

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique  
Université M'hamed Bougara - Boumerdès



Faculté des Sciences  
Département d'Informatique

**Domaine** : Mathématiques Informatique  
**Filière** : Informatique  
**Spécialité** : Ingénierie du logiciel et traitement de l'information

**Mémoire de fin d'études en vu de l'obtention du  
Diplôme de Master Académique**

**Thème**

Application de l'apprentissage profond par transfert en imagerie médicale pour l'aide au diagnostic des lésions thyroïdiennes

**Présenté par :**  
CHAOUCHI Lynda  
DJANI Djouher

**Soutenu le /09/2021 Devant le jury composé de**

Pr. BERRICHI Ali : Président  
Dr. SALHI Dhai Eddine : Examinateur  
Pr. GACEB Djamel : Encadreur  
Dr. TOUAZI Fayçal : Co-Encadreur

## **Remerciements**

Nous remercions l'Eternel Dieu pour le souffle de vie et toutes les capacités physiques et intellectuelles qui nous a donné pour réaliser ce modeste travail.

Nous adressons nos plus sincères remerciements à tous ceux qui nous ont fait progresser dans la réflexion et l'interprétation de ce modeste travail. Plus précisément, nous tenons à remercier Pr. GACEB Djamel ainsi que Dr. TOUAZI Fayçal pour nous avoir initié dans un domaine d'actualité, pour leur confiance et orientation durant ce projet.

Merci pour vos conseils et vos remarques. Merci pour votre engagement et votre dévouement envers les étudiants et pour l'avancement de la recherche.

Nous souhaitons également remercier nos familles et plus particulièrement nos parents pour leur soutien, leur amour et leurs encouragements continus, qui nous ont permis de réaliser les études pour lesquelles nous étions destinées et par conséquent ce mémoire.

À tous, nous présentons nos sincères remerciements et notre profonde gratitude.

## Résumé

La thyroïde fait partie du système endocrinien. C'est une petite glande située à l'avant du cou près de la trachée. Elle se compose de deux feuilles gauche et droite. La glande thyroïdienne est principalement composée de cellules C et de cellules folliculaires, ces dernières produisent des hormones thyroïdiennes. Ces cellules changent parfois, les obligeant à se développer ou à se comporter anormalement. Ce qui entraîne des affections non cancéreuses ou bénignes telles que l'hyperthyroïdie, les nodules thyroïdiens, la thyroïdite et le goitre, contrairement à certains cas, où les cellules modifiées de la thyroïde peuvent devenir cancéreuses. Avec les récents progrès de l'apprentissage profond, les performances de la classification des images médicales ont été considérablement améliorées. Dans ce travail, nous visons à évaluer et comparer les performances plusieurs architectures CNN pour classer les nodules thyroïdiens en deux classes (malin/bénin) sur des images échographiques. A cette fin, nous avons utilisé une base d'images d'une petite taille, composée de 216 cas de lésions bénignes et 427 cas de lésions malignes. Afin de compenser le manque de volume de cette base, le transfert d'apprentissage est appliqué d'une manière progressif sur trois modèles CNN existants (VGG-16, ResNet-50 et EfficientNet). Ces derniers sont pré-entraînés, en premier lieu, sur la base d'images ImageNet (disponible en grande taille mais de nature très différente de notre base d'images), et en second lieu, sur une base d'images radiographiques qui concerne une autre maladie (disponible en taille moyenne avec une nature proche à notre base). Ces modèles sont affinés selon différents niveaux et stratégies de fine-tuning. Pour la classification des lésions en maligne ou bénigne, nous avons utilisé un classifieur personnalisé supervisé softmax sur les trois modèles, à l'exception du modèle VGG-16, où nous avons testé deux autres classificateurs supplémentaires, le SVM et le Random Forest.

Le transfert d'apprentissage progressif en utilisant une base d'images radiologiques comme base intermédiaire a amélioré les performances de diagnostic des nodules thyroïdiens (sur des images échographiques) pour les architectures VGG-16+Softmax et Resnet-50V2, alors que sur les autres architectures, leurs performances ont été proches des résultats du transfert d'apprentissage effectué directement sur un modèle pré-entraîné sur la base ImageNet. Les meilleurs résultats ont été obtenus en utilisant le modèle VGG-16 + RF pré-entraîné directement sur la base d'images ImageNet avec un bon niveau de fine tuning (score F1=87.20%, précision = 82.24%, spécificité = 62.26%, sensibilité = 92.93%).

Ces résultats montrent que le bon choix de l'architecture et du niveau de fine tuning en combinaison avec un classifieur puissant tel que le RF sur une petite base d'images échographiques offre une meilleure adaptation et spécialisation de modèle pré-entraîné, même en prenant comme départ une base très différente telle que la base d'image ImageNet (destinée à l'origine pour la classification de 1000 types d'objets). Ceci nous épargne de l'usage d'une base intermédiaire de taille moyenne plus proche à la base utilisée dans notre cadre d'application. Ceci montre également la possibilité de passer facilement de la classification d'une maladie vers une autre qui dispose d'un nombre d'images limité en profitant des connaissances déjà acquises sur une autre base très large.

## Liste des figures

Figure 1.1 : Glande thyroïde .....	4
Figure 1.2 : Thyroïde Normale et Cancer de la Thyroïde .....	5
Figure 1.3 : Nodule nettement hypo-échogène aux contours irréguliers avec une adénopathie hypo-échogène : TI-RADS 5 Confirmation histologique de carcinome papillaire.....	7
Figure 1.4: (a) <b>Lobe gauche</b> : Un nodule iso échogène aux limites nettes TI-RADS 3 (histologie bénigne)	
(b) <b>Lobe droit</b> : Un nodule nettement hypo-échogène aux contours nets sans adénopathie TI-RADS 4B (histologie carcinome papillaire). ....	7
Figure 1.5 : Intelligence artificielle, apprentissage automatique et profond. ....	8
Figure1. 6 : Principe des SVM linéaire dans un espace 2D. ....	10
Figure 1.7:Principe du SVM dans un espace 3D. ....	10
Figure 1.8: Réseau neuronal artificiel Feed-forward.....	12
Figure 1.9: Neurone artificiel.....	13
Figure 1.10: Schéma d'une convolution avec un noyau de taille 3*3 et un pas de 2 [48].....	16
Figure 1.11 : Structure du CNN. ....	17
Figure 1.12 : Structure de GoogleNet. ....	21
Figure 1.13 : Architecture VGG-16. ....	22
Figure 1.14 : Architecture du réseau de base EfficientNet-B0.....	22
Figure 3. 1 : Organigramme du modèle de prédiction des nodules thyroïdiens VGG-16 basé sur les différents niveaux de Fine-tuning. ....	40
Figure 3. 2 : Organigramme du modèle de prédiction des nodules thyroïdiens ResNet-50V2 basé sur les différents niveaux de Fine-tuning. ....	41
Figure 3. 3 : Organigramme du modèle de prédiction des nodules thyroïdiens EfficientNet basé sur les différents niveaux de Fine-tuning. ....	42
Figure 3. 4 : Exemples d'images échographiques de la base de données DDTI, Ultrasoundcases.info et thyroid.xiaobaishi.com. ....	43
Figure 3. 5: Exemples d'images transformées par les méthodes d'augmentation de données.	44
Figure 3. 6 : Exemples des images radiographiques thoraciques COVID-19.....	45
Figure 4. 1 : Logo TensorFlow.....	54
Figure 4. 2 : Logo Keras.....	55
Figure 4. 3 : Logos Google Colab et Jupyter Notebook.....	55

Figure 4. 4 : Logos Gradio.....	56
Figure 4. 5 : Les courbes ROC des cinq modèles avec le Transfert d'apprentissage sur la base d'images ImageNet.....	60
Figure 4. 6 : Les courbes ROC des cinq modèles avec le Transfert d'apprentissage progressif sur la base d'images radiographique.....	62
Figure 4. 7 :Les courbes de précision et de perte de modèle M4a (EfficientnetB0-MLP) dans le cas de l'entraînement des deux derniers blocs de convolution dans le scénario Transfert d'apprentissage sur la base ImageNet .....	64
En comparant les meilleures valeurs de score F1 du tableau 4.1 ( M1a=87.20% , M2a=84 .65%, M3a=69,37%, M4a=70,54% et M5a=74,58% ) et pour éviter tous faux négatif, on peut clairement dire que M1a (VGG16+RF) suivi par M2a(VGG16+SVM) formés avec des échantillons des trois bases d'images fusionnées, non augmentées et affinées avec la 3 <sup>ème</sup> stratégie en en gelant toute la base de convolution de l'architecture sont les meilleurs modèles et les mieux adaptés à notre cas, avec une valeur AUC ou ROC (figure 4. 8) de 0.85 et 0.83 respectivement. En revanche, la valeur de spécificité (62,26% VS 52,83%), de sensibilité (92,93% VS 91 ,92%) et d'AUC de M1a(VGG16+RF) sont élevées à celles de M2a (VGG16+SVM) ce qui nous permet de le sélectionner comme le meilleur modèle dans le cas de ce scénario. ....	64
Figure 4. 9 : les courbes de précision et de perte de modèle M4b(EfficientnetB0-MLP) dans le cas de l'entraînement du dernier bloc de convolution dans le scénario Transfert d'apprentissage progressif sur la base d'images radiographiques. ....	65
Figure 4. 10 : Courbes de précision et de perte pour le modèle Resnet-50V2.....	66

## Liste des tableaux

Tableau 1.1 : Classification EU-TIRADS [26] .....	6
Tableau 2. 1 : Tableau comparatif des modèles proposé par Sundar et al. [31].....	32
Tableau 3. 1: Répartition des images dans les ensembles de données d'entraînement, validation et de test.....	43
Tableau 3. 2 : La répartition des cas avec et sans augmentation dans les groupes de formation et de validation des trois bases de données. ....	44
Tableau 3. 3 : Matrice de confusion ..... 49	
Tableau 4. 1 : Performance des modèles VGG-16, EfficientnetB0 et Resnet-50V2 sur les données de validation des trois bases fusionnées sans l'augmentation.....	59
Tableau 4. 2 : Performance des modèles sur les données de validation des trois bases fusionnées avec l'augmentation. ....	59
Tableau 4. 3 : Performance des modèles VGG-16, EfficientnetB0 et Resnet-50V2 sur les données de validation des trois BASES fusionnées sans l'augmentation. ....	61
Tableau 4. 4: Performance des modèles sur les données de validation avec l'augmentation. .	61

## **LISTE DES ABRÉVIATIONS**

US	Ultra Son
EU-TIRADS	European Thyroid Imaging-Reporting and Data System
SVM	Support Vector Machine
RF	Random Forest
CNN	Convolutional Neural Network
ANN	Artificial Neural Network
ReLU	Rectified Linear Units
FC	Fully-Connected
TDM	Tomodensitométrie
IRM	Imagerie par Résonance Magnétique
IA	Intelligence Artificielle
FNA	Fine Needle Aspiration
DAO	Diagnostic Assisté par Ordinateur
DL	Deep Learning
ML	Machine Learning
RBF	Radial Basis Function
CAEs	Convolutional Auto-Encoders et spatiales
LBP	Local Binary Pattern
HOG	Histogram of Oriented Gradients
CAD	Computer-Aided Diagnosis
AUC	Area under the ROC Curve
ROC	Receiver Operating Characteristic
DDTI	Digital Database Thyroid Image
CXR	Chest X-Ray

## Table des matières

Introduction générale.....	1
Chapitre 1 : Notion de base sur la maladie de la thyroïde et l'apprentissage automatique .....	4
Introduction .....	4
1.1 Nodules thyroïdiens et cancer de la thyroïde .....	4
1.2 Imagerie médicale .....	5
1.3 Apprentissage automatique .....	8
1.3.1 Machine à Vecteurs de Support.....	9
1.3.2 Arbre de décision.....	11
1.3.3 Forêts aléatoires .....	11
1.3.4 .....	12
Réseaux de neurones artificiels .....	12
1.3.5 Limites de l'apprentissage automatique .....	14
1.4 Apprentissage profond (Deep Learning).....	15
1.4.1 Réseaux de neurones convolutifs.....	15
1.4.2 Apprentissage par transfert (Transfer Learning).....	18
1.4.3 Fine Tuning : Adaptation des CNN pré-entraînés .....	18
1.5 Architectures CNN existantes .....	20
1.5.1 ResNet.....	20
1.5.2 Inception .....	20
1.5.3 VGG16.....	21
1.5.4 EfficientNet.....	22
1.6 Apprentissage profond et imagerie médicale .....	23
1.6.1 Classification.....	23
1.6.1.1 Classification des images.....	23
1.6.1.2 Classification des objets ou des lésions .....	23
1.6.2 Détection .....	24
1.6.2.1 Détection d'objets ou de lésions.....	24
1.7 Conclusion.....	25
Chapitre 2 : Applications de l'apprentissage automatique ou profond au diagnostic médical .	26
2.1 Cas général .....	26
2.1.1 Application de l'apprentissage automatique dans le diagnostic médical.....	26
2.1.2 Application de l'apprentissage profond dans le diagnostic médical.....	27
2.2 Cas particulier .....	28

2.2.1	Application de l'apprentissage automatique dans le diagnostic des nodules thyroïdiens .....	28
2.2.2	Application de l'apprentissage profond dans le diagnostic des nodules thyroïdiens 29	
2.3	Quelques travaux de classification des nodules thyroïdiens .....	33
2.3.1	Diverses méthodologies de classification basées sur les caractéristiques du goulot d'étranglement du VGG16 et Inception-v3 : .....	33
2.3.2	Classification des images échographiques de la thyroïde basée sur le modèle affiné GoogLeNet .....	34
2.4	Comparatif des approches existantes de classification binaire des nodules thyroïdiens 35	
2.5	Conclusion.....	36
	Chapitre 3 : Contribution.....	37
3.1	Méthodologie .....	37
3.2	Approche proposée.....	38
3.3	Bases d'images utilisées .....	42
3.4	Augmentation de données .....	43
3.5	Préparation des images de la base .....	44
3.5.1	Collecte de données pour le Transfert progressif d'apprentissage.....	44
3.5.1.1	Base de données sur la radiographie thoracique COVID19 .....	44
3.6	Sélection des hyper-paramètres de l'approche proposée .....	45
3.6.1	Paramétrage de classifieur SVM.....	46
3.6.2	Hyper-paramètres RF .....	46
3.7	Construction du modèle profond .....	47
3.7.1	L'architecture VGG-16 / ResNet-50 /EfficientNet +classifieur personnalisé supervisé Softmax.....	47
3.7.2	L'architecture VGG-16+le classifieur (Random Forest / SVM) .....	48
3.8	Métriques d'évaluation de la performance de la classification.....	49
3.8.1	Matrice de confusion.....	49
3.8.2	Accuracy .....	50
3.8.3	Précision.....	50
3.8.4	Rappel .....	50
3.8.5	Sensibilité.....	51
3.8.6	Spécificité .....	51
3.8.7	F-mesure (F-score ou F1-score).....	51
	Conclusion.....	51

Chapitre 4 : Expérimentations, résultats et discussion .....	53
4.1    Outils utilisés.....	54
4.1.1    Plateforme TensorFlow.....	54
4.1.2    Environnements virtuels dans Jupyter Notebook.....	55
4.1.3    Gradio : Interfaces graphiques pour les modèles de ML et DL .....	56
4.2    Expérimentations et résultats .....	56
4.2.1    Expérimentation des modèles CNN développés.....	57
4.2.1.1    Expérience 1 : élaboration du plusieurs variantes de modèle VGG16 avec les classificateurs SVM, RF et Softmax.....	57
4.2.1.2    Expérience 2 : élaboration du modèle EfficientnetB0+Softmax .....	57
4.2.1.3    Expérience 3 : élaboration du modèle Resnet-50 + Softmax .....	58
4.2.2    Transfert d'apprentissage sur la base d'images ImageNet .....	58
4.2.3    Transfert d'apprentissage progressif sur l'ensemble de données radiographique ..	60
4.3    Discussion et conclusion .....	62
Conclusion générale .....	69
Références bibliographiques .....	71

## Introduction générale

La médecine est une science qui a pour but de guérir et de préserver le fonctionnement normal du corps humain, des organes qui le constituent ainsi que de la santé physique et psychique des individus, et cela grâce à la bonne pratique du diagnostic, du traitement et de la prévention contre les maladies.

En 1895 avec l'invention des rayons X, l'imagerie médicale a vu le jour. Cette découverte a permis un meilleur diagnostic, et a offert un nouvel espoir de traitement pour plusieurs maladies. L'analyse d'images médicales est la science de l'analyse ou de la résolution de problèmes médicaux à l'aide de différentes techniques permettant une extraction efficace de l'information médicale. Elle est devenue l'un des principaux domaines de recherche de l'ingénierie et de la médecine.

A cause de la complexité des données d'imagerie médicale, il est généralement difficile de déduire des solutions analytiques ou des méthodes simples pour décrire et représenter des objets tels que les lésions et les anatomies. De ce fait, les tâches en imagerie médicale nécessitent un apprentissage à partir d'exemples. Cela constitue l'un des intérêts clés du domaine de l'apprentissage automatique en général et du son sous domaine de l'apprentissage profond en particulier.

Aujourd'hui, on utilise des algorithmes d'apprentissage automatique pour améliorer les performances du système d'analyse d'images médicales et de diagnostic assisté par ordinateur. Contrairement aux méthodes conventionnelles utilisant un apprentissage automatique basé sur un choix manuel de caractéristiques d'images (une tâche très complexe, réservée le plus souvent aux experts du domaine), les approches basées sur l'apprentissage profond offrent une caractérisation automatique plus pertinente et sont capables d'apprendre en développant de larges connaissances sur une base d'images de grand volume. La possibilité de transférer l'apprentissage et de faire un apprentissage incrémental offre un grand avantage à ces approches.

Les méthodes d'apprentissage profond utilisent des réseaux de neurones profond à convolution (CNN). Elles ont été appliquées en analyse d'images médicales avec succès donnant des résultats prometteurs. Le domaine d'application couvre tout le spectre de l'analyse d'images médicales, y compris la détection, la segmentation, la classification et le diagnostic assisté par ordinateur.

Les nodules thyroïdiens font partie des lésions les plus fréquentes en Algérie et dans le monde entier, qui affectent le bon fonctionnement de la glande thyroïdienne. Dans la pratique

clinique, leur diagnostic constitue un problème de base, car malgré qu'ils sont hétérogènes et comportent des composants distincts, leur apparition présente un aspect très similaire à l'arrière-plan dans des images échographiques.

Pour y remédier, un système dédié à la génération de rapports et de données d'imagerie thyroïdienne TI-RADS (Thyroid Imaging Reporting and Data System) a été développé pour classer les nodules thyroïdiens en fonction de leurs probabilités de malignité. Mais sa précision est souvent basée sur l'expérience personnelle des radiologues. Une discrimination automatique ou semi-automatique des nodules bénins et malins est donc nécessaire pour une classification précise des nodules thyroïdiens : Les systèmes (semi)-automatiques d'aide à la discrimination sont les outils de Diagnostic Assisté par Ordinateur.

C'est dans cette optique, que cette étude a été proposée où nous comparons des modèles CNN d'apprentissage profond par transfert pour la classification des images échographiques de la thyroïde sur une base d'images de taille relativement petite. Pour cela, différents modèles CNN pré-entraînés sont testés, plus particulièrement les réseaux VGG-16, ResNet-50V2 et EfficientNetB0. Ces modèles sont élaborés selon différents niveaux de réglage fin (fine-tuning) et validés à l'aide d'une base d'images échographiques composée de deux types d'images thyroïdiennes : nodules thyroïdiens malins et nodules thyroïdiens bénins.

Nous étudions également la capacité de transfert des connaissances et des caractéristiques apprises dans un premier lieu sur la base d'images ImageNet contenant plus de 14 millions images non médicales aux images échographiques de la thyroïde constituant un ensemble d'images de taille relativement petite. De ce fait, nous avons choisi d'explorer différents niveaux transfert d'apprentissage qui peuvent être définis comme un transfert progressif constitué d'un transfert des caractéristiques apprises sur la base ImageNet aux images radiographiques thoraciques pour les cas positifs au COVID-19 ainsi que des images de pneumonie normale et virale composant 20 000 images, puis un transfert des caractéristiques apprises sur cette dernière base de COVID-19 aux images échographiques de la thyroïde. Ce type de transfert peut être utile lorsque les réseaux de neurones profonds sont pré-entraînés sur des ensembles de données très différents du domaine en question et lorsque nous disposons d'une base d'images de taille faible.

Nous évaluons et comparons les performances de différentes architectures d'apprentissage profond pour différencier les nodules malins des nodules bénins et menons des expériences sur notre base d'images en procédant à certaines expérimentations afin de déterminer la meilleure façon de réaliser cette classification.

Ce mémoire est organisé en quatre chapitres :

- **Dans le premier chapitre**, nous présenterons quelques notions de base sur la thyroïde, l'apprentissage automatique et l'apprentissage profond et leur application en reconnaissance de formes et classification d'images.
- **Dans le deuxième chapitre**, nous expliquerons les approches classiques et celles qui sont basées sur les CNNs pour le diagnostic médical et en particulier pour la classification des nodules thyroïdiens.
- **Dans le troisième chapitre**, nous explorons diverses méthodologies de classification basée sur l'apprentissage profond plus particulièrement sur les réseaux VGG-16, ResNet-50 et EfficientNet pré-entraînés et affinés selon les différentes stratégies de réglage fin à l'aide d'une base d'images échographiques de la thyroïde.
- **Et dans le quatrième chapitre**, nous analyserons et discuterons les résultats d'évaluation obtenus de la classification d'une base d'images échographiques de la thyroïde, en évaluant et comparant les performances de nos modèles d'apprentissage profond pour différencier les nodules malins des nodules bénins.

# **Chapitre 1 : Notion de base sur la maladie de la thyroïde et l'apprentissage automatique**

## **Introduction**

Nous présentons dans ce chapitre divers concepts sur la maladie de la thyroïde ainsi que l'apprentissage automatique et l'apprentissage profond ayant une relation avec notre étude.

### **1.1 Nodules thyroïdiens et cancer de la thyroïde**

La glande thyroïde est située dans la partie inférieure et antérieure du cou, devant la trachée avec deux lobes latéraux réunis par un isthme sous forme d'un papillon. (Figure 1.1)

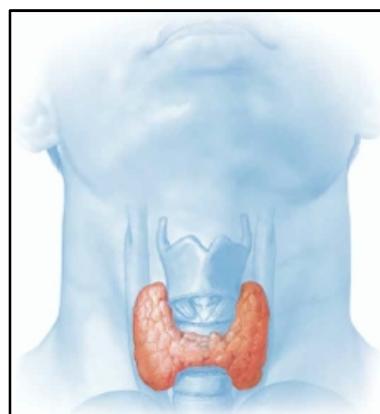


Figure 1.1 : Glande thyroïde

La thyroïde produit deux principales hormones, la tétraïodothyronine “thyroxine” (T4) et la triiodothyronine (T3). Ces hormones contrôlent le métabolisme humain et agissent ainsi sur de nombreux systèmes (muscles, rythme cardiaque et système nerveux).

Les nodules thyroïdiens sont des masses situées dans la glande thyroïde, dans la majorité des cas ils sont bénins.

Les cancers de la thyroïde entraînent généralement peu de symptômes. Ils se manifestent principalement par un nodule, le plus souvent découvert à la palpation du cou ou lors d'examens d'imagerie.

Les nodules thyroïdiens sont indolores et non gênants. Ils peuvent avoir pour causes des tumeurs bénignes (kystes liquides ou adénomes) ou tumeurs cancéreuses (cancer papillaire ou cancer folliculaire).

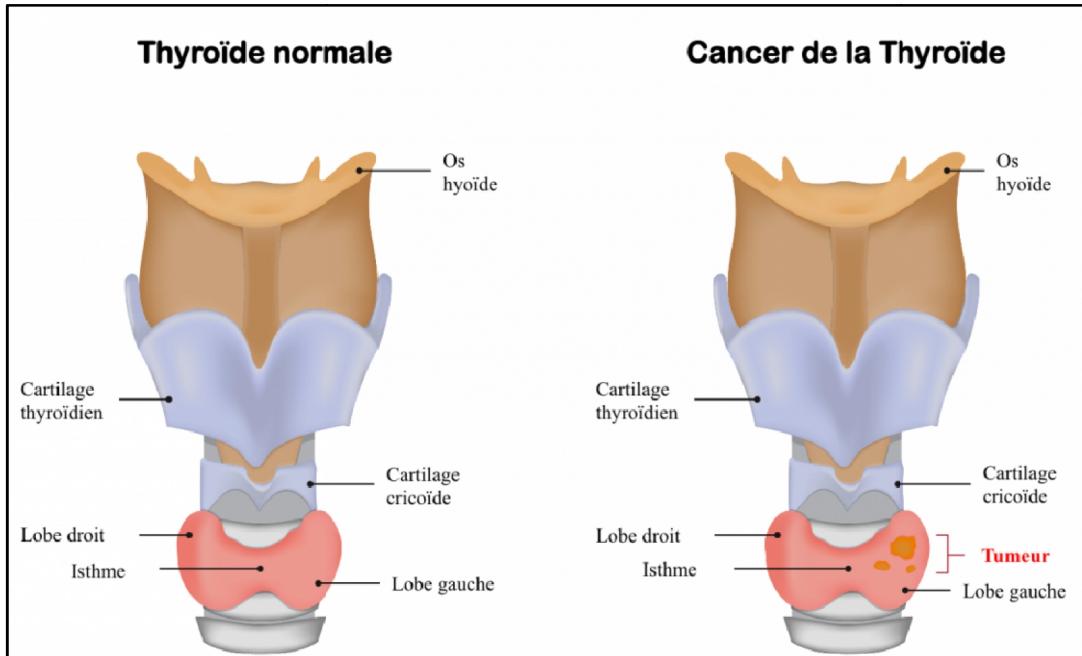


Figure 1.2 : Thyroïde Normale et Cancer de la Thyroïde

## 1.2 Imagerie médicale

L'échographie (US) est devenue un outil important dans l'évaluation diagnostique de la pathologie thyroïdienne et en particulier des nodules thyroïdiens. Elle permet d'explorer la structure tissulaire et l'échogénicité de la thyroïde (hypoéchogène, isoéchogène ou hyperéchogène).

