

APPRENTISSAGE PROFOND EN IMAGERIE MÉDICALE  
PLAN DE COURS - M2 SNS -

**Public Cible :** Étudiants en Master ou Doctorat avec une solide connaissance des CNN pour la vision par ordinateur (architectures de type ResNet, DenseNet, formation par rétropropagation, etc.).

**Objectif du cours :** Fournir aux étudiants les compétences nécessaires pour adapter et développer des modèles d'apprentissage profond performants et fiables pour des tâches d'analyse d'images médicales, en se concentrant sur les défis spécifiques au domaine tels que la rareté des données, la quantification des incertitudes et l'interprétabilité des résultats.

A voir :

- interprétabilité LIME, SHAP
- Federative learning : défis techniques (hétérogénéité des données, communication) et des solutions pour collaborer de manière sécurisée
- articles sur "Fairness and Bias in Medical AI" ou "Ethical considerations for AI in healthcare". Ces ressources sensibiliseraient les étudiants aux risques de créer des modèles qui pourraient exacerber les inégalités en santé.



# Chapitre 1

## Données en Imagerie Médicale

L'imagerie joue un rôle central dans le processus des soins de santé. L'imagerie est cruciale non seulement pour les soins de santé, mais aussi pour la communication et l'éducation médicales, ainsi que pour la recherche. En fait, une grande partie de nos progrès récents, en particulier dans le domaine du diagnostic, peut être attribuée à la disponibilité de techniques d'imagerie de plus en plus sophistiquées qui non seulement montrent la structure du corps avec des détails incroyables, mais aussi la fonction des tissus à l'intérieur du corps.

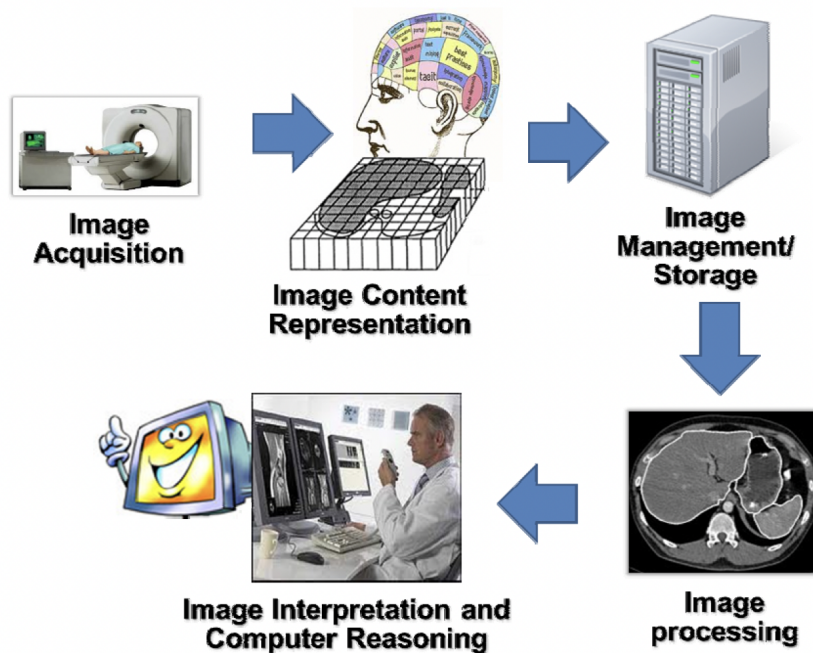


FIGURE 1.1 – Source : [?] The major topics in biomedical imaging informatics include image acquisition, image content representation, management/storage of images, image processing, and image interpretation/computer reasoning (Figure 9.1). Image acquisition is the process of generating images from the modality and converting them to digital form if they are not intrinsically digital

L'imagerie médicale joue un rôle significatif dans diverses applications cliniques, notamment pour la détection précoce, le suivi, le diagnostic et l'évaluation du traitement de nombreuses

conditions médicales. Son utilisation pour le diagnostic est en constante augmentation, venant compléter l'examen clinique et d'autres types d'investigations.

Née il y a un peu plus d'un siècle, l'imagerie médicale a évolué grâce aux grandes découvertes de la physique (rayons X, radioactivité, propriétés magnétiques) et s'appuie sur les avancées en médecine, chimie, mathématiques et informatique. Les progrès récents dans ce domaine sont si importants que de nombreuses spécialités médicales qui dépendent fortement de l'interprétation des images, comme la radiologie, la pathologie et l'ophtalmologie, ont vu un essor de la recherche en intelligence artificielle (IA).

Au-delà de la simple observation de la structure d'un organe, il est maintenant possible de le voir fonctionner grâce à des images fixes ou animées. Les techniques d'imagerie permettent de distinguer les informations anatomiques/structurelles (taille et forme des organes) des informations fonctionnelles (composition ou fonction des tissus). Par exemple, l'IRM peut être utilisée pour l'étude anatomique (structure du cerveau), fonctionnelle (suivi de son fonctionnement) et de diffusion (connaissance de l'organisation tissulaire).

## 1.1 Qu'attend-on des images ?

En général, il y a deux stratégies différentes dans l'imagerie du corps : (1) délimiter la structure anatomique structure (anatomic/structural imaging), et (2) déterminer la composition des tissus ou leur fonction fonction (functional imaging).

**L'imagerie de la structure anatomique** du corps a été et continue d'être la principale application de l'imagerie médicale. Son objectif est de délimiter clairement la position, forme, taille, anomalie, etc des organes.

**L'imagerie fonctionnelle**, qui vise à déterminer la composition ou la fonction des tissus. Il est maintenant possible non seulement d'observer un organe, mais aussi de le voir fonctionner grâce à des images fixes ou animées. Ainsi, la fonction peut être déduite en observant les changements de structure au fil du temps. Ces dernières années, cette capacité à imager la fonction s'est grandement accélérée. Par exemple, l'échographie et l'angiographie sont largement utilisées pour montrer le fonctionnement du cœur.

**Un équilibre entre structure et fonction** : En réalité, la plupart des modalités d'imagerie fournissent des informations sur la morphologie et la fonction. Il n'existe pas de modalité d'imagerie parfaite ; chacune présente un compromis entre la résolution spatiale, la résolution de contraste et la capacité à représenter la fonction physiologique.

Une application particulièrement importante de l'imagerie fonctionnelle est de comprendre l'activité cognitive du cerveau.

