 représentent les matrices de confusion obtenues pour chacun des modèles évalués (RF-DETR, YOLOE, YOLOv8 et YOLO26), où les deux classes cibles (Benign et Malignant) ainsi que le background.

La matrice de confusion du modèle RF-DETR pour la détection de nodules en 2 classes révèle un total de 151 instances mal classées sur 1 140 au total. Il s'agit notamment de 48 instances de la classe `benign_nodule` qui ont été classées à tort comme nodule. Du côté de la classe nodule, 103 instances ont été erronément classées comme `benign_nodule`. En revanche, le modèle a correctement classé 651 instances de la classe `benign_nodule` et 338 instances de la classe nodule (voir Figure 4.24) .

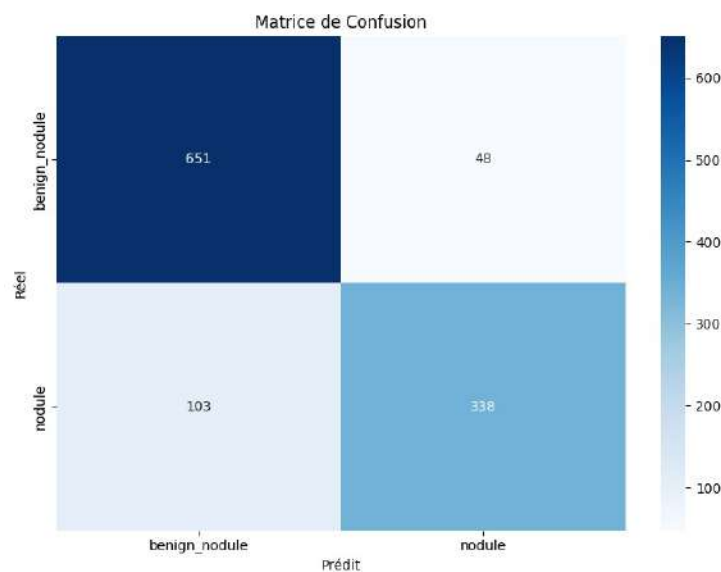


Figure 4.24 – Matrice de confusion du Modèle RF-DETR

Avec l'architecture du Modèle YOLOv8, la matrice de confusion illustrée en Figure 4.25 présente des comptages au niveau des prédictions, le modèle classe correctement 1 357 instances de la classe 0 (benign) et 653 instances de la classe 1 (maligne). S'agissant des erreurs de classification, 158 instances de la classe 1 (maligne) et 421 instances de la classe background ont été incorrectement attribuées à la classe 0 (benign), tandis que 127 instances de la classe 0 (benign) et 260 instances de la classe background ont été mal assignées à la classe 1 (maligne). De surcroît, 24 instances de la classe 0 (benign) et 5 instances de la classe 1 (maligne) ont été erronément prédites comme relevant de la catégorie background.

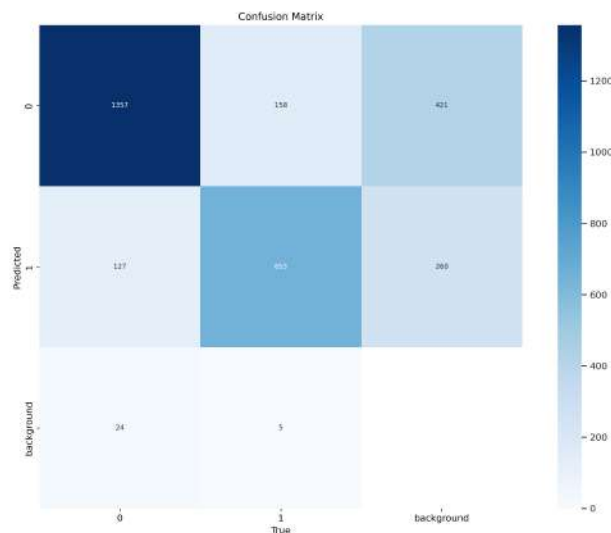


Figure 4.25 – Matrice de confusion du Modèle YOLOv8

Pour le modèle YOLOE, illustre les performances de détection sur trois catégories : benign, malignant et background. Les vrais positifs s'élèvent à 1 368 détections correctes pour les nodules bénins et à 649 pour les nodules malins, indiquant d'une bonne capacité de localisation pour ces deux classes. Concernant les erreurs de classification inter-classes, 153 instances réelles de nodule ont été prédites comme benign nodule, et inversement, 105 instances de benign nodule ont été confondues avec la classe nodule. Par ailleurs, le modèle présente des confusions avec le fond : 147 instances de background ont été détectées comme benign nodule et 121 comme nodule. En sens inverse, 35 véritables benign nodules et 14 véritables nodules ont été considérés comme background, ce qui correspond à des faux négatifs (voir Figure 4.26).

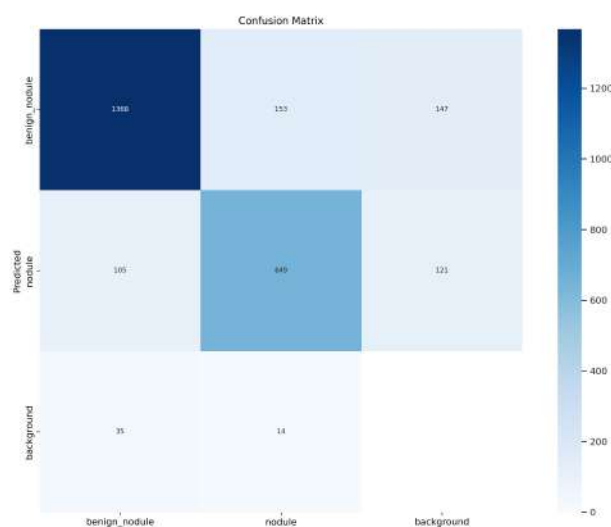


Figure 4.26 – Matrice de confusion du Modèle YOLOE

Pour le modèle YOLO26, la matrice de confusion illustrée en Figure 4.27 montre que le modèle a classifié correctement 959 instances bénignes et 819 instances malignes. Concernant les erreurs de classification, 143 images malignes et 179 instances background ont été incorrectement attribuées à la classe Bénigne, et 122 images bénignes ainsi que 136 images background ont été mal assignées à la classe Maligne. Par ailleurs, 38 instances bénignes et 11 instances malignes ont été incorrectement attribuées à la catégorie background, ce qui témoigne de la difficulté persistante du modèle à distinguer cette classe des deux classes pathologiques, avec des taux de confusion de 57% vers la classe bénigne et de 43% vers la classe maligne.

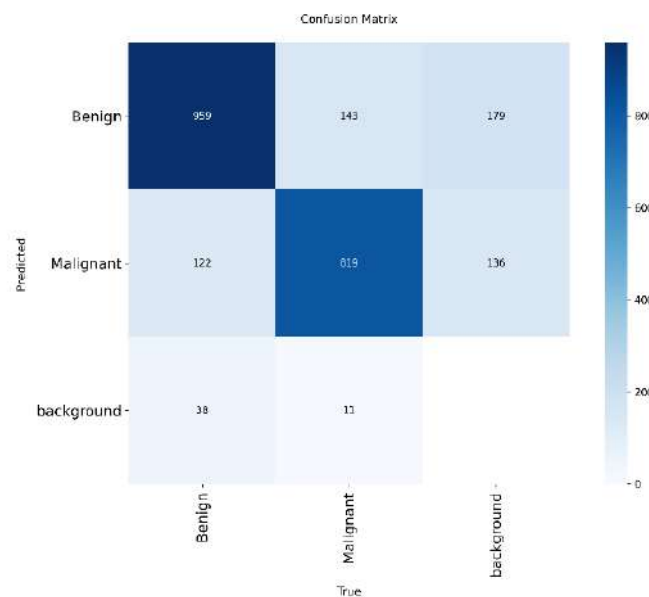


Figure 4.27 – Matrice de confusion du Modèle YOLO26

Le tableau 4.7 présente un récapitulatif des performances de nos quatre modèles sur les données de test du dataset ThyroidXL, où sont reportées les statistiques pour chacune des deux classes (Benign et Malignant), ainsi que la moyenne des performances pour chaque modèle.

Classe	Précision	Rappel	F1-score	Support
Benign_nodule	0.86	0.93	0.90	699
Maligne_nodule	0.88	0.77	0.82	441
Accuracy			0.87	1140
Macro avg	0.87	0.85	0.86	1140
Weighted avg	0.87	0.87	0.87	1140

(a) RF-DETR

Classe	Images	Instances	Precision	Recall	mAP50	mAP@0.50 :0.95
all	2324	2324	0.82	0.853	0.895	0.611
Benign_nodule	1508	1508	0.84	0.883	0.919	0.671
Maligne_nodule	816	816	0.801	0.823	0.872	0.55

(b) YOLOv8

Classe	Images	Instances	Precision	Recall	mAP50	mAP@0.50 :0.95
all	2324	2324	0.821	0.874	0.901	0.61
Benign_nodule	1508	1508	0.835	0.91	0.929	0.668
Maligne_nodule	816	816	0.807	0.838	0.873	0.551

(c) YOLOE

Classe	Images	Instances	Precision	Recall	mAP50	mAP@0.50 :0.95
all	2092	2092	0.817	0.847	0.896	0.587
Benign_nodule	1119	1119	0.81	0.85	0.89	0.626
Maligne_nodule	973	973	0.824	0.844	0.902	0.547

(d) YOLO26

Table 4.7 – Rapports de classification

Dans un contexte médical, les faux négatifs (nodules malins non détectés) sont particulièrement critiques, puisqu'ils peuvent retarder le diagnostic et la prise en charge du patient. Ainsi, au-delà du mAP, il est essentiel d'évaluer la capacité du modèle à détecter correctement cette classe à travers des indicateurs spécifiques comme la sensibilité, qui mesure la capacité du système à identifier correctement ces cas critiques. Le calcul du taux de faux négatifs (FNR) par classe constitue une métrique plus pertinente pour l'évaluation clinique (voir le tableau 4.8).

