



# Rapport de Projet

## 5eme année Ingénierie Informatique et Réseaux

Sous le thème

---

Détection des maladies pulmonaires par  
Deep Learning

**Réaliser par :**

AIT EL QADI Anas

BELKACEM Mouad



# Dédicace

À nos familles et nos amis qui nous ont toujours soutenus. Ce travail est le résultat de nos efforts communs et de notre envie d'apprendre. Nous dédions ce projet à tous ceux qui croient en l'utilité de la technologie pour améliorer la santé et aider les gens.

# Remerciements

Nous souhaitons exprimer notre profonde gratitude à tous ceux qui ont contribué à la réalisation de ce projet de fin d'études. Tout d'abord, nous remercions sincèrement Monsieur le Directeur de l'EMSI pour la formation de qualité qu'il nous a offerte. Grâce à son leadership et à sa vision, nous avons bénéficié d'un environnement académique exceptionnel, propice à notre épanouissement et à notre réussite. Nos sincères remerciements vont également à nos professeurs, dont l'expertise, la patience et l'engagement ont été essentiels à notre apprentissage. Leurs enseignements et leurs conseils nous ont guidés tout au long de notre parcours académique, nous permettant de développer les compétences nécessaires pour relever les défis de demain. À nos amis, merci pour votre soutien constant, vos encouragements et pour les moments de camaraderie qui ont rendu cette expérience encore plus mémorable. Votre présence à nos côtés a été une source de motivation et de réconfort. Enfin, nous souhaitons exprimer notre infinie reconnaissance à nos parents. Votre amour, votre soutien inconditionnel et vos sacrifices ont été le pilier sur lequel nous avons construit nos rêves. Sans vous, rien de tout cela n'aurait été possible. Ce projet est le fruit de la collaboration, du travail acharné et du soutien de chacun d'entre vous. Merci de nous avoir permis de réaliser cet accomplissement.

# Résumé

Ce projet porte sur la conception et le développement d'un système de diagnostic automatisé visant à identifier les pathologies pulmonaires majeures, telles que la pneumonie bactérienne, la pneumonie virale et le COVID-19. L'objectif central est de fournir aux radiologues un outil d'aide à la décision basé sur les réseaux de neurones convolutifs (CNN).

La démarche méthodologique adoptée suit le standard industriel CRISP-DM, structurant le projet depuis l'exploration d'un ensemble de données massif de 11 712 images jusqu'au déploiement final. À la suite d'une phase d'expérimentation comparative entre plusieurs modèles, l'architecture DenseNet121 a été retenue pour ses performances supérieures. Les résultats obtenus sont probants : une précision globale de 85,37 % et un rappel (sensibilité) exceptionnel de 99,2 % pour la détection des infections bactériennes. Ce travail illustre comment le transfert d'apprentissage peut transformer l'analyse de l'imagerie médicale en un processus rapide, fiable et accessible via une interface web moderne.

# Abstract

This project details the design and implementation of an automated diagnostic system dedicated to identifying major lung diseases, including bacterial pneumonia, viral pneumonia, and COVID-19. The primary goal is to provide radiologists with a decision-support tool powered by Convolutional Neural Networks (CNN).

The development process follows the CRISP-DM methodological framework, guiding the project from the exploration of a large dataset of 11,712 images to the final deployment. Following a comparative experimental phase, the DenseNet121 architecture was selected due to its superior performance. The results are highly significant : an overall accuracy of 85.37% and an outstanding recall of 99.2% specifically for bacterial infection detection. This study demonstrates how transfer learning can effectively turn medical imaging analysis into a fast, reliable, and user-friendly process through a modern web interface.

# Table des matières

<b>Dédicace</b>	<b>2</b>
<b>Remerciements</b>	<b>3</b>
<b>Résumé</b>	<b>4</b>
<b>Abstract</b>	<b>5</b>
<b>Introduction générale</b>	<b>10</b>
<b>1 État de l’art</b>	<b>11</b>
1.1 Introduction . . . . .	12
1.2 Étude bibliométrique . . . . .	12
1.2.1 Étude Quantitative . . . . .	12
1.2.2 Étude Qualitative . . . . .	15
Résumé du chapitre . . . . .	16
<b>2 Conception et mise en œuvre</b>	<b>17</b>
2.1 Introduction . . . . .	18
2.2 Techniques Utilisées . . . . .	18
2.3 Processus CRISP-DM . . . . .	18
2.3.1 Compréhension des données . . . . .	18
2.3.2 Préparation des données . . . . .	19
2.3.3 Modélisation . . . . .	19
2.3.4 Temps d’entraînement et ressources matérielles . . . . .	20
2.3.5 Analyse des résultats et comparaison . . . . .	20
2.3.6 Déploiement et analyse des résultats via l’interface utilisateur . . . . .	23
<b>Conclusion générale</b>	<b>26</b>
<b>Perspectives</b>	<b>27</b>
<b>Webographie</b>	<b>28</b>
<b>Bibliographie</b>	<b>29</b>

# Table des figures

1.1	Interface de la plateforme OpenAlex et statistiques de recherche sur les maladies pulmonaires . . . . .	12
1.2	Indicateurs bibliométriques principaux obtenus avec Bibliometrix . . . . .	13
1.3	Nuage de mots des thèmes les plus fréquents . . . . .	13
1.4	Graphique des mots-clés les plus utilisés . . . . .	13
1.5	Analyse de l'évolution et croissance des publications . . . . .	14
1.6	Classement des auteurs ayant le plus publié entre 2020 et 2025 . . . . .	14
2.1	Répartition des images par groupe et par catégorie . . . . .	18
2.2	Évolution de la précision et des erreurs pendant l'entraînement . . . . .	19
2.3	Comparaison des performances globales entre les deux modèles . . . . .	20
2.4	Graphique de comparaison des scores des modèles . . . . .	21
2.5	Résultats détaillés pour chaque type de maladie . . . . .	21
2.6	Tableau complet des mesures de performance pour DenseNet121 . . . . .	22
2.7	Matrices de confusion montrant les erreurs de chaque modèle . . . . .	22



# Liste des tableaux

1	Tableau des sigles et acronymes utilisés dans le rapport . . . . .	9
---	--	---

# Liste des sigles et acronymes

IA	Intelligence Artificielle : utilisation des machines pour simuler l'intelligence humaine.
DL	Deep Learning : technique d'apprentissage automatique basée sur les réseaux de neurones profonds.
CNN	Convolutional Neural Network : réseau de neurones spécialisé dans l'analyse d'images.
X-Ray	Radiographie thoracique : image médicale des poumons utilisée pour le diagnostic.
UML	Unified Modeling Language : langage standard pour représenter des modèles logiciels.
CRISP-DM	Cross Industry Standard Process for Data Mining : méthode pour structurer un projet de data science.
GPU	Graphics Processing Unit : processeur graphique utilisé pour accélérer les calculs de Deep Learning.
CPU	Central Processing Unit : processeur central de l'ordinateur.
RAM	Random Access Memory : mémoire vive permettant de stocker temporairement des données pour le calcul.

