



Royaume du Maroc المملكة المغربية

كلية الطب والصيدلة
+ⴰⴳⴷⵓⴻⴰⴳ ⴰⴳⴷⵓⴻⴰⴳ ⴰⴳⴷⵓⴻⴰⴳ
FACULTÉ DE MÉDECINE ET DE PHARMACIE

Année 2020

Thèse N° 214/20

CRÉATION ET ÉVALUATION D'UN SYSTÈME INTELLIGENT D'AIDE À LA DÉCISION CLINIQUE (SADC) POUR LE DIAGNOSTIC DES LÉSIONS PULMONAIRES À PARTIR DE RADIOGRAPHIES PULMONAIRES

THÈSE

PRÉSENTÉE ET SOUTENUE PUBLIQUEMENT LE 22/12/2020

PAR

Mr. MAMOUN QJIDAA

Né le 05 Août 1993 à Fès

POUR L'OBTENTION DU DOCTORAT EN MÉDECINE

MOTS-CLÉS :

Radiographie pulmonaire- Système aide au diagnostique- Intelligence artificielle-
Apprentissage profond- Réseaux de neurones

JURY

| | |
|---|----------------|
| M. BOUJRAF SAID | PRÉSIDENT |
| Professeur de Biophysique | |
| M. MAAROUFI MUSTAPHA | RAPPORTEUR |
| Professeur de Radiologie | |
| M. LAMRANI ALAOUI MOULAY YOUSSEF | JUGES |
| Professeur agrégé de Radiologie | |
| M. ALAMI BADREDDINE | |
| Professeur agrégé de Radiologie | |
| M. QJIDAA HASSAN | MEMBRE ASSOCIÉ |
| Professeur d'Informatique | |

PLAN

| | |
|--|----|
| LISTE DES FIGURES | 7 |
| LISTE DES TABLEAUX | 15 |
| INTRODUCTION générale | 16 |
| 1. Contexte et problématique | 17 |
| 2. Approche proposée | 19 |
| 3. Plan du manuscrit..... | 21 |
| CHAPITRE I LA RADIOGRAPHIE THORACIQUE..... | 22 |
| 1. Introduction | 23 |
| 2. Qualité Du Cliché..... | 25 |
| 3. Aperçu de la sémiologie radiologique nécessaire a notre étude | 25 |
| 4. LES SYNDROMES PARENCHYMATEUX : | 25 |
| 4.1.Syndrome alvéolaire | 25 |
| 4.2.L'emphysème | 30 |
| 4.3.Atélectasie | 31 |
| 5. SYNDROMES PLEURAUX..... | 34 |
| 5.1.Pleurésie | 34 |
| 5.2.Pneumothorax..... | 38 |
| 5.3.Épaississement pleural | 39 |
| 6. SYNDROME INTERSTITIEL | 40 |
| 6.1.Le verre dépoli | 42 |
| 6.2.Fibrose pulmonaire | 42 |
| 7. SYNDROME TUMORAL..... | 44 |
| 7.1.Les nodules | 44 |
| 7.2.Masses..... | 45 |
| 7.3.Les images cavitaires..... | 46 |
| 8. SYNDROME MÉDIASTINAL | 48 |
| 8.1.Cardiomégalie | 48 |
| 8.2.Les hernies..... | 50 |
| CHAPITRE II INTELLIGENCE ARTIFICIELLE EN SANTE | 53 |
| 1. Introduction | 54 |
| 2. Historique | 55 |
| 3. L'intelligence artificielle dans le secteur de la santé..... | 57 |

| | |
|--|-----|
| 4. Les fonctionnements de l'intelligence artificielle (Machine Learning, Deep Learning)..... | 58 |
| 4.1. Machine Learning | 59 |
| 4.2. Deep Learning | 61 |
| 5. L'aide au diagnostic..... | 63 |
| 6. Neurone humain et artificiel | 64 |
| 7. Qu'est ce que le machine learning (ML)?..... | 65 |
| 7.1. Performance et surapprentissage..... | 67 |
| 7.2. Les différents types de machine learning | 69 |
| 7.3. Régression et classification..... | 69 |
| 7.3.1. La régression linéaire..... | 69 |
| 7.4. Les k plus proches voisins | 71 |
| 7.5. Le classifieur naïf de Bayes | 71 |
| 7.6. k-means | 72 |
| 7.7. Les arbres de décision | 73 |
| 7.8. Les forêts aléatoires | 74 |
| 7.9. Les machines à vecteur de support | 76 |
| 7.10. Le perceptron multicouches | 79 |
| 8. Deep Learning | 81 |
| 8.1. Introduction | 81 |
| 8.2. Historique | 83 |
| 8.3. Pourquoi le Deep Learning? | 84 |
| 8.4. Les différents types de modèles..... | 87 |
| 8.5. Les réseaux de neurones convolutifs | 87 |
| 8.6. Optimisation pour l'apprentissage en Deep Learning | 97 |
| 8.7. Algorithmes d'optimisation de la descente de gradient | 100 |
| 8.8. Limites théoriques de l'optimisation | 102 |
| 8.9. Les fonctions d'activation | 103 |
| 8.10. Conclusion | 106 |
| CHAPITRE III METHODES ET MATERIEL..... | 107 |
| 1. Modèle proposé basé sur le Transfert Learning | 108 |
| 1.1. Réseaux de neurones convolutifs utilisant l'apprentissage par transfert..... | 108 |
| 1.2. Réseaux de neurones à convolution pré entraînés sur la base ImageNet | 110 |

| | |
|---|-----|
| 1.3. Architecture VGG16..... | 110 |
| 1.4. Architecture VGG19..... | 112 |
| 1.5. Architecture MobileNet..... | 113 |
| 1.6. Architecture InceptionV3..... | 113 |
| 1.7. Architecture Xception..... | 114 |
| 1.8. Architecture InceptionResNetV2..... | 115 |
| 1.9. Architecture de DenseNet121..... | 117 |
| 1.10. Modèle d'empilage..... | 118 |
| 2. Carte d'activation de classe de gradient (Grad-CAM)..... | 120 |
| 3. Base de données radiographiques du thorax de l'Institut national de la santé (National Institutes of Health NIH)..... | 122 |
| 3.1. La base de données NIH..... | 122 |
| 3.2. Licence et attribution..... | 124 |
| 3.3. Accès aux données Google Cloud..... | 125 |
| 3.4. Stockage en ligne (Cloud Storage)..... | 125 |
| 3.5. Accès à l'ensemble de données radiographiques du thorax du NIH dans le Cloud Storage..... | 125 |
| 3.6. Libellés supplémentaires..... | 126 |
| 3.7. Processus d'attribution des libellés..... | 126 |
| 3.8. Comment utiliser ces libellés..... | 127 |
| 3.9. Les données..... | 128 |
| 3.10. Statistiques descriptives de la Base de données..... | 129 |
| 3.11. Pathologies selon le sexe..... | 130 |
| 3.12. Répartition des pathologies simples et Multiples..... | 130 |
| 3.13. l'atélectasie..... | 131 |
| 3.14. Cardiomegalie..... | 131 |
| 3.15. Consolidation..... | 131 |
| 3.16. Oedème..... | 132 |
| 3.17. Effusion..... | 133 |
| 3.18. Emphysème..... | 133 |
| 3.19. Fibrose..... | 134 |
| 3.20. Infiltration pulmonaire..... | 134 |
| 3.21. Masse..... | 135 |

| | |
|---|-----|
| 3.22.Nodules..... | 135 |
| 3.23.Epaississement..... | 136 |
| 3.24.Pneumonie..... | 136 |
| 3.25.Pneumothorax..... | 137 |
| CHAPITRE IV resultats et disccussion..... | 138 |
| 1. Matériel utilisé..... | 139 |
| 2. Ensemble d'images pour l'apprentissage et la validation du modèle proposé | 140 |
| 3. Ensemble d'image pour le test..... | 141 |
| 4. Prétraitement et augmentation des données..... | 142 |
| 5. Mesure des performances de notre modèle..... | 143 |
| 5.1.Fonction de perte..... | 144 |
| 5.2.Courbe ROC..... | 145 |
| 6. 1 ^{ère} EXPERIENCE : Résultats pour l'ensemble à 3 classes..... | 145 |
| 6.1.Taille des échantillons..... | 145 |
| 6.2.Exemple d'images de chaque classe..... | 146 |
| 6.3.Pourcentage des images considérées dans les phases d'apprentissage, de validation et de test..... | 147 |
| 6.4.Configuration de notre modèle proposé..... | 147 |
| 6.5.Courbes de perte d'apprentissage et de précision pour les phases d'apprentissage et de validation..... | 148 |
| 6.6.Courbe du ROC pour la phase du test..... | 151 |
| 6.7.Performances mesurées de notre modèle proposé..... | 155 |
| 7. 2 ^{ème} EXPERIENCE : Résultats pour l'ensemble à 4 classes..... | 156 |
| 7.1.Taille des échantillons..... | 156 |
| 7.2.Exemple d'images de chaque classe..... | 156 |
| 7.3.Pourcentage des images considérées dans les phases d'apprentissage, de validation et de test..... | 157 |
| 7.4.Configuration de notre modèle proposé..... | 157 |
| 7.5.Courbes de perte d'apprentissage et de précision pour les phases d'apprentissage et de validation..... | 158 |
| 7.6.Courbe du ROC pour la phase du test..... | 161 |
| 7.7.Performances mesurées de notre modèle proposé..... | 165 |
| 8. 3 ^{ème} Expérience: Résultats pour l'ensemble à 14 classes..... | 166 |

| | |
|--|-----|
| 8.1. Résultats du réseau ResNet..... | 166 |
| 8.2. Résultats du réseau ResNet101 | 168 |
| 8.3. Résultats du réseau DenseNet121 | 169 |
| 9. Carte thermique Grad-CAM pour une localisation grossière des lésions pulmonaires avec la probabilité de classification | 172 |
| 10. Exemple d'images utilise pour le test..... | 176 |
| 11. Discussion..... | 183 |
| CONCLUSION GENERALE..... | 186 |
| RÉSUMÉS | 189 |
| REFERENCES..... | 194 |

LISTE DES FIGURES

| | |
|---|----|
| Figure 1 : Les 4 densités radiologiques..... | 23 |
| Figure 2 : Opacité apparaissant blanche sur le cliché | 24 |
| Figure 3 : Clarté apparaissant noire sur le cliché..... | 24 |
| Figure 4 : Radiographie thoracique normale. | 24 |
| Figure 5 : Comblement des alvéoles pulmonaires par du liquide systématisé : (a) vue latérale , (b) vue de profile | 26 |
| Figure 6. Comblement des alvéoles pulmonaires par du liquide avec un bronchogramme aérien..... | 27 |
| Figure 7: Bronchogramme aérique : (a) radiographie pulmonaire (b) CT image | 28 |
| Figure 8 : Oedeme alvéolaire | 29 |
| Figure 9 : augmentation des espaces clairs, rétro–sternal (*) et rétrocardiaque (+)... | 30 |
| Figure 10 : Emphyseme pulmonaire..... | 31 |
| Figure 11. L'atélectasie | 32 |
| Figure 12 : Les petites atélectasies planes, d'orientation horizontale | 32 |
| Figure 13: Séquence montrant l'évolution vers une atélectasie complète du champ pulmonaire droit..... | 33 |
| Figure 14 Comblement du cul de sac pleural gauche | 35 |
| Figure 15: Epanchement pleural gauche de moyenne abondance | 36 |
| Figure 16 : pleurésie à moyenne abondance avec une limite supérieure concave et souvent floue, remontant vers l'extérieur | 36 |
| Figure 17 : Épanchement pleural | 37 |
| Figure 18: Epanchement pleural gauche complet compressif | 37 |
| Figure 19 : Pneumothorax..... | 39 |
| Figure 20 : L'épaississement pleural..... | 40 |

| | |
|---|----|
| Figure 21 : Syndrome interstitiel entraînant une opacité en verre dépoli..... | 42 |
| Figure 22 : La fibrose pulmonaire..... | 43 |
| Figure 22 : Tumeur pulmonaire sous forme de nodules: (a) cas d'une nodule unique (b) cas de nodules multiples..... | 44 |
| Figure 24 : Nodule pulmonaire..... | 45 |
| Figure 25: Tumeur pulmonaire sous forme de masse..... | 45 |
| Figure 26 : Masse pulmonaire | 46 |
| Figure 27 : Les images cavitaires : (a) image radiographique (b) image CT | 47 |
| Figure 28 : Aspect de miliaire des images cavitaires : (a) image radiographique (b) image CT..... | 47 |
| Figure 29 : La cardiomégalie : e plus grand diamètre horizontal du cœur et le plus grand diamètre horizontal du thorax, mesurés sur une radiographie thoracique conventionnelle | 49 |
| Figure 30 : Cliché montrant une cardiomégalie..... | 49 |
| Figure 31 : Une hernie hiatale (a), (b) Volumineuse | 51 |
| Figure 32 : Schéma récapitulatif du fonctionnement du Machine Learning..... | 60 |
| Figure 33: Schéma récapitulatif du fonctionnement du Deep Learning | 61 |
| Figure 34 : Graphique représentant l'augmentation de la performance d'un système d'apprentissage avec le nombre d'échantillons..... | 62 |
| Figure 35 : Le processus typique du ML..... | 66 |
| Figure 37 : Le sur-apprentissage..... | 68 |
| Figure 38 : La validation croisée | 69 |
| Figure 39 : Régression linéaire représentant prix des maisons en fonction de leur surface | 70 |
| Figure 40 : L'algorithme k -means regroupe les données en k cluster, ici $k = 3$. Les centres de gravité sont représentés par de petit cercle | 73 |

| | |
|---|-----|
| Figure 41 : Création de B bootstrap à partir des exemples d'apprentissage. | 76 |
| Figure 42 : Recherche d'un hyperplan qui divise les observations en deux catégories.(a) les hyperplans possibles avant optimisation, (b) l'hyperplan optimal..... | 78 |
| Figure 44 : Exemple d'un problème non linéairement séparable. La courbe devient une bande linéaire après avoir appliqué la transformation φ non linéaire | 79 |
| Figure 44 : Fonctionnement d'un neurone formel. | 80 |
| Figure 44 : Structure d'un PMC composé d'une couche d'entrée, deux couches cachées et une couche de sortie..... | 80 |
| Figure 45 : La relation entre l'intelligence artificielle, le ML et le deep learning..... | 82 |
| Figure 46 : La différence de performance entre le Deep Learning et la plupart des algorithmes de ML en fonction de la quantité de données..... | 85 |
| Figure 47 : Le processus du ML classique comparé à celui du Deep Learning | 86 |
| Figure 48: Deep Learning vs Machine Learning dans ImageNet challenge | 86 |
| Figure 49 : Résultat d'une convolution 2D | 90 |
| Figure 51: Sparse interactions (a) s est formé par convolution avec un noyau de largeur 3, (b) s est formé par une multiplication matricielle | 91 |
| Figure 50: Exemple image 1000x1000. (a) résultat du non paramètre sharing (b) résultat du paramètre sharing | 92 |
| Figure 52 : exemple d'applications du Pooling, (à gauche) Average pooling, le Max pooling (à droite) | 94 |
| Figure 49 Diagramme blocks d'un réseau de neurones convolutif..... | 95 |
| Figure 53 : Le taux d'erreur dans ImageNet Visual recognition Challenge. Le Deep Learning dépasse la performance humaine..... | 96 |
| Figure 54 : Les oscillations causées par SGD..... | 98 |
| Figure 56 : La fonction d'activation ReLu (à gauche), (à droite) la fonction Leaky ReLu et Parametric ReLu qui corrige l'inconvénient de ReLu. | 105 |

| | |
|---|-----|
| Figure 57 : Les fonctions d'activation Sigmoidale (a) et tangente hyperbolique (b). ... | 106 |
| Figure 58: Processus du modèle proposé | 110 |
| Figure 59.a: architecture du réseau VGG-16 | 111 |
| Figure 59.b: Modèle VGG16 avec des couches formables et des couches gelées | 112 |
| Figure 59.c : Architecture du réseau VGG19 | 112 |
| Figure 60: Modèle MobileNet avec des couches formables et des couches gelées ... | 113 |
| Figure 61.a : Architecture du réseau de neurones à convolution Inception V3. | 114 |
| Figure 61.b: Modèle InceptionV3 avec des couches formables et des couches gelées. | 114 |
| Figure 62.a : Convolution séparable en profondeur | 115 |
| Figure 63.b: Modèle Inception-ResNetV2 avec des couches formables et des couches gelées..... | 117 |
| Figure 64.a : Architecture du réseau DenseNet121 | 117 |
| Figure 64.b: Modèle DenseNet121 avec des couches formables et des couches gelées. | 118 |
| Figure 65: Modèle d'empilement utilisant sept différentes architectures pré-entraînées et vote à la majorité. | 119 |
| Figure 66 : Processus suivi par l'approche du Grad-CAM | 120 |
| Figure 67 : Lien vers la base NIH | 124 |
| Figure 68 : Exemples d'images stockées dans la base NIH | 125 |
| Figure 69: Répartition des pathologies selon le sexe | 130 |
| Figure 70: Répartition des pathologies simples et Multiples | 130 |
| Figure 71: Lésions concomitantes les plus fréquentes avec l'atélectasie | 131 |
| Figure 72: Lésions concomitantes les plus fréquentes avec la Cardiomegalie | 131 |
| Figure 73: Lésions concomitantes les plus fréquentes avec « Consolidation »..... | 132 |
| Figure 74: Lésions concomitantes les plus fréquentes avec l'œdème..... | 132 |

| | |
|---|-----|
| Figure 75 : Lésions concomitantes les plus fréquentes avec la pleurésie 'Effusion' . | 133 |
| Figure 76 : Lésions concomitantes les plus fréquentes avec l'emphysème | 133 |
| Figure 77 : Lésions concomitantes les plus fréquentes avec la Fibrose | 134 |
| Figure 78 : Lésions concomitantes les plus fréquentes avec l'Infiltration pulmonaire | 134 |
| Figure 79 : Lésions concomitantes les plus fréquentes avec les tumeurs (nommées Masse)..... | 135 |
| Figure 70 : Lésions concomitantes les plus fréquentes avec les Nodules | 135 |
| Figure 71 : Lésions concomitantes les plus fréquentes avec l'épaississement..... | 136 |
| Figure 72 : Lésions concomitantes les plus fréquentes avec la Pneumonie | 136 |
| Figure 73 : Lésions concomitantes les plus fréquentes avec le Pneumothorax | 137 |
| Figure 74: Carte graphique NVIDIA Quadro M2200 utilisée pour l'apprentissage.. | 140 |
| Figure 75: Exemple d'images radiographiques à pré traiter | 142 |
| Figure 76: Résultat de l'exécution de notre scripte de prétraitement sur une image | 143 |
| Figure 77 : Exemple de radiographie thoracique due la ere expérience : (a) Masse, (b) Nodule et (c) Non Défini..... | 146 |
| Figure 79: Courbes de perte d'apprentissage et de précision du réseau CNN DenseNet121: (a) courbe de perte (b) courbe de précision | 149 |
| Figure 80: Courbes de perte d'apprentissage et de précision du réseau CNN ResNet50: (a) courbe de perte (b) courbe de précision..... | 150 |
| Figure 81: Courbes de perte d'apprentissage et de précision du réseau CNN InceptionV3: (a) courbe de perte (b) courbe de précision..... | 150 |
| Figure 82: Courbes de perte d'apprentissage et de précision du réseau CNN MobileNet: (a) courbe de perte (b) courbe de précision..... | 151 |

| | |
|--|-----|
| Figure 83 : Courbes ROC obtenues par le réseau Xception pour les classes : Masse, Nodule et Non défini..... | 152 |
| Figure 85: Courbes ROC obtenues par le réseau DENSENet21 pour les classes : Masse, Nodule et Non défini..... | 152 |
| Figure 86: Courbes ROC obtenues par le réseau ResNet50 pour les classes : Masse, Nodule et Non défini..... | 153 |
| Figure 87: Courbes ROC obtenues par le réseau InceptionV3 pour les classes : Masse, Nodule et Non défini..... | 153 |
| Figure 88: Courbes ROC obtenues par le réseau MobileNet pour les classes : Masse, Nodule et Non défini..... | 154 |
| Figure 89: Courbes ROC obtenues par le réseau Ensemble Model pour les classes : Masse, Nodule et Non défini..... | 154 |
| Figure 90: Exemple de radiographie thoracique due la 2 ^{ème} expérience : (a) Masse, (b) Nodule, (c) Effusion et (d) Non Défini..... | 156 |
| Figure 91: Courbes de perte d'apprentissage et de précision du réseau CNN Xception: (a) courbe de précision (b) courbe de perte..... | 159 |
| Figure 92: Courbes de perte d'apprentissage et de précision du réseau CNN DenseNet121: (a) courbe de perte (b) courbe de précision | 159 |
| Figure 93: Courbes de perte d'apprentissage et de précision du réseau CNN ResNet50: (a) courbe de perte (b) courbe de précision..... | 160 |
| Figure 94: Courbes de perte d'apprentissage et de précision du réseau CNN InceptionV3: (a) courbe de perte (b) courbe de précision..... | 160 |
| Figure 95: Courbes de perte d'apprentissage et de précision du réseau CNN MobileNet: (a) courbe de perte (b) courbe de précision..... | 161 |
| Figure 96: Courbes ROC obtenues par le réseau Xception pour les classes : Masse, Nodule, Effusion et Non défini..... | 162 |

| | |
|---|-----|
| Figure 97: Courbes ROC obtenues par le réseau DenseNet121 pour les classes : Masse, Nodule, Effusion et Non défini..... | 162 |
| Figure 98: Courbes ROC obtenues par le réseau ResNet50 pour les classes : Masse, Nodule, Effusion et Non défini..... | 163 |
| Figure 99 Courbes ROC obtenues par le réseau Inception V3 pour les classes : Masse, Nodule, Effusion et Non défini..... | 163 |
| Figure 100: Courbes ROC obtenues par le réseau MobileNet pour les classes : Masse, Nodule, Effusion et Non défini..... | 164 |
| Figure 101: Courbes ROC obtenues par le réseau Ensemble Modèle pour les classes : Masse, Nodule, Effusion et Non défini | 164 |
| Figure 102: la courbe de précision (en haut) et de perte (en bas) sur 25 époques pour ResNet50..... | 166 |
| Figure 103: Courbe ROC avec résultat du score AUC du modèle ResNet50 sur les14 pathologies..... | 167 |
| Figure 104: Courbe ROC micro-moyenne pour le modèle ResNet50..... | 167 |
| Figure 105: la courbe de précision (en haut) et de perte (en bas) sur 25 époques pour ResNet101..... | 168 |
| Figure 107: Courbe ROC avec résultat du score AUC du modèle ResNet101 sur les14 pathologies..... | 168 |
| Figure 108: Courbe ROC micro-moyenne pour le modèle ResNet101..... | 169 |
| Figure 109: Courbe ROC avec résultat du score AUC du modèle DenseNet121 sur les 14 pathologies..... | 170 |
| Figure 110: Courbe ROC micro-moyenne pour le modèle DenseNet121..... | 170 |
| Figure 111 : Exemple de carte thermique Grad-CAM pour chaque classe de maladie pulmonaire | 175 |

LISTE DES TABLEAUX

| | |
|---|------------|
| TABLEAU 1 : LES ÉTAPES MAJEURS DU DEEP LEARNING | 83 |
| TABLEAU 2: RÉPARTITION DES DONNÉES IMAGES DE L'EXPÉRIENCE 1..... | 147 |
| TABLEAU 3: CONFIGURATION DE L'ARCHITECTURE DE CHAQUE RÉSEAU CNN UTILISÉ DANS LE MODÈLE PROPOSÉ | 148 |
| TABLEAU 4: PERFORMANCE (%) DE NOS MODÈLES SUR L'ENSEMBLE DE TEST..... | 155 |
| TABLEAU 5: RÉPARTITION DES DONNÉES IMAGES DE L'EXPÉRIENCE 1..... | 157 |
| TABLEAU 6: CONFIGURATION DES ARCHITECTURE DE CHAQUE RÉSEAU CNN UTILISÉ DANS LE MODÈLE PROPOSÉ | 158 |
| TABLEAU 7: PERFORMANCE (%) DE NOS MODÈLES SUR L'ENSEMBLE DE TEST..... | 165 |
| TABLEAU 8: RESULTATS AUC POUR TOUTES NOS EXPERIENCES. | 171 |
| TABLEAU 9: COMPARAISON ENTRE NOS TRAVAUX ET CEUX DE LA LITTÉRATURE..... | 184 |

INTRODUCTION GÉNÉRALE

Les travaux présentés dans ce rapport de thèse s'inscrivent dans le domaine de l'intelligence artificielle appliquée en médecine et plus précisément la création d'un système intelligent d'aide à la décision clinique (SADC), accessible et peu coûteux pour le diagnostic des lésions pulmonaires chez les populations des zones rurales et isolées à partir de radiographies pulmonaires.

1. CONTEXTE ET PROBLÉMATIQUE

En imagerie médicale, la radiographie est utilisée pour le diagnostic de diverses pathologies. Les médecins utilisent cette technique depuis plusieurs décennies pour explorer et visualiser les lésions dans les différentes parties du corps humain. Cela est dû au fait que les rayons X sont un moyen de diagnostic très efficace pour révéler les altérations pathologiques, en plus d'être non invasifs et peu coûteux à produire.

Les médecins sont régulièrement confrontés à la réalisation et à l'interprétation d'exams radiologiques de base dans leur cabinet. Encore aujourd'hui, les radiographies les plus fréquemment pratiquées sont la radiographie du thorax ainsi que les clichés réalisés pour des problématiques ostéo-articulaires, en particulier des extrémités.

Ainsi, les lésions thoraciques peuvent être visualisées sur les radiographies thoraciques sous différents aspects qui seront par la suite analysés par un médecin. La classification de ces anomalies est cependant considérée comme une tâche difficile. Par conséquent, au cours des dernières décennies, des systèmes de diagnostic assistés par ordinateur ont été développés afin d'aider les médecins. Cependant, ces systèmes n'ont pas atteint un niveau de précision significatif pour prendre des décisions de manière autonome, c'est pourquoi ils ne font qu'aider et orienter les praticiens.

L'émergence de l'intelligence artificielle (IA) dans le domaine médical, dont les premiers concepts datent des années 1950 est la conséquence de trois

bouleversements radicaux : la numérisation des images médicales permettant leur paramétrage, le développement des algorithmes autorisant l'utilisation des données saisies en langage naturel, et l'apprentissage profond (deep learning) permettant à partir de données radiologiques massives de développer des algorithmes de traitement automatique d'images médicales. Ces systèmes permettent dès aujourd'hui la détection automatique de lésion et ouvrent la voie au dépistage du cancer du poumon, du sein ou de la prostate. Leur fiabilité est parfois supérieure aux capacités médicales humaines [1–5].

Récemment, de nouvelles méthodes de reconnaissance très précises ont vu le jour et sont basées sur l'apprentissage profond [6–10]. Ces réseaux profonds ont montré des précisions le plus souvent surhumaines. Ce succès nous a motivé à utiliser ces réseaux pour l'analyse des radiographies thoraciques et les résultats préliminaires ont montré que les réseaux profonds peuvent extraire efficacement les caractéristiques utiles pour distinguer les différentes lésions.

