



Université Mohamed Khider de Biskra  
Faculté des Sciences et de la Technologie  
Département de Génie Électrique

# MÉMOIRE DE MASTER

Sciences et Technologies  
Télécommunication  
Réseaux et Télécommunication

Réf. : Entrez la référence du document

---

Présenté et soutenu par :

**Sidaty Tofla et Ghellab Nessrine**

Le : dimanche 26 juin 2022

## **Évaluation de l'âge osseux à l'aide de modèles de transformateur de vision**

---

Jury :

Dr. Abdelhamid Benakcha	Pr.	Université Biskra	President
Dr. Salim Sbaa	Pr.	Université Biskra	Examineur
Dr. Soraya Zehani	MCA	Université Biskra	Rapporteur

Année universitaire : 2021 – 2022

## **Remerciements**

Nous tenons tout d'abord à remercier Dieu le tout puissant et miséricordieux, qui nous a donné la force et la patience d'accomplir ce Modeste travail.

La première personne que nous tenons à remercier est notre encadreur **Mme. Zehani Soraya** pour l'orientation, la confiance, la patience et ses bonnes explications qui ont constitué un apport considérable sans lequel ce travail n'aurait pas pu être mené au bon port.

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail et de l'enrichir par leurs propositions. Nous tenons à exprimer nos sincères remerciements à tous les professeurs qui nous ont enseigné et qui par leurs compétences nous ont soutenu dans la poursuite de nos études.

Enfin, nous tenons également à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail.

## **Dédicaces**

Je dédie ce modeste travail à : A mes chers parents « Mat Yenja et Mohamed Bouna Sidaty » que nulle dédicace ne puisse exprimer mes sincères sentiments, Pour leur patience illimitée, leur encouragement continu, leur aide, mon succès demeure de loin le fruit de leurs longues années de sacrifices et de patience.

À mes chères sœurs,

À mes frères,

À mes chers amis,

Et à tous ceux qui ont contribué de près ou de loin pour que ce travail soit possible, je vous dis merci.

## Dédicaces

Je dédie ce modeste travail à : A mes chers parents «Ghellab Nour Eddin et Sawla Nassira» que nulle dédicace ne puisse exprimer mes sincères sentiments, Pour leur patience illimitée, leur encouragement continu, leur aide, mon succès demeure de loin le fruit de leurs longues années de sacrifices et de patience.

À mon fiancé,

À mes chères sœurs,

À mes frères,

À mes chers amis,

Et à tous ceux qui ont contribué de près ou de loin pour que ce travail soit possible, je vous dis merci.

## المخلص

منذ سنوات عديدة حقق مجال التصوير الطبي إزدهارا كبيرا مما ساهم في تشخيص وعلاج الكثير من الأمراض، ويعتبر محول الرؤية من أحدث تقنيات تصنيف الصور، الهدف من عملنا هو تصوير وتقييم عمر العظام بواسطة محول الرؤية ونستخدم في هذه الرسالة كل من ViT16 (dataset bone assessment لتصنيف الصور على قاعدة بيانات ResNet50, VGG19).  
الكلمات الافتتاحية: التصنيف، التصوير الطبي، التعلم العميق، التعلم المنقول، محول الرؤية

## Résumé

Pendant de nombreuses années, le domaine de l'imagerie médicale a connu une grande prospérité, ce qui a contribué au diagnostic et au traitement de nombreuses maladies. La Vision transformer est considéré comme une technologie de classification d'images de pointe. Le but de notre travail est de visualiser et d'évaluer l'âge des os de la main avec la vision transformer Dans ce mémoire, nous utilisons à la fois ViT16 VGG19, et ResNet50, pour classer les images sur une base de données (dataset bon age assessment).

**Mots-Clés :** Classification, Imagerie médicale, Apprentissage profond, Apprentissage par transfert, Vision de transformer.

## Abstract

For many years, the field of medical imaging has achieved great prosperity, which has contributed to the diagnosis and treatment of many diseases, and the vision transformer is considered the latest technology for classifying images. The aim of our work is to view and assess the age of the hand bones with a vision transformer, we use both ViT16, VGG19 and ResNet50 to classify images on a database (dataset bon age assessment).

**Keywords:** Classification, Medical imaging, Deep learning, Transfer learning, Vision transformer.

## Table des matières

Résumé.....	IV
Abstract.....	IV
Table des matières.....	V
Liste des figures.....	VII
Listes des tableaux.....	IX
Listes des Acronymes.....	X
Introduction Générale.....	1
I.1 Introduction.....	3
I.2 Historique de l'imagerie médicale.....	3
I.5 Description des différentes techniques d'imagerie médicale.....	5
I.5.3. Image rayons X.....	6
I.5.4. Imagerie nucléaire.....	9
I.5.5. Scintigraphie.....	13
I.5.6. Imagerie thermique.....	15
I.5.7. Imagerie hyperspectrale.....	16
I.6. Standard DICOM (Digital Imaging and Communication Medicine).....	16
I.7. Avantages et inconvénients de l'image médicale.....	16
I.8. Conclusion.....	18
II. Introduction.....	20
II.2. Machine Learning.....	20
II.2.1 Définition.....	20
II.2.2. Différents types d'apprentissage machine.....	21
II.2.3. Algorithmes de l'apprentissage machine.....	21
II.3. Réseaux de neurones.....	23
II.3.1 Des neurones biologiques aux neurones artificiels.....	23
II.3.2 Réseaux de neurones artificiels.....	24
II.3.3 Perceptron multicouches (PMC).....	26
II.4. Apprentissage profond (Deep Learning).....	27
II.4.1. Domaines d'application du Deep Learning.....	28
II.5. Transfer Learning.....	28
II.5.1. Définition.....	28
II.5. 2. Histoire du Transfer Learning.....	29
II.5.3. Comment utiliser l'apprentissage par transfert.....	30
II.5.4. Exemples d'application du Transfer Learning.....	31

II.5.5. Type de Transfer Learning .....	32
II.6. Transformateur de vision (ViT).....	34
II.6.1. Définition .....	34
II.6.2. Histoire de Transformateur de Vision .....	35
II.6. 2.Fonctionnement du Transformateur de Vision .....	36
II.6.3 Domaines d'application.....	40
II.7. Conclusion.....	41
III.1 Introduction.....	43
III.2 Environnement de travail .....	43
III.3 Squelette de la main .....	44
III.4 Structure d'Évaluation de l'âge osseux à l'aide de modèles de vision transformer .	48
III.5 Résultats .....	53
III.6. Conclusion .....	62
Conclusion générale.....	64
Références.....	65

## Liste des figures

### *Chapitre I : Généralités sur l'imagerie médical*

<b>Figure I.2 :</b> Échographie.....	5
<b>Figure.I.3 :</b> IRM-cérébrale-coupes-axiales-pondérées-en-diffusion.....	6
<b>Figure I 4 :</b> Imagerie médical Rayon X.....	7
<b>Figure.I.5 :</b> Radiographie thoracique .....	8
<b>Figure I.6 :</b> Scanner cérébral.....	9
<b>Figure I.7 :</b> Imagerie PA_CHSF-Médecine-Nucléaire .....	11
<b>Figure I.8.:</b> Spect-brain-point tomographie .....	12
<b>Figure I.9 :</b> Scan-TEP lors d'une écoute subjective (a) et analytique (b).....	13
<b>FigureI.10 :</b> Scintigraphie thyroïdienne .....	14
<b>Figure I.11 :</b> La Scintigraphie osseuse .....	14
<b>Figure I.12 :</b> Scintigraphie-myocardique.....	15

### *Chapitre II : Apprentissage par transfert & Vision Transformateur*

<b>Figure II.1 :</b> Machine learning vs deep learning.....	21
<b>Figure II.2 :</b> SVM.....	23
<b>Figure II.3 :</b> Réseaux de neurone biologique.....	24
<b>Figure II.4 :</b> Architecture de réseau neuronal simple.....	25
<b>Figure II. 5 :</b> Fonctions d'activations.....	26
<b>Figure II.6 :</b> Un PMC.....	26
<b>Figure II.7 :</b> Machine learning Vs transfer learning .....	27
<b>Figure II.8 :</b> Transfer Learning .....	29
<b>Figure II.9 :</b> Type de Transfer Learning .....	32
<b>Figure II.10 :</b> Apprentissage inductif et transfert inductif .....	32
<b>Figure II.11 :</b> Aperçu du Modèle ViT .....	34
<b>Figure II.12 :</b> Architecture de codeur transformateur.....	35
<b>Figure II.13 :</b> Les Variantes du modèle Vision Transformer .....	36
<b>Figure II.14 :</b> Image 16x16:transformateurs pour la reconnaissance d'images.....	38

### ***Chapitre III : Implémentation et Résultats***

<b>Figure III.1:</b> Subdivision du squelette de la main vue dorsale et antérieur.....	<b>44</b>
<b>Figure III.2:</b> Squelette de la main vue antérieure .....	<b>45</b>
<b>Figure III.3 :</b> Répartition du nombre des individus male/femelle dans la base de données...46	
<b>Figure III.4 :</b> Extraits du nombre des individus male/femelle dans la base de données.....47	
<b>Figure III.5 :</b> Représentation de la perte en mois.....	<b>53</b>
<b>Figure III.6 :</b> La partie zoom de la perte en mois.....	<b>54</b>
<b>Figure III.7 :</b> Représentation d'erreur absolue a un mois.....	<b>55</b>
<b>Figure III.8 :</b> La partie zoom d'erreur absolue a un mois.....	<b>55</b>
<b>Figure III. 9 :</b> Exemple de répartition par âge et sexe dans le résultat.....	<b>58</b>
<b>FigureIII.10 :</b> La variation d'erreur en fonction des époques pour VGG19.....	<b>59</b>
<b>Figure III.11 :</b> Variation d'erreur en fonction des époques pour ResNet.....	<b>60</b>
<b>FigureIII.12 :</b> Prédiction de l'Age osseux.....	<b>61</b>

## Listes des tableaux

### **Chapitre I : Généralités sur l'imagerie médicale**

Tableau I.1 : Types et Modalités en Imagerie Médicale. ....4

### **Chapitre II : Apprentissage par transfert & Vision transformer**

Tableau II.1 : Différents types d'apprentissage machine.....21

### **Chapitre III : Implémentation et résultats**

Tableau III.1 : Comparaison entre les résultats du modèle.....61

## Listes des Acronymes

**CNN** : Convolution neural network

**DL** : Deep Learning

**IRM** : Imagerie par résonance magnétique

**MLP** : multi-layer perceptron

**NLP** : Natural Language Processing

**NN** : Neural Network

**SVM** : Support Vector Machines

**TEMP** : Tomographie d'Émission Monophonique

**TEP** : Tomographie par Émission de Positons

**ViT** : Vision Transformer

## **Introduction Générale**

Actuellement, le domaine médical a connu un large développement à tous les niveaux, car de nouvelles technologies ont été découvertes qui ont contribué au diagnostic et au traitement de nombreuses maladies, en raison des grands progrès de l'information.

Le terme intelligence artificielle fait référence à des systèmes ou des dispositifs qui simulent l'intelligence humaine pour effectuer des tâches et qui peuvent s'améliorer en fonction des informations qu'ils collectent. L'intelligence artificielle est appliquée dans divers domaines, et le domaine médical est l'un des domaines qui a grandement bénéficié de ces progrès en particulier.

Dans ce contexte, les technologies de l'information sont devenues un outil essentiel et important dans les pratiques médicales modernes en général, telles que le diagnostic, notamment par l'analyse du signal, et l'imagerie médicale.

Le Transformateur de Vision, ou ViT est un transformateur utilisé dans le domaine de la vision par ordinateur, c'est un modèle de classification d'images qui utilise une architecture de type Transformer sur des parties de l'image.

L'objectif de notre travail est de visualiser et d'estimer l'âge des os de la main à l'aide de la transformateur de vision pour ce faire nous allons utiliser les modèles (ViT16, VGG19 et ResNet50) et on utilise la base des données (bone age assessment) qui contient 12611 images training) et 803 images validation

Ce mémoire est organisé en 3 chapitres, le premier chapitre est sur les généralités de l'imagerie médicale, le deuxième chapitre sera une description sur les différents types d'apprentissage (automatique -profond- par transfert) et le transformateur de vision, puis on abordera le dernier chapitre qui est le système réalisé où on va présenter en détails les résultats obtenus.

# *Chapitre I*

## *Généralités sur l'imagerie médicale*

## **I.1 Introduction**

Chaque année, les entreprises médicales révèlent de nombreuses technologies d'imagerie médicale pour les meilleures dans ce domaine en raison de leur grande importance pour la santé humaine. L'imagerie médicale est une méthode de diagnostic utilisée par les médecins pour détecter et traiter les maladies et c'est un diagnostic qui vise à regarder à l'intérieur du corps et à découvrir l'emplacement de la maladie. D'autre part, on peut dire que le terme imagerie médicale désigne les techniques et opérations utilisées pour obtenir des images du corps humain ou de parties de celui-ci à des fins diagnostiques, thérapeutiques et de recherche. Dans ce chapitre nous allons présenter quelques concepts et notions fondamentaux concernant le domaine d'imagerie médicale.

## **I.2 Historique de l'imagerie médicale**

Au 19<sup>ème</sup> siècle, Wilhelm Conrad Röntgen (1845-1923) fût le premier physicien allemand a élaboré la première technique d'imagerie médicale. Cette technique est la radiographie a été découverte par ce physicien par pur hasard dans l'obscurité de son laboratoire, ce dernier étudiait l'électricité dans un tube de verre muni d'électrodes et dans lequel on fait le vide, quand il vit en interposant sa main devant le tube, ses propres os

De nombreuses études ont été réalisées et ont porté sur le principe de la radiographie. Des améliorations ont été effectuées jusqu'à la radiographie aux rayons X.

Pierre - Marie Curie et Henry Becquerel, ont mis naissance à la radioactivité naturelle en 1943, Irène et Frédéric Joliot-Curie, ont développé la médecine nucléaire avec la scintigraphie et la tomographie par émission des positons (TEP) [BIE, 2000].

Inge Elder (1911-2001), un cardiologue suédois a eu l'idée en 1955 d'utiliser le principe du principe de propagation des ultrasons utilisée par les SONAR (Sound Navigation Ranging) dès 1915 pour détecter les icebergs suite au naufrage du Titanic. Ce principe a permis à Elder de diagnostiquer les sténoses mitrales.

À partir des années soixante-dix et précisément en 1972, les radiologues britanniques Allan Mc Kornack (1924-1998) et Godfrey N. Hounsfield (1918), on mit au point un scanner grâce à l'apport de l'informatique et du traitement numérique de l'image.

En 1973, le chimiste américain Paul Lauterbur a réalisé le premier cliché d'un animal en utilisant la méthode d'imagerie par résonance magnétique. Cette méthode découle de la découverte de la résonance des noyaux des atomes (résonance nucléaire) en 1945 par Edward Purcell et Félix Bloch, prix Nobel en 1952. [1].

## I.4 Modalités de l'imagerie médicale

Les différents appareils d'acquisition d'images médicales sont nombreux et permettent d'obtenir des informations différentes selon le procédé physique utilisé pour observer les tissus du corps humain. Certains procédés apportent une information anatomique (Imagerie Structurale) tandis que d'autres détectent la fonctionnalité des organes en offrant une carte d'activité. [2].

Selon la nature des recherches, on distingue comme types des images médicales :

**-L'imagerie structurale** recueille des informations sur la morphologie des organes, tissus ou cellules, comme la taille, le volume, la localisation, ou encore la forme d'une éventuelle lésion, de la zone étudiée (IRM, TDM, RX,).

**-L'imagerie fonctionnelle** encore appelée imagerie métabolique et moléculaire, vise à étudier le fonctionnement et l'activité physiologique des organes, tissus ou cellules par la technologie TEP.

**-L'imagerie interventionnelle** correspond à l'ensemble des actes médicaux réalisés sous le contrôle de dispositifs d'imagerie, comme le repérage, le guidage, le contrôle optimal du geste médical [3].

Le tableau I1 résume les types et les différentes modalités et techniques d'acquisition en imagerie biomédicale.

Type d'imagerie	Modalités		Principaux axes
Structurelle Visualiser la structure et la forme des organes, tissus ou cellules	<b>Rayons X</b> ( <i>Radiologie, angiographie, ostéodensitométrie</i> )	<b>Inf</b> <b>ormat</b> <b>I</b> <b>que</b> <b>M</b> <b>e</b> <b>d</b> <b>i</b> <b>c</b> <b>a</b> <b>l</b>	Neurologie
	<b>Champs magnétiques</b> ( <i>IRM etc.</i> )		Cancérologie
	<b>Médecine nucléaire</b> ( <i>TEP, TEMP, scintigraphie etc.</i> )		Pneumologie
Fonctionnelle Étudier le fonctionnement des organes, tissus ou cellules	<b>Ultrasons</b> ( <i>Échographie, Doppler etc.</i> )		Ophthalmologie
	<b>Rayonnement lumineux</b> ( <i>opt. imagerie optique, fluorescence, etc...</i> )		Gynécologie, obstétrique
Interventionnelle Procédures diagnostiques et thérapeutiques effectuées sous contrôle d'imagerie	<b>Électrophysiologie</b> ( <i>EEG, ECG, MEG, etc.</i> )		Cardio-vasculaire
	<b>Radiations</b> ( <i>Radiothérapie etc...</i> )		Ostéoarticulaire
	<b>Imagerie biologique</b>		Dentaire

Tableau I1 : Types et Modalités en Imagerie Médicale.

## I.5 Description des différentes techniques d'imagerie médicale

### I.5.1 Échographie

Elle repose sur l'utilisation d'ondes ultrasonores, dont la propagation à des vitesses différentes dans des tissus différents, permet après traitement numérique de l'image, la visualisation en temps réel, des organes et tissus. La sonde échographique est un émetteur-récepteur relié à un ordinateur qui numérise les données intégrées par la sonde et les codes en images selon une échelle de gris. L'échographie utilise parfois des produits de contraste. Elle est non invasive, opérateur-dépendant, accessible (coût, infrastructure et disponibilité) et sans contre-indications. L'échographie doppler est le seul examen de routine permettant l'étude des organes et des vaisseaux en temps réel (cœur, tube digestif...), ce qui permet d'ajouter à des critères anatomiques de normalité ou pathologiques, des critères fonctionnels d'une valeur diagnostique non négligeable (déplacement des globules rouges dans un vaisseau). .[4]



Figure I.1 : Échographie [3].

