



Rapport PFE 2025 IA/HPDA

Tiffany Gay-Bellile

Sarah Chaabouni

Encadré par

Yannick LE NIR

Karima EL GANAoui

Juan ANGEL LORENZO DEL CASTILLO

Mai, 2025

Contents

1 Contexte médical	2
1.1 L'imagerie médicale et son intérêt	2
2 Présentation générale du dataset	2
3 Les pathologies	3
4 Analyse exploratoire détaillée des données	5
4.1 Distribution des pathologies	5
4.2 Nombre de pathologies par image	6
4.3 Nombre d'images par patient	6
4.4 Répartition des positions radiologiques (AP / PA)	7
4.5 Co-occurrence entre pathologies	8
4.6 Répartition par âge et genre	9
5 Traitement des données	11
5.1 Déséquilibre de classe	11
5.2 Qualité de l'image	12
5.2.1 CLAHE	13
5.2.2 Région d'intérêt	13
6 Classification	14
6.1 Réseaux convolutifs	14
6.1.1 DenseNet121	14
6.1.2 EfficientNet	16
6.2 Transformers	17
6.3 Meilleur modèle	19
6.4 Explicabilité	19
6.4.1 Attention Rollout	20
6.4.2 Grad-CAM	20
7 Segmentation	23
8 Application	26
9 Axes d'amélioration	27
10 Conclusion	27
11 Références bibliographiques	28

1 Contexte médical

1.1 L'imagerie médicale et son intérêt

L'imagerie médicale est un domaine qui permet de visualiser l'intérieur du corps humain sans intervention chirurgicale. Elle est essentielle dans le diagnostic de maladies, le suivi de traitements et la détection précoce d'anomalies. Parmi les nombreuses techniques disponibles, on trouve par exemple la radiographie, l'échographie ou encore l'IRM.

Face à l'augmentation du volume d'images à analyser, l'intelligence artificielle devient un outil essentiel pour assister les radiologues. Elle permet d'accélérer le diagnostic, de détecter automatiquement certaines pathologies, et de réduire les erreurs humaines, tout en soutenant la prise de décision médicale.

2 Présentation générale du dataset

Le jeu de données utilisé dans ce projet est le **NIH ChestX-ray14**, une base publique de radiographies thoraciques développée par le *National Institutes of Health*. Il contient **112 120 images** correspondant à **30 805 patients**, soit une moyenne de **3,5 images par patient**.

Cependant, cette moyenne masque une grande variabilité : certains patients n'ont qu'une seule image, tandis que d'autres peuvent en avoir jusqu'à **185**, correspondant soit à des *suivis médicaux répétés* (sur plusieurs années), soit à des *séries d'imagerie prises dans un contexte aigu* (sur une même année).

Les informations associées à chaque image sont regroupées dans le fichier `Data_Entry_2017.csv`, structuré en plusieurs colonnes. Voici une description des champs principaux :

- **Image Index** : nom du fichier image, par exemple `0000001_000.png`. Il encode à la fois l'identifiant du patient et le numéro de l'image (de 0 à 184). Ainsi, le fichier `0000001_002.png` désigne la 3^e radiographie du patient 1.
- **Finding Labels** : une ou plusieurs pathologies présentes sur l'image, séparées par le symbole `|`. Par exemple : `Effusion|Infiltration`. Il s'agit donc d'un champ **multi-label**.
- **Follow-up #** : numéro de suivi, permettant d'identifier l'ordre de passage du patient. Ce champ est peu exploité ici, car les dates manquent.
- **Patient ID** : identifiant unique du patient, qui permet de regrouper les images par individu.
- **Patient Age** : âge estimé du patient au moment de la radiographie. Attention : certaines valeurs peuvent être aberrantes (patients très jeunes ou très âgés) et doivent être vérifiées.
- **Patient Gender** : sexe du patient (M pour homme ou F pour femme). Cette donnée est globalement équilibrée dans la base (environ 54 % d'hommes et 46 % de femmes).

- **View Position** : angle de prise de vue, principalement PA (Postéro-Antérieure) ou AP (Antéro-Postérieure). Ce champ est important car il influence la forme apparente des organes.
- **Image Width / Height** : taille de l'image d'origine, en pixels (souvent 1024×1024, mais pas toujours).
- **PixelSpacing X / Y** : résolution physique de l'image (en mm/pixel).

Image Index	Finding Labels	Follow-up #	Patient ID	Patient Age	Patient Gender	View Position	Image Width	Image Height	PixelSpacing X	PixelSpacing Y
0 00000001_000.png	Cardiomegaly	0	1	58	M	PA	2682	2749	0.143	0.143
1 00000001_001.png	Cardiomegaly Emphysema	1	1	58	M	PA	2894	2729	0.143	0.143
2 00000001_002.png	Cardiomegaly Effusion	2	1	58	M	PA	2500	2048	0.168	0.168
3 00000002_000.png	No Finding	0	2	81	M	PA	2500	2048	0.171	0.171
4 00000003_000.png	Hernia	0	3	81	F	PA	2582	2991	0.143	0.143

Figure 1: Data_Entry_2017.csv

3 Les pathologies

Ce fichier est la source principale des informations cliniques associées aux images. Une étape de nettoyage et de renommage des colonnes a été nécessaire, car certaines colonnes comme `OriginalImage[Width ou Height]` étaient mal nommées, et une colonne inutile (`Unnamed: 11`) a été supprimée.

Le dataset NIH ChestX-ray14 couvre 14 pathologies thoraciques couramment rencontrées dans la pratique médicale. Ces pathologies représentent un large éventail de cas cliniques allant des infections bénignes aux complications graves ou chroniques. Elles sont annotées dans le fichier `Data_Entry_2017.csv` à travers une colonne multi-label.

Voici la liste complète avec une brève définition médicale de chacune :

- **Atelectasis** : affaissement partiel ou complet d'un lobe pulmonaire.
- **Consolidation** : remplissage alvéolaire par du liquide ou du pus.
- **Infiltration** : opacité diffuse causée par un liquide inflammatoire ou infectieux.
- **Pneumonia** : infection des alvéoles pulmonaires.
- **Effusion** : épanchement pleural.
- **Edema** : œdème pulmonaire.
- **Emphysema** : destruction des parois alvéolaires.
- **Fibrosis** : cicatrisation pulmonaire chronique.
- **Pleural Thickening** : épaissement de la plèvre.
- **Pneumothorax** : présence d'air dans la cavité pleurale.
- **Cardiomegaly** : augmentation anormale de la taille du cœur.

- **Mass** : opacité arrondie supérieur à 3 cm.
- **Nodule** : opacité arrondie inférieur à 3 cm.
- **Hernia** : passage d'organes abdominaux dans le thorax.

Cette diversité de classes représente un grand intérêt pour l'entraînement de modèles de classification ou détection, mais elle entraîne aussi un fort déséquilibre de classes...

Voici quelques exemples de maladies sur images :

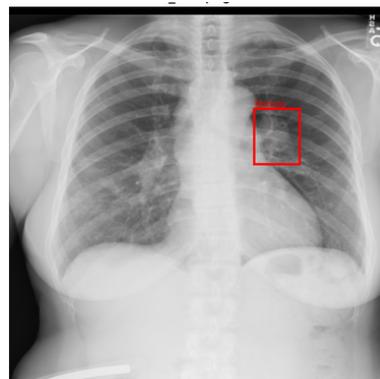


Figure 2: Image d'un Nodule

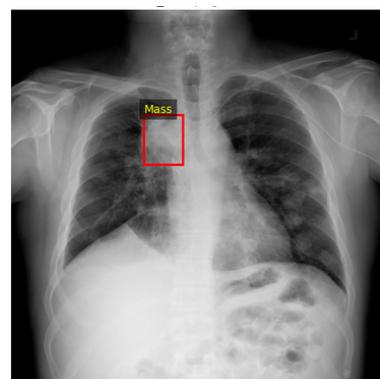


Figure 3: Image d'une Masse

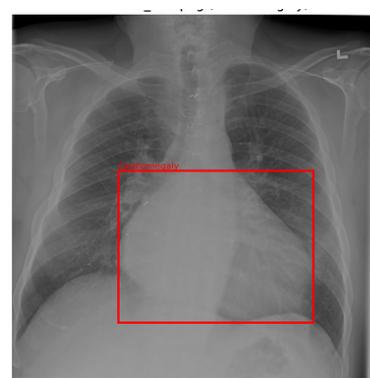


Figure 4: Image d'une Cardiomégalie

4 Analyse exploratoire détaillée des données

L'analyse exploratoire statistique du dataset **NIH ChestX-ray14** a permis de mettre en évidence plusieurs tendances structurantes. Voici l'interprétation des principaux graphiques présentés dans cette section.

4.1 Distribution des pathologies

Le premier graphique montre la fréquence de chaque pathologie présente dans la base de données. Il apparaît que plus de **50 % des images sont annotées comme “No Finding”**, c'est-à-dire qu'aucune pathologie n'a été détectée.