## 1.2 Panorama des modalités d'imagerie

<https://www.cea.fr/comprendre/Pages/sante-sciences-du-vivant/essentiel-sur-imagerie-m.aspx>

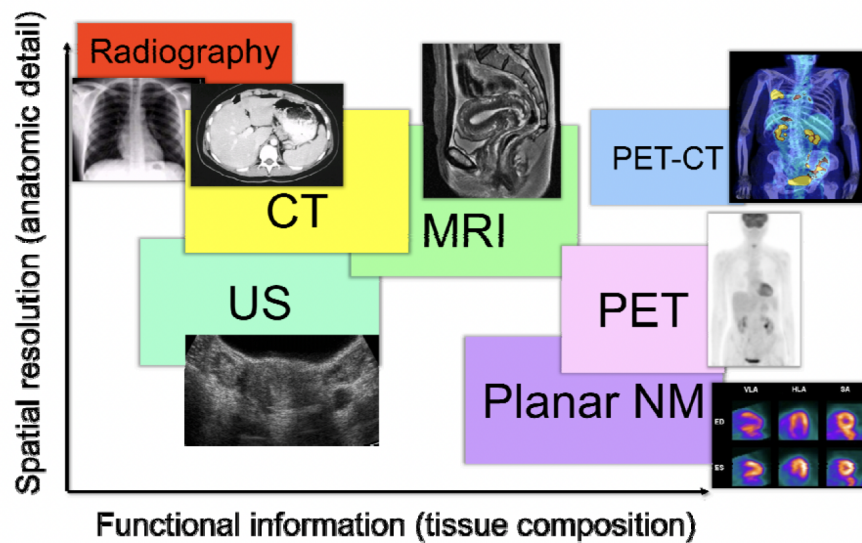


FIGURE 1.2 – Source : [?]

L'imagerie médicale joue un rôle central dans les soins de santé, le diagnostic, le suivi des maladies et la recherche. Il existe une grande variété de techniques, ou modalités, qui diffèrent principalement par la source d'énergie utilisée pour générer les images. Chaque modalité possède des caractéristiques uniques en termes de résolution, de type d'information fournie (anatomique ou fonctionnelle) et d'applications cliniques.

### 1.2.1 Radiographie (Rayons X)

La radiographie est la modalité d'imagerie la plus courante dans le monde, souvent utilisée pour le dépistage et le diagnostic.

**Principe** : Cette technique de projection utilise des rayons X, une forme de rayonnement ionisant. Les rayons X ont des longueurs d'onde plus courtes que la lumière visible, ce qui leur permet de traverser la plupart des tissus mous du corps humain. Cependant, les tissus plus denses comme les os, qui contiennent du calcium, absorbent ou diffusent davantage ces rayons. Un détecteur ou un film placé de l'autre côté du patient capture les rayons X restants, créant une image négative où les zones exposées à plus de lumière (tissus mous comme les poumons) apparaissent plus sombres, et les os apparaissent plus clairs. Des agents de contraste, comme des produits iodés, peuvent être utilisés pour rehausser le contraste de certaines structures.

**Caractéristiques** : La radiographie offre une très haute résolution spatiale. Cependant, comme il s'agit d'une projection, elle entraîne une superposition de toutes les structures traversées par le faisceau, ce qui peut créer des ambiguïtés de localisation. Sa résolution de contraste est relativement faible par rapport à d'autres modalités, et elle fournit des informations fonctionnelles limitées. **Applications** : Elle est couramment utilisée pour la détection de fractures osseuses, de problèmes dentaires, de pneumonies, de tuberculose et de certains types de tumeurs. La fluoroscopie, une variante, permet une visualisation en temps réel. Tomodensitométrie (TDM ou CT - Computed Tomography)

### 1.2.2 Tomodensitométrie (Rayons X)

La TDM, également appelée scanner, est une méthode d'imagerie qui utilise aussi les rayons X pour produire des images en coupe du corps.

**Principe** : Un faisceau de rayons X étroit tourne autour du patient, et de multiples clichés sont pris sous différents angles. Des algorithmes informatiques de reconstruction estiment alors la densité absolue en chaque point (voxel) du volume exploré, éliminant le problème de superposition de la radiographie conventionnelle. L'empilement de ces coupes transversales permet de générer une représentation tridimensionnelle (3D) qui est beaucoup plus informative qu'une radiographie classique.

**Caractéristiques** : La résolution spatiale de la TDM est inférieure à celle de la radiographie, mais sa résolution de contraste est bien supérieure, permettant une meilleure distinction des tissus mous. Elle a une grande capacité de détection et peut révéler de petites lésions.

**Applications** : La TDM est un outil de diagnostic populaire pour identifier des tumeurs ou des lésions dans l'abdomen, localiser des blessures à la tête, et diagnostiquer des fractures osseuses complexes. Elle est fréquemment utilisée pour l'identification de nodules pulmonaires.

### 1.2.3 Imagerie par Résonance Magnétique (IRM - MRI)

L'IRM n'utilise pas de rayonnements ionisants mais un champ magnétique et des ondes de radiofréquence pour créer des images détaillées.

**Principe** : L'IRM force les protons du corps à s'aligner avec un champ magnétique puissant. Une impulsion de radiofréquence perturbe cet alignement. Lorsque l'impulsion est arrêtée, les protons libèrent de l'énergie en se réalignant, et cette énergie varie selon les types de tissus (basé sur les temps de relaxation T1 et T2). Cette variation permet de différencier les différentes régions du corps.

**Caractéristiques** : L'IRM offre à la fois une haute résolution spatiale et des informations fonctionnelles détaillées sur la composition des tissus. Elle est particulièrement efficace pour l'imagerie des tissus mous et non osseux (cerveau, moelle épinière, nerfs, muscles), où elle est souvent supérieure à la TDM.

**Applications** : C'est la modalité de choix pour des tâches telles que la détection de tumeurs cérébrales ou l'identification de lésions tissulaires. Des variantes comme l'angiographie par résonance magnétique (ARM) permettent de visualiser le flux sanguin.

L'IRM permet non seulement d'étudier la morphologie des organes (IRM anatomique), mais aussi de suivre le fonctionnement du cerveau (IRM fonctionnelle) et de connaître l'organisation des tissus, comme les fibres neuronales (IRM de diffusion)

### 1.2.4 Tomographie par Émission de Positrons (TEP - PET)

La TEP est une technique d'imagerie de médecine nucléaire principalement axée sur l'information fonctionnelle.