Xiaosong Wang, dans son travail [11], a utilisé des réseaux neuronaux basés sur l'apprentissage profond pour classer les maladies thoraciques les plus courantes. Deux auto-encodeurs empilés sont examinés dans cette étude pour classer les radiographies thoraciques en deux classes communes : normale et anormale avec différents types de pathologies thoraciques, à savoir, l'atélectasie, la cardiomégalie, l'épanchement, l'Infiltration, la masse ou nodule, la pneumonie, le pneumothorax, l'œdème, l'emphysème, la fibrose, l'épaississement pleural et la hernie diaphragmatique.

2. APPROCHE PROPOSÉE

Dans cette thèse nous utiliserons le Deep Learning pour extraire les caractéristiques graphiques des différentes lésions à partir des images radiographiques thoraciques. Ces types d'images sont plus accessibles pour les patients que les images tomendensitogrammes (scanners) en particulier dans les zones rurales et isolées.

L'architecture que nous allons proposer est basée sur l'apprentissage profond (Deep Learning) par des réseaux de neurones à convolutions que nous allons exploiter pour extraire les caractéristiques graphiques de la maladie à partir des images radiographiques thoraciques. Ces caractéristiques extraites nommées descripteurs seront utilisées pour entraîner notre modèle proposé afin de fournir un diagnostic clinique précoce économisant ainsi un temps critique pour le contrôle des maladies. Pour cela nous allons utiliser l'apprentissage sur une base d'images thoracique en utilisant plusieurs architectures de réseaux neuronaux à convolution tels que le réseau à blocs résiduels (ResNet), le réseau à blocs denses (DenseNet) et le réseau du groupe de recherche VGG (Visual Geometry Group).

Pour augmenter la précision de notre modèle, nous allons choisir dans un premier temps des réseaux déjà pré entraînés sur la base d'images ImageNet constituée par des millions d'images. Dans un deuxième temps nous allons utiliser la notion de transfert d'apprentissage pour construire une nouvelle architecture appelée ensemble modèle (Stacking Model) [7] dans laquelle nous allons transmettre les poids des réseaux déjà pré entraînés aux réseaux du modèle proposé où seule la dernière couche de chaque réseau sera utilisée pour un apprentissage sur notre base d'images radiographiques. De plus le modèle proposé effectuera un tri des différentes prédictions issues de chaque réseau de telle sorte que seule la prédiction qui a obtenu le plus grand score sera prise en considération comme prédiction finale. De cette façon

nous allons augmenter la précision de notre modèle par rapport à la précision de chaque réseau CNN.

Les données radiologiques utilisées seront collectées à partir de la base de donnée du NIH.

Afin de pouvoir suivre la prédiction de notre système en fonction du nombre de classe présentes dans la base du NIH nous avons construit trois ensembles d'images. Un ensemble formé de 3 classes, un ensemble formé de 4 classes et un ensemble formé de 14 classes. Nous avons ainsi mené une première expérience de reconnaissance automatique d'image sur le premier ensemble constitué de 3 classes : Nodule, Masse, non défini. Ensuite nous avons mené une deuxième expérience de reconnaissance sur le deuxième ensemble formé de quatre classes : Nodule, Masse, Effusion, non défini. Enfin une troisième expérience a été menée sur l'ensemble contenant les 14 classes. Pour chaque expérience nous avons mesuré plusieurs critères de performance pour caractériser le pouvoir de reconnaissance de notre modèle proposé.

3. PLAN DU MANUSCRIT

La présente thèse est composée de quatre chapitres. Le premier chapitre est un récapitulatif de la sémiologie radiographique du thorax, décrivant les pathologies intra-thoraciques qu'on va étudier. Le second chapitre est un descriptif du procédé d'intelligence artificielle, ses principes de fonctionnement, en plus de son algorithme de formation. Le troisième chapitre présente la phase de formation et de test du réseau profond. Dans ce chapitre, le réseau est formé sur la base de données disponible sur le web et testé par la même occasion. Les résultats seront rapportés et discutés à partir des simulations du réseau dans le chapitre quatre.

Enfin, la conclusion générale présente une synthèse des contributions apportées ainsi que les pistes définissant des perspectives possibles pour des travaux futurs.

CHAPITRE I

LA RADIOGRAPHIE

THORACIQUE

1. INTRODUCTION

Un test radiologique thoracique est un test de radiologie normal et non invasif qui crée une image de la poitrine et des organes thoraciques. Pour effectuer ce test, le thorax est rapidement soumis au rayonnement d'une machine à rayons X et une image est créée sur un film ou dans un ordinateur informatisé. En fonction de son épaisseur, chaque organe à l'intérieur de la cage thoracique possède des degrés de d'absorption variables, créant des ombres distinctives sur le film.

On distingue ainsi quatre densités radiologiques qui sont de la plus absorbante à la moins absorbante : osseuse, hydrique, grasseuse et aérique. Ainsi les os absorbant les rayons X de manière importante apparaîtront blancs ce qui en terme radiologique se nomme opaque. L'air absorbant peu les rayons X apparaîtra noir ce qui en terme radiologique se dit clair.



Figure 1 : Les 4 densités radiologiques

On parle :

D'opacité pour toute image apparaissant blanche sur le cliché :



Figure 2 : Opacité apparaissant blanche sur le cliché

De clarté pour toute image noire :



Figure 3 : Clarté apparaissant noire sur le cliché

Voici un exemple de radiographie thoracique :



Figure 4 : Radiographie thoracique normale.

2. QUALITÉ DU CLICHÉ

Les critères pour vérifier qu'il est de bonne qualité sont les suivants :

- La distance séparant le bord interne des clavicules aux épineuses est égale à droite et à gauche et l'épineuse de la 3ème vertèbre thoracique est centrée (ainsi le cliché est bien de face).
- Il existe un niveau hydro-aérique dans la poche gastrique (ainsi le patient est bien debout).
- La coupole diaphragmatique droite est au niveau ou sous la partie antérieure du sixième arc costal et les culs de sac costo-diaphragmatiques sont bien visibles (ainsi le cliché a été réalisé en inspiration profonde).
- Le rachis et les vaisseaux sont visibles derrière le cœur (ainsi l'exposition est correcte).
- Les omoplates doivent être bien dégagées.

3. APERÇU DE LA SÉMILOGIE RADIOLOGIQUE NÉCESSAIRE A

NOTRE ÉTUDE

Dans cette section, nous allons définir les notions de sémiologie radiologique nécessaires à notre étude.

4. LES SYNDROMES PARENCHYMATEUX :

4.1. Syndrome alvéolaire

Le syndrome alvéolaire traduit le comblement des alvéoles pulmonaires par du liquide (eau, pus, sang) et/ou de cellules. Il se révèle par une opacité :

- De densité hydrique (de même densité que celle du cœur)
- A limites floues : reflet de la superposition de groupes d'alvéoles remplis alors que d'autres sont encore aérés

- Ayant tendance à confluer en raison d'une diffusion de proche en proche du liquide à travers les pores de Kohn et les canaux de Lambert (petits pertuis faisant communiquer les alvéoles entre elles)
- Systématisée, correspondant à un territoire anatomique bien délimité (segment de lobe, lobe, plus exceptionnellement tout un poumon). L'opacité est volontiers limitée par une scissure, souvent infranchissable (**Figure 5.a**). Le cliché de profil confirme ce caractère systématisé (**Figure 5.b**). Le volume du territoire condensé reste normal à la différence des opacités liées à un trouble de ventilation ou la rétraction donne une perte de volume.

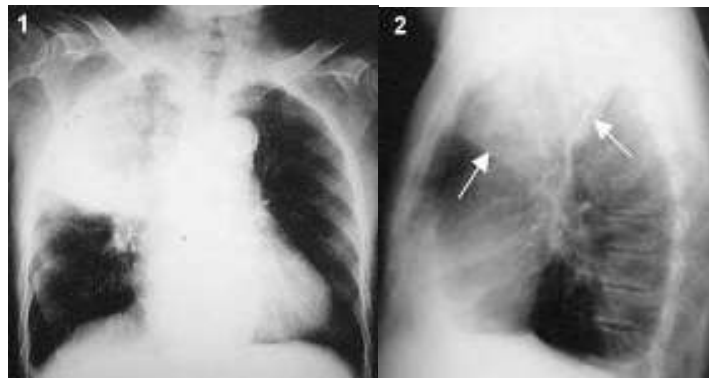


Figure 5 : Comblement des alvéoles pulmonaires par du liquide systématisé : (a) vue latérale , (b) vue de profile

- Siège d'un bronchogramme aérien : au sein des alvéoles comblées, les bronches demeurent souvent aérées. Ainsi elles apparaissent comme des hyperclartés, bifurquant du hile vers la périphérie (Figure 6)



Figure 6. Comblement des alvéoles pulmonaires par du liquide avec un bronchogramme aérien

4.1.1. Pneumonie

A. Définition

La pneumonie est une infection des poumons causée le plus souvent par un virus ou une bactérie. L'infection touche plus précisément les alvéoles pulmonaires, ces minuscules sacs en forme de ballons situés à l'extrémité des bronchioles. Elle touche généralement un seul des 5 lobes du poumon (3 lobes dans le poumon droit et 2 dans le gauche), d'où le terme pneumonie lobaire. Lorsque la pneumonie atteint aussi les bronches, on l'appelle bronchopneumonie. La pneumonie lobaire est caractérisée par un remplissage des espaces alvéolaires par un exsudat inflammatoire, avec peu ou pas de dommage tissulaire.

B. Aspect radiographique

L'aspect typique de pneumopathie aiguë communautaire (PAC) est celui d'une condensation alvéolaire unique, à limites floues, sous-pleurale, bien limitée en juxta-scissural, évoluant vers une opacité systématisée segmentaire, affectant un ou plusieurs segments contigus ou un lobe, avec ou sans bronchogramme aérien (Figure 7).

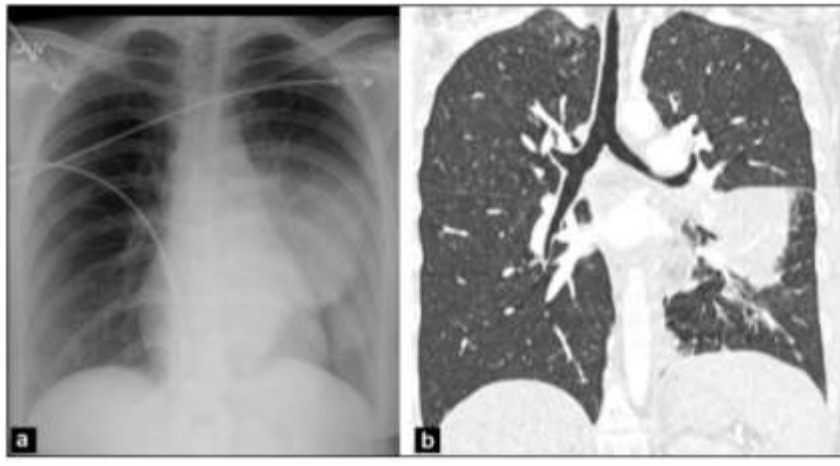


Figure 7: Bronchogramme aérique : (a) radiographie pulmonaire (b) CT image

C. Etiologies

Presque toutes les pneumonies sont causées par des germes, virus ou bactéries. Lorsque le microbe attaque les poumons, le corps réagit en déclenchant une réaction d'inflammation. Les alvéoles se remplissent de pus et de liquide inflammatoire, ce qui entraîne des difficultés respiratoires.

4.1.1.1. Œdème pulmonaire

A. Définition

Il s'agit d'une extravasation brutale de liquide plasmatique dans l'interstitium et les alvéoles avec une augmentation de l'eau pulmonaire extravasculaire. On distingue deux types d'OAP : L'OAP hémodynamique où l'issue du liquide est liée à l'augmentation de la pression capillaire pulmonaire et l'OAP lésionnel où l'extravasation est plutôt en rapport avec une augmentation de la perméabilité au niveau de la membrane alvéolo-capillaire.

B. Aspect radiographique

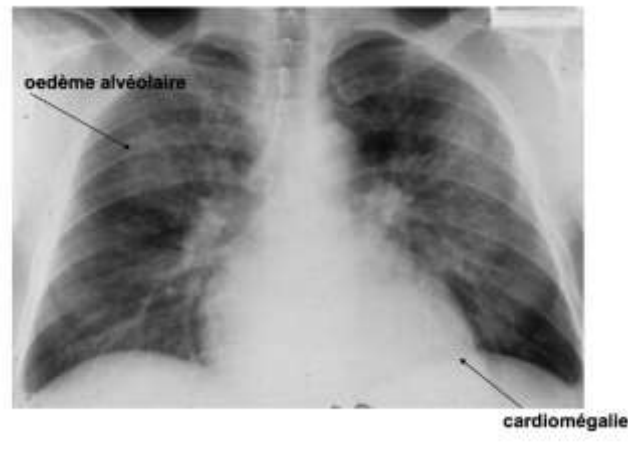


Figure 8 : Oedeme alvéolaire

C. Etiologies

- Cardiopathies hypertensives : HTA essentielle, maligne, éclampsie
- Cardiopathies ischémiques : 60%
- Poussée d'insuffisance rénale
- Cardiopathie chronique : rechercher systématiquement un facteur déclenchant : Infection, fièvre, trouble du rythme, écart de régime ou de traitement, pic d'HTA, ischémie myocardique, intervention chirurgicale, anémie, hypoprotidémie, grossesse
- Myocardites aiguës : toxiques, infectieuses
- Valvulopathies aiguës : ischémiques ou infectieuses
- Les obstructions et les désinsertions de prothèses valvulaires
- Troubles du rythme cardiaque et de la conduction A-V
- L'hypothermie profonde, envenimation scorpionique

Pour l'OAP de type lésionnel, les étiologies sont multiples et dépendent d'un contexte particulier

4.2. L'emphysème

A. Définition

L'emphysème est une maladie pulmonaire entrant dans le cadre des BPCO caractérisé par l'augmentation permanente du volume des espaces aériens situés au-delà de la bronchiole terminale, accompagnée d'une destruction de leurs parois, sans cicatrisation fibreuse.

Sont atteints les trois ordres de bronchioles respiratoires, les canaux alvéolaires et les alvéoles (toutes les structures qui sont au-delà de la bronchiole terminale).

B. Aspect radiographique

Il se traduit par :

- Une hyperclarté diffuse par hypovascularisation
- Une distension thoracique avec :
- Un aplatissement des coupes diaphragmatiques
- Un élargissement des espaces intercostaux
- Une augmentation des espaces clairs, rétro-sternal (*) et rétrocardiaque (+) .(Figure 9)..

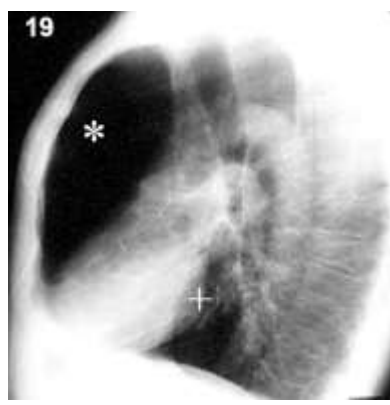


Figure 9 : augmentation des espaces clairs, rétro-sternal (*) et rétrocardiaque (+)



Figure 10 : Emphyseme pulmonaire

C. Etiologies

Les deux plus grandes causes de l'emphysème pulmonaire sont :

- Le tabagisme (grande majorité des cas et donnant surtout des emphysèmes centrolobulaires et paraseptaux) ;
- Déficience en alpha 1-antitrypsine d'origine génétique (très rare) donnant des emphysèmes panlobulaires (Figure 10).

4.3. Atélectasie

A. Définition

L'atélectasie (Figure 11) est une réduction de volume pulmonaire par collapsus des alvéoles. Ce trouble de la ventilation est lié soit à une obstruction endobronchique soit à une compression extrinsèque d'une bronche.

B. Aspects radiologiques

L'atélectasie forme une opacité à limites nettes, systématisée à un segment, un lobe ou tout un poumon selon le siège de l'obstacle bronchique.

Les signes radiologiques indirects sont la conséquence de la diminution de volume du parenchyme atteint :

- Déplacement de la scissure qui devient concave vers le parenchyme sain

- Attraction de diverses structures (hile, coeur, trachée)
- Ascension de la coupole diaphragmatique
- Pincement des espaces intercostaux

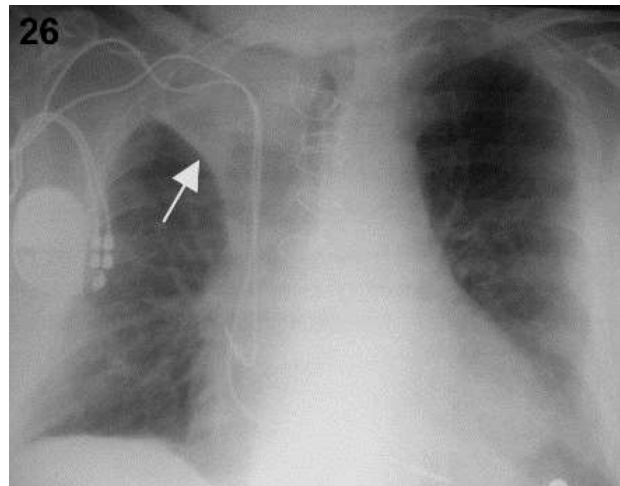


Figure 11. L'atélectasie

La TDM aide à préciser le siège et la cause de l'obstruction bronchique : corps étranger, tumeur, etc...

Les petites atélectasies planes, d'orientation horizontale, sont situées au-dessus des coupes diaphragmatiques (Figure 12). Elles correspondent à des sous-segments mal ventilés.

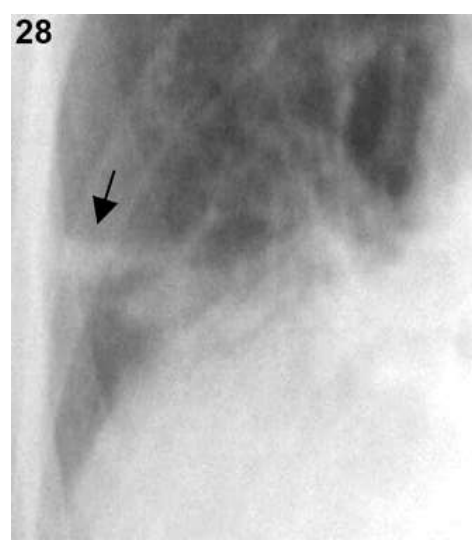


Figure 12 : Les petites atélectasies planes, d'orientation horizontale

Un exemple d'évolution vers une atélectasie complète du champ pulmonaire droit est montré sur la figure 13. On observe une rétraction progressive du poumon droit avec attraction de toutes les structures de voisinage, médiastin, coupole diaphragmatique droite et grill costal droit.



Figure 13: Séquence montrant l'évolution vers une atélectasie complète du champ pulmonaire droit.

C. Etiologies

Les facteurs les plus fréquents à l'origine d'une atélectasie comprennent les suivants:

- Obstruction intrinsèque des voies respiratoires (par un corps étranger, une tumeur, un bouchon muqueux)
- Compression extrinsèque des voies respiratoires (par une tumeur, ou une lymphadénopathie)
- Suppression de la respiration ou de la toux (par une anesthésie générale, une sédation, une douleur)
- Position en décubitus dorsal, en particulier chez les patients obèses et en cas de cardiomégalie
- Compression ou atteinte du parenchyme pulmonaire (par un grand épanchement pleural ou un pneumothorax)

Les chirurgies thoraciques et abdominales sont des causes très fréquentes car elles impliquent une anesthésie générale, l'utilisation d'opiacés (avec possibilité de dépression respiratoire secondaire), et souvent une respiration douloureuse. Une sonde endotrachéale mal positionnée peut causer une atélectasie par occlusion d'une bronche souche.

Des causes moins fréquentes d'atélectasie comprennent une dysfonction du surfactant pulmonaire et des cicatrices ou des tumeurs parenchymateuses.

5. SYNDROMES PLEURAUX

Il s'agit de l'ensemble des signes qui traduisent la présence de liquide, d'air ou de tissus anormaux entre les feuillets pleuraux.

5.1. Pleurésie

Une pleurésie peut être de faible abondance, de moyen abondance ou de grande abondance.

A. Pleurésie de faible abondance

De face, elle réalise une opacité qui émousse les culs de sac (Figure 14). De profil, elle forme un comblement dans le cul de sac postérieur. L'échographie pleurale peut aider à confirmer ou à rechercher ces épanchements de faible abondance.

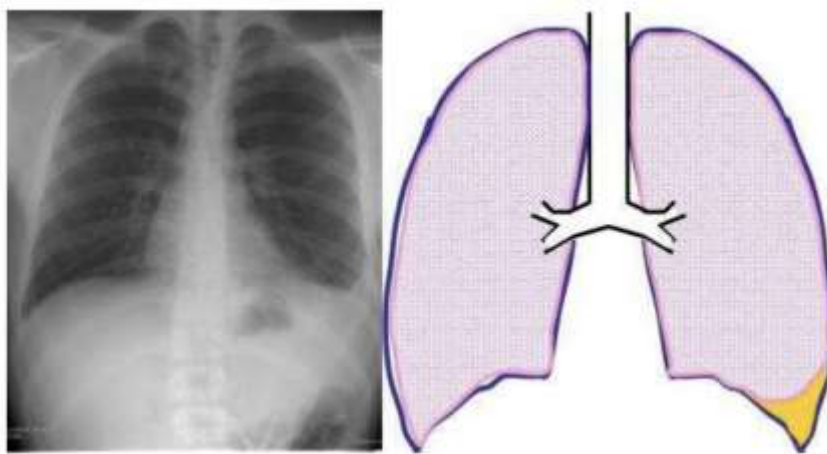


Figure 14 Comblement du cul de sac pleural gauche

B. De moyenne abondance

En position debout, l'épanchement pleural liquidien se collecte au niveau des bases sous la forme d'une opacité en nappe (Figure 15) qui :

- De face, s'étend du médiastin en dedans à la partie latérale du thorax en dehors
- De profil, va de la partie antérieure rétro-sternale jusqu'au rachis
- N'est pas systématisée
- Ne contient pas de bronchogramme aérien
- Déclive en position debout et mobile avec les changements de position du malade ; en position couchée, l'épanchement se caractérise par une asymétrie de transparence des poumons avec vascularisation pulmonaire conservée.

- En haut, a une limite supérieure concave et souvent floue, remontant vers l'extérieur là où le poumon se laisse plus facilement comprimer, réalisant la ligne de Damoiseau (Figure 16).
- En bas, efface la coupole diaphragmatique.

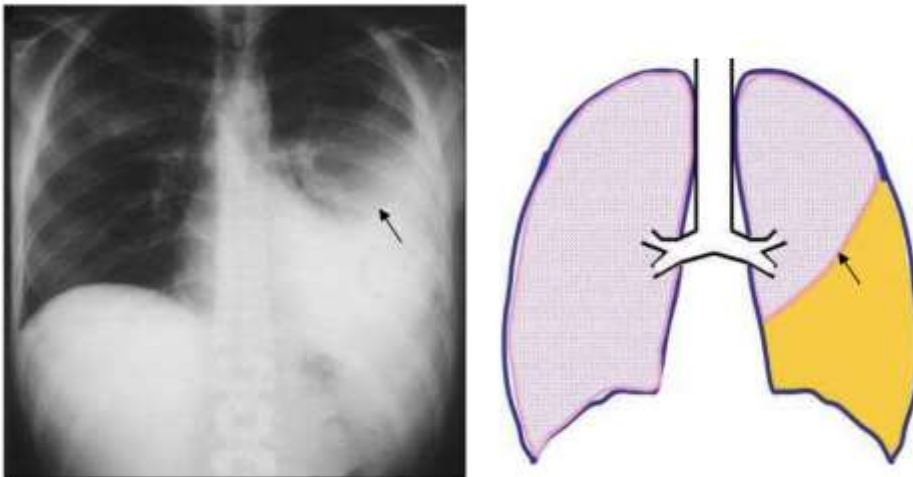


Figure 15: Epanchement pleural gauche de moyenne abondance

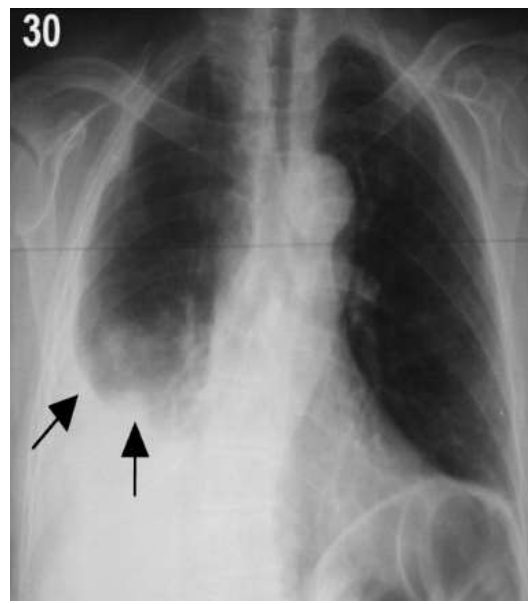


Figure 16 : pleurésie à moyenne abondance avec une limite supérieure concave et souvent floue, remontant vers l'extérieur



Figure 17 : Épanchement pleural

C. De grande abondance

L'épanchement de grande abondance réalise un hémithorax opaque sans bronchogramme aérien, avec refoulement controlatéral du médiastin (Figure 18).

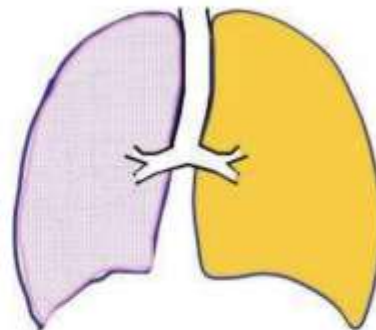


Figure 18: Epanchement pleural gauche complet compressif

D. Etiologies

- Transsudats :
 - Insuffisance Cardiaque Gauche
 - Cirrhose
 - Syndrome néphrotique
 - Atélectasie pulmonaire

- Embolie pulmonaire
- Exsudats
 - Exsudats néoplasiques
 - Exsudats infectieux
 - Embolie pulmonaire
 - Post traumatique
 - Rupture oesophagienne
 - Témoin d'une affection sous diaphragmatique
 - Auto-immune

5.2. Pneumothorax

L'air, en pénétrant de façon anormale dans la cavité pleurale, découple le poumon de la paroi thoracique ; grâce à son élasticité, le poumon se rétracte vers le hile tandis que la paroi thoracique se distend (Figure 19).

Il se traduit sur la radiographie thoracique par :

- Une hyperclarté gazeuse périphérique, limitée en dedans par la plèvre viscérale,
- Visible sous la forme d'un fin liseré et en dehors par la paroi thoracique
- La disparition des vaisseaux au-delà de ce liseré pleural
- La rétraction du moignon pulmonaire plus ou moins complètement au hile.

En cas de doute, un cliché thoracique en expiration forcée ou mieux sur un TDM permet le diagnostic.

La déviation médiastinale controlatérale, parfois associée à un aplatissement de la coupole diaphragmatique et à un élargissement des espaces intercostaux, est fréquemment observé dans les pneumothorax complets. Le pneumothorax compressif est défini comme un pneumothorax mal toléré cliniquement, quel que soit l'importance du décollement du poumon sur la radio du thorax et ce, même en l'absence de

déviations médiastinales controlatérales.