TABLE 1 – Tableau des sigles et acronymes utilisés dans le rapport

# Introduction générale

L'imagerie médicale s'est imposée comme un pilier indispensable du diagnostic clinique moderne. Cependant, l'augmentation constante du volume de données radiographiques et la complexité croissante des pathologies respiratoires représentent des défis majeurs pour les infrastructures de santé mondiales. Dans ce contexte, l'intelligence artificielle, et plus particulièrement l'apprentissage profond (Deep Learning), offre des solutions innovantes pour automatiser et sécuriser l'interprétation des clichés thoraciques.

L'enjeu de ce projet de fin d'année est de développer une solution logicielle capable de classer avec précision les radiographies des poumons en trois classes distinctes : "Normal", "Pneumonie Bactérienne" et "Pneumonie Virale". Pour assurer la robustesse et la généralisation du système, nous avons exploité le potentiel du transfert d'apprentissage (Transfer Learning) en adaptant des modèles de pointe tels que ResNet50 et DenseNet121 à notre domaine d'application spécifique.

Ce rapport est organisé en quatre chapitres structurants :

1. **État de l'art** : Une analyse bibliométrique approfondie pour identifier les technologies dominantes et les tendances actuelles de la recherche en informatique médicale.
2. **Conception et méthodologie** : Une description détaillée de la démarche technique selon le cycle CRISP-DM, incluant le prétraitement des données.
3. **Analyse et Modélisation** : Une étude fonctionnelle s'appuyant sur la modélisation UML pour définir l'architecture du système.
4. **Réalisation et Résultats** : La présentation de l'implémentation technique, l'analyse comparative des performances métriques et la présentation de l'interface de déploiement.

# Chapitre 1

## État de l'art

## 1.1 Introduction

La détection automatique des maladies des poumons grâce aux images de radiologie est un défi important pour l'informatique médicale. À cause des crises sanitaires mondiales et de la charge de travail des médecins, il est devenu nécessaire de créer des outils d'aide au diagnostic utilisant l'intelligence artificielle. Cette partie présente une étude bibliométrique pour comprendre l'état actuel de la recherche. Nous allons identifier les technologies les plus utilisées, comme le Deep Learning, et justifier nos choix pour ce projet grâce à une analyse précise des données scientifiques.

## 1.2 Étude bibliométrique

### 1.2.1 Étude Quantitative

#### Sources des données : La plateforme OpenAlex

Les informations de cette étude proviennent de la plateforme OpenAlex, qui est une base de données scientifique gratuite. Nous avons cherché les travaux sur la détection des maladies pulmonaires par radiographie entre 2020 et 2025.

Comme on peut le voir sur la Figure 1.1, cette recherche a donné 789 résultats. L'analyse montre que 65,7 % des articles sont en libre accès, ce qui aide à partager les connaissances sur des sujets importants comme le COVID-19 (486 publications) ou d'autres infections respiratoires.

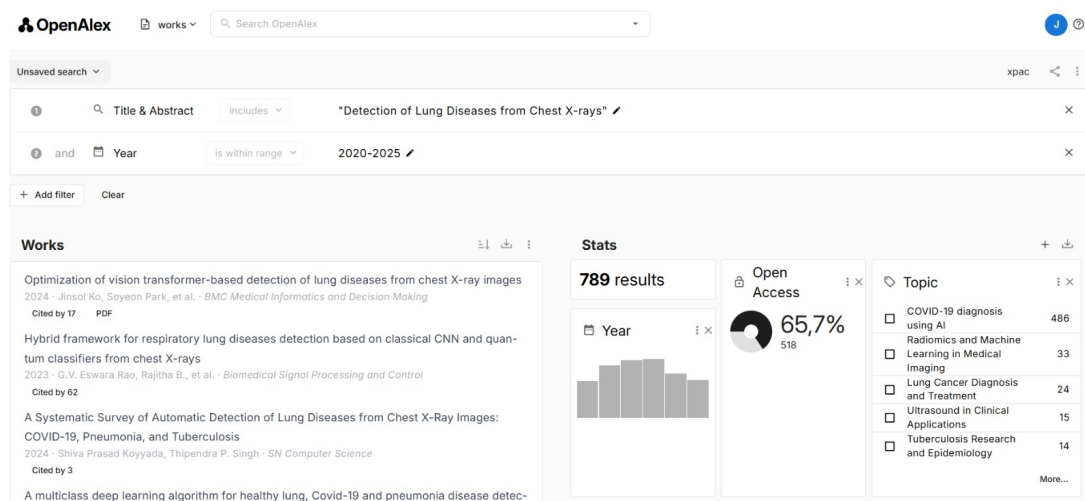


FIGURE 1.1 – Interface de la plateforme OpenAlex et statistiques de recherche sur les maladies pulmonaires

## Outils d'analyse : RStudio et Bibliometrix

Pour analyser ces données, nous avons utilisé le logiciel RStudio avec l'outil Bibliometrix. Les indicateurs principaux de la Figure 1.2 montrent un ensemble de 789 documents provenant de 362 sources différentes, avec un total de 3 235 auteurs. On remarque aussi une forte collaboration entre les pays (21,17 %) et un impact scientifique réel, car chaque document est cité en moyenne 17,2 fois.

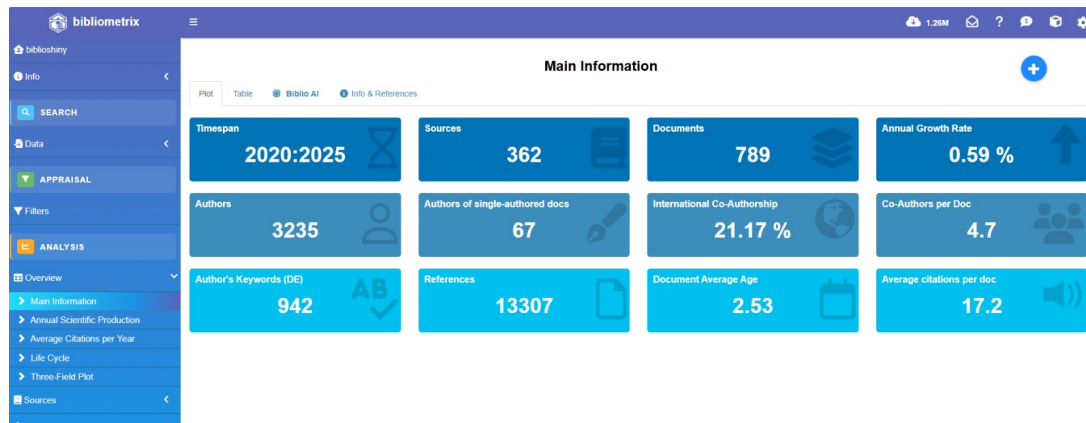


FIGURE 1.2 – Indicateurs bibliométriques principaux obtenus avec Bibliometrix

## Analyse des thèmes et des mots-clés

L'analyse des mots-clés sur la Figure 1.4 et le nuage de mots sur la Figure 1.3 montrent que ce domaine mélange plusieurs spécialités. Les mots les plus fréquents sont la médecine (655 fois), l'informatique (554 fois) et l'intelligence artificielle (528 fois).