L'association européenne de la thyroïde a mis en place une méthode de classification EU-TIRADS acronyme pour European Thyroid Imaging-Reporting and Data System des nodules thyroïdiens. Cette méthode inclue une stratification du risque de malignité des nodules thyroïdiens qui repose sur un score présentant 5 grades (Tableau 1.1). Ceci facilitera le processus décisionnel concernant la nécessité d'effectuer une surveillance ou une cytoponction aspiration à l'aiguille fine dans les cas suspects/non équivoques [14].

TIRADS	SIGNIFICATION	Risque de malignité	ASPECT	CAT
TIRADS 1	Examen normal	/	/	Pas de surveillance
TIRADS 2	Lésion bénigne	0%	Kyste simple. Nodules spongiformes. White knight. Macro calcification isolée. Thyroïdite subaiguë. Amas hypoéchogènes confluents.	Surveillance ECHO
TIRADS 3	Lésion très probablement bénigne	0.25%	zéro signe de forte suspicion Isoéchogène Hyperéchogène	Surveillance ECHO Ponction : > 20 mm ou progression
TIRADS 4A	Faible suspicion de malignité	6%	zéro signe de forte suspicion Modérément hypoéchogène	Ponction
TIRADS 4B	Forte suspicion de malignité	69%	1 ou 2 signes de forte suspicion. Pas d'adénopathie.	Ponction
TIRADS 5	Carcinome quasi certain	100%	3 à 5 signes et/ou adénopathie.	Ponction

Tableau 1.1 : Classification EU-TIRADS [26]

Les cinq (05) scores comptés sur la base de nombreuses normes, reflètent le risque de malignité nodulaire. La description du nodule est basée sur l'analyse de sa forme, de son contour et de son contenu.

Dans un premier temps, il s'agit de rechercher des signes de forte suspicion de malignité tels que les contours irréguliers, les micro-calcifications, une forte hypoéchogénicité et des mensurations caractéristiques (épaisseur supérieure à la largeur).

Les scores 2 et 3 sont attribués aux nodules ne présentant aucun de ces signes : Ces nodules sont considérés comme bénins.

Les scores 4A, 4B et 5 sont attribués aux nodules revêtant un aspect suspect. Lorsqu'un nodule présente trois à cinq des signes de forte suspicion et/ou il est accompagné d'adénopathie, il sera classé TI-RADS 5 (le score le plus élevé) (Figures 1.3 et 1.4) [45].

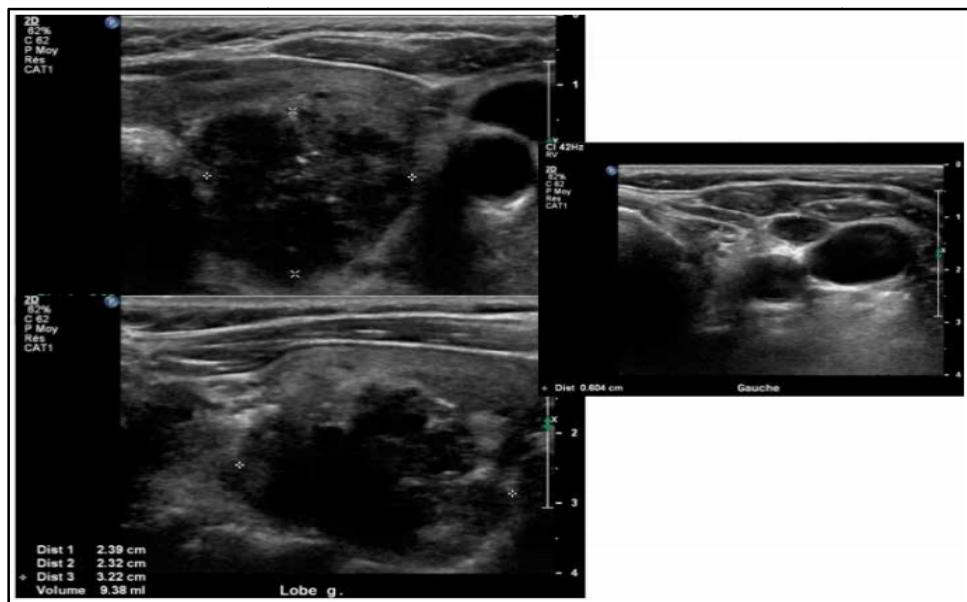


Figure 1.3 : Nodule nettement hypo-échogène aux contours irréguliers avec une adénopathie hypo-échogène : TI-RADS 5 Confirmation histologique de carcinome papillaire.

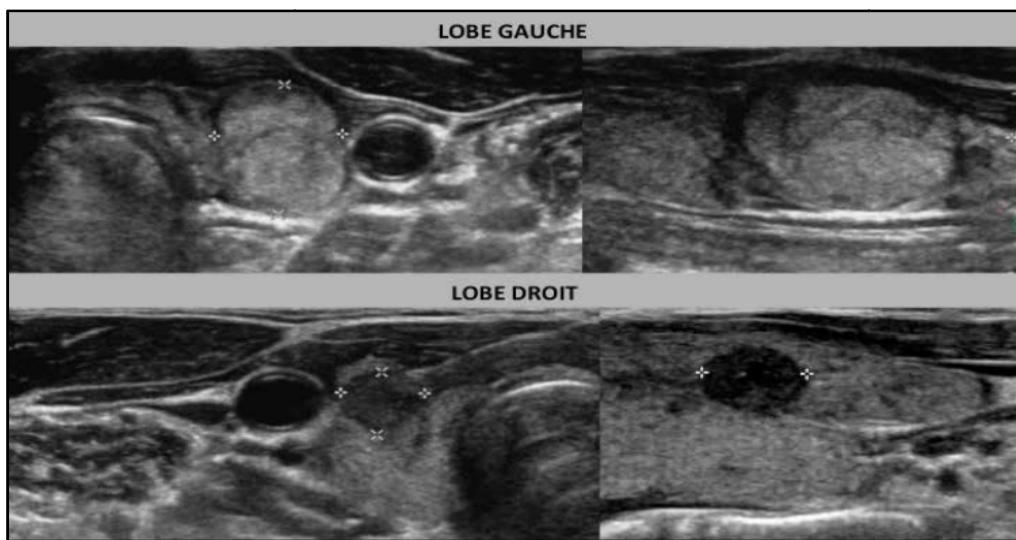


Figure 1.4: (a) **Lobe gauche** : Un nodule iso échogène aux limites nettes TI-RADS 3 (histologie bénigne)

(b) **Lobe droit** : Un nodule nettement hypo-échogène aux contours nets sans adénopathie TI-RADS 4B (histologie carcinome papillaire).

L'immédiateté, le faible coût et l'absence d'effets secondaires forment initialement l'avantage majeur de l'imagerie médicale par ultrasons. Celle-ci est devenue ainsi une méthode d'imagerie irremplaçable pour la détection des nodules thyroïdiens bénins et malins.

### 1.3 Apprentissage automatique

L'apprentissage automatique (Machine Learning), peut être défini comme une branche de l'intelligence artificielle qui a émergé dans la deuxième partie du XXe siècle. Il consiste à laisser des machines découvrir automatiquement des ‘patterns’ à savoir ‘des motifs récurrents’ dans un ensemble de données. Il s'agit d'algorithmes qui analysent un ensemble de données afin de déduire des règles (une étape dite d'entraînement) constituant des connaissances permettant d'analyser de nouvelles situations. De grands ensembles de données (représentatifs) sont nécessaires pour l'entraînement. L'apprentissage profond (deep learning) est une branche de l'apprentissage automatique (Figure 1.5).

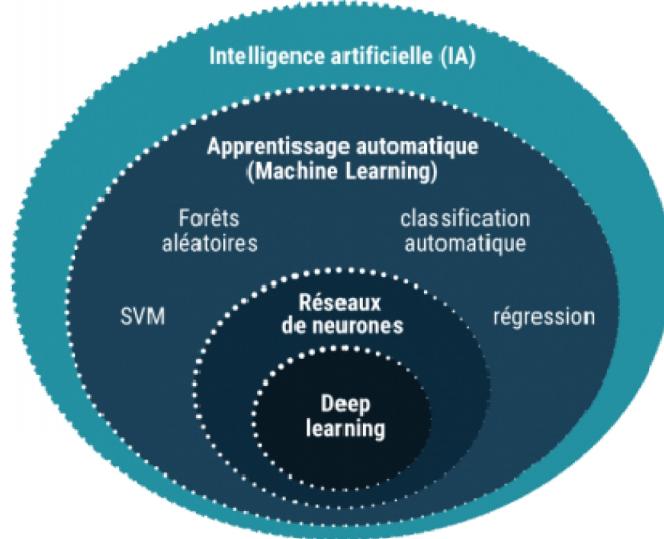


Figure 1.5 : Intelligence artificielle, apprentissage automatique et profond.

Les algorithmes d'apprentissage peuvent se catégoriser selon le mode d'apprentissage qu'ils emploient :

- **Apprentissage non supervisé** : Elle traite les données non-étiquetées et l'algorithme doit découvrir par lui-même la structure plus ou moins *cachée* des données. Le clustering est un algorithme d'apprentissage non supervisé. Exemple : Un épidémiologiste pourrait par exemple dans un ensemble assez large de victimes de cancers du foie tenter de faire émerger des hypothèses explicatives, l'ordinateur peut différentier différents groupes, qu'on pourrait ensuite associer par exemple à leur

provenance géographique, génétique, à l'alcoolisme ou à l'exposition à un métal lourd ou à une toxine telle que l'aflatoxine.

- **Apprentissage supervisé** : Elle s'intéresse aux données étiquetées. L'objectif est de prédire l'étiquette (inconnue) associée à une nouvelle observation  $x$ , à partir de  $N$  observations étiquetées du jeu de données. Exemple : les réseaux de neurones artificiels ou les SVM sont des exemples typiques. Autre exemple : en fonction de points communs détectés avec les symptômes d'autres patients connus (les exemples), le système peut catégoriser de nouveaux patients au vu de leurs analyses médicales en risque estimé (probabilité) de développer telle ou telle maladie.
- **Apprentissage semi-supervisé** : Effectué de manière probabiliste ou non, il vise à faire apparaître la distribution sous-jacente des « exemples » dans leur espace de description. Il est mis en oeuvre quand des données (ou « étiquettes ») manquent... Le modèle doit utiliser des exemples non-étiquetés pouvant néanmoins renseigner (ex. en médecine, il peut constituer une aide au diagnostic ou au choix des moyens les moins coûteux de tests de diagnostics).
- **Apprentissage partiellement supervisé (probabiliste ou non)** : Quand l'étiquetage des données est partiel. C'est le cas quand un modèle énonce qu'une donnée n'appartient pas à une classe A, mais peut-être à une classe B ou C (A, B et C étant 3 maladies par exemple évoquées dans le cadre d'un diagnostic différentiel).
- **Apprentissage par renforcement** : l'algorithme apprend un comportement étant donné une observation. L'action de l'algorithme sur l'environnement produit une valeur de retour qui guide l'algorithme d'apprentissage (exemple : l'algorithme de Q-learning).

Dans le cadre de notre projet, nous nous concentrerons principalement sur l'apprentissage supervisé. Plusieurs algorithmes d'apprentissage supervisé existent, les plus utilisés sont :

### 1.3.1 Machine à Vecteurs de Support

Les SVM ou “*Support Vector Machine*” sont une famille d'algorithmes d'apprentissage automatique supervisé. Ils ont été appliqués avec succès à un certain nombre d'applications, allant de la prédiction de séries chronologiques, à la reconnaissance du visage, au traitement des données biologiques pour le diagnostic médical.

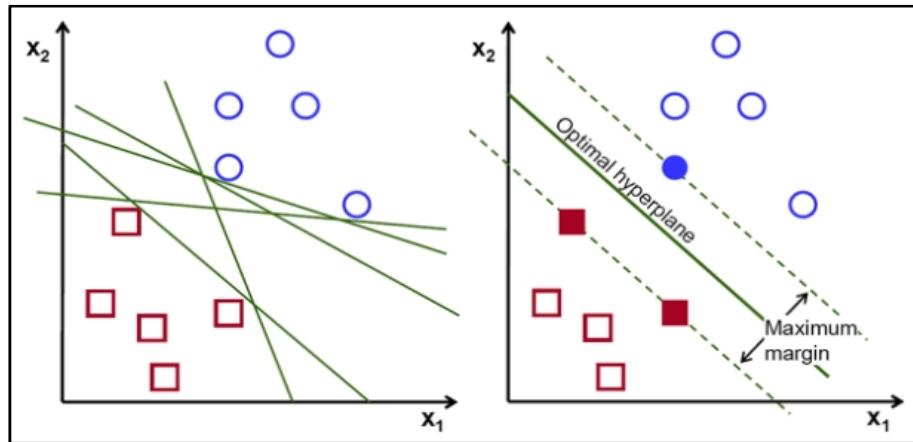


Figure 1.6 : Principe des SVM linéaire dans un espace 2D.

Leur principe de fonctionnement est simple. Il consiste à regrouper les données en classes à l'aide d'une frontière appelée hyperplan, de telle façon à avoir une distance maximale entre les différents groupes de données et l'hyperplan et donc avoir une meilleure robustesse face au bruit. Cette distance est aussi appelée « marge ».

Les SVMs sont qualifiés de « séparateurs à vaste marge » et les « vecteurs de support » sont les données les plus proches de la frontière. Cette notion de frontière suppose que les données soient linéairement séparables, ce qui est rare dans la pratique. C'est pour cela que les SVMs reposent souvent sur l'utilisation de « noyaux » pour classer des données non linéairement séparables. Ces fonctions mathématiques permettent de séparer les données en les projetant dans un espace de caractéristiques de plus grande dimension (Figure 1.7).

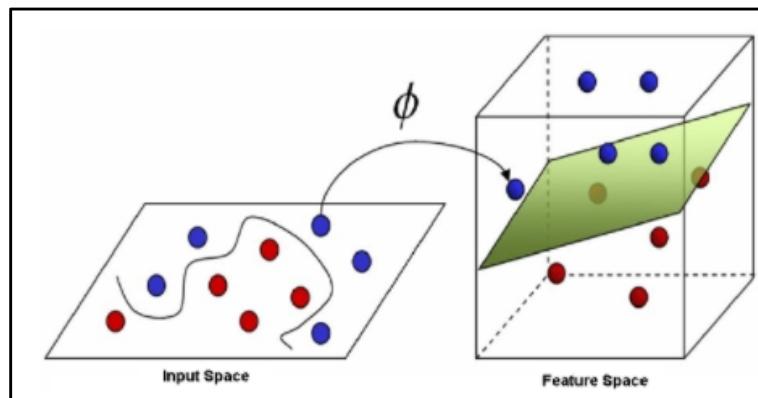


Figure 1.7:Principe du SVM dans un espace 3D.

### 1.3.2 Arbre de décision

C'est l'une des méthodes d'apprentissage supervisé les plus utilisées pour résoudre les problèmes de classification de données. Les décisions possibles sont situées aux extrémités des branches (les « feuilles » de l'arbre) et sont atteintes en fonction des décisions prises à chaque étape. Un arbre de décision fonctionne en appliquant de manière itérative des règles logiques très simples (typiquement des séparations de données par « hyperplan », généralisation d'un plan à plus de 2 dimensions), chaque règle étant choisie en fonction du résultat de la règle précédente. Les arbres de décision sont faciles à interpréter, utilisent des données catégorielles et continues et nécessitent peu de préparation de données. Cependant parfois les arbres générés sont instables : De légers changements aux niveaux des données, produisent des arbres très différents. Les changements des nœuds proches de la racine, affectent beaucoup l'arbre résultant. On dit que les arbres produisent des estimateurs de variance élevée. Pour répondre à ce problème qui n'admet pas de solution par optimisation algorithmique, des approches de type *Bagging* et « Forêts aléatoires » ont été appliquées. Ils seront traités de manière plus détaillée dans la présente étude.

Les arbres de décision sont souvent utilisés dans les applications médicales.

- Cremilleux et Robert présentent le cadre général de l'utilisation des arbres de décision en médecine.
- Kokol et al dans leur article montrent une certaine limitation des arbres de décision dans le domaine médical.
- Zorman et al évaluent différentes stratégies d'induction d'arbres de décision sur des problèmes de données de fracture orthopédique du monde réel. Ils ont testé diverses méthodes de construction d'arbres de décision uni-variés afin de trouver la meilleure stratégie d'induction. Ils ont testé quatre approches classiques, une approche hybride (réseaux de neurones et arbres de décision) et une approche évolutive. Les résultats montrent que toutes les approches avaient des problèmes de précision, de sensibilité ou de taille de l'arbre de décision. La comparaison montre que le meilleur compromis dans la construction d'arbres de décision de problèmes du monde réel est l'approche évolutive [17] .

### 1.3.3 Forêts aléatoires

L'algorithme des « forêts aléatoires » (ou Random Forest parfois aussi traduit par forêt d'arbres décisionnels) est un algorithme de classification qui réduit la variance des prévisions d'un arbre de décision seul, améliorant ainsi leurs performances. Pour cela, il combine de nombreux arbres de décisions dans une approche de type bagging.

L'algorithme des *forêts aléatoires* effectue un apprentissage en parallèle sur de multiples arbres de décision construits aléatoirement et entraînés sur des sous-ensembles de données différents. Le nombre idéal d'arbres, qui peut aller jusqu'à plusieurs centaines voire plus, est un paramètre important. Il est très variable et dépend du problème.

Concrètement, chaque arbre de la forêt aléatoire est entraîné sur un sous ensemble aléatoire de données, avec un sous ensemble aléatoire de caractéristiques (attributs) selon le principe des projections aléatoires. Les prédictions sont ensuite moyennées lorsque les données sont quantitatives (problème de régression) ou utilisées pour un vote pour des données qualitatives (problème de classification). C'est ce que l'on appelle une méthode de *bagging*.

Dans le cas des arbres de classification, l'algorithme des forêts aléatoires est l'un des classificateurs les plus efficaces (nécessitant peu de prétraitement des données). Il a été utilisé dans de nombreuses applications, comme pour la classification d'images de la caméra de console de jeu Kinect dans le but d'identifier des positions du corps [16]. A.Prochazka et S.Gulatil ont utilisé pour la classification des nodules thyroïdiens dans les images échographiques à l'aide de caractéristiques indépendantes de la direction extraites par décomposition binaire à deux seuils. Ses caractéristiques ont ensuite été utilisées dans la machine à vecteurs de support et les classificateurs de forêts aléatoires pour différencier les nodules en classes malignes et bénignes. Leurs résultats ont montré que les deux méthodes sont utiles dans la pratique [23].

### 1.3.4 Réseaux de neurones artificiels

Les réseaux de neurones artificiels ont acquis une large reconnaissance en tant qu'algorithme d'apprentissage automatique efficace, en permettant d'entraîner nos machines sur différents problèmes et sur des tâches plus avancées tels que : L'analyse de sentiment, la détection de cancer, la création artistique, etc.

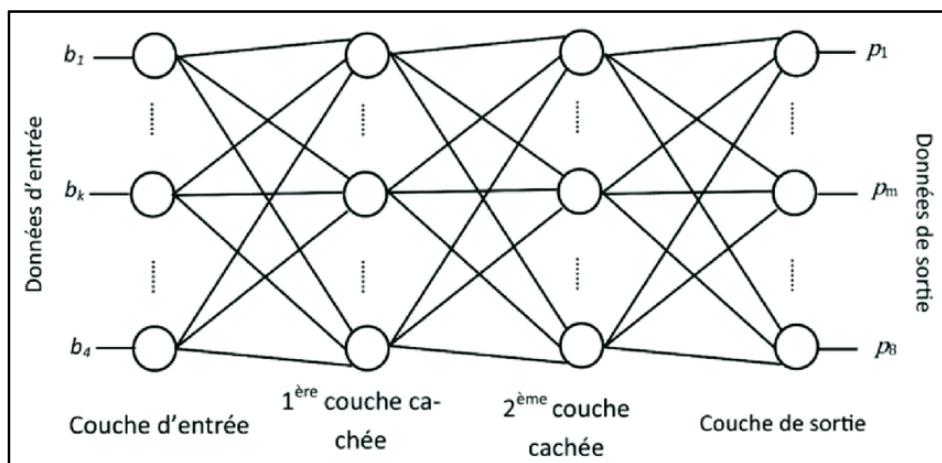


Figure 1.8: Réseau neuronal artificiel Feed-forward.

La figure 1.10 présente la description exacte de la structure d'un réseau de neurones, Le réseau de neurones artificiel, appelé réseau neuronal Feed-forward, permet uniquement aux signaux de voyager de l'entrée à la sortie. On y trouve 3 niveaux : Le premier pour la couche d'entrée qui contient le vecteur d'entités en entrée ; Le second pour les couches cachées qui contiennent les neurones qui se connectent à la fois à l'entrée et à la sortie ; Le troisième pour la couche de sortie qui se compose de la réponse du réseau neuronal.

Les cercles symbolisent les neurones qui représentent techniquement les fonctions d'activation. Pour un réseau de neurones basiques, la *fonction Logistique* est utilisée comme fonction d'activation [10]. Pour comprendre le fonctionnement, on analysera ce qui se passe au niveau d'un neurone.

Le réseau de neurones le plus simple qui existe porte le nom de *perceptron*, inventé en 1957 par Frank Rosenblatt et considéré comme la base des principaux systèmes utilisés aujourd'hui. Le neurone reçoit comme entrée les features  $x$  multipliées par des paramètres que la machine doit apprendre appelés poids synaptique  $w$ , ainsi le neurone fait son calcul en deux étapes, calcul linéaire et calcul non-linéaire.

La première opération réalisée par le neurone consiste en une somme de toutes les entrées pondérées par les coefficients synaptiques, c'est-à-dire la somme :

$$f(x) = \sum_{j=1}^m w_j x_j = w_1 x_1 + \dots + w_m x_m$$

Le résultat généré passe ensuite par une fonction d'activation non-linéaire nommée fonction d'activation ou fonction de transfert afin de renvoyer en sortie une valeur possible comprise dans les intervalles  $[0 ; +1]$  ou  $[-1 ; +1]$ , selon la fonction d'activation.

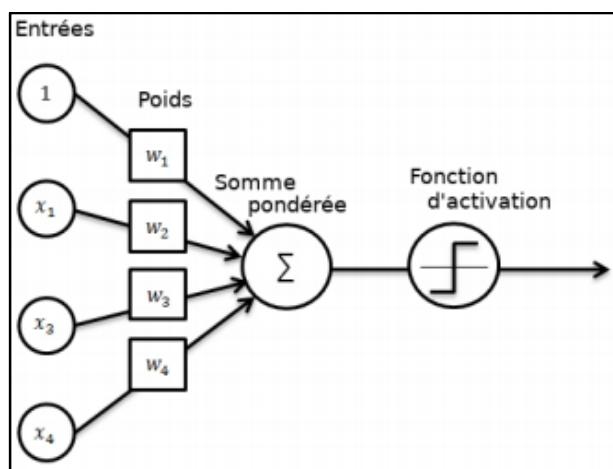


Figure 1.9: Neurone artificiel.

Parmi les principales fonctions d'activation, nous avons :

- Linéaire  $f(x) = x$
- Sigmoïde  $f(x) = \frac{1}{1+e^{-x}}$
- ReLU  $f(x) = \max(0, x)$
- Softmax  $f(x) = \frac{e^x}{\sum_1^k e^x}$
- Seuil  $f(x) = 1_{[1, +\infty]}(x)$

Les modèles linéaires, sigmoïdale, ReLU, softmax sont bien adaptés aux algorithmes d'apprentissage impliquant une rétro-propagation du gradient car leur fonction d'activation est différentiable, ils sont les plus utilisés. Le modèle à seuil est sans doute plus conforme à la réalité biologique mais pose des problèmes d'apprentissage.

Par ailleurs, nous utiliserons les modèles ReLU dans le réseau qui est la fonction d'activation la plus utilisée au monde à l'heure actuelle. Elle est utilisée dans presque tous les réseaux de neurones convolutifs ou le deep learning que nous détaillerons par la suite, ainsi que des modèles softmax plus adaptés pour différencier les classes à la fin du réseau [18].