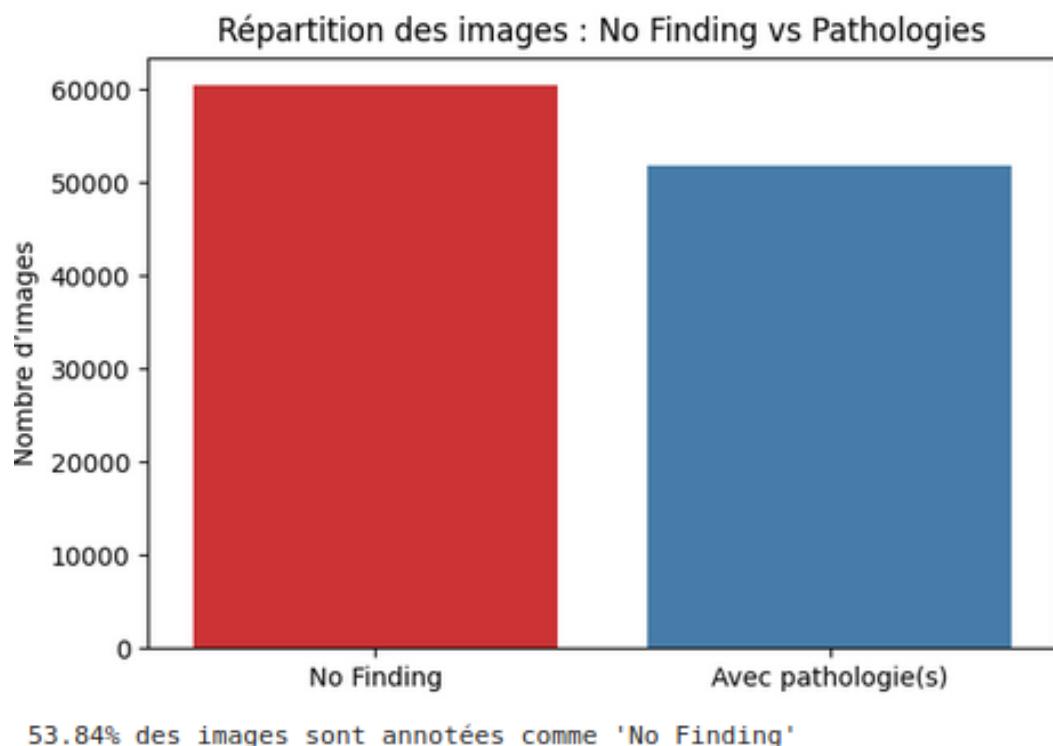


Figure 5: Répartition des images

Ce déséquilibre a un impact direct sur l'entraînement des modèles : un modèle peu performant pourrait atteindre une précision élevée en prédisant simplement “pas de maladie”, sans réellement apprendre à identifier les pathologies.

Parmi les pathologies annotées, les plus fréquentes sont :

- **Infiltration** : environ 18 %
- **Effusion** : environ 12 %
- **Atelectasis** : environ 10 %

Les moins fréquentes sont :

- **Fibrosis, Hernia et Emphysema**, toutes en dessous de 3 %

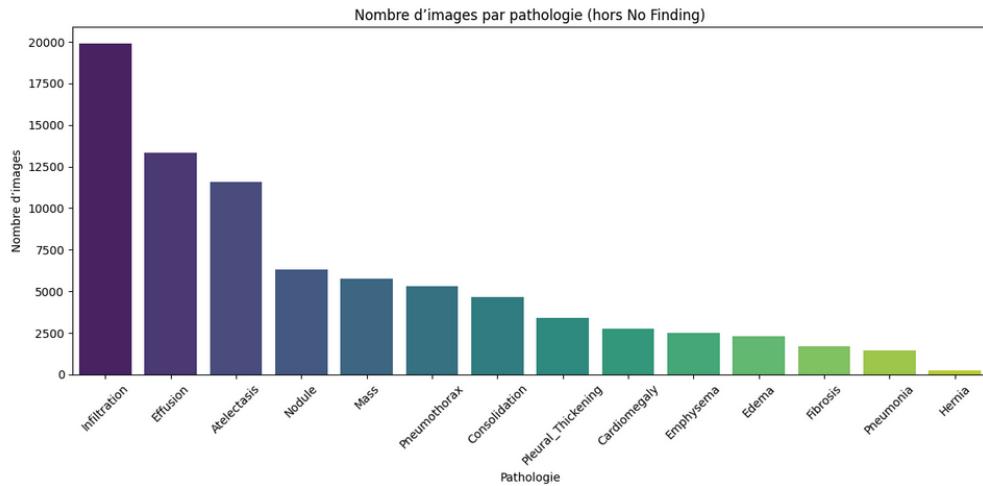


Figure 6: fréquences des pathologies

4.2 Nombre de pathologies par image

Ce graphique montre

Ce graphique indique combien de pathologies peuvent coexister sur une même image. On observe que :

- Une grande proportion des images (environ 50 %) n'ont **aucune pathologie annotée**.
- Environ **30 %** des images présentent une **seule pathologie**.
- Près de **20 %** en contiennent **deux ou plus**.

Ce constat confirme la nature **multi-label** du dataset : une image peut contenir simultanément, par exemple, un *épanchement pleural*, une *cardiomégalie* et une *pneumonie*.

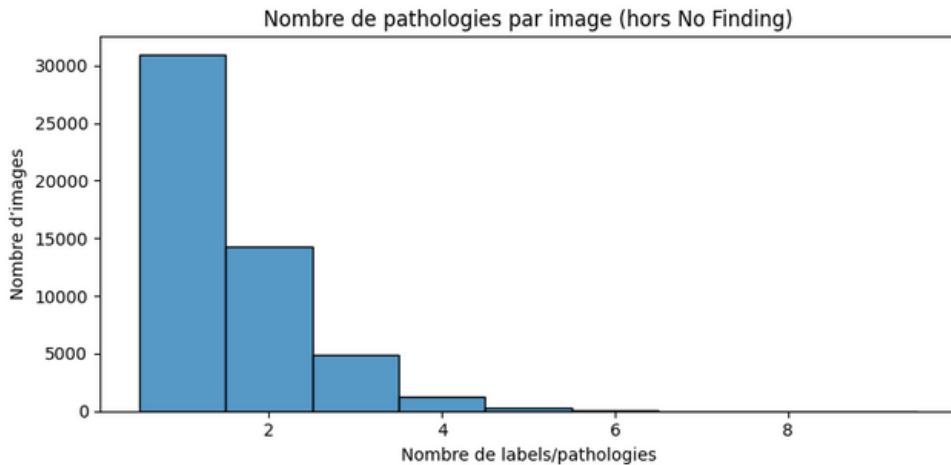


Figure 7: Nombre de pathologies par image

4.3 Nombre d'images par patient

Dans ce graphique, on analyse la distribution du nombre d'images par patient. Bien que la moyenne soit d'environ **3,5 images par patient**, la réalité est beaucoup plus hétérogène :

- Une majorité des patients ont **une seule image**.
- Certains patients possèdent **plusieurs dizaines** d'images.
- Un petit nombre de cas extrêmes vont jusqu'à **185 images pour un seul patient**.

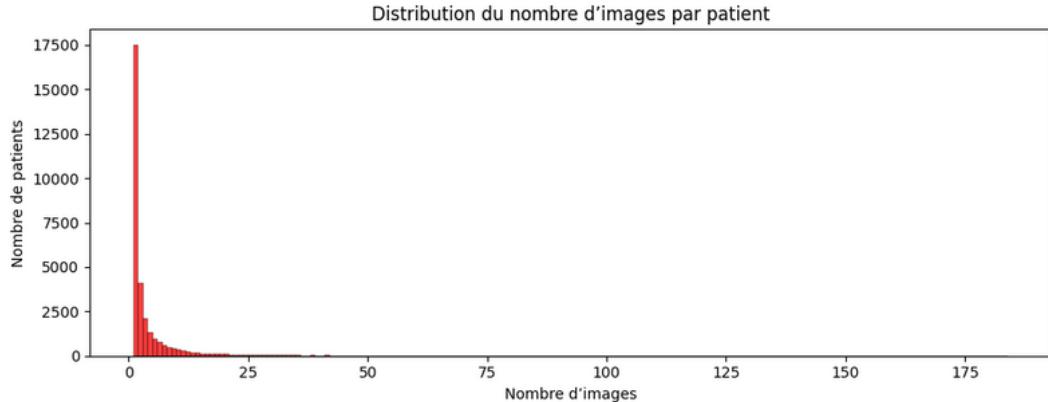


Figure 8: Nombre d'images par patient

4.4 Répartition des positions radiologiques (AP / PA)

Ce graphique représente les positions de prise de vue radiologique : **PA** (Postéro-Antérieure) et **AP** (Antéro-Postérieure). Ces deux types de vues ont des implications cliniques et techniques différentes.

PA (Postéro-Antérieure) Vue standard, utilisée lorsque le patient est debout face à la plaque. Cette vue est plus fiable car :

- le cœur est plus proche du capteur, donc **moins déformé**,
- les poumons sont **mieux étendus**,
- il y a **moins de distorsion anatomique**.

AP (Antéro-Postérieure) Vue utilisée lorsque le patient est alité (urgence, réanimation). Elle induit des artéfacts visuels :

- le cœur semble **plus gros**,
- les infiltrations peuvent être **plus floues**,
- la lecture est **plus difficile** si on ne tient pas compte du type de vue.

Dans notre dataset :

- **PA** représente environ **60 %** des images.
- **AP** représente les **40 %** restants.

Il est donc essentiel de **prendre en compte cette information** dans l'analyse ou l'entraînement des modèles, car le type de vue influe directement sur l'apparence anatomique des structures observées.