L'utilisation fréquente de termes comme Deep Learning (262 fois) ou réseaux de neurones (CNN) confirme que ces techniques sont les plus populaires pour analyser les radios du thorax. Les maladies les plus étudiées sont le COVID-19, la pneumonie et la tuberculose.

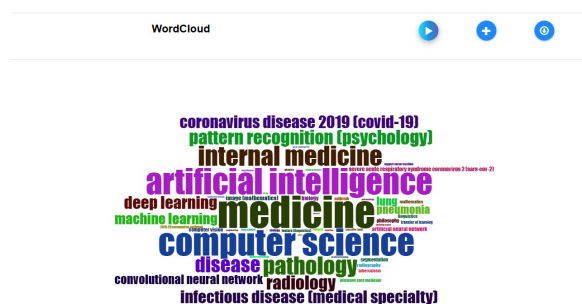


FIGURE 1.3 – Nuage de mots des thèmes les plus fréquents

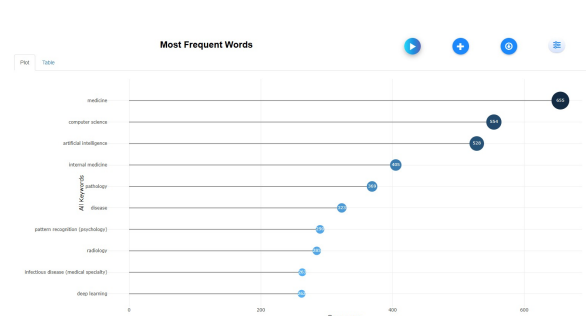


FIGURE 1.4 – Graphique des mots-clés les plus utilisés

## Évolution de la production scientifique

L'évolution dans le temps, montrée sur la Figure 1.5, indique que le nombre de publications a été le plus élevé en 2022 avec 161 articles. La courbe de croissance montre que le domaine arrive maintenant à une certaine maturité. Cela signifie que les chercheurs cherchent désormais à améliorer les modèles existants plutôt qu'à en créer de nouveaux.

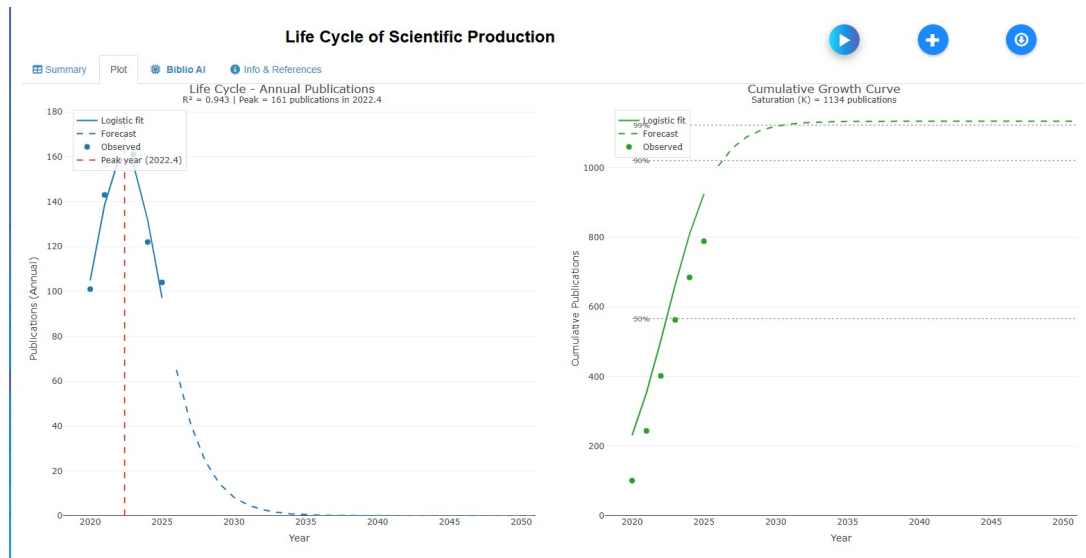


FIGURE 1.5 – Analyse de l'évolution et croissance des publications

## Analyse des auteurs les plus importants

L'identification des auteurs principaux sur la Figure 1.6 montre que des chercheurs comme Amith Khandakar, Muhammad E. H. Chowdhury et Tawsifur Rahman sont très actifs, avec 10 publications chacun sur cette période. Cela prouve qu'il existe des groupes d'experts spécialisés dans l'IA appliquée à la radiographie thoracique.

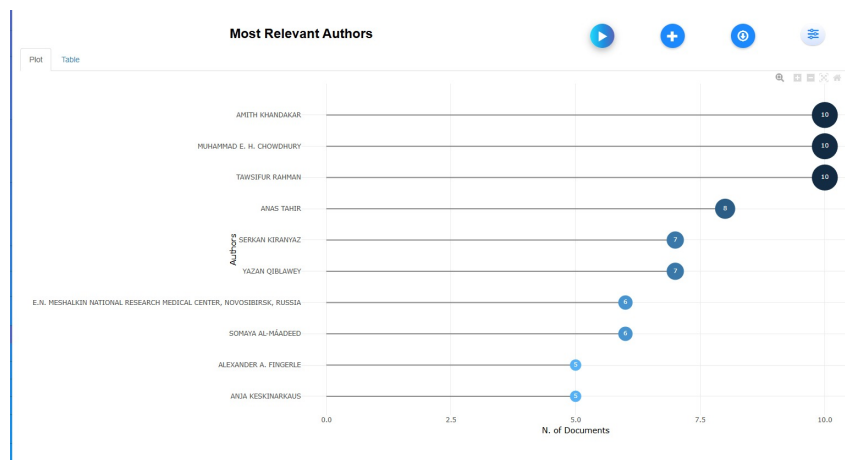


FIGURE 1.6 – Classement des auteurs ayant le plus publié entre 2020 et 2025

### 1.2.2 Étude Qualitative

La détection des maladies pulmonaires par l'imagerie est devenue essentielle pour aider les systèmes de santé. Pour répondre aux besoins de précision et de fiabilité médicale, la recherche a proposé plusieurs méthodes basées sur le Deep Learning. Cette étude qualitative présente une analyse de trois approches importantes de la littérature selon trois axes : la précision des diagnostics, la capacité à traiter plusieurs maladies et l'explication des résultats.

Trois axes complémentaires sont ainsi explorés :

- La recherche d'une précision maximale pour identifier des signes cliniques fins.
- L'utilisation d'architectures capables de classer plusieurs pathologies en même temps.
- L'importance de montrer au médecin les zones suspectes sur la radio pour justifier le diagnostic.