### 1.3.5 Limites de l'apprentissage automatique

Le principal inconvénient de l'*apprentissage automatique* classique est le fait qu'ils nécessitent toujours une expertise humaine en matière d'extraction des caractéristiques, ceci constitue un problème dans le cas du BIG DATA.

Habituellement, une analyse exploratoire approfondie des données est d'abord effectuée. Pour faciliter le traitement, une réduction de la dimension de la base pourrait être effectuée pour rendre les motifs plus visibles et les algorithmes d'apprentissage fonctionnels, vu que leurs performances dépendent fortement des fonctionnalités techniques [16].

Dans le cas du *deep learning*, l'extraction des caractéristiques est faite automatiquement. Le réseau apprend à extraire les caractéristiques dont il aura besoin pour prendre une décision de classification pertinente-.

## 1.4 Apprentissage profond (Deep Learning)

L'apprentissage profond (Deep Learning) est un domaine du machine Learning, c'est une nouvelle approche qui met l'accent sur l'apprentissage des couches successives [8]. Il est caractérisé par l'utilisation d'un réseau neuronal composé de dizaines voire de centaines de « couches » de neurones, chacune reçoit et interprète les informations et les données de la couche précédente.

La spécification de ce qu'une couche fait à ses données d'entrée, est stockée dans les poids de la couche (paramètres de la couche) [8]. L'apprentissage dans ce cas, consiste à trouver la bonne valeur du poids de chaque couche pour avoir un résultat meilleur.

Cependant, un réseau profond peut contenir des dizaines de millions de paramètres et en trouver une valeur correcte est une tâche qui présente de grandes difficultés. En apprentissage profond, l'astuce fondamentale est l'utilisation d'une fonction objective qui calcule le score de distance entre la sortie prédite et la sortie attendue. Ce score est utilisé comme un *feedback* signal pour ajuster la valeur des poids, cet ajustement est le travail de l'optimiseur qui met en œuvre l'algorithme centrale du *deep Learning* " *Back-propagation* " [8].

Les techniques d'apprentissage profond peuvent être adaptées à différents domaines et applications bien plus facilement que les algorithmes de ML classiques. Le transfert d'apprentissage a rendu efficace l'utilisation des réseaux profonds pré-entraînés pour différentes applications et dans de nombreux domaines. Par exemple en vision par ordinateur, les réseaux de classification d'images pré-entraînés sont souvent utilisés pour l'extraction des caractéristiques pour détecter et segmenter les objets. L'utilisation de ces réseaux pré-entraînés facilite l'apprentissage du modèle.

### 1.4.1 Réseaux de neurones convolutifs

Les réseaux convolutifs (CNN) sont une forme particulière de réseaux neuronaux multicouches dont l'architecture des connexions est inspirée de celle du cortex visuel des mammifères. Leur conception suit la découverte de mécanismes visuels dans les organismes vivants. Ces réseaux de neurones artificiels (aussi baptisés réseaux de neurones à convolution, ou CNN) sont capables de catégoriser les informations des plus simples aux plus complexes.

Comme les autres réseaux de neurones, un CNN est composé d'une couche d'entrée, d'une couche de sortie et de nombreuses couches cachées. Les CNN sont spécialement conçus

pour traiter des images en entrée. Leur architecture est alors plus spécifique, elle est composée de deux blocs principaux [9].

C'est le premier bloc qui a fait que les CNN soient si particuliers, car il permet d'extraire les caractéristiques des images, pour en faire des simples. L'image fournie en entrée passe à travers une succession de filtres et crée ainsi de nouvelles images appelées cartes de caractéristiques, les cartes de convolutions obtenues sont concaténées dans un vecteur de caractéristiques appelé code CNN.

Le second bloc se retrouve à la fin de tous les réseaux de neurones, c'est le bloc de classification où le code CNN obtenu en sortie du premier bloc est fourni en entrée.

La distribution hiérarchique la plus courante des couches CNN contient quatre types de couches : La couche de convolution, la couche de pooling, la couche de correction Relu et la couche fully-connected (voir la figure 1.11) [9].

**1- La couche de convolution (CONV) :** est un empilement de convolutions. En effet, l'image est parcourue par plusieurs noyaux de convolution qui donnent lieu à plusieurs cartes de caractéristiques de sorties. Chaque noyau (ou masque) de convolution possède des paramètres spécifiques à l'information recherchée dans l'image, son but est d'extraire les caractéristiques du volume d'entrée.

Le choix des paramètres du noyau de convolution dépend de la tâche à résoudre. Avec les méthodes deep Learning, ces paramètres sont automatiquement appris par l'algorithme à partir des données d'entraînement. La technique de rétro propagation du gradient permet l'ajustement des paramètres en fonction de la valeur du gradient de la fonction de perte. Cette dernière calcule l'erreur entre la valeur prédite et la valeur cible [48].

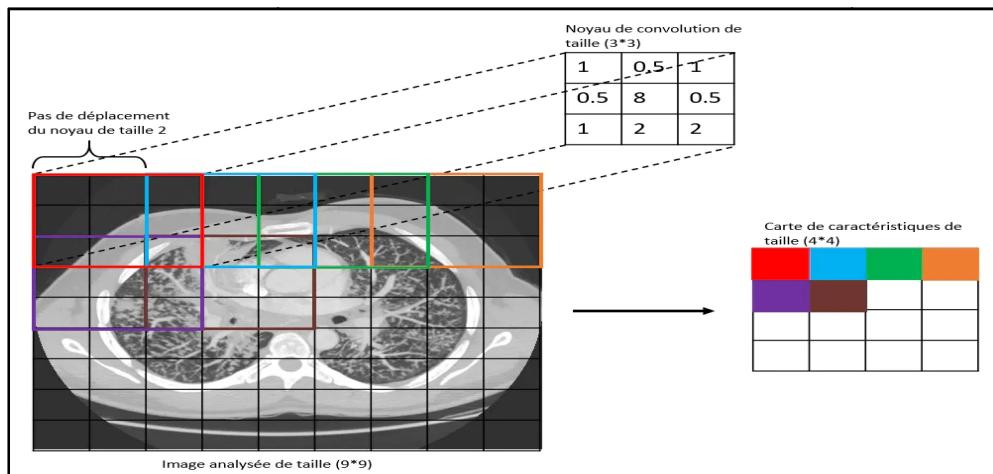


Figure 1.10: Schéma d'une convolution avec un noyau de taille 3\*3 et un pas de 2 [48].

**2- La couche de correction ReLU (*Rectified Linear Units*) :** La couche de correction ou d'activation est l'application d'une fonction non-linéaire aux cartes de caractéristiques en sortie de la couche de convolution. En rendant les données non-linéaires, elle facilite l'extraction des caractéristiques complexes qui ne peuvent pas être modélisées par une combinaison linéaire d'un algorithme de régression. [48]

**3- La couche de pooling (POOL) :** Ce type de couche est souvent placé entre deux couches de convolution. L'opération de pooling consiste à réduire la taille des images tout en préservant leurs principales caractéristiques, accélérer le temps de calcul et diminuer le risque de sur-apprentissage.

L'opération de pooling la plus courante est celle du maximum MaxPool (2\*2, 2). Elle est plus efficace que la moyenne, car elle maximise le poids des activations fortes. Elle est appliquée à la sortie de la couche précédente comme un filtre de convolution de taille (2\*2), et se déplace avec un pas de 2. En sortie de la couche de pooling est obtenue une carte de caractéristique compressée par un facteur de 4.

**4- La dernière couche fully-connected (FC) :** La couche FC applique successivement une combinaison linéaire puis une fonction d'activation afin de classifier l'image d'entrée, elle renvoie un vecteur de taille  $N$ , où  $N$  est le nombre de classes correspondantes au problème étudié.

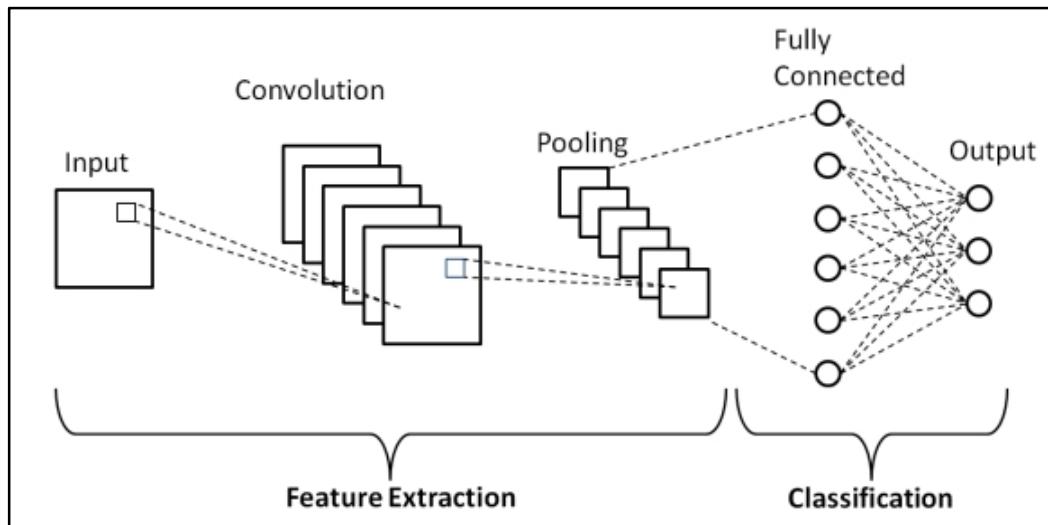


Figure 1.11 : Structure du CNN.

La différence remarquable entre un CNN et un ANN est due à la présence de couches convolutives et du max-pooling. Le concept de convolution a pour rôle de passer un filtre sur une partie des données et additionner les éléments de cette partie. Le concept de pooling

fonctionne de la même manière, mais sur une plus grande zone. La convolution agit plus comme un filtre, le pooling recherche plutôt des comportements dans les entrées.

Créer un nouveau réseau de neurones convolutif est coûteux en termes d'expertise, de matériel et du volume des données annotées nécessaires.

Il s'agit d'abord de fixer l'architecture du réseau, c'est-à-dire le nombre de couches, leur taille et les opérations matricielles qui les connectent. L'entraînement consiste alors à optimiser les coefficients du réseau pour minimiser l'erreur de classification en sortie. Cet entraînement peut prendre plusieurs semaines pour les meilleurs CNN, avec de nombreux GPU travaillant sur des centaines de milliers d'images annotées. Plus les couches s'empilent, plus le nombre de convolutions et de paramètres à optimiser est élevé. L'ordinateur doit être en mesure de stocker plusieurs giga-octets de données et de faire efficacement les calculs [9].

Actuellement il est possible d'utiliser les connaissances acquises par un réseau de neurones lors de la résolution d'un problème afin d'en résoudre un autre plus ou moins similaire. On réalise ainsi un transfert de connaissances ou le transfert Learning [9].

#### **1.4.2 Apprentissage par transfert (Transfer Learning)**

L'apprentissage par transfert est une technique d'apprentissage automatique où un modèle développé pour une tâche particulière est à nouveau utilisé comme point de départ pour développer un modèle pour une autre tâche. C'est une technique très utilisée en pratique et simple à mettre en œuvre. L'utilisation des méthodes de *Transfert Learning* en *Deep Learning*, consiste principalement à exploiter des réseaux de neurones pré-entraînés. Actuellement il est possible de se servir de la puissance des CNNs en évitant toute la complexité et la difficulté de les créer, et cela en adaptant des réseaux pré-entraînés disponibles, publiquement cette technique est appelée le fine tuning.

Du point de vue de l'apprentissage profond, le problème de classification d'images peut être résolu par l'apprentissage par transfert. En fait, plusieurs résultats de pointe en classification d'images sont basés sur des solutions d'apprentissage par transfert.

#### **1.4.3 Fine Tuning : Adaptation des CNN pré-entraînés**

Le Fine Tuning consiste à se servir d'un modèle pré-entraîné comme initialisation pour l'entraînement sur un nouveau problème. L'intérêt est double : on utilise une architecture

optimisée avec soin par des spécialistes, et l'on profite des capacités d'extraction de caractéristiques apprises sur un jeu de données de qualité. Le Fine Tuning sur des images consiste en quelques sortes à prendre un système visuel déjà bien entraîné sur une tâche de classification pour le raffiner sur une tâche similaire.

Le fine tuning est une réutilisation des CNNs pré-entraînés sur une très grande base de données. On les re-entraînent plus finement, de manière supervisée, pour concevoir de nouvelles architectures répondant à des tâches spécifiques, et ainsi tenter de contourner l'exigence de grands ensembles de données pour l'entraînement des réseaux profonds.

Lorsqu'on réutilise un modèle pré-entraîné pour une tâche quelconque, on doit d'abord supprimer le classificateur d'origine, puis ajouter un nouveau classificateur qui correspond aux besoins, et enfin affiner le modèle selon l'une des trois stratégies :

#### **Stratégie 1 : Entraînement de l'ensemble du modèle**

Dans ce cas, on utilise l'architecture du modèle pré-entraîné, et on l'entraîne en fonction du nouveau jeu de données. Pour entraîner le modèle à partir de zéro, on aura besoin d'un grand ensemble de données.

#### **Stratégie 2: Entrainer quelques couches et laisser les autres gelées (freeze)**

En général, les couches inférieures d'un CNN font référence à des caractéristiques générales (indépendantes du problème), tandis que les couches supérieures font référence à des caractéristiques spécifiques (dépendantes du problème). L'idée principale pour cette stratégie est de jouer avec cette dichotomie en choisissant combien de couches à geler qui ne changent pas pendant l'entraînement.

#### **Stratégie 3 : Geler la base convolutive.**

Ce cas correspond à une situation extrême du compromis train/freez. L'idée principale est de conserver la base convolutive dans sa forme d'origine, puis d'utiliser ses sorties pour alimenter le classifieur.

L'idée est d'utiliser le modèle pré-entraîné comme extracteur de caractéristiques, ce cas on l'utilise généralement quand on manque de puissance de calcul, si un ensemble de données est petit et/ou si le modèle pré-entraîné résout un problème très similaire à celui qu'on veut résoudre.

## 1.5 Architectures CNN existantes

On peut comprendre qu'une architecture CNN est un agencement spécifique des couches susmentionnées. De nombreuses variantes de tels agencements se sont développées au fil des ans, aboutissant à plusieurs architectures CNN. Les plus courants d'entre eux sont :

### 1.5.1 ResNet

Le réseau de neurones résiduel (Residual Network) connu par le nom ResNet, il est apparue par rapport au problèmes de rétro-propagation du gradient et l'augmentation de l'erreur d'apprentissage, car plus le réseau est profond plus l'apprentissage devient plus difficile mais avec ResNet, l'optimisation des réseaux profonds est assurée en utilisant des connexions résiduelles. Cela permet aux gradients de passer dans deux couches de convolution, et de passer directement à travers un saut aux couches suivantes [11]. Les ResNets sont constitués de plusieurs blocs résiduels, empilés les uns sur les autres pour former un réseau complet.

Les résultats d'étude basée sur cette classification ont montré que les problèmes du gradient sont résolus en adaptant cette architecture, en fait ResNet présente 5 autres versions qui se diffèrent seulement au niveau du nombre des couches, on cite ResNet-18 (18 couches), ResNet-34 (34 couches), ResNet-50, ResNet-101 et ResNet-152.

### 1.5.2 Inception

Le principal but de la conception de ce réseau c'est d'améliorer les performances des réseaux de neurones profonds à la fois en termes de rapidité et de précision, il existe plusieurs versions de cette architecture les plus populaires sont : Inception-v2, Inception-v3, et Inception-v4. L'architecture de ce réseau se base essentiellement sur des modules d'Inception, un module est le résultat d'une nouvelle version de ce réseau par exemple Inception-v1 a été conçu suite aux 3 problèmes suivants :

- La difficulté de transmettre les mises à jour de gradient à travers un réseau trop profond.
- L'augmentation du temps de calcul suite aux grandes opérations de convolution.
- La difficulté de choisir la bonne taille du noyau pour l'opération de convolution.

GoogleNet (ou Inception Network) est une architecture qui se dérive du réseau **Inception**, se compose de 22 couches (27 couches, y compris les couches de pooling), et une

partie de ces couches est un total de 9 modules d'Inception, il est conçue par des chercheurs de Google.

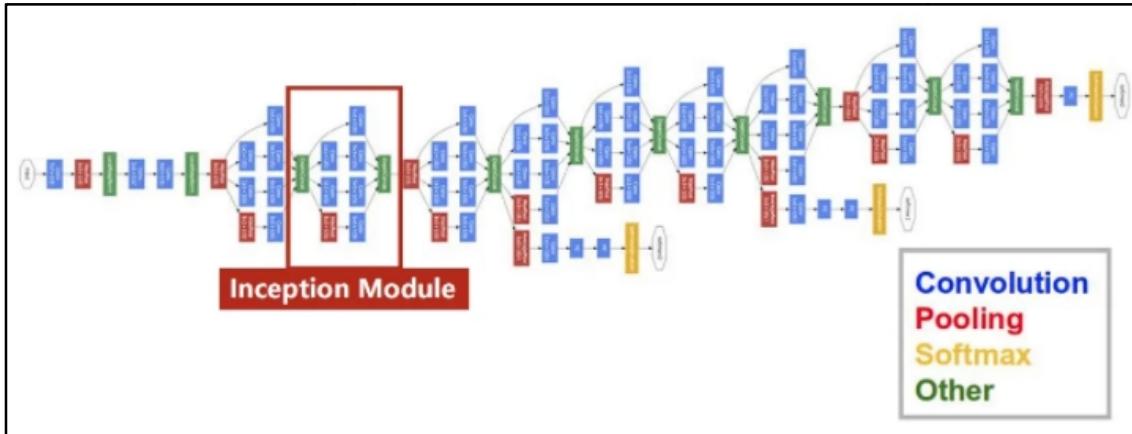


Figure 1.12 : Structure de GoogleNet.

Aujourd'hui, GoogleNet est utilisé pour les tâches de vision par ordinateur telles que la détection de lésion dans le domaine médical et la reconnaissance des visages, la formation antagoniste, etc.

### 1.5.3 VGG16

Il a été présenté par les chercheurs du Visual Graphics Group à Oxford (**VGG**), il existe deux algorithmes disponibles : **VGG16** et **VGG19**. Dans cet article, nous allons nous concentrer sur VGG16.

Vgg16 est un modèle de réseau de neurones convolutif, Il a apporté une amélioration par rapport à AlexNet, il a abordé un autre aspect très important des CNN : la profondeur en remplaçant les grands filtres par plusieurs filtres. Aujourd'hui il est l'une des architectures de reconnaissance d'images les plus utilisées, le réseau se caractérise par sa simplicité : les seuls autres composants étant des couches de mise en commun et une couche entièrement connectée.

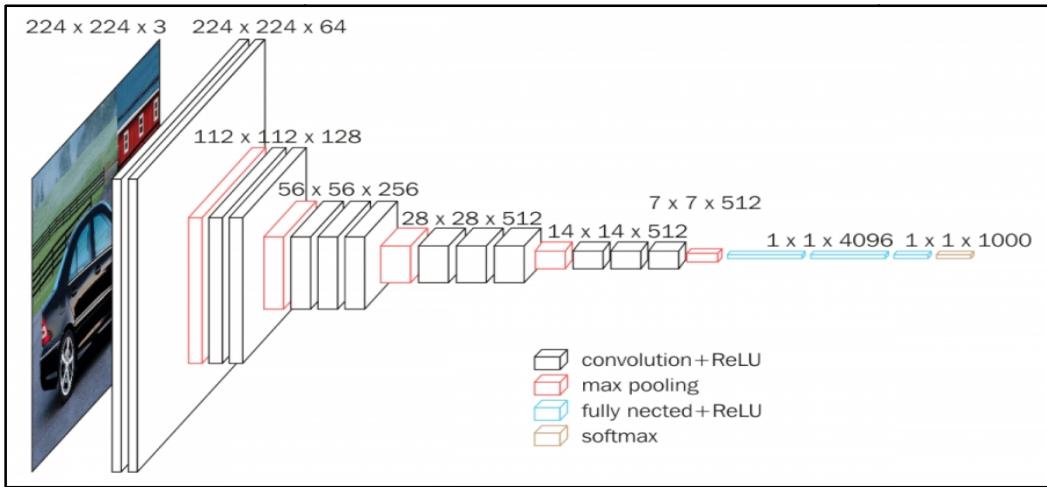


Figure 1.13 : Architecture VGG-16.

#### 1.5.4 EfficientNet

EfficientNet est une architecture de réseau neuronal convolutif créée par Google et une méthode de mise à l'échelle qui met à l'échelle uniformément toutes les dimensions de profondeur/largeur/résolution à l'aide d'un *coefficient composé*. Contrairement à la pratique conventionnelle qui met à l'échelle arbitrairement ces facteurs. Ces CNN offrent non seulement une meilleure précision, mais améliorent également l'efficacité des modèles en réduisant le nombre de paramètres par rapport aux autres modèles de pointe. Le modèle EfficientNet-B0 est une architecture de base de taille mobile simple et basé sur les blocs résiduels de goulot d'étranglement inversés de MobileNetV2 , en plus des blocs de compression et d'excitation. Il est entraîné sur l'ensemble de données ImageNet. EfficientNets transfère également bien et atteint une précision de pointe sur CIFAR-100 (91,7 %), Flowers (98,8 %) et 3 autres ensembles de données d'apprentissage de transfert, avec un ordre de grandeur moins de paramètres.

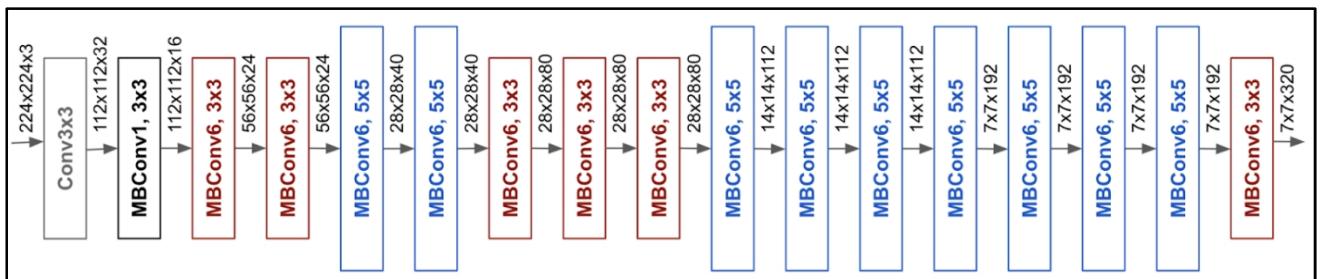


Figure 1.14 : Architecture du réseau de base EfficientNet-B0.

## **1.6 Apprentissage profond et imagerie médicale**

Les tâches de vision par ordinateur telles que la détection, la segmentation et la classification sont généralement effectuées avec des algorithmes basés sur des caractéristiques, des classificateurs et des méthodes d'extraction de formes.

Les approches récentes basées sur l'apprentissage profond représentent un changement de paradigme important où les caractéristiques ne nécessitent pas un choix préalable par un expert, elles sont apprises d'une manière progressive. Ces caractéristiques décrivent, par exemple, l'apparence des organes et des points d'intérêt dans les images médicales. Les classifieurs intègrent des caractéristiques pour générer une décision. L'apprentissage profond fournit une approche de bout en bout où les caractéristiques sont apprises pour maximiser les performances du classifieur [2].

### **1.6.1 Classification**

#### **1.6.1.1 Classification des images**

La classification des images a été l'un des premiers domaines dans lesquels l'apprentissage profond a apporté une contribution majeure à l'analyse d'images médicales. Dans la classification des images, on a généralement une ou plusieurs images en entrée avec une seule variable de diagnostic en sortie (par exemple, maladie présente ou non). Dans un tel contexte, chaque image de diagnostic est un échantillon et les tailles des ensembles de données sont minimes par rapport à celles de la vision par ordinateur (par exemple, des centaines / milliers contre des millions d'échantillons).

#### **1.6.1.2 Classification des objets ou des lésions**

La classification des objets se focalise généralement sur la classification d'une petite partie de l'image médicale en deux classes ou plus (par exemple, la classification des lésions thyroïdiennes dans des images d'échographie médicale). Pour beaucoup de ces tâches, des informations locales sur l'apparence des lésions et des informations contextuelles globales sur la localisation des lésions sont nécessaires pour une classification précise.

Setio et al. (2016) [12] ont utilisé un CNN multi-flux pour classer les points d'intérêt de la tomodensitométrie thoracique en tant que nodule ou non-nodule. Jusqu'à neuf patchs orientés différemment extraits du candidat ont été utilisés dans des flux séparés et fusionnés dans les couches entièrement connectées pour obtenir la sortie de classification finale. En revanche,

Nie et al. (2016) [15] ont exploité la nature 3D de l'IRM en entraînant un CNN 3D pour évaluer la survie chez des patients souffrant de gliomes de haut grade.

La majorité des articles récents préfèrent utiliser un CNN pré-entraîné. Dans certains cas, d'autres architectures sont utilisées, telles que les RBM et les SAE.

Dans l'ensemble, la classification d'objets voit moins l'utilisation de réseaux pré-entraînés que les classifications des images, principalement en raison de la nécessité d'incorporer des informations contextuelles ou tridimensionnelles. Plusieurs auteurs ont trouvé des solutions innovantes pour ajouter ces informations à des réseaux profonds avec de bons résultats, et en tant que tel, nous nous attendons à ce que l'apprentissage profond devienne encore plus important pour cette tâche dans un proche avenir[6].