Metric	RF-DETR	YOLOv8	YOLOE	YOLO26
Recall (Maligne)	0.770	0.823	0.838	0.844
FNR (Maligne)	0.230	0.177	0.162	0.156

Table 4.8 – Analyse du Recall et du FNR pour la classe maligne

Pour une analyse plus approfondie, le tableau 4.9 présente les performances de notre modèle sur la même base de données.

Modèle	map@0,50	mAP@0,50 :0,95	Précision	Rappel	F1-score
RF-DETR	84%	51%	87%	87%	87%
YOLOv8	89.5%	61.1%	82%	85.3%	83.6%
YOLOE	90.1%	61%	82.1%	87.4%	84.7%
YOLO26	89.6%	58.7%	81.7%	84.7%	83.2%

Table 4.9 – Comparaison des performances des modèles de détection

Ce tableau présente une comparaison des performances de quatre modèles de détection d'objets selon plusieurs métriques clés. Parmi eux, YOLOE et YOLOv8 se distinguent comme les plus performants. YOLOE obtient le meilleur map@0,50 (90.1%) ainsi qu'un rappel élevé (87.4%), démontrant une excellente capacité de détection, tandis que YOLOv8 atteint le meilleur mAP@0,50 :0,95 (61%). Les autres modèles présentent des performances plus intermédiaires, avec RF-DETR montrant un bon équilibre mais une robustesse limitée à IoU élevé. Partageant une architecture YOLO compatible, YOLOE et YOLOv8 constituent les meilleurs candidats pour la conception d'un modèle hybride. Leur fusion permettrait d'associer la forte capacité de détection de YOLOE à la précision de localisation de YOLOv8, afin d'obtenir un détecteur plus robuste, réduisant à la fois les faux négatifs et les erreurs de localisation.

4.8 Les Approches Hybrides Proposées

Dans le cadre de cette étude, ce travail explore des approches hybrides combinant les architectures YOLOv8 et YOLOE selon des stratégies d'intégration distinctes : la fusion pondérée des boîtes (Weighted Boxes Fusion ,WBF), la suppression non-maximale en ensemble (Non-Maximum Suppression ,NMS), l'architecture en cascade, ainsi qu'une approche combinaison WBF et NMS. Chaque pipeline intègre un seuillage asymétrique par classe favorisant la détection des nodules malins, et est évalué selon un protocole unifié incluant les métriques map@0,50 , mAP@0,50 :0,95 , précision , rappel et F1-score, complétées par l'analyse des matrices de confusion et la visualisation des prédictions sur le jeu de test ThyroidXL.

4.8.1 Approches en cascade

4.8.1.1 Cascade YOLOv8 → YOLOE

L'approche cascade développée dans ce travail combine deux modèles complémentaires pour la détection des nodules thyroïdiens sur images échographiques. Le pipeline fonctionne en deux étapes séquentielles : YOLOv8 intervient d'abord comme localisateur global, détectant les régions d'intérêt (ROI) sur l'image complète avec un seuil de confiance de 0.25, qui sont ensuite transmises à YOLOE, qui se charge de la classification fine entre nodules bénins et malins.

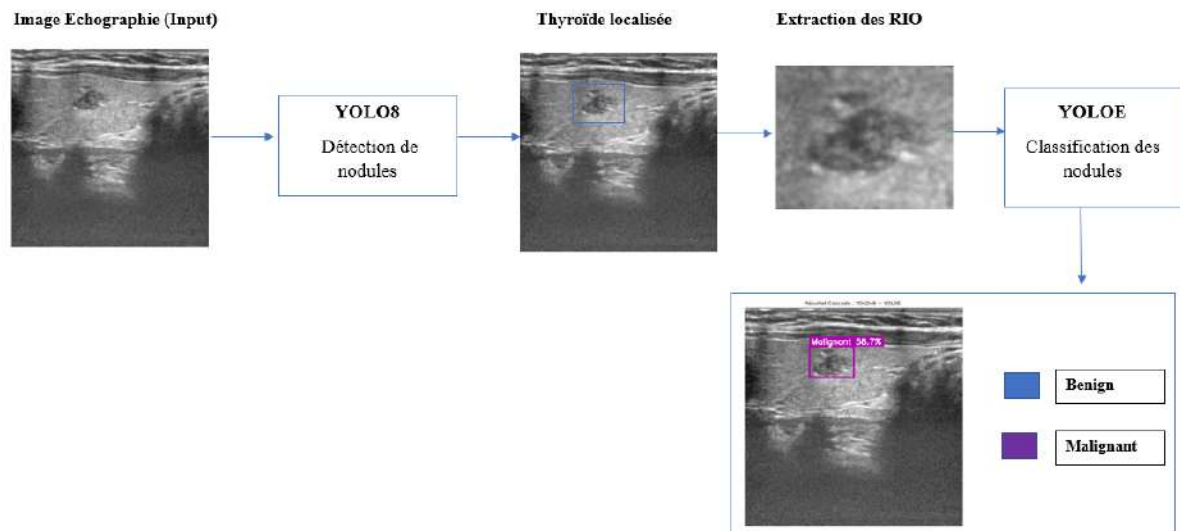


Figure 4.28 – Architecture Hybride en cascade YOLOv8 → YOLOE

4.8.1.2 Cascade YOLOE → YOLOv8

L'approche cascade inverse propose une architecture complémentaire où YOLOE joue le rôle de détecteur primaire, exploitant ses capacités de généralisation issues de la distillation de connaissances pour proposer des régions candidates sur l'image échographique complète avec un seuil de confiance de 0.25. Ces régions d'intérêt sont ensuite transmises à YOLOv8, qui intervient en second étage comme classificateur de précision entre nodules bénins et malins.

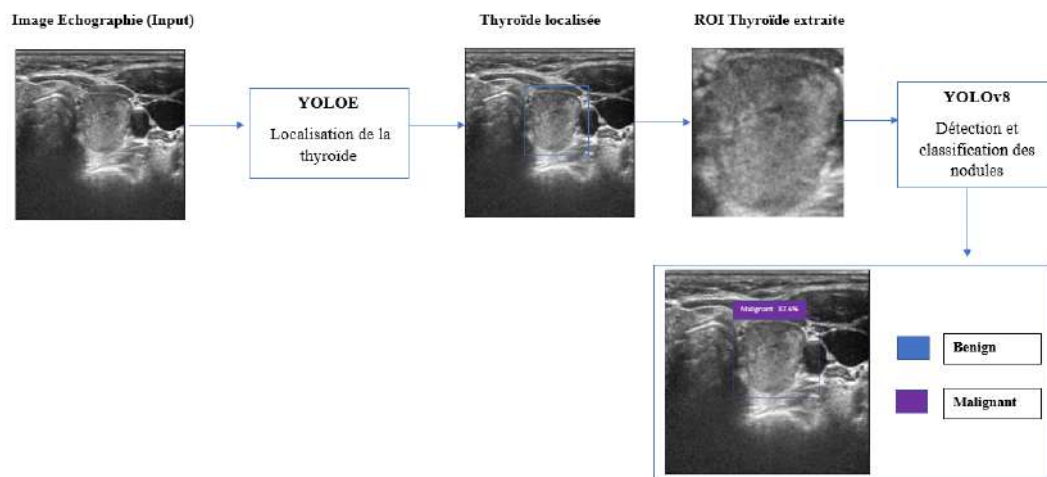


Figure 4.29 – Architecture Hybride en cascade YOLOE → YOLOv8

4.8.2 Approches basée sur la fusion des prédictions

4.8.2.1 Fusion par Non-Maximum Suppression (NMS)

Le NMS Ensemble est un algorithme de post-traitement utilisé en détection d'objets qui fusionne les prédictions de YOLOv8, entraîné sur la détection des nodules thyroïdiens et de YOLOE issu d'un processus de distillation de connaissance, en concaténant leurs boîtes englobantes, puis applique un NMS par classe avec un seuil de confiance de 0.25 et un seuil IoU de 0.50, afin d'éliminer les détections redondantes inter-modèles et ne conserver que les plus pertinentes.