**Principe** : Un traceur radioactif est chimiquement lié à une molécule biologiquement active (comme un analogue du glucose) et injecté au patient. Ce composé s'accumule dans les tissus métaboliquement actifs (par exemple, les tumeurs) et émet des rayonnements (positrons). Ces

rayonnements sont détectés par un scanner qui crée une image montrant la concentration du traceur dans le corps.

**Caractéristiques :** La TEP fournit des informations fonctionnelles de haute qualité, visualisant l'activité moléculaire au niveau des tissus. Cependant, sa résolution spatiale est faible. Pour pallier ce défaut, elle est souvent combinée à un scanner TDM dans une machine hybride (TEP/TDM), qui fusionne l'information fonctionnelle de la TEP avec la localisation anatomique précise de la TDM.

**Applications :** La TEP est largement utilisée en oncologie pour le diagnostic, la stadification et le suivi de la réponse au traitement des cancers.

### 1.2.5 Mammographie (MG)

La mammographie est une technique de radiographie à faible dose spécifiquement conçue pour l'imagerie du sein.

**Principe :** Elle utilise des rayons X pour visualiser la structure interne du sein. C'est l'outil le plus fiable et le plus courant pour la détection précoce du cancer du sein.

**Caractéristiques :** Le défi principal en mammographie est la détection de tumeurs qui ne représentent qu'une petite partie de l'image. Des systèmes de classification standardisés comme le BI-RADS (Breast Imaging Reporting and Data System) sont utilisés pour décrire les observations.

**Applications :** Dépistage et diagnostic du cancer du sein, notamment par la détection de masses ou de microcalcifications.

### 1.2.6 Échographie (Ultrasound)

L'échographie utilise des ondes sonores à haute fréquence pour créer des images des structures corporelles.

**Principe :** Une sonde émet des impulsions sonores qui traversent le corps. Lorsque ces ondes rencontrent des interfaces entre différents types de tissus (avec des impédances acoustiques différentes), une partie est réfléchi sous forme d'écho. La sonde capte ces échos, et un ordinateur les utilise pour construire une image en temps réel. Le temps de retour de l'écho détermine la profondeur de la structure, et l'amplitude de l'écho sa luminosité sur l'image.

**Caractéristiques :** L'échographie n'utilise pas de rayonnement ionisant. Elle fournit des informations fonctionnelles, notamment sur la composition des tissus et le flux sanguin (via les méthodes Doppler). Sa résolution spatiale est généralement plus faible que celle d'autres modalités. Elle permet une imagerie en temps réel (cineloops).

**Applications :** Visualisation du mouvement du cœur, étude du flux sanguin dans les vaisseaux, imagerie fœtale, etc.

### 1.2.7 Images d'histopathologie

L'histopathologie est l'étude de tissus au microscope pour identifier des maladies. **Principe :** Des échantillons de tissus sont prélevés, fixés sur des lames de verre, puis colorés (par exemple, avec de l'hémaroxyline et de l'éosine, H&E) pour mettre en évidence des structures spécifiques comme les noyaux cellulaires. La pathologie numérique (ou digitale) implique la numérisation

de ces lames à très haute résolution pour créer des images de lames entières (WSI - Whole Slide Images).

**Caractéristiques :** Ces images sont massives (plusieurs gigaoctets) et permettent une analyse détaillée au niveau cellulaire. Les méthodes d'apprentissage profond (Deep Learning) sont de plus en plus utilisées pour analyser ces images, notamment pour la détection de noyaux, la segmentation de cellules ou la classification de tissus.

**Applications :** Diagnostic et classification des cancers (rein, poumon, sein, etc.), évaluation du grade tumoral.

## 1.3 Nature des données

Les images sont des données non structurées. Bien qu'un ordinateur puisse facilement stocker et récupérer les données visuelles brutes (les pixels), il ne peut pas accéder directement à leur contenu sémantique ou "contenu de connaissance" (reconnaître l'anatomie, les anomalies, etc.). De même, les rapports de radiologie, rédigés en texte libre par exemple, sont également non structurés et difficiles à exploiter par des systèmes informatiques. Le développement d'ontologies et de terminologies contrôlées est une approche fondamentale pour rendre ce contenu de connaissance explicite et accessible aux machines.

**Dimensions des images :** Les données peuvent être des images 2D ou des volumes 3D.

- Les images 2D, comme les radiographies ou les images de microscopie, sont représentées par une matrice bidimensionnelle de pixels (éléments d'image).
- Les volumes 3D, typiques de l'IRM, de la TDM ou de la TEP, sont représentés par une matrice tridimensionnelle de voxels (éléments de volume). Un volume 3D peut être vu comme un empilement d'images 2D (coupes).

**Type de valeurs :** La plupart des images médicales sont en niveaux de gris (grayscale). Chaque pixel ou voxel a une valeur numérique représentant une intensité (par exemple, l'atténuation des rayons X en TDM). Pour la TDM, ces valeurs (unités Hounsfield) couvrent une large plage, et les opérateurs utilisent une technique de "fenêtrage" pour ajuster le contraste et la luminosité afin de visualiser des tissus d'intérêt.

**Autres caractéristiques des données :**

**Isotropie :** Les voxels sont dits isotropes s'ils ont les mêmes dimensions dans toutes les directions (x, y, z). Si ce n'est pas le cas (par exemple, l'épaisseur de coupe est plus grande que la résolution dans le plan), ils sont anisotropes, ce qui peut compliquer les reconstructions dans des plans arbitraires.

**Métadonnées :** Chaque fichier d'image médicale contient des métadonnées essentielles stockées dans son en-tête, telles que le nom du patient, la date de l'examen, la modalité utilisée, l'épaisseur de coupe, etc.. Le standard DICOM (Digital Imaging and Communications in Medicine) est largement utilisé pour normaliser le format de ces fichiers et la communication entre les appareils d'imagerie.

**Données multi-dimensionnelles et temporelles :** Les données peuvent avoir plus de trois dimensions. Par exemple, une IRM peut inclure plusieurs séquences (T1, T2), créant un "vecteur de caractéristiques" pour chaque voxel. De même, les vidéos (comme en échographie cardiaque) ajoutent une dimension temporelle.