### I.5.2. Imagerie par résonance magnétique (IRM)

Il s'agit d'une imagerie liée à l'émission de signal (par restitution d'énergie), par des protons qui résonnent en phase dans un champ magnétique, sous l'impulsion d'une onde de radiofréquence (rf) ; la fréquence de Larmor de cette onde rf doit être égale à la fréquence de

résonance des protons dans un champ magnétique donné. En application clinique l'intensité de ce champ magnétique varie de 0,3 à 1,5 Tesla (1 tesla (T) = 10000 Gauss. À titre de comparaison l'intensité du champ magnétique de l'écorce terrestre est de 1 Gauss (G). Cet examen est non irradiant mais a de potentiels effets secondaires non encore évalués du fait du manque de recul. L'IRM a des capacités d'études anatomiques dans un plan donné de l'espace (coronal, axial, sagittal) sans nécessité de reconstruction comme en TDM. [4]

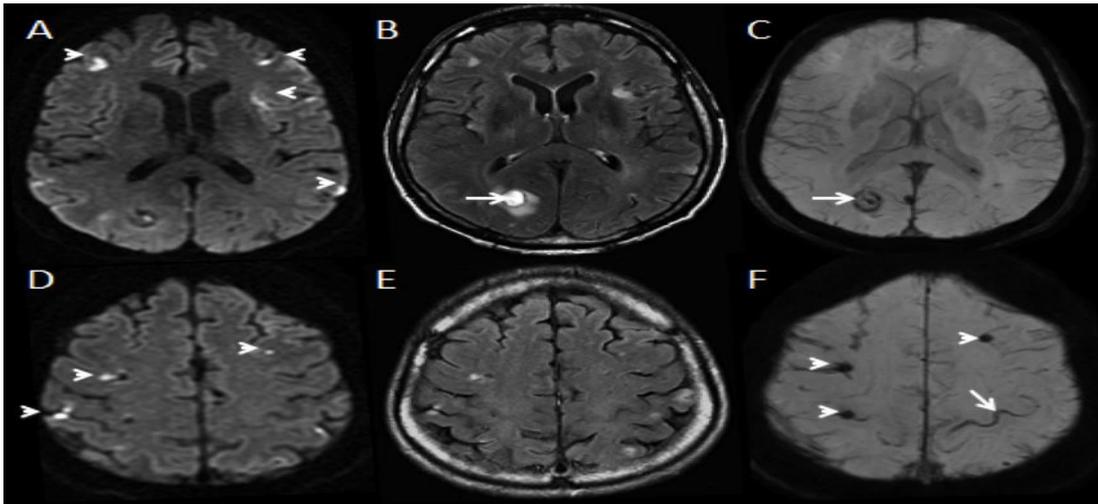


Figure.I.2 : IRM-cérébrale-coupes-axiales-pondérées-en-diffusion

### I.5.3. Image rayons X

Les rayons X sont une forme de rayonnement électromagnétique, au même titre que la lumière visible, l'ultra-violet, l'infra-rouge, les micro-ondes, les ondes radio ou les rayons gamma. Les premières versions de tubes à rayons X voient le jour au début du XX<sup>ème</sup> siècle et, depuis maintenant un siècle, le principe physique régissant leur fonctionnement reste le même. Les matériaux et la technologie ont, par contre, largement évolué et permettent des régimes de fonctionnement plus

Soutenus, ainsi que de meilleurs rendements de conversions énergétiques. Peuvent être produits de deux manières différentes. La première manière est celle liée aux changements d'orbite des électrons. Des rayons X sont produits par des transitions électroniques qui font intervenir les couches internes proches du noyau. Ces transitions peuvent être données lors d'un rayonnement incident de rayons X vers l'atome ou bien par un bombardement d'électrons, fait qui provoquera l'excitation de l'atome (éjection des électrons). Si l'électron éjecté est proche du cœur, un électron de la périphérie va occuper sa position (Processus de désexcitation) en

émettant un photon qui va appartenir au domaine X. La deuxième façon de produire des rayons X est basée sous le principe d'accélération (Freinage et changement de trajectoire) d'électrons. Un tube à rayons X sera alors nécessaire. Le principe est le suivant : Lorsqu'on applique une haute tension électrique (de l'ordre de 20 à 400 kV) entre deux électrodes, un courant d'électrons, de la cathode vers l'anode (appelée aussi anticathode ou cible), est produit. Lorsque le faisceau d'électrons avance vers la cible, ils sont freinés par les atomes de celle-ci, en provoquant un rayonnement continu de freinage ou de Bremsstrahlung (phénomène décrit plus bas), dont une partie du spectre est dans le domaine des rayons X. En même temps, les rayons résultants provoquent l'excitation des atomes de la cible qui, telle que décrit au début de cette section, réémettent un rayonnement X grâce au phénomène de la fluorescence X. Le tube donnera un spectre résultant de la superposition du rayonnement de freinage et de la fluorescence X de la cible [5]



**Figure I.3 :** Imagerie médical Rayon X. [dataset bon age training]

#### **a. Radiographie conventionnelle et numérique**

Elle se base sur l'utilisation du rayonnement gamma ionisant, potentiellement toxique à faible dose et manifestement toxique à haute dose. Elle utilise un tube à rayons X comme émetteur. Le faisceau de rayons X est atténué dans l'organisme, selon la densité et l'épaisseur du tissu traversé ; le recueil des données sur film-écran ou par un système de numérisation est à l'origine de la formation de l'image (différence d'atténuation).[6]

### **b. radiographie interventionnelle**

L'imagerie permet aujourd'hui la réalisation des gestes thérapeutiques guidés avec précision à l'intérieur de l'organisme. Ainsi elle permet de faire des ponctions dirigées ou des cathétérismes endovasculaires, la mise en place de "stents", la réalisation des gestes thérapeutiques comme la nucléolyse etc. [4]

### **c. La radiographie**

La radiographie est la formation, sur un film photographique, d'une image d'un objet qui a été interposé entre le film et une source de Rayon X. Le radiogramme est l'image obtenue de cette manière. Le terme radiologie (terme issu du latin radius : Rayon et du grec logos : discours) est un terme plus global qui désigne l'utilisation des Rayons X dans un but diagnostique radiographique ou thérapeutique (cancer). Le terme de radiographie est maintenant utilisé dans un sens plus large d'imagerie médicale (remplaçant de ce fait la radiologie diagnostique). Elle est notamment utilisée en orthopédie, en rhumatologie et en orthodontie ou elle permet d'étudier les traumatismes osseux (fractures, etc.), les déformations du squelette ou les implantations dentaires. La pneumologie y a aussi recours (radio de poumons). Chez la femme, la radiographie du sein (mammographie) est devenue un examen systématique de prévention du cancer du sein. Il est aussi possible de visualiser certains organes ou parties creuses, habituellement invisibles aux rayons X, en les « remplissant » d'un produit de contraste, opaque aux rayons X : c'est la radiographie de contraste.[7].

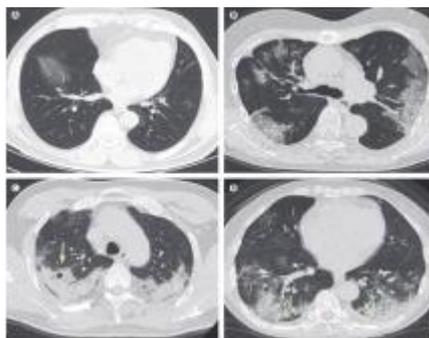


**Figure I.4** : Radiographie thoracique.[9].

#### d. Tomodensitométrie (scanner)

Un scanner, appelé tomographie axiale assistée par ordinateur (TAO), est un appareil radiographique qui combine des rayons X pour imager le plan de coupe d'un objet, en particulier le corps humain. Inventé par le britannique Godfrey Hounsfield dans les années 1970. Un scanner est une technique utilisée pour créer une image en "section" continue d'une zone numérisée. Ces images sont le résultat d'un traitement informatique et de reconstruction des informations reçues par les capteurs de l'appareil. Les scanners peuvent montrer des changements de volume ainsi que des anomalies structurelles dans les tissus ou les organes. La technologie est utilisée dans divers domaines d'application :

- La cancérologie pour la détection de nodules, le contrôle de la réponse à la chimiothérapie et le bilan d'extension des cancers (pulmonaires, bronchiques, pelviens, ORL),
- La chirurgie pour localiser précisément la zone où l'on envisage l'intervention,
- Traumatologie pour la recherche sur les traumatismes et les fractures. Les scanners d'aujourd'hui continuent d'évoluer pour tenter d'aller plus loin dans le diagnostic médical. Pour tendre vers cet objectif, ce mode doit être ouvert aux présentations en trois démontions. Cette nouvelle visualisation apporte de nombreux avantages au diagnostic médical. Mais cette nouvelle technologie n'est possible qu'avec une console de post-traitement. Ces consoles sont des ordinateurs qui utilisent des algorithmes spécifiques pour pouvoir reconstruire des images numérisées en trois dimensions. Aujourd'hui, ils jouent un rôle de plus en plus important et sont à la base des données acquises avec les scanners. **.[8]**.



**Figure I.5** : Imagerie médicale d'un scanner.[8].

#### I.5.4. Imagerie nucléaire

L'invention de la médecine nucléaire s'est reposée sur la découverte des isotopes radioactifs par Frédérique et Irène Joliot-Curie en 1934 et sur le développement de la première gamma-caméra en 1957 par Hal Anger. La médecine nucléaire regroupe plusieurs techniques

comme la scintigraphie : on injecte au patient un produit radioactif, porté par une molécule ayant des propriétés d'association aux tissus biologiques. En effet, ces isotopes s'accumulent dans certains organes précis que le médecin cherche à étudier. Ils émettent alors des rayonnements, gamma le plus souvent (mais d'autres techniques peuvent être utilisées), détectés par une gamma-caméra à scintillation. Les radio-isotopes sont des éléments non-stables subissant des désintégrations ou des désexcitations pour arriver à un état stable. Ils sont donc des isotopes radioactifs. Le plus utilisé en médecine est le technétium de faible demi-vie. Il provient de la désintégration du molybdène, un métal de transition. Une fois injecté au patient, le technétium se désexcite en technétium stable et émet des rayons gammas d'énergie 140 Kev -idéale puisqu'assez énergétique pour traverser les tissus vivants, et donc pouvant être captés par le gamma caméra-. Tout d'abord, les photons passent par un collimateur à trous parallèles qui permet leur sélection. En effet, seuls les photons frappant le collimateur de façon perpendiculaire à sa surface sont utiles pour la formation de l'image. Sans collimateur, il est impossible de déterminer l'origine des photons et ainsi de reconstruire une image. Ensuite, le cristal scintillant convertit les photons gammas ou bêtas en lumière visible. Enfin des photomultiplicateurs augmentent le nombre de photons et convertissent le signal optique en signal numérique. L'électronique et l'informatique calculent la position et l'énergie des photons gammas et reconstruisent numériquement l'image. On peut améliorer la résolution de l'image par une collimation énergétique : seuls les photons ayant l'énergie caractéristique du radio-traceur sont retenus pour la formation de l'image.[8].

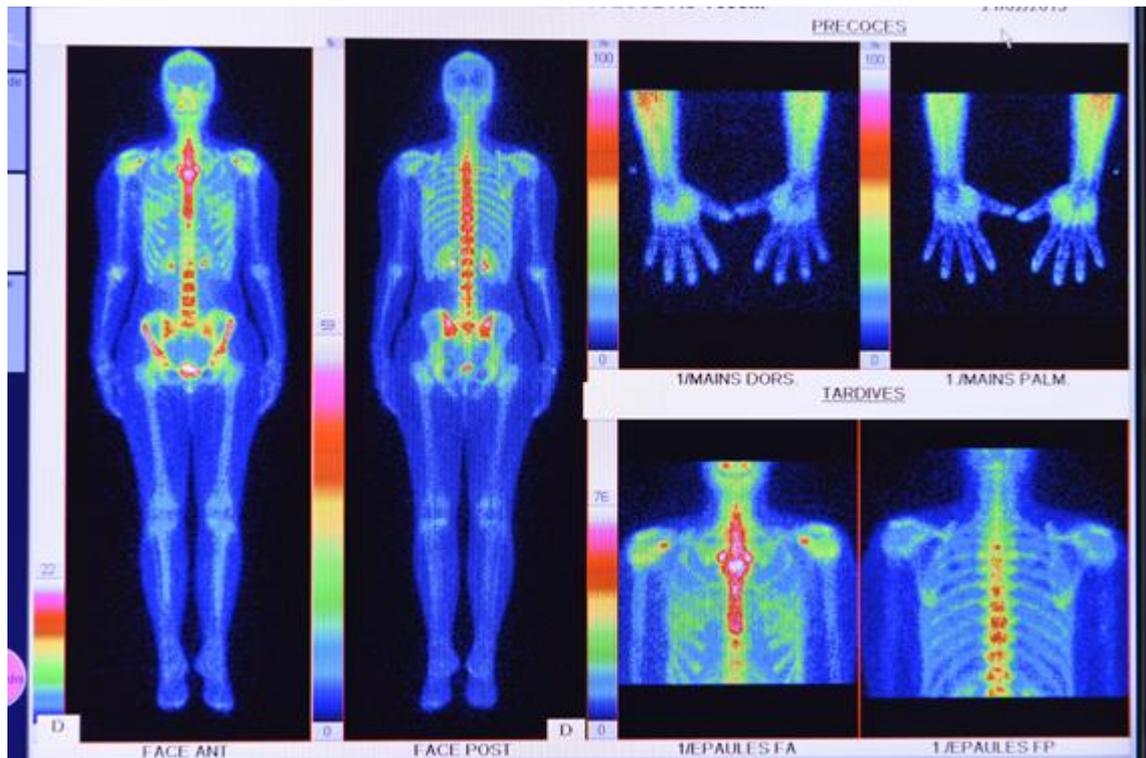
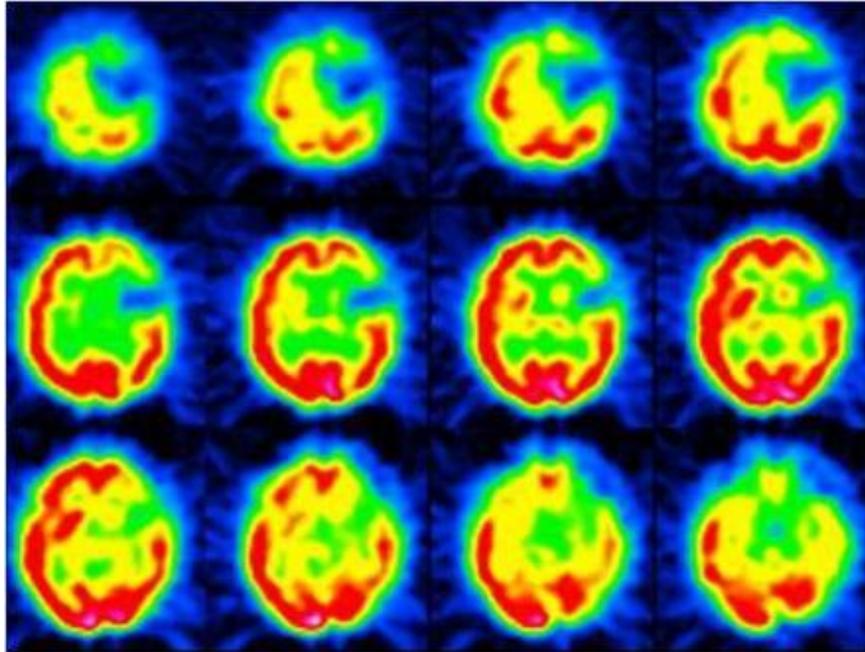


Figure I.6 : Imagerie PA\_CHSF-Médecine-Nucléaire

- **Tomographie D'émission Monophonique (TEMP)**

Le principe de la TEMP est de suivre l'évolution dans le corps humain d'un radioélément, qui est dans ce cas un émetteur naturel de simples photons  $\gamma$ . Le radioélément est en général fixé sur une molécule (appelée radio-traceur). Bien qu'elle possède une moins bonne résolution spatiale que la TEP en raison de la détection de photons simples (résolution spatiale de l'ordre du cm), la TEMP est une technique intéressante et nettement moins onéreuse que la TEP. De ce fait, elle est plus facilement disponible sur un grand nombre de sites. La qualité des images reconstruites en TEMP résulte classiquement d'un compromis entre sensibilité de détection et donc niveau du rapport signal / bruit et résolution spatiale. Différents paramètres d'acquisition permettent d'ajuster ce compromis : la dose de traceur injectée, la taille du pixel (échantillonnage spatial), le temps d'acquisition, le nombre de projections, l'algorithme de reconstruction et le filtrage. En moyenne, les systèmes actuels permettent d'obtenir des résolutions spatiales des données reconstruites entre 10 et 15mm. [8].



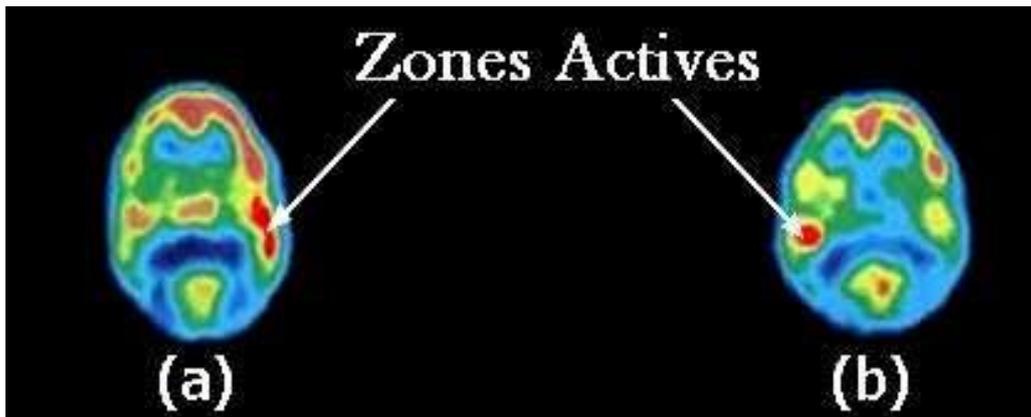
**Figure I.7 :** Spect-brain-p point tomographie. [imageetmultimedia.wordpress.].

- **La Tomographie par Émission de Positons (TEP)**

En médecine nucléaire, la tomographie par émission de positons (TEP) est une méthode qui permet de mesurer l'activité métabolique d'un organe grâce aux émissions produites par les positons issus de la désintégration d'un produit radioactif injecté au préalable [9]. La TEP repose sur le principe général de la scintigraphie qui consiste à injecter un traceur qui est généralement sous forme

Sous forme de glucose [10]. Ce traceur est marqué par un atome radioactif qui

Émet des positons dont l'annihilation produit elle-même des photons [9]. La détection de la trajectoire de ces photons par le collimateur de la caméra TEP permet de localiser le lieu de leur émission et donc la concentration du traceur en chaque point de l'organe. Cette information quantitative est représentée sous la forme d'une image faisant apparaître les zones de forte concentration du traceur. Ainsi la TEP permet de visualiser les activités du métabolisme des cellules : on parle d'imagerie fonctionnelle par opposition aux techniques d'imagerie dite structurale comme celles basées sur les rayons X (radiologie ou scanner) qui réalisent des images de l'anatomie. Par conséquent, la tomographie par émission de positons est un outil diagnostique qui permet de déceler certaines pathologies qui se traduisent par une altération de la physiologie normale comme les cancers ; en effet, Comme les cellules tumorales sont plus actives que les cellules normales, elles consomment davantage de glucose. La TEP permet ainsi d'obtenir des images précises de la répartition du glucose radioactif dans l'organisme et donc de localiser des cellules cancéreuses [9]



**Figure I.8 :** Scan-TEP lors d'une écoute subjective (a) et analytique (b) [9].

La TEP est aussi utilisée en recherche biomédicale, par exemple en imagerie cérébrale où elle permet de révéler les régions actives du cerveau lors de différentes activités cognitives. Dans la figure I.8 nous remarquons que l'écoute subjective ou analytique d'une même pièce de musique par un même sujet active respectivement l'hémisphère droit ou l'hémisphère gauche.