## 1.6.2 Détection

Le domaine de l'intelligence artificielle a connu de nombreuses avancées grâce au Deep Learning et au traitement des images. Il est maintenant possible de reconnaître des images ou même de détecter (localiser) des objets à l'intérieur d'une image, c'est un processus de décision binaire ; Il s'agit de donner une estimation positive ou négative de la présence d'un objet sur l'image.

### 1.6.2.1 Détection d'objets ou de lésions

La détection d'objets d'intérêt ou de lésions sur les images médicales est un problème important dans l'analyse d'imagerie médicale. La plupart des travaux antérieurs se concentrent sur la détection et la segmentation d'une catégorie spécialisée de lésions (lésion pulmonaires, thyroïdiennes etc..). Dans la pratique clinique, les radiologues sont chargés de trouver tous les types d'anomalies possibles qui présente l'élément clé du diagnostic et l'une des plus exigeantes en main d'œuvre pour les cliniciens. En règle générale, les tâches consistent en la localisation et l'identification de petites lésions avec un cadre de délimitation dans tout l'espace de l'image. Il existe une longue tradition de recherche dans les systèmes de détection assistée par ordinateur qui sont conçus pour détecter automatiquement les lésions, améliorant la précision de détection ou diminuant le temps de lecture des experts humains. Fait intéressant, le premier système de détection d'objets utilisant des CNN a déjà été proposé en 1995, utilisant un CNN à quatre couches pour détecter les nodules dans les images radiographiques (Lo et al., 1995) [13].

La plupart des systèmes de détection d'objets d'apprentissage en profondeur publiés utilisent encore des CNN pour effectuer la classification des pixels (ou voxels), après quoi une certaine forme de post-traitement est appliquée pour obtenir des objets candidats. Comme la tâche de classification effectuée à chaque pixel est essentiellement une classification d'objets, l'architecture et la méthodologie CNN sont très similaires à celles de la classification des objets ou des lésions.

Les défis dans l'application significative des algorithmes d'apprentissage en profondeur dans la détection d'objets sont donc pour la plupart similaires à ceux de la classification d'objets. Nous nous attendons à ce que l'accent soit mis davantage sur ces domaines dans un proche avenir [6].

## 1.7 Conclusion

Nous avons présenté au cours de ce chapitre la maladie de la thyroïdeet des connaissances théoriques sur l'apprentissage automatique et les réseaux de neurones tels que le perceptron simple. Ensuite, nous nous sommes intéressés à la théorie de l'apprentissage profond et particulièrement au recours des CNNs, leur mécanisme d'apprentissage et l'évolution des architectures proposées pour réaliser une classification d'images.

Dans le chapitre suivant, nous allons présenter un état de l'art sur les approches basées sur les CNNs ayant été réalisées pour la classification des nodules thyroïdiens.

## **Chapitre 2 : Applications de l'apprentissage automatique ou profond au diagnostic médical**

L'objectif général de ce projet est de classer des nodules thyroïdiens comme bénins ou malins grâce à des images échographiques en ayant recours à l'apprentissage profond. À cet égard, nous allons présenter dans ce chapitre les différentes applications des deux familles d'apprentissage (automatique et profond) pour l'aide au diagnostic médical en générale, puis nous nous intéresserons plus particulièrement au diagnostic des nodules thyroïdiens.

### **2.1 Cas général**

#### **2.1.1 Application de l'apprentissage automatique dans le diagnostic médical**

L'apprentissage automatique apporte de nouvelles solutions aux professionnels de santé pour gagner du temps et optimiser le bon diagnostic. Il ouvre de nouvelles perspectives dans le repérage des maladies. Par exemple, il peut aider les médecins à détecter plus facilement les anomalies des images radiologiques des patients. Parmi les travaux ayant utilisé les techniques du machine Learning on trouve le travail de Ozciift [35] qui a utilisé le meilleur algorithme des forêts d'arbres décisionnels (ou forêts aléatoires de l'anglais random forest) pour sélectionner les caractéristiques optimales de quatre ensembles de données médicales : le cancer du côlon, le cancer de la leucémie, le cancer du sein et le cancer du poumon. Le modèle proposé avec la précision des caractéristiques extraites a été comparé à 15 classificateurs largement utilisés formés avec toutes les caractéristiques et a montré une précision de classification améliorée [34].

Nguyen et coll. [36] ont utilisé le même classifieur “aléatoire des forêts” combiné à une sélection de caractéristiques pour le diagnostic du cancer du sein, le modèle proposé a atteint une précision de classification de 99,82%. On trouve aussi d'autres travaux qui ont utilisé ce classificateur par exemple pour diagnostiquer la maladie de lymphadénopathie abdominale [37], la maladie d'Alzheimer et le risque cardiovasculaire.

Dans la littérature on y trouve pas mal de recherches sur le diagnostic médical du cancer du sein à l'aide des SVM et la plupart d'entre elles ont rapporté des précisions de classification élevées par exemple l'étude [38] où K. Polat et S. Güneş ont utilisé le least square SVM pour le diagnostic médical du cancer du sein avec une précision de 98,53%.

Ce classifieur est largement utilisé dans l'approche hybride pour le diagnostic médical de plusieurs maladies, par exemple la combinaison du Genetic + Fuzzy + SVM pour le diagnostic du diabète, des maladies du foie et du cœur [39], un autre exemple est l'algorithme hybride ANN + SVM pour le diagnostic du cancer de la prostate [40] et le ANFIS + SVM pour l'identification de la douleur [41].

Cependant, l'obstacle pratique du modèle de classification basé sur SVM est sa nature de boîte noire. Une solution possible à ce problème est l'utilisation de techniques d'extraction de règles SVM ou l'utilisation d'un modèle hybride SVM combiné avec d'autres modèles plus interprétables.

### **2.1.2 Application de l'apprentissage profond dans le diagnostic médical**

L'apprentissage profond s'est rapidement imposé comme un standard dans plusieurs domaines en pulvérifiant les records des précédentes méthodes de l'état de l'art. Ses domaines sont principalement l'analyse et la reconnaissance d'images ainsi que le traitement du langage naturel. Un des enjeux majeurs de cette approche est son application à la santé.

Les approches d'apprentissage automatique ont récemment gagné en popularité et de nombreux travaux actuels tendent à utiliser les réseaux de neurones convolutifs pour concevoir un système de prédiction et de diagnostic médical, ces systèmes se concentrent en général sur les CNNs car ils offrent de bonnes performances [1], des résultats très intéressants ont ainsi été produits en 2017 et 2018 pour les pathologies rétiennes, cardiaques, abdominales, musculo-squelettiques.

L'étude d'Antony et al. [19] a obtenu une performance de 57.6% pour déterminer les grades de l'entorse du genou en utilisant l'apprentissage par transfert avec le modèle CNN AlexNet. Kim et al. [20] ont présenté une approche pour l'extraction de vecteur des caractéristiques avec un transfert d'apprentissage pour la classification cytopathologie avec une précision de 70.5%. Le Deep Learning est également utilisé pour lutter contre le diabète avec une précision de 98%, il a aidé aussi au diagnostic de la maladie d'Alzheimer avec une précision de 95.35 % [21]. Google s'est investi aussi pour lutter contre le cancer du poumon en développant un algorithme permettant de détecter le cancer de façon beaucoup plus précise qu'un humain. En comparant leurs résultats avec ceux des plus grands cancérologues américains, ils se sont aperçus que leur technologie a permis de détecter 5% de cancer en plus. Elle a aussi réduit de 11% les faux-positifs [22].

D'autres contextes cliniques avec des algorithmes d'apprentissage profond attirent l'attention des chercheurs grâce à leurs excellentes performances dans les tâches de d'analyse et de compréhension d'images, telles que la segmentation des métastases hépatiques par tomodensitométrie (TDM) [4], la segmentation des tumeurs cérébrales avec imagerie par résonance magnétique (IRM), la classification de la maladie pulmonaire interstitielle avec tomodensitométrie thoracique à haute résolution [2].

L'apprentissage profond offre tellement d'opportunités qu'il serait intéressant de pouvoir l'intégrer dans les hôpitaux pour soigner les patients. Mais si l'algorithme suggère un traitement et qu'il se trompe, les conséquences peuvent être catastrophiques. C'est en partie pour cela que l'on est encore très limité sur ce que l'on peut faire avec l'IA sans supervision d'un expert dans la santé. C'est aussi le cas pour d'autres domaines critiques comme la justice [22].

## 2.2 Cas particulier

De nombreux nodules thyroïdiens étant détectés incidemment, il est important d'identifier autant de nodules malins que possible tout en excluant ceux qui risquent fort d'être bénins par une biopsie par aspiration à l'aiguille fine (FNA) ou une chirurgie. Cette section présente quelques systèmes existants de diagnostic assisté par ordinateur (DAO) permettant de détecter et classer les nodules thyroïdiens sur des images échographiques en utilisant les techniques de ML puis DL.

### 2.2.1 Application de l'apprentissage automatique dans le diagnostic des nodules thyroïdiens

De nombreuses études utilisant la technique ML ont développé des systèmes de DAO basés sur les caractéristiques des images échographiques à savoir : la composition, la forme, la marge, l'échogénicité et les calcifications et ont démontré leur potentiel dans le diagnostic du cancer de la thyroïde. A titre d'exemple, Chang et al. [32] ont développé un système semi-automatique de diagnostic assisté par ordinateur (DAO) pour détecter le cancer de la thyroïde à l'aide d'images échographiques bidimensionnelles, pour cela, un ensemble de 118 images échographiques présentant des nodules malins et bénins est collecté. Un logiciel a été développé pour extraire des caractéristiques quantitatives de ces images puis en se basant sur ces dernières, un classificateur de machine à vecteurs de support (SVM) a été utilisé pour différencier les nodules bénins et malins, dans cette étude, les auteurs ont rapporté une aire sous la courbe de fonctionnement du récepteur (AUROC) jusqu'à 0,986 avec le SVM, ce qui

était similaire aux résultats obtenus par inspection visuelle par des radiologues (AUROC, 0,979).

Nous retrouvons aussi l'étude de Hongxun et Zhaohong [46] qui ont développé différents algorithmes d'apprentissage automatique, à utiliser dans le diagnostic différentiel des nodules thyroïdiens suspects à savoir le classificateur naïf de Bayes, la machine à vecteurs de support (SVM) et la fonction de base radiale (RBF) - réseau de neurones (NN). Cette étude a inclus 970 nodules thyroïdiens prouvés histo-pathologique chez 970 patients. Deux radiologues ont revu les images échographiques, et les nodules ont été classés selon un système de notation échographique à cinq niveaux. Des variables statistiquement significatives basées sur les observations d'un radiologue expérimenté ont été obtenues avec une optimisation des attributs en utilisant une validation croisée quintuple et appliquées comme nœuds d'entrée pour construire des modèles pour prédire la malignité des nodules. Les résultats ont montré que les algorithmes d'apprentissage automatique sont sous-performés par rapport aux lectures du radiologue expérimenté utilisées pour les construire, et le RBF-NN a surpassé les autres modèles d'algorithmes d'apprentissage automatique atteignant la sensibilité la plus élevée de 92,31 % avec une valeur de l'AUC égal à 0,9103.

Dans [33], une autre étude a été menée dans le but de développer un modèle de diagnostic moins subjectif aidé par l'apprentissage automatique cependant un total de 2064 nodules thyroïdiens ont été collectés. Pour effectuer cette tâche, plusieurs modèles ont été développés en utilisant 60% d'échantillons sélectionnés au hasard sur la base de neuf algorithmes couramment utilisés et validés en utilisant les 40% des cas restants. Cette étude a montré que le classifieur Random Forest peut mieux diagnostiquer les nodules thyroïdiens malins que les radiologues (ASC = 0,924 vs 0,834).

Par conséquent, de nombreux chercheurs ont suggéré que les systèmes de DAO basés sur la ML pourraient jouer un rôle dans la génération d'un deuxième avis pour les radiologues.

## **2.2.2 Application de l'apprentissage profond dans le diagnostic des nodules thyroïdiens**

Dès que les réseaux de neurones profonds sont entrés pour la première fois dans le champ de vision du public, quelques travaux ont utilisé ce nouveau formalisme du Deep Learning pour la détection et la reconnaissance automatiques des nodules thyroïdiens. Le travail publié dans [25] fut un des premiers articles proposant d'appliquer les CNNs pour la

détection et la reconnaissance automatiques des nodules thyroïdiens, cet article a mentionné que l'utilisation de l'imagerie par ultrasons thyroïdiens est une méthode clinique courante pour le diagnostic des nodules dans la région thyroïdienne. Ce système détecte d'abord la thyroïde sur l'image échographique, extrait ensuite la région d'intérêt, puis l'analyse avec un réseau neuronal convolutif pour intégrer les caractéristiques d'identification de chaque dimension.

Ma et al. [29] ont proposé une méthode hybride pour le diagnostic des nodules thyroïdiens en échographie en utilisant le transfert Learning, cette étude consiste à combiner deux réseaux de modèles CNN pré-entraînés avec différentes couches convolutives et des couches entièrement connectées. Les cartes de caractéristiques des deux réseaux obtenues après raffinement sont fusionnées et utilisées comme entrée du classifieur softmax pour classer les nodules thyroïdiens. La méthode proposée est validée sur 15000 images échographiques collectées dans deux hôpitaux locaux. Les résultats des expériences montrent que les méthodes basées sur CNN proposées peuvent diagnostiquer avec une précision de 83,02%.

Liu et al. [28] ont proposé une autre approche de classification basée sur l'extraction des caractéristiques des images échographiques en utilisant VGG-F et en le transférant dans le domaine de l'image échographique, pour générer des caractéristiques sémantiques profondes pour les combiner par la suite avec les caractéristiques classiques tels que l'histogramme de gradient orienté (en anglais, histogram of oriented gradients ou HOG), les motifs binaires locaux (LBP), LCM (Gray Level Co-Occurrence Matrix), SIFT (Scale Invariant Feature Transform), VLAD (Vector of Locally Aggregated Descriptors) pour former un espace de caractéristiques hybride. Les résultats expérimentaux sur 1037 images montrent que la précision du modèle généré est de 93 %. Ce qui dépasse l'approche précédente de 10%.

D'autres études ont tenté d'améliorer leurs prédecesseurs, comme l'étude de Mei et al. [27], qui présente une méthode basée sur l'extraction des caractéristiques CAEs (Convolutional Auto-Encoders) et spatiales (LBP et HOG), pour la classification en maligne et bénigne, un classificateur SVM est utilisé en recevant en entrées toutes les caractéristiques de l'image (CAE, LBP et HOG) et les caractéristiques de TIRADS, les résultats de l'expérience été validés sur 427 images, montrant que cette approche atteint une précision de 94%. La deuxième étude de l'année 2017 est la [5], l'objectif était de Développer un système complet de classification d'images échographiques thyroïdiennes basé sur un Modèle *GoogLeNet* raffiné et en utilisant l'algorithme d'apprentissage automatique "Cost-

*sensitiveRandom Forest".* Les expérimentations de cet article, soulignent que le prétraitement des échantillons d'image et leur augmentation en incluant à la fois les textures de nodule et les textures voisines peuvent améliorer efficacement les performances de réglage fin d'un DCNN. Le modèle proposé avait une précision de classification de 96,34%.

Dehua et al. [30] ont proposé une approche dans le même sens, ils ont utilisé un classifieur du texte échographique basé sur l'apprentissage profond pour prédire les nodules thyroïdiens bénins et malins. Cette méthode a donné des taux de précision de 93% et 95% à la fois sur la base d'images médicales réel et sur la base d'images standard UCI. Khachnaoui et al, ont développé dans [7] un système de DAO de classification des nodules thyroïdiens dans des images échographiques basé sur le fine-tuning d'une architecture de réseau résiduel avec 50 couches (ResNet-50), les résultats de leur expérience validé sur un ensemble de 814 images échographiques montrent que le 'raffinement' du ResNet-50 peut être utilisé efficacement pour classer les nodules thyroïdiens avec une précision de classification de 97,33%.

Les recherches ne se sont pas arrêtées là, en 2019 Sundar et al. [31], ont exploré diverses méthodologies de classification pour les comparer par la suite, les images échographiques de la thyroïde sont classées à l'aide d'un petit CNN entraîné à partir de zéro, de transfert Learning et d'affinement des modèles VGG-16 et Inception-v3, en tout ils ont généré 7 modèles. La première méthode consiste à entraîner le CNN en utilisant de petites données médicales, et une autre méthode se focalise sur l'apprentissage par transfert et fine-tuning des modèles Inception-v3 et VGG 16. Cependant pour la première méthode avec le petit CNN, les caractéristiques de la couche FC ont été classées en utilisant une fonction sigmoïde régulière dans les deux classes bénignes ou malignes pour la seconde méthode les caractéristiques de *bottleneck* de VGG-16 et d'Inception-v3 ont été obtenues et utilisées pour entraîner le CNN en premier, puis les utiliser comme entrée pour le classifieur SVM, pour les deux derniers modèles (le fine-tuning de VGG-16 et d'Inception-v3), ils ont utilisé le classifieur Softmax. Le tableau 2.1 résume tous les modèles générés ainsi que leur précision.

	Modèles proposés dans l'étude [31]	Accuracy (%)
1	Le CNN proposé dans l'étude [31] entraîné à partir de zéro	82
2	VGG-16 + CNN	79
3	Inception-v3 + CNN	93
4	VGG-16 + SVM	94
5	Inception-v3 + SVM	89
6	VGG-16 Fine-tuning	89
7	Inception-v3 Fine-tuning	79

Tableau 2. 1 : Tableau comparatif des modèles proposé par Sundar et al. [31].

On remarque que la combinaison d'un classifieur linéaire SVM avec un CNN VGG-16 était le meilleur modèle par rapport au reste avec une précision de 94 %.

Dans [44], Xiaowen et al. se sont concentrés sur le développement d'un système CAD multi-organes basé sur un modèle CNN pour classer les nodules thyroïdiens et mammaires (les caractéristiques de GoogLeNet et CaffeNet ont été exploitées dans la classification). Les auteurs ont étudié aussi l'impact de ce système sur l'efficacité de diagnostic par différentes méthodes. Les datasets d'apprentissage et de validation comprennent des images sélectionnées au hasard de nodules thyroïdiens et mammaires. Selon différentes méthodes de prétraitement (selon les méthodes de segmentation et de classification) les données ont été réparties en 4 modèles. Par comparaison avec les directives échographiques, un ensemble de données prospectives a été sélectionné pour vérifier la valeur clinique du modèle CNN. L'efficacité du diagnostic a été évaluée sur la base de courbes ROC. Les résultats de l'étude ont montré que parmi ces 4 modèles, le modèle CNN qui utilise des images segmentées pour la classification a donné les meilleurs résultats. Pour l'ensemble de validation, la sensibilité, la spécificité, la précision et l'aire sous la courbe (AUC) du modèle CNN proposé sont respectivement de 84,9%, 69,0%, 75,0% et 0,769%. Il n'y a pas de différence statistiquement significative entre le modèle CNN et le guide échographique. Par rapport à l'usage unique, la combinaison de ces

deux méthodes permet d'obtenir une efficacité diagnostique plus élevée. Cependant l'utilisation d'images segmentées et la classification selon la nature de la maladie est le principal facteur responsable de l'amélioration du modèle CNN. De plus, la combinaison du modèle CNN et du guide à ultrasons conduit à de meilleures performances de diagnostic, ce qui contribuera à améliorer l'efficacité du diagnostic du système CAO.

Corina et al. ont proposé dans [47] une approche qui combine deux modèles d'apprentissage profond pour la classification des images de nodules thyroïdiens, l'un basé sur le réseau de neurones convolutifs et l'autre basé sur l'apprentissage par transfert. Pour le premier modèle, appelé 5-CNN, ils ont développé un modèle entraîné de bout en bout efficace avec cinq couches convolutives, tandis que pour le deuxième modèle, l'architecture VGG-19 pré-entraînée a été réutilisée, optimisée et entraînée. Leurs modèles ont été entraînés et validés à l'aide d'un ensemble de données d'images échographiques composé de quatre types d'images thyroïdiennes : auto-immune, nodulaire, micro-nodulaire et normale. Les résultats ont montré que la méthode d'ensemble CNN-VGG a surpassé les modèles 5-CNN et VGG-19 avec 97,35% pour la précision globale du test, une spécificité globale de 98,43 % et sensibilité de 95,75 %. L'étude décrite ci-dessus était la première tentative d'utilisation des méthodes d'apprentissage profond pour analyser et classer plusieurs classes (auto-immunes, micronodulaires, nodulaires et normales) d'images thyroïdiennes.

## **2.3 Quelques travaux de classification des nodules thyroïdiens**

Dans cette section, nous présenterons en détail certaines méthodes de classification des nodules thyroïdiens orientées apprentissage profond citées dans la section précédente. Ce choix est basé d'une part sur les bons résultats, et d'autre part sur la disponibilité d'informations suffisamment détaillées sur leur travail.

### **2.3.1 Diverses méthodologies de classification basées sur les caractéristiques du goulot d'étranglement du VGG16 et Inception-v3 :**

Dans l'article [31], les auteurs ont exploré diverses méthodologies de classification sur une base d'images de petite taille, vu que la disponibilité des données médicales est limitée. Les images sont classées ensuite à l'aide d'un petit CNN entraîné à partir de zéro, d'un apprentissage par transfert et du réglage fin de modèles Inception-v3 et VGG-16.

La première méthode consistait à entraîner un CNN à partir de zéro en utilisant les données médicales. L'architecture des couches est constituée de 3 couches convolutionnelles avec des noyaux  $3 \times 3$  de nombres 32, 32 et 64 respectivement.

La deuxième méthode consistait en premier lieu à obtenir les caractéristiques de goulot d'étranglement des VGG-16 et Inception-v3 puis entraînées sur le CNN à partir de la méthode précédente. Dans VGG-16, les trois dernières couches FC ont été supprimées et le modèle CNN utilisé dans la première méthode a été alimenté par ces caractéristiques. Les poids pré-entraînés sur ImageNet ont été chargés dans les modèles et après le passage vers l'avant de l'image à travers les caractéristiques de goulot d'étranglement du réseau ont été enregistrées.

Par la suite, le classifieur de CNN a été remplacé par un classifieur SVM linéaire qui est alimenté par les caractéristiques de goulot d'étranglement pour la classification. Les CNN profonds sont connus pour être d'excellents extracteurs d'entités et l'utilisation d'un classificateur linéaire pour utiliser ces caractéristiques s'avère être un excellent moyen de s'attaquer à des ensembles de données de petite taille.

La troisième méthode, porte sur le réglage fin du modèle Inception-v3. Les auteurs ont obtenu les points de terminaison "pré-logits" qui sont une couche antérieure à la dernière couche de l'architecture et la couche FC personnalisée qui porte sur un classifieur binaire de type Softmax.

Pour un réglage fin du VGG-16, les trois dernières couches FC qui contribuent à d'énormes calculs ont été supprimées et une nouvelle couche FC a été attachée après la couche «pool-5», toutes les couches au-dessus de la conv 5\_2 ont été gelées. Donc, essentiellement les trois dernières couches (à l'exclusion des couches FC) et la nouvelle couche FC personnalisée ont été formées à cette approche.

### **2.3.2 Classification des images échographiques de la thyroïde basée sur le modèle affiné GoogLeNet**

L'étude de [5] présente un système de diagnostic assisté par ordinateur (DAO) pour classer les nodules thyroïdiens dans les images échographiques. Les chercheurs ont utilisé des méthodes d'apprentissage profond pour extraire les caractéristiques des images échographiques de la thyroïde. Les images échographiques sont d'abord prétraitées puis utiliser pour affiner le modèle GoogLeNet pré-entraîné afin d'obtenir une excellente extraction de caractéristiques. Les caractéristiques extraites des images échographiques de la

thyroïde sont envoyées à un classificateur de forêt aléatoire sensible aux coûts pour classer les images comme cas malins et bénins.

Les auteurs ont utilisé un processus de réglage fin de GoogLeNet. Les nouveaux exemples d'images sont envoyés au GoogLeNet pré-entraîné, modifiant les trois couches de classification (de 1000 à 2) pour n'apprendre que deux sorties correspondantes aux deux classes, c'est-à-dire bénignes ou malignes. Ces couches ont été renommées pour qu'au lieu de choisir les poids dans le modèle GoogLeNet pré-entraîné, les couches puissent commencer à s'entraîner avec des poids aléatoires. Après avoir affiné GoogLeNet, les caractéristiques des images ont d'abord été extraites puis transmises à un classifieur supervisé, qui les a classés comme bénignes ou malignes, en utilisant l'algorithme d'apprentissage automatique "Cost-sensitive Random Forest".

## **2.4 Comparatif des approches existantes de classification binaire des nodules thyroïdiens**

Le tableau ci-dessous, fourni uniquement à titre indicatif, montre quelques résultats obtenus pour la sous-tâche de classification des images de nodules thyroïdiens bénins et malins. La précision de classification a été utilisée comme mesure d'évaluation sur l'ensemble d'images de test. Les résultats sont triés dans l'ordre décroissant.