## 1.4 Spécificités des données médicales

Voir incertainties in Dojat's paper

1. Grande dimension
2. La rareté des données étiquetées et le coût de l'annotation par des experts.
3. La variabilité des données : bruit intrinsèque (artefacts, faible rapport signal/bruit) et variabilité inter-annotateurs [?].  
Variabilité inter-patient pour certains organes (ex vessie)
4. Le décalage de distribution (domain shift) entre les hôpitaux, les machines et les populations de patients.
5. La nécessité d'une IA digne de confiance pour une adoption clinique, au-delà de la performance pure.

### 1.4.1 Qualité des images

La qualité et la valeur diagnostique d'une image médicale sont principalement évaluées selon trois caractéristiques fondamentales. Aucune modalité d'imagerie n'est parfaite ; chacune représente un compromis entre ces différents aspects.

**Résolution Spatiale** C'est une mesure de la **netteté de l'image**, c'est-à-dire la capacité de la modalité à distinguer des points très proches sur l'objet imagé. Pour une image numérique, elle est généralement liée au nombre de pixels par unité de surface.

- La **radiographie** offre une très haute résolution spatiale.
- L'**IRM** a également une haute résolution spatiale.
- À l'inverse, l'**échographie** et l'**imagerie de médecine nucléaire** (TEP, TEMP) ont une faible résolution spatiale en raison d'un flux de photons ou d'ondes plus faible.

**Résolution de Contraste** Elle mesure la capacité à distinguer de **faibles différences d'intensité** entre différentes régions de l'image. Pour les images numériques, elle est liée au nombre de bits par pixel. Une bonne résolution de contraste est essentielle pour différencier les tissus mous.

- La **Tomodensitométrie (TDM)** et l'**IRM** ont une résolution de contraste supérieure à celle de la radiographie conventionnelle.
- Des **agents de contraste** (produits iodés en TDM, gadolinium en IRM) sont souvent utilisés pour rehausser artificiellement le contraste de certaines structures.

**Résolution Temporelle** Il s'agit du **temps nécessaire pour créer une image**. Une haute résolution temporelle est cruciale pour les applications en temps réel, comme l'imagerie du cœur battant, qui nécessite au moins 30 images par seconde pour éviter le flou. L'échographie est un bon exemple de modalité avec une haute résolution temporelle.

### 1.4.2 Variabilité des données

Les sources de variabilité et d'erreurs en imagerie médicale sont multiples et complexes. Elles proviennent à la fois du processus physique d'acquisition, des caractéristiques des patients, des méthodes d'annotation et de la manière dont les ensembles de données sont constitués.

La prise en compte de ces facteurs est essentielle pour développer des solutions d'IA qui soient non seulement performantes mais aussi fiables et dignes de confiance en pratique clinique.

### **Variabilité liée aux Images d'Entrée**

Cette catégorie concerne les erreurs et variations inhérentes au processus de capture des données et aux caractéristiques physiques des images. Elle correspond en grande partie à ce que l'on nomme l'incertitude aléatoire (aleatoric uncertainty), qui décrit le bruit et les effets aléatoires intrinsèques aux données

**Bruit et Artefacts Techniques** Bruit intrinsèque au processus d'acquisition : Les images médicales peuvent présenter un faible rapport signal/bruit, des artefacts (dus aux mouvements du patient, aux implants métalliques, etc.) ou un effet de volume partiel, où un seul voxel peut contenir plusieurs types de tissus, rendant les frontières floues.

Incohérences de qualité : Les données peuvent varier en termes de résolution et de contraste, en raison des procédures cliniques utilisées. Un exemple concret est un signal d'auscultation cardiaque qui peut être corrompu par du bruit lors de sa capture.

**Décalage de Domaine (Domain Shift) lié aux Protocoles et Équipements** C'est l'une des sources de variabilité les plus importantes. Les modèles d'IA ont souvent du mal à généraliser leurs performances lorsqu'ils sont confrontés à des données provenant d'un environnement différent de celui de l'entraînement.

Différences d'équipements et d'institutions : Il existe une variation significative entre les images acquises dans différents hôpitaux ou avec différentes machines. Les ensembles de données sont souvent composés de sous-ensembles provenant de différents appareils (par exemple, stéthoscopes Welch Allyn Meditron, 3M Littmann) et de différents environnements de capture (laboratoire vs hôpital), ce qui crée des distributions de données distinctes.

Protocoles d'acquisition non standardisés : L'absence de standardisation dans la manière dont les images sont acquises est une limitation majeure en analyse d'images médicales.

**Caractéristiques Spécifiques aux Patients** Des dissimilitudes importantes existent entre les données de patients différents en fonction de leurs caractéristiques démographiques, de leur âge, de leur sexe, etc.. Par exemple, des différences substantielles ont été identifiées dans les données d'EEG au repos d'enfants provenant de différentes régions géographiques.

**Variations liées au Flux de Travail Clinique** La manière dont les images sont manipulées peut introduire des erreurs. Par exemple, l'utilisation de photos de radiographies prises avec un smartphone introduit des artefacts qui ne sont pas présents dans les images numériques originales. Ces artefacts incluent des angles de vue modifiés, des conditions d'éclairage variables, des reflets (glare), des effets de moiré, des rotations, des translations et du flou.

### **1.4.3 Variabilité et Erreurs dans les Annotations (Vérité Terrain)**

Cette variabilité affecte directement la qualité des étiquettes utilisées pour entraîner et évaluer les modèles d'IA supervisés. Elle est également une composante de l'incertitude aléatoire.

**Variabilité Inter-Annotateurs** C'est un phénomène courant où plusieurs experts humains (radiologues, pathologistes) peuvent être en désaccord sur l'interprétation ou la délimitation d'une même image.

Un exemple typique est la segmentation de la prostate sur des images IRM, où l'incertitude est particulièrement observée aux frontières de l'organe en raison des divergences entre les annotations de différents experts.

**Divergence entre les Rapports Textuels et les Labels d'Images** Les étiquettes pour de grands ensembles de données sont souvent extraites automatiquement de rapports radiologiques textuels. Cependant, il existe un désaccord significatif entre les labels issus des rapports et ceux attribués par des radiologues regardant directement les images.