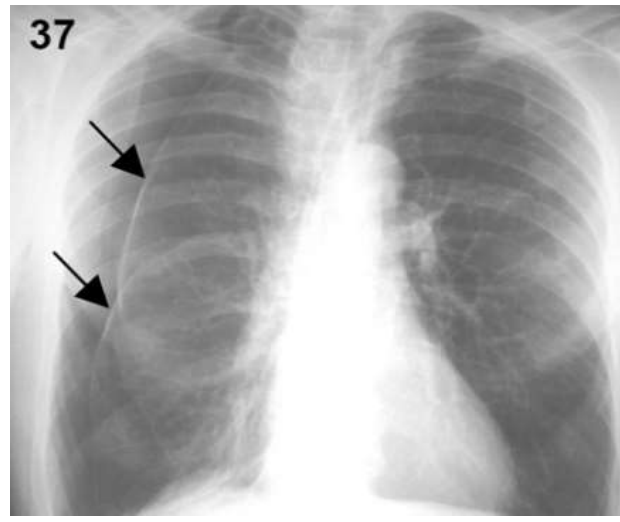


Figure 19 : Pneumothorax

5.3. Épaississement pleural

L'épaississement pleural est une réaction non spécifique à une agression dont les étiologies sont nombreuses et variées, de nature inflammatoire, infectieuse, ischémique, tumorale ou traumatique. La grande majorité des épaississements pleuraux sont secondaires à l'organisation et la résorption d'un épanchement pleural.

Il peut être diffus ou localisé et se traduit souvent par un effacement du cul de sac costophrénique latéral et/ou postérieur qu'il ne faut pas confondre avec un comblement du cul de sac par un épanchement liquidien de faible abondance.

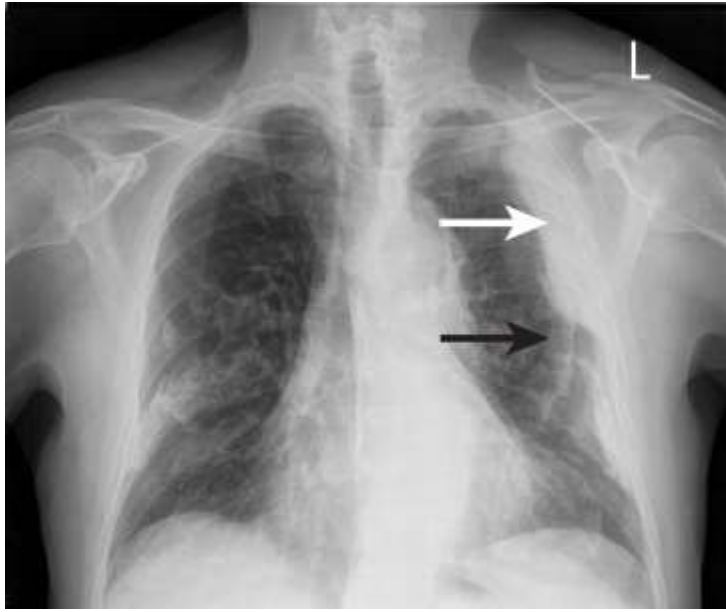


Figure 20 : L'épaississement pleural

E. Etiologies

L'épaississement pleural est une réaction non spécifique à une agression dont les étiologies sont nombreuses et variées, de nature inflammatoire, infectieuse, ischémique, tumorale ou traumatique. La grande majorité des épaississements pleuraux sont secondaires à l'organisation et la résorption d'un épanchement pleural. L'épaississement pleural peut aussi être localisé et suspendu, en particulier lorsqu'il est en regard d'un infarctus pulmonaire ou en regard d'une cavité pulmonaire préexistante où il peut être annonciateur de l'apparition d'une greffe aspergillaire. En cas de radiothérapie, il apparait entre 6 et 24 mois après l'irradiation et se localise en regard du champ irradié. Enfin, l'épaississement pleural peut être la traduction d'une colonisation par des cellules tumorales, envahissement à partir d'un cancer bronchopulmonaire, métastases pleurales ou mésothéliome.

6. SYNDROME INTERSTITIEL

Il traduit la présence de liquide et/ou de cellules anormales dans l'interstitium pulmonaire. Ces images peuvent s'associer mais une lésion interstitielle pure n'est

visible qu'en l'absence de syndrome alvéolaire surajouté.

Les lésions interstitielles sont des opacités à :

- Limites nettes
- Non confluentes
- Non systématisées
- Sans bronchogramme aérien
- De topographie ni vasculaire ni bronchique
- A évolution lente

6.1. Le verre dépoli

Le verre dépoli (figure 21) est une hyperdensité du parenchyme respectant les contours des vaisseaux pulmonaires et des parois bronchiques. Il peut être localisé, multifocal ou diffus, réalisant parfois un aspect en mosaïque.



Figure 21 : Syndrome interstitiel entraînant une opacité en verre dépoli

Cette image correspond à une hypertrophie inflammatoire du tissu interstitiel (parois alvéolaires et interstitium interlobulaire) qui devient radiologiquement individualisable.

Elle peut aussi correspondre au comblement partiel de la lumière de certaines alvéoles par des débris cellulaires ou par des membranes hyalines.

6.2. Fibrose pulmonaire

A. Définition

La fibrose pulmonaire (figure 22) est une lésion des poumons caractérisée par une fibrose, c'est-à-dire la présence d'un excès de tissu conjonctif fibreux. De nombreuses pathologies sont en cause, parmi les pneumopathies interstitielles.

B. Aspect radiographique

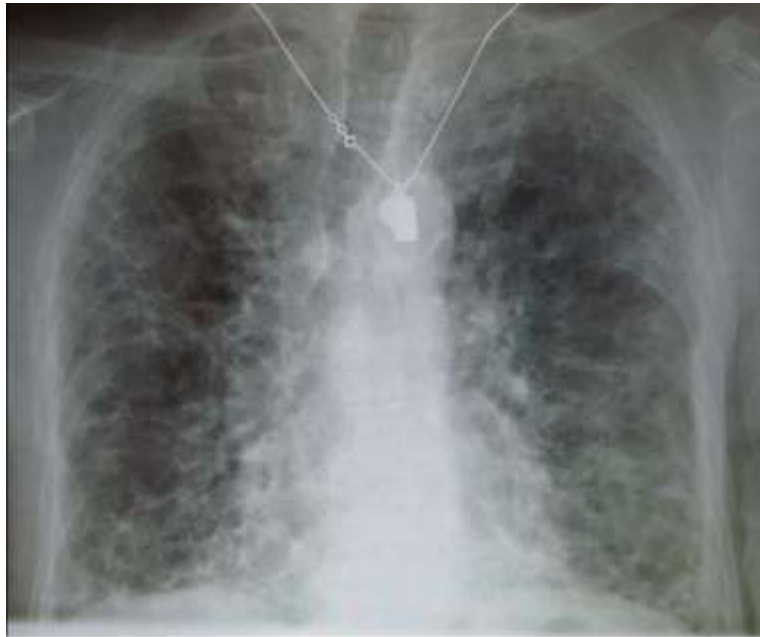


Figure 22 : La fibrose pulmonaire

C. Etiologies

C'est un stade évolutif de plusieurs maladies, la plupart d'entre elles étant classées dans les pneumopathies interstitielles.

Les maladies et les conditions qui peuvent causer la fibrose pulmonaire comprennent entre autres l'inhalation de polluants environnementaux ou sur le lieu de travail, comme dans le cas de l'asbestose ou de la silicose, ainsi que l'exposition à certains gaz; les mineurs, les marins et les sableurs, entre autres, sont des catégories à risque plus élevé. Une pneumopathie d'hypersensibilité résulte le plus souvent de l'inhalation de poussières contaminées par des sources bactériennes, fongiques ou animales. Le tabagisme peut accroître le risque ou aggraver les symptômes. Certaines maladies auto-immunes comme la polyarthrite rhumatoïde, le lupus érythémateux disséminé et la sclérodermie. D'autres maladies du tissu conjonctif comme le Syndrome d'Ehlers-Danlos, la sarcoïdose et la granulomatose de Wegener des infections ainsi que certains médicaments comme l'amiodarone, la bléomycine, le busulfan, le méthotrexate et le nitrofurantoïne et

la radiothérapie mammaire.

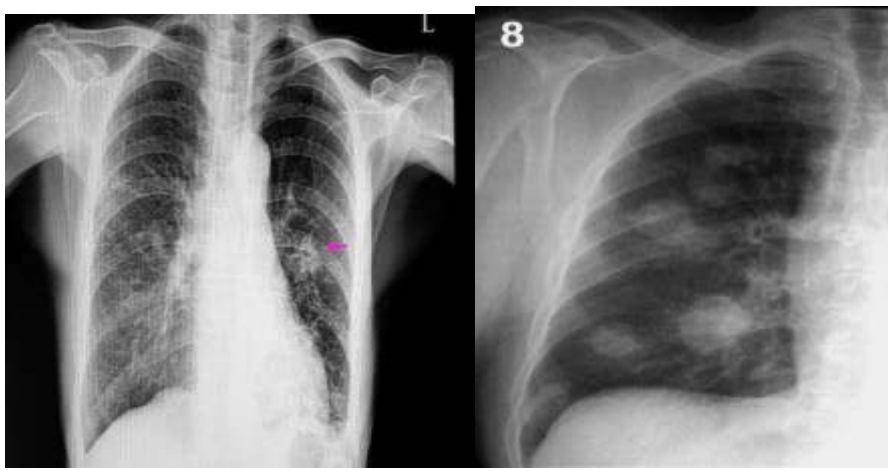
Dans certains cas, la fibrose pulmonaire peut apparaître sans cause connue. Il s'agit de pneumopathie interstitielle diffuse idiopathique, dont la principale forme est la fibrose pulmonaire idiopathique.

7. SYNDROME TUMORAL

7.1. Les nodules

A. Définition

Les nodules sont des opacités rondes dont le plus grand diamètre a une taille comprise entre 5 mm et 3 cm. Ils peuvent être uniques (figure 22) ou multiples réalisant un aspect dit de "lâcher de ballons".



(a)

(b)

Figure 22 : Tumeur pulmonaire sous forme de nodules: (a) cas d'une nodule unique
(b) cas de nodules multiples

B. Aspect radiographique (figure 24)

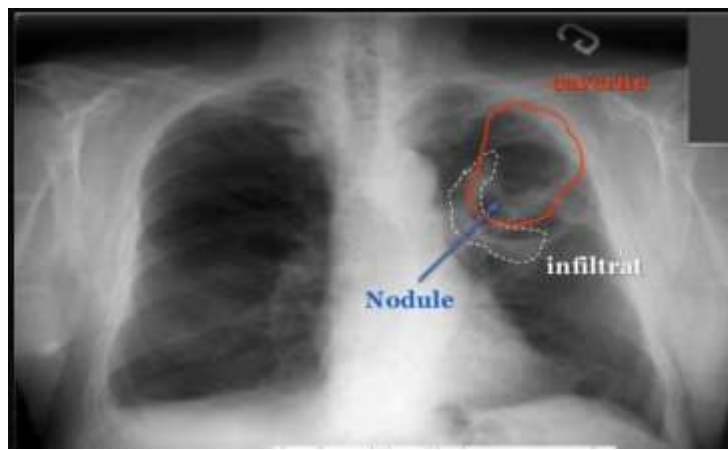


Figure 24 : Nodule pulmonaire

7.2. Masses

Les masses correspondent à une image arrondie ou ovalaire dont le diamètre est supérieur à 3 cm. Son contour est régulier ou spiculé. La TDM permet de juger de la topographie exacte de la masse et de son extension éventuelle aux structures médiastinales ou bien à la paroi. Elles sont parfois le siège de calcifications.



Figure 25: Tumeur pulmonaire sous forme de masse

A. Aspect radiographique (figure 26)

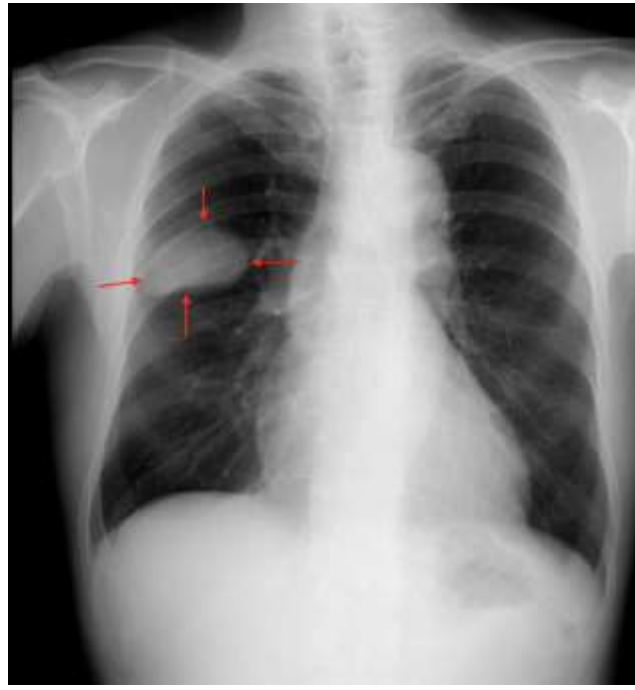


Figure 26 : Masse pulmonaire

B. Etiologie

- Néoplasique : Carcinome bronchique primitif, Métastase, Lymphome
- Tumeur bénigne : Hamartome, chondrome
- Infectieuse : Granulome, Abscess, Pneumonie chronique
- Inflammatoire
- Vasculaire : Infarctus, malformation artério-veineuse

7.3. Les images cavitaires

A. Définition

Les cavités (figure 27) résultent de la destruction du parenchyme. Leur paroi est épaisse permettant de les distinguer des kystes et des bulles dont la paroi est fine. Un niveau liquidien est parfois présent au sein de cette excavation. L'image en grelot est une opacité arrondie plus ou moins volumineuse située à l'intérieur de la cavité. Elle évoque la présence d'un champignon dans la cavité (aspergillus).

B. Aspect radiographique

Leur dissémination donne un aspect de miliaire. La tomographie, plus sensible, permet d'étudier leur répartition, leur profusion, leur rapport avec la plèvre, les bronches ou les vaisseaux. Parfois, ces micronodules sont calcifiés (figure 28).

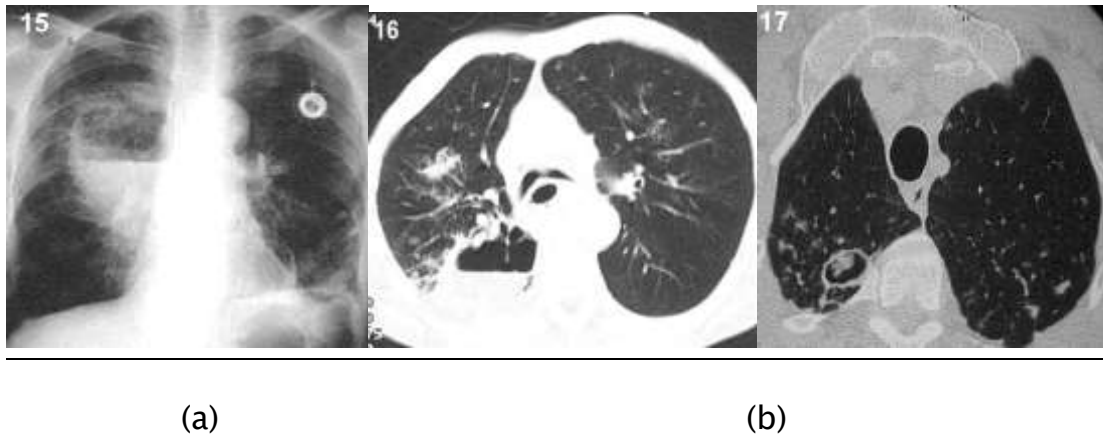


Figure 27 : Les images cavitaires : (a) image radiographique (b) image CT



Figure 28 : Aspect de miliaire des images cavitaires : (a) image radiographique (b) image CT

C. Étiologies

- Abscès pulmonaire
- Tuberculose et mycobactéries atypiques
- Kyste hydatique
- Pneumatocèles post-infectieux
- Métastases pulmonaires
- Bulles d'emphysème
- Bronchectasies kystiques
- Mycoses
- Lymphome
- Maladies auto-immunes

8. SYNDROME MÉDIASTINAL

8.1. Cardiomégalie

A. Définition

La cardiomégalie (figure 30) est une augmentation anormale de la taille du cœur. Elle peut être dépistée par une radiographie thoracique de face, en position debout, montrant un index cardio thoracique supérieur à 0,5 ou 0,55. Cet index est égal au rapport entre le plus grand diamètre horizontal du cœur et le plus grand diamètre horizontal du thorax, mesurés sur une radiographie thoracique conventionnelle.

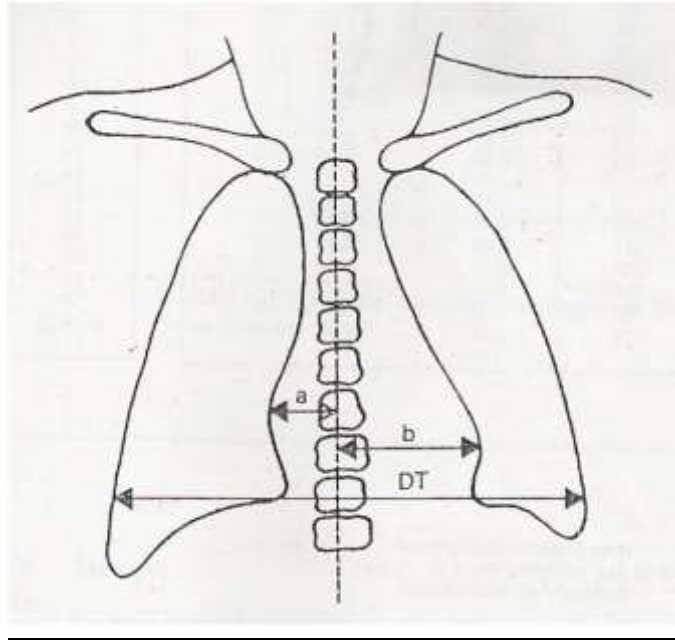


Figure 29 : La cardiomégalie : e plus grand diamètre horizontal du cœur et le plus grand diamètre horizontal du thorax, mesurés sur une radiographie thoracique conventionnelle

B. Aspect radiographique (figure 30)

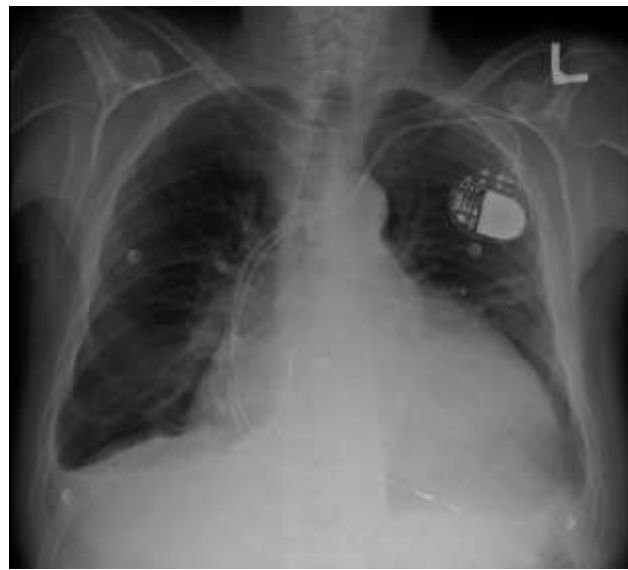


Figure 30 : Cliché montrant une cardiomégalie

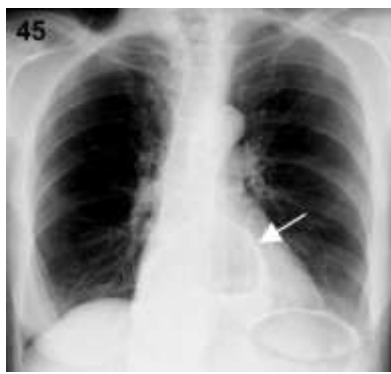
C. Etiologies

Un « gros cœur » peut être du, soit à une augmentation d'une ou plusieurs cavités cardiaques, par exemple dans le cadre d'une cardiomyopathie dilatée, soit à un épaississement important des parois des ventricules cardiaques (hypertrophie ventriculaire gauche dans le cadre d'une cardiomyopathie hypertrophique). Il peut également parfois témoigner d'un épanchement péricardique. Ces différents diagnostics sont tranchés par une échocardiographie.

8.2. Les hernies

Il s'agit de l'ascension d'organes abdominaux dans le thorax à travers un orifice diaphragmatique. Une volumineuse hernie hiatale (figure 31) est détectée par la radiographie thoracique sous la forme d'une opacité rétrocardiaque, parfois associée à un niveau hydroaérique très évocateur.

A. Aspect radiologique



(a)



(b)

Figure 31 : Hernie hiatale (a) et (b)

B. Etiologies

Il existe deux sortes de hernie hiatale, avec des causes probables différentes :

La hernie par glissement : dans 90 % des cas, il y a glissement vers le haut du cardia (partie haute de l'estomac) qui quitte ainsi l'abdomen et se trouve dans le thorax. Les causes en sont par exemple une augmentation de la pression dans l'abdomen, une obésité, un raccourcissement congénital de l'œsophage.

La hernie par roulement (paraœsophagienne) : dans 10 % des cas, il y a roulement de la grosse tubérosité de l'estomac qui passe à travers l'orifice diaphragmatique alors que le bas œsophage et le cardia restent en place. La cause en est le plus souvent un défaut congénital du hiatus du diaphragme, il peut s'agir aussi d'un traumatisme.

Les facteurs de risque de la hernie hiatale sont ceux du reflux gastro-œsophagien.

CHAPITRE II

INTELLIGENCE ARTIFICIELLE

EN SANTE

1. INTRODUCTION

L'intelligence artificielle (IA) est un domaine de recherche en pleine expansion et promis à un grand avenir. Ses applications, qui concernent toutes les activités humaines, permettent notamment d'améliorer la qualité des soins. L'IA est en effet au cœur de la médecine du futur, avec les opérations assistées, le suivi des patients à distance, les prothèses intelligentes, les traitements personnalisés grâce au recoupement d'un nombre croissant de données (big data), etc.

L'Intelligence artificielle (IA) change notre façon de travailler et de vivre. L'expansion révolutionnaire dans l'IA, au cours des cinq dernières années, est directement liée aux progrès majeurs de la recherche en apprentissage automatique, qui permet aux ordinateurs d'apprendre sans être explicitement programmés à cet effet. Ces avancées représentent de grandes opportunités pour l'amélioration de notre système de santé, de la recherche fondamentale aux applications en cliniques.

Dans ce chapitre nous commençons par présenter un bref historique de l'IA. Ensuite nous présenterons les grandes approches en intelligence artificielle et les éléments requis pour les mettre en œuvre ainsi que les méthodes de validation appropriées pour évaluer la performance des outils générés. Nous présentons ensuite notre modèle basé sur le transfert du deep learning, la base de données utilisée en apprentissage et l'approche de localisation des lésions pulmonaires.

2. HISTORIQUE

Le terme d'intelligence artificielle a été évoqué pour la première fois en 1950 par Alan Turing. Ce mathématicien a élaboré un test permettant de vérifier la capacité d'un système à faire preuve d'intelligence humaine. De nos jours, le test de Turing est toujours utilisé pour mesurer l'intelligence de certaines machines, mais son efficacité est souvent remise en cause. Il faudra attendre 1956 pour que l'un de ses créateurs, Marvin Lee Minsky, définisse pour la première fois le terme d'intelligence artificielle comme étant « la construction de programmes informatiques qui s'adonnent à des tâches qui sont, pour l'instant, accomplies de façon plus satisfaisante par des êtres humains car elles demandent des processus mentaux de haut niveau tels que : l'apprentissage perceptuel, l'organisation de la mémoire et le raisonnement critique ».

Au fil des années, les chercheurs ont dissocié le concept d'intelligence artificielle en deux grandes approches : une intelligence artificielle symbolique et une autre connexionniste. L'approche qualifiée d'intelligence symbolique représente tous systèmes dotés de mécanismes de raisonnement capables d'interpréter des données symboliques qui constituent une base de connaissance. Cette approche permet de traiter des questions de logique formelle pour prendre des décisions intelligentes en fonction de règles, faits et raisonnements établis par le savoir humain.

La seconde approche, qualifiée d'intelligence connexionniste, s'inspire du fonctionnement des neurones biologiques. En effet, cette méthode regroupe tous les systèmes composés de sous-élément interconnectés capables de traiter et de communiquer des informations entre eux. En outre, cette intelligence artificielle connexionniste permet d'extraire des règles implicites contenues dans des bases de données volumineuses. Ces réseaux neuronaux artificiels sont généralement optimisés par des méthodes **d'apprentissage afin d'apprendre de leurs erreurs par comparaison**

statistiques avec les résultats attendus.

Depuis les années 1990, les chercheurs développent des modèles hybrides combinant ces deux approches d'intelligence artificielle qui comportent des caractères complémentaires. Ces systèmes sont constamment adaptés aux dernières avancées technologiques afin qu'ils soient toujours plus performants et innovants.

Parmi les domaines de l'intelligence artificielle en plein essor, on retrouve :

La reconnaissance et interprétation d'informations. Les données peuvent être de diverses natures nécessitant des traitements spécifiques. Ces systèmes sont exploités dans la reconnaissance de parole (logiciel de dictée vocale), la reconnaissance d'écriture manuscrite, la reconnaissance de visage, le traitement d'images, le diagnostic ou la compréhension de signaux.

L'aide à la décision permettant d'assister une personne dans ses choix lorsque les informations sont diverses et incertaines. Ces outils sont utilisés dans de nombreux secteurs d'activité tels que le médical, la finance, l'industrie, l'armée etc.

La planification d'actions est utilisée afin d'exécuter précisément une série d'actions préenregistrées dans le but de réaliser des tâches complexes et fastidieuses dans un environnement connu.

3. L'INTELLIGENCE ARTIFICIELLE DANS LE SECTEUR DE LA SANTÉ

Le développement général des volumes de données disponibles et le perfectionnement des nouvelles technologies, plus particulièrement des algorithmes et du matériel informatique, font de la santé un secteur de développement immense et extrêmement diversifié. Le croisement entre l'intelligence artificielle et la médecine permet de concevoir des systèmes plus préventifs et personnalisés, apportant une amélioration du suivi médical des patients et une assistance considérable aux professionnels de santé.

Les principaux domaines d'application de l'intelligence artificielle sont :

- La médecine prédictive de maladie et de leurs évolutions
- La médecine de précision permettant d'établir des recommandations de traitements personnalisés
- L'aide au diagnostic accompagnant les médecins dans leurs prises de décisions thérapeutiques les robots compagnons pour les personnes âgées ou fragiles
- La chirurgie assistée par ordinateur ; la prévention et l'anticipation d'épidémies
- L'aide au diagnostic prend une place considérable dans le développement des solutions d'intelligence artificielle dans le secteur de la santé.

L'imagerie, et plus généralement l'analyse de données médicales, et sans doute le domaine de la santé où l'apport de l'intelligence artificielle est le plus apparent. La santé est un domaine de choix pour la pleine exploitation des images, soit que leurs détails soient trop fins ou subtils pour être interprétés, soit que leur résolution médiocre demande un travail d'extrapolation, soit que la bonne interprétation demande une expérience et une connaissance d'une montagne de cas à des degrés inhumains.