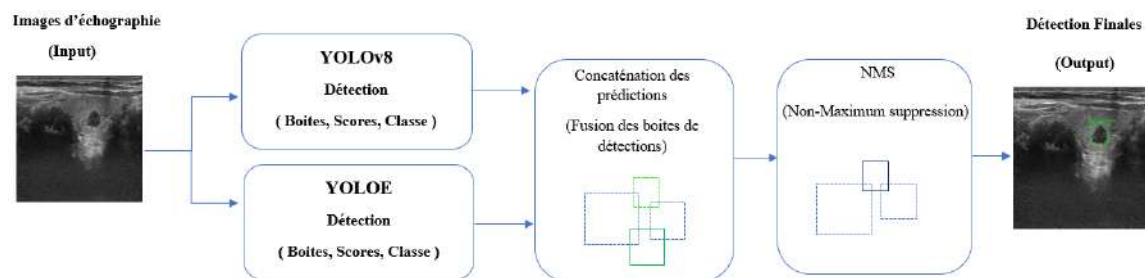


Figure 4.30 – Architecture Hybride basée sur la fusion NMS

4.8.2.2 Fusion par Weighted Boxes Fusion (WBF)

Le pipeline hybride WBF combine deux modèles de détection YOLOv8 et YOLOE pour classifier les nodules thyroïdiens. Chaque modèle prédit indépendamment des boîtes englobantes sur l'image, puis ces prédictions sont normalisées en coordonnées $[0, 1]$ et fusionnées via WBF (seuil IoU = 0.5), qui calcule une boîte moyenne pondérée par les scores de confiance plutôt que de supprimer les détections redondantes. Cette fusion tire profit des complémentarités des deux modèles pour produire des détections plus robustes.

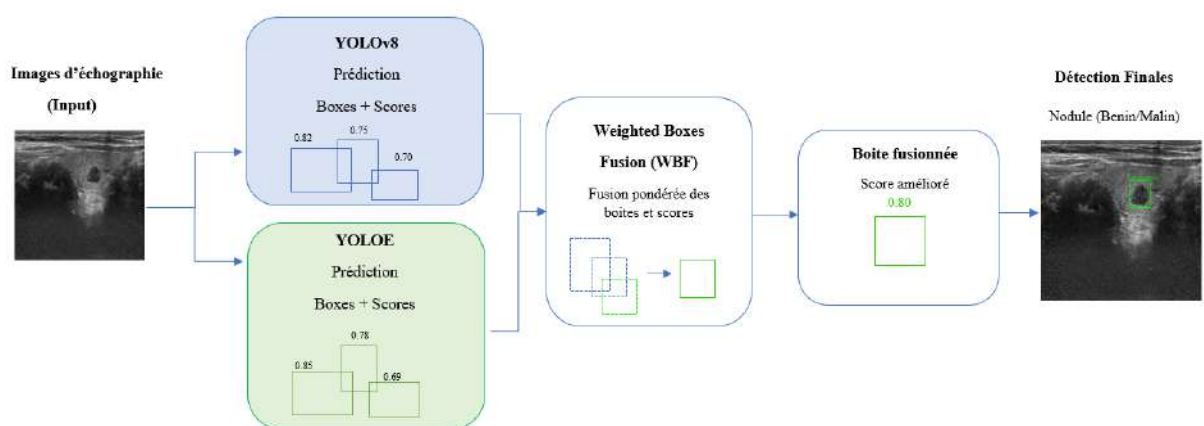


Figure 4.31 – Architecture Hybride basée sur la fusion WBF

4.8.2.3 Fusion Hybride WBF + NMS

L'architecture WBF+NMS repose sur la combinaisons des prédictions a l'aide de la technique de fusion pondérée des boîtes (WBF) afin d'affiner la localisation des boîtes englobantes, puis applique un NMS pour éliminer les redondances .Enfin, un filtrage basé sur un seuil de confiance pour ne retenir que les prédictions les plus fiables, aboutissant à une détection finale classifiant la lésion comme Bénigne ou Maligne.

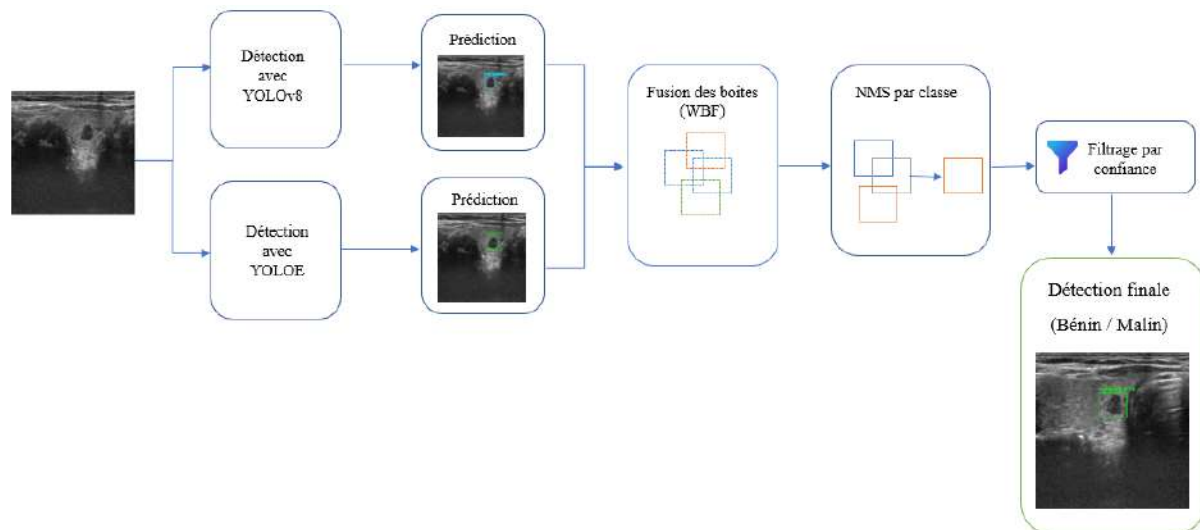


Figure 4.32 – Architecture Hybride basée sur la fusion WBF+NMS

4.8.3 Évaluation expérimentale des approches hybrides

Les Figures 4.33, 4.34 représentent les matrices de confusion obtenues pour chaque approches hybrides proposées.

La matrice de confusion de L'approche hybride cascade YOLOv8 →YOLOE (voir Figure 4.33 (a)) classe correctement 1 334 nodules bénins et 682 malins, mais génère 209 faux négatifs malins et 277 faux positifs malins, avec des fausses détections sur background (296 et 277 respectivement).

Pour l'approche inverse YOLOE →YOLOv8 (voir Figure 4.33 (b)) améliore légèrement la détection bénigne (1 363 vrais positifs) mais dégrade le rappel malignité, avec 306 faux négatifs malins contre 209 pour la configuration précédente, indiquant que YOLOv8 en reclassificateur tend à sous-estimer la classe maligne.

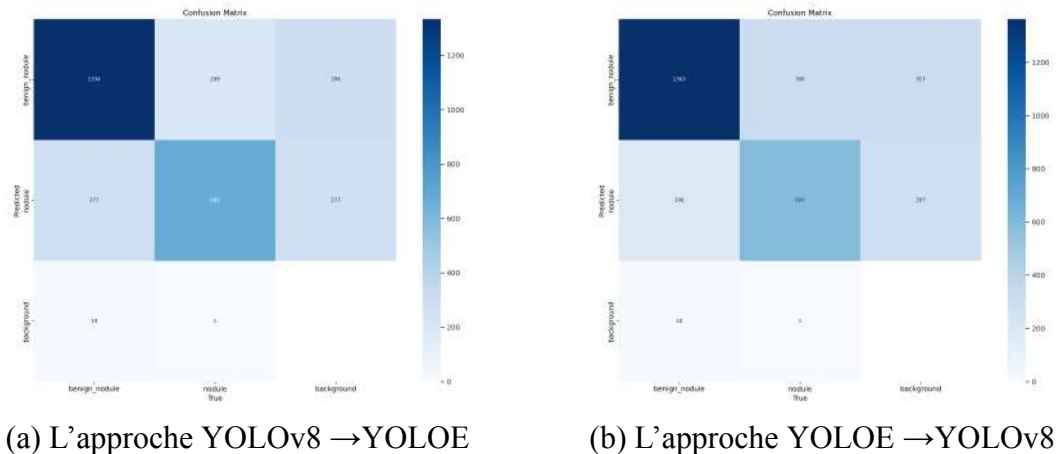


Figure 4.33 – Matrice de confusion de L'approche hybride cascade

Approche hybride basée sur la fusion NMS (voir Figure 4.34 (a)) offre 343 vrais positifs malins et 598 bénins correctement classifiés, mais présente 162 faux négatifs malins. Sa nature sélective et non pondérée limite sa capacité à préserver les prédictions concordantes sur des nodules malins difficiles.

Approche hybride basée sur la fusion par WBF (voir Figure 4.34 (b)) réduit significativement les faux négatifs malins à 94, avec 410 vrais positifs malins, au prix d'un taux de faux positifs légèrement plus élevé (162).

La matrice de confusion de l'architecture combinée (voir Figure 4.34 (c)) révèle 582 vrais négatifs et 50 faux positifs pour les nodules bénins, contre 404 vrais positifs et 102 faux négatifs pour les nodules malins. Ce taux de faux négatifs, bien que non négligeable, demeure acceptable dans un contexte clinique où la priorité est accordée à la détection des cas malins.