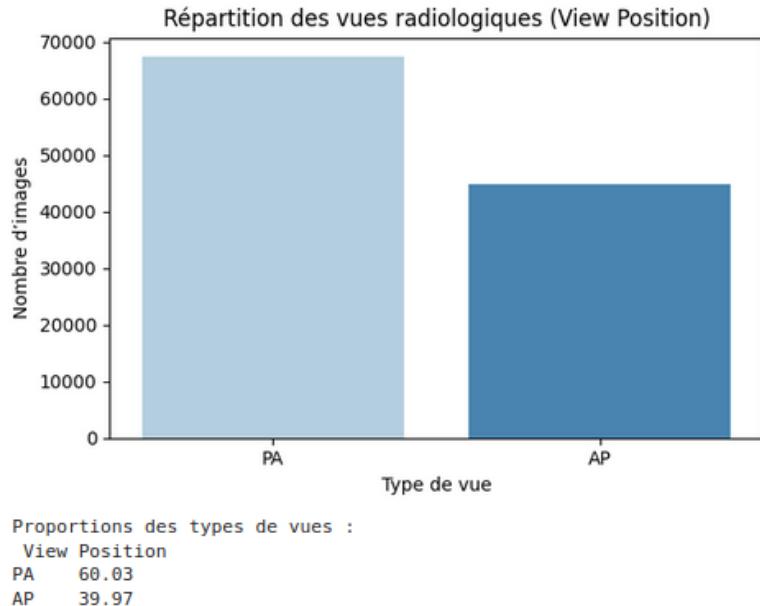


Figure 9: Répartition des positions radiologiques (AP / PA)

4.5 Co-occurrence entre pathologies

Nous avons produit une **matrice de co-occurrence** sur les colonnes binaires représentant la **probabilité qu'une pathologie B soit présente, sachant que la pathologie A l'est**.

Les corrélations observées sont généralement **faibles à modérées**, ce qui est cohérent avec la diversité des cas cliniques rencontrés. On note toutefois quelques associations notables :

- **Effusion → Atelectasis** : 0.58
- **Mass → Nodule** : 0.44
- **Pneumonia → Infiltration** : 0.79

Ces corrélations présentent un double intérêt :

- Elles confirment des **liens cliniquement établis**
- Elles permettent aux modèles de **tirer parti de dépendances structurelles** entre pathologies, ce qui peut améliorer la performance en classification multi-label.

Cependant, la **faible intensité globale des corrélations** montre que chaque pathologie garde une **signature distincte**, ce qui est favorable à une modélisation fine et indépendante de chaque classe.

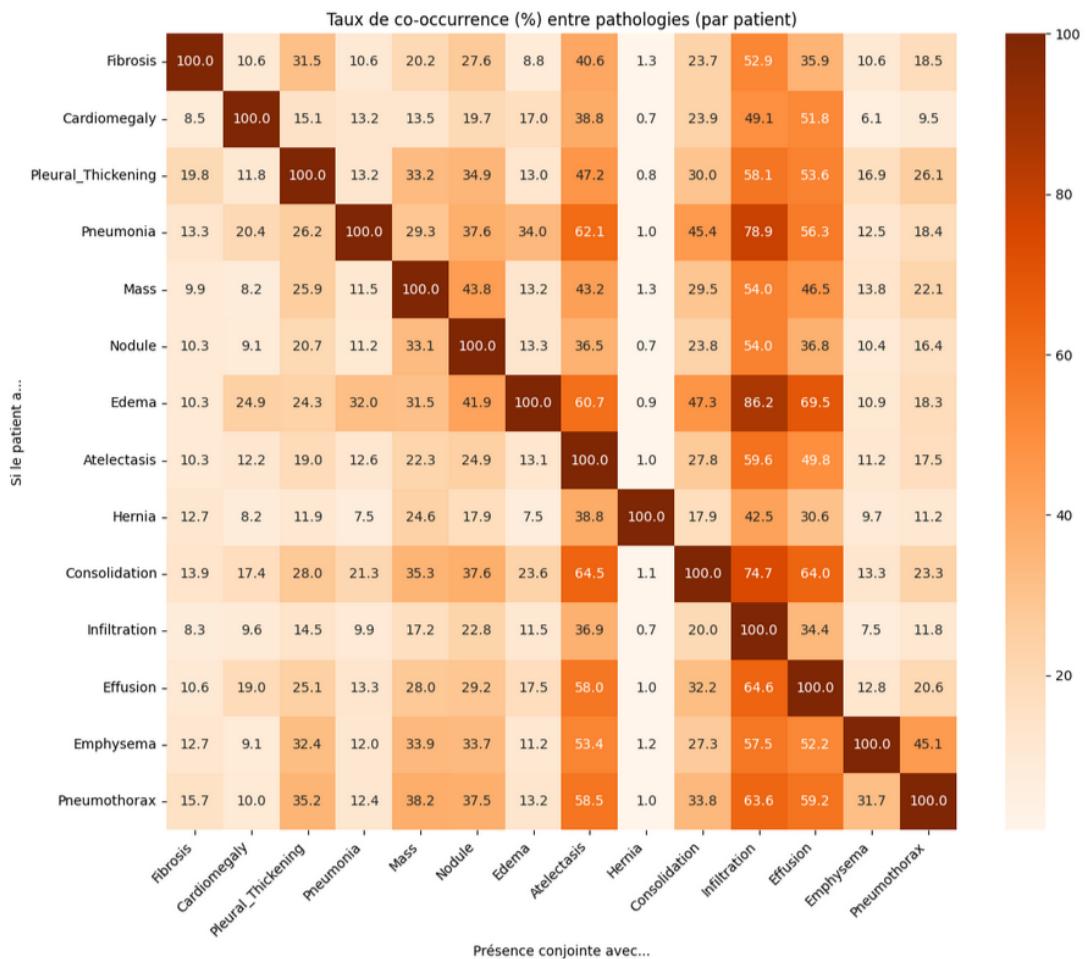


Figure 10: Matrice de corrélation entre les différentes pathologies

4.6 Répartition par âge et genre

L'âge des patients du dataset varie de quelques mois à plus de 90 ans. La majorité des cas se situe entre 40 et 70 ans, ce qui correspond à la population à risque pour de nombreuses pathologies pulmonaires chroniques.

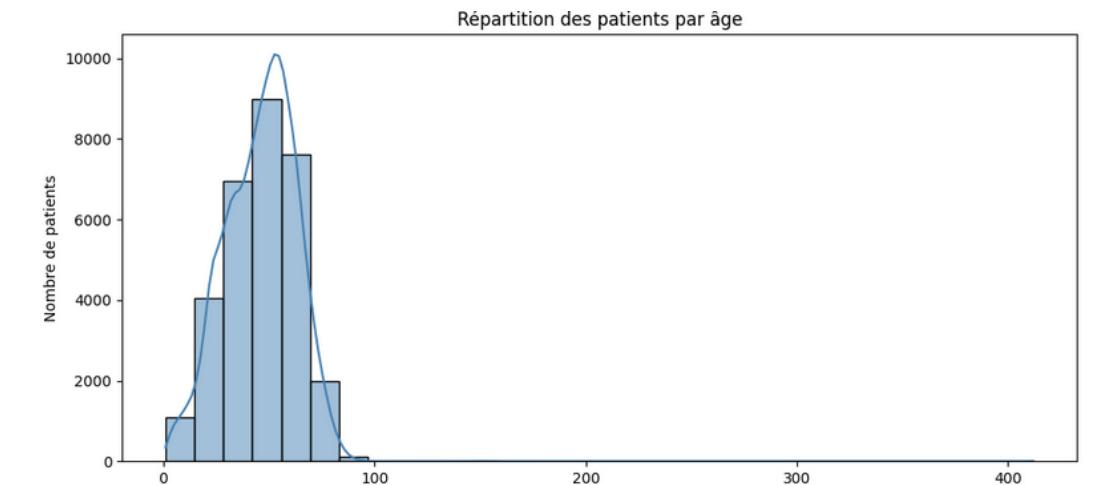


Figure 11: Répartition par âge

L'étude croisée entre l'âge des patients et la présence de maladies montre globalement un risque

d'être malade qui augmente avec l'âge :

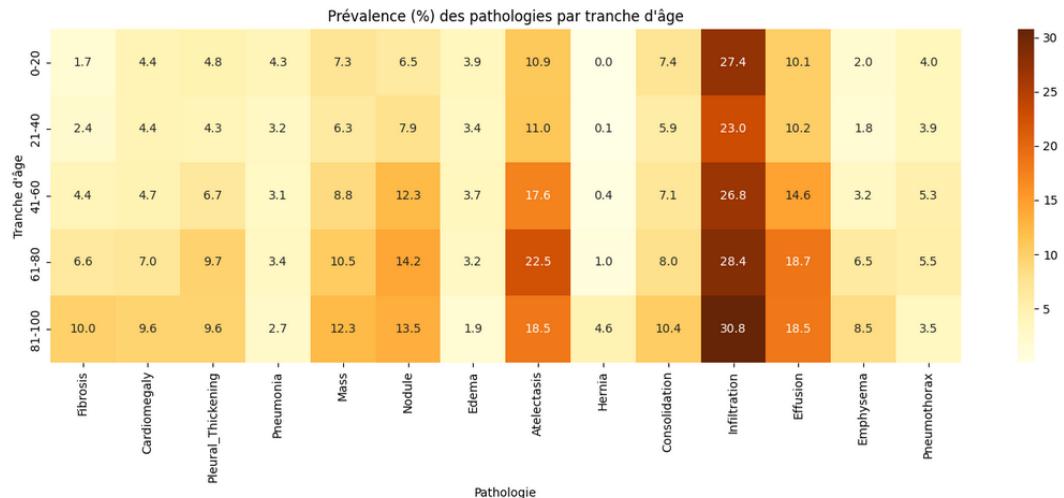


Figure 12: Prévalence des pathologies par tranche d'âge

Dans ce dataset la répartition des hommes et des femmes est à peu près équivalente :