4. LES FONCTIONNEMENTS DE L'INTELLIGENCE ARTIFICIELLE **(MACHINE LEARNING, DEEP LEARNING)**

Aujourd'hui, l'intelligence artificielle est dissociée en deux catégories : l'IA faible ou descendante, et l'IA forte ou ascendante. Chacune de ces catégories possède des méthodes de fonctionnement bien distinctes.

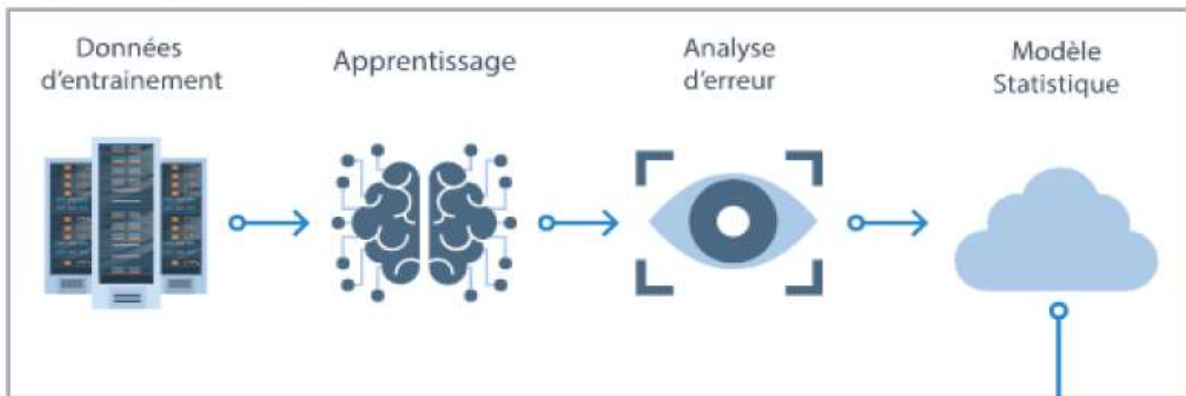
L'intelligence artificielle faible consiste à reproduire à l'identique une tâche spécifique à partir d'un programme informatique. Cette intelligence artificielle se focalise uniquement sur le résultat de l'action à réaliser, en délaissant son fonctionnement. De ce fait, ces systèmes effectueront uniquement ce qui leur a été demandé, sans possibilité d'évolution.

L'intelligence artificielle forte, quant à elle, a la capacité de produire un comportement intelligent et de concevoir une réelle conscience de soi. Contrairement à l'intelligence artificielle faible, ces programmes sont en constante évolution et peuvent acquérir et développer des comportements face à des situations non prévues lors de leur conception. Parmi cette intelligence forte se distinguent deux systèmes d'apprentissage qui sont : le Machine Learning et le Deep Learning.

4.1. Machine Learning

Le Machine Learning ou « apprentissage automatique » en français, représente l'ensemble des systèmes informatiques (Figure 32) qui reproduisent des comportements, non pas en les programmant, mais en leur fournissant des techniques statistiques capables d'apprendre de manière autonome. Lors de la phase d'apprentissage, les ingénieurs analysent des données de référence et créent des modèles statistiques pour pérenniser les résultats. De ce fait, lorsque de nouvelles données sont intégrées, les systèmes sont capables de prédire automatiquement des résultats en se basant sur les précédents modèles statistiques. Plus la base de données initiale importée est volumineuse, plus sa capacité de différenciation sera supérieure.

Étape 1 : Apprentissage



Étape 2 : Réalisation de la mission



Figure 32 : Schéma récapitulatif du fonctionnement du Machine Learning

Actuellement, de nombreuses entreprises de la santé développent et commercialisent des logiciels équipés de Machine Learning permettant d'identifier une pathologie, de la caractériser et d'établir un diagnostic. En se basant sur de grandes bases de données de diagnostic, les algorithmes peuvent détecter des anomalies sur une image. Le fonctionnement de ces algorithmes sera détaillé dans la seconde sous-partie.

4.2. Deep Learning

Le Deep Learning ou « apprentissage profond » en français, est un sous ensemble du Machine Learning. Les systèmes d'apprentissage profond utilisent de nombreuses couches de neurones (figure 33), capables d'interagir entre elles pour analyser des données non structurées et prédire des résultats. Ces systèmes sont capables d'apprendre et de s'améliorer de manière autonome. En effet, dès lors qu'une couche constate une erreur, les données sont alors systématiquement éliminées et renvoyées vers les précédents niveaux de couches afin de rectifier le modèle mathématique. Par la suite, lorsque ce modèle sera confronté à ces nouvelles données, il sera capable de les assimiler et de les dissocier sans que personne ne lui indique.

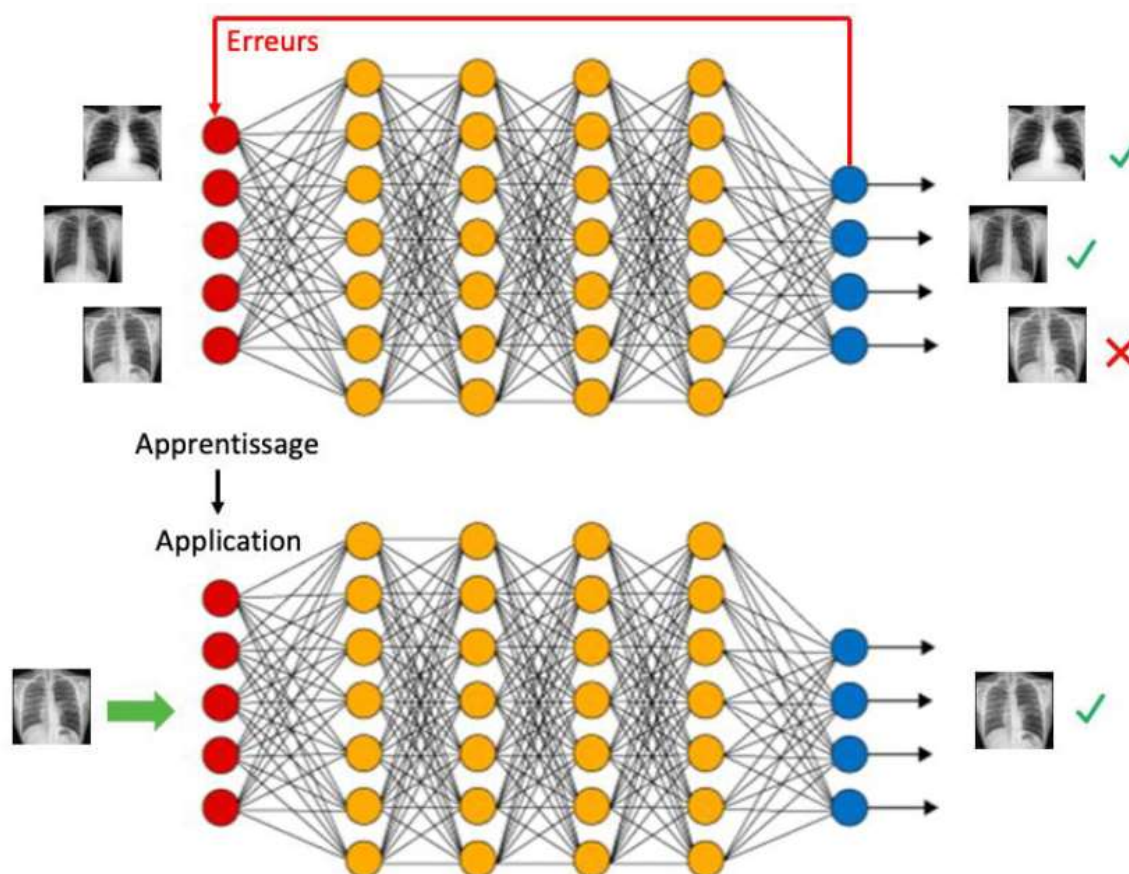


Figure 33: Schéma récapitulatif du fonctionnement du Deep Learning

L'intérêt des réseaux de neurones artificiels est d'imiter la logique de décision humaine. Par conséquent, les performances de ces systèmes vont dépendre considérablement de la quantité des échantillons (figure 34). Plus la base de données d'un système sera conséquente, plus la performance du système sera acceptable.

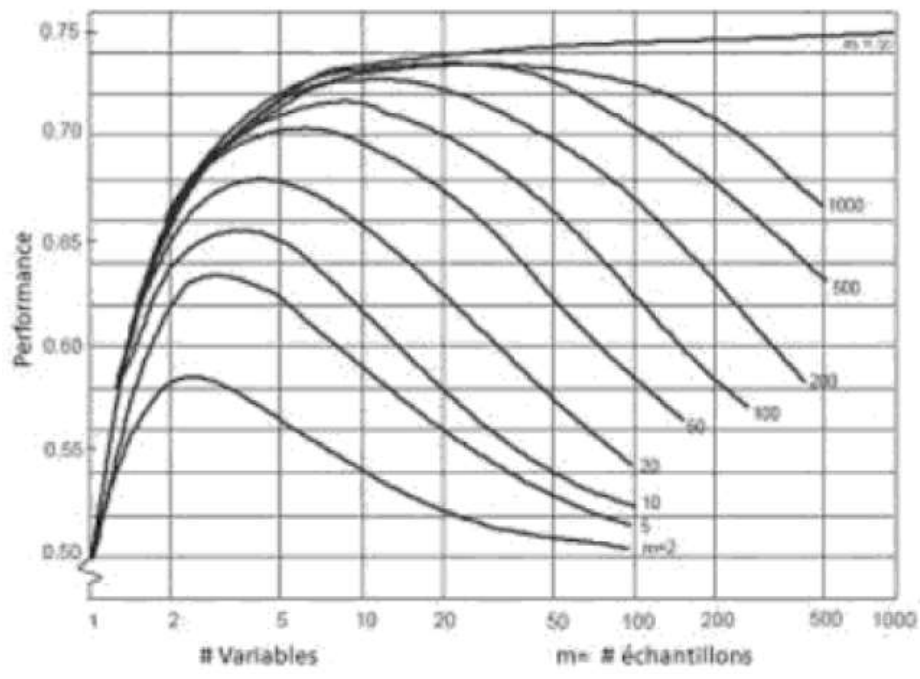


Figure 34 : Graphique représentant l'augmentation de la performance d'un système d'apprentissage avec le nombre d'échantillons

5. L'AIDE AU DIAGNOSTIC

Les systèmes d'aide au diagnostic, aussi nommés SADM (Systèmes d'Aide à la Décision Médicale), ont été définis comme étant « des applications informatiques dont le but est de fournir aux cliniciens en temps et lieux utiles les informations décrivant la situation clinique d'un patient ainsi que les connaissances appropriées à cette situation, correctement filtrées et présentées ».

Ces outils informatiques accompagnent et assistent de façon immédiate les praticiens dans leurs démarches diagnostiques lorsqu'ils sont confrontés à des situations d'urgence, à des pathologies rares ou des pathologies fréquentes de formes atypiques. De plus, certaines tâches s'avèrent fastidieuses et ennuyeuses pour les radiologues, par conséquent ces systèmes d'aide à la décision peuvent être un gain de temps considérable et d'une reproductibilité accrue. Ces algorithmes représentent ainsi un outil précieux permettant de prendre des décisions plus rapides et objectives afin d'accroître la qualité des soins et l'état de santé des patients.

Plus concrètement, ce sont les CAD (Computed Aided Diagnosis) qui analysent automatiquement les données recueillies par les praticiens, dans le but d'observer des irrégularités et d'établir des recommandations cliniques. Pour cela, ces CAD sont composés de base de données alimentée par un grand nombre d'images (correspondant à la structure anatomique étudiée) associées, ou non, à un état pathologique. Ainsi, à partir de ces bases de données, les algorithmes sont capables de répertorier automatiquement des images dans des catégories diagnostiques et de proposer des recommandations qui éclaireront les praticiens dans leurs démarches (Deep Learning).

Ces systèmes ont d'ores et déjà connu un relatif succès dans les hôpitaux américains avec des CAD capables de détecter des nodules pulmonaires, quantifier des lésions actives de sclérose en plaques, détecter des foyers de microcalcifications en mammographie etc...

6. NEURONE HUMAIN ET ARTIFICIEL

Au sein du cerveau humain, chaque neurone reçoit environ 100 000 signaux électriques des autres neurones. Chaque neurone en activité peut produire un effet excitant ou inhibiteur sur ceux auxquels il est connecté. Au sein d'un réseau artificiel, le principe est similaire. Les signaux voyagent entre les neurones. Toutefois, au lieu d'un signal électrique, le réseau de neurones assigne un certain poids à différents neurones. Un neurone qui reçoit plus de charge exercera plus d'effet sur les neurones adjacents. La couche finale de neurones émet une réponse à ces signaux.

Pour comprendre comment fonctionne le Deep Learning, prenons un exemple concret de reconnaissance d'images. Imaginons que le réseau de neurones soit utilisé pour reconnaître les photos qui comportent au moins un chat. Pour pouvoir identifier les chats sur les photos, l'algorithme doit être en mesure de distinguer les différents types de chats, et de reconnaître un chat de manière précise quel que soit l'angle sous lequel il est photographié.

Afin d'y parvenir, le réseau de neurones doit être entraîné. Pour ce faire, il est nécessaire de compiler un ensemble d'images d'entraînement pour pratiquer le Deep Learning. Cet ensemble va regrouper des milliers de photos de chats différents, mélangés avec des images d'objets qui ne sont pas des chats. Ces images sont ensuite converties en données et transférées sur le réseau. Les neurones artificiels assignent ensuite un poids aux différents éléments. La couche finale de neurones va alors rassembler les différentes informations pour déduire s'il s'agit ou non d'un chat.

Le réseau de neurones va ensuite comparer cette réponse aux bonnes réponses indiquées par les humains. Si les réponses correspondent, le réseau garde cette réussite en mémoire et s'en servira plus tard pour reconnaître les chats. Dans le cas contraire, le réseau prend note de son erreur et ajuste le poids placé sur les différents neurones pour corriger son erreur. Le processus est répété des milliers de fois jusqu'à ce que le réseau soit capable de reconnaître un chat sur une photo dans toutes les circonstances. Cette technique d'apprentissage est appelée « supervised learning » ou apprentissage supervisé.

Une autre technique d'apprentissage est celle de l'« unsupervised learning », ou apprentissage non supervisé. Cette technique repose sur des données qui ne sont pas étiquetées. Les réseaux de neurones doivent reconnaître des patterns au sein des ensembles de données pour apprendre par eux-mêmes quels éléments d'une photo peuvent être pertinents.

7. QU'EST CE QUE LE MACHINE LEARNING (ML)?

Le ML dont le processus est décrit par la figure 35 est une discipline de l'IA qui offre aux ordinateurs la possibilité d'apprendre à partir d'un ensemble d'observations que l'on appelle ensemble d'apprentissage.

Chaque observation, comme par exemple « j'ai mangé tels et tels aliments à tel moment de la journée pendant telle période ce qui a causé telle maladie » est décrite au moyen de deux types de variables :

Les premières sont appelées les variables prédictives (ou attributs ou caractéristiques), dans notre exemple mon âge, mon dossier médical, mes antécédents médicaux. Ce sont les variables à partir desquelles on espère pouvoir faire des prédictions. Les n variables prédictives associées à une observation seront notées comme un vecteur $x=(x_1, \dots, x_n)$ à n composantes. Un ensemble de M observations sera

constitué de M tels vecteurs $x(1), \dots, x(M)$.

Une variable cible dont on souhaite prédire la valeur pour des événements non encore observés. Dans notre exemple, il s'agirait de la maladie contractée. On notera y cette variable cible.

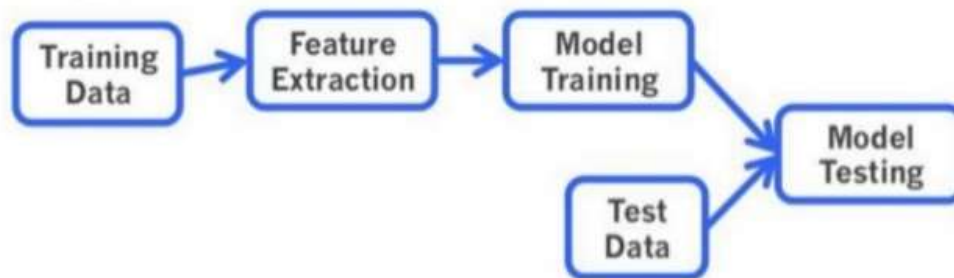


Figure 35 : Le processus typique du ML

En résumé, la valeur de la variable y dépend de :

Une fonction $F(x)$ déterminée par les variables prédictives.

Un bruit $\epsilon(x)$ qui est le résultat d'un nombre de paramètres dont on ne peut pas tenir compte.

Aussi bien F que ϵ ne seront jamais connues mais l'objectif d'un modèle de ML est d'obtenir la meilleure approximation possible de F à partir des observations disponibles (figure 36). Cette approximation sera notée f , on l'appelle la fonction de prédiction.

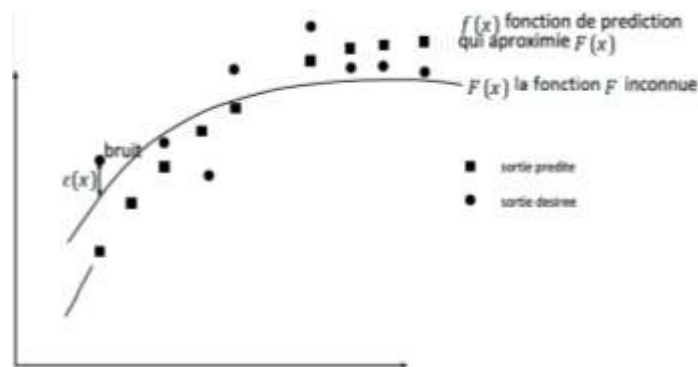


Figure 36 : un modèle de ML qui essaye d'obtenir la meilleure approximation possible de F

7.1. Performance et surapprentissage

On peut penser que la performance d'un modèle sera en fonction des prédictions correctes faites sur l'ensemble d'observation utilisées pour l'apprentissage, plus elle est élevée, mieux c'est. Pourtant c'est complètement faux, ce qu'on cherche à obtenir du ML n'est pas de prédire avec exactitude les valeurs des variables cibles connues puisqu'elles ont été utilisées pour l'apprentissage mais bien de prédire celles qui n'ont pas encore été observées. Par conséquent, la qualité d'un algorithme de ML se juge sur sa capacité à faire les bonnes prédictions sur les nouvelles observations grâce aux caractéristiques apprises lors de la phase d'entraînement. Il faut donc éviter le cas où on a un modèle de ML tellement trop entraîné qu'il arrive à prédire à la perfection les données d'apprentissage mais qui n'arrive pas à généraliser sur les données de test. On l'appelle le sur-apprentissage (figure 37).

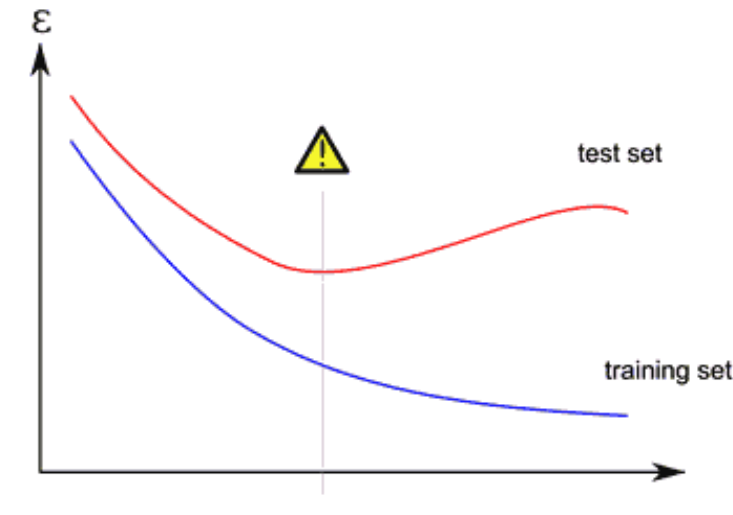


Figure 37 : Le sur-apprentissage

Le graphe montre l'évolution de l'erreur commise sur l'ensemble de test par rapport à celle commise sur l'ensemble d'apprentissage, les deux erreurs diminuent mais dès que l'on rentre dans une phase de sur-apprentissage, l'erreur d'apprentissage continue de diminuer alors que celle du test augmente.

Pour résoudre ce problème, on divise les données disponibles en deux groupes distincts. Le premier sera l'ensemble d'apprentissage, et le deuxième sera l'ensemble de test.

Pour avoir une bonne séparation des données en données d'apprentissage et données de test, on utilise la validation croisée (figure 38). L'idée c'est de séparer aléatoirement les données dont on dispose en k parties séparées de même taille. Parmi ces k parties, une fera office d'ensemble de test et les autres constitueront l'ensemble d'apprentissage. Après que chaque échantillon ait été utilisé une fois comme ensemble de test. On calcule la moyenne des k erreurs moyennes pour estimer l'erreur de prédiction.

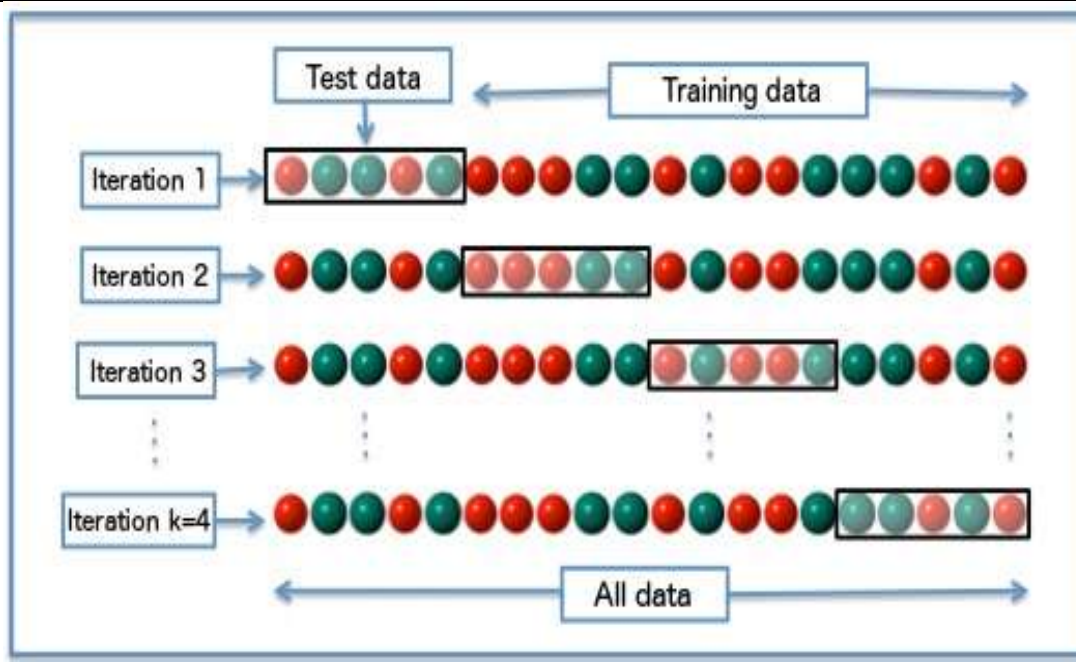


Figure 38 : La validation croisée

7.2. Les différents types de machine learning

L'apprentissage supervisé nécessite un étiquetage préalable des données. En effet, Soit l'ensemble d'apprentissage composé de N exemples de pair entrée-sortie : $(x(1),y(1)),(x(2),y(2)),\dots,(x(M),y(M))$ chaque $y(j)$ a été généré par une fonction $F(x) = y$ inconnue, découvrir la fonction f qui se rapproche de F

L'apprentissage non supervisé ou clustering ne demande aucun étiquetage préalable des données. Le but est que le modèle réussisse à regrouper les observations disponibles en catégories par lui même

7.3. Régression et classification

Un modèle de classification est un modèle de ML dont les sorties y appartiennent à un ensemble fini de valeurs (exemple : bon, moyen, mauvais)

Un de modèle de régression est un modèle de ML dont les sorties y sont des nombres (exemple : la température de demain)

7.3.1. La régression linéaire

Une régression linéaire est un modèle de ML supervisé, avec x en entrée et y en

sortie elle est de la forme $y = w_1x + w_0$ ou w_0 et w_1 sont des valeurs réelles à apprendre. On définit w comme le vecteur $[w_0, w_1]$, et on définit :

$$f(x) = w_1x + w_0 \quad (1)$$

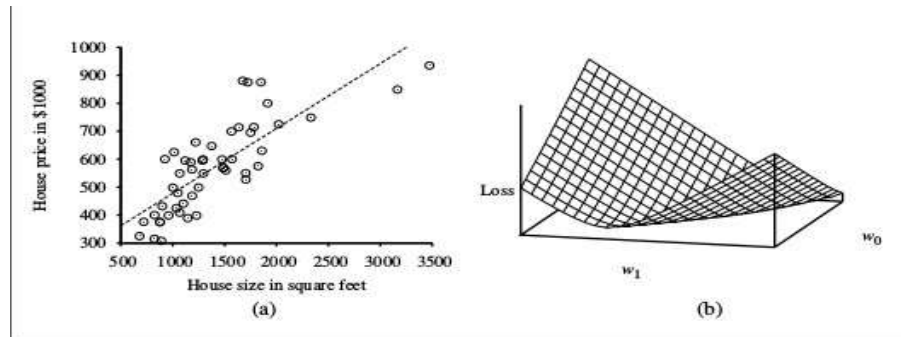


Figure 39 : Régression linéaire représentant prix des maisons en fonction de leur surface

La figure 28 - (a) représente le prix des maisons en fonction de leur surface qui sont en vente à Berkeley. Le prix est représenté par la fonction linéaire qui minimise l'erreur :

$y = 0.232x + 246$. (b) représente la fonction coût $\sum (w_1x_i + w_0 - y_i)^2$ pour différentes valeurs de w_1 et w_0 .

La figure 28(b) montre un exemple d'ensemble d'apprentissage de n points sur le plan x, y , chaque point représente la surface et le prix d'une maison en vente. La tâche de trouver f qui convient le mieux à ces données s'appelle la régression linéaire. Pour trouver f il suffit de trouver les valeurs $[w_1, w_0]$ qui minimise l'erreur empirique.

$$\text{Loss}(f) = \sum (y_i - (w_1x_i + w_0))^2 \quad (2)$$

Avantages :

L'avantage avec ce modèle est que l'apprentissage se résume à trouver la solution de l'équation $w^* = \text{argmin}_w \text{Loss}(f)$ qui est exacte :

$$w^* = (X^T X)^{-1} X^T y \quad (3)$$

Le modèle est facile à interpréter

Inconvénients

Sensible aux bruits

Négligence des interactions entre les variables prédictives

7.4. Les k plus proches voisins

L'algorithme des K-Nearest Neighbors (KNN) (K plus proches voisins) est un algorithme de classification supervisé. Chaque observation de l'ensemble d'apprentissage est représentée par un point dans un espace à n dimensions ou n est le nombre de variables prédictives. Pour prédire la classe d'une observation, on cherche les k points les plus proches de cet exemple. La classe de la variable cible, est celle qui est la plus représentée parmi les k plus proches voisins. Il existe des variantes de l'algorithme ou on pondère les k observations en fonction de leur distance à l'exemple dont on veut classer [12], les observations les plus éloignées de notre exemple seront considérées comme moins importantes.