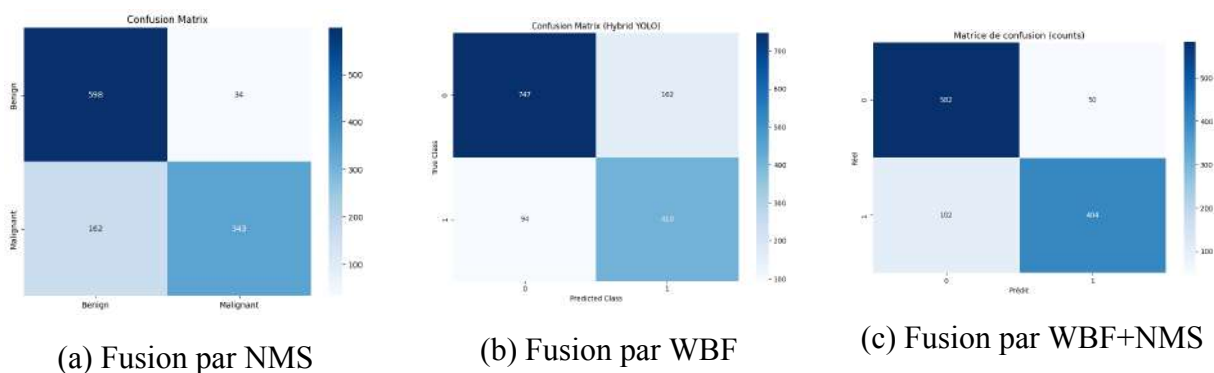


Figure 4.34 – Matrices de confusion des différentes approches de fusion.

Le Tableau 4.10 présente les métriques d'évaluation obtenues par les approches hybrides sur le jeu de test, comparées aux deux modèles de base YOLOv8 et YOLOE utilisés individuellement.

Classe	P	R	F1	map@0,50	mAP@0,50 :0,95
All	74.83 %	87.40 %	80.56 %	88.07 %	72.00 %
Benign Nodule	77.20 %	92.60 %	84.20 %	93.00 %	78.00 %
Malignant Nodule	72.50 %	82.20 %	77.05 %	83.10 %	66.00 %

(a) Cascade YOLOE–YOLO8

Classe	P	R	F1	map@0,50	mAP@0,50 :0,95
All	76.30 %	87.40 %	81.50 %	87.60 %	73.70 %
Benign Nodule	83.50 %	87.10 %	85.30 %	92.50 %	73.70 %
Malignant Nodule	69.00 %	87.60 %	77.20 %	82.80 %	68.00 %

(b) Cascade YOLOv8–YOLOE

Classe	P	R	F1	map@0,50	mAP@0,50 :0,95
All	80.5 %	97.43 %	88.14 %	83.29%	54.25%
Benign Nodule	90.4 %	97.2 %	93.7 %	85.29%	60.04%
Malignant Nodule	65.9 %	97.9 %	78.8 %	81.28%	48.46%

(c) Fusion par NMS

Classe	P	R	F1	map@0,50	mAP@0,50 :0,95
All	80.54 %	97.51 %	88.22 %	97.32 %	64.26 %
Benign Nodule	79.25 %	96.94 %	87.21 %	97.81 %	69.63 %
Malignant Nodule	82.22 %	98.25 %	89.52 %	96.83 %	58.89 %

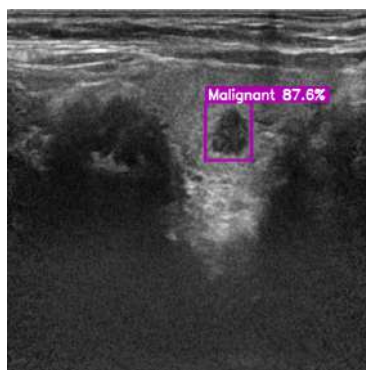
(d) Fusion par WBF

Classe	P	R	F1	map@0,50	mAP@0,50 :0,95
All	88.63%	97.51%	92.86%	83.56%	55.39%
Benign Nodule	93.83%	96.88%	95.33%	84.05%	60.75%
Malignant Nodule	81.80%	98.49%	89,37%	83.07%	50.03%

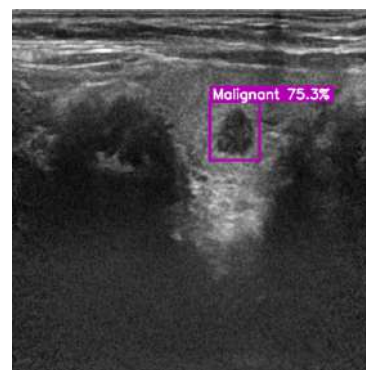
(e) Fusion par WBF + NMS

Table 4.10 – Les Rapports de classification des approches hybrides

Afin d'évaluer qualitativement le comportement de chaque approche hybride, nous présentons dans cette section des exemples de prédictions effectuées sur une même image d'échographie thyroïdienne présentant un nodule malin. Les approches hybrides développées produisent des prédictions concordantes, attribuant unanimement la classe Malignant au nodule détecté. Parmi l'ensemble des approches, le pipeline en cascade YOLOv8→YOLOE se distingue comme la meilleure configuration, atteignant le score de confiance le plus élevé de 87.6% pour la détection du nodule malin.



(a) Cascade YOLOv8–YOLOE



(b) Cascade YOLOE–YOLOv8

Figure 4.35 – Visualisation des prédictions de l'approche en cascade

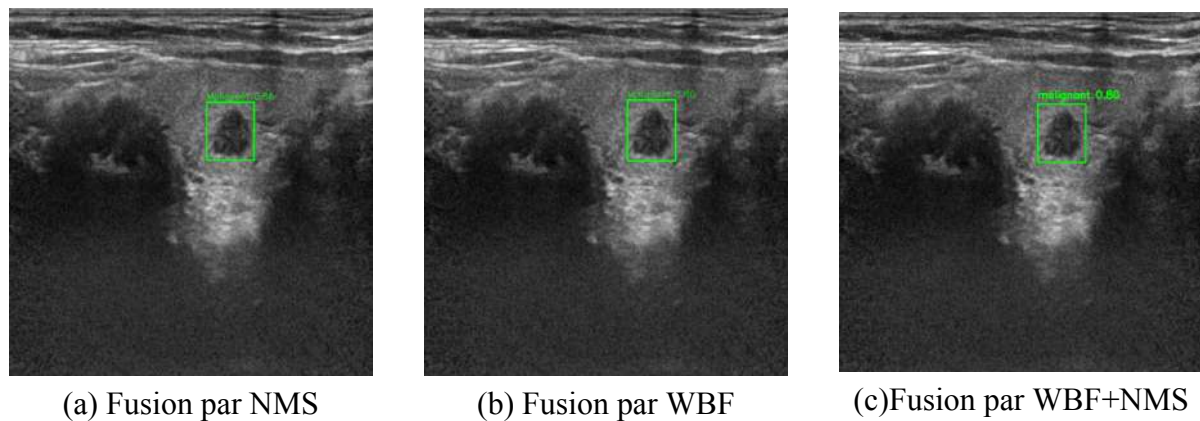


Figure 4.36 – Visualisation des prédictions de l'approche basée sur la fusion

Discussion et Analyse Comparative des Approches Hybrides

Les résultats expérimentaux présentés dans la section précédente permettent de procéder à une analyse comparative des cinq approches hybrides évaluées, à savoir deux architectures séquentielles en cascade (YOLOE→YOLOv8 et YOLOv8→YOLOE) et trois stratégies de fusion des détections issues des deux modèles (NMS, WBF, et la combinaison WBF+NMS). Cette discussion vise, dans un premier temps, à dégager les tendances qui se dégagent de cette comparaison, puis à interpréter ces tendances à la lumière des mécanismes propres à chaque stratégie, et enfin à identifier l'approche la plus pertinente au regard de l'objectif, à savoir la maximisation conjointe de la sensibilité de détection et de la fiabilité des prédictions, en particulier pour la classe des nodules malins.

Approche	F1(global)	map@0,50	mAP@0,50 :0,95	F1(malin)	R(malin)
(a) Cascade YOLOE→YOLOv8	80,56 %	88,07 %	72,00 %	77,05 %	82,20 %
(b) Cascade YOLOv8→YOLOE	81,50 %	87,60 %	73,70 %	77,20 %	87,60 %
(c) Fusion NMS	88,14 %	83,29 %	54,25 %	78,80 %	97,90 %
(d) Fusion WBF	88,22 %	97,32 %	64,26 %	89,52 %	98,25 %
(e) Fusion WBF+NMS	92,90 %	83,56 %	55,39 %	89,37 %	98,49 %

Table 4.11 – Synthèse comparative des indicateurs de performance des cinq approches hybrides évaluées (classe « Toutes catégories » et classe « Nodule malin »)

Compromis entre architectures séquentielles et stratégies de fusion

L'analyse des indicateurs rapportés au Tableau 4.11 met en évidence un compromis structurant entre les deux familles d'approches considérées. Les architectures en cascade, qui consistent à appliquer successivement les deux détecteurs

selon un ordre prédéfini, atteignent les valeurs de $mAP@0,50 : 0,95$ les plus élevées de l'étude (72,00 % pour la configuration YOLOE→YOLOv8 et 73,70 % pour la configuration YOLOv8→YOLOE), ce qui témoigne d'une localisation spatiale relativement précise des boîtes englobantes sur l'ensemble de la plage des seuils d'intersection sur union (IoU). Ce résultat s'explique vraisemblablement par le fait que chaque détecteur conserve, dans une architecture en cascade, sa géométrie de prédiction propre, sans opération de fusion susceptible d'altérer la précision des contours détectés.