Méthode	Modèle	Test	Accuracy	Description de la base de données
Chi et al. [5] (2017)	Modèle affiné GoogLeNet + Classificateur de forêt aléatoire sensible aux coûts	693	99.13%	BASE public [4] et BASE locale
Khachnaoui et al. [7] (2018)	ResNet-50	814	97.33%	DDTI (Digital Database Thyroid Image) [49] et Ultrasoundcases.info [50].
Sundar et al. [31] (2019)	VGG-16 + SVM	613	94%	BASE public [4] et BASE locale de GE LOGIQ P9

Koh et al. [3] (2020)	Ensemble AlexNet- GoogLeNetSqueezeNet- InceptionResNetv2 (CNNE2).	200	85%	BASE locale de 4 hôpitaux : Severance Hospital, Samsung Medical Center, Bundang Medical Center et Kyung Hee Hospital.
--------------------------	---	-----	-----	---

Tableau 2. 2 : Tableau comparatif des modèles.

## 2.5 Conclusion

À travers cette revue bibliographique, nous constatons que la classification des nodules thyroïdiens bénins et malins suscite beaucoup d'intérêt au sein de la communauté scientifique. La technique classique, en utilisant l'apprentissage automatique, a apporté les premières solutions pour le problème de classification des nodules en utilisant les images échographiques. Le seul inconvénient réside dans la nécessité d'une expertise humaine pour le choix d'un jeu de caractéristiques discriminantes, cependant les chercheurs se sont penchés vers une autre méthode qui porte sur l'apprentissage profond qui a surmonté ce problème en se basant sur la puissance des réseaux de neurones convolutifs pour extraire de façon automatique les caractéristiques trouvées sur les images échographiques.

Dans la dernière partie, nous avons mis en revue avec détails quelques travaux récents des chercheurs dans le domaine de la classification des images de nodules thyroïdiens bénins et malins basé sur l'apprentissage profond et particulièrement le réseau CNN. Les résultats obtenus ont été intéressants. Les expériences de l'étude [5] sur l'apprentissage profond, en particulier le réglage fin d'un réseau d'apprentissage profond existant, soulignent l'avantage de nécessiter moins d'échantillons d'entraînement ce qui est en notre faveur vu le manque d'imageries médicales et l'avantage d'améliorer les performances de l'évaluation des nodules thyroïdiens en tant que système de diagnostic assisté par ordinateur. Tandis que l'étude [31], la combinaison de classificateurs linéaires comme le SVM à l'architecture (en particulier le VGG-16) s'est avéré être utile lorsque les réseaux de neurones profonds sont entraînés sur des ensembles de données très différents du domaine en question, et à notre connaissance aucune des études proposées jusqu'à présent n'a eu recours aux niveaux de transfert qui peuvent être définis comme un transfert progressif. Nous présenterons notre contribution dans le cadre de ce travail dans le chapitre suivant.

## Chapitre 3 : Contribution

Dans le cadre de notre recherche, nous avons traité le cas de la classification des lésions thyroïdiennes. Ce traitement permet d'extraire la nature de la lésion qui s'exprime en maligne ou bien bénigne (cas d'une classification binaire). Les données d'entrée que nous avons utilisées sont des images échographiques extraites de trois data-sets [65, 66, 81] pour entraîner et tester notre modèle de DL. Cependant le choix de ce type d'imagerie médicale est dicté par le fait que l'échographie est la modalité la plus sensible et la plus fondamentale pour le diagnostic des nodules thyroïdiens comme nous avons déjà discuté au chapitre 1 et 2, elle est utilisée pour évaluer la structure de la thyroïde. C'est un complément précieux au test clinique pour déterminer la taille et l'anatomie de la thyroïde et pour identifier les nodules. De plus, le développement rapide des techniques d'échographie a déterminé leur utilisation comme méthode alternative pour le diagnostic et le suivi des nodules thyroïdiens en raison de leurs caractéristiques non invasives en temps réel. Ce présent chapitre résumera notre contribution en décrivant l'approche utilisée.

### 3.1 Méthodologie

Dans ce travail, nous explorons les différents modèles d'apprentissage par transfert pour la classification d'un ensemble de données d'images échographiques de la thyroïde comme malins ou bénins à l'aide des différentes architectures d'apprentissage profond telles que les réseaux de neurones convolutifs, ainsi, on observera les limites de chaque architecture.

Nous étudions dans un premier lieu la capacité de transfert des caractéristiques apprises sur la base d'images ImageNet contenant plus de 14 millions images non médicales aux images échographiques de la thyroïde constituant un ensemble de données de taille assez petite. Par conséquent, nous avons choisi d'explorer les niveaux de transfert qui peuvent être définis comme un transfert progressif constitué d'un transfert des caractéristiques apprises sur la base d'images ImageNet aux images radiographiques thoraciques pour les cas positifs au COVID-19 ainsi que des images de pneumonie normale et virale composant 20 000 images puis un transfert des caractéristiques apprises sur cette dernière base de données COVID-19 aux images échographiques de la thyroïde. Ce type de transfert peut être utile lorsque les réseaux de neurones profonds sont pré-entraînés sur des ensembles de données très différents du domaine en question et lorsque nous disposons d'une base de données assez petite en raison du manque de données médicales pour entraîner un algorithme d'apprentissage profond à partir de zéro. D'ailleurs, notre base de données d'entraînement est constituée d'images

échographiques de la thyroïde étiquetées comme négatives (bénignes) et positives (malignes), voir la Section 3.1. L'exécution d'un algorithme d'apprentissage profond avec des millions de paramètres et moins de données entrave la formation avec un sur-ajustement. Le modèle a donc tendance à échouer lors de la généralisation de l'apprentissage, ce qui donne une faible précision sur l'ensemble de données de test. La régularisation peut réduire la variance élevée dans une certaine mesure, mais la formation d'un cadre d'apprentissage profond à partir de zéro reste interdite. L'augmentation des données peut être utilisée pour accroître davantage la taille de l'ensemble de données (Section 3.2). Un prétraitement (Section 3.3) est appliqué sur l'ensemble de ces images en calibrant l'échelle de toutes les images pour avoir la même taille.

Pour atteindre notre objectif nous avons utilisé trois architectures VGG-16, ResNet-50V2 et EfficientNet pré-entraînées, en premier lieu sur les jeux de données ImageNet, affinés selon les différentes stratégies de réglage fin à l'aide de nos échantillons d'images, et en second lieu sur l'ensemble de données radiographiques, également affiné, par la suite les évaluer selon des mesures basées sur la précision, la spécificité, la sensibilité, le score F1 et la courbe ROC en comparant leurs performances diagnostiques.

### 3.2 Approche proposée

L'apprentissage profond nécessite un volume de données très important comme déjà cité, qui sont difficiles à obtenir avec l'imagerie médicale actuelle. Pour cette raison, nous avons utilisé un apprentissage par transfert « progressif » qui adopte un CNN pré-entraîné avec une énorme quantité d'images médicales et l'avons appliqué. C'est pourquoi, nous avons proposé dans ce travail d'utiliser et d'affiner trois modèles CNN pré-entraînés (VGG-16, ResNet-50V2 et EfficientNet):

- en premier lieu sur les jeux de données ImageNet, comme extracteur de caractéristiques et affinées selon les deux stratégies (2<sup>ème</sup> et 3<sup>ème</sup> stratégies) de réglage fin mentionnée au chapitre 1 en utilisant un ensemble de données contenant 795 images échographiques de la thyroïde,

- et en second lieu pré-entraînées et affiné sur une base d'images radiographiques.

Pour la classification maligne et bénigne, nous avons utilisé le *classificateur personnalisé supervisé softmax*, à l'exception du modèle VGG-16, nous avons utilisé 3 types de classificateurs *SVM*, *random forest* et un *classificateur personnalisé supervisé softmax*. La méthode proposée est réalisée en deux étapes, la première étant d'extraire les caractéristiques des images thyroïdiennes ; telles que ces dernières soient passées comme entrées des 3 réseaux

neuronaux, pour en sortir un vecteur de caractéristiques qui dépend de ces images. La seconde étape consiste à passer ce vecteur comme entrée de l'un des trois classificateurs. En résumé, cette stratégie consiste à passer en entrée pour chaque architecture des images échographiques prétraitées de nodules thyroïdiennes, puis générer des prédictions (malignes et bénignes) en sortie, les figures 3.1, 3.2 et 3.3 montrent l'organigramme suivi pour chaque architecture.

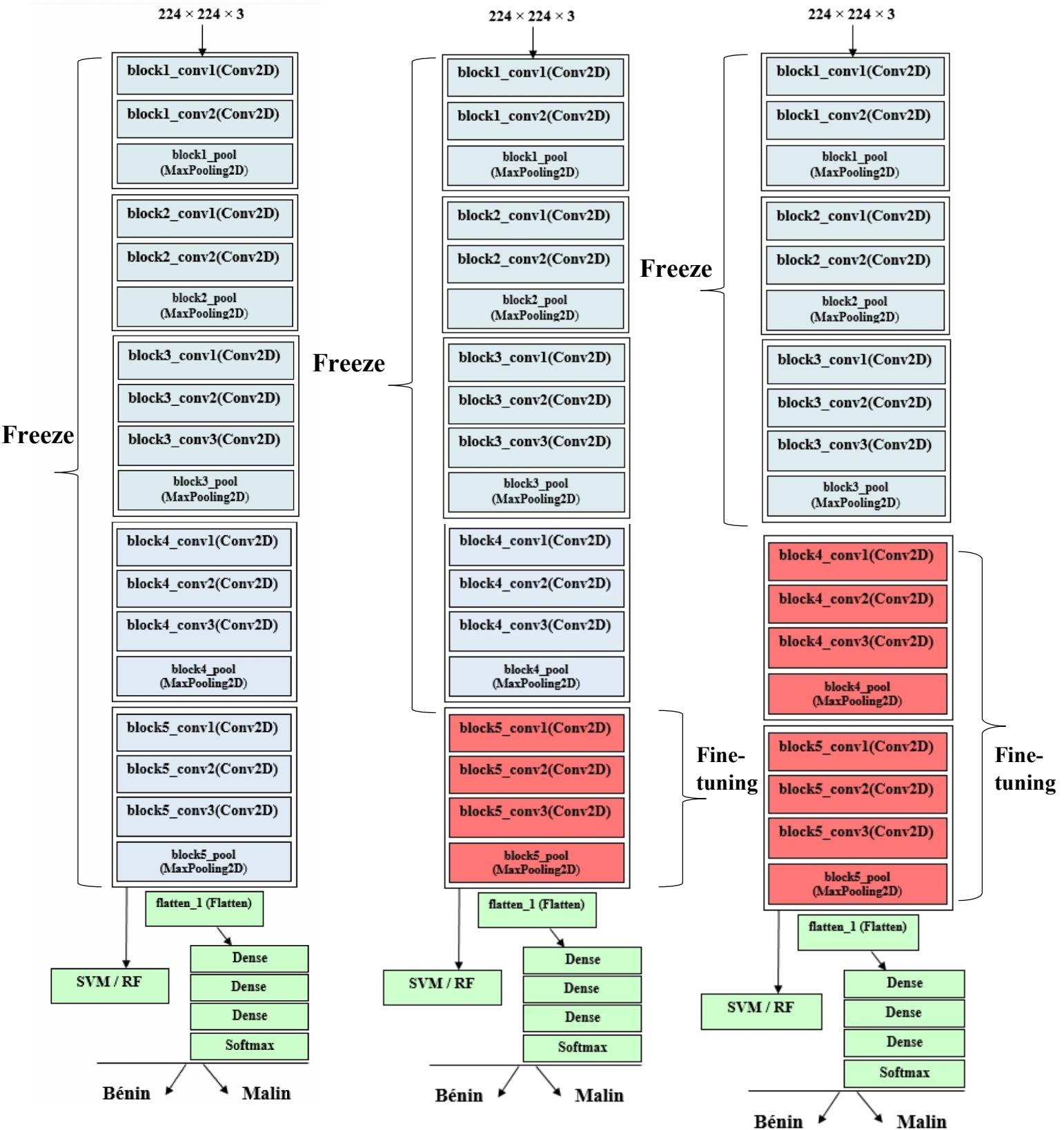


Figure 3. 1 : Organigramme du modèle de prédiction des nodules thyroïdiens VGG-16 basé sur les différents niveaux de Fine-tuning.

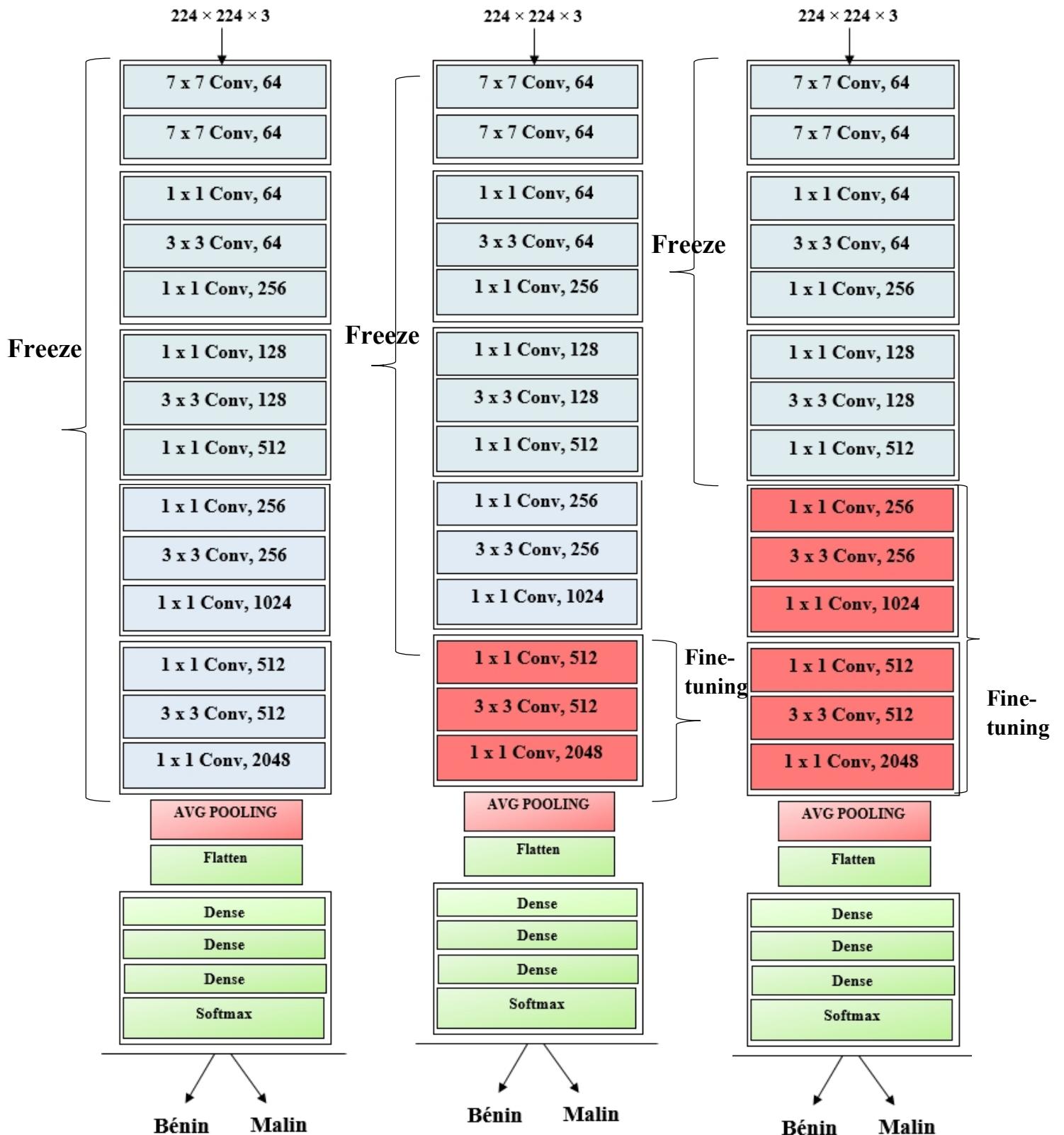


Figure 3. 2 : Organigramme du modèle de prédiction des nodules thyroïdiens ResNet-50V2 basé sur les différents niveaux de Fine-tuning.

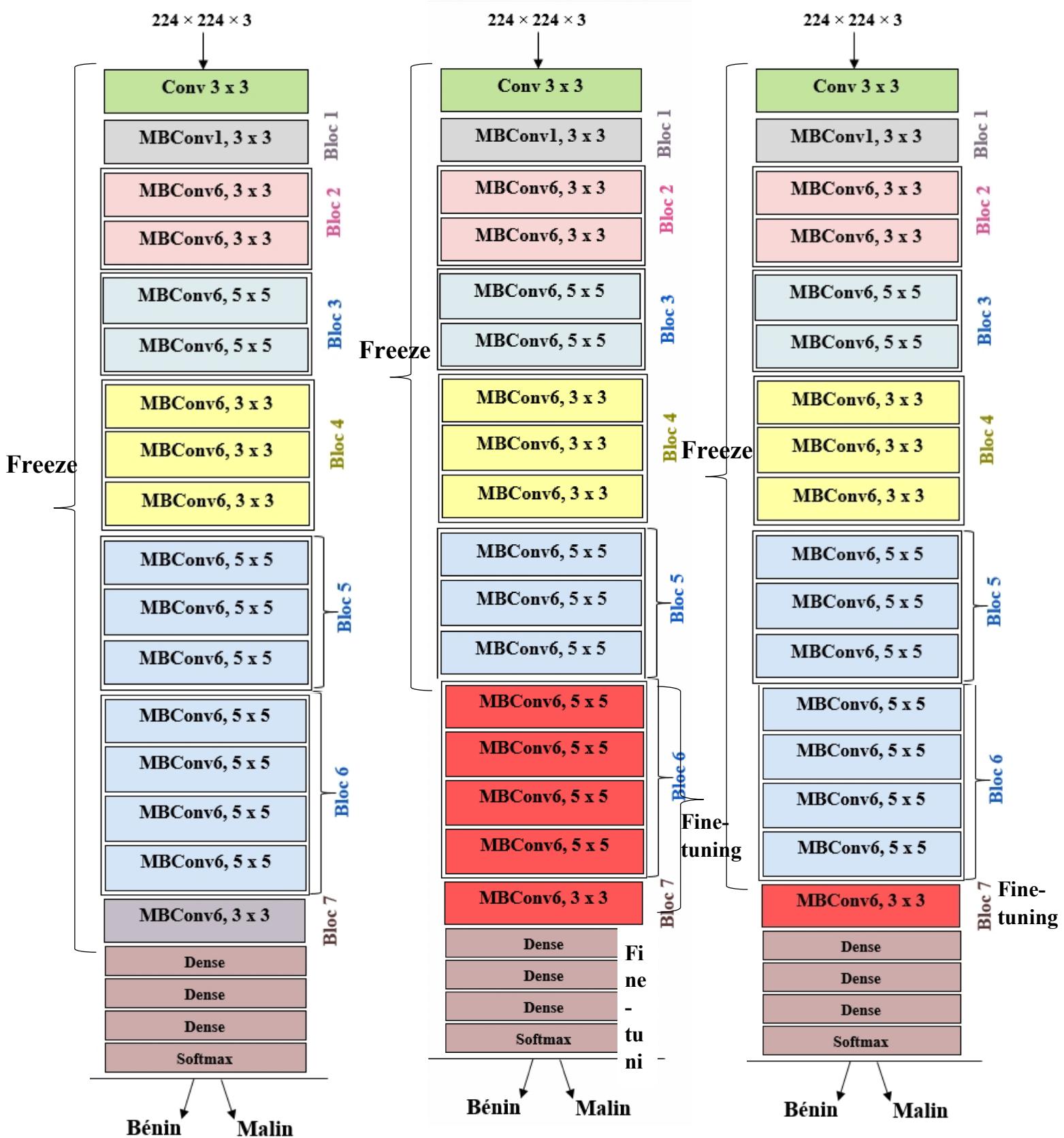


Figure 3. 3 : Organigramme du modèle de prédiction des nodules thyroïdiens EfficientNet basé sur les différents niveaux de Fine-tuning.

### 3.3 Bases d'images utilisées

L'utilisation d'une base d'images bien définie est la clé de la recherche sur la détection/classification des nodules thyroïdiens. Pour cela, nous avons réalisé l'entraînement en utilisant trois bases de données publiques d'images échographiques thyroïdiennes : DDTI (Digital Database Thyroid Image) [49], Ultrasoundcases.info [50] et thyroid.xiaobaishiji.com [51].

**La première base :** DDTI, il s'agit d'une base de données en accès libre, d'images échographiques de nodules thyroïdiens. Elle comprend 349 images échographiques thyroïdiennes, de taille  $560 \times 360$ , dont 289 sont étiquetées comme malignes et confirmées par une procédure de biopsie, tandis que 60 images sont étiquetées comme bénignes.

**La deuxième base :** Elle est accessible publiquement et constituée d'un ensemble de données d'images échographiques contenant un grand nombre de cas d'échographie générale. Cette base de données contient 303 images échographiques thyroïdiennes dont 170 images qui sont étiquetées comme malignes et 133 qui sont étiquetées comme bénignes.

**La troisième base :** Est un ensemble de données d'images échographiques accessible publiquement sur la plate-forme cloud de diagnostic d'intelligence artificielle des nodules thyroïdiens. Nous avons pu télécharger 177 images échographiques thyroïdiennes, dont 158 sont étiquetées comme malignes, tandis que 96 images sont étiquetées comme bénignes. Nous avons fusionné les trois bases de données pour obtenir un total de 795 d'images échographiques de nodules thyroïdiens dont 526 (80%) sont malignes et 269 (20%) sont bénignes, afin de garantir suffisamment d'échantillons d'apprentissage pour chaque classe.

La répartition des images pour le diagnostic est décrite dans le tableau 3.1 :

Base d'images	Entrainement	Validation	Test	Total d'échantillon
Les 3 BASES : 1ère base: DDTI + 2ème base: Ultrasoundcases.info + 3ème base :	446	190	159	795

thyroid.xiaobaishiji.com				
--------------------------	--	--	--	--

Tableau 3. 1: Répartition des images dans les ensembles de données d'entraînement, validation et de test.

La figure 3.4 Exemples des images échographiques de la base de données DDTI, Ultrasoundcases.info et thyroid.xiaobaishiji.com.

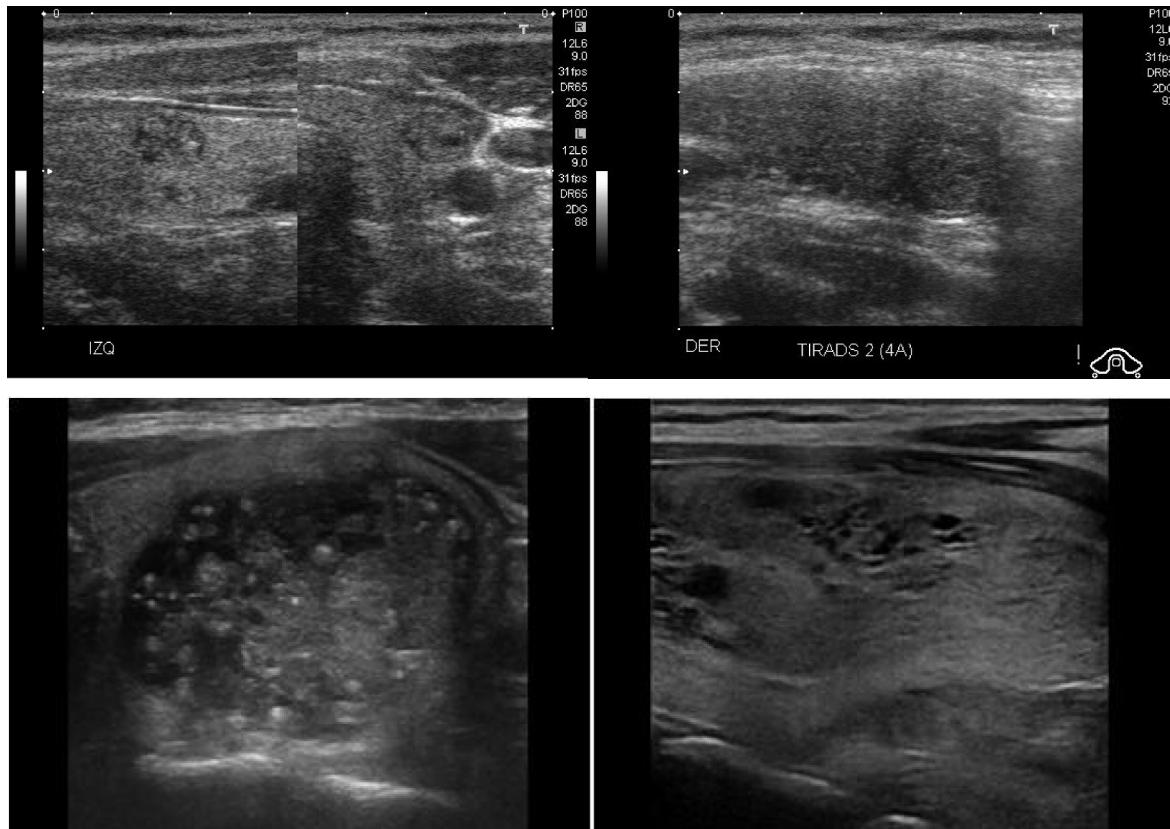


Figure 3. 4 : Exemples d'images échographiques de la base de données DDTI, Ultrasoundcases.info et thyroid.xiaobaishiji.com.