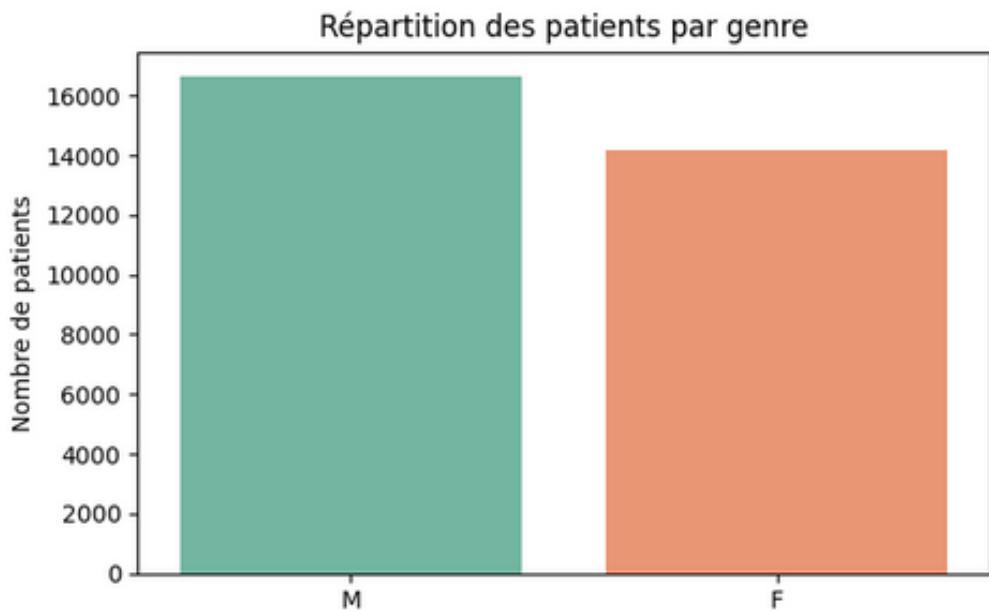


Figure 13: Répartition par genre

On observe que les hommes sont légèrement plus malades que les femmes :

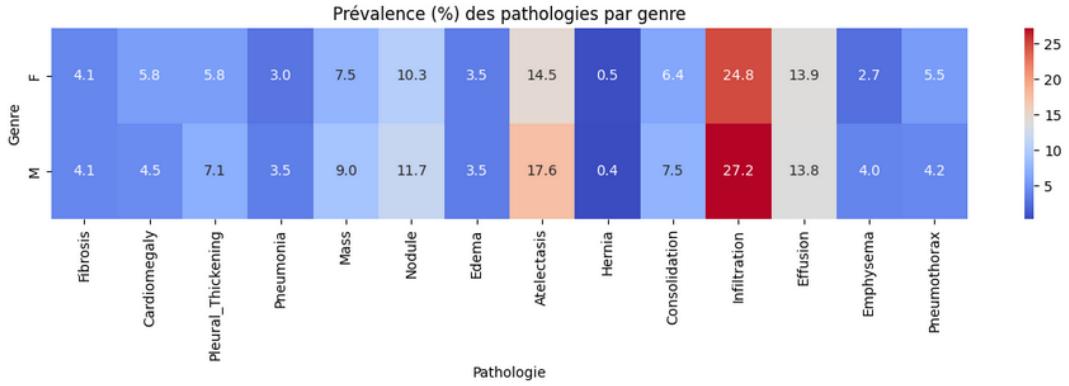


Figure 14: Prévalence des pathologies par genre

5 Traitement des données

Le prétraitement des données constitue une étape cruciale dans tout projet de deep learning. Cela permet à la fois d'améliorer la qualité des données en entrée, de rendre les modèles plus robustes et d'assurer une meilleure généralisation. Afin de trouver les meilleures transformations et traitements, nous avons réalisé des tests sur une partie des données (Atelectasis PA) avant de les appliquer à l'ensemble du dataset. L'idée était de tester rapidement ce qui fonctionne le mieux, sans perdre de temps ni surcharger le GPU durant les phases d'entraînement.

5.1 Déséquilibre de classe

Le dataset utilisé comporte 15 classes, dont 14 pathologies et une classe supplémentaire "No Finding". Cette dernière est de loin la plus représentée, avec plus de 60 000 images, soit environ 54% de l'ensemble des données. Ce déséquilibre est un premier défi, car un modèle non régularisé aura tendance à prédire cette classe majoritaire, au détriment de la détection des pathologies.

Lorsque l'on se concentre uniquement sur les 14 pathologies, un second niveau de déséquilibre apparaît. La pathologie la plus fréquente, l'Infiltration, représente 17.7% des images, alors que des pathologies comme la Fibrose ou la Hernie représentent moins de 1% du dataset (respectivement 0.6% et 0.2%). Ce type de déséquilibre est particulièrement problématique en médecine, car il peut amener le modèle à négliger des maladies rares mais potentiellement graves.

Pour atténuer ces effets, nous avons eu recours à de la data augmentation, en appliquant diverses transformations (flips, rotations, zooms) sur les classes minoritaires afin d'augmenter artificiellement leur représentativité. Toutefois, cette méthode a ses limites : elle ne permet pas de créer une vraie diversité de données cliniques, et ne peut compenser un déséquilibre aussi important.

Afin de renforcer la robustesse de l'entraînement et de limiter les biais introduits par des classes ultra-minoritaires, nous avons pris la décision de supprimer 7 pathologies parmi les moins représentées dans le dataset :

Pleural Thickening, Cardiomegaly, Emphysema, Edema, Pneumonia, Fibrosis, Hernia.

Pathologie	Nombre d'images	Pourcentage approx.
No Finding	60 361	53.8 %
Infiltration	19 894	17.7 %
Effusion	13 317	11.9 %
Atelectasis	11 278	10.0 %
Nodule	6 331	5.6 %
Pneumothorax	5 331	4.8 %
Mass	5 318	4.7 %
Consolidation	4 667	4.2 %
Pleural Thickening	3 385	3.0 %
Cardiomegaly	2 775	2.5 %
Emphysema	2 516	2.2 %
Edema	2 303	2.1 %
Pneumonia	1 431	1.3 %
Fibrosis	727	0.6 %
Hernia	227	0.2 %

Figure 15: Répartition des maladies

Cette réduction du nombre de classes permet de recentrer l'apprentissage sur les pathologies les plus représentées, améliorant ainsi la stabilité du modèle et la fiabilité des prédictions sur les cas les plus courants.

Malgré l'exclusion des pathologies les moins représentées et l'application de techniques de data augmentation, le déséquilibre entre les classes restantes reste significatif. Certaines pathologies continuent de dominer le dataset (par exemple, Infiltration ou Effusion), tandis que d'autres, comme Mass ou Pneumothorax, restent peu représentées.

Afin d'atténuer ce biais structurel et d'éviter que le modèle n'apprenne à sur-prédire les classes majoritaires, nous avons mis en œuvre une stratégie de sous-échantillonnage. Cela consiste à réduire volontairement le nombre d'images dans les classes les plus fréquentes, de manière à rééquilibrer approximativement la distribution globale des classes. Cette méthode est appliquée uniquement pour les "No Finding" car toujours sur-représentés.

5.2 Qualité de l'image

La qualité des images médicales est un facteur déterminant dans la performance d'un modèle de classification. En effet, des images floues, à faible contraste ou contenant des zones non pertinentes peuvent nuire à la capacité du modèle à apprendre des caractéristiques discriminantes.

C'est particulièrement vrai pour les radiographies thoraciques, où des détails fins peuvent être essentiels pour identifier certaines pathologies.

5.2.1 CLAHE

Afin d'améliorer la lisibilité des structures internes sur les radiographies, nous avons appliqué la méthode CLAHE (Contrast Limited Adaptive Histogram Equalization). CLAHE agit localement sur des sous-régions de l'image (appelées "tiles"), ce qui permet de renforcer le contraste dans les zones peu visibles tout en limitant l'amplification du bruit.

Cette opération est particulièrement utile pour mettre en évidence des anomalies subtiles, comme l'atelectasie ou l'infiltration, qui peuvent être difficilement perceptibles sur une image brute.

5.2.2 Région d'intérêt

Une radiographie thoracique contient souvent des éléments hors du champ d'intérêt pour le diagnostic, comme les épaules, les bras, ou le texte annoté. Pour améliorer la pertinence de l'information visuelle, nous avons mis en œuvre un algorithme de détection de la région d'intérêt (ROI) afin d'extraire uniquement la zone pulmonaire.

Cette étape permet :

- De réduire le bruit visuel en supprimant les zones non informatives.
- De centrer l'attention du modèle sur les parties réellement diagnostiques.