### I.5.5. Scintigraphie

Parfois appelée gammagraphie, découverte en 1955, permettant de visualiser, après injection d'une substance radioactive, l'intérieur de l'organisme, afin de détecter d'éventuels troubles de fonctionnement des organes. La scintigraphie consiste à observer, à l'aide d'une caméra sensible au rayonnement radioactif, le déplacement dans le corps humain d'une substance radioactive, appelée scintillant. Celle-ci, administrée par voie orale ou par injection, se fixe sélectivement sur l'organe qui intéresse le praticien. On peut ainsi réaliser des scintigraphies osseuses, cérébrale, rénale, etc. Après traitement informatique du signal vidéo, une image apparaît sur un écran, à partir de laquelle on peut observer la structure de l'organe choisi ou le mouvement d'un fluide à l'intérieur de celui-ci (le sang dans le cas d'une scintigraphie cardiaque, par exemple). [8]

#### a) Scintigraphie thyroïdienne

La scintigraphie thyroïdienne (Figure I.9) repose sur l'utilisation d'un produit radioactif qui se fixe sélectivement sur la thyroïde pour une recherche d'une anomalie morphologique ou fonctionnelle. Deux radioéléments sont utilisés pour cet examen : l'iode-123 et le technétium-99m. À l'aide de la scintigraphie thyroïdienne, il est possible de visualiser des régions de la

thyroïde qui captent moins le traceur (hypofixations appelées nodules froids) ou qui captent plus le traceur (hyperfixations appelées nodules chauds). [9].



**Figure I.9 :** La scintigraphie thyroïdienne.

#### **b) Scintigraphie Osseuse**

La scintigraphie osseuse repose sur la fixation dans les structures osseuses de molécules phosphatées marquées au technétium-99m. Le radiopharmaceutique est injecté par voie intraveineuse et aucune préparation particulière du patient n'est nécessaire avant l'examen. Le traceur circule dans le sang et sa captation par le squelette est maximale au bout de trois heures, ce qui impose un temps d'attente équivalent entre l'injection et le début de l'examen. [9].



**Figure I.10 :** La Scintigraphie osseuse.

### c) Scintigraphie Cardiaque

La scintigraphie cardiaque (Figure I.11) permet d'évaluer la circulation du sang au niveau du muscle cardiaque (évaluation de la perfusion) et donne des renseignements sur sa fonction et ses capacités de contraction. La comparaison de la fixation du radiotraceur dans le muscle cardiaque pour deux examens, un d'effort et un de repos, permet de diagnostiquer ou d'exclure des maladies cardiaques. Les traceurs radioactifs couramment utilisés pour cet examen sont le Thallium-201 et le Technétium-99m. [9].

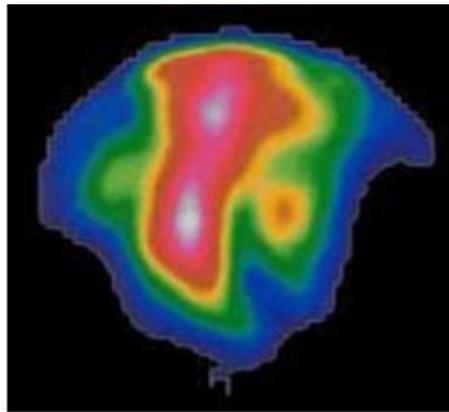


Figure I.11 : Scintigraphie-cardiaque

### I.5.6. Imagerie thermique

L'imagerie thermique est largement utilisée dans le domaine militaire pour ses capacités de vision diurne et nocturne et sa longue portée d'observation. Cette technologie est basée sur la détection passive dans l'infrarouge. En conditions météorologiques dégradées ou quand la cible est partiellement dissimulée par du feuillage ou des filets de camouflages militaires, elle devrait être à court terme de plus en plus complétée par un système d'imagerie active. Cette technologie est essentielle pour l'imagerie à longue portée. La technique d'imagerie dite flash 2D est basée sur une source laser impulsionnel qui illumine la scène et sur une caméra rapide synchronisée qui constitue le système d'imagerie. Ces deux technologies sont bien éprouvées en présence de conditions météorologiques claires. Les modèles TRM4 (imagerie thermique) et PERFIMA (imagerie active) sont capables de prédire correctement les performances de tels systèmes par beau temps. En revanche, en conditions dégradées telle que la pluie, le brouillard ou la neige, ces modèles deviennent non pertinents [11].

### **I.5.7. Imagerie hyperspectrale**

L'imagerie hyperspectrale est une technique qui permet d'obtenir une image d'un sujet dans des centaines de bandes spectrales contiguës ce qui permet de produire un spectre de réflectance complet pour chacun des éléments individuels (pixels) composant l'image. La résolution spatiale d'une image hyperspectral est moyenne mais par contre la résolution spectrale est généralement inférieure à 10 nanomètres. Cette technique se différencie de l'imagerie spectrale, technique ancienne qui restitue une image dans un nombre de bandes spectrales limité (jusqu'à 15), larges et discontinues<sup>1</sup>. L'imagerie hyperspectral est utilisée principalement en lumière visible et proche infrarouge (0,4-2,5  $\mu\text{m}$ ). Mais elle peut également être utilisée dans le moyen infrarouge (3–5  $\mu\text{m}$  et 8–12  $\mu\text{m}$ , bandes II et III) et l'infrarouge thermique et des prototypes sont évalués pour l'observation des micro-ondes. [12].

### **I.6. Standard DICOM (Digital Imaging and Communication Medicine)**

Digital Imaging and Communication Medicine couramment abrégé DICOM, est une norme standard pour la gestion informatique des données issues de l'imagerie médicale. Un fichier DICOM peut être vu comme une enveloppe contenant les informations sur un patient, sur l'examen qu'il a subi, et contenant les divers clichés réalisés. Il contient des informations de type divers dans son en-tête, suivies par les données brutes. Le préambule contient 128 octets généralement mis à zéro suivi de 4 octets pour y Écrire "DCM ». À la suite du préambule commence une succession d'informations diverses : l'en-tête qui contient des données textuelles concernant le patient (âge, poids, sexe.), l'examen (région du corps explorée), la technique utilisée (IRM, scanner,), le nombre de clichés ainsi que leur taille. [2].

### **I.7. Avantages et inconvénients de l'image médicale [2].**

#### **-Radiologie**

La radiologie a pour avantage d'être rapide, simple à utiliser et peu coûteuse. Il s'agit d'un des moyens les plus utilisés. Cependant, la qualité d'image n'est pas la meilleure. En effet, la radiologie fournit des images des organes dans toute leur épaisseur ce qui ne permet pas de voir toutes les liaisons. La radiologie comporte des risques pour le patient et le manipulateur. Pour limiter ces risques :

- le patient ne doit pas réaliser des examens radiologiques trop fréquemment car cela peut provoquer des altérations graves au niveau de certains tissus.

- les manipulateurs doivent se protéger en utilisant des protections plombées (tablier, gants, vitres...).

### **-Scanner**

L'avantage du scanner est de pouvoir fournir une image en 3D en quelques secondes et de très bonne qualité. L'image est très détaillée et permet d'apprendre beaucoup sur l'anatomie.

Il y a néanmoins deux inconvénients :

- le premier est le cout de l'examen, qui est Élevé a cause du cout de l'appareil.
- Le deuxième est celui de l'exposition aux rayons X, qui peut se révéler dangereux sur le long terme.

### **-Echographie**

L'échographie médicale ne présente aucun danger "elle peut dès lors être répétée sans problème", c'est la seule technique permettant d'avoir une image du fœtus

Sans nuisances. Il n'y a pas d'allergie ni de contre-indication concernant cette technique.

- Elle est indolore pour le patient, et ne nécessite, sauf cas a part, ni hospitalisation, ni anesthésie.

### **-IRM**

L'IRM n'est pas une méthode invasive c'est à dire qu'elle n'a aucun effet sur les patients contrairement à la radiologie qui irradie légèrement ces derniers. L'IRM nécessite un produit de contraste qui est en majorité bien supporté par les patients on dénombre moins de 1% d'allergie au produit de contraste.

L'IRM est une technique qui permet de faire des clichés d'une très haute qualité. En revanche l'IRM ne présente pas que des avantages :

Il ne faut pas être claustrophobe car l'examen peut durer un certain temps. De plus le principe de l'IRM repose sur la quantité d'atomes d'hydrogène des milieux observés. Par conséquent, l'IRM n'est utile que pour observer des tissus mous O2 en eau donc certes extrêmement utiles pour les zones cérébrales et cardiaques mais les parties osseuses pauvres en hydrogène ne sont pas observables à l'IRM. Les patients devront se tourner vers d'autres techniques d'imagerie médicales.

## **I.8. Conclusion**

L'imagerie médicale est l'un des domaines qui a connu un grand développement ces dernières années. En raison de l'évolution rapide des techniques d'imagerie médicale, les médecins disposent de nombreuses possibilités d'examens, adaptées aux besoins spécifiques du patient. Une bonne utilisation de l'imagerie médicale est essentielle pour un bon diagnostic ou un bon suivi. Nous avons abordé dans ce chapitre le volet généralité de l'imagerie médicale, et ses différents types et modalités. Dans le chapitre suivant nous présentons l'apprentissage au profond et plus spécialement l'apprentissage par transfert et le transformateur de vision en détail.

*Chapitre III*  
*Apprentissage par transfert*  
*& Vision transformer*

## **II. Introduction**

L'avènement de l'apprentissage par transfert a révolutionné le domaine de l'apprentissage artificiel. Il a particulièrement permis d'améliorer la capacité d'apprentissage de la machine grâce à l'utilisation de données massives et hétérogènes. L'apprentissage par transfert (Transfer Learning) est l'un des champs de recherche de l'apprentissage automatique qui vise à transférer des connaissances d'une ou plusieurs tâches sources vers une ou plusieurs tâches cibles. Nous commençons donc ce chapitre par donner quelques notions de base sur l'apprentissage automatique (Machine Learning) et l'apprentissage profond (Deep Learning), en suite nous présenterons l'apprentissage par transfert et après on passera à La transformateur de vision (Vision Transformer (ViT)) qui est un modèle de classification d'images utilisant une architecture de type Transformer sur des parties de l'image. Étant le sujet autour duquel ce chapitre se basera sur l'utilité du ViT.

### **II.2. Machine Learning**

#### **II.2.1 Définition**

L'apprentissage automatique (ou machine Learning) est un sous-domaine de l'intelligence artificielle qui permet aux systèmes de comprendre grâce à son algorithme. Il est basé sur L'apprentissage d'un algorithme à partir de données et Utilisez ces données pour faire des prédictions et apprendre à résoudre des tâches avec ces ordinateurs Concrètement, il n'est pas nécessaire de les programmer. L'objectif de L'apprentissage automatique est d'identifier des structures dans des données souvent trop complexes Difficiles à détecter ou mesurer manuellement. À partir de ces structures, on peut chercher à classer individus, objets, prédire la valeur d'une variable dans une plage, expliquer présence ou absence de fonctionnalités [8].

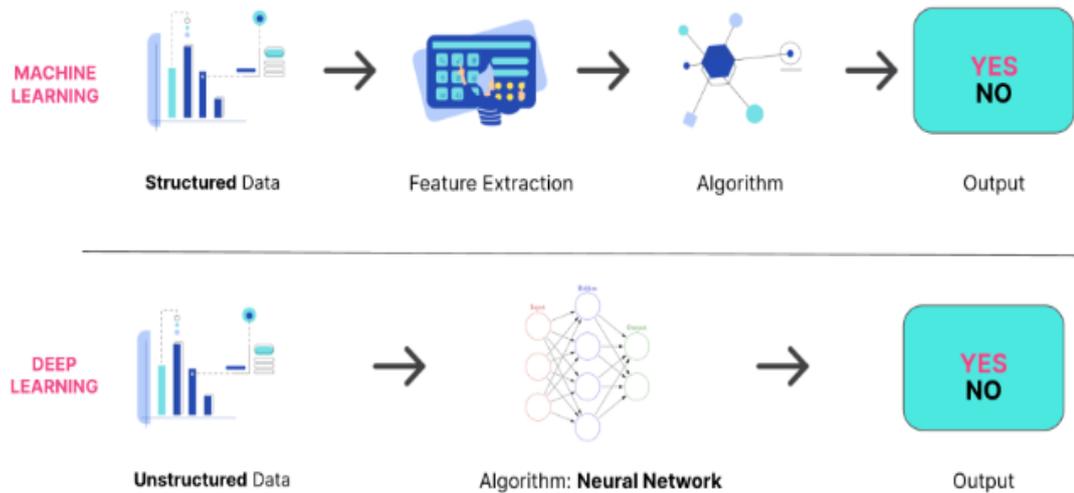


Figure II.1: Machine learning Vs deep learning[jedha.com].

### II.2.2. Différents types d'apprentissage machine

Le tableau suivant montre les différents types de l'apprentissage machine

	Apprentissage supervise	Apprentissage non supervise	Apprentissage par renforcement
Définition	L'algorithme apprend à partir de données labellisées	L'algorithme est entraîné à partir de données non labellisées sans indications particulières	L'algorithme interagit avec son environnement en réalisant des actions et en apprenant de ses erreurs et ses succès
Type de problèmes	Régression et classification	Association et clustering	Bases sur un système de récompense
Type de données	Données labellisées	Données non labellisées	Pas de données fournies au préalable
Approche	Étudie les relations sous-jacentes qui lient les données en entrée aux labels	Découvre les motifs communs au sein des données d'entrées	Apprend une stratégie de comportement en fonction d'expériences passées et des récompenses perçues

Tableau II.1 : Différents types d'apprentissage machine.

### II.2.3. Algorithmes de l'apprentissage machine

- Support Vector Machine (SVM) classificateur
- Arbre de décision
- Gradient Boosting

- Les algorithmes Apriori
- Régression linéaire
- La répartition en K-moyennes (K-means)
- Réseaux antagonistes génératifs
- Classificateur bayésien naïf
- Q-Learning
- Algorithme basé sur un modèle (model-based)
- Random Forest

### **a) Régression linéaire**

Sans aucun doute possible, les algorithmes de régression linéaire sont les plus utilisés par les équipes de data science. Il s'agit d'effectuer des corrélations simples entre deux variables dans un jeu de données. Un ensemble d'entrées et les sorties correspondantes sont examinés et quantifiés pour montrer une relation, par exemple comment le changement d'une variable affecte une autre. Les régressions linéaires sont représentées sous forme de lignes sur un graphique. La popularité de la régression linéaire s'explique par sa simplicité. L'algorithme est facilement explicable, relativement transparent et il y a peu de paramètres à configurer. Bien connu dans la pratique des statistiques, ce type d'algorithmes est souvent utilisé pour prévoir des ventes ou des risques. La régression linéaire est la meilleure solution lorsque « vous cherchez à prédire votre valeur ou une classe », déclare Shekhar Vemuri, CTO de Clairvoyant, une société de conseil en analytique [8].

### **b) Support Vector Machine (SVM) classificateur**

SVM est une méthode dont l'objectif principal est de trouver des séparateurs linéaires qui se séparent entre différentes classes. Le SVM trouve un hyperplan avec la plus grande marge possible. Le SVM prend en entrée un grand nombre d'exemples et les affecte à différentes classes. Les données formées qui sont des exemples considérés comme des points dans l'espace. Le point que nous avons obtenu appartient à l'une des classes qui sont pré-séparées et la distance obtenue est appelée hyperplan. Quand nous avons une nouvelle donnée d'entrée, nous la mettons dans le SVM. Elle sera assignée à l'une des classes après une séparation par la marge. Plus la marge est grande, plus l'erreur de généralisation du classificateur est faible [15].

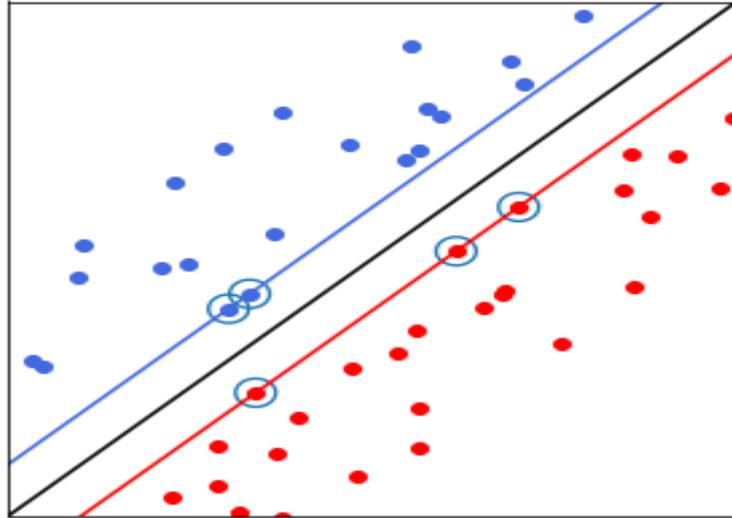


Figure II.2 : SVM.

## II.3. Réseaux de neurones

### II.3.1 Des neurones biologiques aux neurones artificiels [13].

Les réseaux de neurones sont utilisés pour simuler le système nerveux humain. Les tâches d'apprentissage automatique sont accomplies en traitant des unités de calcul dans un modèle appris de la même manière que les neurones humains. La grande vision des réseaux de neurones est de créer L'intelligence artificielle simule des machines informatiques en construisant son architecture système nerveux humain. Prend en compte la nature gourmande en données et en calcul des réseaux de neurones comme un obstacle à son utilisation. Enfin, la force est grandement améliorée L'informatique et une plus grande disponibilité des données a apporté le succès depuis les années 1990 réseau neuronal ajouté, et ce domaine renaît sous le nouveau label de "deep learning". Les réseaux de neurones sont théoriquement capables d'apprendre n'importe quelle fonction de mon sujet avec suffisamment de données d'entraînement, et ils sont devenus des techniques d'apprentissage automatique populaires pour simuler les mécanismes d'apprentissage des organismes vivants. La biologie. Le système nerveux humain contient des cellules appelées neurones Les neurones sont connectés les uns aux autres grâce à l'utilisation d'axones et de dendrites, et les régions de connexion entre les axones et les dendrites sont appelées synapses.