### 3.4 Augmentation de données

Le total de 795 d'images échographiques de nodules thyroïdiens fusionnées ne suffit pas par rapport à l'exigence de quantité d'images pour affiner un DCNN. Nous avons donc appliqué les méthodes d'augmentation les plus courants telles que le redimensionnement, la rotation et l'ajustement du contraste aux données dans chaque groupe d'images. Le tableau 3.2 montre la distribution des échantillons dans les groupes d'entraînement et de validation des trois bases de données fusionnées.

	Total	
	Cas	Echantillons augmentés
Entrainement	446	757
Validation	190	325

Tableau 3. 2 : La répartition des cas dans les groupes d'entraînement et de validation des trois bases de données.

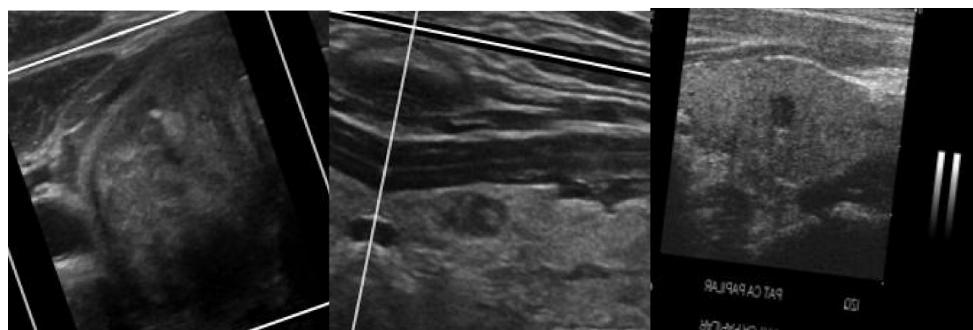


Figure 3. 5: Exemples d'images transformées par les méthodes d'augmentation de données.

### 3.5 Préparation des images de la base

Les images des trois bases sont normalisées à une taille identique et adéquate puis utilisées comme images d'entraînement et de validation pour affiner directement le DCNN.

#### 3.5.1 Collecte de données pour le Transfert progressif d'apprentissage

##### 3.5.1.1 Base de données sur la radiographie thoracique COVID19

Une équipe de chercheurs de l'Université de Doha au Qatar et de l'Université de Dhaka au Bangladesh, ainsi que des collaborateurs et des médecins du Pakistan et de Malaisie, ont créé une base de données d'images radiologiques thoraciques de cas positifs au COVID-19 et de pneumonie normale et virale au format de fichier Portable Network

Graphics (PNG) et résolution de 299\*299 pixels constituant un ensemble de données de 20 022 images. Cet ensemble de données COVID-19, normales et autres infections pulmonaires est publié par phases. Dans la première version, ils ont publié 219 images COVID-19, 1341 images normales et 1345 radiographies pulmonaires (CXR) de pneumonie virale. Dans la première mise à jour, ils ont augmenté la classe COVID-19 à 1200 images CXR. Dans la deuxième mise à jour, ils ont augmenté la base de données à 3616 cas positifs au COVID-19 ainsi que 10 192 images normales, 6012 images d'opacité pulmonaire (infection pulmonaire non COVID) et 1345 images de pneumonie virale. Cette base de données est enrichie au fur et à mesure que d'autres images radiographiques pour les patients atteints de pneumonie COVID-19 sont disponibles [70, 71].



Figure 3. 6 : Exemples des images radiographiques thoraciques COVID-19.

### 3.6 Sélection des hyper-paramètres de l'approche proposée

L'ajustement des hyper-paramètres ou *L'hyperparameter tuning*, est une étape clé de l'apprentissage automatique. Le fonctionnement normal des algorithmes ML dépend de leur optimisation. Les hyper-paramètres sont des paramètres de réglage des algorithmes d'apprentissage profond. Leurs valeurs ne sont pas optimisées généralement, c'est au data scientiste de les choisir. En fait, sans ces hyper-paramètres, les algorithmes du DL ne peuvent pas fonctionner. Leur réglage (tuning) nécessite un gros travail statistique pour déterminer les hyper-paramètres pouvant fournir les meilleurs résultats.

Il existe plusieurs méthodes, dont la recherche par grille : construire une liste de possibilités pour chaque hyper-paramètre. Le modèle sera entraîné pour chaque combinaison et son score sera calculé à chaque fois. La combinaison qui obtient le score le plus élevé sera l'hyper-paramètre sélectionné.

Pour notre cas nous avons choisi empiriquement 50 % d'images de notre ensemble de données comme ensemble d'entraînement, 30% d'images de notre ensemble d'entraînement comme ensemble de validation, tandis que le reste (20%) sont utilisées comme ensemble de validation, nous avons formé par la suite nos modèles avec un taux d'apprentissage égal à  $10^{-5}$  afin d'éviter la distorsion rapide des poids du modèle préformé.

### **3.6.1 Paramétrage de classifieur SVM**

Les SVM sont connus par leur séparation des points de données appartenant à différentes classes avec une limite de décision. Lors de la détermination de la limite de décision, la marge souple SVM tente de résoudre le problème d'optimisation en augmentant la distance entre la limite de décision et les classes, en maximisant le nombre de points correctement classés dans l'ensemble d'entraînement. L'influence des hyper-paramètres n'affecte le noyau linéaire que dans une certaine mesure, pour les noyaux non linéaires, l'influence des hyper-paramètres devient plus évidente. L'un des noyaux non linéaires les plus couramment utilisés est le noyau gaussien (RBF), par exemple le paramètre gamma de RBF contrôle la distance de l'influence d'un seul point d'entraînement.

En ce qui concerne ce travail, nous avons choisi de paramétriser le classifieur SVM employé avec un noyau gaussien de paramètre gamma =  $10^{-3}$ .

### **3.6.2 Hyper-paramètres RF**

La forêt aléatoire est un algorithme des problématiques de prédiction, comme déjà cité. Il est particulièrement apprécié lorsqu'il existe un grand nombre de variables explicatives. L'ajustement et l'optimisation des hyper-paramètres de cet algorithme complexe consistent notamment à choisir le nombre d'arbres qu'il faut créer et le nombre de variables à être utilisé pour chaque division d'un nœud. La recherche par grille peut être utilisée pour tester une gamme de paramètres, comparer les résultats et déterminer les meilleurs paramètres. Pour notre cas on a choisi un nombre d'arbre égal à 100.

### **3.7 Construction du modèle profond**

#### **3.7.1 L'architecture VGG-16 / ResNet-50 /EfficientNet +classifieur personnalisé supervisé Softmax**

Les étapes de l'algorithme pour définir et entraîner les différentes architectures combinées au classifieur personnalisé supervisé Softmax ont été décrites dans l'algorithme 1 :

---

***Algorithme 1.*** La description des modèles[VGG-16 / ResNet-50 / EfficientNet] + classifieur personnalisé supervisé softmax

---

Entrée : Image de la thyroïde de taille 560 pixels et 360 pixels obtenue à partir de l'ensemble de données du train.

Sortie : Probabilités de prédiction pour chaque classe de diagnostic (malin, bénin).

1. pour chaque image de l'ensemble de données
  2. Redimensionner l'image à 224 px et 224 px.
  3. Utiliser un processing spécifique au VGG16 au lieu de normaliser les valeurs des pixels de l'image entre [0, 1] avec un rescale 1. /255.
  4. fin
  5. Charger les modèles VGG-16 / ResNet-50 / EfficientNet pré-entraînés sans le classifieur final.
  6. Ajouter une couche Flatten sur la sortie du modèle pour obtenir un tableau 1D d'entités.
  7. Ajouter trois couches entièrement connectées de 128 neurones.
  8. Ajouter une couche de régularisation pour éviter le sur-ajustement.
  9. Ajouter une couche entièrement connectée avec deux unités cachées et une fonction d'activation softmax.
  10. Appliquer les 3 niveaux de fine-tuning selon le modèle.
  11. Compiler et optimiser le modèle avec l'optimiseur Adam avec un taux d'apprentissage égale à  $10^{-5}$ .
-

12. Entraîner le modèle pour 80 époques.

13. Sauvegarder le modèle final.

14. Évaluer le modèle.

---

### **3.7.2 L'architecture VGG-16+le classifieur (Random Forest / SVM)**

Les étapes de l'algorithme pour définir et entraîner VGG-16 + le classifieur (RF / SVM) ont été décrites dans l'algorithme 2 :

---

#### ***Algorithme 2. La description du modèle VGG-16 + RF / SVM***

---

Entrée : Image de la thyroïde de taille 560 pixels et 360 pixels obtenue à partir de l'ensemble de données du train.

Sortie : Probabilités de prédiction pour chaque classe de diagnostic (malin, bénin).

1. Pour chaque image de l'ensemble de données.

2. Redimensionner l'image à 224px et 224 px.

3. Normaliser les valeurs des pixels de l'image entre [0, 1] en divisant chaque valeur de pixel par 255.

4. Appliquer un codage à chaud unique pour les variables de sortie.

5. fin

6. Charger le modèle VGG-16 pré-entraîné sans le classificateur final.

7. Appliquer les 3 niveaux de fine-tuning.

8. Ajouter une couche Flatten sur la sortie du modèle pour obtenir un tableau 1D d'entités.

9. Compiler et Optimiser le modèle avec l'optimiseur Adam avec un taux d'apprentissage égale à  $10^{-5}$ .

10. Entraîner le modèle pour 50 époques.

---

11. Faire passer la sortie du modèle en entrée du classifieur RF / SVM.

12. Entraîner le classifieur (SVM / RF).

13. Évaluer le modèle.

---

### 3.8 Métriques d'évaluation de la performance de la classification

Une fois les différentes caractéristiques extraites et la classification réalisée, il existe diverses techniques (métriques) permettant de vérifier la fiabilité et la pertinence de l'approche de classification utilisée.

Les métriques d'évaluation les plus utilisées pour les modèles de classification binaire sont les suivantes : *Matrice de confusion*, *Accuracy*, Précision (*Precision*), Rappel (*Recall*) et *F1-score*.

#### 3.8.1 Matrice de confusion

*La matrice de confusion ou matrice d'erreur* est l'un des concepts clés lorsque nous parlons de problèmes de classification. Cette matrice est un tableau à deux dimensions (« réelle » et « prédictive ») et des ensembles de « classes » dans les deux dimensions. Nos classifications réelles sont des colonnes et celles prédictives sont des lignes comme le montre le tableau ci-dessous :

		Réelle	
		Positives (1)	Négatives (0)
Prédite	Positives (1)	TP	FP
	Négatives (0)	FN	TN

Tableau 3. 3 : Matrice de confusion

Presque toutes les mesures de performance sont basées sur la matrice de confusion et les chiffres qu'elle contient.

- **Termes associés à la matrice de confusion :**
- **Vrais positifs (TP)** : les vrais positifs sont les cas où la classe réelle de la donnée était 1 (True) et la prédiction est également 1 (True).
- **Vrais négatifs (TN)** : les vrais négatifs sont les cas où la classe réelle de la donnée était 0 (Faux) et la prédiction est également 0 (Faux).
- **Faux positifs (FP)** : les faux positifs sont les cas où la classe réelle de la donnée était 0 (False) et la prédiction est 1 (True). « Faux » est dû au fait que le modèle a prédit de manière incorrecte et « positive » parce que la classe prédictive était positive (1).
- **Faux négatifs (FN)** : les faux négatifs sont les cas où la classe réelle de donnée était à 1 (Vrai) et la classe prédictive est à 0 (Faux). « Faux » est dû au fait que le modèle a prédit de manière incorrecte et « négative » parce que la classe prédictive était négative (0).

### 3.8.2 Accuracy

L'Accuracy est une mesure d'évaluation courante pour les problèmes de classification. Elle est calculée comme un nombre de prédictions correctes divisé par le nombre total d'échantillons. Lorsque nous le multiplions par 100, nous obtenons une précision en pourcentage. L'Accuracy est une bonne mesure lorsque les classes de variables cibles dans les données sont presque équilibrées.

$$\text{Accuracy} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total d'échantillons}} = \frac{(T_{\text{P}} + T_{\text{N}})}{(T_{\text{P}} + T_{\text{N}} + F_{\text{P}} + F_{\text{N}})}$$

### 3.8.3 Précision

La précision est une métrique très utile et elle contient plus d'informations que l'Accuracy. Essentiellement, avec précision, nous répondons à la question : « Lorsque le modèle prédit positif, à quelle fréquence est-il correct? ». Elle est calculée pour chaque classe séparément avec la formule :  $P = \frac{T_{\text{P}}}{T_{\text{P}} + F_{\text{P}}}$

Et sa valeur peut aller de 0 à 1.

### 3.8.4 Rappel

Rappel (Recall) peut être décrit comme la capacité du classificateur à trouver tous les échantillons positifs. Avec cette métrique, nous essayons de répondre à la question : "Quelle proportion de positifs réels a été correctement identifiée ?" Il est défini comme la fraction

d'échantillons d'une classe spécifique qui sont correctement prédicts par le modèle ou mathématiquement :

$$R = \frac{T}{T + F}$$

**Différence entre Précision et Recall dans le cadre de classification :** Recall nous donne des informations sur les performances d'un classifieur par rapport aux faux négatifs (combien en avons-nous manqué), tandis que la précision nous donne des informations sur ses performances par rapport aux faux positifs (combien en avons-nous attrapés).

### 3.8.5 Sensibilité

La sensibilité est une mesure de la proportion de cas positifs réels qui ont été prédicts comme positifs (ou vrais positifs). La sensibilité est également appelée rappel « **Recall** ».

### 3.8.6 Spécificité

Elle est définie comme la proportion de négatifs réels, qui ont été prédicts comme négatifs (ou vrais négatifs). Elle peut être calculée à l'aide de l'équation :

$$\text{Spécificité} = \frac{T}{T + F}$$

### 3.8.7 F-mesure (F-score ou F1-score)

Moyenne harmonique de la précision et du rappel. Mesure la capacité du système à donner toutes les solutions pertinentes et à refuser les autres.

$$\text{F-mesure} = \frac{2P \times R}{P + R}$$

## Conclusion

Dans ce chapitre, nous avons décrit la contribution de notre travail et présenté les choix que nous avons pris pour explorer les diverses méthodologies de la classification d'un ensemble de données d'images échographiques de la thyroïde comme malins ou bénins à l'aide des architectures d'apprentissage profond.

Dans le prochain chapitre, nous analyserons et discuterons les résultats obtenus par nos modèles de classification, en évaluant et comparant leurs performances diagnostiques

pour différencier les nodules malins des nodules bénins et pour confirmer nos hypothèses : Recourir au transfert d'apprentissage pour réduire la nécessité d'avoir un grand nombre d'images échographiques de la thyroïde, transfert d'apprentissage progressif, réglage fin des architectures et fusion de classificateurs linéaires comme le SVM ou RF à l'architecture (en particulier le VGG-16), afin d'essayer de surpasser le problème des réseaux de neurones profonds pré-entraînés sur des ensembles de données très différents du domaine.

## **Chapitre 4 : Expérimentations, résultats et discussion**

Dans ce chapitre nous détaillerons les différentes expérimentations menées et nous discuterons les résultats d'évaluation obtenus de la classification d'une base d'images échographiques de la thyroïde comme malins ou bénins en présentant les outils utilisés. Une analyse des performances des cinq modèles pré-entraînés (VGG-16 / ResNet-50 / EfficientNet + classificateur personnalisé supervisé Softmax et VGG-16 + le classifieur (Random Forest / SVM)) sera faite à travers les études comparatives suivantes :

- 1- Classification avec et sans augmentation des images avec les cinq modèles pré-entraînés avec un transfert d'apprentissage sur la base d'images ImageNet et affinés avec la 2<sup>ème</sup> et 3<sup>ème</sup> stratégie du réglage fin.
- 2- Classification avec et sans augmentation des images avec les cinq modèles pré-entraînés avec un transfert d'apprentissage progressif sur la base d'images radiographique et affinés avec les deux stratégies (2<sup>ème</sup> et 3<sup>ème</sup> stratégies) du réglage fin.

L'évaluation des performances d'un modèle est un élément essentiel de la création d'un modèle d'apprentissage profond efficace. L'évaluation de performance des modèles proposés est basée sur les valeurs :

- De la précision de classification (elle mesure le pourcentage de diagnostic correct des nodules thyroïdiens).
- De la spécificité (elle mesure le pourcentage de nodules bénins correctement identifiés comme bénins).
- De la sensibilité ou rappel (elle mesure le pourcentage de nodules malins correctement identifiés comme malins).
- Score F1 (il mesure la moyenne harmonique des valeurs de précision et de rappel)
- La courbe des caractéristiques de l'opérateur du récepteur (ROC) (elle est une mesure d'évaluation, elle trace le taux vrai positif par rapport au taux faux positif à diverses valeurs de seuil).

## 4.1 Outils utilisés

Pour l'implémentation du système, plusieurs bibliothèques et technologies ont été nécessaires.

### 4.1.1 Plateforme TensorFlow

Nous avons utilisé dans notre travail **TensorFlow** qui est une bibliothèque open source pour le calcul numérique et l'apprentissage automatique à grande échelle. Crée par l'équipe Google Brain, TensorFlow regroupe une multitude de modèles et d'algorithmes d'apprentissage automatique et d'apprentissage profond (alias réseau neuronal) et les rend utiles au moyen d'une métaphore commune.



Figure 4. 1 : Logo TensorFlow.

Certaines des caractéristiques principales de TensorFlow sont décrites ci-dessous :

- **Prend en charge la création de modèles** : il propose différents niveaux de concepts pour former et créer des modèles.
- **Robuste et indépendant de la plateforme** : Il permet de déployer nos applications très facilement, quel que soit le langage ou la plateforme que nous utilisons.
- **Recherche efficace** : TensorFlow nous permet de contrôler l'API de sous classement de modèles et l'API fonctionnelle **Keras** pour créer des topologies difficiles.

Keras , d'autre part, est une interface de programmation d'applications (API) de réseau neuronal de haut niveau efficace écrite en Python qui s'exécute sur TensorFlow, CNTK et Theano. Cette bibliothèque de réseaux de neurones open source est conçue pour fournir une expérimentation rapide avec les réseaux de neurones profonds.

Keras se concentre sur la modularité, la convivialité et l'extensibilité. Il ne gère pas les calculs de bas niveau ; au lieu de cela, il les transmet à une autre bibliothèque appelée Backend. Keras a été adopté et intégré à TensorFlow à la mi-2017. Les utilisateurs peuvent y accéder via le module `tf.keras`. Cependant, la bibliothèque Keras peut toujours fonctionner séparément et indépendamment.



Figure 4. 2 : Logo Keras.

#### 4.1.2 Environnements virtuels dans Jupyter Notebook

**Google Colab** : Google Colab ou Colaboratory est un environnement de notebook Jupyter gratuit qui s'exécute sur les serveurs cloud de Google et destiné à la formation et à la recherche dans l'apprentissage automatique. Cette plateforme permet à l'utilisateur d'exploiter le matériel backend comme les GPU et les TPU et d'entraîner des modèles de Machine Learning et Deep Learning directement dans le cloud. Sans donc avoir besoin d'installer quoi que ce soit sur notre ordinateur à l'exception d'un navigateur.

**Jupyter Notebook** : Jupyter Notebook est une application Web Open Source permettant de créer et de partager des documents contenant du code (exécutable directement dans le document), des équations, des images et du texte. Son nom jupyter vient du langage de programmation pris en charge par le noyau qu'il prend en charge, à savoir Julia, Python et R. Avec cette application il est possible de faire du traitement de données, de la modélisation statistique, de la visualisation de données, du Machine Learning, du Deep Learning etc. Elle est disponible par défaut dans la distribution Anaconda.



Figure 4. 3 : Logos Google Colab et Jupyter Notebook.

### 4.1.3 Gradio : Interfaces graphiques pour les modèles de ML et DL

Gradio est une bibliothèque python open source qui permet de créer rapidement des parties d'interface utilisateur en ligne en quelques lignes de code. Il est rapide, simple à configurer, prêt à être utilisé et partageable en tant que connexion publique que tout le monde peut obtenir pour exécuter le modèle à distance et en parallèle sur votre machine. Gradio fonctionne avec un large éventail de médias-texte, images, vidéo et son. Mis à part les modèles ML et DL. Il peut être intégré aux modèles TensorFlow et PyTorch pour une meilleure compréhension.



Figure 4. 4 : Logos Gradio.

## 4.2 Expérimentations et résultats

Pour étudier la pertinence de l'apprentissage profond pour la classification binaire des nodules thyroïdiens, nous avons comparé plusieurs scénarios de classification avec et sans augmentation des images avec le transfert d'apprentissage sur notre petit base d'images échographiques de la thyroïde. Pour cela, nous avons adapté et affiné des modèles CNN, pré-entraînés directement sur la base ImageNet (de grande taille), selon la 2<sup>ème</sup> et 3<sup>ème</sup> stratégie. Nous avons affiné aussi d'une manière progressif des modèles CNN, pré-entraînés sur une base des images radiologiques (de taille moyenne) selon la 2<sup>ème</sup> et 3<sup>ème</sup> stratégie. Les derniers modèles entraînés sont à l'origine issus d'un transfert d'apprentissage des modèles pré-entraînés sur ImageNet. Ceci nous permet d'étudier la meilleure façon de compenser la petite taille de notre base et de comparer l'impact de transférer l'apprentissage à partir d'une base d'images très différente de notre base (ImageNet) à d'une base proche à la notre (images radiologiques).

Dans nos expériences qui seront présenté ci-dessous, notre base d'images échographiques a été préparée de la façon suivante :

A partir des trois bases présentées précédemment, nous avons créée une base d'apprentissage de 643 images (216 cas bénins et 427 cas malins) et une base de test de 152 images (53 cas bénins et 99 cas malins).

#### **4.2.1 Expérimentation des modèles CNN développés**

A l'applications des scénarios présenté précédemment, nous élaborons et comparons les performances de VGG-16, ResNet-50 et EfficientNet sur notre base d'image échographiques selon les différentes expériences suivantes.

##### **4.2.1.1 Expérience 1 : élaboration du plusieurs variantes de modèle VGG16 avec les classifieurs SVM, RF et Softmax**

Le modèle VGG16 pré-entraîné sur la base ImageNet a été élaboré selon trois variantes de modèles : M1=VGG-16 + [RF], M2=VGG-16 + [SVM], et M3= VGG-16 avec [Softmax]. L'apprentissage par transfert des trois modèles M1, M2 et M3 est fait pour 50 époques sur notre base d'images échographiques à partir d'ImageNet (M1a, M2a et M3a ) et 80 époques lorsqu'ils sont pré-entraînés sur une base intermédiaire des images radiographiques (M1b, M2b et M3b).

D'après les résultats expérimentaux présentés dans les tableaux 4.1 et 4.3, on a observé que les modèles M1, M2 et M3 avaient une bonne capacité à distinguer les images malignes des images bénignes sur les images de la base de test. De plus, on a constaté que le modèle M1 = VGG-16 + [RF] arrive à prédire les cas bénin mieux que tous les autres modèles. Les résultats de classification pour les images augmentées résumés dans les tableaux 4.2 et 4.4 étaient inférieurs aux résultats de classification des échantillons non augmentés.

##### **4.2.1.2 Expérience 2 : élaboration du modèle EfficientnetB0+Softmax**

Le modèle M4 = EfficientnetB0 avec Softmax pré-entraîné sur la base d'images ImageNet a été entraîné pour 50 époques sur l'ensemble de nos données d'entraînement (M4a) et 80 époques lorsqu'il est pré-entraîné sur l'ensemble de données radiographiques (M4b). Les résultats expérimentaux présentés dans les tableaux 4.1 et 4.3 ont été obtenus sur l'ensemble de données de validation et ont montré que le modèle EfficientnetB0 pré-entraîné sur l'ensemble de données radiographiques, avait une grande capacité à différencier les images bénignes des images malignes que lorsqu'il est pré-entraîné sur la base d'images ImageNet.

Les résultats de classification pour les images augmentées résumés dans les tableaux 4.2 et 4.4 étaient également inférieurs aux résultats de classification des échantillons non augmentés.

#### **4.2.1.3 Expérience 3 : élaboration du modèle Resnet-50 + Softmax**

Le modèle M5= Resnet-50 avec softmax, pré-entraîné sur la base d'images ImageNet (M5a) a été entraîné pour 50 époques sur notre base d'images échographiques d'apprentissage et 80 époques lorsqu'il est pré-entraîné sur la base d'images radiographiques (M5b). Les résultats expérimentaux ont montré que le modèle M5a avait également une plus grande capacité à distinguer les images malignes des images bénignes. Cependant, le modèle M5b arrive à prédire davantage les cas bénins par rapport au modèle M5a. Les résultats sont résumés dans les tableaux 4.1 et 4.3.

#### **4.2.2 Transfert d'apprentissage sur la base d'images ImageNet**

Le tableau 4.1 résument les résultats obtenus par les 5 modèles étudiés : sans augmentation en incluant le transfert d'apprentissage sur la base d'images ImageNet, avec un réglage fin selon les deux stratégies (2<sup>ème</sup> et 3<sup>ème</sup> stratégies) citées précédemment lors de l'entraînement sur nos données, exprimés en terme de précision, spécificité, sensibilité (rappel) et Score F1.