#### Étude 1 : Performance des réseaux de neurones profonds

Cette première étude s'intéresse à l'utilisation de modèles classiques comme ResNet ou VGG pour l'analyse d'images médicales.

- **Contexte et objectifs** : L'idée est de dépasser les capacités humaines pour détecter des maladies difficiles à voir, comme les débuts de pneumonie, en utilisant des modèles déjà entraînés sur des millions d'images.
- **Méthodologie** : Les chercheurs utilisent souvent la technique du Transfer Learning. Cela permet d'adapter un modèle informatique existant aux spécificités des radiographies thoraciques sans repartir de zéro.
- **Résultats et analyse** : Les résultats montrent que ces modèles sont très fiables et peuvent servir de "deuxième avis" pour les radiologues. Cependant, ils demandent beaucoup de puissance de calcul.

#### Étude 2 : Détection multi-classes et modèles unifiés

Cette étude analyse des travaux comme ceux de CheXNet, qui permettent de détecter plusieurs maladies sur une seule image.

- **Contexte et problématique** : Souvent, une radiographie peut montrer plusieurs problèmes en même temps. Il faut donc un système capable de donner plusieurs probabilités en un seul passage.
- **Méthodologie** : Les auteurs utilisent des architectures comme DenseNet. Ces réseaux permettent une meilleure circulation de l'information entre les couches, ce qui aide à mieux comprendre les détails complexes des tissus pulmonaires.
- **Résultats et analyse** : Cette approche a prouvé qu'une intelligence artificielle peut atteindre le niveau d'un radiologue expérimenté pour détecter la pneumonie. Elle montre que la structure du réseau est très importante pour la réussite du projet.



### Étude 3 : Interprétabilité et aide à la décision

Cette étude s'intéresse à la manière dont l'IA explique son choix au médecin.

- **Contexte et défis** : En médecine, il ne suffit pas de donner un résultat, il faut expliquer pourquoi. Le médecin doit voir où l'IA a détecté une anomalie.
- **Méthodologie** : On utilise des techniques de visualisation comme les cartes de chaleur (Heatmaps). Ces cartes colorent les zones de la radio qui ont poussé l'IA à prendre sa décision.
- **Résultats et compromis** : Même si cela n'augmente pas toujours la précision brute, cela donne plus de confiance aux médecins. Cela rend l'outil beaucoup plus facile à utiliser dans un vrai hôpital.

### Synthèse comparative et choix du modèle pour le projet

L'analyse comparative montre trois points importants. Les réseaux très profonds sont très précis mais lourds. Les modèles unifiés comme DenseNet sont parfaits pour gérer plusieurs maladies. Enfin, les outils de visualisation sont nécessaires pour que le système soit accepté par les professionnels.

Sur la base de cette étude qualitative, nous avons choisi d'utiliser l'architecture DenseNet121 pour notre projet. Ce modèle offre un excellent équilibre entre la précision et la rapidité. De plus, il est très efficace pour apprendre à partir de bases de données de radiographies thoraciques, ce qui correspond exactement à notre objectif de détection de la pneumonie et du COVID-19.

## Résumé

Ce chapitre a permis d'analyser le contexte scientifique du projet à travers deux approches complémentaires :

- **L'étude quantitative** : Grâce à l'analyse de 789 publications (2020-2025), nous avons constaté que le domaine est en pleine maturité. L'intelligence artificielle, et plus précisément le Deep Learning, est devenue l'outil standard pour traiter l'augmentation massive des données médicales depuis la crise du COVID-19.
- **L'étude qualitative** : L'examen de la littérature montre que les modèles de type CNN (réseaux de neurones convolutionnels) offrent les meilleurs résultats. Nous avons identifié trois critères de réussite : la précision du diagnostic, la capacité à gérer plusieurs maladies simultanément, et l'importance de l'explicabilité (montrer au médecin pourquoi l'IA a pris sa décision).

En conclusion, ces analyses justifient notre choix d'utiliser l'architecture **DenseNet121**. Ce modèle est reconnu pour sa grande précision tout en restant efficace en termes de calcul, ce qui correspond parfaitement à notre objectif de détection de la pneumonie et du COVID-19.

## Chapitre 2

### Conception et mise en œuvre

## 2.1 Introduction

Ce chapitre explique les détails techniques et les choix que nous avons faits pour créer notre système de diagnostic. Nous avons utilisé une méthode sérieuse pour être certains que le modèle de détection des maladies des poumons soit efficace et fiable.

## 2.2 Techniques Utilisées

Pour développer ce projet, nous avons utilisé des technologies modernes de l'apprentissage profond (Deep Learning). Nous avons travaillé avec l'outil TensorFlow et la bibliothèque Keras pour construire nos modèles.

Nous avons choisi d'utiliser le transfert d'apprentissage (Transfer Learning). Cette technique permet de réutiliser des modèles qui ont déjà appris sur des millions d'images, comme ResNet50 ou DenseNet121. Cela nous aide à obtenir de bons résultats même si nous n'avons pas énormément d'images médicales, et cela permet aussi de gagner du temps pendant les calculs.

## 2.3 Processus CRISP-DM

Pour organiser notre travail, nous avons suivi les étapes de la méthode CRISP-DM. Ce processus nous guide de la compréhension des besoins jusqu'à l'installation de la solution finale.

### 2.3.1 Compréhension des données

L'objectif est de classer les images de radiographies en trois groupes : les poumons normaux, les pneumonies causées par des bactéries et les pneumonies causées par des virus. Nous voulons atteindre une précision d'au moins 80 % pour que le projet soit une réussite.

Notre base de données contient 11 712 images. La répartition de ces images est montrée dans la Figure 2.1.

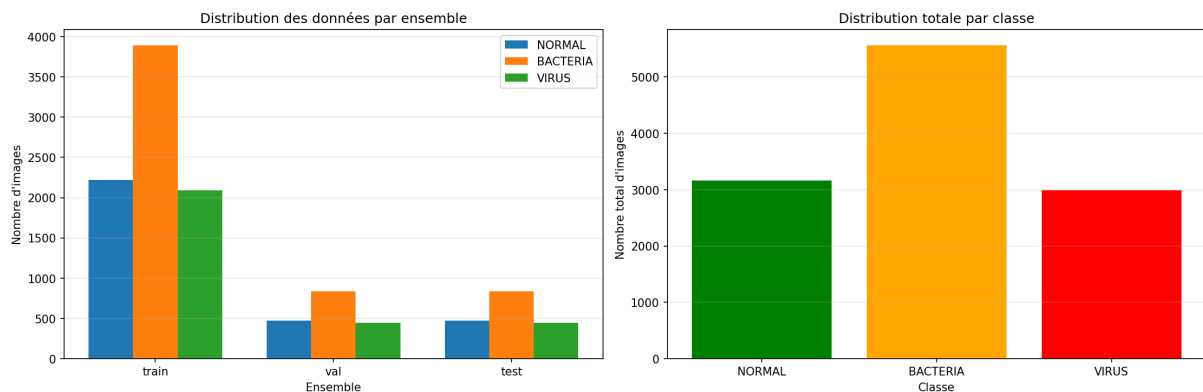


FIGURE 2.1 – Répartition des images par groupe et par catégorie

La Figure 2.1 nous donne des informations importantes sur notre base de données. On voit que nous avons bien divisé les données en trois ensembles : 70 % pour l'entraînement, 15 % pour la validation et 15 % pour le test. Sur le graphique de droite, on remarque que la catégorie des bactéries contient beaucoup plus d'images que les catégories normales ou virales. Il a fallu faire attention à ce déséquilibre pour que le modèle n'apprenne pas seulement à reconnaître les bactéries.