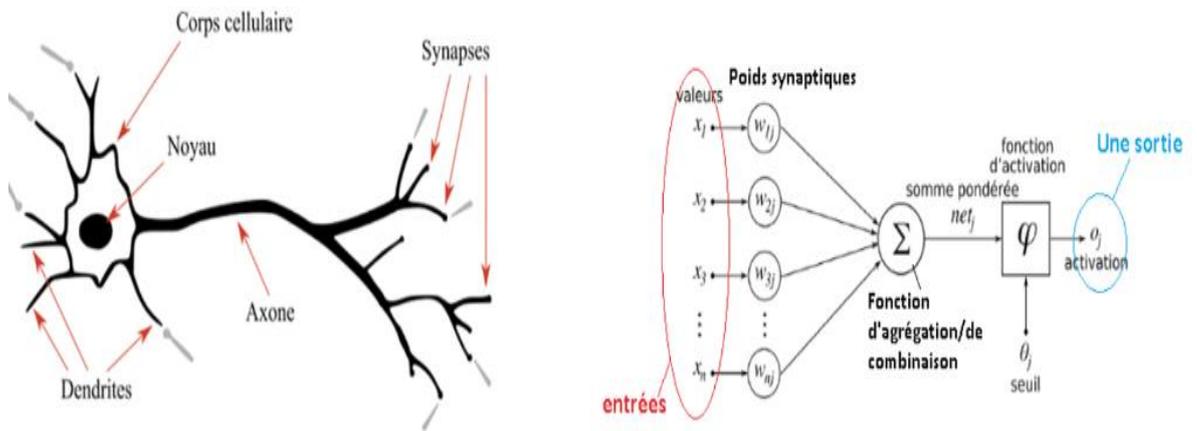


Figure II.3 : Neurone biologique VS Neurone Artificiel.

### II.3.2 Réseaux de neurones artificiels [16].

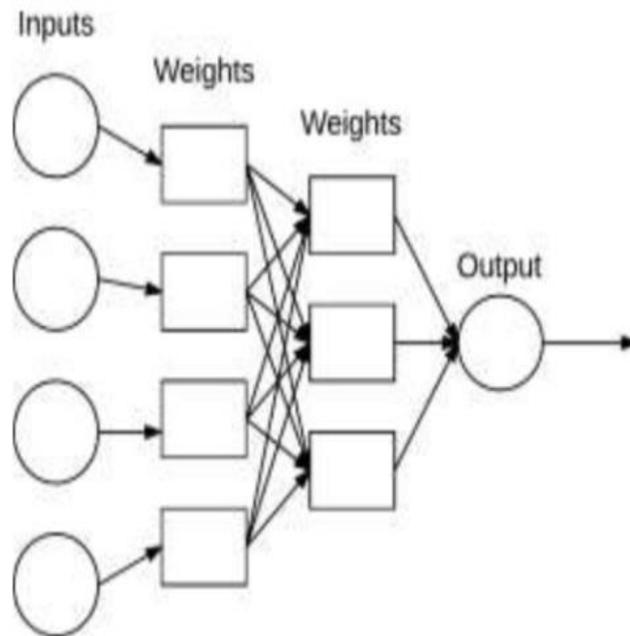
Un réseau de neurones artificiels est un système informatique qui essaie d'imiter ou du moins de s'inspirer des connexions neuronales de notre système nerveux. Réseau de neurones artificiels est aussi appelé réseau de neurones. Les domaines d'application des réseaux neuronaux sont souvent caractérisés par une relation entrée-sortie de la donnée d'information :

- La reconnaissance d'image
- Les classifications de textes ou d'images
- Identification d'objets
- Prédiction de données
- Filtrage d'un set de données

#### ➤ Système de neurone artificiel : [17].

Pour qu'un système soit considéré comme un NN (neural network), il doit contenir une structure de graphe orientée étiquetée où chaque nœud du graphe effectue un calcul simple. Selon la théorie des graphes, un graphe orienté consiste en un ensemble de nœuds (c'est-à-dire sommets) et un ensemble de connexions (c'est-à-dire des arêtes) qui relient des paires de nœuds. Dans (Figure II.4). Chaque nœud effectue un calcul simple. Ensuite, chaque connexion transporte un signal (c'est-à-dire des sorties de calculs) d'un nœud à un autre, étiquetés avec des poids, notés  $w$ . Le degré auquel le signal est amplifié ou atténué. Certaines connexions ont des poids positifs. L'important est d'amplifier le signal, ce qui montre que le signal est très important lors de la lecture. D'autres ont des poids négatifs, ce qui réduit la force du signal, alors spécifiez les. Les sorties des nœuds sont moins importantes dans le classement final. Un

Le système est un réseau Si le nerf artificiel consiste en une structure graphique (comme le montre la figure II.4) Utiliser un algorithme d'apprentissage pour modifier les poids de connexion.



**Figure II.4:** Architecture de réseau neuronal simple.

➤ **Fonctions d'activation**

- ✓ Fonction à seuil 'step'
- ✓ Fonction 'sigmoid'
- ✓ La tangente hyperbolique
- ✓ Fonction 'ReLU'
- ✓ Fonction 'Leaky ReLUs'
- ✓ Fonction Unités linéaires exponentielles 'Exponential Linear Units ELUs'
- ✓ Fonction 'Softmax'

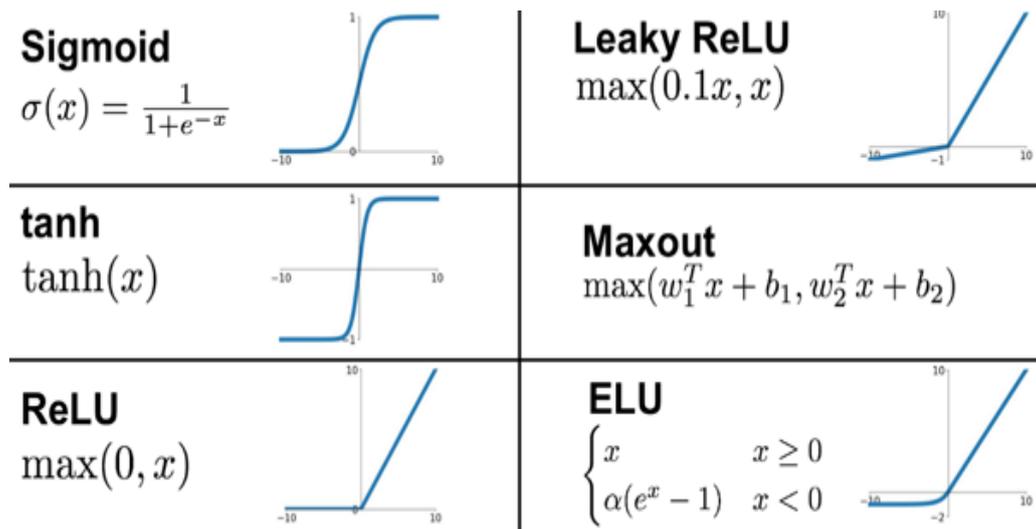


Figure II. 5 : Fonctions d'activation.

### II.3.3 Perceptron multicouches (PMC) [18]

Le perceptron multicouches (PMC) est un réseau composé de couches successives. Une couche est un ensemble de neurones n'ayant pas de connexion entre eux. Une couche d'entrée lit les signaux entrant, un neurone par entrée  $x_j$ , une couche en sortie fournit la réponse du système. Selon les auteurs, la couche d'entrée qui n'introduit aucune modification n'est pas comptabilisée. Une ou plusieurs couches cachées participent au transfert. Dans un perceptron, un neurone d'une couche cachée est connecté en entrée à chacun des neurones de la couche précédente et en sortie à chaque neurone de la couche suivante.

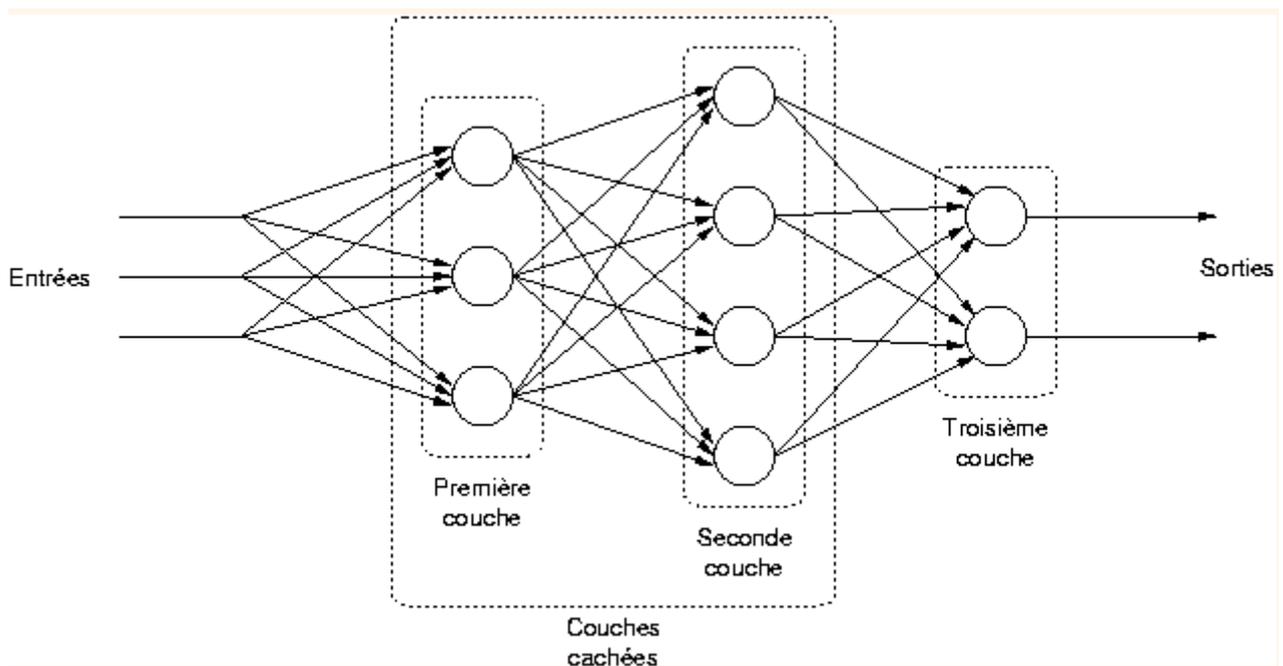
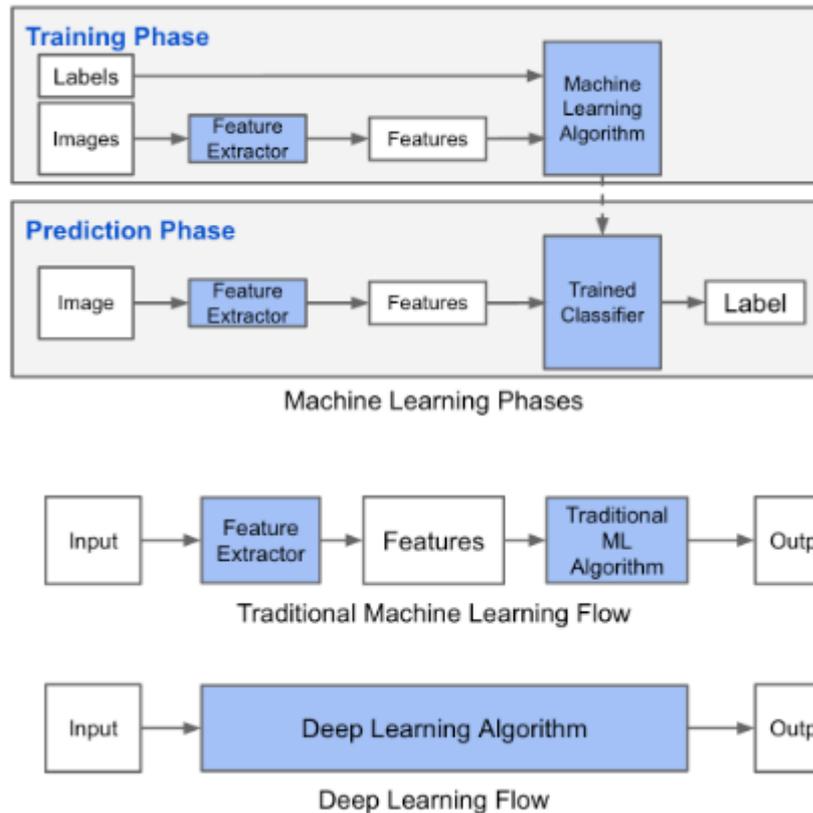


Figure II.6 : Un PMC. [researchgate.net]

## II.4. Apprentissage profond (Deep Learning)

Dans les techniques d'apprentissage classiques, le but est de trouver les caractéristiques représentatives qui contiennent l'information utile pour la classification, ainsi pour chaque problème étudié, plusieurs types de descripteurs peuvent être combinés pour trouver les bons éléments. Le problème est que généralement ces descripteurs sont incomplets, redondants et très complexes à trouver. La figure II.7 montre la différence entre le deep Learning et les méthodes conventionnelles du machine Learning.



**Figure II.7:** Machine learning vs Transfer learning

Par contre, le deep Learning (DL) est une branche d'apprentissage automatique. Il donne la capacité à la machine d'apprendre comment réaliser des tâches de classification directement à partir des données brutes (image, texte ou d'audio). Le principe du DL repose sur un apprentissage hiérarchique couche par couche, entre chaque couche interviennent des transformations non-linéaires et chaque couche reçoit en entrée la sortie de la couche précédente. L'un des atouts du deep Learning est qu'il remplace la phase d'extraction des caractéristiques qui se fait manuellement dans la technique classique par des algorithmes d'extraction de descripteurs hiérarchiques. Le deep Learning a été introduit dans les années 1980, mais ses capacités n'ont été exploitées que récemment à cause de deux points essentiels

: premièrement la nécessité de grandes bases de données labellisées pour l'apprentissage et deuxièmement la puissance de calcul considérable nécessaire

Le terme profond est lié au nombre de couches cachées, contrairement aux réseaux de neurones multicouches qui ne comporte que 2 à 3 couches, les réseaux profonds peuvent avoir jusqu'à une centaine de couches cachées. Les systèmes de classification à base du deep Learning ont montré qu'ils peuvent atteindre des taux de reconnaissance exceptionnelle en utilisant de très grandes bases de données pour l'apprentissage et des modèles de réseaux de neurones artificiels à plusieurs couches cachées [19].

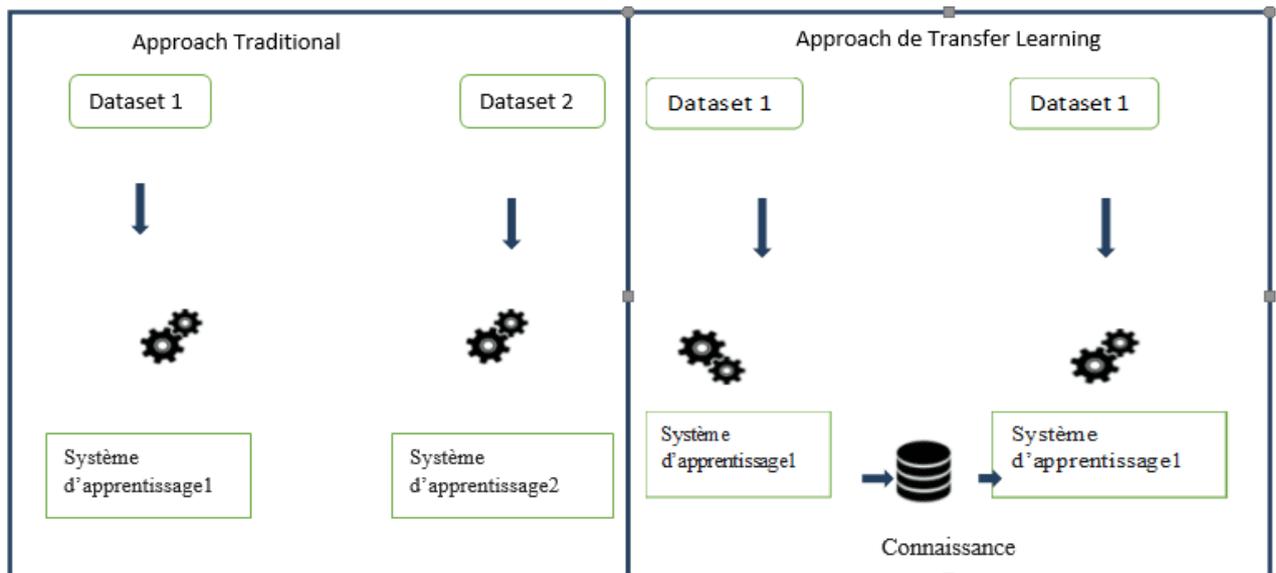
### II.4.1. Domaines d'application du Deep Learning

L'importance de l'apprentissage profond découle des résultats très performants obtenus dans divers secteurs, actuellement on trouve que le Deep Learning a été intégré dans les nouvelles voitures autonomes pour la détection des panneaux de signalisation, des piétons et des obstacles dangereux sur la route, dans l'aéronautique (drones), dans la téléphonie (capteurs photo intelligents), dans l'identification et la détection des objets à partir des images satellitaires de grandes dimensions ainsi que dans le domaine médical tel que la détection des cellules cancéreuses. Dans le domaine de la reconnaissance d'activités humaines, le deep Learning a permis aux chercheurs de franchir la barrière de la nécessité de conception des descripteurs en rendant cette étape automatisée et aussi d'atteindre des taux de classification inédits [19].

## II.5. Transfer Learning

### II.5.1. Définition

L'apprentissage par transfert (Transfer Learning en anglais) est l'un des champs de recherche de l'apprentissage automatique qui vise à transférer des connaissances d'une ou plusieurs tâches sources vers une ou plusieurs tâches cibles. Il peut être vu comme la capacité d'un système à reconnaître et appliquer des connaissances et des compétences, apprises à partir de tâches antérieures, sur de nouvelles tâches ou domaines partageant des similitudes [20].



**Figure II.8:** Transfert Learning

L'apprentissage par transfert est une méthode d'apprentissage automatique où un modèle développé pour une tâche est réutilisé comme point de départ pour un modèle sur une deuxième tâche [20].

## II.5. 2. Histoire du Transfer Learning

- En 1976, Stevo Bozinovski et Ante Fulgosi ont publié un article traitant explicitement de l'apprentissage par transfert dans la formation aux réseaux de neurones. L'article donne un modèle mathématique et géométrique de l'apprentissage par transfert. En 1981, un rapport a été présenté sur l'application de l'apprentissage par transfert dans la formation d'un réseau de neurones sur un ensemble de données d'images représentant des lettres de terminaux informatiques. L'apprentissage par transfert positif et négatif a été démontré expérimentalement.
- En 1993, Lorien Pratt a publié un article sur le transfert dans l'apprentissage automatique, formulant l'algorithme de transfert basé sur la discriminabilité (DBT).
- En 1997, Pratt et Sebastian Thrun ont édité un numéro spécial de Machine Learning consacré à l'apprentissage par transfert, et en 1998, le domaine avait progressé pour inclure l'apprentissage multi-tâches, ainsi qu'une analyse plus formelle de son fondations. Apprendre à apprendre, édité par Thrun et Pratt, est une revue de 1998 sur le sujet.

- L'apprentissage par transfert a également été appliqué dans les sciences cognitives , avec Pratt en tant qu'invité d'édition d'un numéro de Connection Science sur la réutilisation des réseaux de neurones par transfert en 1996
- Andrew Ng a déclaré dans son tutoriel NIPS 2016 que le TL sera le prochain moteur du succès commercial du ML après un apprentissage supervisé pour souligner l'importance du TL. [21].

### II.5.3. Comment utiliser l'apprentissage par transfert ?[22].

Vous pouvez utiliser l'apprentissage par transfert sur vos propres problèmes de modélisation prédictive.