Les raisons de cette divergence incluent :

1. L'accès à l'historique clinique par le radiologue qui rédige le rapport, ce qui peut biaiser son interprétation, alors que celui qui annote l'image ne dispose pas de ce contexte.
2. L'utilisation de différentes sections du rapport (par exemple, la section "Impression" vs "Constatations") qui peuvent ne pas contenir les mêmes informations.
3. Les hiérarchies d'étiquettes, où un radiologue peut mentionner une pathologie spécifique (enfant) sans mentionner la catégorie plus générale (parent).
4. Le bruit inhérent à la tâche d'étiquetage, incluant des erreurs humaines ou des seuils de diagnostic différents entre experts.

#### 1.4.4 Variabilité liée à la Constitution des Ensembles de Données

Enfin, la manière dont les ensembles de données sont créés et structurés est une source majeure de variabilité et de biais potentiels.

**Rareté des Données et Déséquilibre des Classes** La rareté des données est une limitation majeure, en particulier pour les modèles d'apprentissage profond qui nécessitent de grands volumes de données annotées.

Les ensembles de données médicaux sont souvent très déséquilibrés, avec beaucoup plus d'échantillons "normaux" que "anormaux", ce qui peut biaiser l'entraînement du modèle vers la classe majoritaire.

**Manque de Diversité et "Sur-Curation" des Données** De nombreux jeux de données publics sont "hautement organisés" (highly curated), collectés dans des environnements contrôlés et des cadres restreints. Par conséquent, ils ne capturent pas la distribution et la diversité réelles des données cliniques. Un modèle entraîné sur des données trop "propres" peut échouer lorsqu'il est confronté à la complexité et au bruit du monde réel.

Les modèles peuvent aussi être confrontés à des cas non vus ou des maladies rares qui n'étaient pas présentes dans l'ensemble d'entraînement, ce qui constitue une autre forme de décalage de distribution.



# Chapitre 2

## Tâches et Architectures Clés en Apprentissage Profond Médical

Ce module présente les applications fondamentales de l'apprentissage profond (DL) dans ce domaine.

### 2.1 Les grandes tâches de l'analyse d'images médicales

1. Classification : Détecter la présence ou l'absence d'une pathologie.
2. Détection : Localiser des anomalies (ex : nodules pulmonaires) (voir table ??).
3. Segmentation : Délimiter précisément des organes ou des tumeurs.

#### 2.1.1 Classification

L'objectif de la classification est d'attribuer une ou plusieurs étiquettes à une image entière. Il s'agit de répondre à la question : "De quoi s'agit-il sur cette image?". Le modèle est entraîné pour détecter la présence ou l'absence d'une maladie ou pour classer l'image dans une catégorie prédéfinie (par exemple, normale ou pathologique).

Exemples

- détecter la présence de pneumonie sur une radiographie pulmonaire.
- classer des lésions cutanées comme bénignes ou malignes.
- Identifier le type de cancer du poumon (par exemple, adénocarcinome ou carcinome épidermoïde) à partir d'une image d'histopathologie.
- Classer la densité mammaire sur une mammographie.

La sortie d'un algorithme de classification est donc globale, au niveau de l'image. Le modèle produit un vecteur de probabilités pour l'ensemble des classes possibles (par exemple,  $P(\text{Normal})$ ,  $P(\text{Pneumonie})$ ,  $P(\text{Épanchement pleural})$ ). La décision finale est une étiquette unique ou un ensemble d'étiquettes pour toute l'image.

Pour résoudre des problèmes de classification d'images médicales par des outils d'apprentissage profond, les architectures standards de classification d'images sont couramment utilisées. Elles

se terminent généralement par des couches entièrement connectées qui réduisent les cartes de caractéristiques spatiales à un vecteur de probabilités de classe : AlexNet, GoogLeNet, VGG-16, ResNet, DenseNet.

### 2.1.2 Segmentation

L'objectif de la segmentation est de partitionner une image en différentes régions sémantiques, c'est-à-dire de délimiter précisément des objets d'intérêt. Il s'agit de répondre à la question : "Où se trouvent exactement les structures d'intérêt sur cette image?". La tâche consiste à assigner une classe à chaque pixel (ou voxel en 3D) de l'image.

Exemples :

- Délimiter les contours d'une tumeur cérébrale sur une IRM.
- Segmenter le foie et les tumeurs hépatiques à partir de volumes CT.
- Identifier et délimiter les noyaux cellulaires sur des images d'histopathologie.
- Segmenter les reins et les tumeurs rénales sur des images CT.

La sortie d'un algorithme de segmentation est une carte de segmentation (ou un "masque") qui a les mêmes dimensions que l'image d'entrée. Chaque pixel de cette carte est coloré ou étiqueté selon la classe à laquelle il appartient (par exemple, pixel de fond, pixel de tumeur, pixel d'organe sain). L'analyse se fait donc au niveau du pixel, offrant une localisation précise. En apprentissage profond, les architectures de segmentation sont souvent de type encodeur-décodeur pour préserver l'information spatiale. L'encodeur comprime l'image en une représentation de bas niveau, et le décodeur reconstruit une carte de segmentation à partir de cette représentation. U-Net est l'architecture la plus emblématique et la plus utilisée pour la segmentation d'images biomédicales. Ses variantes comme H-DenseUNet ou 3D U-Net sont également très populaires.

### 2.1.3 Détection d'anomalies

**Détection d'Anomalies Supervisée** Classification Binaire (*R-CNN (Region-based CNN)*, *Faster CNN* : mammographie, nodule pulmonaires), Approche par Détection d'Objets (*CluRNN*, *mRFCN (poumon)*, *RetinaNet (détection d'ulcères dans les images d'endoscopie)*)

**Apprentissage multitâches** : Par exemple, une tâche secondaire d'identification du patient peut être couplée à la tâche primaire de détection d'anomalies pour mieux gérer les variations inter-sujets. Un autre exemple est l'apprentissage conjoint de la détection d'images suspectes et de la segmentation des régions d'intérêt dans ces images.