### 2.3.2 Préparation des données

Avant de donner les images au modèle, nous devons les préparer. Voici les étapes importantes :

- Nous changeons la taille des images en 224x224 pixels pour qu'elles soient toutes identiques.
- Nous normalisons les pixels pour que les calculs mathématiques du modèle soient plus rapides et plus stables.
- Nous utilisons l'augmentation de données, comme des petites rotations ou des zooms, pour que le modèle voie des images légèrement différentes à chaque fois et n'apprenne pas par cœur.

### 2.3.3 Modélisation

Nous avons testé deux modèles différents : ResNet50 et DenseNet121. Nous avons ajouté à ces modèles des couches spécifiques, comme une couche de sortie avec trois neurones et une couche de sécurité appelée Dropout pour éviter les erreurs. La Figure 2.2 montre l'évolution du modèle pendant les 30 étapes d'apprentissage.

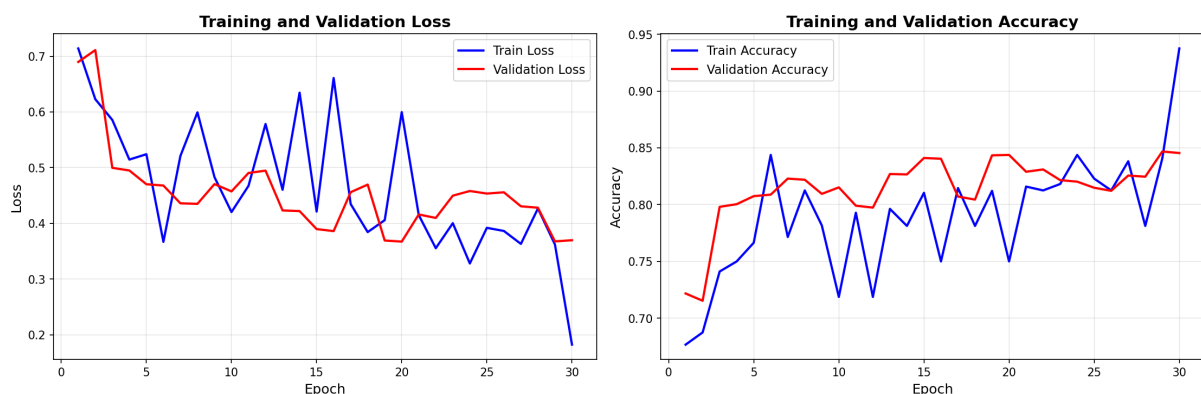


FIGURE 2.2 – Évolution de la précision et des erreurs pendant l'entraînement

Sur la Figure 2.2, nous pouvons voir deux graphiques. Celui de gauche montre que l'erreur (loss) diminue régulièrement, ce qui est une très bonne chose. Celui de droite montre que la précision (accuracy) augmente progressivement. On voit que les résultats pour l'entraînement et pour la validation sont très proches, ce qui signifie que le modèle est capable de bien fonctionner sur des images qu'il n'a jamais vues.

### 2.3.4 Temps d'entraînement et ressources matérielles

L'apprentissage complet du modèle **DenseNet121**, incluant la phase d'entraînement et la phase de test, a nécessité environ **22 à 23 heures**. Ce temps d'exécution s'explique par la taille importante de la base de données (**11 712 images**), l'utilisation du transfert d'apprentissage (*Transfer Learning*) et la complexité de l'architecture DenseNet121.

Les expérimentations ont été réalisées sur une machine disposant des ressources matérielles suivantes :

- **Processeur (CPU)** : Intel Core i5 de 11<sup>e</sup> génération
- **Carte graphique (GPU)** : NVIDIA GTX 1650 Super
- **Mémoire vive (RAM)** : 16 Go

Malgré ce coût en temps de calcul, cette durée reste acceptable dans un contexte médical, car l'entraînement du modèle est effectué une seule fois. En revanche, la phase d'inférence sur une nouvelle radiographie est quasi instantanée, ce qui permet une utilisation efficace du système en pratique clinique.

### 2.3.5 Analyse des résultats et comparaison

Le modèle DenseNet121 est le plus performant. Comme on peut le voir dans le tableau de la Figure 2.3, il obtient une précision de 85,37 % alors que ResNet50 s'arrête à 72,15 %.

**Comparaison des Modèles: ResNet50 vs DenseNet121**

Metric	ResNet50	DenseNet121	Difference
Accuracy	0.7215	0.8537	+0.1322
Precision (Macro)	0.6946	0.8434	+0.1488
Recall (Macro)	0.7007	0.8542	+0.1535
F1-Score (Macro)	0.6872	0.8480	+0.1608

FIGURE 2.3 – Comparaison des performances globales entre les deux modèles

Ce tableau montre clairement que DenseNet121 est meilleur pour toutes les catégories. La Figure 2.4 permet de visualiser cette différence avec des barres de couleur. On remarque que DenseNet121 a des scores plus élevés pour la précision, le rappel et le score F1.