Deux approches courantes sont les suivantes :

- Développer une approche modèle
- Approche de modèle préformé

#### a) Développement d'une approche modèle

1-Sélectionnez Tâche source. Vous devez sélectionner un problème de modélisation prédictive connexe avec une abondance de données où il existe une relation entre les données d'entrée, les données de sortie et/ou les concepts appris lors du mappage des données d'entrée aux données de sortie.

2- Développer le modèle source. Ensuite, vous devez développer un modèle habile pour cette première tâche. Le modèle doit être meilleur qu'un modèle naïf pour garantir qu'un certain apprentissage des fonctionnalités a été effectué.

3- Réutiliser le modèle. L'ajustement du modèle sur la tâche source peut ensuite être utilisé comme point de départ pour un modèle sur la deuxième tâche d'intérêt. Cela peut impliquer d'utiliser tout ou partie du modèle, selon la technique de modélisation utilisée.

4- Modèle d'accord. En option, le modèle peut devoir être adapté ou affiné sur les données de pair entrée-sortie disponibles pour la tâche d'intérêt.

#### b) Approche de modèle pré-formé

1- Sélectionnez Modèle source. Un modèle source pré-entraîné est choisi parmi les modèles disponibles. De nombreux instituts de recherche publient des modèles sur des ensembles de données volumineux et complexes qui peuvent être inclus dans le pool de modèles candidats parmi lesquels choisir.

- 2- Réutiliser le modèle. Le modèle préformé peut ensuite être utilisé comme point de départ pour un modèle sur la deuxième tâche d'intérêt. Cela peut impliquer d'utiliser tout ou partie du modèle, selon la technique de modélisation utilisée.
- 3- Modèle d'accord. En option, le modèle peut devoir être adapté ou affiné sur les données de pair entrée-sortie disponibles pour la tâche d'intérêt.

Ce deuxième type d'apprentissage par transfert est courant dans le domaine de l'apprentissage en profondeur.

#### **II.5.4. Exemples d'application du Transfer Learning [23].**

Les exemples les plus connus d'apprentissage par transfert sont dans les domaines du traitement du langage naturel (NLP) et de la vision par ordinateur (computer vision).

##### **a) Apprentissage par transfert en Computer Vision**

La vision par ordinateur a vraiment bénéficié des différents avantages de l'apprentissage par transfert. Et pour quoi ?

En fait, entraîner des modèles de computer vision requiert une puissance de calcul conséquente, ou même l'accès à certains composants matériels (Graphics Processing Unit (GPU), (Tensor Processing Unit (TPU))). En effet, il est assez fastidieux d'entraîner un modèle sur jeu de données composés de millions d'images.

Alors, il est courant de ré-utiliser un modèle de machine learning déjà entraîné (sur une grande quantité d'images/vidéos).

Ces modèles peuvent être téléchargés et incorporés directement dans la création de nouveaux modèles qui utilisent des images en entrée.

Voici trois exemples de modèles de ce type :

- Le célèbre modèle d'Oxford le VGG
- Le modèle de Google Inception
- Le ResNet de Microsoft

##### **b) Apprentissage par transfert en traitement du langage naturel (NLP)**

Le Transfer Learning est aussi de plus en plus présent en NLP (traitement du langage naturel : Natural Language Processing (NLP)).

Il existe plusieurs approches, la première est l'utilisation d'Embeddings. En effet, cette technique permet de représenter chaque mot d'un dictionnaire par un vecteur de nombres réels. L'idée sera alors d'entraîner un modèle sur un vaste corpus de textes afin d'avoir une représentation vectorielle de la sémantique réelle des mots. On pourra alors utiliser ces modèles pour de l'analyse de sentiments, la classification de texte... Comme exemple populaire nous avons Word2vec et FastText.

En outre, une nouvelle approche est née avec des modèles plus récents tels que les Transformers, ELMO, BERT, ...

Ces nouvelles approches présentent certainement une myriade de possibilités (Question Answering, Détection de contenu toxique et de fakeNews).

### II.5.5. Type de Transfer Learning [24].

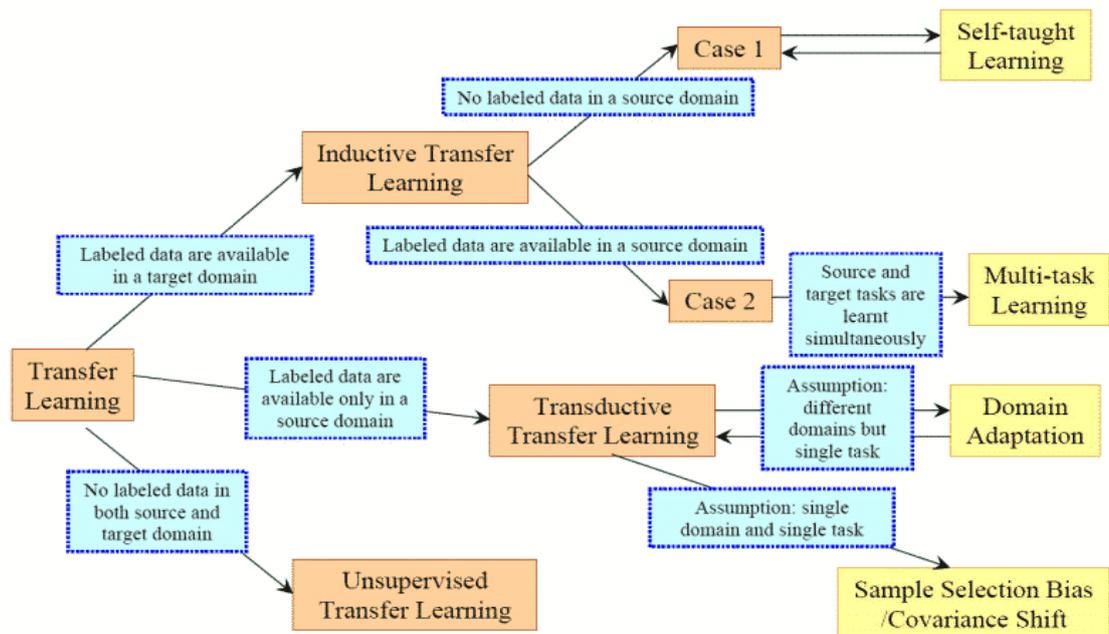


Figure II.9 : Type de transfert Learning. [meritis.fr].

#### a) Apprentissage par transfert inductif

Dans cette configuration, les domaines source et cible sont les mêmes (même données), mais les tâches source et cible sont différentes mais proches. L'idée consiste alors à utiliser les modèles existants pour réduire de manière avantageuse le champ d'application des modèles possibles (biais de modèle) comme illustré dans la figure ci-dessous.

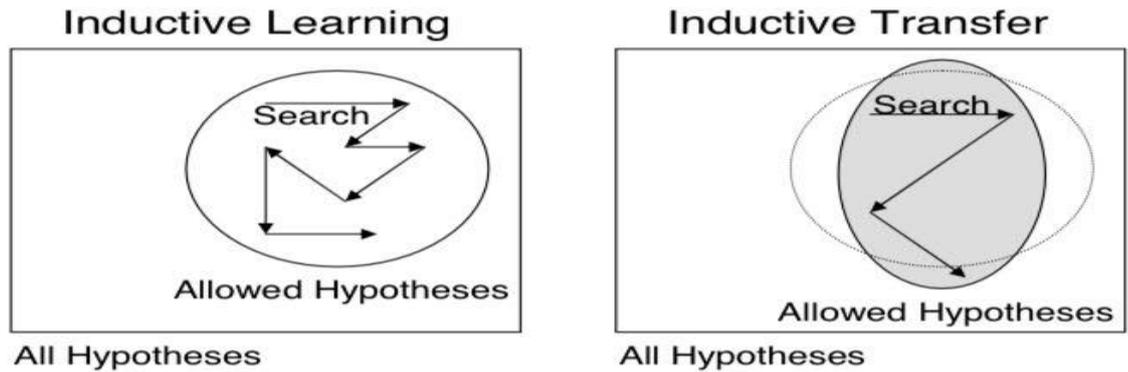


Figure II.10: Apprentissage inductif et transfert inductif [24].

### b) Apprentissage par transfert non supervisé

Comme dans le cas de l'apprentissage par transfert inductif, les domaines source et cible sont similaires, mais les tâches sont différentes. Toutefois, les données des deux domaines ne sont pas labellisées.

Il est souvent plus facile d'obtenir de grandes quantités de données non labellisées, à partir de bases de données et de sources sur le web par exemple, que des données labellisées. C'est pourquoi l'idée d'utiliser l'apprentissage non supervisé en combinaison avec le Transfer Learning suscite un grand intérêt.

À titre d'exemple, le Self-taught clustering est une approche qui permet de réaliser le clustering de petites collections de données cibles non labellisées, avec l'aide d'une grande quantité de données sources non labellisées. Cette approche s'avère plus performante que les approches de pointe traditionnellement utilisées, lorsque les données cibles sont labellisées de manière non pertinente.

### c) Apprentissage par transfert transductif

Dans cette configuration, les tâches sources et cibles sont similaires, mais les domaines correspondants sont différents soit en termes de données ou de distributions de probabilités marginales.

Par exemple, les modèles de NLP, tels que ceux utilisés pour l'étiquetage morpho-syntaxique de mots, Part-Of-Speech Tagger (POS Tagger) en anglais, sont généralement entraînés et testés sur des données d'actualité comme celles du Wall Street Journal. Ils peuvent être adaptés aux données issues des réseaux sociaux dont le contenu est différent mais proche de celles des journaux.

## II.6. Transformateur de vision (ViT)

### II.6.1. Définition [25].

Le transformateur de vision (ViT) est un transformateur utilisé dans le domaine de la vision par ordinateur qui fonctionne sur la base de la nature fonctionnelle des transformateurs utilisés dans le domaine du traitement du langage naturel.

Un transformateur destiné aux tâches de traitement de la vision telles que la reconnaissance d'images.

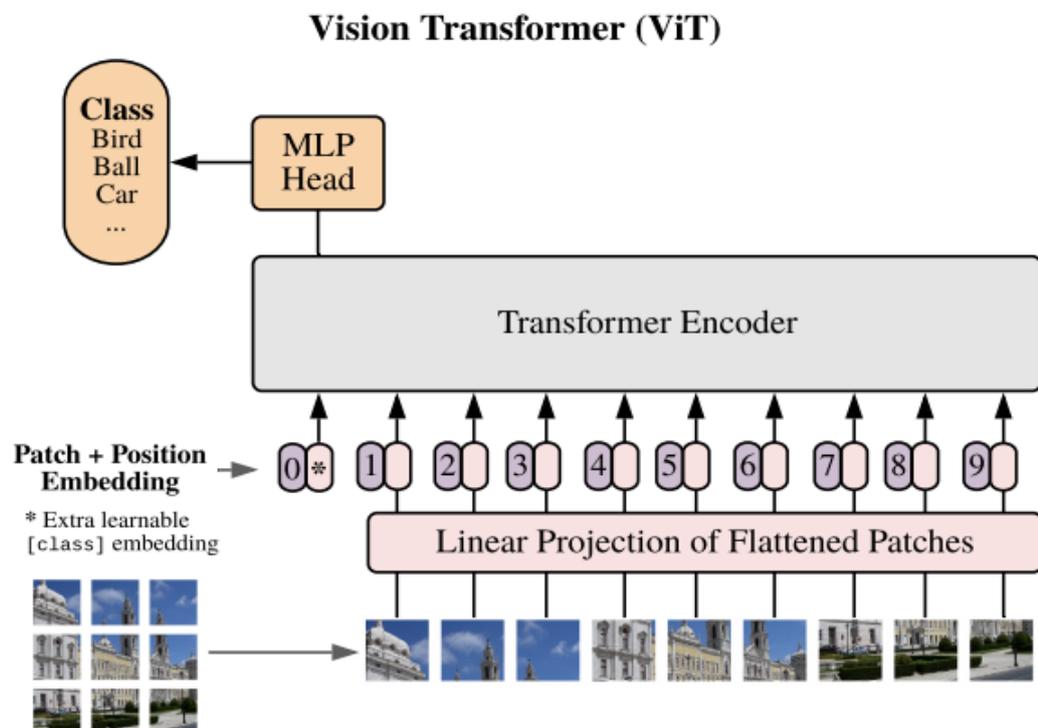
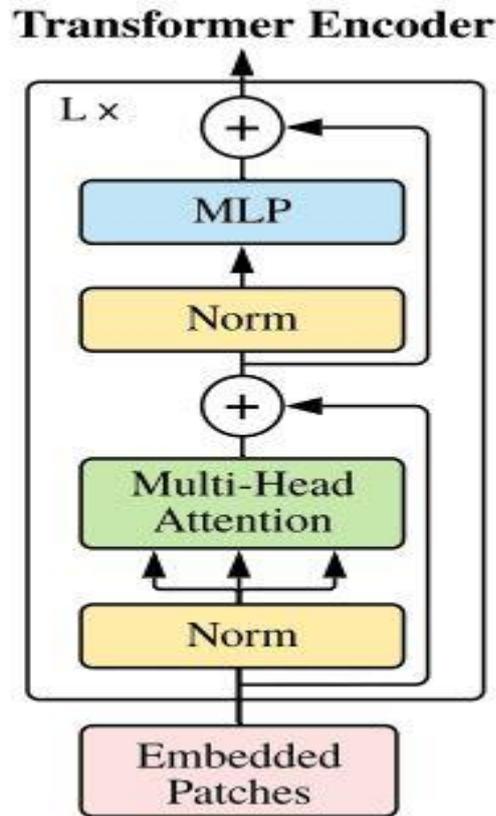


Figure II.11: Aperçu du modèle Vit. [28].



**Figure II.12** : Architecture de codeur transformateur [28].

## II.6.2. Histoire de Transformateur de Vision

- En 2017 les transformateurs initialement introduits dans le célèbre article "Attention is All You Need", le Transformer s'est largement répandu dans le domaine du traitement du langage naturel et il est rapidement devenu l'une des architectures les plus utilisées et les plus prometteuses dans le domaine. [26].

- En 2020, les transformateurs de vision ont ensuite été utilisés pour des tâches de vision par ordinateur, l'article "Une image vaut 16x16 mots". L'idée est essentiellement de décomposer l'image d'entrée en une série de patches qui, une fois convertis en vecteurs, sont traités comme des mots dans un convertisseur normal. La même année, d'importantes variantes de transducteurs visuels ont été proposées. [26].

-En 2021 Le modèle Vision Transformer (ViT) a été introduit dans un article de recherche publié en tant qu'article de conférence à l'ICLR intitulé « Une image vaut 16\*16 mots : transformateurs pour la reconnaissance d'images à l'échelle ». Il a été développé et publié par Neil Houlsby, Alexey Dosovitskiy et 10 autres auteurs de l'équipe Google Research Brain. [27].

- En 2022, le Vision Transformer (ViT) est apparu comme une alternative compétitive aux réseaux de neurones convolutifs (CNN) qui sont actuellement à la pointe de la technologie en vision par ordinateur et donc largement utilisés dans différentes tâches de reconnaissance d'images. Les modèles ViT surpassent l'état de l'art actuel (CNN) de près de x4 en termes d'efficacité et de précision de calcul. [27].

## II.6. 2.Fonctionnement du Transformateur de Vision [28].

Voici comment fonctionne la vision transformer étape par étape

- Découpage d'une image en patches
- Aplatissement des patches
- Production d'incorporations linéaires de dimension inférieure à partir des patches aplatis
- Ajouter des intégrations positionnelles
- Introduire la séquence comme entrée d'un codeur transformateur standard
- Pré-entraînement du modèle avec des étiquettes d'image (entièrement supervisé sur un énorme ensemble de données)
- Ajustement fin sur l'ensemble de données en aval pour la classification d'images.

La seule chose qui change est le nombre de ces blocs. Pour prouver qu'avec plus de données, il est possible d'entraîner de plus grandes variantes de ViT, 3 modèles ont été proposés :

Model	Layers	Hidden size $D$	MLP size	Heads	Params
ViT-Base	12	768	3072	12	86M
ViT-Large	24	1024	4096	16	307M
ViT-Huge	32	1280	5120	16	632M

**Figure II.13** : Variantes du modèle Vision Transformer [28].

- **Heads** fait référence à l'attention multi-têtes.
- **MLP size** fait référence au module bleu de la figure, **MLP (multi-layer perceptron)** signifie perceptron multicouche, mais il s'agit en fait d'un ensemble de couches de transformation linéaire.
- **Hidden size  $D$**  est la taille d'intégration, qui est maintenue fixe tout au long des couches. Pour que nous puissions utiliser des connexions de saut résiduelles courtes

la taille d'intégration est gardée fixe. il n'y a pas de décodeur Juste une couche linéaire supplémentaire pour la classification finale appelée tête MLP.

➤

➤ **Représentation d'une image comme une séquence de patches**

Pour une image d'entrée  $(\mathbf{x}) \in \mathbb{R}^{H \times W \times C}$  et la taille du patch  $p$ , nous voulons créer  $N$  patches d'image désignés par les termes suivants :  $(x)_p \in \mathbb{R}^{N \times (p^2 \times C)}$  où  $N = \frac{HW}{p^2}$ .  $N$  est la longueur de la séquence, similaire aux mots d'une phrase.

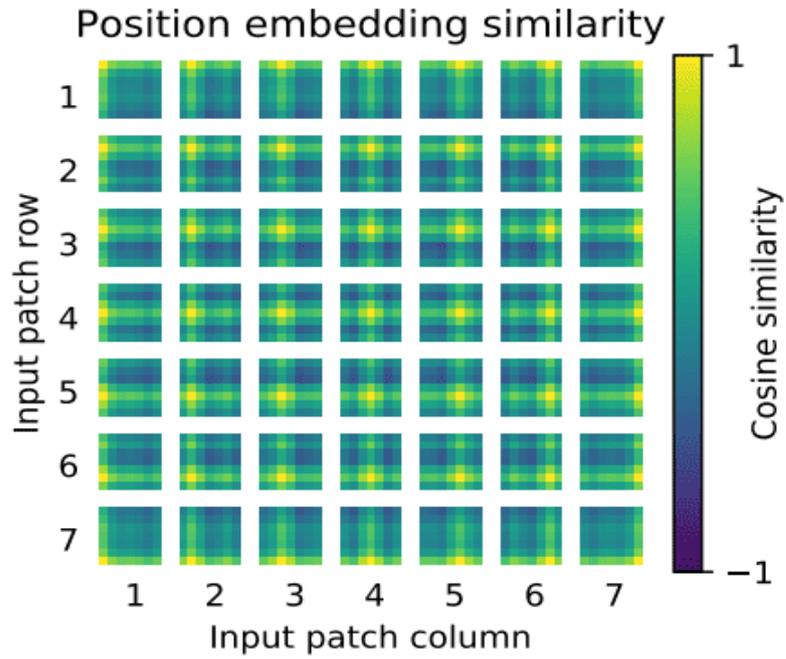
Pour passer du patch à l'intégration c'est juste une couche de transformation linéaire qui prend une séquence de  $P^2 C$  et les sorties  $D$ .