Figure 16: Avant transformations

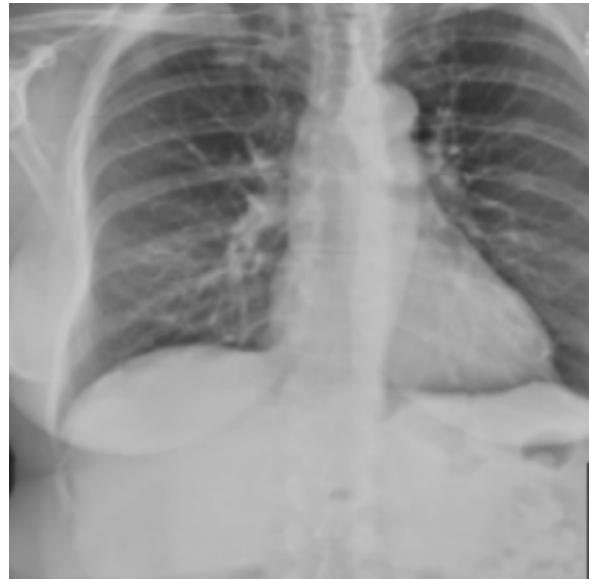


Figure 17: Après transformations

6 Classification

La classification est une tâche fondamentale en apprentissage automatique supervisé, qui consiste à attribuer une ou plusieurs étiquettes à une image en fonction de son contenu. Dans le cadre de l'imagerie médicale, il s'agit de prédire la présence ou l'absence de certaines pathologies à partir d'images radiographiques. Cependant, elle ne permet pas de savoir où se trouve la maladie.

Comme pour le pré-traitement des données, afin de gagner du temps et préserver nos ressources, nous avons réalisé des tests sur le même échantillons du dataset que précédemment. Les résultats qui vont suivre dans cette partie sont uniquement sur cet échantillon.

6.1 Réseaux convolutifs

Les Convolutional Neural Networks (CNN) sont les modèles les plus utilisés pour l'analyse d'images. Ils reposent sur l'utilisation de filtres convolutifs capables d'extraire automatiquement des motifs visuels à différentes échelles (bords, textures, formes...).

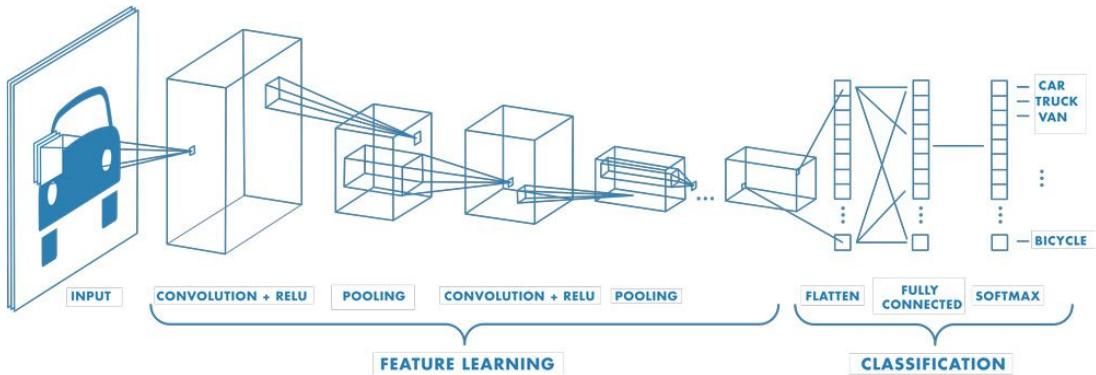


Figure 18: Structure d'un réseau convolutif

6.1.1 DenseNet121

DenseNet (Densely Connected Convolutional Networks) est une architecture de réseau de neurones convolutifs introduite en 2017. Elle établit des connexions directes entre toutes les couches. Ainsi, chaque couche reçoit comme entrée les features cumulées de toutes les couches précédentes, ce qui :

- Renforce la propagation de l'information et des gradients
- Encourage une réutilisation des caractéristiques
- Réduit le nombre de paramètres par rapport à des architectures profondes classiques

La version DenseNet121 désigne une architecture composée de 121 couches. Elle est particulièrement populaire car elle offre un bon équilibre entre performance et efficacité. Elle est moins lourde que d'autres CNN comme ResNet50 ou VGG16 et est très performante en classification d'images médicales. Elle est capable de capturer des détails fins, ce qui est essentiel dans l'analyse de radiographies.

Nous avons entraîné notre DenseNet121 sur 2210 Atelectasis et 2210 No Finding. Le modèle a montré des résultats contrastés sur les différentes métriques. Il atteint un recall très élevé de 0.9053, ce qui signifie qu'il a détecté la grande majorité des cas positifs dans le jeu de validation. Cela indique une excellente sensibilité, essentielle dans un contexte médical où il est important de ne pas passer à côté d'une pathologie.

Le F1-score de 0.7569 confirme que le modèle atteint un bon compromis entre rappel et précision. En revanche, la précision reste modérée (0.6503), traduisant une tendance à générer des faux positifs.

L'accuracy de 0.7014 est correcte, mais elle doit être interprétée avec prudence, en prenant en compte les autres métriques.

Enfin, l'analyse des courbes de perte révèle un surapprentissage clair. Tandis que la loss d'entraînement diminue régulièrement, la validation loss augmente fortement dès les premières époques, pour dépasser 1.6 à l'époque 10. Cela montre que le modèle s'adapte trop au jeu d'entraînement et perd en généralisation sur des données nouvelles.

Métrique	Valeur
Accuracy	0.7014
Précision	0.6503
Recall	0.9053
F1-score	0.7569

Figure 19: Métriques DensetNet121

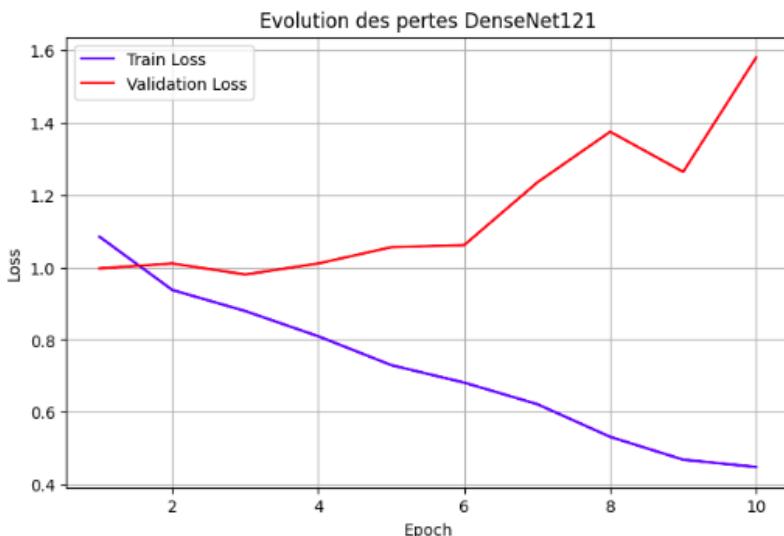


Figure 20: Courbes des pertes DensetNet121

6.1.2 EfficientNet

EfficientNet est une famille d'architectures de réseaux de neurones convolutifs introduite par Google en 2019. Son innovation principale réside dans l'utilisation d'une méthode automatique de recherche d'architecture (AutoML), suivie d'un principe de mise à l'échelle uniforme appelé compound scaling.

La version EfficientNet-B0 est la version de base. Elle sert de point de départ pour les autres variantes plus grandes.

Elle est idéale pour :

- Des projets nécessitant rapidité et légèreté
- Des données limitées
- Une utilisation sur GPU modeste

EfficientNet-B0 a très peu de paramètres, environ 5 millions. Ce modèle est un bon compromis entre précision et temps d'entraînement.

EfficientNet-B2 est une version légèrement plus grande que B0, avec plus de couches, plus de largeur, une résolution d'entrée plus élevée (260×260 au lieu de 224×224). Elle permet une meilleure capture des détails visuels et une meilleure précision sur des anomalies fines, au prix d'un temps de calcul légèrement supérieur.

Le modèle EfficientNet-B0 présente des résultats équilibrés et cohérents. Il atteint un F1-score de 0.7492, proche de celui de DenseNet121, avec une meilleure précision (0.7101). Cela signifie que le modèle fait moins de fausses alertes que DenseNet121, avec des prédictions plus fiables lorsqu'il identifie une pathologie.

Le recall reste bon (0.7930), bien qu'un peu en retrait par rapport à DenseNet121, ce qui montre qu'il détecte la majorité des cas positifs, tout en réduisant les faux positifs. L'accuracy de 0.7274 est également solide.

En ce qui concerne la courbe de loss, on observe une diminution régulière de la perte d'entraînement, tandis que la validation loss reste globalement stable jusqu'à l'époque 3-4, avant de commencer à augmenter. Cette dynamique suggère un léger surapprentissage, mais moins marqué que chez DenseNet121. Le modèle généralise mieux, ce qui le rend fiable pour une utilisation équilibrée entre sensibilité et précision.

EfficientNet-B2 affiche des résultats très élevés en précision (0.7643) et accuracy (0.7676), ce qui en fait le modèle le plus fiable pour minimiser les faux positifs. Il est donc très performant lorsqu'il s'agit de confirmer avec confiance qu'une pathologie est bien présente.

En revanche, le recall est plus faible (0.7070), indiquant que le modèle a laissé passer davantage de cas positifs non détectés par rapport aux autres modèles. Cela se reflète dans un F1-score légèrement inférieur (0.7346), traduisant une légère perte d'équilibre entre sensibilité et précision.

Du côté des pertes, on constate que la train loss diminue régulièrement, mais la validation loss augmente progressivement dès les premières époques, avec un écart de plus en plus marqué. Ce comportement suggère un surapprentissage plus net que celui observé avec EfficientNet-B0. Le modèle s'adapte très bien aux données d'entraînement, mais généralise moins bien.