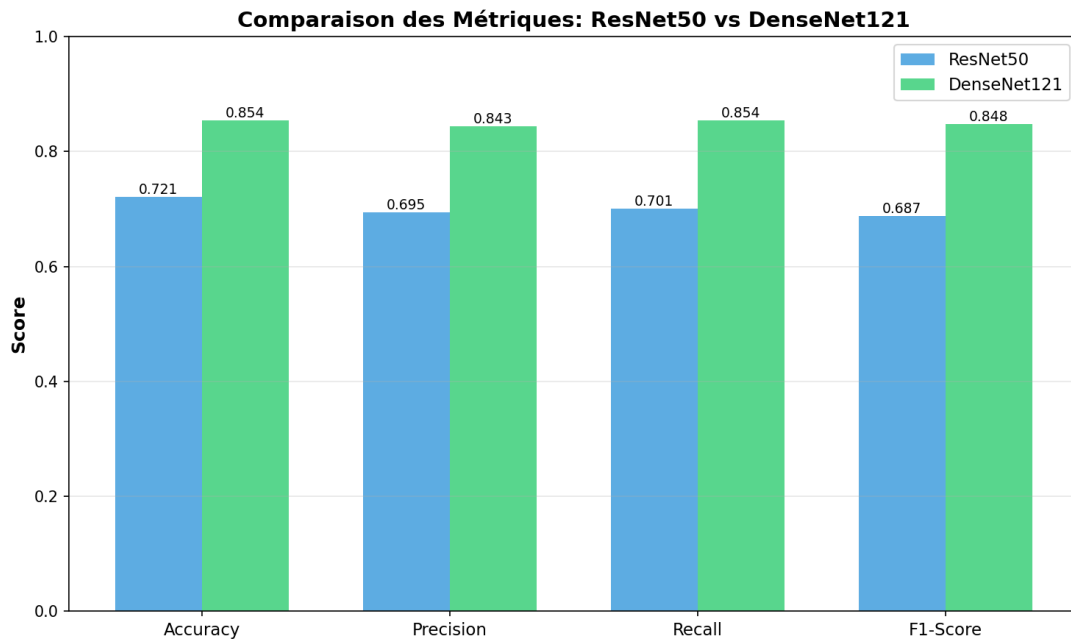


FIGURE 2.4 – Graphique de comparaison des scores des modèles

Pour comprendre les détails, nous avons regardé les résultats par type de maladie (Figure 2.5) et les chiffres précis (Figure 2.6).

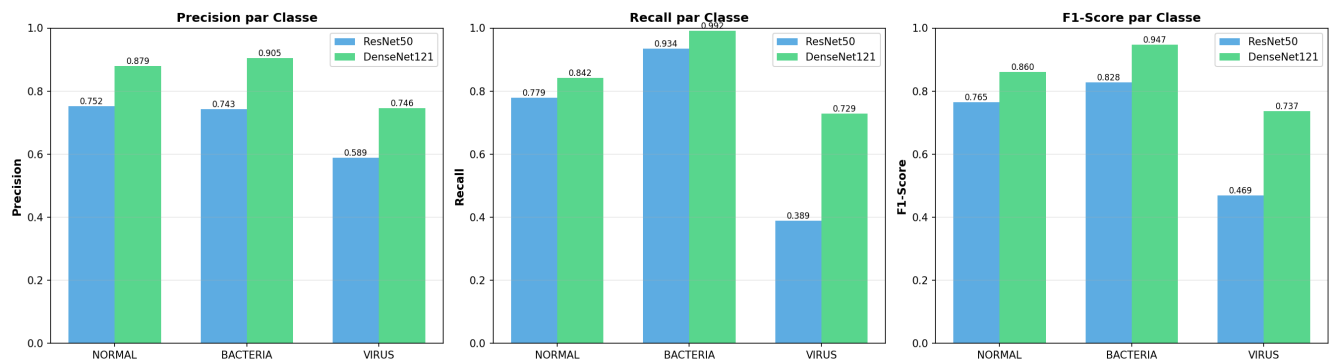


FIGURE 2.5 – Résultats détaillés pour chaque type de maladie

### Classification Metrics

Class	Precision	Recall	F1-Score	Support
NORMAL	0.8791	0.8421	0.8602	2780
BACTERIA	0.9054	0.9918	0.9466	1583
VIRUS	0.7457	0.7287	0.7371	1493
<b>AVERAGE (Macro)</b>	<b>0.8434</b>	<b>0.8542</b>	<b>0.8480</b>	<b>5856</b>

FIGURE 2.6 – Tableau complet des mesures de performance pour DenseNet121

La Figure 2.5 montre que le modèle est extrêmement efficace pour la catégorie des bactéries, avec un rappel de 99,2 %. Cela veut dire que le modèle ne manque presque aucune infection bactérienne. Par contre, le score pour les virus est plus bas (72,9 %), car les virus ressemblent parfois aux images normales. La Figure 2.6 confirme ces résultats avec le nombre exact d'images testées pour chaque catégorie.

Enfin, les matrices de confusion de la Figure 2.7 nous montrent où le modèle se trompe.

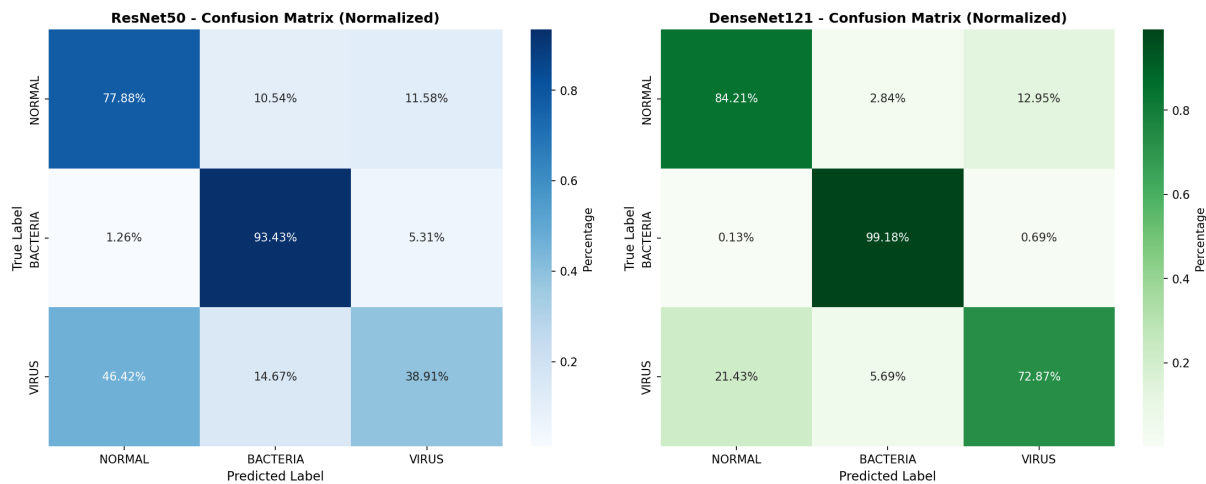
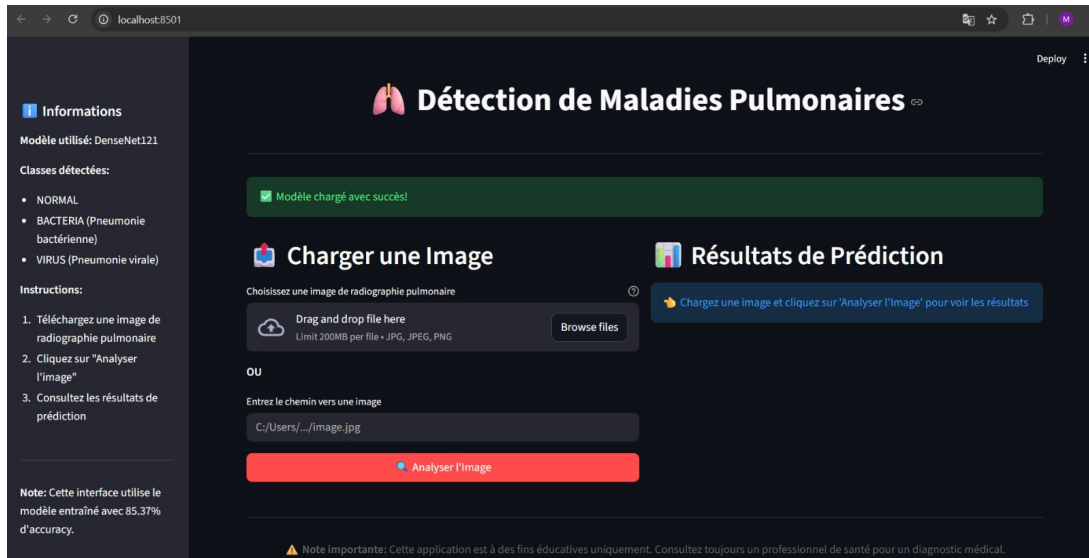