Avantages :

- Simple à concevoir

Inconvénients

- Sensible aux bruits
- Pour un nombre de variable prédictives très grands, le calcul de la distance devient très coûteux

7.5. Le classifieur naïf de Bayes

Le classifieur naïf de Bayes est un algorithme supervisé probabiliste qui suppose que l'existence d'une caractéristique pour une classe, est indépendante de l'existence d'autres caractéristiques, raison pour laquelle on utilise l'adjectif «naïf» . Une personne peut être considérée comme un homme s'il pèse un certain poids et mesure une certaine taille. Même si ces caractéristiques sont liées dans la réalité, un classifieur bayésien naïf déterminera que la personne est un homme en considérant

indépendamment ces caractéristiques de taille et de poids.

Malgré des hypothèses de base extrêmement simplistes, ce classifieur conduit à de très bons résultats dans beaucoup de situations réelles complexes.

Avantages :

- L'algorithme offre de bonne performance

Inconvénients

La prédiction devient erronée si L'Hypothèse indépendance conditionnelle est invalide

7.6. k-means

L'algorithme des k-moyennes (k-means) est un algorithme non supervisé. Chaque observation est représentée par un point dans un espace à n dimensions ou n est le nombre de variables descriptives. À partir d'un ensemble d'apprentissage de M observations $x(1), \dots, x(M)$, cet algorithme va répartir ces observations en k clusters de manière à ce que la distance euclidienne qui sépare les points au centre de gravité du groupe auquel ils sont affectés soit minimale. Les étapes de l'algorithme sont :

- Choisir k points qui représentent la position moyenne des clusters
- Répéter jusqu'à stabilisation des points centraux :
- Affecter chacun des M points au plus proche des k points centraux
- Mettre à jour les points centraux en calculant les centres de gravité des k cluster

Avantages :

- Implémentable pour des grands volumes de données

Inconvénients

- Le choix du paramètre k n'est pas découvert mais choisi par l'utilisateur

La solution dépend des k centre de gravité choisi lors de l'initialisation

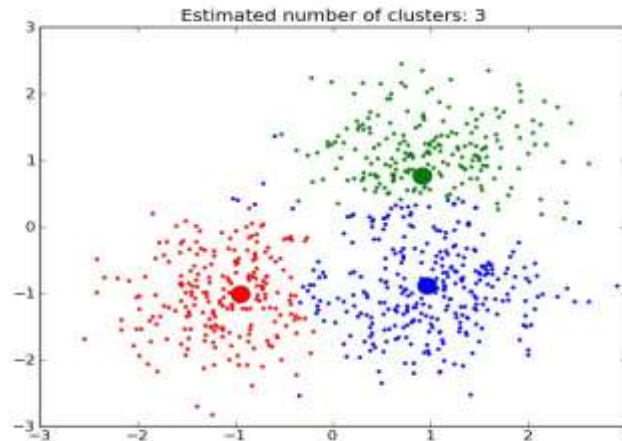


Figure 40 : L'algorithme k -means regroupe les données en k cluster, ici $k = 3$. Les centres de gravité sont représentés par de petit cercle

7.7. Les arbres de décision

Les arbres de décision sont des modèles de ML supervisés, pouvant être utilisés aussi bien pour la classification que pour la régression.

Un arbre de décision représente une fonction qui prend comme entrée un vecteur d'attributs et retourne une décision qui est une valeur unique. Les entrées et les sorties peuvent être discrètes ou continues.

Un arbre de décision prend ses décisions en exécutant une séquence de test, chaque nœud interne de l'arbre correspond à un test de la valeur d'un attribut et les branches qui sortent du nœud sont les valeurs possibles de l'attribut. La classe de la variable cible est alors déterminée par la feuille dans laquelle parvient l'observation à l'issue de la séquence de test.

La phase d'apprentissage consiste à trouver la bonne séquence de test. Pour cela, on doit décider des bons attributs à garder. Un bon attribut divise les exemples en ensembles homogènes c'est à dire qu'ils ne contiennent que des observations appartenant à la même classe, alors qu'un attribut inutile laissera les exemples avec presque la même proportion de valeur pour la variable cible.

Ce dont on a besoin c'est d'une mesure formelle de "bon" et "inutile". Pour cela, il existe des métriques standards homogénéisées avec lesquels on peut mesurer l'homogénéité d'un ensemble. Les plus connus sont l'indice de diversité de Gini et l'entropie.

En général l'entropie d'une variable aléatoire V avec des valeurs v_k chacune avec une probabilité $P(v_k)$ est définie comme :

$$\text{Entropie : } H(V) = -\sum P(v_k) \log_2 P(v_k) \quad (4)$$

Avantages :

- C'est un modèle boîte blanche, simple à comprendre et à interpréter
- Peu de préparation des données
- Les variables prédictives en entrée peuvent être aussi bien qualitatives que quantitatives
- Performant sur de grands jeux de données

7.8. Les forêts aléatoires

Combiner un ensemble de classifieurs individuels faibles pour former un unique système de classification appelé Ensemble de Classifieurs a suscité un intérêt grandissant de la communauté scientifique. L'efficacité des combinaisons de classifieurs repose principalement sur leur capacité à tirer parti des complémentarités des classifieurs individuels, dans le but d'améliorer autant que possible les performances en généralisation de l'ensemble.

Parmi les différentes approches de construction d'ensembles de classifieurs, l'algorithme des forêts aléatoires, composé d'un ensemble de classifieurs élémentaires de type arbres de décision, le but de l'algorithme est de conserver les avantages des arbres de décision tout en éliminant leurs inconvénients et particulièrement leur vulnérabilité au sur-apprentissage. C'est un algorithme qui peut être utilisé aussi bien pour la classification que pour la régression.

L'algorithme repose sur trois idées principales :

Pour M observations de l'ensemble d'apprentissage, chacune décrite par n variables prédictives, on crée B nouveau échantillon de même taille M par tirage avec remise. Cette technique s'appelle le bootstrap. Chacun des B échantillon servira à l'apprentissage d'un arbre de décision.

Pour n caractéristiques, un nombre $k < n$ (généralement n) est tiré aléatoirement de sorte qu'à chaque nœud de l'arbre, un sous-ensemble de k caractéristiques soit tiré aléatoirement, parmi lesquelles la meilleure est ensuite sélectionnée pour le partitionnement.

Pour classer une nouvelle observation, on procède par vote majoritaire. On fait passer cette observation par les B arbres et sa classe c'est la classe majoritaire parmi les B prédictions

Quelques travaux de recherche se sont intéressés au nombre d'arbres de décision à construire au sein d'une forêt. Quand Breiman [13] a introduit les forêts aléatoires, il démontra également qu'au-delà d'un certain nombre d'arbres, en ajouter d'autres ne permettait pas forcément d'améliorer les performances de l'ensemble. Ce résultat indique que le nombre d'arbres d'un forêt aléatoire ne doit pas nécessairement être le plus grand possible pour produire un classifieur performant (figure 41) où la création de B bootstrap à partir des exemples d'apprentissage servira à entraîner un arbre de décision. Si c'est un cas de classification, la classe finale est attribuée par vote majoritaire. Si c'est un cas de régression, on fait la moyenne des prédictions.

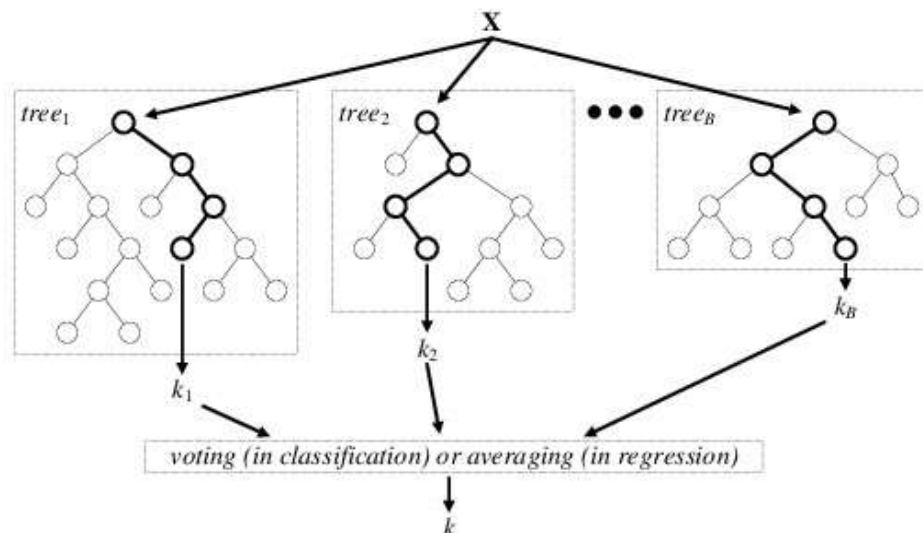


Figure 41 : Création de B bootstrap à partir des exemples d'apprentissage.

Avantages :

- C'est un des meilleurs algorithmes pour ce qui est de la précision.
- Incorporation de la validation croisée.
- Il conserve une bonne puissance de prédiction même s'il y a des données manquantes
- Il ne souffre pas du sur-apprentissage

Inconvénients

- Une implémentation difficile

La solution dépend des k centre de gravité choisi lors de l'initialisation

7.9. Les machines à vecteur de support

Les Support Vector Machine (SVM) (machines à vecteur de support) sont des algorithmes de classification binaire non linéaire très puissante.

Le principe des SVM consiste à construire une bande séparatrice non linéaire de largeur maximale qui sépare deux ensembles d'observations et à l'utiliser pour faire des prédictions. L'astuce des SVM pour y parvenir consiste à utiliser une transformation ϕ non linéaire qui envoie les points $x(1), \dots, x(M)$ de l'espace original à n dimensions (n

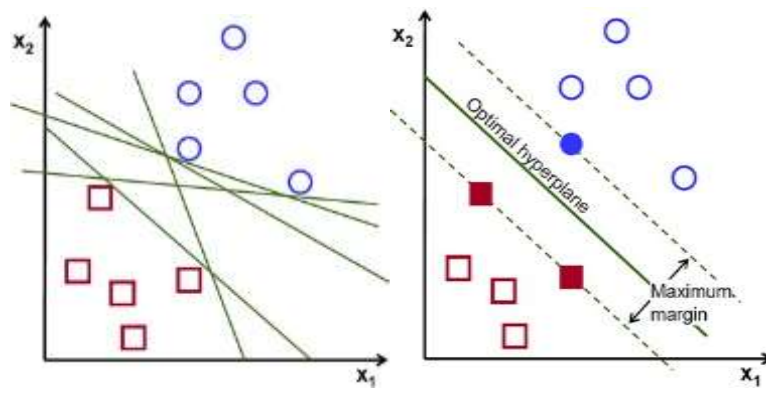
est le nombre de variable prédictives) vers des nouveaux points $\varphi(x(1)), \dots, \varphi(x(M))$ dans un espace de dimension plus grand que n ou ils seront plus facile à séparer.

Les SVM sont des classificateurs qui reposent sur deux idées clés :

La première idée consiste à trouver un séparateur linéaire de largeur maximale, c'est la notion de marge maximale. La marge est la distance entre la frontière de séparation et les échantillons les plus proches. Ces derniers sont appelés vecteurs supports. Le problème est de trouver cette frontière séparatrice optimale.

Dans le cas où le problème est linéairement séparable, le choix de l'hyperplan séparateur n'est pas évident. Il existe en effet une infinité d'hyperplans séparateurs, dont les performances en phase d'apprentissage sont identiques, mais dont les performances en phase de test peuvent être très différentes. Pour résoudre ce problème, Vapnik a été montré [14], qu'il existe un unique hyperplan optimal, défini comme l'hyperplan qui maximise la marge entre les échantillons et l'hyperplan séparateur.

Il existe des raisons théoriques à ce choix. Vapnik a montré [14] que la capacité des classes d'hyperplans séparateurs diminue lorsque leur marge augmente.



(a)

(b)

Figure 42 : Recherche d'un hyperplan qui divise les observations en deux catégories.(a) les hyperplans possibles avant optimisation, (b) l'hyperplan optimal

Afin de pouvoir traiter des cas où les données ne sont pas linéairement séparables, la deuxième idée clé des SVM est de transformer l'espace de représentation des données d'entrées en un espace de plus grande dimension, dans lequel il est probable qu'il existe une séparation linéaire. Ceci est réalisé grâce à une fonction noyau, qui doit respecter les conditions du théorème de Mercer et qui a l'avantage de ne pas nécessiter la connaissance explicite de la transformation à appliquer pour le changement d'espace (figure 43). Les fonctions noyaux permettent de transformer un produit scalaire dans un espace de grande dimension, ce qui est coûteux, en une simple

évaluation ponctuelle d'une fonction. Cette technique est connue sous le nom de kernel trick. Les deux fonctions noyau les plus utilisées sont le noyau polynomial et le noyau gaussien.

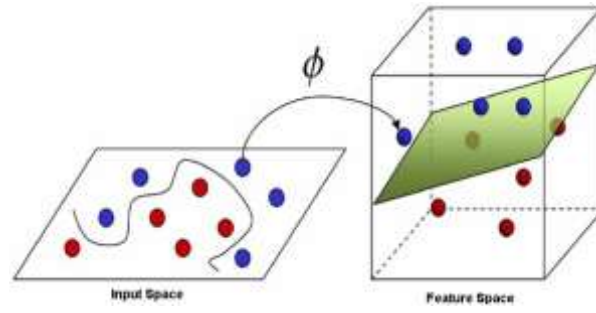


Figure 44 : Exemple d'un problème non linéairement séparable. La courbe devient une bande linéaire après avoir appliqué la transformation ϕ non linéaire

Avantages :

- Il permet de traiter des problèmes de classification non linéaire complexe.
- Les SVM constituent une alternative aux réseaux de neurones car plus faciles à entraîner.

Inconvénients

- Les SVM sont souvent moins performants que les forêts aléatoires.

7.10. Le perceptron multicouches

Le perceptron multicouches (Multi Layer Perceptron (MLP) [15]) est un classifieur de type réseaux de neurones qui est organisé en plusieurs couches, chaque couche étant formée d'un ou plusieurs neurones formels. Utilisé dans un cas d'apprentissage supervisé, il utilise l'algorithme de retro propagation de gradient. C'est un composant calculatoire qui fait la somme pondérée des signaux reçus en entrée, puis on leur applique une fonction d'activation afin d'obtenir Y (figure 44).

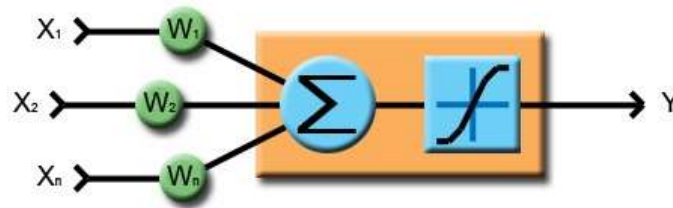


Figure 44 : Fonctionnement d'un neurone formel.

Les caractéristiques du modèle sont:

- Il comporte une seule couche d'entrée et une seule couche de sortie.
- Il peut comporter une ou plusieurs couches cachées.
- Chaque neurone est relié uniquement à tous les neurones de la couche suivante.
- Chaque lien de la couche i vers la couche suivante j sert à propager l'activation a_i de i jusqu'à j et qui possède un poids w_{ij} qui détermine l'intensité du signal de la connexion. Chaque unité de la couche j calcul la somme pondérée de ses entrées :

$$in_j = \sum w_{ij} a_i \quad (5)$$

Puis on applique une fonction d'activation :

$$a_j = g(in_j) = g(\sum w_{ij} a_i) \quad (6)$$

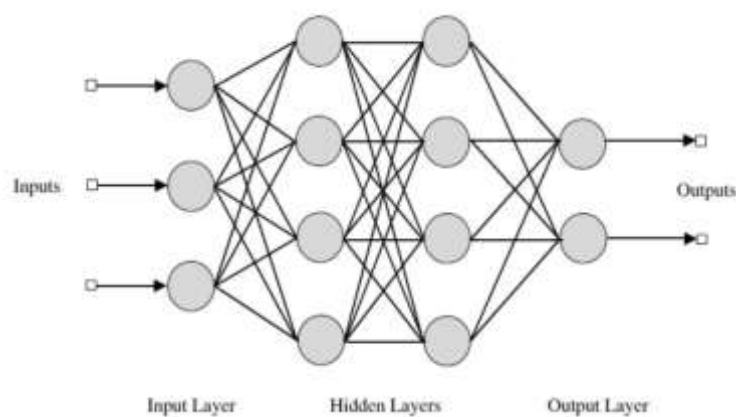


Figure 44 : Structure d'un PMC composé d'une couche d'entrée, deux couches cachées et une couche de sortie

L'algorithme d'apprentissage consiste à présenter au réseau des entrées et on lui demande de modifier sa pondération de façon à retrouver la sortie correspondante. D'abord on propage vers l'avant les entrées jusqu'à obtenir une sortie calculée par le réseau puis on compare cette sortie avec la sortie désirée, enfin on modifie les poids de telle sorte qu'à la prochaine itération l'erreur commise entre les sorties réelles et les sortie désirées soit minimisée. On répète ce processus jusqu'à ce qu'on obtienne une erreur de sortie négligeable.

Avantages :

- Capacité à découvrir les dépendances par lui-même.
- Résistance aux bruits

Inconvénients

- C'est un modèle boîte noire qui n'explique pas ses décisions.

8. DEEP LEARNING

8.1. Introduction

Le Deep Learning est un nouveau domaine de recherche du ML, qui a été introduit dans le but de rapprocher le ML de son objectif principal : l'intelligence artificielle (figure 45). Il concerne les algorithmes inspirés par la structure et le fonctionnement du cerveau. Ils peuvent apprendre plusieurs niveaux de représentation dans le but de modéliser des relations complexes entre les données. C'est aussi cette partie de l'IA qui va nous intéresser dans le cadre de notre travail.

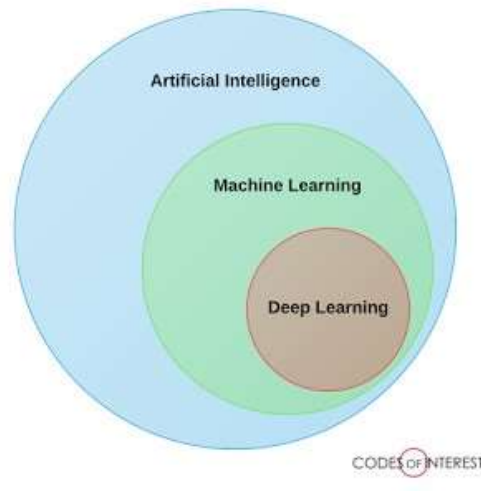


Figure 45 : La relation entre l'intelligence artificielle, le ML et le deep learning

Le Deep Learning est basé sur l'idée des réseaux de neurones artificiels et il est taillé pour gérer de larges quantités de données en ajoutant des couches au réseau. Un modèle de deep learning a la capacité d'extraire des caractéristiques à partir des données brutes grâce aux multiples couches de traitement composé de multiples transformations linéaires et non linéaires et apprendre sur ces caractéristiques petit à petit à travers chaque couche avec une intervention humaine minimale.

Sur les cinq dernières années, le deep learning est passé d'un marché de niche ou seulement une poignée de chercheurs s'y intéressait au domaine le plus prisé par les chercheurs. Les recherches en relation avec le deep learning apparaissent maintenant dans les meilleurs journaux. Le deep learning a aujourd'hui appris à conduire une voiture, diagnostiquer le cancer et l'autisme, et est même devenu un artiste.

Le terme "Deep Learning" a été introduit pour la première fois au ML par Dechter (1986) [16] et aux réseaux neuronaux artificiels par Aizenberg et al (2000) [17].

8.2. Historique

Tableau 1 : Les étapes majeurs du Deep Learning

| année | Contributeur | Contribution |
|-----------|----------------------|--|
| 300 AC | Aristote | introduction de l'associationnisme, début de l'histoire des humains qui essayent de comprendre le cerveau |
| 1873 | Alexander Bain | introduction du Neural Groupings comme les premiers modèles de réseaux de neurones |
| 1943 | McCulloch and Pitts | introduction du McCulloch–Pitts (MCP) modèle considéré comme L'ancêtre des réseaux de neurones artificielles |
| 1949 | Donald Hebb | considérer comme le père des réseaux de neurones, il introduit la règle d'apprentissage de Hebb qui servira de fondation pour les réseaux de neurones modernes |
| 1958 | Frank Rosenblatt | introduction du premier perceptron |
| 1974 | Paul Werbos | introduction de la retro propagation |
| 1980 | Teuvo Kohonen | introduction des cartes auto organisatrices |
| 1980 | Kunihiko Fukushima | introduction du Neocognitron, qui a inspiré les réseaux de neurones convolutif |
| 1982 | John Hopfield | introduction des réseaux de Hopfield |
| 1985 | Hilton and Sejnowski | introduction des machines de Boltzmann |
| 1986 | Paul | introduction de Harmonium, |

| | | |
|------|----------------------------------|--|
| | Smolensky | qui sera connu plus tard comme machines de Boltzmann restreintes |
| 1986 | Michael I. Jordan | définition et introduction des réseaux de neurones récurrent |
| 1990 | Yann LeCun | introduction de LeNet et montra la capacités des réseaux de neurones profond |
| 1997 | Schuster and Paliwal | introduction des réseaux de neurones récurrent bidirectionnelles |
| 1997 | Hochreiter and Schmidhuber | introduction de LSTM, qui ont résolu le problème du vanishing gradient dans les réseaux de neurones récurrent |
| 2006 | Geoffrey Hinton | introduction des Deep belief Network |
| 2009 | Salakhutdinov and Hinton | introduction des Deep Boltzmann Machines |
| 2012 | Alex Krizhevsky | introduction de AlexNet qui remporta le challenge ImageNet |

8.3. Pourquoi le Deep Learning?

Les algorithmes de ML décrits dans la première partie fonctionnent bien pour une grande variété de problèmes. Cependant ils ont échoué à résoudre quelques problèmes majeurs de l'IA telle que la reconnaissance vocale et la reconnaissance d'objets.

Le développement du Deep Learning fut motivé en partie par l'échec des algorithmes traditionnels dans de telles tâches de l'IA.

Mais ce n'est qu'après que de plus grandes quantités de données ne soit disponibles grâce notamment au Big Data et aux objets connectés et que les machines de calcul soient devenues plus puissantes qu'on a pu comprendre le potentiel réel du

Deep Learning.

Une des grandes différences entre le Deep Learning et les algorithmes de ML traditionnels c'est qu'il s'adapte bien, plus la quantité de données fournie est grande plus les performances d'un algorithme de Deep Learning sont meilleures. Contrairement à plusieurs algorithmes de ML classiques qui possèdent une borne supérieure à la quantité de données qu'ils peuvent recevoir des fois appelée "plateau de performance", les modèles de Deep Learning n'ont pas de telles limitations (théoriquement) et ils sont même allés jusqu'à dépasser la performance humaine dans des domaines comme l'image processing (figure 46).

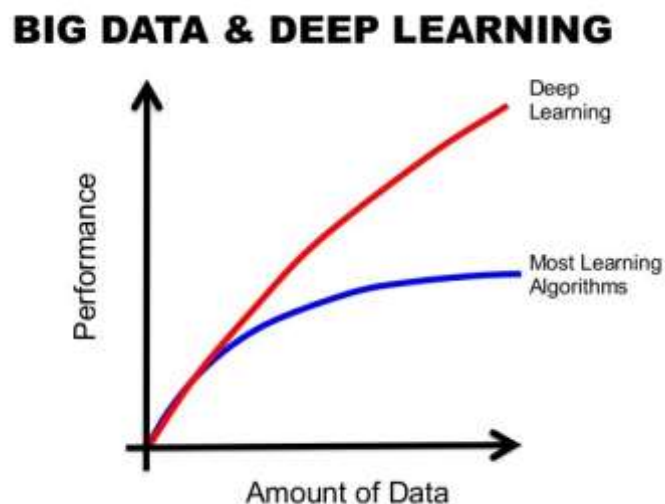


Figure 46 : La différence de performance entre le Deep Learning et la plupart des algorithmes de ML en fonction de la quantité de données

Autre différence entre les algorithmes de ML traditionnelles et les algorithmes de Deep Learning c'est l'étape de l'extraction de caractéristiques (figure 47). Dans les algorithmes de ML traditionnelles l'extraction de caractéristiques est faite manuellement, c'est une étape difficile et coûteuse en temps et requiert un spécialiste en la matière alors qu'en Deep Learning cette étape est exécutée automatiquement par l'algorithme. De plus le Deep Learning a démontré une robustesse vis-à-vis de l'erreur

de reconnaissance par rapport au ML lors de la compétition de reconnaissance de forme image de la base ImageNet (figure 48).

Les différents types de modèles

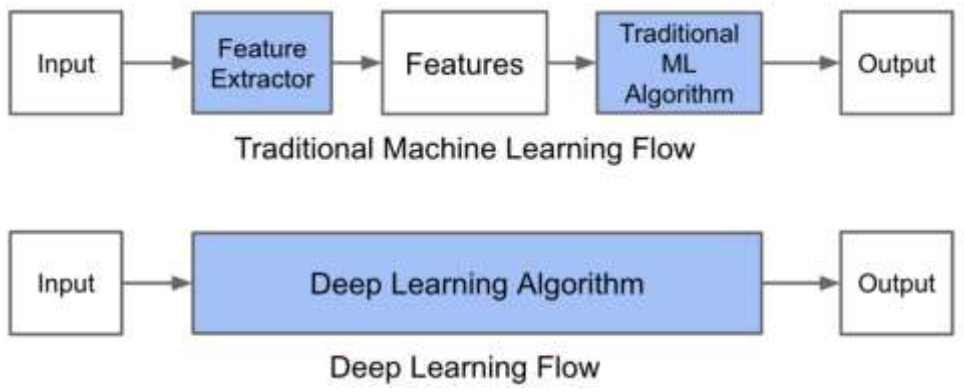


Figure 47 : Le processus du ML classique comparé à celui du Deep Learning



Figure 48: Deep Learning vs Machine Learning dans ImageNet challenge

8.4. Les différents types de modèles

Il existe un grand nombre de variantes d'architectures profondes. La plupart d'entre elles sont dérivées de certaines architectures parentales originales. Il n'est pas toujours possible de comparer les performances de toutes les architectures, car elles ne sont pas toutes évaluées sur les mêmes ensembles de données. Le Deep Learning est un domaine à croissance rapide, et de nouvelles architectures, variantes ou algorithmes apparaissent toutes les semaines.

8.5. Les réseaux de neurones convolutifs

Les réseaux de neurones convolutifs (Convolutional Neural Network (CNN)) sont un type de réseau de neurones spécialisés pour le traitement de données ayant une topologie semblable à une grille. Les exemples comprennent des données de type série temporelle, qui peuvent être considérées comme une grille 1D en prenant des échantillons à des intervalles de temps réguliers et des données de type image, qui peuvent être considérées comme une grille 2D de pixels. Les réseaux convolutifs ont connu un succès considérable dans les applications pratiques. Le nom « réseau de neurones convolutif » indique que le réseau emploie une opération mathématique appelée convolution. La convolution est une opération linéaire spéciale. Les réseaux convolutifs sont simplement des réseaux de neurones qui utilisent la convolution à la place de la multiplication matricielle dans au moins une de leurs couches.