Métrique	EfficientNet-B0	EfficientNet-B2
Accuracy	0.7274	0.7676
Précision	0.7101	0.7643
Recall	0.7930	0.7070
F1-score	0.7492	0.7346

Figure 21: Métriques EfficientNet

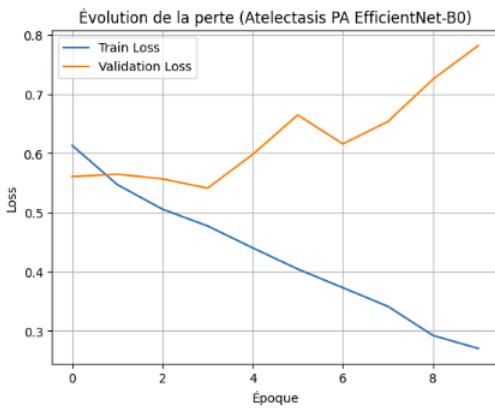


Figure 22: Courbes des pertes B0

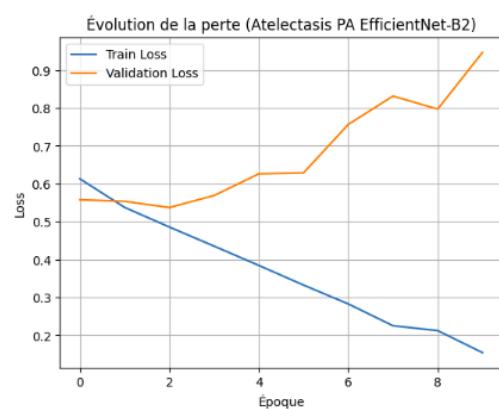


Figure 23: Courbes des pertes B2

6.2 Transformers

Le Vision Transformer (ViT) est un modèle d'apprentissage profond introduit par Google en 2020, qui applique les principes des transformers, initialement conçus pour le traitement du langage naturel, au domaine de la vision par ordinateur. Contrairement aux CNN qui utilisent des convolutions locales, ViT divise l'image en petits patchs, les transforme en vecteurs, et les traite comme une séquence, à l'aide d'un mécanisme d'attention globale.

Cette architecture permet au modèle de :

- Capturer des relations longues distance entre les parties de l'image
- Apprendre des représentations globales sans biais inductif spatial comme les CNN
- Être plus flexible et potentiellement plus performant sur des tâches où le contexte global est important

Cependant, ViT nécessite en général plus de données pour s'entraîner efficacement, car il ne bénéficie pas de la structure inductive des CNN.

Le modèle ViT atteint un recall de 0.8975, l'un des plus élevés parmi tous les modèles testés. Cela signifie qu'il a su détecter presque tous les cas positifs, ce qui en fait un modèle très sensible, capable d'identifier la majorité des pathologies présentes dans les images.

Cependant, la précision reste limitée (0.5561), indiquant que le modèle a également généré beaucoup de faux positifs. Il prédit donc souvent la présence d'une pathologie, même quand il n'y en a pas, ce qui peut conduire à des alertes inutiles.

Le F1-score atteint 0.6867, ce qui est raisonnable compte tenu de l'écart entre précision et rappel. Cela montre que le modèle compense sa précision plus faible par une très bonne capacité à détecter les vrais cas positifs.

En ce qui concerne la courbe de loss, le train loss et la validation loss diminuent de façon régulière et restent proches tout au long de l'entraînement. Ce comportement indique que le modèle généralise bien et n'est pas en surapprentissage. Cette stabilité est un point fort important, surtout comparé à d'autres modèles qui présentent une perte de validation en hausse rapide.

Métrique	Valeur
Loss	0.6711
Val Loss	0.6807
Accuracy	0.6316
Précision	0.5561
Recall	0.8975
F1-score	0.6867

Figure 24: Métriques Vit

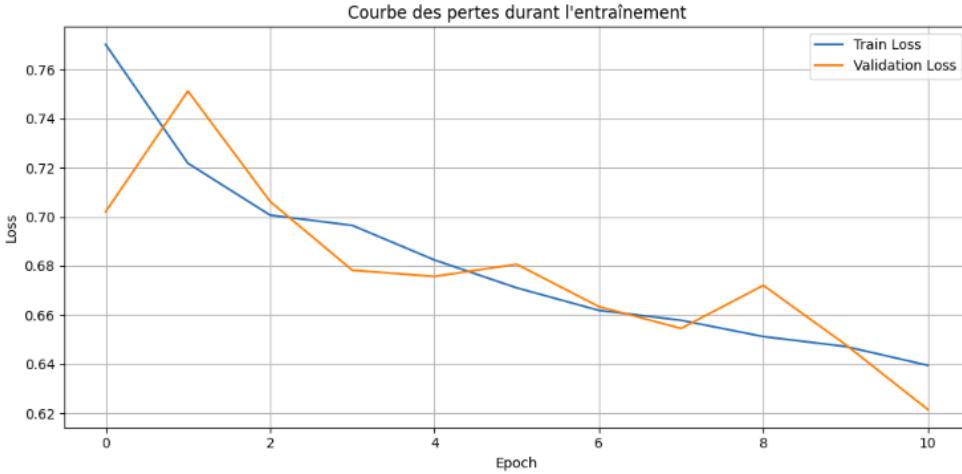


Figure 25: Courbes des pertes Vit

6.3 Meilleur modèle

Chaque modèle a ses points forts et points faibles. Les résultats obtenus montrent que chaque modèle présente des avantages spécifiques, et qu'aucun ne surpassé les autres sur tous les critères. Le choix du modèle optimal dépend donc fortement de la priorité clinique ou du contexte d'usage.

Si l'objectif est de maximiser la détection des pathologies, c'est-à-dire de minimiser le risque de passer à côté d'un cas positif, alors DenseNet121 ou ViT sont les modèles les plus adaptés. Tous deux affichent un rappel supérieur à 89%, ce qui en fait des candidats solides pour des systèmes d'aide au diagnostic où la sensibilité est prioritaire. Cependant, DenseNet121 a un surapprentissage important. ViT serait donc à privilier sur ce point.

En revanche, si l'objectif est de limiter les faux positifs, alors EfficientNet-B2 est le meilleur choix. Il offre la meilleure précision (0.7643) et la meilleure accuracy (0.7676), ce qui en fait un modèle fiable pour confirmer un diagnostic avec certitude. Cependant, son rappel plus faible (0.7070) implique qu'il peut laisser passer certains cas positifs.

Pour une utilisation plus équilibrée, où l'on cherche à combiner bonne détection et fiabilité dans les prédictions, EfficientNet-B0 se distingue comme le modèle le plus stable. Il affiche un F1-score élevé, une précision correcte, et montre moins de surapprentissage que les autres modèles. Il constitue donc un bon compromis général entre sensibilité, précision et robustesse.

6.4 Explicabilité

Dans un contexte médical, il est essentiel de comprendre sur quelles régions de l'image le modèle s'appuie pour poser son diagnostic. Pour cela, des méthodes d'explicabilité visuelle sont utilisées afin de rendre le processus décisionnel des modèles de deep learning plus transparent et interprétable.

6.4.1 Attention Rollout

Pour le modèle ViT, nous avons appliqué la méthode Attention Rollout, qui exploite le mécanisme d’attention multi-têtes des Transformers. Cette méthode permet de propager les poids d’attention à travers toutes les couches du ViT, afin de produire une carte d’attention globale indiquant quelles zones ont le plus influencé la décision finale.

La visualisation obtenue montre une attention répartie sur toute l’image, avec des zones colorées de manière plus diffuse et pas nettes. Nous n’avons pas une zone en particulier qui nous explique la sortie du modèle. Ce qui est un problème. Nous avons des pixels qui n’ont rien à voir entre eux. Ceci empêche de déterminer ce qui a influencé notre modèle.

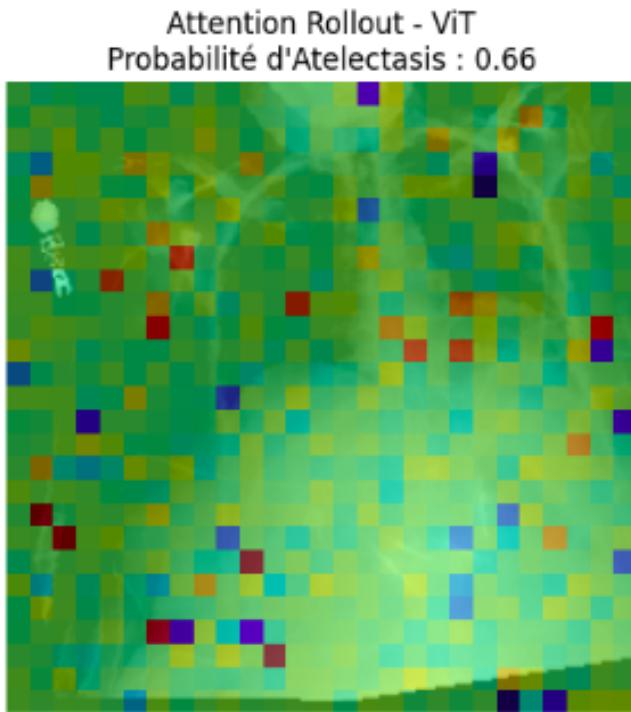


Figure 26: Attention Rollout

6.4.2 Grad-CAM

Pour le modèle EfficientNet-B0, une architecture de type CNN, nous avons utilisé la méthode Grad-CAM (Gradient-weighted Class Activation Mapping). Cette technique repose sur les dernières couches convolutionnelles du réseau, et calcule des poids dérivés des gradients pour générer une carte de chaleur (heatmap) superposée à l’image.