**Détection d'Anomalies non Supervisée** : Auto-Encodeurs (Sparse AE, denoising AE, variational AE), GAN, RNN.

La détection d'anomalies est très souvent abordée comme un problème non supervisé. Dans ce cas, le modèle est entraîné uniquement sur des données considérées comme "normales". Il apprend ainsi la distribution des données normales, et tout ce qui s'écarte de cette distribution est signalé comme une anomalie. Des architectures comme les Auto-Encodeurs (AE) ou les Réseaux Génératifs Antagonistes (GANs) sont typiques de cette approche. L'avantage est qu'il

TABLE 2.1 – Tableau récapitulatif des différences entre Classification et Segmentation

Caractéristique	Classification	Segmentation
Objectif principal	Attribuer une ou plusieurs étiquettes à l'image entière.	Partitionner une image en délimitant des régions d'intérêt <b>pixel par pixel</b> .
Nature de la sortie	Une étiquette de classe ou un vecteur de probabilités (ex : "Pneumonie").	Une <b>carte de segmentation</b> ou un masque de la même taille que l'image d'entrée.
Échelle d'analyse	<b>Globale</b> (au niveau de l'image).	<b>Locale</b> (au niveau du pixel ou du voxel).
Question posée	"Qu'est-ce que cette image représente ?" ou "Y a-t-il une pathologie ?"	"Où se trouvent précisément les différentes structures ?"
Architectures typiques	<b>CNN classiques</b> (ex : ResNet, DenseNet, VGG).	<b>Encodeur-Décodeur</b> , principalement <b>U-Net</b> et ses variantes.

n'est pas nécessaire de disposer d'exemples étiquetés pour chaque type d'anomalie possible [?].

Des AE basés sur des GRU pour la détection de maladies respiratoires, des CNN-AE pour l'analyse d'EEG, et des 3D-CNN-AE pour l'analyse de scanners CT volumiques ont été proposés.

Un GAN est entraîné pour générer des images "normales" réalistes. Pour la détection d'anomalies, lorsqu'une image anormale est présentée au modèle, celui-ci ne parvient pas à la mapper correctement dans l'espace latent et à la reconstruire, révélant ainsi une anomalie. Exemple : Le framework AnoGAN a été utilisé pour la détection d'anomalies dans des images de tomographie par cohérence optique (OCT). Un autre exemple est une méthode basée sur les GAN pour la prédiction de crises d'épilepsie, où le discriminateur est ensuite adapté pour la classification normal/anormal.

## 2.2 Quelques architectures de référence

1. Pour la classification : Réseaux de neurones convolutifs (CNN) de type ResNet et DenseNet.
2. Pour la segmentation : L'architecture U-Net et ses variantes, un standard pour la segmentation biomédicale.
3. Pour la détection d'anomalies

TABLE 2.2 – Tableau comparatif entre la Classification et la Détection d'Anomalies

Caractéristique	Classification	Détection d'Anomalies
<b>Objectif Principal</b>	Attribuer une étiquette parmi un ensemble de classes <b>connues</b> .	Identifier les données <b>inhabituelles</b> ou <b>hors-distribution</b> par rapport à une norme.
<b>Données d'Entraînement</b>	Principalement <b>supervisée</b> , avec des étiquettes pour chaque classe.	Souvent <b>non supervisée</b> , entraînée uniquement sur des données "normales".
<b>Type de Problème</b>	Apprend à <b>discriminer</b> entre plusieurs classes spécifiques.	Apprend à <b>modéliser la normalité</b> et à détecter les déviations.
<b>Gestion des Classes</b>	Opère dans un " <b>monde fermé</b> " : ne reconnaît que les classes vues à l'entraînement.	Opère dans un " <b>monde ouvert</b> " : conçu pour identifier les données nouvelles ou non vues.
<b>Cas Particulier</b>	N/A	Peut être formulée comme une <b>classification binaire supervisée</b> (normal vs anormal), surtout en médecine.

### 2.2.1 UNet pour la segmentation d'images

On peut voir U-Net comme une architecture encodeur-décodeur spécialisée (inspirée des autoencodeurs) conçue pour la segmentation sémantique, mais ce n'est pas un autoencodeur classique, car : il ne reconstruit pas nécessairement l'image d'entrée ; son décodeur prédit un masque de segmentation (ou une sortie similaire) plutôt qu'une copie exacte de l'entrée. Contrairement à un autoencodeur classique, U-Net ajoute des connexions directes entre les couches correspondantes de l'encodeur et du décodeur. Ces connexions transmettent les informations spatiales fines perdues lors du sous-échantillonnage, ce qui est crucial pour des tâches comme la segmentation d'images.

**Image d'entrée** : Le processus commence par l'injection dans le réseau d'une image médicale (ou autre), généralement en niveaux de gris.

**Encodeur (chemin contractant)** : le côté gauche du U compresse l'image d'entrée grâce à des couches de convolution + pooling, ce qui extrait des caractéristiques de plus en plus abstraites tout en réduisant la résolution spatiale (semblable à la partie « encodeur » d'un autoencodeur). À chaque niveau, la taille spatiale diminue tandis que le nombre de canaux de caractéristiques augmente, ce qui permet au modèle de capturer des motifs de plus haut niveau.

**Traitement au goulot d'étranglement (Bottleneck)** : il s'agit de la partie centrale du



réseau, où l'image est la plus réduite. Elle contient une version compacte mais très informative de l'image, qui capture les principales caractéristiques.

**Décodeur (chemin expansif) :** Le décodeur commence à reconstruire la taille originale de l'image par suréchantillonnage. À chaque niveau, il combine les caractéristiques du décodeur avec celles de l'encodeur correspondantes grâce aux connexions de saut, afin de conserver les détails spatiaux fins.

**Connexions de saut pour la précision :** Les connexions de saut aident à préserver la précision spatiale en transférant les caractéristiques détaillées des couches précédentes. Elles sont particulièrement utiles lorsque le modèle doit distinguer les frontières dans des tâches de segmentation.

**Prédiction finale :** Une convolution  $1 \times 1$  en sortie convertit les cartes de caractéristiques affinées en une carte de segmentation finale, où chaque pixel est classé dans une catégorie spécifique (par ex. avant-plan ou arrière-plan). Ce résultat a la même résolution spatiale que l'image d'entrée.

## 2.2.2 Yolo la segmentation d'images

YOLO (You Only Look Once) est utilisé pour les images médicales, notamment dans la recherche et certaines applications cliniques. C'est un modèle de détection d'objets, ce qui correspond bien à des tâches comme repérer une tumeur, un polype, un nodule ou une cellule. Il est rapide, donc utile pour des applications en temps réel (ex. endoscopie, chirurgie assistée). Il fournit à la fois une localisation (boîtes englobantes) et une classification de la région d'intérêt.