Ils ont de larges applications dans la reconnaissance de l'image et de la vidéo, les systèmes de recommandations et le traitement du langage naturel.

Inspiration

Dans le ML, un réseau convolutif est un type de réseau de neurones feed-forward, il a été inspiré par des processus biologiques. Le modèle de connectivité entre les neurones d'un CNN s'inspire de l'organisation du cortex visuel animal.

Les neurones corticaux individuels répondent aux stimuli dans une région restreinte de l'espace connu sous le nom de champ réceptif. Les champs réceptifs des différents neurones se chevauchent partiellement de sorte qu'ils couvrent le champ visuel.

La réponse d'un neurone individuel aux stimuli dans son champ réceptif peut être approchée mathématiquement par une opération de convolution.

L'opération de convolution

Dans sa forme la plus générale, la convolution est une opération sur deux fonctions d'argument réel. Pour comprendre la motivation derrière la convolution, nous commençons par des exemples de deux fonctions qu'on pourrait utiliser.

Supposons que nous suivons l'emplacement d'un vaisseau spatial avec un capteur laser. Notre capteur laser fournit une sortie $x(t)$ qui est la position du vaisseau spatial au moment t . x et t sont des réelles, c'est-à-dire que nous pouvons obtenir une lecture différente du capteur laser à tout moment.

Supposons maintenant que notre capteur laser soit quelque peu bruité. Pour obtenir une estimation moins bruitée de la position du vaisseau spatial, nous aimerions combiner plusieurs mesures. Bien sûr, les mesures plus récentes sont plus pertinentes, donc nous voulons que ces mesures soient une moyenne pondérée et donner plus de poids aux mesures récentes. Nous pouvons le faire avec une fonction de pondération $w(a)$, où a est l'âge de la mesure.

Si nous appliquons une telle opération de moyenne pondérée à chaque instant, nous obtenons une nouvelle fonction qui fournit une estimation lisse de la position du vaisseau spatial :

$$s(t) = \int x(a)w(t - a)da \quad (7)$$

Cette opération s'appelle convolution. L'opération de convolution est généralement désignée par un astérisque :

$$s(t) = (x * w)(t) \quad (8)$$

Dans notre exemple, l'idée d'un capteur laser qui peut fournir des mesures à chaque instant dans le temps n'est pas réaliste. Habituellement, lorsque nous travaillons avec des données sur l'ordinateur, le temps sera discrétisé (numérisé), et notre capteur fournira des données à des intervalles réguliers. Dans notre exemple, il est plus réaliste de supposer que notre laser fournit une mesure une fois par seconde. L'index de temps t ne peut alors prendre que des valeurs entières. Si on suppose maintenant que x et w sont des entiers, on peut définir la convolution discrète :

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t - a) \quad (9)$$

Enfin, nous utilisons souvent des convolutions sur plus d'un axe à la fois. Par exemple, si nous utilisons une image bidimensionnelle x comme entrée :

$$s(i, j) = (x * w)(i, j) = \sum_m \sum_n x(m, n)w(i - m, j - n) \quad (10)$$

Dans la terminologie du réseau convolutif, le premier argument (dans cet exemple, la fonction x) de la convolution est souvent appelé l'entrée (input) et le second argument (dans cet exemple, la fonction w) comme noyau (kernel). La sortie est parfois appelée feature map.

Couche convolutive

Les réseaux de neurones traditionnels utilisent la multiplication matricielle par une matrice de paramètres avec des paramètres séparés décrivant l'interaction entre chaque unité d'entrée et chaque unité de sortie. Cela signifie que chaque unité de sortie interagit avec chaque unité d'entrée ce qui n'est pas le cas des réseaux de neurones convolutifs. Ceci est accompli en rendant le noyau plus petit que l'entrée. Par exemple,

lors du traitement d'une image, l'image d'entrée peut avoir des milliers ou des millions de pixels, mais nous pouvons détecter de petites caractéristiques significatives telles que les bords avec des noyaux qui n'occupent que des dizaines ou des centaines de pixels. Cela signifie que nous pouvons stocker moins de paramètres, ce qui réduit les besoins en matière de mémoire du modèle et améliore son efficacité. Cela signifie également que le calcul des sorties nécessite moins d'opérations. Ces améliorations en matière d'efficacité sont généralement assez importantes. (figure 49).

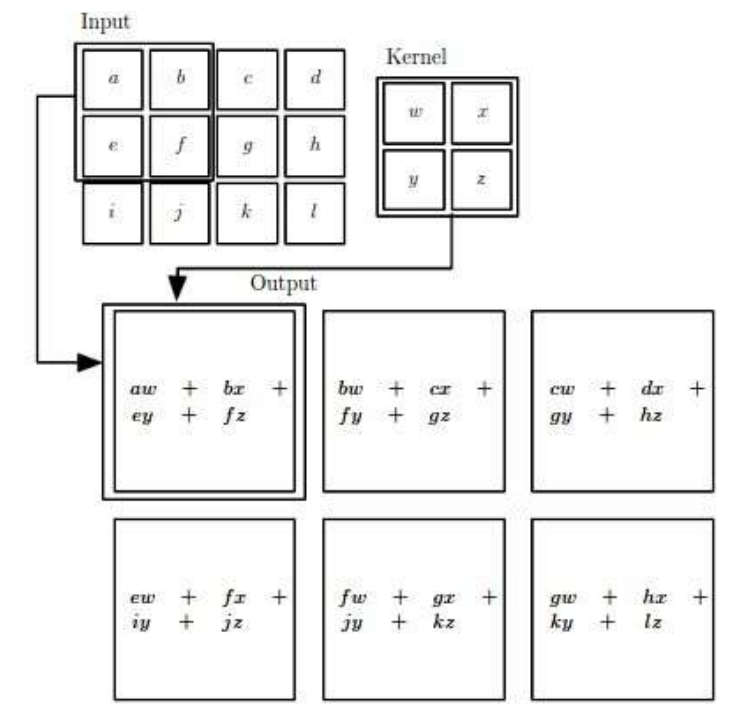


Figure 49 : Résultat d'une convolution 2D

La convolution s'appuie sur trois idées importantes qui peuvent aider à améliorer un système de ML : interactions clairsemées (sparse interactions), partage de paramètres, (parameter sharing) et représentations équivariantes (equivariant representations).

A. Interactions clairsemées (sparse interactions)

Ce phénomène est mis en évidence par la figure 51 suivant la nature de l'unité d'entrée, x_3 , et les unités de sortie qui sont affectées par cette unité. Lorsque s'est formé par convolution avec un noyau de largeur 3, seules trois sorties sont affectées par x (Figure 51 a). Lorsque s'est formé par une multiplication matricielle, toutes les sorties sont atteintes par x_3 (figure 51 b).

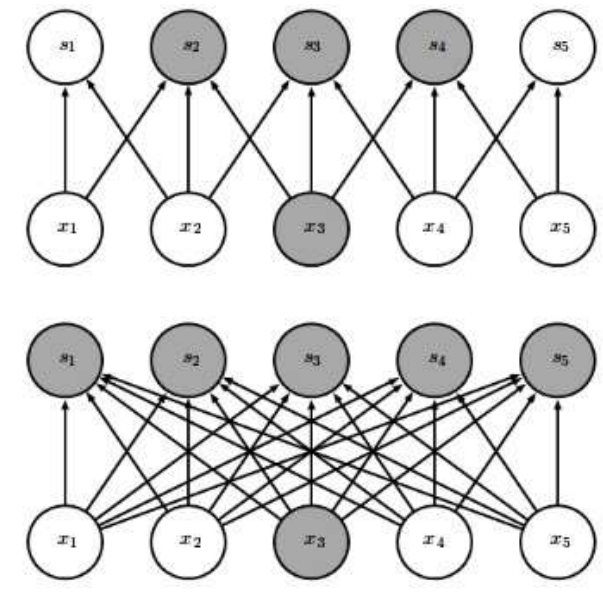


Figure 51: Sparse interactions (a) s est formé par convolution avec un noyau de largeur 3, (b) s est formé par une multiplication matricielle

B. Partage de paramètre (parameter sharing)

Le partage des paramètres se réfère à l'utilisation du même paramètre pour plus d'une fonction dans un modèle. Dans un réseau de neurones convolutif, chaque élément du noyau est utilisé à chaque position de l'entrée (sauf peut-être quelques-uns des pixels des bords, selon le choix de conception concernant la frontière). Le paramètre sharing utilisé par l'opération de convolution signifie que, plutôt que d'apprendre un ensemble de paramètres distincts pour chaque emplacement, nous n'apprenons qu'un ensemble ce qui réduit encore davantage les exigences de stockage du modèle. Dans l'exemple de la figure 50 on a pris une image 1000x1000 (figure 50

a). Le non paramètre sharing nous oblige à concevoir une couche cachée de 10^6 neurones, chaque neurone est connecté à 10^6 pixels, en tout ça fait 10^{12} paramètres. Avec le paramètre sharing nous utilisons un noyau 10×10 et une couche cachée de 10^6 neurones, le nombre de paramètres est de 10^8 (figure 50 b).

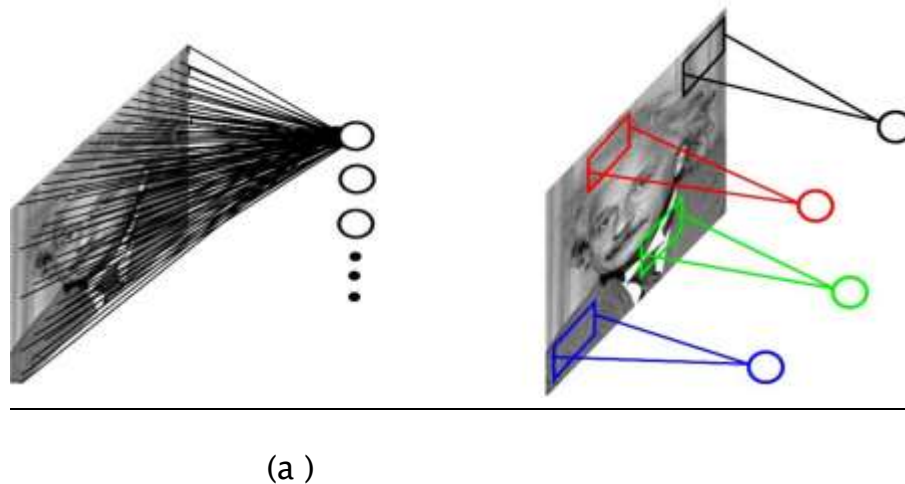


Figure 50: Exemple image 1000x1000. (a) résultat du non paramètre sharing (b) résultat du paramètre sharing

C. Représentation équivariante (equivariant representations)

Dans le cas de la convolution, la particularité du parameter sharing fait que la couche possède une propriété appelée equivariant representations. Pour dire qu'une fonction est équivariante signifie que si l'entrée change, la sortie change de la même manière. Spécifiquement, une fonction $f(x)$ est équivariante à la fonction g si $f(g(x)) = g(f(x))$. Dans le cas du traitement d'image, la convolution crée une carte 2-D de certaines caractéristiques apparaissant dans l'entrée. Si nous déplaçons l'objet dans l'entrée, sa représentation va se déplacer de la même façon dans la sortie. Par exemple, dans le cas d'une image, Il est utile de détecter les arêtes dans la première couche d'un réseau convolutif. Les mêmes arêtes apparaissent plus ou moins partout dans l'image, donc il est pratique de partager des paramètres sur l'image entière.

Couche de pooling

Une architecture atypique d'un réseau convolutif se compose de trois types de couches différentes. D'abord une couche convolutive pour générer un ensemble d'activations linéaires ensuite, on les fait passer à travers une couche d'activation non linéaire telle que Rectified Linear Unit (ReLU), enfin on utilise la fonction pooling. Ce dernier permet :

- De réduire progressivement la taille des représentations afin de réduire la quantité de paramètres et de calcul dans le réseau et, par conséquent, de contrôler également le sur-apprentissage.
- L'invariance aux petites translations
- D'être utile lorsqu'on préfère savoir si une caractéristique est présente plutôt que la région de sa présence.
- Plusieurs type de pooling différent (MAX pooling (très populaire), AVG pooling, ...)

Un exemple de résultat d'application de l'Average pooling et du Max pooling est donné par la figure 52. A gauche de la figure on a appliqué l'Average pooling : chaque case correspond à la moyenne du carré d'entrée de la même couleur, ex de la case jaune : $(1+3+1+3)/4 = 2$. A droite un exemple de l'application du Max pooling: chaque case correspond à la valeur maximum du carré d'entrée de la même couleur, ex de la case bleu : $\max(5,7,5,7) = 7$

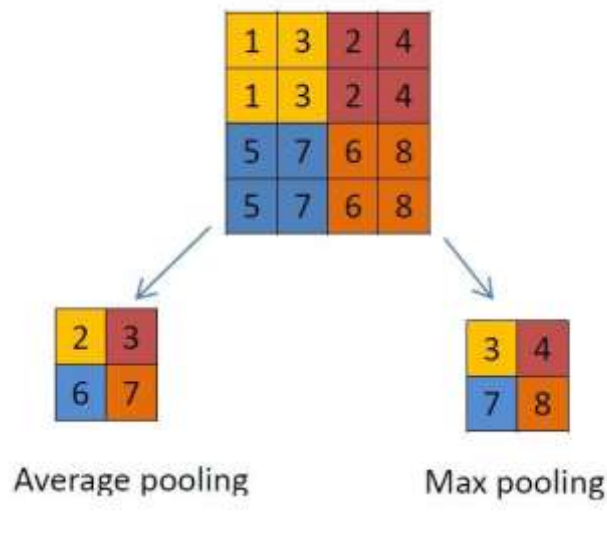


Figure 52 : exemple d'applications du Pooling, (à gauche) Average pooling, le Max pooling (à droite)

Perceptron

Après avoir extrait les caractéristiques des entrées, on attache à la fin du réseau un perceptron ou bien un MLP. Le perceptron prend comme entrée les caractéristiques extraites et produit un vecteur de N dimensions où N est le nombre de classe ou chaque élément est la probabilité d'appartenance à une classe. Chaque probabilité est calculée à l'aide de la fonction softmax dans le cas où les classes sont exclusivement mutuelles.

$$softmax(z)_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (11)$$

Schéma Réseau de neurones convolutif

Un schéma généralisé d'un réseau de neurones convolutif qui reçoit une image 2D comme entrée et qui est composé d'une couche convolutive, une fonction d'activation non linéaire, une couche MAX pooling et enfin un perceptron multi couche est donné par la figure 53.

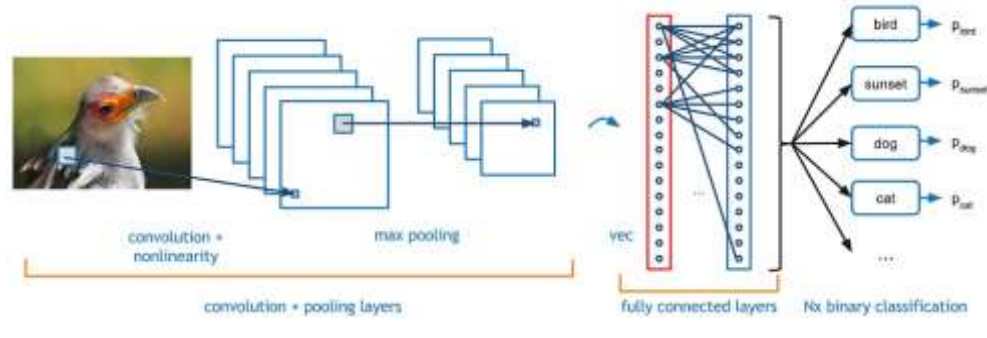


Figure 49 Diagramme blocks d'un réseau de neurones convolutif

Quelques réseaux convolutifs célèbres

Le Deep Learning dépasse la performance humaine (figure 53) Parmi les réseaux qui ont déjà participé à ce challenge on trouve :

- LeNet [18] : Les premières applications réussies des réseaux convolutifs ont été développées par Yann LeCun dans les années 1990. Parmi ceux-ci, le plus connu est l'architecture LeNet utilisée pour lire les codes postaux, les chiffres, etc.
- AlexNet[19] : Le premier travail qui a popularisé les réseaux convolutifs dans la vision par ordinateur était AlexNet, développé par Alex Krizhevsky, Ilya Sutskever et Geoff Hinton. AlexNet a été soumis au défi ImageNet ILSVRC en 2012 et a nettement surpassé ses concurrents. Le réseau avait une architecture très similaire à LeNet, mais était plus profond, plus grand et comportait des couches convolutives empilées les unes sur les autres (auparavant, il était commun de ne disposer que d'une seule couche convolutifs toujours immédiatement suivie d'une couche de pooling).
- ZFnet[20] : Le vainqueur de ILSVRC challenge 2013 était un réseau convolutif de Matthew Zeiler et Rob Fergus. Il est devenu ZFNet (abréviation de Zeiler et Fergus Net). C'était une amélioration de AlexNet en ajustant les hyper-paramètres de l'architecture, en particulier en élargissant la taille des couches convolutifs et en réduisant la taille du noyau sur la première couche.

- GoogLeNet[21] : Le vainqueur de ILSVRC challenge 2014 était un réseau convolutif de Szegedy et al. De Google. Sa principale contribution a été le développement d'un *module inception* qui a considérablement réduit le nombre de paramètres dans le réseau (4M, par rapport à AlexNet avec 60M). En outre, ce module utilise le global AVG pooling au lieu du PMC à la fin du réseau, ce qui élimine une grande quantité de paramètres.
- ResNet[22]: Residual network développé par Kaiming He et al. A été le vainqueur de ILSVRC 2015. Il présente des sauts de connexion et une forte utilisation de la batch normalisation. Il utilise aussi le global AVG pooling au lieu du PMC à la fin.

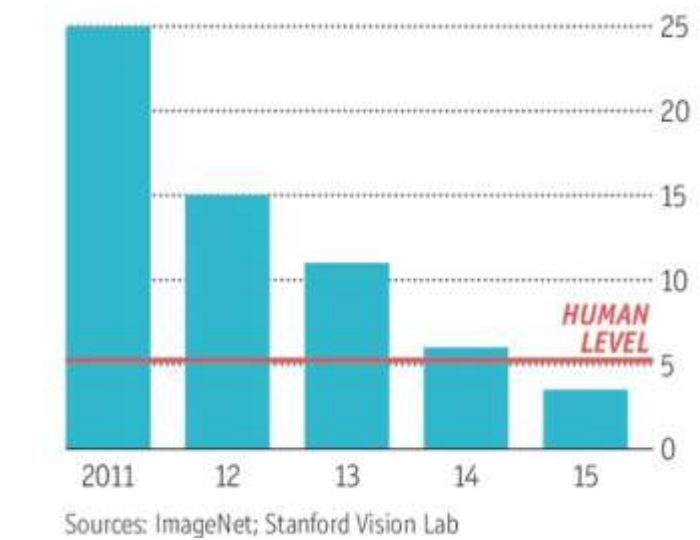


Figure 53 : Le taux d'erreur dans ImageNet Visual recognition Challenge. Le Deep Learning dépasse la performance humaine.

8.6. Optimisation pour l'apprentissage en Deep Learning

Les variantes de la descente de gradient

La descente de gradient est la méthode la plus célèbre pour optimiser un réseau de neurones où on doit minimiser une fonction objective $J(\theta)$ caractérisée par les paramètres θ en les mettant à jour dans la direction opposée de celle de gradient de la fonction objective. Le taux d'apprentissage η détermine l'importance de l'étape qu'on va prendre pour atteindre le minimum local.

Il existe trois variantes de cette méthode. En fonction de la quantité de données, nous ferons un compromis entre la précision de la mise à jour des paramètres et le temps d'exécution de cette mise à jour.

A. Descente de gradient par lots (Batch gradient descent)

C'est la descente de gradient classique, on calcule le gradient de la fonction coût aux paramètres θ pour tout l'ensemble d'apprentissage

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta) \quad (18)$$

Comme on doit calculer le gradient pour tout l'ensemble de données pour exécuter juste une seule mise à jour, cette méthode peut être très lente et irréalisable si les données ne peuvent pas être stockées en mémoire.

Avec cette méthode on est sûr de converger vers l'optimum global si la fonction cout est convexe et vers un optimum local si elle est non convexe.

B. B Descente de gradient stochastique

La descente de gradient stochastique (SGD) met à jour les paramètres pour chaque exemple de l'ensemble de données $x(i)$ et label $y(i)$.

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x(i); y(i)) \quad (19)$$

Cette méthode est plus rapide mais les mises à jours des paramètres trop fréquentes causent à la fonction objectif des oscillations (figure 54), ces oscillations d'une part permettent d'atterrir dans des minimums locaux potentiellement meilleurs mais d'autre part rendent la convergence plus difficile. Cependant il a été montré qu'en baissant le taux d'apprentissage η , SGD montre la même convergence que la batch gradient descent.

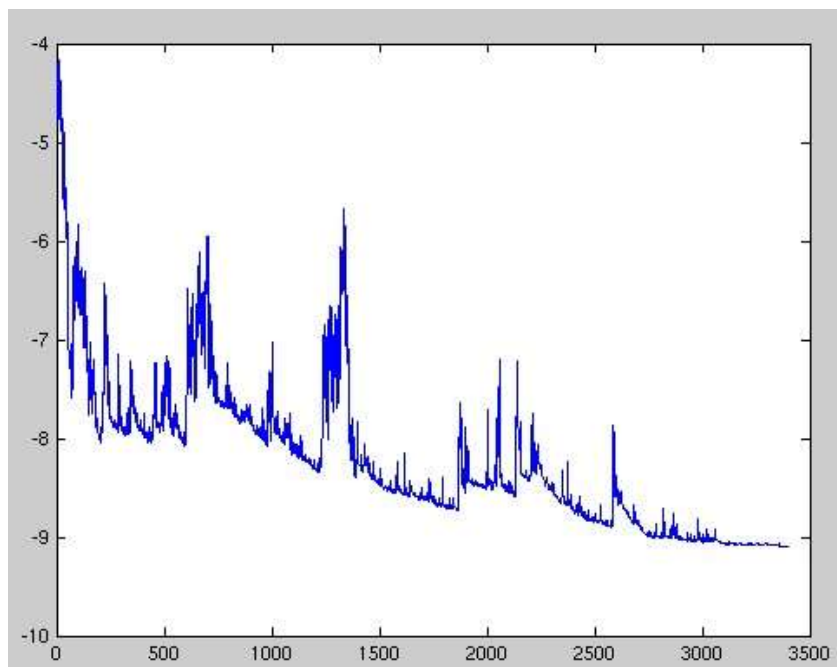


Figure 54 : Les oscillations causées par SGD

C. Mini-batch gradient descent

Cette méthode prend le meilleur des deux méthodes et met à jour les paramètres pour chaque mini groupes de n exemples :

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x(i:i+n); y(i:i+n)) \quad (20)$$

Cette méthode réduit la variance des mises à jour des paramètres ce qui conduit à une convergence plus stable. Généralement les mini groupes contiennent de 50 à 256 exemples. C'est une méthode de choix pour entraîner un réseau de neurones.

La descente de gradient classique n'offre pas toujours une bonne convergence et pose quelques challenges qui ont besoin d'être résolus :

- Le choix du taux d'apprentissage est difficile, s'il est trop petit cela peut emmener à une convergence trop lente, s'il est trop grand il peut causer des oscillations à la fonction coût voir même ne pas converger du tout.
- Le même taux d'apprentissage est appliqué à tous les paramètres. Si les caractéristiques des observations dont on dispose ont une fréquence différente, on peut vouloir appliquer une plus grande mise à jour aux paramètres pour les caractéristiques qui reviennent plus rarement.
- Un autre challenge de la minimisation des fonctions coût non convexe très courantes dans les réseaux de neurones c'est éviter d'être piégé dans des optimums locaux. Dauphin et al. [23] soutiennent que la difficulté ne survient pas des minimum locaux mais des points col (saddle point) (Figure 55).

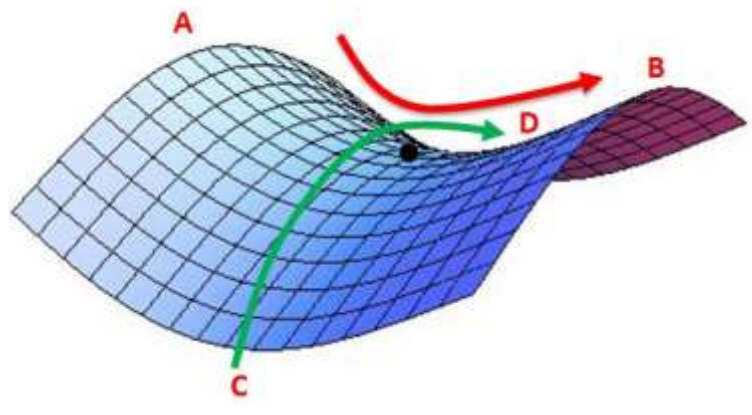


Figure 55 : Saddle point : Le point noir joue le rôle du minimum par rapport à l'axe A-B et du maximum par rapport à l'axe C-D

8.7. Algorithmes d'optimisation de la descente de gradient

Dans ce qui va suivre, on exposera quelques algorithmes largement utilisés par la communauté du Deep Learning pour résoudre les challenges cités précédemment.

8.7.1. Adagrad

Adagrad [24] est un algorithme d'optimisation basé sur la descente de gradient qui ne fait qu'adapter le taux d'apprentissage aux paramètres, effectuant de plus grandes mises à jour pour les caractéristiques peu fréquentes et de plus petites pour les caractéristiques plus fréquentes. Dean et al. [24] ont trouvé que Adagrad a grandement amélioré la robustesse de SGD et utilisé pour l'apprentissage de grands réseaux de neurones chez Google qui –entre autres– ont appris à reconnaître les chats dans les vidéos Youtube. De plus Pennington et al. [25] ont utilisé Adagrad pour l'apprentissage de GloVe word embeddings, comme les mots les moins fréquents demandent de plus grandes mises à jour que ceux qui sont plus fréquents.

Précédemment, il a été effectué les mises à jour des paramètres θ en utilisant le même taux d'apprentissage η . Adagrad utilise un taux d'apprentissage différent pour chaque paramètre θ_i et à chaque étape t .

Soit $g_{t,i}$ le gradient de la fonction objective sachant le paramètre θ_i à l'étape t :

$$g_{t,i} = \nabla J(\theta_i) \quad (23)$$

La mise à jour de chaque paramètre θ_i à l'étape t devient :

$$\theta_{t+1,i} = \theta_{t,i} - \eta \cdot g_{t,i} \quad (24)$$

Dans sa règle de mise à jour, Adagrad modifie le taux d'apprentissage général η à chaque étape t pour chaque paramètre θ_i basé sur les gradients passés calculés pour θ_i :

$$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{G_{t,i} + \epsilon}} \cdot g_{t,i} \quad (25)$$

$G_t \in \mathbb{R}^{d \times d}$ est une matrice diagonale où chaque élément i,i est la somme des

carrés de gradient sachant θ_t à l'étape t et ϵ généralement fixé à $1e-8$ est utilisé pour éviter la division par zéro. Fait intéressant, sans l'opération racine carrée, l'algorithme produit des résultats plus mauvais.

Un des avantages de Adagrad c'est qu'il élimine le besoin de modifier manuellement le taux d'apprentissage.