Sur l’image issue d’EfficientNet-B0, la heatmap est bien localisée, concentrée sur les zones pulmonaires. Cela montre que le modèle s’appuie sur des régions qui semblent cohérentes, ce qui renforce la fiabilité de son interprétation et suggère qu’il ne s’appuie pas sur des zones non pertinentes.

Malgré les limites de Grad-CAM pour les Transformers, nous avons voulu utiliser cette méthode pour notre modèle ViT. Contrairement à EfficientNet-B0, la carte obtenue avec Grad-CAM sur

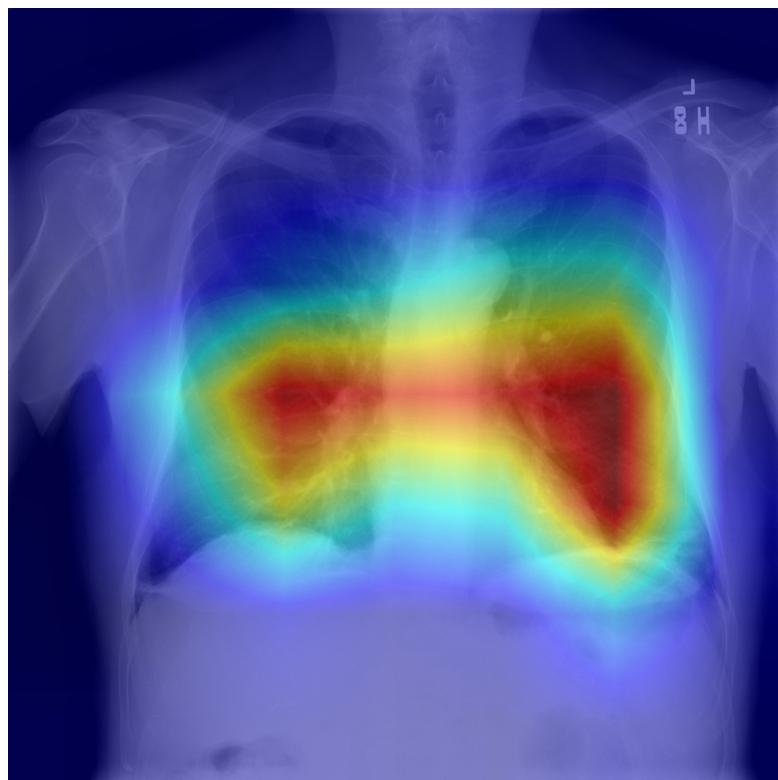


Figure 27: Grad-CAM EfficientNet-B0

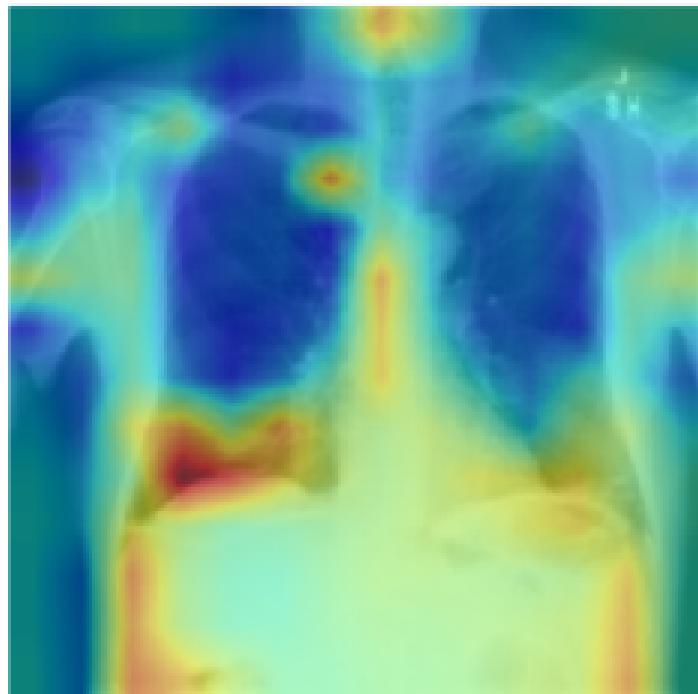


Figure 28: Grad-CAM ViT

ViT est moins précise. Cette comparaison montre que, dans notre cas, EfficientNet-B0 offre une explicabilité plus utile et plus exploitable en contexte médical.

7 Segmentation

Dans le cadre de notre projet, nous avons souhaité expérimenter une approche de segmentation d’images médicales, consistant à prédire un masque binaire indiquant, pixel par pixel, les zones pathologiques sur une radiographie thoracique.

Notre première intuition a été d’exploiter le fichier `BBox_List_2017.csv`, qui contient des annotations sous forme de *bounding boxes* pour certaines images. L’idée était de transformer ces *bounding boxes* en masques pour entraîner un modèle de segmentation supervisée.

Cependant, seules 900 images sont annotées avec des *bounding boxes*, soit une infime fraction des 112 000 images disponibles. De plus, ces boîtes englobent souvent du tissu sain, ce qui nuit à la qualité du masque pour l’apprentissage. Nous avons tout de même essayé de les transformer en masque mais les résultats n’ont pas été concluants :

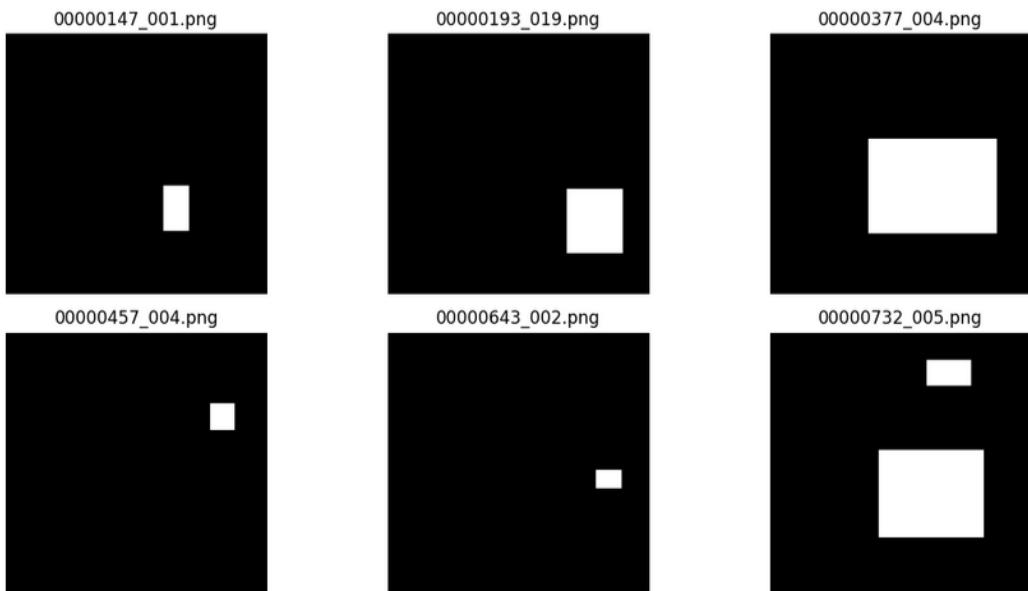


Figure 29: BBox en masques

Nous avons donc envisagé une alternative : l’annotation manuelle précise de certaines masses visibles. Nous avons sélectionné 25 images annotées comme “Mass” dans le fichier `BBox_List_2017.csv` et les avons segmentées manuellement à l’aide de l’outil *VGG Image Annotator (VIA)*.

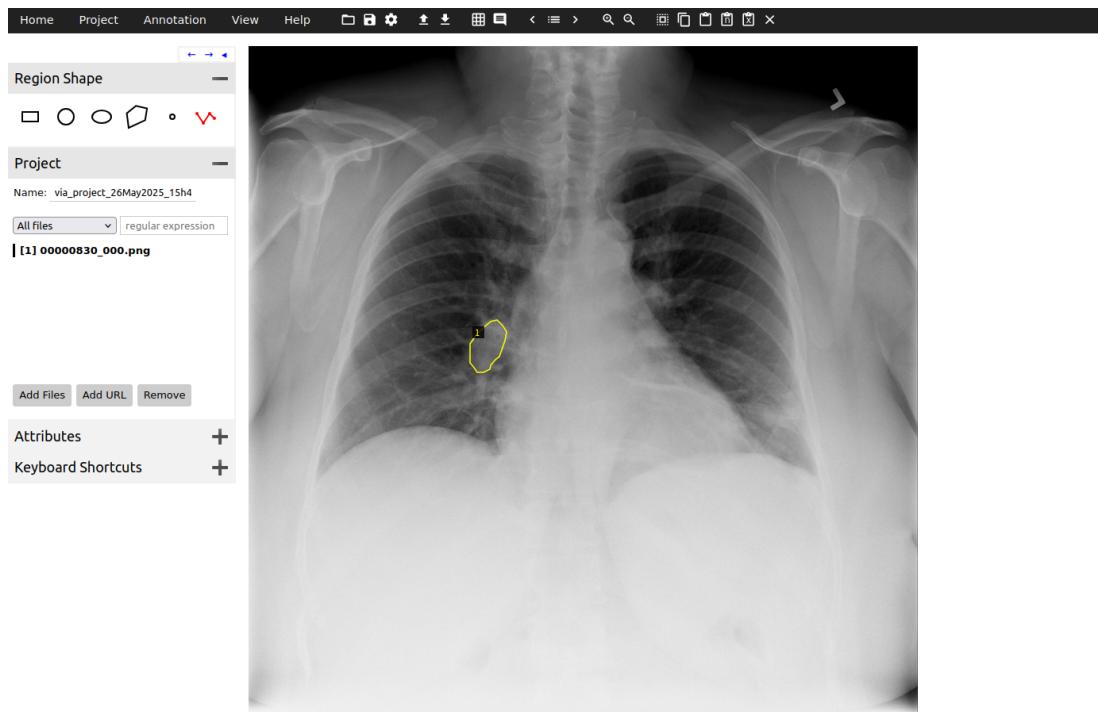


Figure 30: Exemple d'image segmentée avec VIA

Nous avons ensuite transformés ces fichiers JSON en masques :