Ces deux architectures présentent toutefois une limite notable du point de vue du rappel sur la classe des nodules malins, qui demeure compris entre 82,20 % et 87,60 % selon la configuration. Dans le contexte d'une tâche de dépistage à visée diagnostique, cette limitation revêt une importance particulière : une sous-détection, même partielle, des lésions malignes constitue un risque clinique plus préjudiciable qu'un excès de faux positifs, ces derniers pouvant être écartés lors d'un examen complémentaire alors qu'un faux négatif expose à un retard de prise en charge.

À l'inverse, les trois stratégies de fusion, qui exploitent conjointement les sorties des deux détecteurs plutôt que de les enchaîner de manière séquentielle, permettent d'obtenir un rappel quasi systématique sur la classe maligne, compris entre 97,90 % et 98,50 % selon la méthode de fusion retenue, ainsi qu'un F1-score global sensiblement supérieur à celui des architectures en cascade (de 88,14 % à 92,90 %, contre 80,56 % et 81,50 % respectivement). Cette amélioration substantielle de la sensibilité de détection s'accompagne cependant d'une diminution notable du $mAP@0,50 : 0,95$, qui se situe entre 54,25 % et 64,26 % pour les trois variantes de fusion, soit un écart de l'ordre de quinze à vingt points par rapport aux architectures en cascade.

Interprétation du comportement du $mAP@0,50 : 0,95$ pour les méthodes de fusion

La baisse conjointe du $mAP@0,50 : 0,95$ observée pour les trois stratégies de fusion peut être rapportée aux mécanismes intrinsèques de ces méthodes. La fusion pondérée des boîtes (WBF) procède par moyennage géométrique des coordonnées des boîtes candidates issues des deux détecteurs, pondéré par leur score de confiance respectif, tandis que la suppression des recouvrements (NMS) sélectionne, parmi un ensemble de détections redondantes, celle dont le score est le plus élevé, au détriment des autres. Dans les deux cas, la boîte résultante, bien que correctement positionnée sur la lésion considérée, ne correspond pas néces-

sairement à l'optimum géométrique qu'aurait pu produire un détecteur unique parfaitement calé sur le contour réel de l'objet. Cet écart géométrique, généralement mineur, devient sensible aux seuils d'IoU élevés (0,75 à 0,95), ce qui explique la chute du $mAP@0,50 : 0,95$ sans que cela traduise nécessairement une dégradation de la capacité de détection à proprement parler, comme en atteste le maintien, voire l'amélioration, du $map@0,50$ et du rappel pour ces mêmes approches.

Ce constat invite à nuancer l'interprétation du $mAP@0,50 : 0,95$ comme indicateur unique de performance : dans une application à finalité diagnostique, où la détection de la présence d'une lésion prime sur la précision millimétrique de son contour, un score de rappel élevé associé à un $map@0,50$ satisfaisant peut constituer un critère de décision plus pertinent que le $mAP@0,50 : 0,95$ seul.

Identification de l'approche la plus performante

Au regard de l'ensemble des indicateurs considérés, et en accordant une importance particulière au F1-score global ainsi qu'au rappel sur la classe des nodules malins, l'approche par fusion WBF+NMS se distingue comme l'approche la plus performante et pertinente parmi les cinq testées. Elle obtient en effet le meilleur F1-score global de l'étude (92,90 %) ainsi que le meilleur rappel sur la classe maligne (98,50 %), tout en conservant un niveau de précision satisfaisant sur les deux classes (88,6 % pour l'ensemble des catégories et 81,8 % pour la classe maligne). Ce résultat traduit le meilleur équilibre observé entre sensibilité diagnostique et fiabilité des prédictions parmi l'ensemble des configurations évaluées. L'approche par fusion WBF seule constitue néanmoins une alternative digne d'intérêt, dans la mesure où elle obtient un F1-score sur la classe maligne quasiment équivalent (89,52 % contre 89,37 % pour WBF+NMS) ainsi que le meilleur $map@0,50$ de l'ensemble de l'étude (97,32 %). Cette configuration pourrait être privilégiée dans l'hypothèse où le $map@0,50$ constituerait le critère de sélection prioritaire. Les architectures en cascade conservent, quant à elles, un intérêt méthodologique en tant que références de comparaison, permettant de quantifier l'apport spécifique des stratégies de fusion en matière de rappel et d'équilibre précision/rappel.

Limites de l'étude et perspectives

Plusieurs limites doivent être soulignées dans des travaux ultérieurs :

- L'écart important entre le $map@0,50$ (97,32 %) et le $mAP@0,50 : 0,95$ (64,26 %) observé pour l'approche WBF seule mériterait une analyse complémentaire, notamment par l'examen de la distribution des valeurs d'IoU des

boîtes fusionnées, afin de déterminer si cet écart traduit une limite intrinsèque de la méthode ou un effet propre à la composition du jeu de données évalué.

- Une validation croisée, par exemple selon un schéma de validation à k plis, ou une évaluation sur un jeu de données externe et indépendant, permettrait de confirmer la stabilité du classement observé entre les cinq approches et d'en renforcer la portée généralisable.
- Le coût computationnel et le temps d'inférence associés à chaque approche n'ont pas été comparés dans le cadre de cette étude. Cette dimension revêt une importance pratique non négligeable dans la perspective d'un déploiement clinique, en particulier si une exploitation en temps réel ou sur des ressources matérielles limitées est envisagée.
- Une analyse qualitative des erreurs résiduelles (faux positifs et faux négatifs) de l'approche WBF+NMS permettrait d'identifier d'éventuels biais systématiques associés à certains sous-types morphologiques de nodules, et d'orienter d'éventuelles améliorations ciblées du pipeline de fusion.

En définitive, les résultats obtenus mettent en évidence la supériorité globale des stratégies de fusion sur les architectures en cascade pour la tâche de détection et de classification des nodules thyroïdiens considérée, et plus particulièrement la pertinence de la stratégie de fusion WBF+NMS, qui réalise le meilleur compromis entre sensibilité de détection, fiabilité des prédictions et performance globale. Cette approche sera retenue comme configuration de référence pour la suite de ce travail, l'approche WBF seule étant présentée à titre de variante comparative permettant d'illustrer la robustesse des méthodes de fusion sur la classe diagnostiquement la plus critique.

4.9 Difficultés rencontrées

Pendant la réalisation de notre système de détection automatique des nodules thyroïdiens, plusieurs contraintes techniques liées aux ressources matérielles des plateformes en ligne ont été rencontrées, notamment Google Colab. En effet, bien que cette plateforme offre initialement une durée d'utilisation du GPU d'environ 12h par jour, elle a été réduite à moins d'une heure quotidienne en raison de l'utilisation intensive ainsi que de la complexité des modèles pré-entraînés utilisés. Nous avons alors décidé de migrer vers la plateforme Lightning AI grâce aux ressources fournies (RAM plus élevée et un accès de 80 h GPU/mois), malgré

ces améliorations, des erreurs de type CUDA Out Of Memory sont survenues, la mémoire GPU restant insuffisante pour supporter à la fois le volume des données et la complexité des modèles. Afin de pallier ce problème, nous avons réduit la taille des images à 640×640 pixels et limité la taille du batch à 16. Par ailleurs, le chargement des données depuis la plateforme Roboflow a engendré des temps de transfert significatifs, en raison du volume important du jeu de données.

4.10 Conclusion

Lors de la conception de notre système de détection des nodules thyroïdiens, nous avons souligné l'importance cruciale du choix du dataset ainsi que des modèles de détection. Par ailleurs, la sélection des architectures de base telles que YOLOv8 et YOLOE s'est révélée essentielle pour construire une approche d'ensemble et permettre la fusion des prédictions à travers différentes stratégies hybrides. Le bon ajustement des hyperparamètres, notamment le seuil de confiance, le seuil d'IoU, la taille des images, le nombre d'époques, le batch size, optimiseur... etc, pour garantir des performances de qualité. Les résultats obtenus démontrent l'efficacité de l'approche hybride proposée, confirmant sa capacité à détecter et localiser avec précision les nodules thyroïdiens.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

La détection automatique des maladies thyroïdiennes est l'un des sujets les plus importants dans le domaine de la vision par ordinateur appliquée à la médecine. Les pathologies thyroïdiennes, notamment les nodules et le cancer, résultent de la transformation anarchique de cellules initialement normales, pouvant former des tumeurs malignes ou, plus fréquemment, des lésions bénignes susceptibles d'entraîner de graves complications en l'absence de diagnostic précoce. Face à ces enjeux cliniques, l'analyse automatisée des anomalies thyroïdiennes s'impose comme une problématique centrale en endocrinologie, suscitant un intérêt croissant de la communauté scientifique. La détection d'objets par apprentissage automatique en vision par ordinateur, offre des perspectives prometteuses pour détecter et de classifier avec précision les anomalies, palliant ainsi les limites de l'interprétation manuelle des images échographiques, tâche complexe, chronophage et exposée aux erreurs humaines.

L'objectif de ce projet est de développer un système de détection des nodules thyroïdiens à partir d'images médicales, à la fois performant et capable de généraliser à d'autres types de données n'a pas été démontrée, en utilisant les techniques d'apprentissage profond. Pour cela, nous avons d'abord comparé plusieurs modèles de détection modernes, à savoir YOLOv8, YOLOE, YOLO26 et RF-DETR afin d'évaluer leurs performances. Ensuite, en nous appuyant sur les meilleurs résultats obtenus, nous avons proposé des approches hybrides combinant une architecture en cascade et différentes méthodes de fusion (NMS, WBF et WBF+NMS), afin de comparer et analyser leur efficacité.