L'inconvénient majeur de Adagrad est l'accumulation des carrés des gradients dans le dénominateur. Comme chaque terme ajouté est positif, l'accumulation des sommes continue d'augmenter au cours de l'apprentissage ce qui va faire baisser le taux d'apprentissage et éventuellement devenir infinitésimalement petit au point que l'algorithme n'est plus capable d'acquérir de nouvelles connaissances.

8.7.2. Adam optimizer

Adaptive Moments (ADAM) optimizer [26] est une autre méthode qui calcule un taux d'apprentissage adaptatif pour chaque paramètre. En plus de stocker une moyenne décroissante exponentielle des précédents carrés des gradients v_t comme RMSprop, ADAM conserve aussi moyenne décroissante exponentielle des précédents gradients m_t

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \end{aligned} \quad (27)$$

m_t et v_t sont respectivement des estimations d'ordre 1 (la moyenne) et d'ordre 2 (la variance) de gradient. Comme m_t et v_t sont initialisés comme des vecteurs de 0, les auteurs ont observé qu'ils sont biaisés vers zéro surtout durant les premières étapes, et surtout lorsque le coefficient de décroissance est petit (β_1, β_2 proche de 1) alors ils ont calculé une correction des biais pour les estimations du premier et du

deuxième ordre :

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (28)$$

Puis vient la mise à jour des

paramètres

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (29)$$

Les auteurs proposent des valeurs par default de 0.9 pour β_1 et 0.999 pour β_2

8.8. Limites théoriques de l'optimisation

Plusieurs résultats théoriques montrent qu'il existe une limite de performance de n'importe quel algorithme d'optimisation qu'on peut concevoir pour un réseau de neurones.

Certains résultats théoriques s'appliquent uniquement au cas où les unités de réseaux ne produisent que des valeurs discrètes. Des résultats théoriques révèlent qu'il existe des classes de problèmes insolubles, mais il peut être difficile de dire si un problème particulier appartient à cette classe. D'autres résultats montrent que trouver une solution pour un réseau d'une taille donnée est insoluble, mais en pratique, nous pouvons trouver une solution facilement en utilisant un réseau plus large pour lequel de nombreux paramètres correspondent à une solution acceptable. En outre, dans le contexte de l'apprentissage d'un réseau de neurones, nous ne nous soucions généralement pas de trouver le minimum exact d'une fonction, mais nous cherchons seulement à réduire sa valeur de manière suffisante pour obtenir une bonne erreur de test. L'analyse théorique de savoir si un algorithme d'optimisation peut atteindre cet objectif est extrêmement difficile.

8.9. Les fonctions d'activation

La fonction d'activation est une fonction mathématique appliquée à un signal en sortie d'un neurone artificiel. Le terme de "fonction d'activation" vient de l'équivalent biologique "potentiel d'activation", seuil de stimulation qui, une fois atteint entraîne une réponse du neurone. La fonction d'activation est souvent une fonction non-linéaire. Son but est de permettre aux réseaux de neurones d'apprendre des fonctions plus complexes qu'une simple régression linéaire car le fait de multiplier les poids d'une couche cachée est juste une transformation linéaire.

8.9.1. Caractéristiques

- Non-linéarité: Quand une fonction est non linéaire, un réseau neuronal à deux couches peut être considéré comme un approximateur de fonction universel.
- Partout différentiable : Cette propriété permet de créer des optimisations basées sur les gradients.
- Étendue : Quand la plage d'activation est finie, les méthodes d'apprentissage basées sur les gradients sont plus stables (impact sur un nombre de poids limités). Quand la plage est infinie, l'apprentissage est généralement plus efficace (impact sur davantage de poids).
- Monotone : Lorsque la fonction est monotone, la surface d'erreur associée avec un modèle monocouche est certifié convexe.
- Dérivée monotone : Les fonctions à dérivée monotone ont été montrées comme ayant une meilleure capacité à généraliser dans certains cas. Ces fonctions permettent d'appliquer des principes comme le rasoir d'Ockham.
- Identité en 0 : Ces fonctions permettent de faire un apprentissage rapide en initialisant les poids de manière aléatoire. Si la fonction ne converge pas

vers l'identité en 0, alors un soin spécial doit être apporté au lors de l'initialisation des poids.

8.9.2. Quelque fonction d'activation

A. Unités linéaires rectifiées (Rectified Linear Units (ReLU))

Les ReLu utilisent la fonction d'activation

$$g(z) = \max\{0, z\} \quad (31)$$

ReLU sont faciles à optimiser car ils sont tellement similaires aux unités linéaires.

La seule différence entre une unité linéaire et ReLu est que ReLu délivre zéro sur tous R- ceci rend les dérivées via ReLu importantes chaque fois que l'unité est active.

Des recherches montrent que ReLu sont un modèle encore meilleur de neurones biologiques et produisent une performance égale ou meilleure que la tangente hyperbolique et un de leur avantage majeur est la réduction de la probabilité du vanishing gradient. Cela se produit lorsque $z > 0$. Dans ce cas, le gradient a une valeur constante puisque la dérivée est égale à 1 partout. En revanche, le gradient des sigmoïdes devient de plus en plus petit car la valeur absolue de z augmente. Le gradient constant de ReLu entraîne un apprentissage plus rapide.

Lors de l'initialisation des paramètres il peut être judicieux d'initialiser les valeurs des biais à une petite valeur positive, telle que 0,1. Il est très probable que ReLu seront initialement actives pour la plupart des entrées de l'ensemble d'apprentissage et permettent aux dérivées de passer à travers.

Il existe plusieurs généralisations de ReLu. La plupart de ces généralisations fonctionnent de manière comparable et de temps à autre, se comportent mieux.

Un inconvénient de ReLu est qu'ils ne peuvent pas apprendre par la descente de gradient sur des exemples pour lesquels leur activation est nulle "*dying ReLu*" *problem*.

Une variété de générations de ReLu garantit qu'elles reçoivent le gradient partout.

Les généralisations de ReLu (figure 56) sont basées sur l'utilisation d'une pente

non nulle α_i lorsque $z_i < 0$: $h_i = g(z, \alpha) = \max(z_i, 0) + \alpha_i \min(0, z_i)$. Absolute value rectification fixe $\alpha_i = -1$ pour obtenir $g(z) = |z|$. Leaky ReLu (LReLU) fixe α_i à une petite valeur comme 0.01 alors que parametric ReLu (PReLU) traite α_i comme un paramètre apprenable.

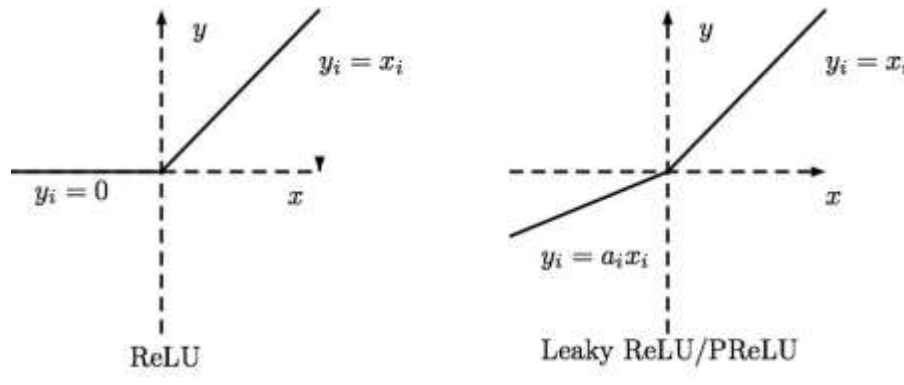


Figure 56 : La fonction d'activation ReLu (à gauche), (à droite) la fonction Leaky ReLu et Parametric ReLu qui corrige l'inconvénient de ReLu.

B. Sigmoïde et tangente hyperbolique

Avant l'introduction de ReLu, la plupart des réseaux neuronaux utilisent la Sigmoïde ou tanh :

la Sigmoïde :

$$g(z) = \frac{1}{1 + e^{-\lambda z}} \quad (32)$$

tanh :

$$g(z) = \frac{1 - e^{-2z}}{1 + e^{-2z}} \quad (33)$$

Ce qu'on reproche aux unités sigmoïdales c'est qu'elles saturent dans la plupart de leur domaine : elles saturent à une valeur élevée lorsque z est très positif et à une valeur faible lorsque z est très négatif et ne sont que très sensibles à leur entrée lorsque z est proche de 0. La saturation généralisée des unités sigmoïdales peut rendre l'apprentissage par la méthode de la descente de gradient très difficile. Pour cette raison, leur utilisation dans les couches cachées est maintenant déconseillée.

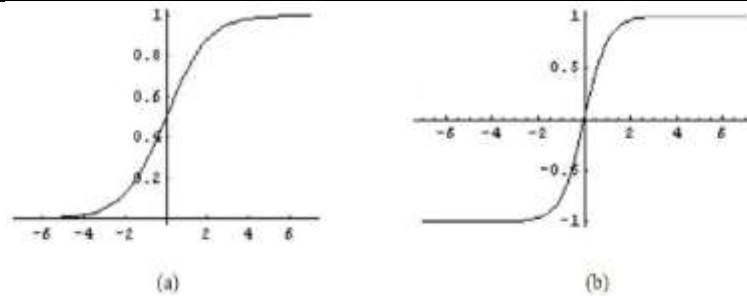


Figure 57 : Les fonctions d'activation Sigmoïde (a) et tangente hyperbolique (b).

8.10. Conclusion

Dans cette partie, nous avons vu ce qu'est le Deep Learning et comment il se différencie des algorithmes de ML traditionnelles. Nous avons vu quelques étapes majeures de son évolution et les exploits qui ont été accomplis avec. Nous avons introduit quelques méthodes utilisées par la communauté du Deep Learning et nous avons expliqué le principe de chacun. Nous avons également présenté les réseaux convolutifs qui font encore l'objet de recherches intensives.

CHAPITRE III

METHODES ET MATERIEL

1. MODÈLE PROPOSÉ BASÉ SUR LE TRANSFERT LEARNING

Dans ce travail, nous avons proposé un système intelligent d'aide à la décision clinique pour la détection précoce de maladies pulmonaires à partir de radiographies pulmonaires. Ce modèle utilise l'apprentissage en profondeur grâce à sept réseaux de neurones à convolution extraits de la base de données ImageNet déjà pré-entraînés. Chaque réseau prédira une classe comme image d'entrée. L'avantage de notre modèle est qu'il attribuera un score à chaque prédiction de classe et n'enverra que la prédiction qui obtient le score le plus élevé.

1.1. Réseaux de neurones convolutifs utilisant l'apprentissage par transfert

Dans un réseau CNN traditionnel, l'image d'entrée passe généralement par une séquence de convolution et de mise en commun (Pooling) pour compresser la largeur et la hauteur de l'image en augmentant la profondeur de la caractéristique. Le but de cette couche de convolution et de regroupement empilé est d'apprendre la représentation de l'image ou des entités. La fonction apprise sera alors l'entrée des couches entièrement connectées pour la classification. Dans de nombreuses architectures précédentes, les chercheurs ont tenté d'empiler de plus en plus de couches de convolution et de mise en commun, d'où le terme Deep Learning, avec l'hypothèse que des réseaux plus profonds apprendront une meilleure représentation de l'objet d'intérêt. Cependant, en pratique, cette conception souffre d'un problème de gradient de disparition et les performances se dégradent au fur et à mesure que les réseaux deviennent plus profonds. Souvent, la formation d'un CNN à partir de zéro est généralement difficile car ce processus nécessitait de grandes données de formation ainsi qu'une expertise significative pour sélectionner une architecture de modèle appropriée pour une convergence appropriée.

Pour résoudre ce problème, nous avons utilisé un modèle d'apprentissage par transfert. Ce dernier a été utilisé pour différents types d'applications [27–28]. Ce modèle offre une alternative prometteuse pour affiner un CNN déjà pré-formé sur un grand ensemble de données comme ImageNet en utilisant ses pondérations. Cela nous aide à accélérer le processus de convergence pendant la formation. L'apprentissage par transfert est une technique très utile et a également obtenu des résultats significatifs dans les domaines de la vision par ordinateur et du traitement d'image.

Le modèle proposé est composé de deux modes. Un mode de démarrage pré-entraîné qui transforme les images d'entrée en vecteurs descripteurs. Un autre mode consiste en plusieurs classificateurs fortement connectés entre eux où chaque classificateur donnera à sa sortie sa propre prédiction. De cette manière, la prédiction qui a obtenu un score maximum est celle qui sera retenue par le système global à sa sortie. Plus précisément, le modèle proposé se compose de trois processus principaux comme le montre la figure 58. Le premier processus utilisé est celui du prétraitement d'image suivi d'une augmentation des données. Le prétraitement consistait à éliminer tout ce qui se trouvait à l'extérieur des zones, délimitant les deux poumons tout en gardant une image de 224x224. Par conséquent, l'augmentation du nombre d'images a été obtenue grâce à la rétroaction horizontale, au recadrage aléatoire et à la variation de l'intensité de chaque image.

Le deuxième processus utilisé est celui de l'extraction et de l'apprentissage des fonctionnalités. À cette étape, le processus utilise sept CNN déjà préformés extraits de la base de données ImageNet. De plus, l'utilisation de ces modèles pré-entraînés évite de passer par le long processus d'apprentissage des modèles à partir de zéro. Bien que ces modèles aient déjà été formés sur le jeu de données ImageNet, nous les formerons également sur notre base d'images de radiographie pulmonaire que nous avons prises.

Le troisième processus utilisé est celui de la classification et de la prédiction à

travers un réseau CNN entièrement connecté composé de plusieurs classificateurs. En sortie du réseau, un vecteur de prédiction est obtenu. Cependant, la prédiction finale retenue par le système global est celle qui a obtenu un score maximum.

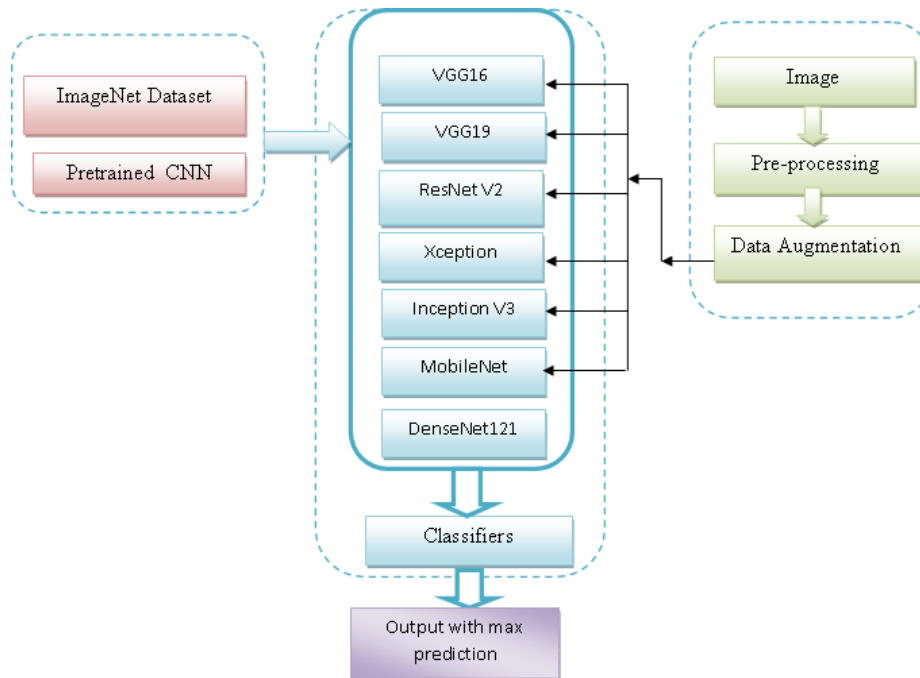


Figure 58: Processus du modèle proposé

1.2. Réseaux de neurones à convolution pré entraînés sur la base ImageNet

Le principal avantage du modèle proposé est d'éviter le passage par le long processus d'apprentissage des modèles à partir de zéro grâce à l'utilisation des poids des couches de modèles déjà préformés extraits de la base de données ImageNet. Dans un modèle de transfert, ces poids déjà prédéfinis seront attribués aux premières couches placées juste après l'entrée dans le modèle, tandis que les poids des dernières couches placées à la sortie du modèle seront entraînés sur les images du nouveau domaine. Dans les sections suivantes, nous présenterons différents modèles d'architecture basés sur un réseau de neurones convolutifs avec différentes couches de profondeur.

1.3. Architecture VGG16

C'est un réseau neuronal convolutif, du Groupe de Géométrie Visuelle (Visual Geometry Group (VGG), département des sciences de l'ingénierie, Université d'Oxford [21]. En 2014. Il a été soumis au Défi de reconnaissance visuelle à grande échelle 2014 (ILSVRC2014) et le modèle atteint 92,7% de précision dans ImageNet. Il existe deux types de réseau neuronal VGG: VGG16 et VGG19.

VGG16 est un modèle de réseau de neurones convolutif contenant 16 couches, proposé par K. Simonyan et A. Zisserman de l'Université d'Oxford dans le document « Very Deep Convolutional Networks for Large-Scale Image Recognition ». Le modèle atteint une précision de test de 92,7% dans le top 5 dans ImageNet, qui est un ensemble de données de plus de 14 millions d'images appartenant à 1000 classes (Figure 59.a).

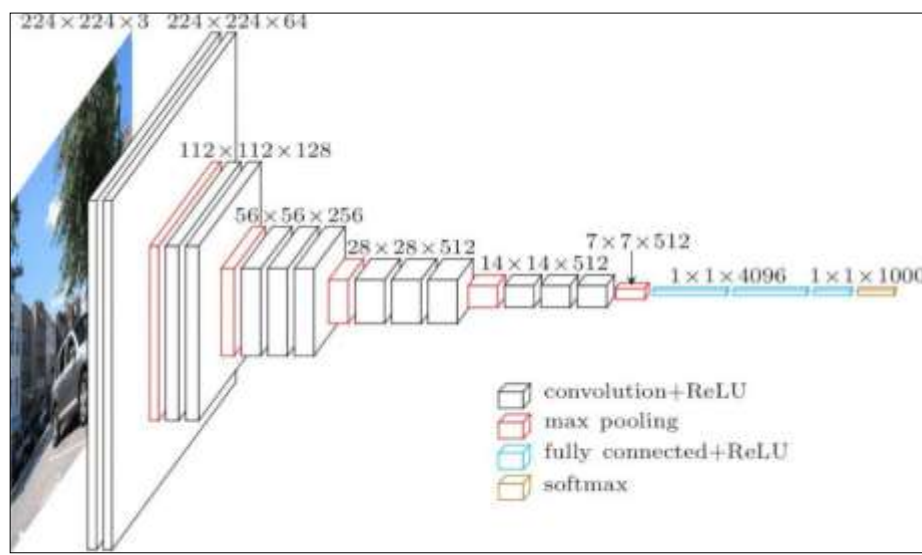


Figure59.a: architecture du réseau VGG-16

Dans notre modèle proposé, nous avons utilisé un modèle avec des couches formables et des couches gelées basé sur un modèle de réseau de neurones convolutif proposé par K. Simonyan [17]. Nous avons utilisé les couches de CONV1-1 à CONV4-3 comme couches pré-entraînées avec des poids déjà fixés. Par conséquent, les couches restantes ont été utilisées pour la formation sur notre ensemble de données d'images radiographiques pulmonaires déjà en place, comme le montre la figure 59.b.

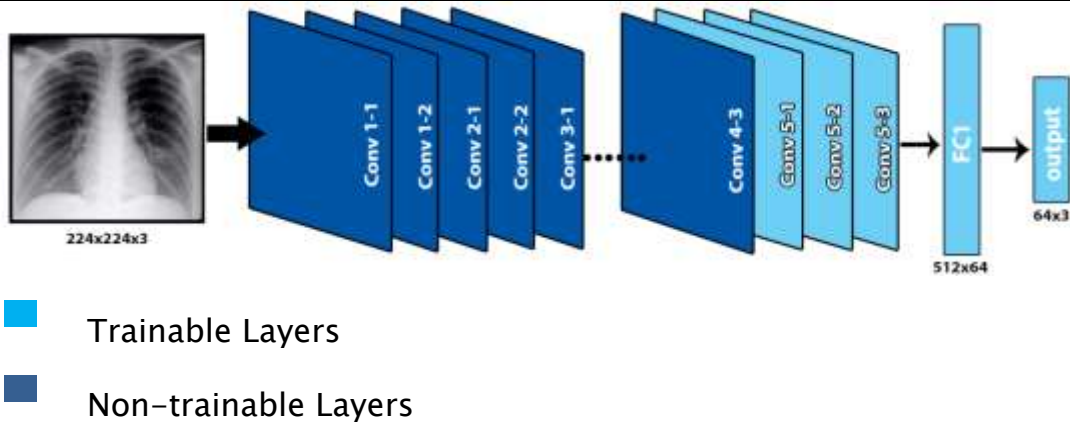


Figure 59.b: Modèle VGG16 avec des couches formables et des couches gelées

1.4. Architecture VGG19

Le modèle VGG-19 est un réseau neuronal convolutif formé basé sur l'architecture VGG-16. Le nombre 19 représente le nombre de couches avec des poids entraînaibles. Au total, nous avons 16 couches convolutives et 3 couches entièrement connectées. Dans le modèle proposé, nous avons utilisé les couches de block1_conv1 à block5_conv1 comme couches pré-entraînées avec des poids déjà fixés. Les couches restantes ont été utilisées pour la formation sur notre jeu de données.

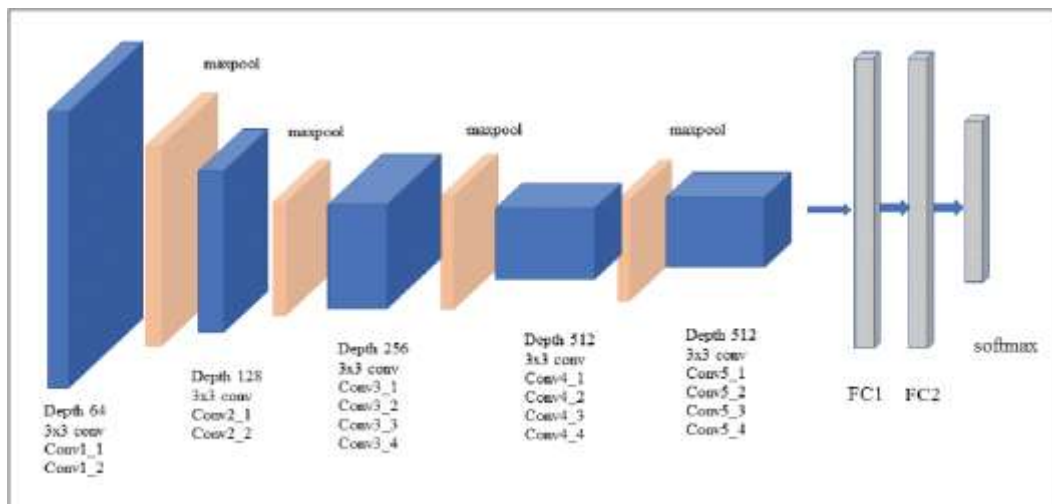


Figure 59.c : Architecture du réseau VGG19

1.5. Architecture MobileNet

L'architecture MobileNet a été proposée par Google. Dans notre modèle proposé, nous avons utilisé les couches de 1 à 75 comme couches pré-entraînées avec des poids déjà fixés. Par conséquent, comme le montre la figure 60, les couches restantes ont été utilisées pour la formation sur notre ensemble de données d'images pulmonaires radiographiques déjà en place.

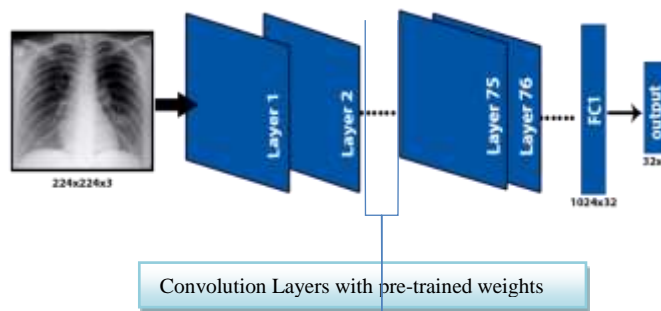


Figure 60: Modèle MobileNet avec des couches formables et des couches gelées

1.6. Architecture InceptionV3

Inception-v3 est un réseau neuronal convolutif d'une profondeur de 48 couches. Il s'agit de la troisième édition du Inception Convolutional Neural Network de Google, initialement présenté lors du Défi de reconnaissance ImageNet.

Inception v3 est un modèle de reconnaissance d'image largement utilisé qui atteint une précision supérieure à 78,1% sur l'ensemble de données ImageNet. Le modèle est l'aboutissement de nombreuses idées développées par plusieurs chercheurs au fil des ans. Il est basé sur l'article original: « Repenser l'architecture de lancement pour la vision par ordinateur » par Szegedy, et. Al [22]. La figure ci-dessous montre l'architecture InceptionV3.

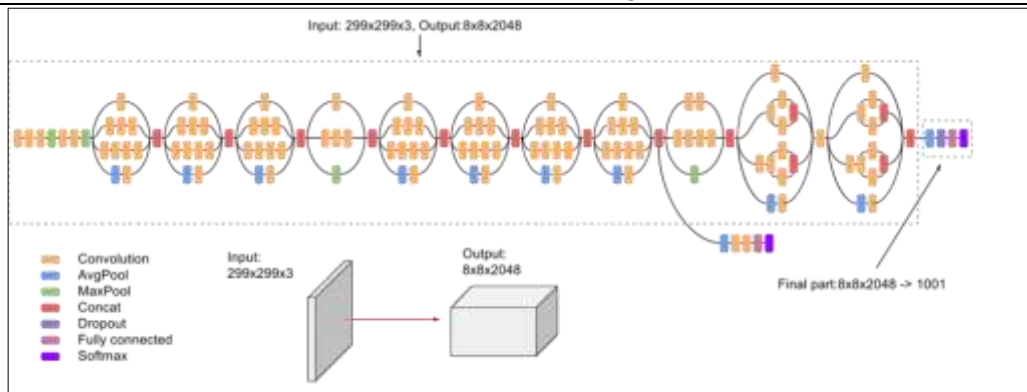
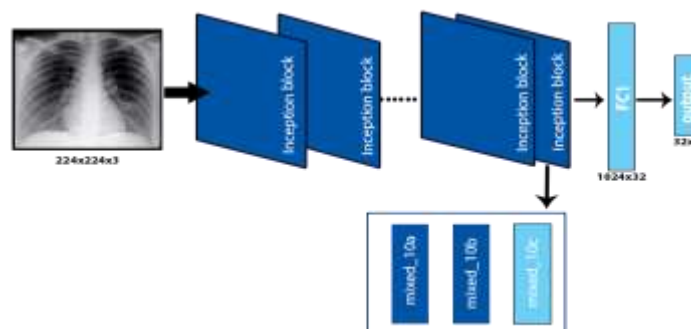


Figure 61.a : Architecture du réseau de neurones à convolution Inception V3.



- Trainable Layers
- Non-trainable Layers

Figure 61.b: Modèle InceptionV3 avec des couches formables et des couches gelées.

1.7. Architecture Xception

Xception est une architecture de réseau de neurones à convolution profonde qui implique des convolutions séparables en profondeur. Elle contient 71 couches, proposées par François Chollet [18]. Xception est une extension de l'architecture inception qui remplace les modules Inception standard par des convolutions séparables en profondeur.