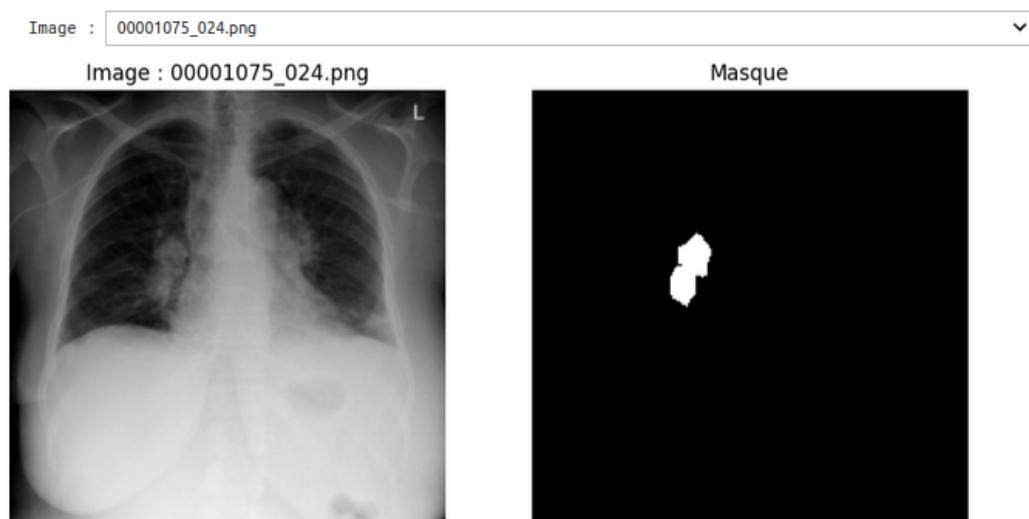


Figure 31: Exemple d'image segmentée transformée en masque

Ces masques ont été exportés en PNG pour l'entraînement d'un U-Net. Toutefois, les résultats n'ont pas été satisfaisants en raison du faible nombre d'exemple.

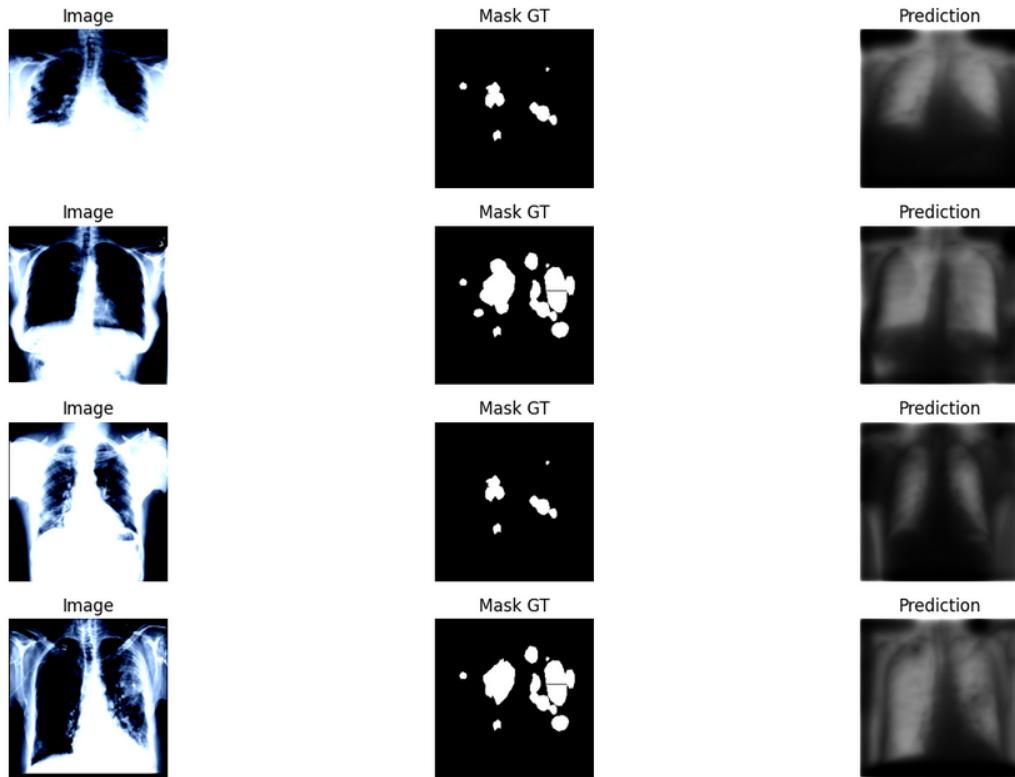


Figure 32: Résultats Unet

Nous avons donc mis cette approche en pause, en conservant les outils pour un usage ultérieur si notre base d'annotations venait à s'enrichir.

8 Application

Afin de conclure nos travaux, nous avons créé une application qui utilise nos modèles de prédiction sur des images qu'un utilisateur peut charger.



Figure 33: Site 1

L'utilisateur peut sélectionner une ou plusieurs maladies à détecter. Un modèle sera lancé pour chacune.



Figure 34: Site 2

La première ligne d'images explicatives sont réalisées avec le modèle EfficientNet-B0. Les probabilités de maladies sont affichées avec les endroits sur lesquels il s'est appuyé.

La dernière ligne utilise ViT uniquement pour la maladie Atelectasis.

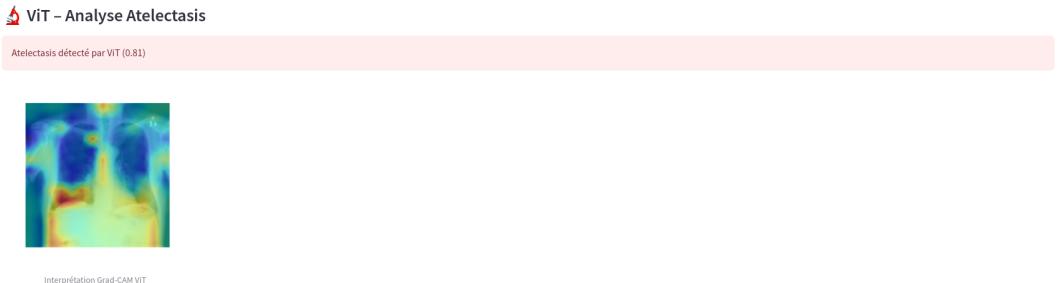


Figure 35: Site 3

9 Axes d'amélioration

L'étude du dataset NIH ChestX-ray14 a mis en évidence plusieurs limites qui restreignent la portée des analyses et des modèles développés. Les axes d'amélioration principaux sont les suivants :

- **Ajout des dates de prise d'image** : L'absence de temporalité empêche de distinguer une évolution réelle d'une pathologie dans le temps chez un patient. Cette information est cruciale pour différencier un suivi médical d'une simple variation d'angle ou de visibilité.
- **Enrichissement en bounding boxes** : Le nombre très restreint de radiographies annotées limite fortement les performances des modèles de segmentation. Une extension du fichier `BBox_List_2017.csv` ou la création de nouveaux masques manuels améliorerait considérablement l'apprentissage.
- **Ajout de métadonnées cliniques** : Des informations comme les symptômes, le contexte clinique ou les antécédents médicaux permettraient une analyse exploratoire plus fine et la construction de modèles multimodaux plus puissants.

10 Conclusion

Ce projet a mis en évidence que l'intelligence artificielle peut constituer une aide précieuse pour les médecins, en apportant à la fois un gain de temps et une meilleure précision dans la détection des anomalies. Elle permet de repérer des signes subtils qui pourraient échapper à l'œil humain. Cependant, l'IA reste un outil d'assistance et ne doit pas remplacer un avis médical.

Nous avons également observé que deux modèles peuvent donner la même prédiction, tout en s'appuyant sur des zones d'attention différentes. Cela montre la nécessité de rester critique face aux résultats générés et de ne pas prendre une décision médicale uniquement sur les sorties d'un modèle. L'interprétation humaine reste donc indispensable pour garantir la fiabilité du diagnostic.

11 Références bibliographiques

- *The Ultimate Guide to Preprocessing Medical Images* – AI Education, Radiography, CRO, Academia
- Ultimate Guide for Machine Learning in Medical Imaging – GitHub Repository
- Imagerie thoracique – Manuel MSD, édition professionnelle
- Institut de Radiologie de Paris. (n.d.). *Imagerie thoracique*.
<https://www.institut-de-radiologie.com/thorax/>
- Wikipédia. (n.d.). *Imagerie par résonance magnétique*.
https://fr.wikipedia.org/wiki/Imagerie_par_r%C3%A9sonance_magn%C3%A9tique
- Wikipédia. (n.d.). *Tumeur de Pancoast-Tobias*.
https://fr.wikipedia.org/wiki/Tumeur_de_Pancoast-Tobias
- Radiopaedia.org. (n.d.). *The Radiology Resource*.
<https://radiopaedia.org>
- Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). *ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases*. Kaggle.
<https://www.kaggle.com/datasets/nih-chest-xrays/data>