➤ **Intégration d'images**

L'incorporation d'images lit les images et les télécharge ou les évalue sur un serveur distant ou localement. Un vecteur de caractéristiques est attribué à chaque image à l'aide d'algorithmes d'apprentissage profond. Elle renvoie un tableau de données qui a été enrichi de colonnes supplémentaires (descripteurs d'images). L'incorporation d'images comprend une variété d'incorporeurs, chacun ayant été entraîné pour une tâche spécifique. Les images sont soit transmises à un serveur, soit évaluées localement sur l'ordinateur de l'utilisateur, et des représentations vectorielles sont alors créées.

➤ **Incorporations positionnelles**

Bien que de nombreux schémas d'incorporation positionnelle aient été appliqués, aucune différence significative n'a été constatée. Cela est probablement dû au fait que l'encodeur transformateur fonctionne au niveau du patch. L'apprentissage d'enchâssements qui capturent les relations d'ordre entre les patches (information spatiale) n'est pas si crucial. Il est relativement plus facile de comprendre les relations entre les patches de  $P \times P$  que d'une image complète Hauteur x Largeur Par conséquent, après la projection linéaire de faible dimension, une intégration de position entraînable est ajoutée aux représentations de patch. Il est intéressant de voir à quoi ressemblent ces incorporations de position après l'entraînement :



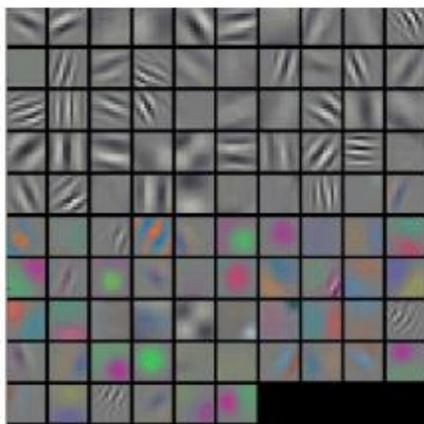
**Figure II.14 :** Image 16x16 : transformateurs pour la reconnaissance d'images [28].

**Résultats clés**

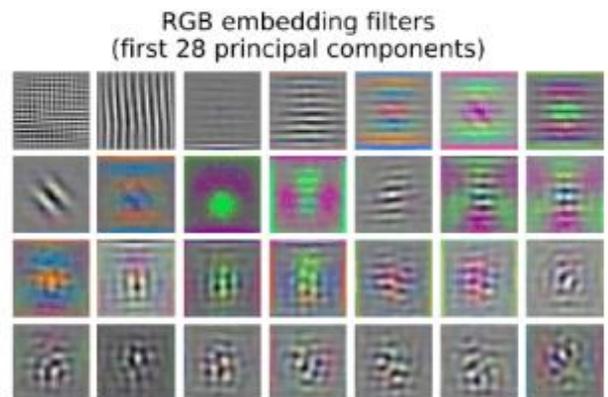
Au début, nous avons l'habitude de visualiser les premières couches.

Parce que nous pensons que les réseaux bien entraînés présentent souvent des filtres agréables et lisses.

**Alexnet 1st conv filters**



**ViT 1st linear embedding filters**



**Figure II.15 :** gauche : visualisation des filters Alexnet. À droite: Filtres appris par ViT. [28].

Remarquez que les poids de la première couche sont très beaux et lisses, ce qui indique un réseau bien convergé. Les caractéristiques de couleur/échelle de gris sont regroupées parce que l'AlexNet contient deux flux de traitement séparés, et une conséquence apparente de cette architecture est qu'un flux développe des caractéristiques d'échelle de gris à haute fréquence et l'autre des caractéristiques de couleur à basse fréquence. Pour de telles visualisations, l'ACP est utilisé. De cette façon, les représentations des premières couches peuvent partager des caractéristiques similaires.

➤ **Interaction de l'apprentissage non-local**

Pour une taille de patch  $P$ , le maximum  $P \times P$ , qui est dans notre cas est de 128, même à partir de la 1ère couche, on n'a plus besoin de couches successives de convolutions pour arriver à des pixels éloignés de 128. Avec des convolutions sans dilatation, le champ réceptif est augmenté linéairement. En utilisant l'auto-attention, nous avons une interaction entre les représentations des pixels dans la 1ère couche et les paires de représentations dans la 2ème couche et ainsi de suite.



Right: Image generated using *Fomoro AI calculator* Left: Image by *Alexey Dosovitskiy et al 2020*

**Figure II.16** : Droite : Image générée à l'aide de la calculatrice Fomoro A Gauche : Image par Alexey Dosovitskiy et al 2020 [28].

En se basant sur le diagramme de gauche de ViT, on peut argumenter que :

-Il y a effectivement des têtes qui s'occupent de l'ensemble du patch dès les premières couches.

-On peut justifier le gain de performance par l'accès précoce aux interactions entre pixels. Il semble plus important que les premières couches aient accès à l'ensemble du patch

(information globale). En d'autres termes, les têtes qui appartiennent à la partie supérieure gauche de l'image peuvent être la raison principale de la performance supérieure.

-Il est intéressant de noter que la distance d'attention augmente avec la profondeur du réseau, comme pour le champ réceptif des opérations locales.

-Il existe également des têtes avec des distances d'attention constamment faibles dans les couches inférieures. À droite, un réseau de 24 couches avec des convolutions standard 3x3 a un champ réceptif de moins de 50. Nous aurions besoin d'environ 50 couches de convolutions pour assister à un champ réceptif d'environ 100, sans dilatation ni regroupement de couches.

Pour renforcer cette idée de têtes d'attention hautement localisées, les auteurs ont expérimenté des modèles hybrides qui appliquent un ResNet avant le Transformer. Ils ont trouvé des têtes moins

fortement localisées, comme prévu. En plus de la visualisation du filtre, cela suggère qu'il pourrait avoir une fonction similaire à celle des premières couches convolutives dans les RNC.

#### ➤ **Distance d'attention et visualisation**

La distance d'attention est calculée comme la distance moyenne entre le pixel de la requête et le reste du patch, multipliée par le poids de l'attention. C'est analogue au champ réceptif, mais pas exactement pareil.

### **II.6.3 Domaines d'application [27]**

Les transformateurs de vision ont de nombreuses applications dans les tâches de reconnaissance d'images courantes telles que la détection d'objets, la segmentation, la classification d'images et la reconnaissance d'actions. De plus, les ViT sont appliquées dans la modélisation générative et les tâches multi-modèles, y compris l'ancrage visuel, la réponse visuelle aux questions et le raisonnement visuel.

La prévision vidéo et la reconnaissance d'activité sont toutes des parties du traitement vidéo qui nécessitent ViT. De plus, l'amélioration d'image, la colorisation et la super-résolution d'image utilisent également des modèles ViT. Enfin, ViTs a de nombreuses applications dans l'analyse 3D, telles que la segmentation et la classification des nuages de points.

## **II.7. Conclusion**

Dans ce second chapitre, nous sommes intéressés au domaine de l'apprentissage machine, et plus spécialement de l'apprentissage par transfert, tout en faisant le tour de ses différents points sous-jacent. En effet, nous avons vu le transformateur de vision (Vision Transformer (ViT)) et son fonctionnement

Le prochain chapitre, traite les détails de la conception, ainsi que la méthode et les outils utilisés pour la réalisation de notre application.

*Chapitre III*

*Implémentations et*

*résultats*

### III.1 Introduction

Dans ce chapitre, nous allons concevoir notre application ; La classification des images de l'os de la main pour l'estimation et la reconnaissance de l'âge de chaque image. Nous présenterons également notre environnement de travail en utilisant le langage Python. D'abord une introduction au langage de programmation choisi, puis la base de données utilisée et enfin la présentation des résultats notre application.

### III.2 Environnement de travail

#### ➤ Matériel et logiciel

**Note** : Nous avons utilisé la plateforme kaggle pour exécuter notre programme car il nécessite l'utilisation de GPU et une grande mémoire pour effectuer les calculs en raison de l'utilisation.

**-Kaggle** est une plateforme web qui accueille la plus grande communauté de Data Science au monde, avec plus de 536 000 membres actifs dans 194 pays et reçoit près de 150 000 soumissions par mois, et qui lui fournit des outils et des ressources puissants pour aider à atteindre tous les progrès de science des données. Kaggle, de la même manière que Datascientest, offre un environnement Jupyter Notebooks personnalisable et sans configuration. Sont accessibles gratuitement des GPU et une grande quantité de données et de codes publiés par la communauté. À l'intérieur de Kaggle, vous trouverez tout le code et les données dont vous avez besoin pour réaliser vos projets de science des données. Il y a plus de 50 000 jeux de données publics et 400 000 notebooks publics disponibles pour tous.

#### ➤ Language de programmation « Python »

Python est un langage de programmation de haut niveau polyvalent et largement utilisé. Il a été créé par "Guido van Rossum" en 1991 et développé par la "Python Software Foundation". Il a été conçu en mettant l'accent sur la lisibilité du code, sa syntaxe permet aux programmeurs d'exprimer leurs concepts avec moins de lignes de code. Python est un langage de programmation qui vous permet de travailler rapidement et d'intégrer le système est plus efficace.

#### ➤ Bibliothèques

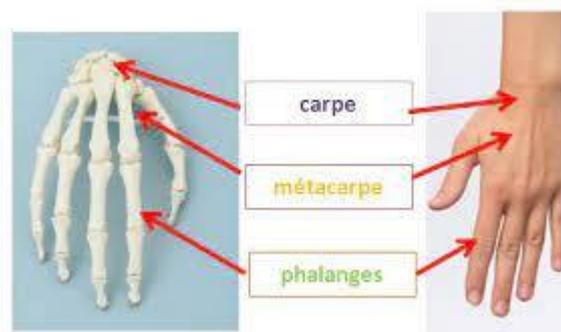
Nous avons installé les différentes bibliothèques nécessaires pour notre projet à l'aide de la commande cmd, les bibliothèques installées sont les suivantes :

- **TensorFlow** est une plateforme open source de bout en bout pour l'apprentissage automatique. Elle dispose d'un écosystème complet et flexible d'outils, de bibliothèques et de ressources communautaires qui permet aux chercheurs de faire progresser l'état

de l'art de l'apprentissage automatique et aux développeurs de créer et de déployer facilement des applications basées sur l'apprentissage automatique.

- **PyTorch(torch)** est un package Python qui fournit deux fonctionnalités de haut niveau comme le calcul tensoriel (comme NumPy) avec une forte accélération GPU et aussi des réseaux neuronaux profonds construits sur un système autograd basé sur des bandes magnétiques.
- **Linformer**. Il s'agit d'une attention dont la complexité n'est que linéaire, ce qui permet de traiter de très longues séquences (1mil+) sur du matériel moderne.
- **Scikit-learn (Sklearn)** est la bibliothèque la plus utile et la plus robuste pour l'apprentissage automatique en Python. Elle fournit une sélection d'outils efficaces pour l'apprentissage automatique et la modélisation statistique, notamment la classification, la régression, le regroupement et la réduction de la dimensionnalité, via une interface cohérente en Python.
- **Vit\_pytorch** : Le transformateur de vision est un modèle de classification d'images qui utilise une architecture de type transformateur sur des patches de l'image.
- **Matplotlib** est une bibliothèque de traçage pour le langage de programmation Python et son extension de mathématiques numériques NumPy. Elle fournit une API orientée objet permettant d'intégrer des tracés dans des applications utilisant des boîtes à outils graphiques à usage général.
- **System libraries** : built-in sont d'autres bibliothèques qui permettent de manipuler des fichiers et d'exécuter des commandes système comme : os, glob, itertools, random et zipfile.

### III.3 Squelette de la main [30]



**Figure III.1** : Subdivision du squelette de la main vue dorsale squelette de la main vue antérieur.

- La main est une structure anatomique merveilleuse mais complexe

- Organe de fonction et d'information qui donne au membre supérieur son importance et sa singularité

- segment terminal du membre thoracique, accommodée pour la préhension, elle a un rôle sensitif.

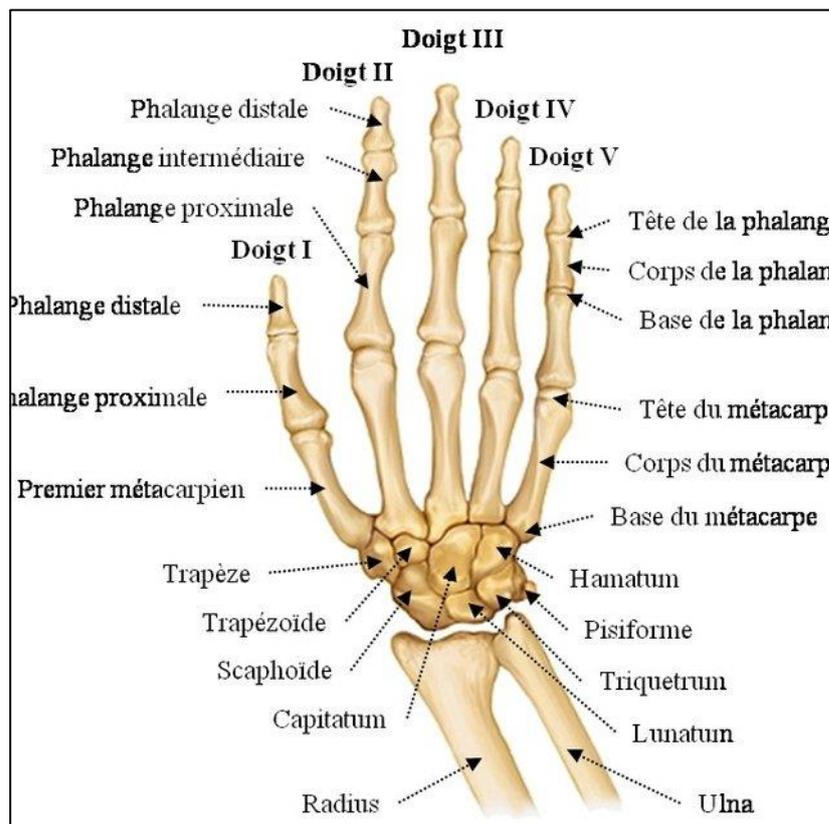
- Elle est paire et asymétrique.

DEFINITION : • Complexe Osseux fait de 27os constant réparti en trois groupes :

- ✓ Carpe
- ✓ Métacarpe
- ✓ Phalanges

- Elle peut posséder également de nombreux osselets inconstants (les plus fréquents étant les sésamoïdes du pouce).

- Ce squelette représente le support architectural des parties molles dont la longueur, la forme et l'orientation ont un rôle dans la fonction de la main



**Figure III.2 :** Squelette de la main vue antérieure [30].

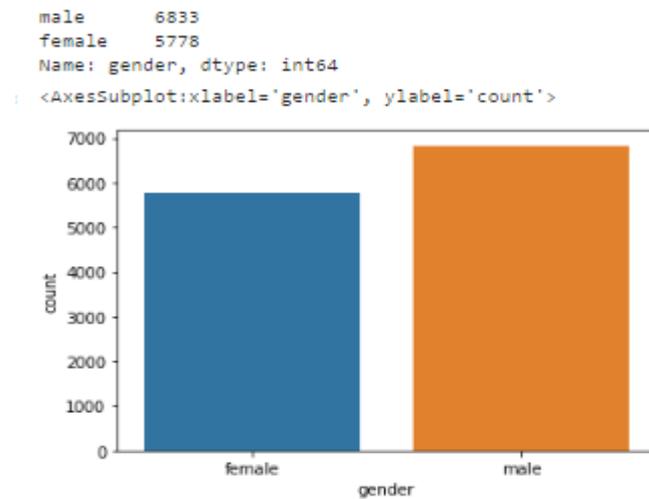
### Base des données utilisée

La base de données dataset contient 12611 images dans le dossier bone+age+training+set et 803 images dans le dossier bone+age+validation+set.

L'ensemble des données contient des images des os de la main pour différents âges mâle et femelle.

Dataset	Nombres d'images	Tailles des dossiers
Train	12611	9,12 Go
Validation	803	1,62 Go

La figure ci-dessous montre la répartition par sexe dans la base des données



**Figure III.3** : Répartition du nombre des individus male/femelle dans la base de données.

La figure suivante représente un exemple des images de l'os de la main extrait de la base de données selon l'Age et le genre (mâle ou femelle).



9 mois Male



1 Ans et 2 mois Male



2 Ans femelle



7 Ans et 8mois



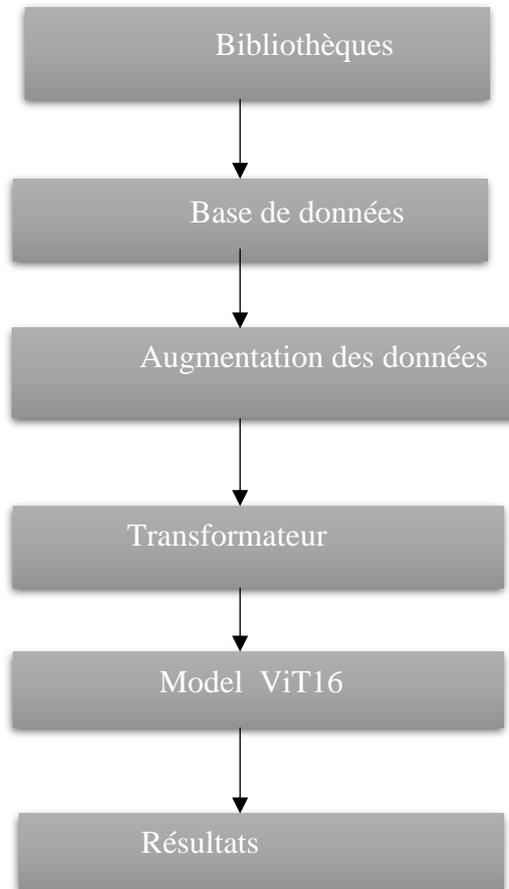
10 Ans Male



12 Ans Femelle

**Figure III.4 :** Extrais du nombre des individus male/femelle dans la base de données.

### III.4 Structure d'Évaluation de l'âge osseux à l'aide de modèles de vision transformer



Cette partie de programme affiche les bibliothèques utilisées dans le programme :

```
from __future__ import print_function
import glob
from itertools import chain
import os
import random
import zipfile
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import torch
import torch.nn as nn
import torch.nn.functional as F
import torch.optim as optim
from linformer import Linformer
from PIL import Image
from sklearn.model_selection import train_test_split
from torch.optim.lr_scheduler import StepLR
from torch.utils.data import DataLoader, Dataset
from torchvision import datasets, transforms
from tqdm.notebook import tqdm
from vit_pytorch.efficient import ViT
```

**-Import Glob :** Le module glob est une partie utile de la bibliothèque standard Python. glob (abréviation de global) est utilisé pour renvoyer tous les chemins de fichiers qui correspondent à un modèle spécifique.

Nous pouvons utiliser glob pour rechercher un modèle de fichier spécifique, ou peut-être plus utilement, rechercher des fichiers dont le nom de fichier correspond à un certain modèle en utilisant des caractères génériques.