Ce manuscrit présente les différentes étapes de notre démarche et contribution. La première étape consiste à charger les données, suivie d'une étape de prétraitement à l'aide de la plateforme Roboflow. L'étape la plus importante consiste à choisir des réseaux de base et les bons hyperparamètres pour obtenir un modèle robuste et précis. Enfin, la dernière étape consiste à détecter et localiser les anomalies thyroïdiennes sur les images échographiques. Les expériences sont menées sur une base de données ThyroidXL, où nous comparons divers modèles selon des métriques de détection. Les résultats ont montré que les approches hybrides offrent les performances les plus robustes et précises. En particulier, L'approche Fusion WBF+NMS est la plus performante globalement, avec le meilleur F1 (92,9 %) et le meilleur rappel sur le malin (98,5 %), offrant le meilleur compromis précision/rappel sur les deux classes. L'approche Fusion WBF seule reste

une alternative solide, en particulier pour le mAP@50 (97,32 %) et un F1 malin très proche (89,52 %).

En conclusion, ce travail confirme le potentiel significatif de la fusion de décision sur l'amélioration de la capacité d'aide à la detection et propose une approche méthodologique robuste qui rapproche les avancées de l'apprentissage profond aux besoins cliniques réels, contribuant ainsi au développement et de solutions d'imagerie médicale plus fiables. Afin d'améliorer davantage les performances obtenues, utiliser une intégration d'autres types de données multimodales (les données cliniques, les analyses biologiques ou encore les scintigraphies...). De plus, l'utilisation de bases de données plus large et diversifiées. Enfin, la conception d'une version mobile représenterait une étape intéressante afin de fournir aux professionnels de la santé un outil intelligent, accessible et complet d'aide au diagnostic des maladies thyroïdiennes.

Bibliographie

- [1] Fondation de l'Avenir. L'état de la recherche en imagerie médicale. URL : <https://www.fondationdelavenir.org/champs-action/nouvelles-technologies-en-sante/imagerie-medicate/>. (visité le 14/02/2026).
- [2] L'image numérique. URL : <https://www.nsi-lycee.fr/snt-2nde/images/image-numerique>. (visité le 05/07/2026).
- [3] Sara Khalfallah. « introduction à l'imagerie médicale ». In : Mémoire de master, Université Mohamed Khider de Biskra, (2020).
- [4] Cedric-Lelievre. Le traitement de l'image : de la couleur aux niveaux de gris. URL : https://cedric-lelievre.canoprof.fr/eleve/2SNT/05%20Module_La_Photographie_Num%C3%A9rique/co/Le_traitement_de_l_image_de_la_couleur_aux_niveaux_de_gris.html. (visité le 14/02/2026).
- [5] C. Houassine. «segmentation d'images par une approche biomimétique hybride ». In : Mémoire de Magister, Université M'hamed Bougara – Boumerdes, 2012.
- [6] S. Hassani and T. Issaoun. «segmentation d'images couleur par analyse d'histogrammes bidimensionnels 2d». Mém. de mast, Faculté de Génie Électrique et d'Informatique, Université Mouloud Mammeri de Tizi-Ouzou, 2017.
- [7] Zoubir Amair and Fouad Dehane. «segmentation d'images par la transformée de hough». Mém. de mast, Université de Ghardaïa, 2018.
- [8] Liv Hospital. 7 technologies d'imagerie médicale expliquées : un guide complet. URL : <https://www.livhospital.com/fr/7-technologies-dimagerie-medicate-expliquees:-un-guide-complet>. (visité le 19/02/2026).
- [9] *Qu'est-ce qu'une radiographie? L'essentiel sur cet examen.* URL : <https://www.sante.fr/quest-ce-quune-radiographie-lessentiel-sur-cet-examen>. (visité le 20/02/2026).

- [10] *Radiographie*. URL : <https://www.larousse.fr/encyclopedie/divers/radiographie/85277>. (visité le 20/02/2026).
- [11] *Scanner*. URL : <https://www.doctissimo.fr/html/sante/imagerie/scanner.htm>. (visité le 19/02/2026).
- [12] L'essentiel sur l'imagerie médicale. URL : <https://www.cea.fr/comprendre/Pages/sante-sciences-du-vivant/essentiel-sur-imagerie-medicale.aspx>. (Visité le 22/06/2026).
- [13] Irm – imagerie par résonance magnétique. URL : <http://concilio.com/orthopedie-irm-imagerie-par-resonance-magnetique>. (visité le 20/02/2026).
- [14] Imagerie par résonance magnétique (irm). URL : <https://www.msmanuals.com/fr/accueil/sujets-particuliers/examens-d-imagerie-courants/imagerie-par-r%C3%A9sonance-magn%C3%A9tique-irm>. (visité le 20/02/2026).
- [15] L'imagerie par résonance magnétique (irm). URL : <https://www.cancer.fr/personnes-malades/diagnostic/examens-de-diagnostic/irm>. (visité le 20/02/2026).
- [16] *Échographie (ultrason)*. URL : <https://www.hug.ch/radiologie/echographie-ultrason>. (visité le 20/02/2026).
- [17] *L'essentiel sur l'imagerie médicale*. URL : <https://www.cea.fr/comprendre/Pages/sante-sciences-du-vivant/essentiel-sur-imagerie-medicale.aspx>. (visité le 20/02/2026).
- [18] *La microscopie*. URL : <https://fac.umc.edu.dz/snv/faculte/BA/2020/5%20La%20microscopie.pdf>. (visité le 20/02/2026).
- [19] *Surgical Microscopes*. URL : <https://www.alternup-medical.com/en/109-surgical-microscopes>. (visite le 20/02/2026).
- [20] *Scintigraphie – Tomographie à émission monophotonique*. URL : <https://www.vidal.fr/sante/examens-tests-analyses-medicales/scintigraphie-tomographie-emission-monophotonique.html>. (visité le 20/02/2026).
- [21] *Elastography*. URL : <https://www.apollohospitals.com/fr/diagnostics-investigations/elastography>. (visité le 20/02/2026).
- [22] *iLIVTouch*. URL : <https://www.cdimed.com/ilivtouch>. (visité le 20/02/2026).

- [23] *Épidémiologie des maladies de la thyroïde.* URL : <https://www.elsevier.com/fr-fr/connect/epidemiologie-des-maladies-de-la-thyroide>. (visité le 08/03/2026).
- [24] H. Boukheris et al. Changes in the demographic and clinicopathological characteristics of thyroid cancer : A population-based investigation in algeria, 1993-2013. *Journal of Cancer Epidemiology*, 2020.
- [25] *Glande thyroïde : Anatomie, histologie, hormones.* URL : <https://www.kenhub.com/fr/library/anatomie/glande-thyroide>. (visité le 08/03/2026).
- [26] *Pourquoi la thyroïde se dérègle chez la femme et l'homme?* URL : <https://sante.journaldesfemmes.fr/fiches-maladies/2518726-dereglement-thyroidien-femme-homme-causes-symptomes-traitement> (visité le 08/03/2026).
- [27] *Glande thyroïde : rôle, schéma, symptômes d'un problème.* URL : <https://sante.journaldesfemmes.fr/fiches-anatomie-et-examens/2538816-glande-thyroide-role-anatomie-localisation-probleme/>. (visité le 08/03/2026).
- [28] Y. Khan et A. Farhana. Histology, thyroid gland. *StatPearls [Internet]*, 2020.
- [29] K. MEKKI et S. DJERDJOUR. « incidences comparées des différents types de goitre et propriétés des cas opérés chez un groupe des patients à la wilaya de chlef ». In : Mémoire de master, Université Hassiba Benbouali de Chlef, 2017.
- [30] *Présentation de la thyroïde.* URL : <https://www.msmanuals.com/fr/accueil/troubles-hormonaux-et-m%C3%A9taboliques/troubles-de-la-thyro%C3%AFde/pr%C3%A9sentation-de-la-thyro%C3%AFde>. (visité le 10/03/2026).
- [31] *La thyroïde en mal de prévention.* URL : <https://sante.lefigaro.fr/actualite/2010/10/31/10511-thyroide-mal-prevention>. (visité le 10/03/2026).
- [32] Bruno Baudin. Iode et hormones thyroïdiennes. *Revue Francophone des Laboratoires*, (2024). p.38-44.
- [33] A. Ryndak-Swiercz. Ontogenèse, anatomie, histologie et physiologie de la thyroïde. *EMC - Endocrinologie-Nutrition*, (2022).