Tandis que le tableau 4.2 résume les résultats obtenus par les 5 modèles étudiés et affinée avec la 2<sup>ème</sup> stratégie du réglage fin (entraînement du dernier bloc de convolution de chaque architecture utilisée) lors de l'entraînement sur nos données (les trois bases) : avec augmentation en incluant le transfert d'apprentissage sur la base d'images ImageNet.

Les trois bases					
Modèle Niveau fine-tuning	M2a= VGG-16 + SVM	M1a= VGG-16 + RF	M3a = VGG-16+ Softmax	M4a= EfficientnetB0+ Softmax	M5a= Resnet-50+ Softmax
Geler toute la base de convolution	Accuracy:78.29% Specificity:52.83% Sensitivity:91.92% Score F1 :84.65%	Accuracy:82.24% Specificity:62.26% Sensitivity:92.93% Score F1:87.20%	Accuracy:77.37% Specificity:42.03% Sensitivity:59.50% Score F1:61.80%	Accuracy:84.21% Specificity:33.33% Sensitivity:75.21% Score F1:70.54%	Accuracy:82.11% Specificity:26.09% Sensitivity:69.42% Score F1:65.62%
Entrainer le dernier bloc de convolution	Accuracy: 75.66% Specificity:50.94% Sensitivity :88.89% Score F1 :82.63%	Accuracy:78.29% Specificity:62.26% Sensitivity:86.87% Score F1:83.90%	Accuracy:81.58% Specificity:18.84% Sensitivity:77.69% Score F1:69.37%	Accuracy: 85.79% Specificity:28.99% Sensitivity:75.21% Score F1 : 69.73%	Accuracy: 81.05% Specificity:34.78% Sensitivity:67.77% Score F1 : 66.13%
Entrainer les deux derniers bloc	Accuracy :77.63 % Specificity:50.94% Sensitivity :91.92% Score F1 :84.26%	Accuracy: 79.60% Specificity:60.38% Sensitivity:89.90% Score F1 : 85.17%	Accuracy: 79.47% Specificity:34.78% Sensitivity:61.98% Score F1 : 62.24%	Accuracy: 86.84% Specificity:33.33% Sensitivity:71.07% Score F1 : 67.98%	Accuracy: 70.00% Specificity:07.25% Sensitivity:90.91% Score F1 : 74.58%

Tableau 4. 1 : Performance des modèles VGG-16, EfficientnetB0 et Resnet-50V2 sur les données de validation des trois bases fusionnées sans l'augmentation.

Les trois bases					
Modèle Niveau fine-tuning	M2a = VGG-16 + SVM	M1a= VGG-16 + RF	M3a = VGG-16+ Softmax	M4a= EfficientnetB0+ Softmax	M5a= Resnet-50+ Softmax
Entrainer le dernier bloc de convolution	Accuracy:39.81% Specificity:13.97% Sensitivity:58.51% Score F1:53.01%	Accuracy:42.28% Specificity:11.03% Sensitivity:64.89% Score F1:56.61%	Accuracy:67.89% Specificity:33.33% Sensitivity:66.94% Score F1:65.32%	Accuracy:73.68% Specificity:30.43% Sensitivity:71.07% Score F1:67.45%	Accuracy:62.11% Specificity:18.84% Sensitivity:85.95% Score F1:74.02%

Tableau 4. 2 : Performance des modèles sur les données de validation des trois bases fusionnées avec l'augmentation.

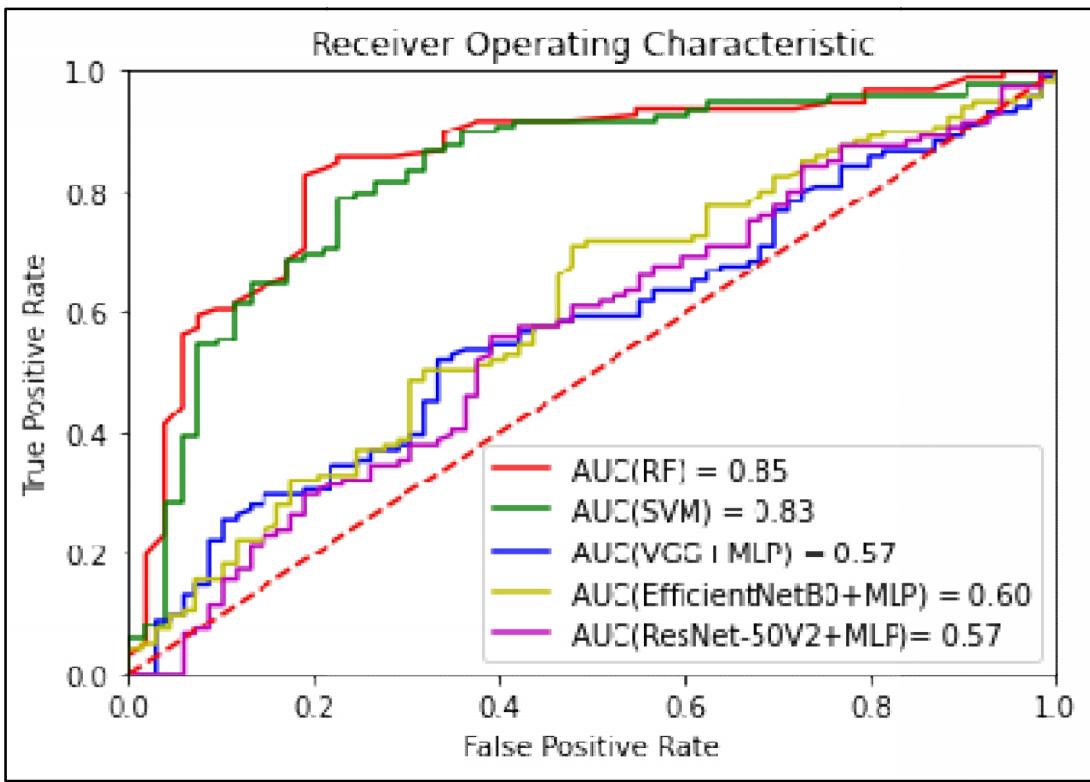


Figure 4.5 : Les courbes ROC des cinq modèles avec le Transfert d'apprentissage sur la base d'images ImageNet.

La figure 4.5 montre les courbes ROC correspondantes au modèles M1 (rouge), M2 (vert), M3 (bleu), M4 (jaune) et M5 (violet).

#### 4.2.3 Transfert d'apprentissage progressif sur l'ensemble de données radiographique

Les résultats obtenus par les 5 modèles étudiés : sans augmentation en incluant le transfert d'apprentissage sur l'ensemble de données radiographiques, avec un réglage fin selon la 2<sup>ème</sup> et 3<sup>ème</sup> stratégie lors de l'entraînement sur nos données sont résumés dans le tableau 4.3. Alors que le tableau 4.4 résume les résultats obtenus par les 5 modèles étudiés et affinée avec la 2<sup>ème</sup> stratégie du réglage fin (entraînement du dernier bloc de convolution de chaque architecture utilisée) lors de l'entraînement sur nos données (les trois bases) avec augmentation.

### Les trois bases

<b>Modèle</b>	<b>M2 b=</b>	<b>M1b=</b>	<b>M3b =</b>	<b>M4b=</b>	<b>M5b=</b>
<b>Niveau fine-tuning</b>	<b>VGG-16 + SVM</b>	<b>VGG-16 + RF</b>	<b>VGG-16+ Softmax</b>	<b>EfficientnetB0+ Softmax</b>	<b>Resnet-50+ Softmax</b>
<b>Geler tout la base de convolution</b>	Accuracy:76.97% Specificity:45.28% Sensitivity:93.94% Score F1 : 84.16%	Accuracy:80.92% Specificity:62.26% Sensitivity:90.91% Score F1: 86.12%	Accuracy:79.47% Specificity:26.09% Sensitivity:70.25% Score F1 : 66.15%	Accuracy:82.11% Specificity:46.38% Sensitivity:56.20% Score F1 : 60.18%	Accuracy:82.63% Specificity:44.93% Sensitivity:66.12% Score F1 : 66.95%
<b>Entrainer le dernier bloc de convolution</b>	Accuracy:77.63% Specificity:58.49% Sensitivity:87.88% Score F1: 83.65%	Accuracy: 79.61% Specificity:64.15% Sensitivity:87.88% Score F1 : 84.88%	Accuracy:83.16% Specificity:36.23% Sensitivity:60.33% Score F1 : 61.34%	Accuracy: 86.32% Specificity:30.43% Sensitivity:59.50% Score F1 : 59.75%	Accuracy:85.26% Specificity:33.33% Sensitivity:63.64% Score F1 : 63.11 %
<b>Entrainer les deux derniers bloc</b>	Accuracy: 75.66 % Specificity:71.70% Sensitivity:77.78% Score F1 : 80.63%	Accuracy: 73.68 % Specificity:52.83% Sensitivity:84.85% Score F1 : 80.77%	Accuracy: 84.21% Specificity:31.88% Sensitivity:71.07% Score F1 : 67.72%	Accuracy: 83.16% Specificity:23.19% Sensitivity:56.20% Score F1 : 56.20%	Accuracy: 81.58% Specificity:30.43% Sensitivity:62.81% Score F1 : 62.04%

Tableau 4. 3 : Performance des modèles VGG-16, EfficientnetB0 et Resnet-50V2 sur les données de validation des trois BASES fusionnées sans l'augmentation.

### Les trois bases

<b>Modèle</b>	<b>M2b =</b>	<b>M1b=</b>	<b>M3 b=</b>	<b>M4b= EfficientnetB0+ Softmax</b>	<b>M5b= Resnet-50+ Softmax</b>
<b>Niveau fine- tuning</b>	<b>VGG-16 + SVM</b>	<b>VGG-16 + RF</b>	<b>VGG-16+ Softmax</b>		
<b>Entrainer le dernier bloc de convolution</b>	Accuracy: 42.90% Specificity:16.18% Sensitivity :62.23% Score F1 :40.61%	Accuracy: 42.59% Specificity:07.35% Sensitivity :68.09% Score F1 :57.92%	Accuracy: 72.11% Specificity:43.48% Sensitivity :66.12% Score F1 : 66.67%	Accuracy: 65.26% Specificity:01.45% Sensitivity:97.52% Score F1 : 76.87%	Accuracy: 64.21% Specificity:07.25% Sensitivity :97.52% Score F1 :77.89%

Tableau 4. 4: Performance des modèles sur les données de validation avec l'augmentation.

La figure 4.6 montre les courbes ROC correspondantes au modèles M1 (rouge), M2 (vert), M3 (bleu), M4 (jaune) et M5 (violet) avec l'apprentissage profond par transfert progressif.

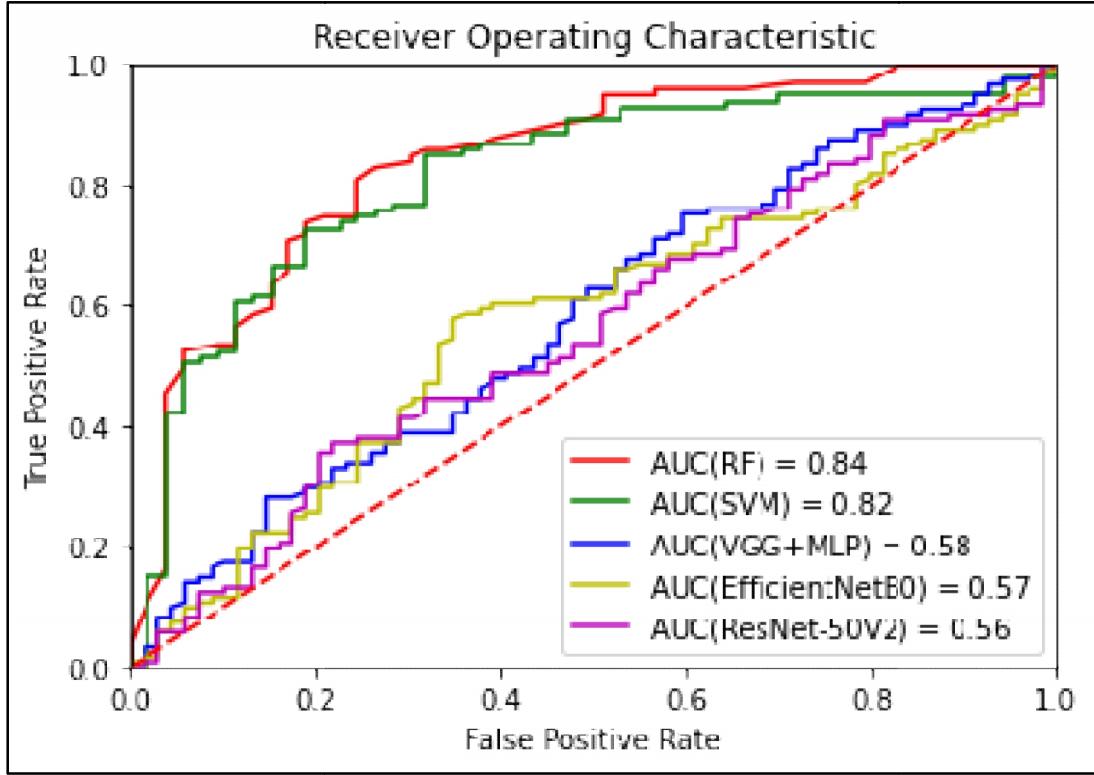


Figure 4. 6 : Les courbes ROC des cinq modèles avec le Transfert d'apprentissage progressif sur la base d'images radiographique.

### 4.3 Discussion et conclusion

La classification a été effectuée sur les trois bases d'images combinées en raison du problème de manque d'échantillons dans chaque base d'images échographiques thyroïdiennes. Des études antérieures sur l'impact de la taille de la base d'images dans le transfert d'apprentissage à l'aide de CNN ont montré qu'une taille appropriée de base de données pour le transfert d'apprentissage pourrait aller de 500 à 1 000 échantillons par classe [52] ; par conséquent, nous avons préféré de ne pas réaliser la classification sur les trois bases d'images séparément ainsi pour éviter le sur-apprentissage.

Il a été observé que les trois bases d'images publiques combinées donnaient une sensibilité élevée et une faible spécificité. Cela est dû au grand nombre d'échantillons malins qui est largement plus important que le nombre d'échantillons bénins dans la base d'images combinée.

Pour les images augmentées, les cinq modèles pré-entraînés et affinés fonctionnent mal, cela est dû aux méthodes d'augmentation impliquant une déformation de l'image de la lésion ce qui entraîne un sur-ajustement, au fur et à mesure que nous augmentons le nombre des échantillons de nos données, on constate que la performance des modèles CNN diminue.

La comparaison des performances de ces cinq architectures pour en sélectionner la meilleure dans notre cas est basée sur la valeur de F1-score, cela est dû aux coûts des faux positifs et des faux négatifs qui diffèrent en nombre [53]. On obtient des faux positifs supplémentaires (fausses alerte) plutôt que de sauvegarder des faux négatifs (c'est-à-dire, les personnes atteintes du cancer de la thyroïde qui seront qualifiées comme personnes seines).

Dans les tableaux 4.1 et 4.3, nous pouvons voir que le réglage fin des modèles M3 =VGG-16 + Softmax, M4= EfficientnetB0 et M5= Resnet-50 en entraînant le classifieur et quelques couches supérieures de la base convective avec les images échographiques thyroïdiennes des trois bases d'images fusionnées, a conduit à une précision de validation supérieur à celle lorsque les modèles pré-entraînés sont exécutés en tant qu'extracteur de caractéristiques. Donc, les M3b, M4b et M5b sont meilleurs que les modèles M3a, M4a et M5a grâce au transfert progressif de l'apprentissage.

En revanche les modèles M1a =VGG-16 + RF et M2a= VGG-16 + SVM, leur performance est meilleure que les modèles M1b, M2b, M3, M4 et M5 quand ils sont utilisés comme mécanisme d'extraction de caractéristiques. Ceci est grâce à la puissance des classificateurs RF et SVM combinés au modèles CNN.

Ces résultats sont donc importants, car ils montrent que le bon choix de niveaux de fine tuning (pour la spécialisation des caractéristiques) et de classifieur (pour la prédiction) peut résoudre le problème de la petite taille de la base d'images et nous épargne de passer par une base intermédiaire.

En analysant les résultats obtenus du tableau 4.1, nous remarquons que les meilleures valeurs de précision sont de 82.24%, 78.29%, 81.58%, 86.84% et 82.11%, pour les modèles M1a=VGG16+RF, M2a=VGG16+SVM, M3a=VGG16+MLP, M4a=EfficientnetB0+MLP et M5=Resnet50V2+MLP respectivement, cependant le modèle EfficientnetB0+MLP l'emporte aux autres architectures avec une valeur de précision de 86,84% en entraînant ses deux derniers blocs, la précision de ce modèle s'est améliorée tout au long du processus du fine-tuning. La figure 4.7 montre la courbe de précision et de perte de modèle M4a

(EfficientnetB0+MLP) dans le cas de l'entraînement des deux derniers blocs de convolution. Une précision élevée et une perte plus faible indique un modèle plus performant.

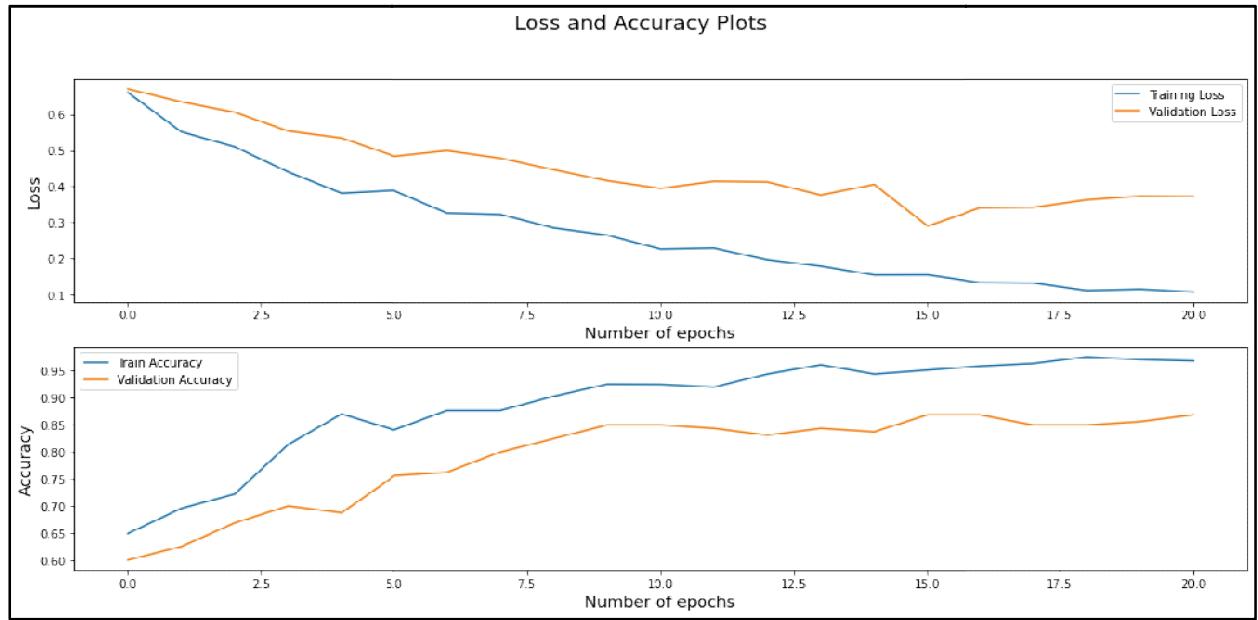


Figure 4.7 :Les courbes de précision et de perte de modèle M4a (EfficientnetB0-MLP) dans le cas de l'entraînement des deux derniers blocs de convolution dans le scénario Transfert d'apprentissage sur la base ImageNet.

Une dégradation de précision est constatée pour le modèle M5a(Resnet50V2+MLP) contrairement à son score F1 qui s'est amélioré au fur et à mesure qu'on applique le fine-tuning, pour le M3a(vgg16-MLP), une précision de 77,37 % est noté en gelant tout le bloc de convolution du modèle suivie par une précision de 81,58% et 78,47% en entraînant le dernier bloc, et les deux derniers blocs respectivement.

En comparant les meilleures valeurs de score F1 du tableau 4.1 ( M1a=87.20% , M2a=84 .65%, M3a=69,37%, M4a=70,54% et M5a=74,58% ) et pour éviter tous faux négatif, on peut clairement dire que M1a (VGG16+RF) suivi par M2a(VGG16+SVM) formés avec des échantillons des trois bases d'images fusionnées, non augmentées et affinées avec la 3<sup>ème</sup> stratégie en en gelant toute la base de convolution de l'architecture sont les meilleurs modèles et les mieux adaptés à notre cas, avec une valeur AUC ou ROC (figure 4.8) de 0.85 et 0.83 respectivement. En revanche, la valeur de spécificité (62,26% VS 52,83%), de sensibilité (92,93% VS 91 ,92%) et d'AUC de M1a(VGG16+RF) sont élevées à celles de M2a (VGG16+SVM) ce qui nous permet de le sélectionner comme le meilleur modèle dans le cas de ce scénario.

Tandis que le tableau 4.3, dans le scénario du transfert d'apprentissage progressif sur la base d'images radiographiques, nous observons une amélioration de la valeur de précision de M3b(VGG16+MLP) et M5b(Resnet50V2+MLP) par rapport à la précision de M3a et M5a (tableau 4.2). Les meilleures valeurs de précision que nous avons obtenues sont de pour, M1b(VGG16+RF)= 80.92%, M2b(VGG16+SVM) =77.97%, M3b(VGG16+MLP)= 84.21% , M4b(EfficientnetB0+MLP)= 86.32% et M5b(Resnet50V2+MLP)= 85.26% respectivement.

De plus nous remarquons une seconde fois que le modèle M4b(EfficientnetB0+MLP) est le meilleur dans cette catégorie (Mib) (en termes de précision en entraînant le dernier bloc, avec un score F1 de 59,75%, ce qui ne met pas en valeur encore une fois ce modèle pour résoudre notre problème.

La figure 4 .8 montre le tracé de la courbe de précision et de perte du modèle M4b(EfficientnetB0+MLP) dans le scénario du transfert d'apprentissage progressif en entraînant le dernier bloc du modèle, ayant la valeur de précision la plus élevée.

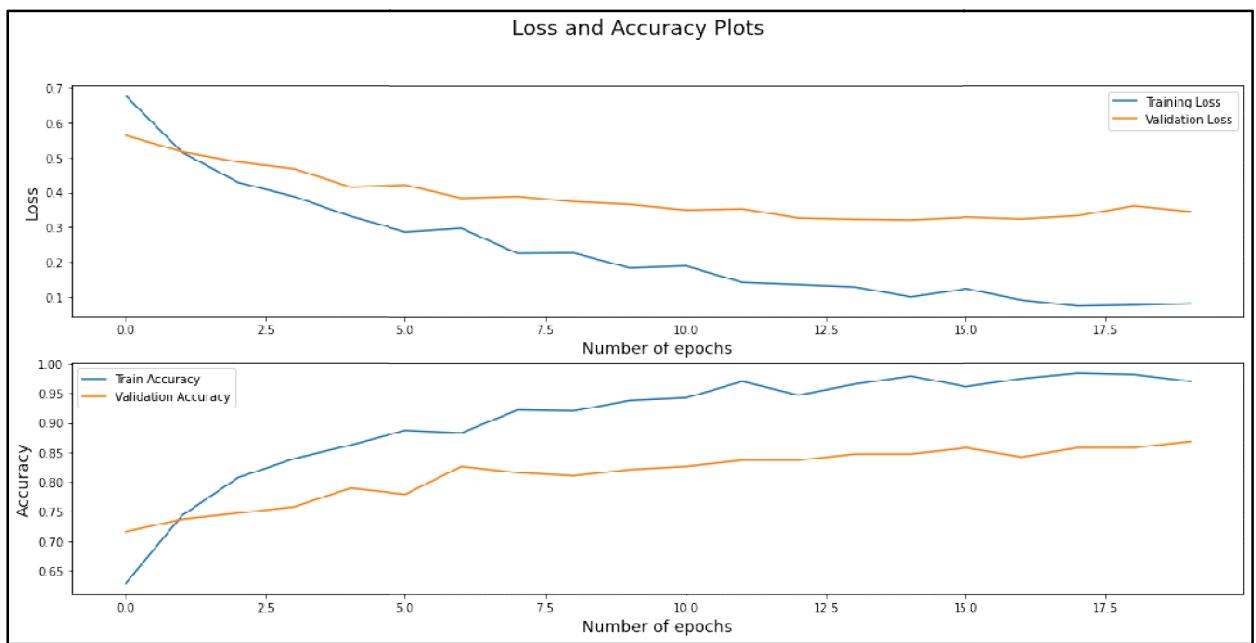


Figure 4. 9 : les courbes de précision et de perte de modèle M4b(EfficientnetB0-MLP) dans le cas de l'entraînement du dernier bloc de convolution dans le scénario Transfert d'apprentissage progressif sur la base d'images radiographiques.