-Et ici nous définissons les paramètres de 'training'

```
# Training settings
batch_size = 64
epochs = 20
lr = 3e-5
gamma = 0.7
seed = 42
def seed_everything(seed):
    random.seed(seed)
    os.environ['PYTHONHASHSEED'] = str(seed)
    np.random.seed(seed)
    torch.manual_seed(seed)
    torch.cuda.manual_seed(seed)
    torch.cuda.manual_seed_all(seed)
    torch.backends.cudnn.deterministic = True
seed_everything(seed)
device = 'cuda'
#Load Data
os.makedirs('data', exist_ok=True)
train_dir = '../train'
test_dir = '../test'
train_list = glob.glob(os.path.join(train_dir, '*.jpg'))
test_list = glob.glob(os.path.join(test_dir, '*.jpg'))
labels = [path.split('/')[-1].split('.')[0] for path in train_list]
```

- **Batch size :** représente la taille du groupe définit le nombre d'échantillons qui seront propages dans le réseau elle représente le nombre d'exemples d'entraînement dans une passe avant/arrière. Plus la taille du lot est élevée, plus vous aurez besoin d'espace mémoire.
- **Époques :** une époque est un terme utilise dans l'apprentissage automatique et indique le nombre de passage de l'ensemble de données d'apprentissage que l'algorithme de l'apprentissage automatique a effectué.
- **Seed :** ce terme désigne l'état d'utilisation un générateur de nombre de nombre pseudo aléatoires

- **Random Seed:** La fonction `random.seed` initialise le générateur de nombres aléatoires avec la valeur donnée.
- **Torch :** Le package Torch contient des structures de données pour des tenseurs multidimensionnels et définit des opérations mathématiques sur ces tenseurs. De plus, il fournit de nombreux utilitaires pour une sérialisation efficace des Tensors et des types arbitraires, ainsi que d'autres utilitaires utiles.

Ici nous définissons le device

-**Data split :** c'est une fonction qui divisera une donnée à trainig data et test data

```
#loading dataframes
train_df = pd.read_csv('/kaggle/input/rsna-bone-age/boneage-training-dataset.csv')
test_df = pd.read_csv('/kaggle/input/rsna-bone-age/boneage-test-dataset.csv')

#appending file extension to id column for both training and testing dataframes
train_df['id'] = train_df['id'].apply(lambda x: str(x)+'.png')
test_df['Case ID'] = test_df['Case ID'].apply(lambda x: str(x)+'.png')

train_df.head()
```

### Augmentations des données

L'augmentation est utile pour améliorer les performances et les résultats des modèles d'apprentissage automatique en formant des exemples nouveaux et différents pour former des ensembles de données. Si l'ensemble de données d'un modèle d'apprentissage automatique est riche et suffisant, le modèle fonctionne mieux et plus précisément.

```

#Image Augumentation
train_transforms = transforms.Compose(
    [
        transforms.Resize((224, 224)),
        transforms.RandomResizedCrop(224),
        transforms.RandomHorizontalFlip(),
        transforms.ToTensor(),
    ]
)
val_transforms = transforms.Compose(
    [
        transforms.Resize(256),
        transforms.CenterCrop(224),
        transforms.ToTensor(),
    ]
)
test_transforms = transforms.Compose(
    [
        transforms.Resize(256),
        transforms.CenterCrop(224),
        transforms.ToTensor(),
    ]
)

```

-est un processus de création de nouvelles formations à partir de celles existantes

```

class Bone_age(Dataset):
    def __init__(self, file_list, transform=None):
        self.file_list = file_list
        self.transform = transform

    def __len__(self):
        self.filelength = len(self.file_list)
        return self.filelength

    def __getitem__(self, idx):
        img_path = self.file_list[idx]
        img = Image.open(img_path)
        img_transformed = self.transform(img)

        label = img_path.split("/")[-1].split(".")[0]
        label = 1 if label == "dog" else 0

        return img_transformed, label

train_data = Bone_age(train_list, transform=train_transforms)
valid_data = Bone_age(valid_list, transform=test_transforms)
test_data = Bone_age(test_list, transform=test_transforms)
train_loader = DataLoader(dataset = train_data, batch_size=batch_size, shuffle=True )
valid_loader = DataLoader(dataset = valid_data, batch_size=batch_size, shuffle=True)
test_loader = DataLoader(dataset = test_data, batch_size=batch_size, shuffle=True)

```

Appelez la classe qui est déjà créée afin d'appliquer le traitement et la transformation de l'image, c'est-à-dire le format et les dimensions corrects pour être prêt pour la phase d'apprentissage.

Comme nous le savons, les données sont l'objet principal de l'apprentissage automatique, ce qui nous permet d'obtenir le modèle approprié avec un minimum d'erreurs.

-les données améliorent la qualité des données et protègent les applications des erreurs potentielles telles que les valeurs nulles.

### Training

```
#Training
# loss function
criterion = nn.mae()
# optimizer
optimizer = optim.Adam(model.parameters(), lr=lr)
# scheduler
scheduler = StepLR(optimizer, step_size=1, gamma=gamma)
```

```
for epoch in range(epochs):
    epoch_loss = 0
    epoch_accuracy = 0
    .
    for data, label in tqdm(train_loader):
        data = data.to(device)
        label = label.to(device)
        .
        output = model(data)
        loss = criterion(output, label)
        .
        optimizer.zero_grad()
        loss.backward()
        optimizer.step()
        .
        acc = (output.argmax(dim=1) == label).float().mean()
        epoch_accuracy += acc / len(train_loader)
        epoch_loss += loss / len(train_loader)
    .
```

```
with torch.no_grad():
    epoch_val_accuracy = 0
    epoch_val_loss = 0
    for data, label in valid_loader:
        data = data.to(device)
        label = label.to(device)

        val_output = model(data)
        val_loss = criterion(val_output, label)

        acc = (val_output.argmax(dim=1) == label).float().mean()
        epoch_val_accuracy += acc / len(valid_loader)
        epoch_val_loss += val_loss / len(valid_loader)

    print(
        f'Epoch : {epoch+1} - loss : {epoch_loss:.4f} - acc: {epoch_accuracy:.4f} - val_loss : {epoch_val_loss:.4f} - val_acc: {epoch_val_accuracy}
    )
```

```

model_final.load_weights('Inception_model.h5')
pred = mean_bone_age + std_bone_age*(model_final.predict(test_X, batch_size = 32, verbose = True))
test_months = mean_bone_age + std_bone_age*(test_Y)

ord_ind = np.argsort(test_Y)
ord_ind = ord_ind[np.linspace(0, len(ord_ind)-1, 8).astype(int)] # take 8 evenly spaced ones
fig, axs = plt.subplots(4, 2, figsize = (15, 30))
for (ind, ax) in zip(ord_ind, axs.flatten()):
    ax.imshow(test_X[ind, :, :0], cmap = 'bone')
    ax.set_title('Age: %fY\nPredicted Age: %fY' % (test_months[ind]/12.0,
                                                pred[ind]/12.0))
    ax.axis('off')
fig.savefig('trained_image_predictions.png', dpi = 300)

```

### III.5 Résultats

Nous avons choisi de tester un modèle de vision transformer qui est le ViT 16, et deux autres du CNN qui sont le ResNet50 et le VGG19 pour appliquer à l'évaluation dès l'âge de os de la main afin de comparer et choisir le plus optimal de notre problématique. Pour montrer les résultats obtenus pour ces architectures, nous illustrons dans ce qui suit les résultats en termes de précision et d'erreur pour chacune des architectures.

- **ViT 16**

**ViT-L/16** désigne la variante "Large" avec une taille de patch d'entrée de  $16 \times 16$ . Notez que la longueur de la séquence du transformateur est inversement proportionnelle au carré de la taille du patch, et les modèles avec une taille de patch plus petite sont plus coûteux en calcul.

La figure représente le modèle ViT 16 sur python

```

#Vision Transformer model 16
model 16 = ViT(
    dim=128,
    image_size=265,
    patch_size=16,
    num_classes=1,
    transformer=efficient_transformer,
    channels=3,
).to(device)

```

-Les figures suivantes représentent la perte en mois durant des époques différentes

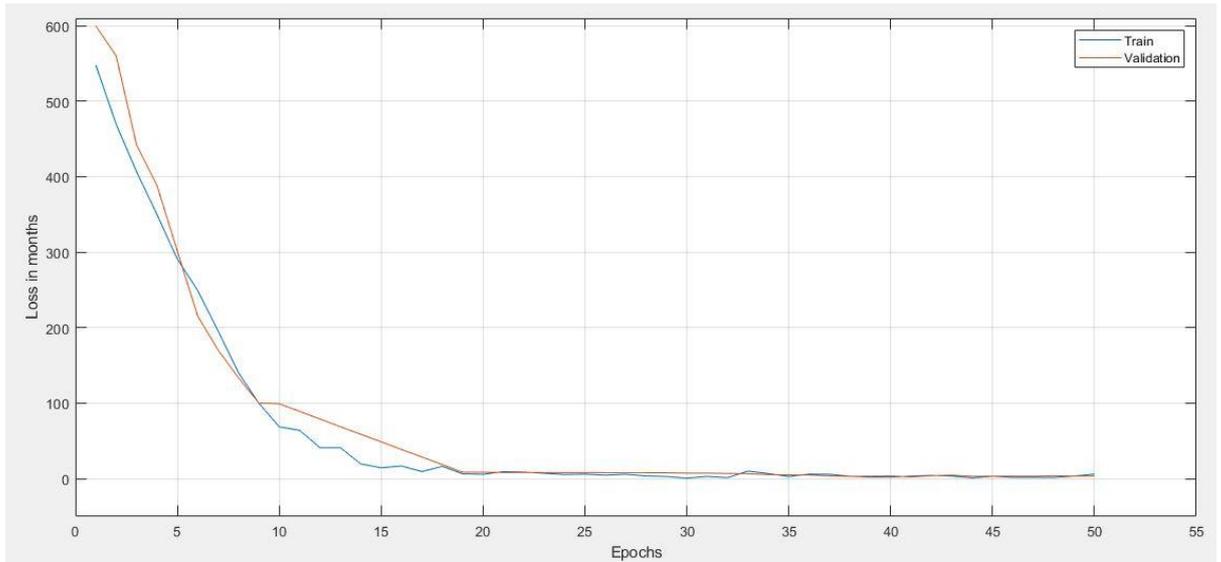


Figure III.5 : Représentation de la perte en mois.

-cette figure represente la partie zoom du figure III.5

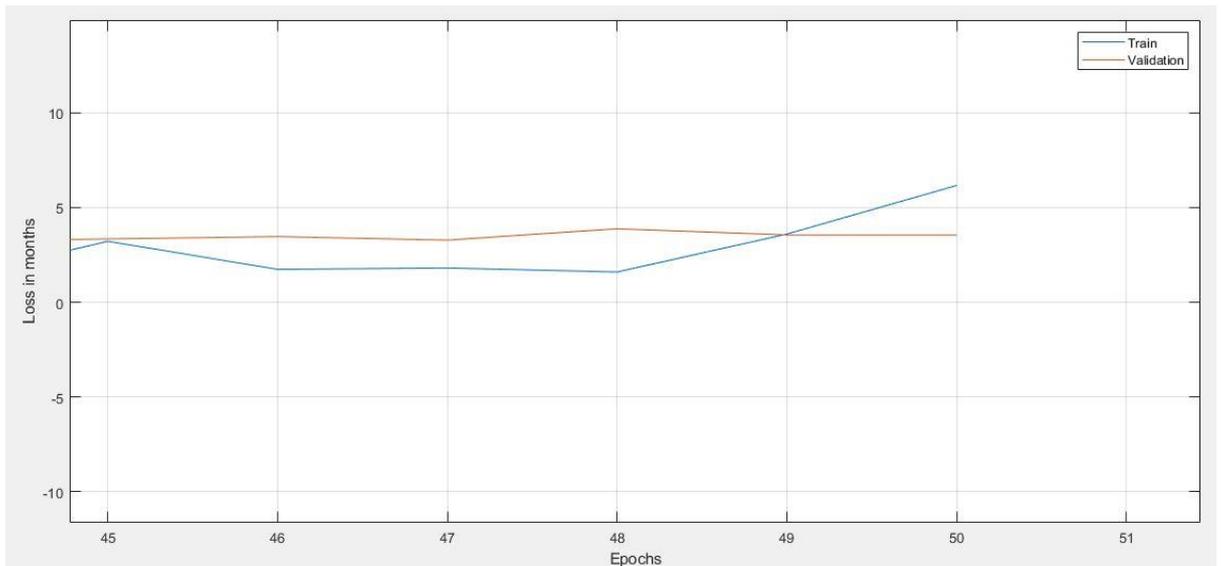


Figure III.6 : La partie zoom de la perte en mois.

Le graphe représente la perte en mois par rapport aux nombres des époques

-la couleur rouge validation

-la couleur bleu train

Pour validation :

Est près que constante durant toutes les époques

Pour train :

- de l'époque 45 le taux de perte est augmentée
- après l'époque 45 jusqu'à 48 le taux de perte est constant
- de l'époque 48 jusqu'à 50 le taux augmente

La courbe représente l'erreur absolue a un mois qui montre la variation d'erreur moyen de l'âge (en mois) par rapport au nombre des époque c à dire que si le nombre d'itération augment l'erreurs diminue.

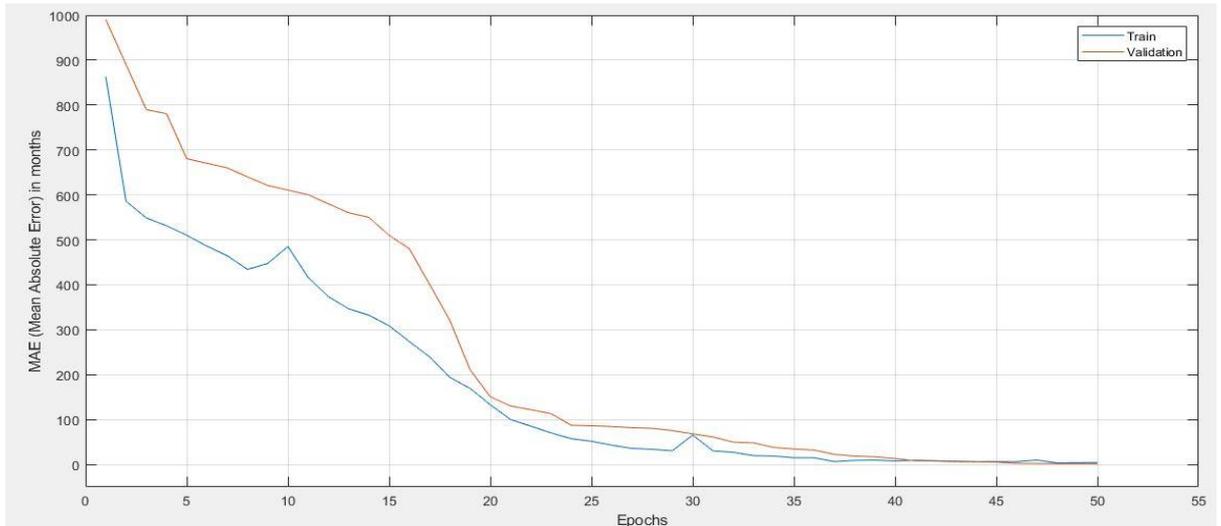


Figure III.7 : Représentation d'erreur absolue à un mois

-on dit que si le nombre de époques augmente l'erreur diminue et plus que l'époque augmente plus que le modèle s'entraîne bien.

-la partie zoom du figure III.7

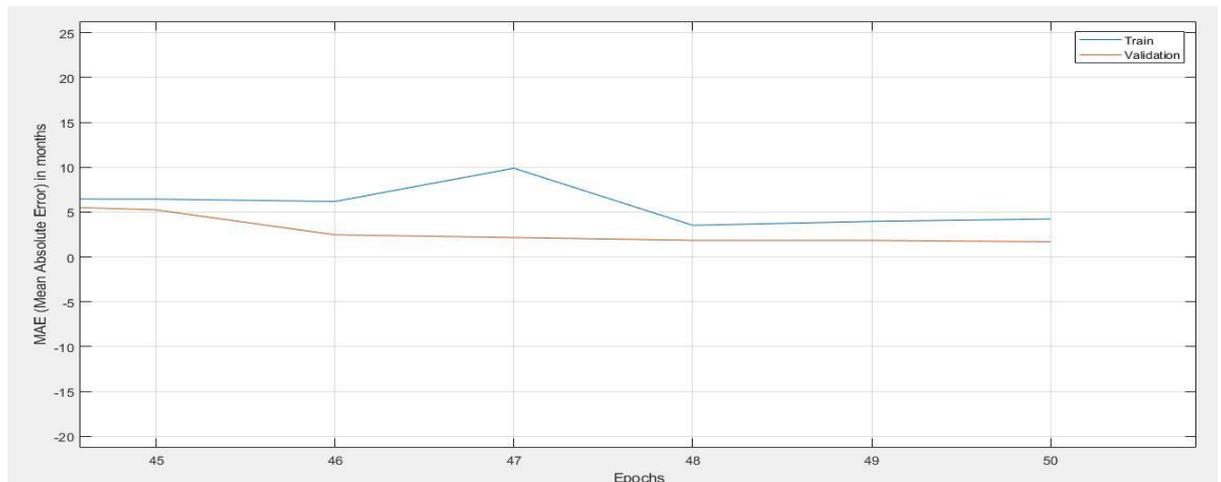


Figure III.8 : Partie zoom d'erreur absolue à un mois.

- **ViT 32**

Les images sont présentées dans ce modèle sous la forme d'une séquence de patches de taille fixe (résolution 32x32), qui sont intégrés de manière linéaire. On ajoute également un jeton [CLS] au début d'une séquence pour l'utiliser pour des tâches de classification. On ajoute également des plongements de position absolue avant d'envoyer la séquence aux couches de l'encodeur Transformer.

-Le modèle ViT a été préformé sur ImageNet-21k, un jeu de données composé de 14 millions d'images et de 21k classes, et affiné sur ImageNet, un jeu de données composé de 1 million d'images et de 1k classes.