- [34] *19 signes de problèmes de thyroïde.* URL : <https://www.medicoverhospitals.in/fr/articles/19-signs-of-thyroid-problems>. (visité le 09/03/2026).
- [35] *Bilan Thyroïdien : TSH, T3, T4, Anticorps - Guide Complet d'Interprétation 2025.* URL : <https://monbilansang.fr/blog/bilan-thyroidien-tsh-t3-t4-guide-complet>. (visité le 08/03/2026).
- [36] *Aspiration à l'aiguille fine (FNA) : qu'est-ce que c'est, procédure et résultats.* URL : <https://gesundmd.com/sante/aspiration-a-laiguille-fine-fna-quest-ce-que-cest-procedure-et-re> (visité le 13/03/2026).
- [37] *Maladies de la thyroïde.* URL : <https://www.hug.ch/endocrinologie-diabetologie-metabolisme/maladies-thyroide>. (visité le 13/03/2026).
- [38] *Les thyroïdites : causes, symptômes et traitements.* URL : https://www.doctissimo.fr/html/sante/encyclopedie/sa_694_thyroidites.htm. (visité le 13/03/2026).
- [39] *Nodule thyroïdien : définition, symptômes et facteurs favorisants.* URL : <https://www.ameli.fr/assure/sante/themes/nodule-thyroidien/definition-symptomes-facteurs-favorisants>. (visité le 20/04/2026).
- [40] *Goitre plongeant : à propos de 86 cas.* Thèse de médecine, Faculté de Médecine et de Pharmacie, Université Mohammed V, Rabat, 2016.
- [41] W. T. SAYAH and K. DOUADI. Profil épidémiologique des cancers de la thyroïde au niveau de l'eph mohamed boudiaf ouargla (janvier 2020 – décembre 2023). Mémoire de doctorat en médecine, Université Kasdi Merbah Ouargla, Faculté de Médecine, 2024.
- [42] *Nodules thyroïdiens : guide sur les nodules et le cancer de la thyroïde.* URL : <https://www.cliniqueorl.com/blog/thyroide/nodules-thyroidiens/>. (visité le 09/03/2026).
- [43] *Nodule de la thyroïde : comment l'explorer?* URL : https://www.has-sante.fr/jcms/p_3287556/fr/nodule-de-la-thyroide-comment-l-explorer. (visité le 09/03/2026).
- [44] *Nodule thyroïde : cancer, symptômes, femme, homme, traitement.* URL : <https://sante.journaldesfemmes.fr/fiches-maladies/>

- 2538286-nodule-thyroïde-cancer-symptomes-femme-homme-traitement/. (visité le 20/04/2026).
- [45] *Prise en charge d'un patient qui présente un nodule thyroïdien.* URL : <https://www.msdmanuals.com/fr/professional/troubles-endocriniens-et-m%C3%A9taboliques/troubles-thyro%C3%AFdiens/prise-en-charge-d-un-patient-qui-pr%C3%A9sente-un-nodule-thyro%C3%AFdien>. (visité le 24/03/2026).
- [46] *ACR Thyroid Imaging Reporting and Data System (ACR TI-RADS).* URL : <https://radiopaedia.org/articles/acr-thyroid-imaging-reporting-and-data-system-acr-ti-rads>. (visité le 14/03/2026).
- [47] *Cancer de la thyroïde.* URL : <https://www.msdmanuals.com/fr/professional/troubles-endocriniens-et-m%C3%A9taboliques/troubles-thyro%C3%AFdiens/cancers-de-la-thyro%C3%AFde>. (visité le 14/03/2026).
- [48] *Cancer de la thyroïde – Définition et traitement.* URL : <https://www.elsan.care/fr/pathologie-et-traitement/cancers/cancer-de-la-thyroïde>. (visité le 14/03/2026).
- [49] *L'irathérapie dans le cancer thyroïdien différencié.* URL : <https://www.cancerologie-pratique.com/journal/article/005965-liratherapie-cancer-thyroidien-differencie>. (visité le 20/04/2026).
- [50] *Guide ALD 30 « Cancer de la thyroïde ».* URL : https://onconormandie.fr/wpcontent/uploads/2024/01/ALD30_thyroïde_med_HAS_2010.pdf. (visité le 14/03/2026).
- [51] *Endocrinologist ultrasound doctor sonographer examine patient on ultrasound.* URL : <https://www.istockphoto.com/vector/endocrinologist-ultrasound-doctor-sonographer-examine-patient-on-ultrasound>. (visité le 14/03/2026).
- [52] *Cytoponction thyroïdienne.* URL : <https://www.imagerie-medicale-paris-est.com/examens/cytoponction-thyroidienne>. (visité le 14/03/2026).
- [53] *Intelligence artificielle.* URL : https://fr.wikipedia.org/wiki/Intelligence_artificielle. (visité le 21/02/2026).
- [54] *Intelligence artificielle : définition et utilisation.* URL : <https://www.europarl.europa.eu/topics/fr/article/20200827ST085804/>

- [intelligence-artificielle-definition-et-utilisation](#).
(visité le 21/02/2026).
- [55] « *Qu'est-ce que l'apprentissage automatique? Définition, types et exemples*». URL : <https://www.coursera.org/fr-FR/articles/what-is-machine-learning?isNewUser=true>. (visité le 21/02/2026).
- [56] *IA contre apprentissage automatique contre apprentissage profond : Comprendre les différences clés*. URL : <https://www.addingvalue.fr/post/ia-contre-apprentissage-automatique-contre-apprentissage-profond-c>
(visité le 21/02/2026).
- [57] *Apprentissage profond vs. apprentissage automatique — Quelle est la différence?* URL : <https://medium.com/levity/deep-learning-vs-machine-learning-whats-the-difference-e367803bb90>
(visité le 21/02/2026).
- [58] V.-C. Jofack Sokeng, F. K. Kouamé, H. Dibi N'da, B. Tankoano, L. Akpa You, and B. Ngounou Ngatcha. Cartographie de l'occupation de sol des hauts plateaux de l'ouest cameroun par réseaux de neurones appliqués à une image landsat 8 oli. *International Journal of Innovation and Scientific Research*, 23(2) :443–454, 2016.
- [59] *Les réseaux de neurones convolutifs*. URL : <https://www.datasciencetoday.net/index.php/fr/deep-learning/173-les-reseaux-de-neurones-convolutifs>. (visité le 21/02/2026).
- [60] *Présentation des réseaux de neurones convolutionnels (CNN)*. URL : <https://encord.com/blog/convolutional-neural-networks-explained/>. (visité le 22/02/2026).
- [61] *Réseaux convolutifs (CNN) : comment ça marche?* URL : <https://blent.ai/blog/a/cnn-comment-ca-marche>. (visité le 21/02/2026).
- [62] Ludivine Morvan. *Prédiction de la progression du myélome multiple par imagerie TEP : Adaptation des forêts de survie aléatoires et de réseaux de neurones convolutionnels*. PhD thesis, École Centrale de Nantes, 2021.
- [63] *Mieux comprendre le deep learning appliqué à la reconnaissance d'images*. URL : <https://www.devoteam.com/fr/expert-view/mieux-comprendre-le-deep-learning-applique-a-la-reconnaissance-dim>
(visité le 23/02/2026).

- [64] *Fonction d'activation*. URL : <https://deeplylearning.fr/cours-theoriques-deep-learning/fonction-dactivation/>. (visité le 23/02/2026).
- [65] *Approach to build realistic models for estimating project effort/cost in an uncertain environment : application to the software development field*. URL : https://www.researchgate.net/publication/303447451_Approach_to_build_realistic_models_for_estimating_project_effortcost_in_an_uncertain_environment_application_to_the_software_development_field. (visité le 23/02/2026).
- [66] Dong Zhonglei and Sha Nu. Recherche sur la situation actuelle et contre-mesures pour développer les talents dans le sport récréatif sous la perspective de l'intelligence artificielle. *Applied Mathematics and Nonlinear Sciences*, 2024.
- [67] Yingzi Huo, Kai Jin, Jiahong Cai, Huixuan Xiong, and Jiacheng Pang. Vision transformer (vit)-based applications in image classification. In *Proc. IEEE BigDataSecurity, HPSC, IDS Conf.*, 2023.
- [68] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words : Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*, 2021.
- [69] Hanhua Long. Hybrid design of cnn and vision transformer : A review. In *Proceedings of the 2024 7th International Conference on Computer Information Science and Artificial Intelligence (CISAI)*, pages 121–127, 2024.
- [70] Ruizhe Yao, Ning Wang, Peng Chen, Di Ma, and Xianjun Sheng. A cnn-transformer hybrid approach for an intrusion detection system in advanced metering infrastructure. *Multimedia Tools and Applications*, 2022.
- [71] Haozheng Yin, Yu Cao, Linlin Liu, Dan Chen, and Qiong Zhang. Meteorological observation research based on an improved efficientnetv2 model. *Environmental Modelling & Software*.
- [72] Zhangcai Huang, Xiaoxiao Jiang, Shaodong Huang, Sheng Qin, and Su Yang. An efficient convolutional neural network-based diagnosis system for citrus fruit diseases. *Frontiers in Genetics*.