### Exemples d'applications

- Radiologie : détection de nodules pulmonaires sur radios thoraciques ou coupes CT.
- Pathologie : identification de cellules cancéreuses sur des lames histopathologiques.
- Gastroentérologie : détection de polypes sur vidéos de coloscopie en temps réel.
- Dermatologie : détection de lésions cutanées.
- Mammographie : repérage de masses ou microcalcifications.

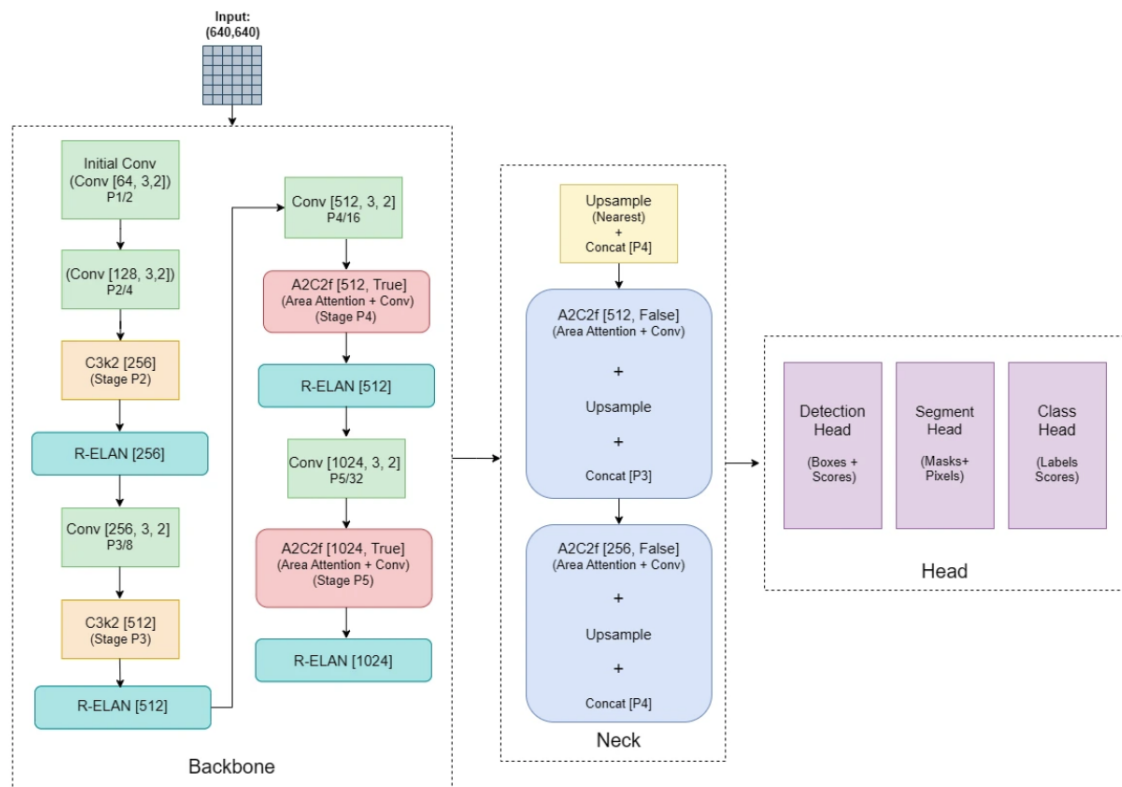
Les limites de YOLO sont les suivantes. YOLO ne fait pas de segmentation détaillée au pixel près ; pour cela on préfère des modèles comme U-Net ou Mask R-CNN. Les ensembles de données médicales sont souvent petits et requièrent du transfer learning (pré-entraînement sur ImageNet puis fine-tuning sur données médicales).

On trouve des exemples sous kaggle, par exemple pour détecter des cellules sanguines <https://www.kaggle.com/code/alaagaberh/blood-cell-detection-using-yolov12>

## 2.3 Stratégies d'Apprentissage en Données Limitées

Objectifs : Maîtriser la technique la plus courante pour pallier le manque de données médicales.  
Contenu :

1. Le paradigme du pré-entraînement sur ImageNet : bénéfices et limites.
2. Analyse critique : Pourquoi la performance sur ImageNet ne se transfère-t-elle pas toujours bien aux tâches médicales ?.
3. Stratégies de fine-tuning.



**YOLOv12 Architecture**

FIGURE 2.1 – Architecture de YOLOv12. Source <https://fr.so-development.org/a-simple-yolov12-tutorial-from-beginners-to-experts/>

4. Efficacité des architectures : Les modèles plus grands sont-ils toujours meilleurs pour les tâches médicales ? Étude de cas sur les familles d'architectures (DenseNet, ResNet).
5. Techniques d'optimisation comme la troncature de modèles pour améliorer l'efficacité sans perte de performance significative.

## 2.4 Apprentissage Auto-Supervisé (Self-Supervised Learning)

Objectifs : Explorer comment extraire des représentations pertinentes d'images médicales non étiquetées.

Contenu :

1. Principe de l'apprentissage contrastif : maximiser l'accord entre les vues d'une même image et minimiser celui entre des images différentes.
2. Étude de cas : MoCo-CXR. Adaptation de la méthode MoCo aux radiographies pulmonaires.
3. Stratégies d'augmentation de données adaptées aux images médicales (rotations, retournements) au lieu des recadrages ou modifications de couleur.
4. Démonstration que les représentations apprises sont de meilleure qualité, surtout avec très peu de données étiquetées.
5. Utilisation des métadonnées patient pour améliorer la sélection des paires positives et négatives.

TP : Fine-tuning et Apprentissage Auto-Supervisé en Pratique

Mettre en œuvre et comparer les stratégies d'apprentissage en données limitées.

Activités :

- Partie A : Fine-tuner un modèle pré-entraîné sur ImageNet (ex : DenseNet121) sur une petite fraction du jeu de données CheXpert pour une tâche de classification.
- Partie B : Utiliser un modèle pré-entraîné avec une méthode auto-supervisée comme MoCo-CXR et le fine-tuner sur la même petite fraction de données.
- Comparer les courbes d'apprentissage et les performances finales pour quantifier le bénéfice de l'apprentissage auto-supervisé spécifique au domaine.