### 2<sup>eme</sup> code

### Structure d'Évaluation de l'âge osseux à l'aide de vision transformer



## 1er étape importation des bibliothèques

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import tensorflow as tf
import datetime, os
import math
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from tensorflow import keras
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Conv2D, MaxPool2D, Flatten, Dropout, BatchNormalization
from tensorflow.keras.layers import GlobalAveragePooling2D, Dense, Flatten
from tensorflow.keras.callbacks import TensorBoard, ModelCheckpoint, EarlyStopping, ReduceLRonPlateau
from tensorflow.keras import Sequential
from tensorflow.keras.applications.inception_v3 import InceptionV3
from tensorflow.keras import Model
```

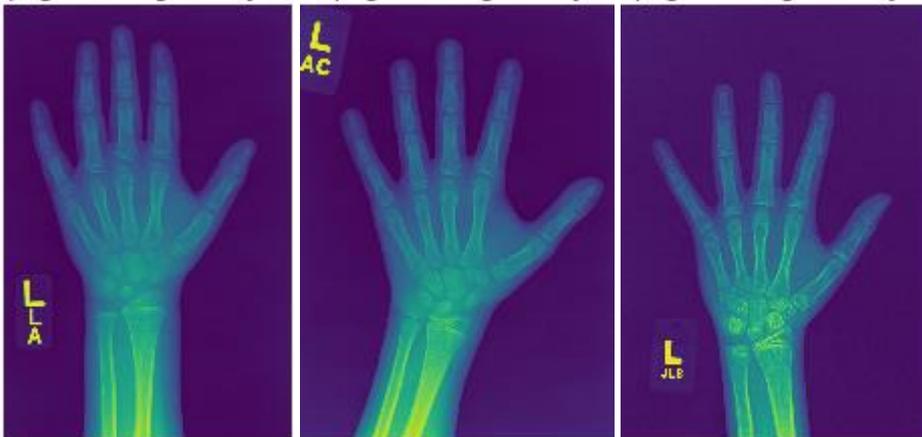
- **Panda** est une bibliothèque écrite pour le langage de programmation python permettant la manipulation et l'analyse des données. Elle propose en particulier des structures de données et des opérations de manipulation de tableaux numérique et de séries temporelles. Panda est un logiciel libre sous licence BSD.
- **TensorFlow** c'est une bibliothèque open source qui est compatible avec Python c'est une technologie qui rend l'Intelligence Artificielle.
- **Scikit-learn(sklearn)** est une bibliothèque d'apprentissage automatique gratuite pour Python. Il comporte divers algorithmes tels que la machine à vecteurs de support, les forêts aléatoires et les k-voisins, et il prend également en charge les bibliothèques numériques et scientifiques Python telles que NumPy et SciPy.
- **loading dataframe** :Cet article explique comment insérer des données SQL dans une trame de données Pandas à l'aide du package pyodbc en Python

```
#loading dataframes
train_df = pd.read_csv('/kaggle/input/rsna-bone-age/boneage-training-dataset.csv')
test_df = pd.read_csv('/kaggle/input/rsna-bone-age/boneage-test-dataset.csv')

#appending file extension to id column for both training and testing dataframes
train_df['id'] = train_df['id'].apply(lambda x: str(x)+'.png')
test_df['Case ID'] = test_df['Case ID'].apply(lambda x: str(x)+'.png')

train_df.head()
```

```
import matplotlib.image as mpimg
for filename, boneage, gender in train_df[['id','boneage','gender']].sample(4).values:
    img = mpimg.imread('/kaggle/input/rsna-bone-age/boneage-training-dataset/boneage-training-dataset/'+ filename)
    plt.imshow(img)
    plt.title('Image name:{} Bone age: {} years Gender: {}'.format(filename, boneage/12, gender))
    plt.axis('off')
    plt.show()
```



13,5 ans mal

9 ans mal

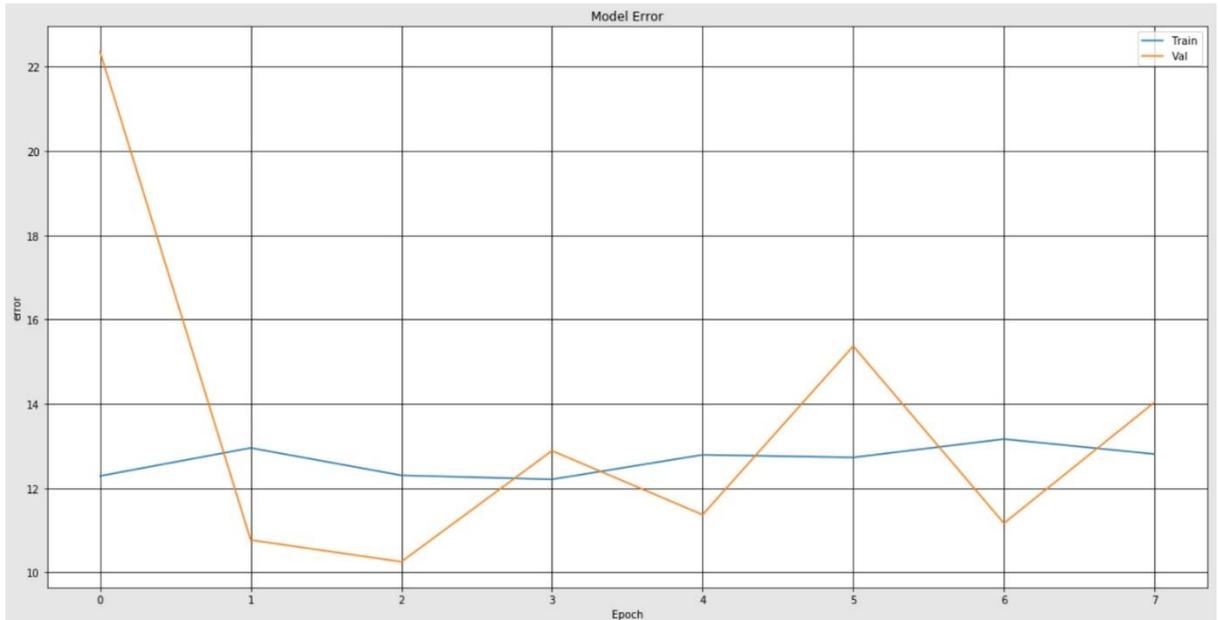
11 ans femelle

**Figure III. 9 :** Exemple de répartition par âge et sexe dans le résultat.

- **VGG19**

**VGG19** est une variante du modèle VGG qui, en résumé, se compose de 19 couches (16 couches de convolution, 3 couches entièrement connectées, 5 couches MaxPool et 1 couche SoftMax). Il existe d'autres variantes de VGG comme VGG11, VGG16 et d'autres. Le VGG19 a 19,6 milliards de FLOPs.

```
from keras.applications.vgg19 import VGG19 #VGG19
model_3=VGG19(input_shape=(256,256,3),
              include_top=False,
              weights='imagenet') #VGG19
for layer in model_3.layers:
    layer.trainable=False
x=model_3.output
x=GlobalAveragePooling2D()(x)
x=Flatten()(x)
x=Dense(512,activation='relu')(x)
predictions=Dense(1,activation='linear')(x)
model_final_3=Model(model_3.input,predictions)
```



**Figure III.10** : La variation d'erreur en fonction des époques pour VGG19.

- **Interprétation**

Disons que nous remarquons une valeur élevée d'erreurs (22 erreurs) dans le cas de l'utilisation de val à la valeur époque 0, contrairement à la valeur de train, elle est faible dans la plage de 12 erreurs, et nous remarquons une stabilité lorsque la valeur d'époque a augmenté.

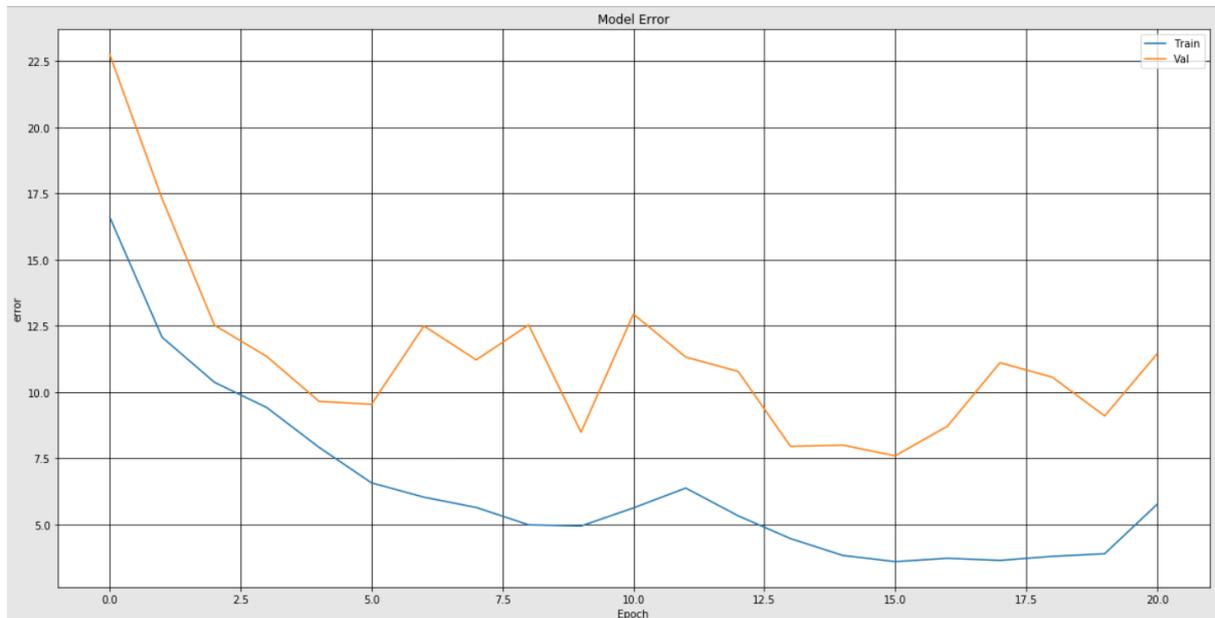
- **ResNet50**

**ResNet50** est une variante du modèle ResNet qui possède 48 couches de convolution ainsi qu'une couche MaxPool et une couche Average Pool. Il compte  $3,8 \times 10^9$  opérations en virgule flottante.

```

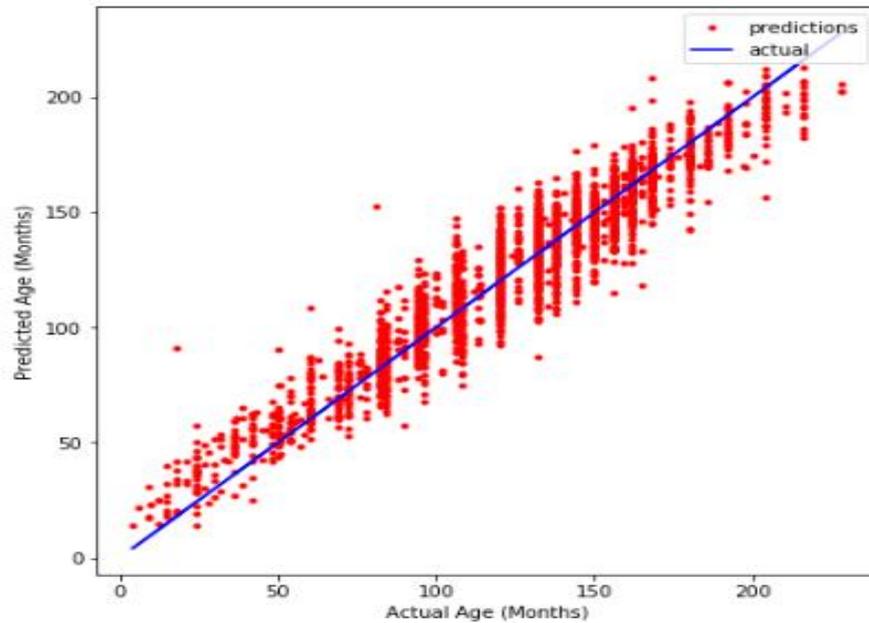
from tensorflow.keras.applications import ResNet50 #ResNet50
model_2=ResNet50(input_shape=(256,256,3),
                 include_top=False,
                 weights='imagenet') #VGG19
for layer in model_2.layers:
    layer.trainable=False
x=model_2.output
x=GlobalAveragePooling2D()(x)
x=Flatten()(x)
x=Dense(512,activation='relu')(x)
predictions=Dense(1,activation='linear')(x)
model_final_2=Model(model_2.input,predictions)

```



**Figure III. 11 :** Variation d'erreur en fonction des époques pour ResNet.

```
fig, ax = plt.subplots(figsize = (7,7))
ax.plot(test_months, pred, 'r.', label = 'predictions')
ax.plot(test_months, test_months, 'b-', label = 'actual')
ax.legend(loc = 'upper right')
ax.set_xlabel('Actual Age (Months)')
ax.set_ylabel('Predicted Age (Months)')
```



**Figure III.12 :** Prédiction de l'Age osseux.

### Interprétation

À partir de cette image, il devient également de plus en plus clair que notre algorithme fonctionne mieux dans la partie d'âge moyen de notre ensemble de données. Il sous-estime particulièrement les âges élevés et surestime l'âge de nos très jeunes êtres humains. Cependant, cela ne devrait pas nous surprendre puisque l'ensemble de données sur lequel notre modèle a été formé contient moins de données pour les tranches d'âge les plus extrêmes (comme vous pouvez le voir par le nombre des points en rouge).

### Comparaison entre les modèles ViT16, ResNet50 et VGG19

On remarque que l'erreur dans le cas du VGG19 est le plus proche de 0 par rapport à celle du ViT16.

À cause de la faible précision du ViT16 lorsque on utilise une dataset insuffisante l'erreur.

Types des modèles	Accuracy	Erreur
ViT16	76%	24%
ResNet50	90.1%	9.9%
VGG19	99.8%	2%

**Tableau 1 :** Comparaison entres les résultats du modèle.

### **III.6. Conclusion**

Nous avons présenté dans ce chapitre notre système utilisé pour la l'évaluation de l'âge osseux par les modèles de vision transformer, Pour cela on a utilisé les modèles suivants : ViT 16, ResNet50 et VGG19. Durant plusieurs essayé nous avons varié les paramètres d'optimisations "Training options" et le nombre d'itération (Epochs). Pour comparer nos résultats, nous avons aussi fixé les optimisations "Training options" et varié le nombre d'itération (Epochs) seulement, puis on a affiché et analyser les différents résultats pour chaque modèle. La comparaison des résultats nous montre que le VGG19, donne les meilleurs résultats pour l'évaluation de l'âge de os par rapport à autres modèles.

## **Conclusion générale**

En raison des progrès de l'intelligence artificielle, tels que l'apprentissage automatique et l'apprentissage en profondeur, permet de traiter des problèmes plus complexes et d'obtenir de bons résultats, ce qui est très Important dans le domaine de l'imagerie médicale où la précision peut sauver des vies humaines. L'intelligence artificielle offre de nombreux avantages pour explorer le domaine de la santé. Cet Engagement envers l'innovation dans les systèmes de santé pour un avenir meilleur.

Dans ce travail nous avons utilisé des modèles pour la classification et évaluation des âges osseux de la main en testant des modèles de transformateur de vision et en sélectionnant le plus optimal et le plus compatible pour notre problématique.

Dans notre cas nous avons utilisé ViT16, ResNet50, VGG19, et nous avons trouvé comme accyarcy 76% pour le ViT16, VGG19 99.8% et ResNet est de 90.1%.

D'après les expériences vues dans les articles ou par notre propre expérience, et en com on remarque que pour une base de donnes petit le modèle préférer est le ViT16.

En tant que pour le transformateur de vision, plus la base des données est volumineuse est grand, plus le résultat est précis et meilleur.

Enfin, avant de poursuivre, ce travail nous permet d'élargir notre champ connaissances, comparant la formation théorique à la pratique, et le temps passé à lire des articles comme un bon début de recherche. En effet, la réalisation de cet article a été une expérience très riche. Ce dernier est une l'occasion de mettre en pratique les connaissances acquises et Projets de recherche intéressants et précieux en intelligence artificielle et en médecine.

## Références

- [1] Bendaoud Mohammed Habib « Développement De Méthodes D'extraction De Contours Sur Des Images À Niveaux De Gris » Thèse Doctorat 2016/2017 U.S.T.O
- [2] Khalfallah Sara « Introduction À L'imagerie Médicale » Master En Mathématiques Univ Biskra 2020
- [3] Larbi Messaouda « méthode de segmentation par ensembles de niveaux (level set) avec contrainte stochastique en imagerie de tomodensimètre » thèse doctorat 2020 univ batna2.
- [4] M. Karim Togola « indicateurs de conformité des demandes d'examens d'imagerie dans le service de radiologie et de médecine nucléaire du chu point g » thèse de doctorat en médecine université de bamako faculté de médecine et d'odonto-stomatologie 2014
- [5] <https://Www.Cea.Fr/Comprendre/Pages/Sante-Sciences-Du-Vivant/Essentiel-Sur-Imagerie-Medicale.AspX>
- [6] Alvarez C. éthique et pratiques médicales dans le domaine de l'imagerie. [Université rené descartes (paris v)] 2010.
- [7] S. Ouddane, « compression des images médicales 3d ». Mémoire de magister département d'informatique, usto-mb, 2009.
- [8] Mazouz Chahrazed « Deep Learning Pour La Classification Des Images Médical À Rayon X » Mémoire De Master Univ Biskra 2021
- [9] R. Bentata, « Segmentation D'images Tomographiques Par Émission De Positions ». Memoir De Magister, UABO, 2011
- [10] Elizabeth Adams, Karen Flynn-Positron Emission Tomography-Management Decision and Research Center-1998
- [11] Erwan Bernard « comparaison théorique et expérimentale de l'imagerie active et de l'ir2 après traitement en conditions dégradées » Thèse doctorat de l'université de Toulouse 2015
- [12]. [https://fr.wikipedia.org/wiki/Imagerie\\_hyperspectrale](https://fr.wikipedia.org/wiki/Imagerie_hyperspectrale)
- [13] Hacene Bellahmer Mémoire de Master « Implémentation et évaluation d'un modèle d'apprentissage automatique pour l'estimation de la valeur marchande de propriétés immobilières »2019 /2020 De Tizi-Ouzou
- [14] [http://cazencott.info/dotclear/public/lectures/IntroML\\_Azencott.pdf](http://cazencott.info/dotclear/public/lectures/IntroML_Azencott.pdf)
- [15] Medjdoubi Abdelkader « L'analyse du sentiment utilisant le deep learning » mémoire de master université saida 2019

- [16] <https://www.juripredis.com/fr/blog/id-19-demystifier-le-machine-learning-partie-2-les-reseaux-de-neurones-artificiels>
- [17] Rosebrock, Adrian. Deep Learning for Computer Vision with Python: Starter Bundle. PyImage Search, (2017).
- [18] <https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-rn.pdf>
- [19] Khelalef Aziz « Reconnaissance d'activités humaines en utilisant les descripteurs spatio-temporels 2D/3D » thèse de doctorat, Université Batna 2020
- [20] [https://fr.wikipedia.org/wiki/Apprentissage\\_par\\_transfert](https://fr.wikipedia.org/wiki/Apprentissage_par_transfert)
- [21] [https://stringfixer.com/fr/Inductive\\_transfer](https://stringfixer.com/fr/Inductive_transfer)
- [22] <https://machinelearningmastery.com/transfer-learning-for-deep-learning/>
- [23] <https://ledatascientist.com/introduction-au-transfer-learning/>
- [24] <https://datascientest.com/transfer-learning>
- [25] <https://analyticsindiamag.com/hands-on-guide-to-using-vision-transformer-for-image-classification/>
- [26] <https://arxiv.org/abs/1706.03762>
- [27] <https://viso.ai/deep-learning/vision-transformer-vit/>
- [28] <https://theaisummer.com/vision-transformer/>
- [29] <https://viso.ai/deep-learning/vision-transformer-vit/>
- [30] [https://medicaldocs.org/1/anatomie/Cours/pdf/Squelette\\_de\\_la\\_main\\_](https://medicaldocs.org/1/anatomie/Cours/pdf/Squelette_de_la_main_)