Les résultats du tableau 4.3 montrent que le modèle M1b (VGG16+RF) et M2b(VGG16+SVM) affinés avec la 3<sup>ème</sup> stratégie ; en gelant toute la base de convolution de l'architecture, ont atteint les performances de prédictions les plus élevées en termes de score F1 (M1b=86.12%, M2b= 84,16%), avec une précision de (M1b=80.92%, M2b=76,97% respectivement). Notons que ces valeurs se sont légèrement diminuées par rapport au scenario

du Transfer d'apprentissage sur la base ImageNet. En comparant VGG16+RF et VGG16+SVM en termes de valeur AUC (Figure 4.6), 0.84 VS 0.82 et le score F1 86.12% VS 84,16%, nous pouvons dire que VGG16+RF est bien meilleur que le VGG16+SVM dans ce scénario.

Cependant, dans les tableaux 4.2 et 4.4 , nous avons observé que l'étape d'augmentation des images n'améliore pas la performance des modèles CNN dans le cas du scénario du transfert d'apprentissage progressif sur la base d'images radiographiques comme dans le cas du scénario du transfert d'apprentissage sur la base ImageNet. Les résultats de classification de VGG16+RF pour les images augmentées étaient inférieurs aux résultats de classification des échantillons non augmentés dans les deux cas. La figure 4.9, montre un exemple de tracé de courbe de précision et de perte du modèle Resnet-50, montrant les performances du modèle (précision et perte) évaluées sur les ensembles de données d'entraînement et de validation à chaque époque tout au long du processus d'entraînement. A partir de la figure 4.9, nous observons un léger écart entre la précision de la formation et de la validation. De plus, bien que la précision soit élevée, le modèle Resnet-50 prédit mal sur les données de test, cela signifie qu'il y a peu de données.

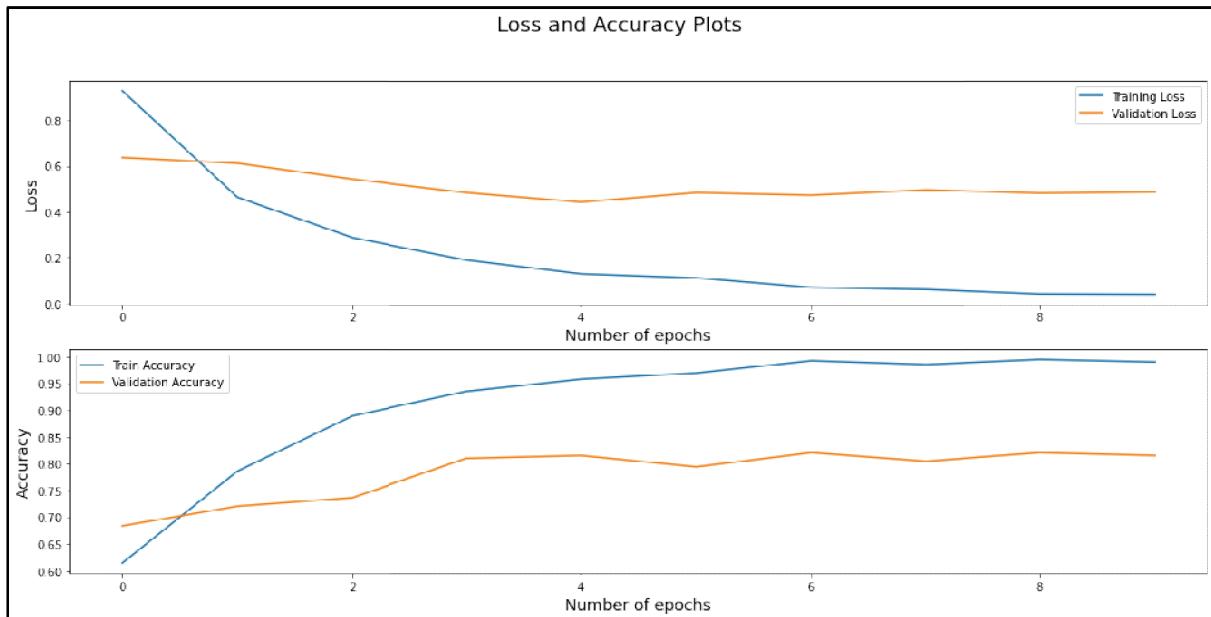


Figure 4. 10 : Courbes de précision et de perte pour le modèle Resnet-50V2.

Le transfert d'apprentissage progressif sur la base d'images radiographiques et le réglage fin ont donné des résultats dont l'amélioration est légèrement significative dans la classification des images de nodules thyroïdiens par rapport à ceux du transfert

d'apprentissage sur la base d'images ImageNet pour les deux architectures Resnet-50V2 et VGG16+MLP.

En comparant les résultats des deux scénarios, nous constatons que le VGG16+RF est sélectionné comme meilleur modèle dans les deux cas. Cependant ce modèle pré-entraîné sur la base ImageNet est le plus pertinent pour l'appliquer à notre problématique avec une précision de 82.24 %, un score F1 de 87.20%, une spécificité de 62.26%, une sensibilité de 92.93%, et une valeur AUC de 0.85.

Les figures 4.10 et 4.11 montrent quelques exemples de l'interface graphique pour la prédiction.

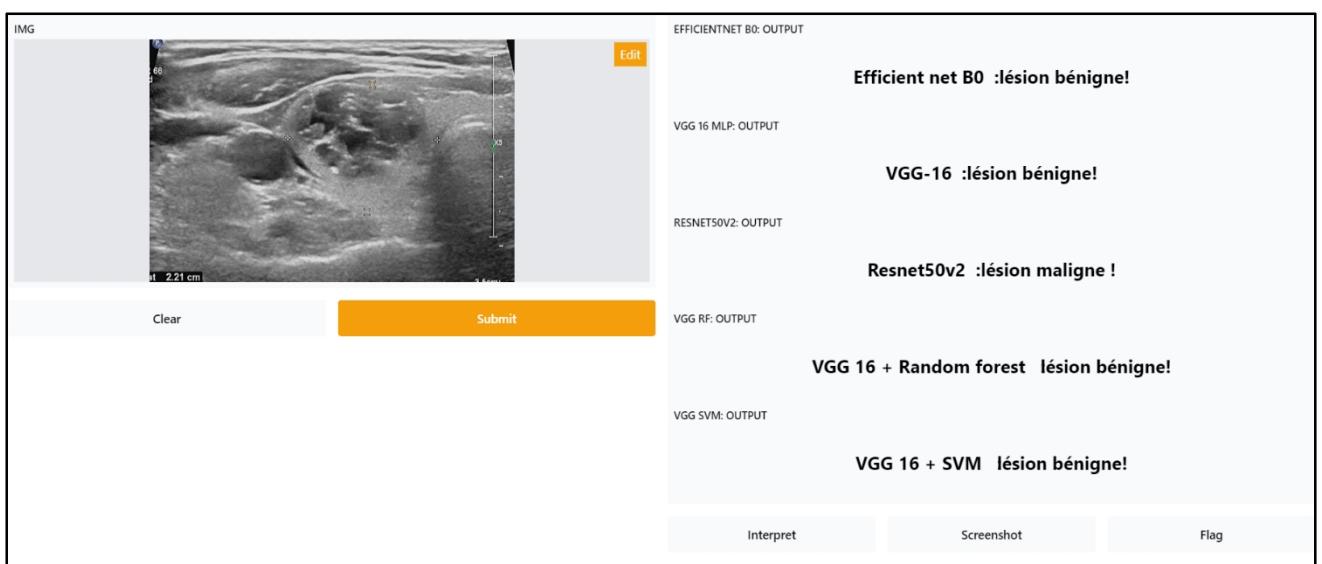


Figure 4.10 : Exemple de prédiction d'une image de lésion bénigne.

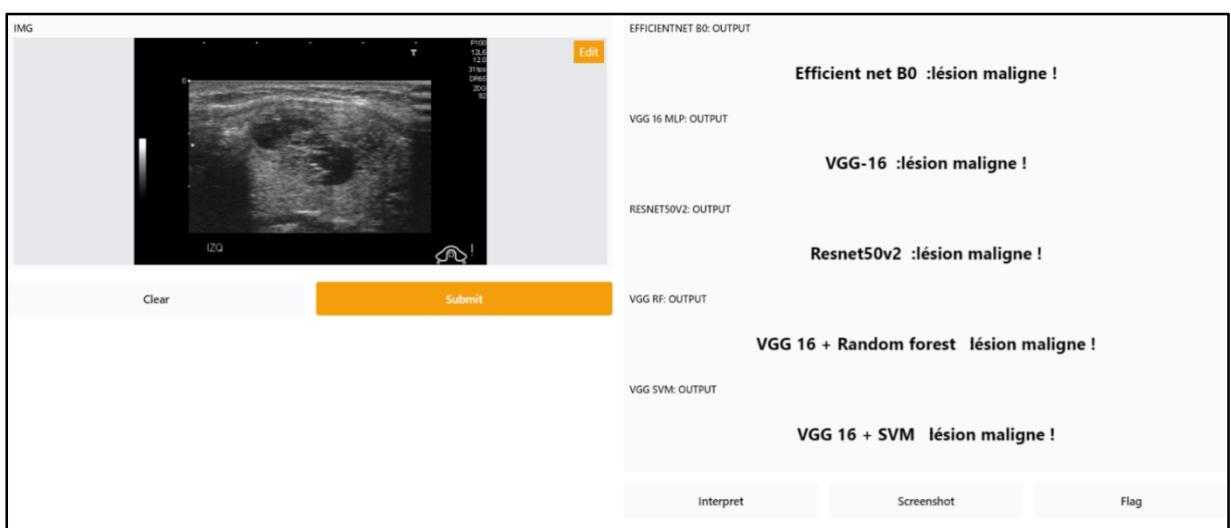


Figure 4.11 : Exemple de prédiction d'une image de lésion maligne.

Nous sommes conscients que notre travail comporte certaines limites. Premièrement, la collecte d'images, nous n'avons qu'une petite base d'images échographiques thyroïdiennes publique et l'apprentissage profond nécessite une base d'images de taille importante. Deuxièmement, le modèle de classification n'est pas assez robuste pour prédire relativement bien les cas bénins, cela est dû à la nature des images dans lesquelles la base d'images publique avait un biais en faveur des échantillons cancéreux. Troisièmement, les performances obtenues n'ont pas été comparées aux méthodes de la littérature. Ceci est principalement dû à l'unicité de la base d'images.

## Conclusion générale

Dans le présent travail, nous avons principalement présenté les différentes méthodes basées sur les CNN pour la classification d'images échographiques thyroïdiennes avec un entraînement supervisé pour différencier les nodules malins des nodules bénins.

Pour la tâche de classification, sur notre base d'images, nous avons analysé les performances de différentes méthodes de classification basées sur des modèles CNN dans deux scénarios. À cette fin, nos expériences ont révélé que le scénario du transfert d'apprentissage progressif à partir de la base d'images radiographiques pour classer les nodules thyroïdiens améliore les performances des architectures VGG-16+Softmax et Resnet-50V2, par contre les performances des autres architectures sont proche des résultats du scénario du transfert d'apprentissage à partir d'ImageNet annoté à grande échelle. De plus, une étude comparative des différents modèles a montré que le VGG-16+RF pré-entraîné sur la base ImageNet, surpassait les autres modèles d'apprentissage profond avec de bonnes performances en termes de score F1 (87.20%), de précision(82.24%), de spécificité (62.26%), de sensibilité (92.93%) et de valeur de ROC (0.85).

Par conséquent, le transfert progressif de l'apprentissage, le réglage fin de l'architecture et la fusion de classificateurs linéaires tels qu'eRF à l'architecture sont utiles pour essayer de surmonter le problème des réseaux de neurones profonds pré-entraînés sur de très grands ensembles de données différents du domaine en question. Bien que l'intelligence artificielle ne remplace pas le médecin dans les années à venir, mais cela le débarrasse de toute une série de travaux, pour qu'il puisse passer plus de temps avec ses patients. L'expert clinique peut apprendre les bases de l'innovation en intelligence artificielle et comment les structures basées sur l'IA les aideront à offrir de plus grands avantages à ses patients au travail. En pratique clinique, notre modèle d'apprentissage profond pourrait offrir un deuxième avis dans le processus de diagnostic.

Dans les perspectives de ce travail, il serait intéressant de développer un système qui permet de classer les nodules thyroïdiens en se servant des cinq modèles proposés d'une manière automatique, le choix du modèle dépend fortement des caractéristiques de chaque image, dans le but de développer un système robuste qui peut classer les nodules thyroïdiens en bénin ou malin tout en minimisant toute fausse classification. Nous souhaitons également

que toute suite de ce travail aura la possibilité d'utiliser une dataset qui contient des milliers d'images thyroïdiennes pour avoir de meilleurs résultats.

## Références bibliographiques

- [1] Guorong Wu, Dinggang Shen, Mert R. Sabuncu; Machine Learning and Medical Imaging. 2016, A volume in The Elsevier and MICCAI Society Book Series, pp.114.
- [2] Gabriel Chartrand, PhD Phillip M. Cheng, MD, MS Eugene Vorontsov, BSc Eng Sci Michal Drozdzal, PhD Simon Turcotte, MD, MSc Christopher J. Pal, PhD Samuel Kadoury, PhD An Tang, MD, MSc; Deep Learning: A Primer for Radiologists1. 2017, RadioGraphics journal, pp. 37:2113–2131.
- [3] Jieun Koh , Eunjung Lee , Kyunghwa Han , Eun-Kyung Kim , Eun Ju Son , Yu-Mee Sohn , Mirinae Seo , Mi-ri Kwon , Jung HyunYoon , Jin Hwa Lee , Young Mi Park , Sungwon Kim , Jung Hee Shin & JinYoung Kwak; Diagnosis of thyroid nodules on ultrasonography by a deep convolutional neural network. 2020, journal of scientific reports, Vol.10, Num.15245.
- [4] Michal Drozdzal, Gabriel Chartrand, Eugene Vorontsov, Lisa Di Jorio, An Tang, Adriana Romero, Yoshua Bengio, Chris Pal, Samuel Kadoury; Learning Normalized Inputs for Iterative Estimation in Medical Image Segmentation. 2017, PMC article, PMID: 29169029, DOI: 10.1016/j.media.2017.11.005.
- [5] Jianning Chi, Ekta Walia, Paul Babyn, Jimmy Wang, Gary Groot & Mark Eramian ;Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. 2017, journal of Digital Imaging, Vol.30, pp. 477-486.
- [6] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, Clara I. Sánchez; A survey on deep learning in medical image analysis. 2017, PMC article, PMID: 28778026, DOI: 10.1016/j.media.2017.07.005.
- [7] AMINA 2018 Khachnaoui H. et al; Classification automatique des nodules thyroïdiens dans les images échographiques par apprentissage profond. 2018, thèse de doctorat, université de Tunis El Manar.
- [8] François Chollet; Deep Learning with Python. 2018, livre, pp. 31-34.

- [9] Pascal Monasse, Kimia Nadjahi; Découvrez les différentes couches d'un CNN. 2020. <https://openclassrooms.com/fr/courses/4470531-classez-et-segmentez-des-donnees-visuelles/5083336-decouvrez-les-differentes-couches-dun-cnn>. Accédé en Juillet 2021.
- [10] Guillaume Saint-Cirgue; Apprendre le machine learning en une semaine. 2019, livre, pp. 13-63.
- [11] Khouloud Dahmane ; Analyse d'images par méthode de Deep Learning appliquée au contexte routier en conditions météorologiques dégradées.2020, thèse de doctorat en Electronique et Systèmes.
- [12] Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J van Riel, Mathilde Marie Winkler Wille, Matiullah Naqibullah, Clara I Sanchez, Bram van Ginneken; Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. 2016, IEEE Trans Med Imaging 35 (5), pp.1160-1169.
- [13] Lo SB, Lou SA, Lin JS, Freedman MT, Chien MV, Mun SK; Artificial convolution neural network techniques and applications for lung nodule detection. 1995, journal, IEEE Trans Med Imaging 14, pp. 711-718.
- [14] Riviera Imagerie Médicale, membre du réseau Vidi ; Echographie de la thyroïde. <https://www.riviera-imagerie.fr/page/echographie-de-la-thyroide/#:~:text=L%C3%A9chographie%20thyro%C3%AFdienne%20est%20indolore,selon%20un%20mouvement%20de%20balayage>. Accédé en Juillet 2021.
- [15] Dong Nie, Xiaohuan Cao, Yaozong Gao, Li Wang , Dinggang Shen; Estimating CT image from MRI data using 3D fully convolutional networks. 2016, journal, In: DLMIA. Vol. 10008 of Lect Notes Comput Sci. pp. 170–178.
- [16] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake; Real-time human pose recognition in parts from single depth images.2013, journal, *Communications of the ACM*, 56(1), pp.116-124.
- [17] Vili Podgorelec, Peter Kokol, Bruno Stiglic, Ivan Rozman; Decision trees: an overview and their use in medicine. October 2002, journal of Medical Systems, Kluwer Academic/Plenum Press, Vol. 26, Num. 5, pp. 445-463.

- [18] Vanhoof, Antoine; Indicateurs de généralisation au sein des couches d'un réseau de neurones convolutifs. 2019, mémoire de Master, école polytechnique de Louvain.
- [19] Joseph Antony, Kevin McGuinness, Noel E O Connor, Kieran Moran; Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. In Pattern Recognition (ICPR). 2016, 23rd International Conference on. IEEE, pp.1195-1200.
- [20] Sangheum Hwang, Hyo-Eun Kim, Jihoon Jeong MD., Hee-Jin Kim; A novel approach for tuberculosis screening based on deep convolutional neural networks. 2016, journal, proc. SPIE 9785, Medical Imaging.
- [21] Yagoub Djamel Eddine; Analyse d'images médicales, application à la catégorisation des types de tuberculose pulmonaire. 2018, mémoire de Master.
- [22] Ilyes Talbi; L'intelligence artificielle au service de la santé. 2020. Accédé en Juillet 2021.
- [23] Antonin Prochazka , Sumeet Gulati , Stepan Holinka et Daniel Smutek; Classification des nodules thyroïdiens dans les images échographiques à l'aide de caractéristiques indépendantes de la direction extraites par décomposition binaire à deux seuils. 2018, PMC article, PMID: 30774015, PMCID: PMC6379796, DOI: 10.1177/1533033819830748.
- [24] Ammar Mahmood, Ana Giraldo Ospina, Mohammed Bennamoun, Senjian An, Ferdous Sohel, Farid Boussaid, Renae Hovey, Robert B. Fisher et Gary A. Kendrick; Automatic Hierarchical Classification of Kelps Using Deep Residual Features. 2020, PMC article, PMID: 31941132, PMCID: PMC7013955, DOI: 10.3390/s20020447.
- [25] Wenfeng Song, Shuai Li, Ji Liu, Hong Qin, Bo Zhang, Shuyang Zhang, Aimin Hao; Multitask Cascade Convolution Neural Networks for Automatic Thyroid Nodule Detection and Recognition. 2019, IEEE J Biomed Health Inform, pp. 1215-1224.
- [26] Valérie JULIEN; Classification TIRADS ponction écho guidée. [https://www.courseexercices.com/PDF\\_Cours\\_Exercices\\_Telecharger.php?q=forum+nodule+vascularis%C3%A9](https://www.courseexercices.com/PDF_Cours_Exercices_Telecharger.php?q=forum+nodule+vascularis%C3%A9). Accédé en Juillet 2021.
- [27] Mei Xueyan,Xiaomeng Dong, Timothy Deyer, Jingyi Zeng, Theodore Trafalis and Yan Fang; Thyroid Nodule Benignity Prediction by Deep Feature Extraction. 2017, IEEE 17th International Conference on Bioinformatics and Bioengineering (BIBE).

- [28] Tianjiao Liu, Shuaining Xie, Jing Yu, Lijuan Niu and Weidong Sun; Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features. 2017, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
- [29] Ma Jinlian, Fa Wu, Jiang Zhu, Dong Xu and Dexing Kong. A pretrained convolutional neural network based method for thyroid nodule diagnosis. 2016, Ultrasonics, pp.221–230.
- [30] Dehua Chen, Jinxuan Niu, Qiao Pan, Yue Li and Mei Wang; A Deep-Learning Based Ultrasound Text Classifier for Predicting Benign and Malignant Thyroid Nodules. 2017, International Conference on Green Informatics 2.
- [31] K V Sai Sundar, Kumar T Rajamani, S. Siva Sankara; Exploring Image Classification of Thyroid Ultrasound Images using Deep Learning. 2018, Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering, pp.1635-1641.
- [32] Yongjun Chang, Anjan Kumar Paul, Namkug Kim , Jung Hwan Baek , Young Jun Choi, et al; Computer-aided diagnosis for classifying benign versus malignant thyroid nodules based on ultrasound images: a comparison with radiologist-based assessments. 2016, PMC article, PMID: 26745948, DOI: 10.1111/1.4939060.
- [33] Bin Zhang, Jie Tian, Shufang Pei, Yubing Chen, Xin He, Yuhao Dong , et al; Machine learning-assisted system for thyroid nodule diagnosis. 2018, PMC article, PMID: 30929637, DOI: 10.1089/thy.2018.0380.
- [34] Anju Jain Asst, Hisar; Machine learning techniques for medical diagnosis: a review. 2015, 2<sup>nd</sup> international conference on Science, Technology and Management, India.
- [35] Akin Ozciit; Enhanced Cancer Recognition System Based on Random Forests Feature Elimination Algorithm. 2011, journal of medical systems, Vol. 36, Num. 4, pp. 2577–2585.
- [36] Cuong Nguyen, Yong Wang, and Ha Nam Nguyen; Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic. 2013, journal of Biomedical Science and Engineering, Vol. 06, Num. 05, pp. 551–560.
- [37] Kevin Cherry, Shijun Wang, Evrim Turkbey, and Ronald Summers; Abdominal lymphadenopathy detection using random forest. March 2014, proceedings of the SPIE, Vol.90351G, medical imaging.

- [38] Kemal Polat; Breast cancer diagnosis using least square support vector machine. Jul. 2007, Digit. Signal Process, Vol. 17, Num. 4, pp. 694–701.
- [39] Bo Jin, Yan-Qing Zhang; Support vector machines with genetic fuzzy feature transformation for biomedical data classification. 2007, Information Sciences, Vol. 177, Num. 2, pp. 476–489.
- [40] Murat Çınar, Mehmet Engin, Erkan Zeki Engin, and Y. Ziya Ateşçi; Early prostate cancer diagnosis by using artificial neural networks and support vector machines. Apr. 2009, Expert Syst. Appl., Vol. 36, Num. 3, Part 2, pp. 6357–6361.
- [41] Maryam Vatankhah, Vahid Asadpour, and Reza Fazel-Rezai; Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage. May 2013, Applied Soft Computing, Vol. 13, Num. 5, pp. 2537– 2546.
- [42] Waqas Ahmed and Mark G. Eramian; Automated detection of grayscale bar and distance scale in ultrasound images. 2010, proceedings of the SPIE, Vol.7623.
- [43] Nikhil S. Narayan, Pina Marziliano, and Christopher G. L. Hobbs; Automatic removal of manually induced artefacts in ultrasound images of thyroid gland. 2013, 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3399–3402.
- [44] Xiaowen Liang, Jinsui Yu, Jianyi Liao, and Zhiyi Chen; Convolutional Neural Network for Breast and Thyroid Nodules Diagnosis in Ultrasound Imaging. 2020, BioMed Research International, article Id: 1763803.
- [45] D. Millon, E. Coche; le rôle de l'échographie dans la prise en charge des nodules thyroïdiens la classification ti-rads. 2016, article, session pathologie thyroidienne.
- [46] Hongxun Wu, Zhaozhong Deng, Bingjie Zhang, Qianyun Liu; Classifier Model Based on Machine Learning Algorithms: Application to Differential Diagnosis of Suspicious Thyroid Nodules via Sonography. 2016, American Journal of Roentgenology, Vol. 207, Num.4.
- [47] Corina Maria Vasile, Anca Loredana Udrîștoiu, Alice Elena Ghenea, et al; Intelligent Diagnosis of Thyroid Ultrasound Imaging Using an Ensemble of Deep Learning Methods. 2021, Medicina article, Vol. 57, Issue. 4.

- [48] Marina Chane; Classification des images medicales comprendre le reseau de neurones convolutifs CNN. 2021. <https://www.imaios.com/fr/Societe/blog/Classification-des-images-medicales-comprendre-le-reseau-de-neurones-convolutifs-CNN>. Accédé en Juillet 2021.
- [49] Universidad Nacional de Colombia. DDTI: An open access database of thyroid ultrasound images. <http://cimalab.intec.co/?lang=en&mod=project&id=31>. Accédé en Juillet 2021.
- [50] Hitachi Medical Systems Europe and Radiology Department of the Gelderse Vallei Hospital. Welcome to Ultrasoundcases.info. <http://www.ultrasoundcases.info/Default.aspx>. Accédé en Juillet 2021.
- [51] xiaobaishiji; AI for Thyroid Imaging. <http://thyroid.xiaobaishiji.com>. Accédé en Aout 2021.
- [52] Soekhoe Deepak, van der Putten Peter, Plaat Aske; On the impact of data set size in transfer learning using deep neural networks. 2016, International symposium on intelligent data analysis, p. 50-60.
- [53] Salma Ghoneim; Précision, rappel, précision, score F et spécificité, sur lesquels optimiser? 2019. <https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124>. Accédé en Aout 2021.