FIGURE 2.7 – Matrices de confusion montrant les erreurs de chaque modèle

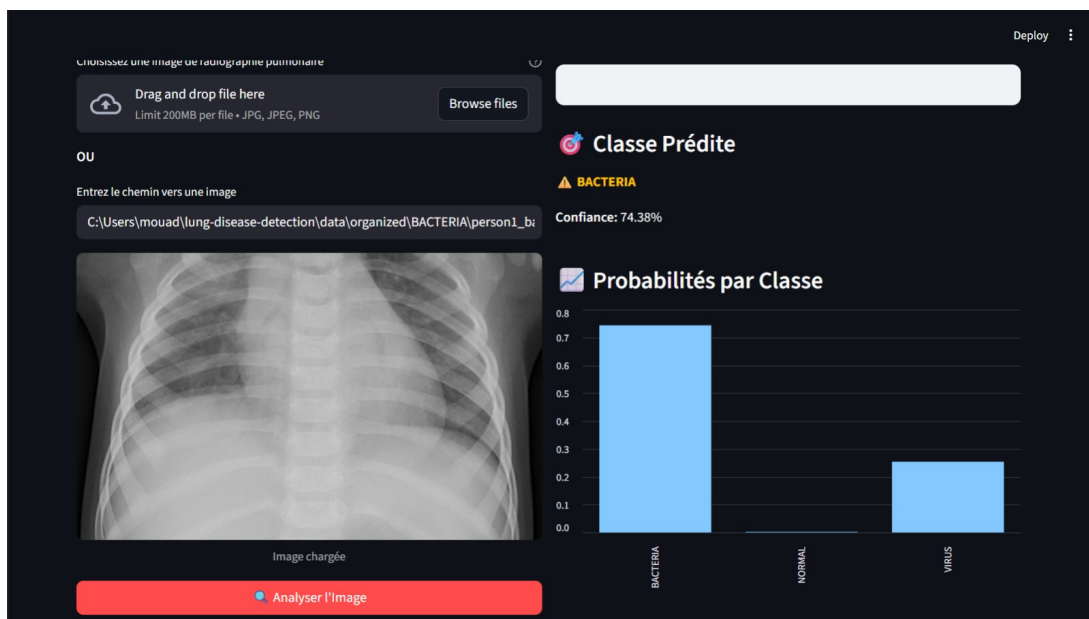
Sur la Figure 2.7, on voit que DenseNet121 classe presque parfaitement les bactéries (seulement 16 erreurs sur 1952 images). C'est beaucoup mieux que ResNet50 qui fait plus de 200 erreurs pour cette même catégorie. Cela prouve que notre choix final est le bon.

### 2.3.6 Déploiement et analyse des résultats via l'interface utilisateur

Une fois le modèle DenseNet121 entraîné et validé, une interface graphique a été développée avec la bibliothèque Streamlit afin de faciliter l'utilisation du système de diagnostic. Cette interface permet à l'utilisateur de charger une radiographie thoracique et d'obtenir une prédiction accompagnée de probabilités.

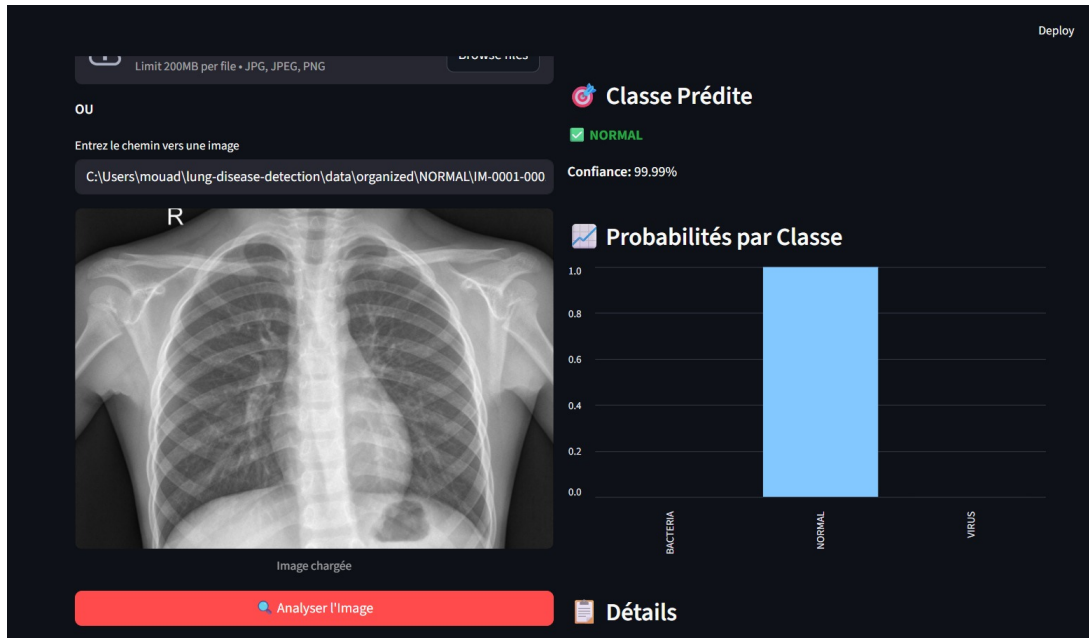


Cette figure présente l'interface principale de l'application. La barre latérale affiche les informations générales du modèle ainsi que les classes de diagnostic disponibles. La partie centrale permet le chargement de la radiographie et le lancement de l'analyse.