# Chapitre 3

## Construire une IA de Confiance : Quantification de l'Incertitude

Cette partie est entièrement dédiée à la quantification de l'incertitude (UQ), un concept clé pour gérer les erreurs de mesure et la fiabilité des modèles.

### 3.1 Fondamentaux de l'Incertitude en Deep Learning

Objectifs : Comprendre les différentes sources d'incertitude et leur importance clinique.

Contenu

1. Distinction entre l'incertitude aléatoire (aleatoric) et épistémique (epistemic).
2. Incertitude aléatoire : liée au bruit et à la variabilité intrinsèque des données. C'est la manifestation directe des erreurs de mesure, des artefacts et de la variabilité inter-annotateurs.
3. Incertitude épistémique : liée aux limites du modèle, souvent due à un ensemble d'entraînement limité. C'est la conséquence directe des petits échantillons.
4. Pourquoi l'UQ est-elle cruciale ? Pour détecter les erreurs silencieuses, identifier les cas à référer à un expert humain et améliorer la confiance de l'utilisateur.

### 3.2 Méthodes de Quantification de l'Incertitude (Partie 1)

Objectifs : Étudier les méthodes d'UQ les plus populaires basées sur l'échantillonnage.

Contenu

1. Monte Carlo Dropout (MC Dropout) : Approcher l'inférence bayésienne en activant le dropout lors de l'inférence. C'est la méthode la plus populaire en imagerie médicale (plus d'un tiers des implémentations).
2. Deep Ensembles : Agréger les prédictions de plusieurs modèles entraînés indépendamment pour quantifier l'incertitude. Connus pour leur robustesse et la qualité de leurs estimations d'incertitude.

3. Test-Time Augmentation (TTA) : Générer plusieurs prédictions en appliquant des transformations à l'image d'entrée au moment du test. Ces méthodes permettent de capturer à la fois l'incertitude aléatoire et épistémique.

Ces méthodes permettent de capturer à la fois l'incertitude aléatoire et épistémique.

### 3.3 Méthodes de Quantification de l'Incertainitude (Partie 2)

Objectifs : Explorer d'autres paradigmes d'UQ, notamment ceux qui ne nécessitent pas d'échantillonnage multiple.

Contenu :

1. Probabilités Softmax et Calibration : Utiliser directement la sortie du modèle, mais en s'assurant que les probabilités sont bien calibrées.
2. Evidential Deep Learning (EDL) : Modéliser explicitement une masse de croyance pour chaque classe ainsi qu'une masse d'incertitude globale.
3. Conformal Prediction : Une approche statistique qui fournit des ensembles de prédiction avec des garanties statistiques de couverture, sans hypothèse sur le modèle.
4. Modèles Génératifs (ex : Probabilistic U-Net) pour modéliser la distribution des segmentations possibles.

# Chapitre 4

## Évaluation, Robustesse et Déploiement Clinique

Cette dernière partie se concentre sur l'évaluation rigoureuse des modèles et les défis de leur généralisation dans des conditions réelles.

### 4.1 L'Incertitude Structurée : Du Pixel à la Lésion

Objectifs : Aligner les estimations d'incertitude avec l'attention clinique qui se porte sur des structures (lésions, organes) plutôt que sur des pixels.

Contenu :

1. Agrégation des incertitudes au niveau pixel pour obtenir un score de confiance par lésion ou par cas.
2. Utilisation de l'incertitude pour le contrôle qualité (QC) :
  - QC en entrée : Détection d'images non conformes ou hors-distribution (OOD).
  - QC en sortie : Détection de segmentations de mauvaise qualité.

### 4.2 Évaluation de la Robustesse face au Décalage de Distribution

Objectifs : Évaluer systématiquement la capacité du modèle à généraliser.

Contenu :

1. Le problème du décalage de distribution (domain shift) est omniprésent en médecine.
2. Types de décalage : nouvelles machines, nouveaux hôpitaux, nouvelles populations de patients.
3. Protocoles d'évaluation OOD : comment tester la réaction du modèle face à des données qu'il n'a jamais vues ?
4. Étude de cas : analyse de la performance de modèles de radiographie pulmonaire sur des photos de radios ou des données d'une institution externe.
5. L'importance d'évaluer la robustesse des estimations d'incertitude elles-mêmes sous un décalage de domaine.

## 4.3 Généralisation aux Maladies Rares et Imprévues

Objectifs : Comprendre comment les modèles se comportent face à des pathologies pour lesquelles ils n'ont pas été entraînés.

Contenu :

1. Le monde clinique est un "monde ouvert" ; un modèle ne peut pas être entraîné sur toutes les maladies existantes.
2. Étude systématique : Comment un modèle entraîné pour détecter un sous-ensemble de maladies (vues) réagit-il en présence de maladies non vues ?.
3. Constat : les modèles ont tendance à classer les maladies non vues comme "absence de maladie", ce qui peut être dangereux.
4. Stratégies de détection de maladies non vues en utilisant les représentations internes du réseau (ex : avant-dernière couche).
5. Lien entre incertitude et explicabilité (XAI) pour une IA médicale transparente.

TP : Évaluation de la Robustesse du Modèle

Objectifs : Mesurer la dégradation des performances d'un modèle face à un décalage de distribution.

Activités

1. Prendre un modèle entraîné sur un jeu de données (ex : CheXpert).
2. L'évaluer sur un jeu de données externe (ex : ChestX-ray14 de NIH) ou sur des données synthétiquement bruitées/transformées (ex : CheXphoto).
3. Quantifier la baisse de performance (AUC, F1-score).
4. Analyser si les estimations d'incertitude (obtenues via MC Dropout) permettent de détecter les échantillons OOD sur lesquels le modèle échoue.



# Chapitre 5

## Travaux Pratiques

### 5.1 Implémentation et Évaluation de l'Incertitude

Objectifs : Mettre en œuvre une méthode d'UQ pour visualiser et quantifier la confiance du modèle.

Activités :

1. Implémenter le MC Dropout sur un modèle CNN de segmentation (ex : U-Net).
2. Pour une image donnée, générer N prédictions et calculer des cartes de variance ou d'entropie pour visualiser l'incertitude au niveau du pixel.
3. Analyse : Mettre en corrélation l'incertitude prédite avec l'erreur de segmentation réelle (ex : Dice score) pour vérifier que les zones les plus incertaines sont bien celles où le modèle se trompe le plus.