- [73] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [74] Chuan Wang Chang, Santanu Santra, Jun Wei Hsieh, Pirdiansyah Hendri, and Chi Fang Lin. Multi-fusion feature pyramid for real-time hand detection. In : *Multimedia Tools and Applications*, (2022).
- [75] Comment le transfer learning révolutionne l'apprentissage automatique avec l'intelligence transférée? URL : <https://www.nexa.fr/post/transfer-learning-definition>. (visité 26/05/2026).
- [76] Transfer learning vs fine-tuning vs multitask learning vs federated learning. URL : <https://www.geeksforgeeks.org/machine-learning/transfer-learning-vs-fine-tuning-vs-multitask-learning-vs-federated-learning/> (visité 26/05/2026).
- [77] Qu'est-ce que le surajustement? URL : <https://www.ibm.com/fr-fr/think/topics/overfitting>. (visité le 03/03/2026).
- [78] Overfitting and Underfitting : Regularization (L1, L2), Dropout, Early Stopping and Data Augmentation. URL : <https://medium.com/@fraidoonomarzai99/overfitting-and-underfitting-regularization-l1-l2-drop-out-early-stopping-and-data-augmentation> (visité le 03/03/2026).
- [79] Qu'est-ce que la détection d'objets? URL : <https://www.ibm.com/fr-fr/think/topics/object-detection>. (visité le 21/02/2026).
- [80] S. M. Asif, T. Anil, S. Tangudu, GH. Sai Keertan, and S. Don. Traffic sign detection using ssd mobilenet & faster rcnn. In : *IEEE Xplore*, (2023).
- [81] D. Zhipeng, S. Hao, Z. Shilin, Z. Juanping, L. Lin, and Z. Huanxin. Multi-scale object detection in remote sensing imagery with convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018.
- [82] Ch. Chunling, Zh. Ziyue, X. Tongyu, G. Shuang, F. Shuai, Y. Weixiang, and L. Yubin. Yolo-based uav technology : A review of the research and its applications. In : *Drones*, (2023).
- [83] R. R. Karmarkar and V.N. Honmane. Object detection system for the blind with voiceguidance. In : *International Journal of Engineering Applied Sciences and Technology*, (2021).

- [84] P. Hidayatullah et R. Tubagus. «yolo26 : A comprehensive architecture overview and key improvements». In : *arXiv preprint arXiv :2602.14582*, (2026).
- [85] *Full Form of YOLO*. URL : <https://www.geeksforgeeks.org/computer-vision/full-form-of-yolo/>. (visit  le 23/02/2026).
- [86] *YOLOv8 Architecture : A Deep Dive into its Architecture*. URL : <https://yolov8.org/yolov8-architecture/>. (visit  le 07/03/2026).
- [87] *Ultralytics YOLO26*. URL : <https://docs.ultralytics.com/models/yolo26/>. (visit  le 07/03/2026).
- [88] *YOLOE : un mod le plus rapide pour la d tection d'objets*. URL : <https://fr.console-linux.com/?p=36380>. (visit  le 07/03/2026).
- [89] «DETR — Transformer for Object Detection And My Experience Using It». URL : <https://medium.com/@binaydhakal35/detr-transformer-for-object-detection-and-my-experience-using-it-8>. (Visit  le 24/02/2026).
- [90] *Understanding “DETR : End-to-End Object Detection with Transformers”*. URL : <https://medium.com/@shashankag14/understanding-detr-end-to-end-object-detection-with-transformers-6>. (Visit  le 24/02/2026).
- [91] *RF-DETR : Transformer-Based Object Detector*. URL : <https://www.emergentmind.com/topics/rf-detr>. (visit  le 24/02/2026).
- [92] *Real-Time Object Detection with RF-DETR and Amazon SageMaker*. URL : <https://medium.com/data-science-collective/real-time-object-detection-with-rf-detr-and-amazon-sagemaker-0263>. (visit  le 24/02/2026).
- [93] *Getting started with Classification*. URL : <https://www.geeksforgeeks.org/machine-learning/getting-started-with-classification/>. (visit  le 02/06/2026).
- [94] Yi Tao, Yanyan Yu, Tong Wu, Xiangli Xu, Quan Dai, Hanqing Kong, Lei Zhang, Weidong Yu, Xiaoping Leng, Weibao Qiu, and Jiawei Tian. Deep learning for the diagnosis of suspicious thyroid nodules based on multi-modal ultrasound images. *Frontiers in Oncology*, (2022). p. 1012724.
- [95] Ching-Wei Wang, Kuan-Yu Lin, Yi-Jia Lin, Muhammad-Adil Khalil, Kai-Lin Chu, and Tai-Kuang Chao. A soft label deep learning to assist breast cancer target therapy and thyroid cancer diagnosis. *Cancers*, (2022). p. 5312.

- [96] Li Zhang, Xinyi Wang, Qian Chen, Jing Liu, Yifan Huang, and Ming Li. Diagnosis of thyroid disease using deep convolutional neural network models applied to thyroid scintigraphy images : a multicenter study. *Frontiers in Endocrinology*, (2023). p. 1224191.
- [97] Truong Thanh Hung Nguyen, Van Binh Truong, Vo Thanh Khang Nguyen, Quoc Hung Cao, and Quoc Khanh Nguyen. Towards trust of explainable ai in thyroid nodule diagnosis. *arXiv preprint arXiv :2303.04731*, (2023).
- [98] Nadeen Amgad, Hadiy Haitham, Mona Alabrak, and Ammar Mohamed. Enhancing thyroid cancer diagnosis through a resilient deep learning ensemble approach. *IEEE*, (2024).
- [99] P Archana, Suhana Anjum S, Keerthana S M, and Srusti P D. Detection of thyroid diseases using deep learning. (2024).
- [100] Wen Wen, Tingrui Zhang, Haina Zhao, Jingyan Liu, Heng Jiang, Yushuang He, and Zekun Jiang. Multimodal model enhances qualitative diagnosis of hypervascular thyroid nodules : integrating radiomics and deep learning features based on b-mode and pdi images. *Gland Surgery*, (2025).
- [101] A. A. Hanani, T. B. Donmez, M. Kutlu, and M. Mansour. Predicting thyroid cancer recurrence using supervised catboost : A shap-based explainable ai approach. *Medicine*, 2025. p. e42667.
- [102] Aya Rashed, T. Medhat, and Ahmed Elgarayhi. Enhancing automatic diagnosis of thyroid nodules from ultrasound scans leveraging deep learning models. *Scientific Reports*, (2025). p. 40364.
- [103] Tathagat Banerjee, Davinder Paul Singh, Debabrata Swain, Shubham Mahajan, Seifedine Kadry, and Jungeun Kim. A novel hybrid deep learning approach combining deep feature attention and statistical validation for enhanced thyroid ultrasound segmentation. *Scientific Reports*, 2025. p. 27207.
- [104] Xiaoxue Wang, Yupeng Niu, Hongli Liu, Fa Tian, Qiang Zhang, Yimeng Wang, Yeju Wang, and Yijia Li. Thyronet-x4 genesis : an advanced deep learning model for auxiliary diagnosis of thyroid nodules' malignancy. *Scientific Reports*, (2025). p. 4214.
- [105] Sanaz Vahdati, Bardia Khosravi, Kathryn A. Robinson, Pouria Rouzrokh, Mana Moassefi, Zeynettin Akkus, and Bradley J. Erickson. A multi-view deep learning model for thyroid nodules detection and characterization in ultrasound imaging. *Scientific Reports*, 2024.

- [106] *Qu'est-ce que Python?* URL : <https://aws.amazon.com/fr/what-is/python/>. (visité le 12/04/2026).
- [107] *Google Colab : le guide complet pour coder en Python dans le cloud.* URL : <https://www.jeuxetredatascientist.fr/google-collab-tout-savoir/>. (visité le 12/04/2026).
- [108] *Lightning AI.* URL : <https://aitools.aiting.com/fr/ai/lightning>. (visité le 12/04/2026).
- [109] *The ultimate end-to-end computer vision platform.* URL : <https://www.ultralytics.com/>. (visité le 12/04/2026).
- [110] *Roboflow – Outils IA.* URL : <https://aikiwi.com/outils-ia/roboflow/>. (visité le 12/04/2026).
- [111] *TensorFlow : définition, fonctionnement et cas d'usage.* URL : https://www.lebigdata.fr/tensorflow-definition-tout-savoir#google_vignette. (visité le 12/04/2026).
- [112] *PyTorch.* URL : <https://pytorch.org/projects/pytorch/>. (visité le 12/04/2026).
- [113] *About OpenCV.* URL : <https://opencv.org/about/>. (visité le 12/04/2026).
- [114] *Matplotlib : tout savoir sur la bibliothèque de visualisation Python.* URL : <https://liora.io/matplotlib-tout-savoir>. (visité le 12/04/2026).
- [115] *Introduction à la librairie Matplotlib en Python.* URL : <https://blog-mes-outils-automatiques.fr/introduction-librairie-matplotlib-python/>. (visité le 12/04/2026).
- [116] *NumPy : The Absolute Basics for Beginners.* URL : https://numpy.org/doc/stable/user/absolute_beginners.html. (visité le 12/04/2026).
- [117] *Qu'est-ce que Scikit-learn en machine learning?* URL : <https://www.nexa.fr/blog/quest-ce-que-scikit-learn-en-machine-learning>. (visité le 12/04/2026).
- [118] *seaborn : statistical data visualization.* URL : <https://seaborn.pydata.org/>. (visité le 14/04/2026).
- [119] *ThyroidXL Dataset.* URL : <https://huggingface.co/datasets/hungl007/ThyroidXL>. (visité le 25/04/2026).

- [120] *Hugging Face*. URL : <https://huggingface.co/>. (visité le 25/04/2026).
- [121] I. Robinson et P. Robicheaux et M. Popov et D. Ramanan et N. Peri. « rf-detr : Neural architecture search for real-time detection transformers ». In : *International Conference on Learning Representations (ICLR)*, (2026).
- [122] Mazouzi wissam et Kabene Yassamine. « improving object detection in medical application ». In : Mémoire de master, Université Hassiba Benbouali de Chlef, (2025).
- [123] A. Wang et L. Liu et H. Chen et Z. Lin et J. Han et G. Ding. « yoloe :real-time seeing anything ». In : *arXiv preprint arXiv :2503.07465*, (2025).