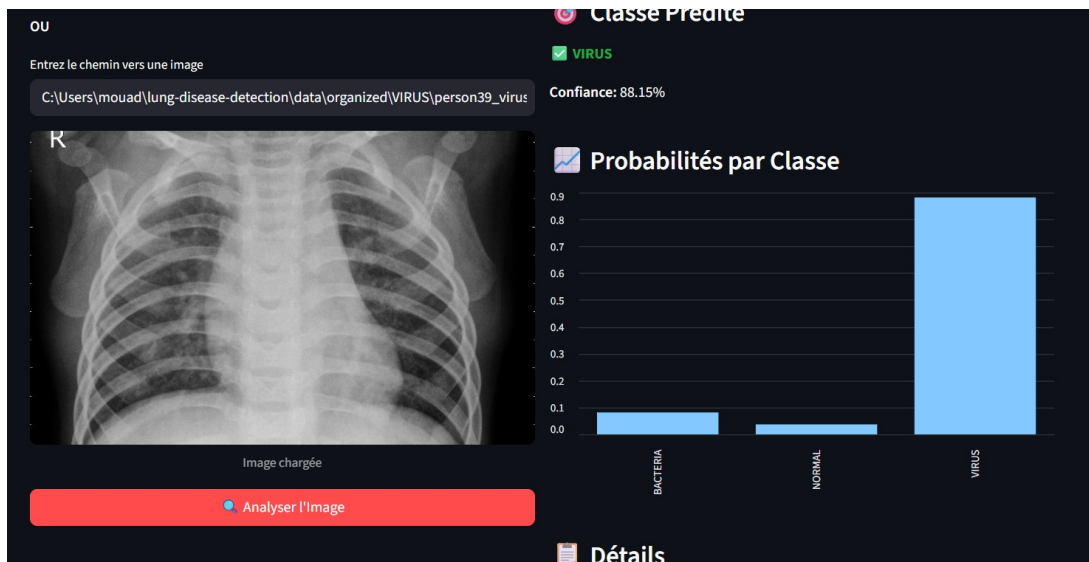




Cette figure montre le résultat obtenu pour une radiographie atteinte d'une pneumonie bactérienne. Le modèle attribue une probabilité de 74,38% à cette classe. Le graphique des probabilités indique que cette catégorie est largement dominante par rapport aux autres.



Cette figure présente le cas d'un poumon sain. La classe normale est détectée avec une probabilité de 99,99%, ce qui montre une très forte confiance du modèle dans sa prédiction.



Cette figure illustre la détection d'une pneumonie virale avec une probabilité de 88,15%. Une faible probabilité reste associée à la classe bactérienne, ce qui montre que certaines infections peuvent présenter des caractéristiques visuelles proches.

Ces exemples montrent que l'interface permet une lecture simple et claire des résultats. Les probabilités affichées aident l'utilisateur à mieux comprendre la décision proposée par le système et à l'utiliser comme un outil d'aide au diagnostic.

## Conclusion

Ce chapitre a présenté la conception et la mise en œuvre d'un système de diagnostic automatique des maladies pulmonaires à partir de radiographies thoraciques. Le projet utilise le Deep Learning et la technique du transfert d'apprentissage pour analyser une base de 11 712 images, préparées par redimensionnement, normalisation et augmentation.

Deux modèles ont été testés, ResNet50 et DenseNet121, ce dernier offrant les meilleures performances avec une précision globale de 85,37 % et un excellent rappel pour les pneumonies bactériennes. Malgré un temps d'entraînement de 22 à 23 heures, l'inférence sur de nouvelles images est quasi instantanée.

Une interface interactive développée avec Streamlit permet à l'utilisateur de charger une radiographie et d'obtenir facilement la prédiction et les probabilités associées. Ce système constitue ainsi un outil fiable et pratique d'aide au diagnostic, prêt à être utilisé ou amélioré pour des applications médicales réelles.

# Conclusion générale

Pour conclure, ce projet nous a permis de franchir toutes les étapes de la création d'une application d'intelligence artificielle appliquée à la santé. Nous avons réussi à transformer un besoin médical complexe en un système fonctionnel capable de traiter des milliers d'images avec une précision remarquable.

Le processus de développement nous a confrontés aux réalités de la science des données : la nécessité d'un nettoyage rigoureux des données, l'importance de l'augmentation d'images pour lutter contre le surapprentissage (overfitting) et l'ajustement précis des hyperparamètres pour optimiser la convergence des modèles. Les tests finaux valident notre approche, montrant que le modèle DenseNet121 offre une sensibilité très élevée, critère crucial en médecine pour éviter les faux négatifs.

L'aboutissement de ce travail est la mise en ligne d'une interface réalisée avec Streamlit, qui rend l'utilisation du modèle intuitive pour un utilisateur non-technicien. Cette expérience enrichissante a consolidé nos compétences techniques en Python et en Deep Learning, tout en soulignant l'importance de l'éthique et de la rigueur scientifique dans le développement d'outils numériques destinés à la médecine.

# Perspectives

Bien que les résultats actuels soient très satisfaisants, plusieurs pistes d'amélioration pourraient enrichir ce projet à l'avenir :

- **Explicabilité du modèle** : L'intégration d'algorithmes de visualisation comme Grad-CAM permettrait de générer des "cartes de chaleur" sur les poumons, montrant précisément au médecin quelles zones de l'image ont influencé la décision de l'IA.
- **Élargissement du diagnostic** : Nous pourrions étendre la capacité du modèle à détecter d'autres anomalies comme l'atélectasie, l'œdème pulmonaire ou le pneumothorax.
- **Optimisation pour le mobile** : En utilisant des techniques de quantification ou de distillation de modèles, nous pourrions rendre l'application assez légère pour fonctionner sur des tablettes ou des smartphones, facilitant ainsi son usage dans des zones isolées.
- **Apprentissage continu** : Mettre en place un système de retour d'expérience où les corrections des médecins servent à ré-entraîner périodiquement le modèle pour améliorer ses performances sur les cas les plus difficiles.

# Webographie

1. Pour AI2 (Allen Institute for Artificial Intelligence) : <https://allenai.org/>
2. Pour ASTA (logiciel de planification et gestion de projet) : <https://www.astapowerproject.com/>
3. Pour OpenAlex (base de données ouverte de publications scientifiques) : <https://openalex.org/>
4. Pour PubMed (moteur de recherche bibliographique en sciences biomédicales) :  
<https://pubmed.ncbi.nlm.nih.gov/>
5. Pour SciSpace (assistant IA pour la recherche scientifique et l'analyse d'articles) :  
<https://www.scispace.com/>

# Bibliographie

La réalisation de ce projet s'est appuyée sur une base documentaire solide et variée :

- **Ouvrages techniques** : Étude des fondamentaux des réseaux de neurones convolutifs (CNN) et des architectures profondes via la documentation officielle de TensorFlow et Keras.
- **Articles scientifiques** : Consultation de publications académiques sur les architectures ResNet et DenseNet, ainsi que sur l'application du Transfer Learning à l'imagerie médicale.
- **Méthodologie** : Utilisation des guides de référence sur le processus CRISP-DM pour la gestion de projets de Data Science.
- **Ressources en ligne** : Analyse de cas pratiques sur Kaggle et GitHub pour l'optimisation des fonctions de perte (Loss functions) comme la Categorical Crossentropy et l'utilisation de l'optimiseur Adam.