Les convolutions séparables en profondeur sont des alternatives aux convolutions classiques qui sont supposées être beaucoup plus efficaces en termes de temps de calcul. La figure 62.a. montre à quoi ressemble la convolution séparable en profondeur.

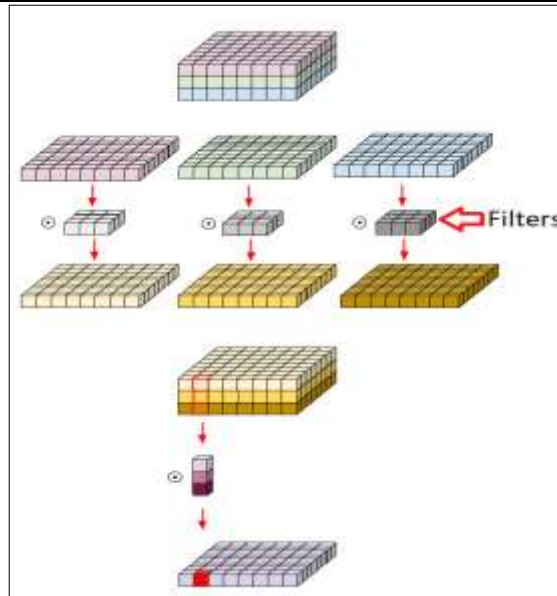
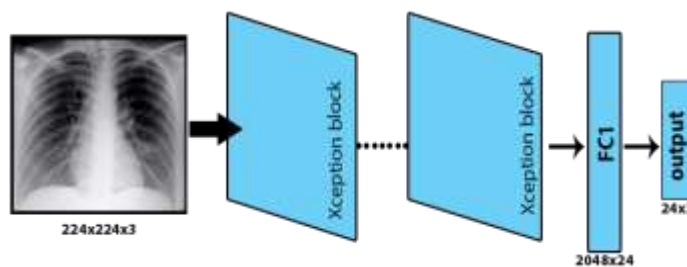


Figure 62.a : Convolution séparable en profondeur

Dans notre modèle proposé, nous avons utilisé toutes les couches pour la formation sur notre ensemble de données d'images pulmonaires radiographiques en utilisant des poids pré-entraînés d'ImageNet comme point de départ, comme le montre la figure 62.b.



■ Trainable Layers

Figure 62.b: Modèle X-ception avec des couches formables et des couches gelées.

1.8. Architecture InceptionResNetV2

Alors que l'architecture InceptionResNetV2 est obtenue suite aux modifications apportées par les chercheurs sur la troisième version de lancement des réseaux CNN. Les chercheurs se sont appuyés sur des connexions résiduelles du réseau Microsoft ResNet pour offrir une version plus approfondie, simple et significative de l'architecture Inception (figure 63.a).

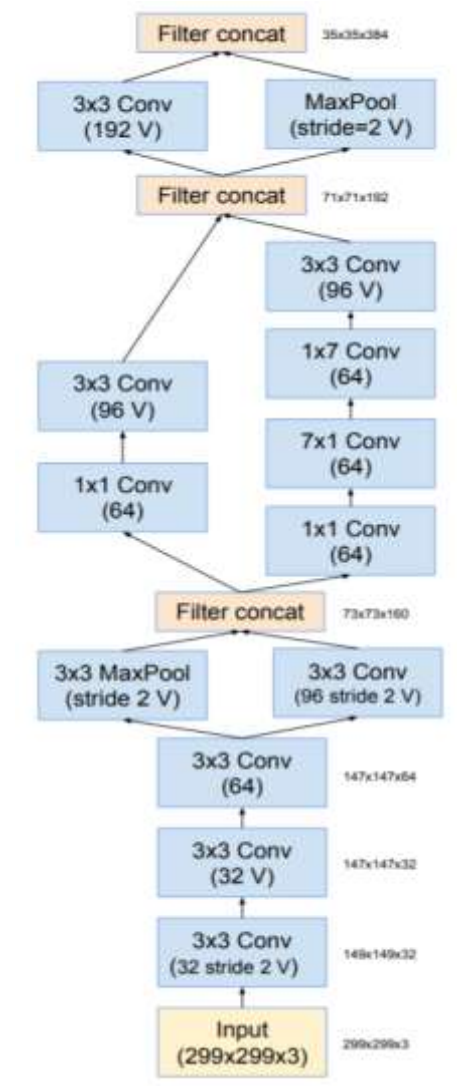


Figure 63.a : Architecture Inception-ResNetV2.

Dans notre modèle proposé, nous avons utilisé les couches de la première à conv2D-58 couches en tant que couches pré-entraînées avec des poids déjà fixés. De cette façon, les couches restantes ont été utilisées pour la formation sur notre ensemble de données d'images radiographiques pulmonaires déjà en place, comme le montre la figure 63.b.

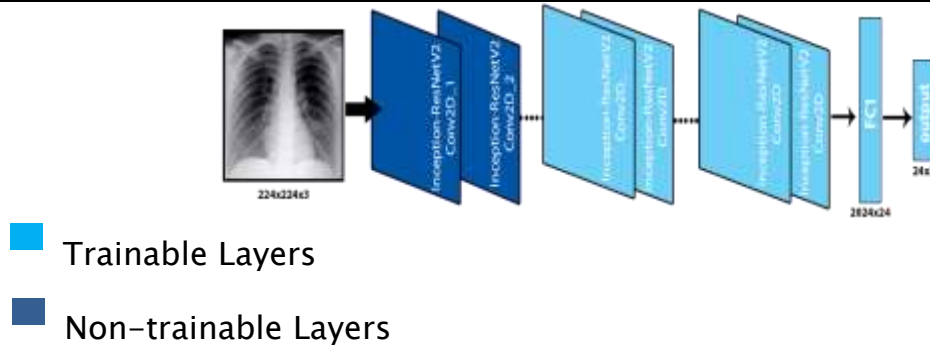


Figure 63.b: Modèle Inception-ResNetV2 avec des couches formables et des couches gelées.

1.9. Architecture de DenseNet121

Huang et coll. [20] ont proposé une nouvelle architecture appelée DenseNet qui exploite davantage les effets des connexions de raccourci. Il connecte toutes les couches directement les unes avec les autres. Dans cette nouvelle architecture, l'entrée de chaque couche se compose des cartes de caractéristiques de toutes les couches antérieures, et sa sortie est transmise à chaque couche suivante. Les cartes d'entités sont agrégées avec une concaténation de profondeur (figure 64.a.).

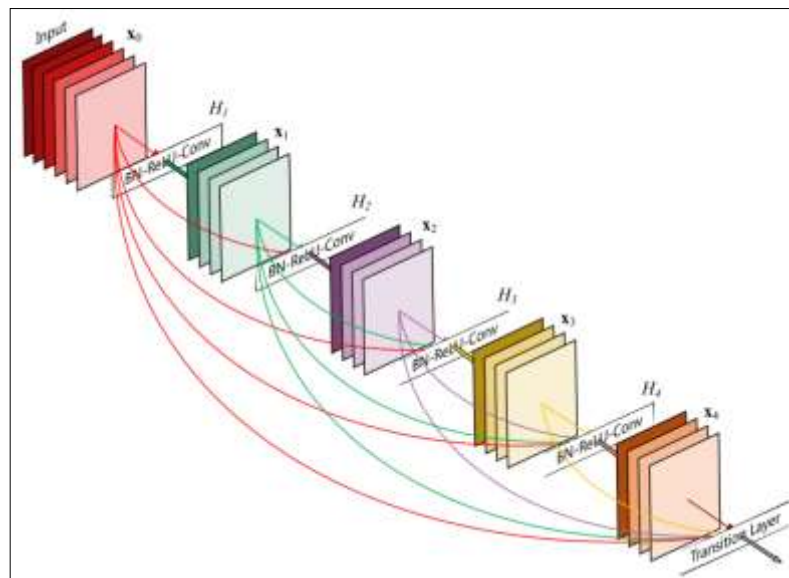
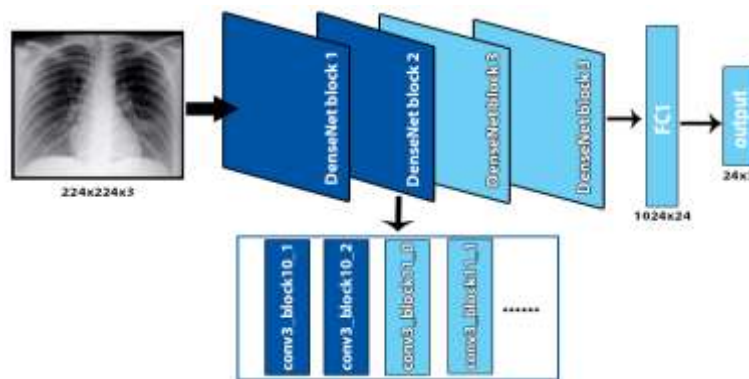


Figure 64.a : Architecture du réseau DenseNet121

Enfin, dans l'architecture DenseNet121, chaque couche a en entrée toutes les sorties des couches qui la précèdent, rendant l'architecture du réseau très dense, permettant une supervision approfondie.

Dans notre modèle proposé, nous avons utilisé les couches de la première à conv3-block10-2-conv couches en tant que couches pré-entraînées avec des poids déjà fixés. Alors que, les couches restantes ont été utilisées pour la formation sur notre ensemble de données d'images radiographiques pulmonaires déjà en place, comme le montre la figure 64.b.



- Trainable Layers
- Non-trainable Layers

Figure 64.b: Modèle DenseNet121 avec des couches formables et des couches gelées.

1.10. Modèle d'empilage

Dans la section restante, nous présenterons notre nouveau modèle d'empilage que nous avons développé. Ce dernier est basé sur les sous-modèles de réseaux de neurones intégrés dans un modèle d'empilement global. Notre nouveau modèle nous a permis de trouver le meilleur moyen de combiner les prédictions de plusieurs modèles existants déjà pré-entraînés. Par conséquent, le modèle d'empilement effectue un apprentissage rapide, réduit le sur-ajustement et produit de meilleures prédictions. Par conséquent, nous développerons un modèle d'empilement utilisant 7

réseaux de neurones comme sous-modèle et un classificateur méta-apprenant. L'idée de base de cette approche est de considérer la prédiction de chaque réseau en lui attribuant un score. Une fois que les sept réseaux ont terminé leur prédiction, le score obtenu par chaque prédiction est compté. Seule la prédiction avec le score le plus élevé est sélectionnée comme sortie du modèle, comme le montre la figure 65.

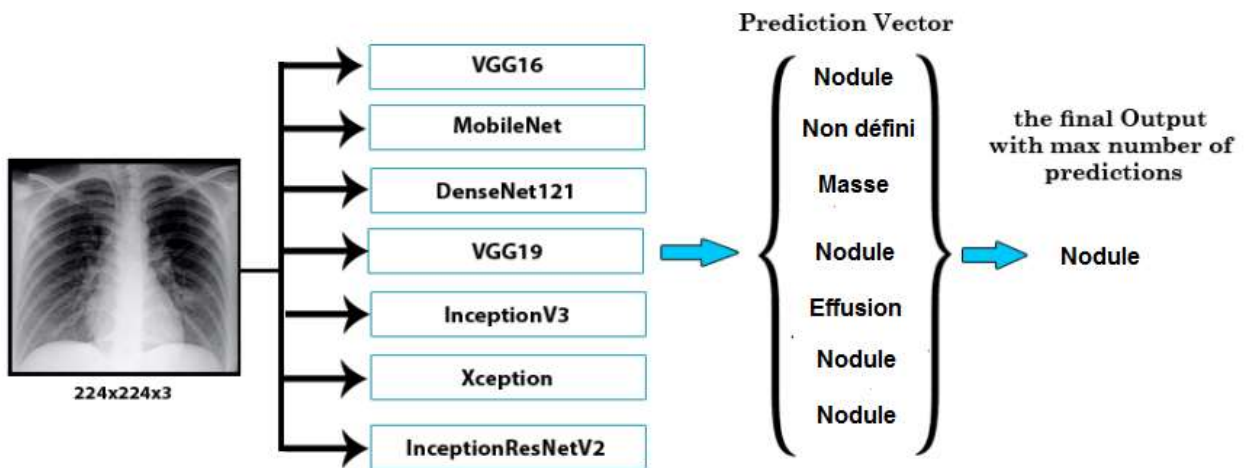


Figure 65: Modèle d'empilement utilisant sept différentes architectures pré-entraînées et vote à la majorité.

2. CARTE D'ACTIVATION DE CLASSE DE GRADIENT (GRAD-CAM)

Dans cette section, nous parlerons de la technique Grad-CAM qui rend les modèles de réseau de neurones convolutifs plus transparents et interprète en visualisant les régions d'entrée sur lesquelles le modèle s'est davantage concentré pour classer et prendre la décision.

2.1. Concept de Grad-CAM

L'idée du Grad-CAM est de conserver les informations de localisation spatiale de l'objet qui sont perdues dans une couche entièrement connectée. Ainsi, la dernière couche de convolution est utilisée car ce sont les neurones qui identifient les parties spécifiques à cette classe. L'architecture de Grad-CAM est montrée par la figure ci-dessous.

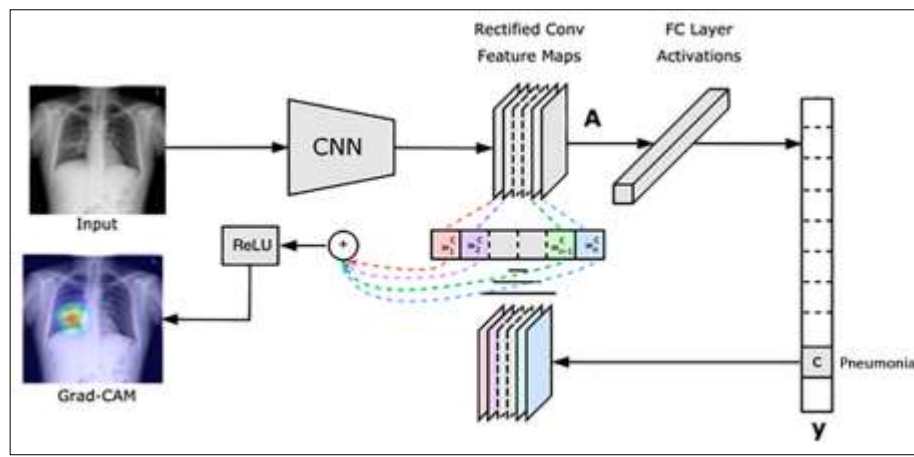


Figure 66 : Processus suivi par l'approche du Grad-CAM

Pour localiser les lésions sur les images radiographiques thoraciques causées par la maladie, nous mettons en évidence les points d'intérêt qui ont fortement contribué à la classification de notre modèle. Pour cela, nous effectuons d'abord une combinaison linéaire des cartes de dégradés de toutes les couches du modèle. La région d'intérêt de dimension $m \times n$ pour une classe (i) est obtenue en calculant d'abord le gradient de la probabilité de présence P_i de chaque classe pour une couche donnée du modèle C_k . La

moyenne globale de ces gradients fixera la pondération des neurones par rapport à la classe (i) [19].

$$\beta_k^i = \frac{1}{Z} \sum_r^m \sum_s^n \frac{\partial P^i}{\partial B_{rs}^k}$$

Ensuite nous appliquons la fonction:

$$L^i = ReLU \left(\sum_k B_k^i B^k \right)$$

Les régions d'intérêt obtenues par notre algorithme correspondent aux types de classes : sont représentées sur la figure 5. Nous avons représenté en jaune les pixels qui présentaient un fort gradient et qui ont énormément contribué à la classification de notre modèle. La couleur bleue a été attribuée aux pixels qui présentaient un dégradé faible et qui n'ont pas contribué à la classification de notre modèle. A noter que dans le cas des images de patients présentant des lésions pulmonaires, notre algorithme s'est focalisé sur la région montrant l'opacité de la lésion.

3. BASE DE DONNÉES RADIOGRAPHIQUES DU THORAX DE L'INSTITUT NATIONAL DE LA SANTÉ (NATIONAL INSTITUTES OF HEALTH NIH)

Nous allons profiter de la publication récente du centre clinique des instituts nationaux de la santé (NIH Clinical Center) une base de données de 112 120 images de radiographie pulmonaire anonymes et leurs données correspondantes pour entraîner notre modèle proposé. Nous allons dans cette section présenter les caractéristiques, le contenu et les différentes annotations apportées à cette base.

3.1. La base de données NIH

Le centre clinique des instituts nationaux de la santé (NIH Clinical Center) a récemment publié plus de 100 000 images de radiographie pulmonaire anonymes et leurs données correspondantes à la communauté scientifique. Cette publication permet aux chercheurs du pays et du monde entier d'accéder librement aux ensembles de données et d'accroître leur capacité à enseigner aux ordinateurs comment détecter et diagnostiquer les maladies. En fin de compte, ce mécanisme d'intelligence artificielle peut amener les cliniciens à prendre de meilleures décisions diagnostiques pour les patients.

Le NIH a compilé l'ensemble de données de plus de 30 000 patients, dont beaucoup avec une maladie pulmonaire avancée. Les patients du NIH Clinical Center, le plus grand hôpital du pays entièrement consacré à la recherche clinique, sont des partenaires de recherche et s'inscrivent volontairement pour participer à des essais cliniques. La confidentialité des patients étant primordiale, l'ensemble de données a été rigoureusement examiné pour supprimer toutes les informations personnellement identifiables avant la publication.

La lecture et le diagnostic des images radiographiques pulmonaires peuvent être une tâche relativement simple pour les radiologues mais, en fait, il s'agit d'un problème de raisonnement complexe qui nécessite souvent une observation attentive et une connaissance des principes anatomiques, de la physiologie et de la pathologie. Ces facteurs augmentent la difficulté de développer une technique cohérente et automatisée pour lire les images radiographiques thoraciques tout en considérant simultanément toutes les maladies thoraciques courantes.

En utilisant cet ensemble de données gratuit, on espère que les établissements universitaires et de recherche à travers le pays seront en mesure d'enseigner à un ordinateur à lire et à traiter de très grandes quantités de scans, pour confirmer les résultats que les radiologues ont trouvés et potentiellement identifier d'autres découvertes qui pourraient avoir été négligé.

En outre, cette technologie informatique avancée peut également être en mesure de:

- ✓ Aider à identifier les changements lents survenant au cours de plusieurs radiographies thoraciques qui pourraient autrement être négligés
- ✓ Apporter une valeur ajoutée aux patients des pays en développement qui n'ont pas accès à des radiologues pour lire leurs radiographies pulmonaires
- ✓ Créer un résident virtuel en radiologie qui pourra plus tard apprendre à lire des images plus complexes comme la tomodensitométrie et l'IRM à l'avenir

Avec un engagement continu envers le partage de données, l'hôpital de recherche NIH prévoit d'ajouter un vaste ensemble de données de tomodensitométrie qui sera également disponible dans les mois à venir.

Les données sont fournies par le NIH Clinical Center et sont disponibles via le site de téléchargement du NIH : <https://nihcc.app.box.com/v/ChestXray-NIHCC> (Figure 67–68).

3.2. Licence et attribution

Il n'y a pas de restrictions concernant l'utilisation des images de radiographie du thorax du NIH. Cependant, l'ensemble de données est soumis aux exigences d'attribution suivantes:

- Fournir un lien vers le site de téléchargement du NIH : <https://nihcc.app.box.com/v/ChestXray-NIHCC>.
- Inclure une citation dans l'article du CVPR 2017: [1] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald Summers, ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases, IEEE CVPR, pp. 3462–3471, 2017
- Confirmer que le NIH Clinical Center est le fournisseur de données.

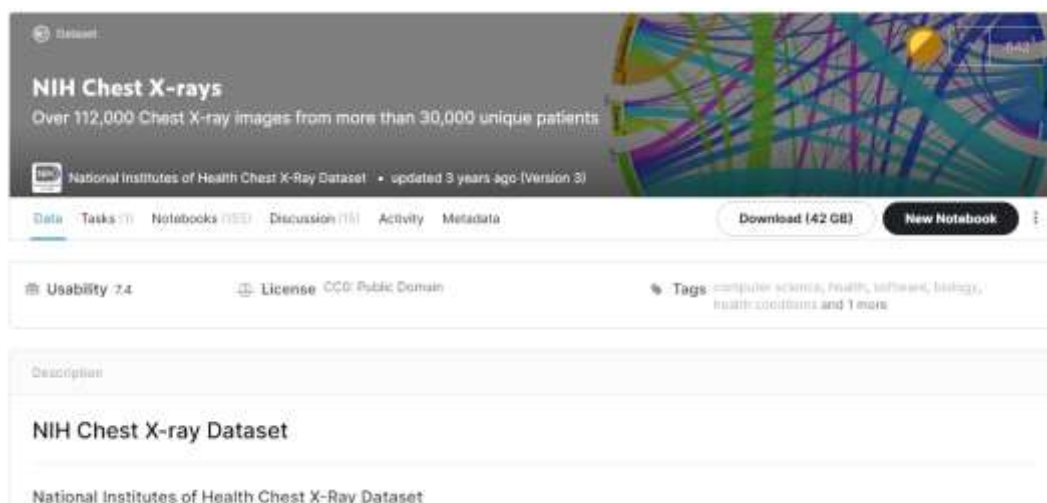


Figure 67 : Lien vers la base NIH

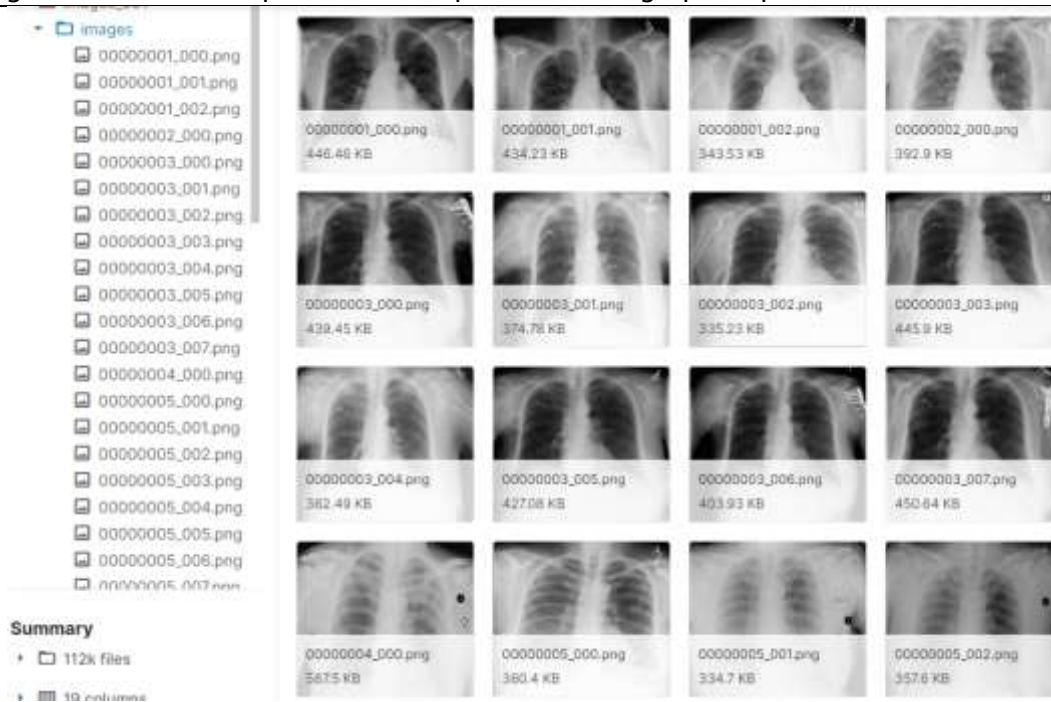


Figure 68 : Exemples d'images stockées dans la base NIH

3.3. Accès aux données Google Cloud

Nous pouvons obtenir les images radiographiques du thorax du NIH à partir de Cloud Storage, BigQuery ou de l'API Cloud Healthcare (figure 68).

3.4. Stockage en ligne (Cloud Storage)

Les données radiographiques du thorax du NIH sont disponibles dans le bucket Cloud Storage suivant: `gs://gcs-public-data--healthcare-nih-chest-xray`

3.5. Accès à l'ensemble de données radiographiques du thorax du NIH dans le Cloud Storage

Le bucket inclut des chemins d'accès aux fichiers PNG d'origine, ainsi qu'aux instances DICOM.

La commande d'accès aux fichiers PNG (fourni par le NIH) est: `gs://gcs-public-data--healthcare-nih-chest-xray/png/FILENAME.png`

La commande d'accès aux fichiers DICOM (fourni par Google) est: `gs://gcs-public-data--healthcare-nih-chest-xray/dicom/FILENAME.dcm`

3.6. Libellés supplémentaires

Des libellés supplémentaires pour les données radiographiques du thorax du NIH sont disponibles dans le bucket Cloud Storage suivant: `gs://gcs-public-data--healthcare-nih-chest-xray-labels`

3.7. Processus d'attribution des libellés

Les libellés définitifs de chaque image ont été attribués par trois radiologues, à l'issue d'un examen décisionnel. Chaque image a été examinée de manière indépendante par trois radiologues. Pour chaque image de l'ensemble de test, les radiologues ont été sélectionnés au hasard parmi une cohorte de 11 radiologues agréés par l'American Board of Radiologie. Pour l'ensemble de validation, les trois radiologues ont été sélectionnés parmi une cohorte de 13 personnes, comprenant des radiologues certifiés et des internes en radiologie.

Les libellés n'ont été déclarés définitifs à l'issue du premier examen que sous réserve de l'accord de tous les lecteurs. Les images suscitant des désaccords concernant les libellés ont été renvoyées pour examen supplémentaire. Les libellés anonymes et les notes émises au cours des séries d'examen précédentes étaient également disponibles lors de chaque examen itératif. Le processus décisionnel s'est poursuivi jusqu'à l'obtention d'un consensus ou pendant un maximum de cinq séries d'examen. En cas d'absence de consensus concernant un petit nombre d'images, le libellé remportant la majorité des voix a été utilisé.

Les seules informations disponibles pendant l'examen étaient l'âge du patient et l'observation de l'image (AP/PA). Aucune information supplémentaire n'a été fournie. En cas de nodule, de masse ou de pneumothorax, les libellés possibles étaient "présent", "absent" ou "Hedge" (en cas d'incertitude). En ce qui concerne les opacités et les fractures, les seules valeurs de libellé possibles étaient "présent" ou "absent".

3.8. Comment utiliser ces libellés

Dans le fichier CSV intitulé `individual_readers.csv`, chaque ligne correspond au libellé de chacune des quatre conditions fournies par un seul lecteur pour une seule image. Cela signifie que chaque ID d'image et le résultat du processus décisionnel correspondant sont répétés sur plusieurs lignes (une ligne par lecteur). L'ID du lecteur est fourni afin d'établir un lien stable entre les images. Une valeur de cellule "YES" signifie "présent", la valeur "NO" signifie "absent" et la valeur "HEDGE" signifie "incertain".

Dans les fichiers CSV nommés `validation_labels.csv` et `test_labels.csv`, quatre colonnes ont été ajoutées pour les métadonnées fournies dans le cadre de l'ensemble de données radiographiques du thorax du NIH, une pour le libellé attribué pour chacune des quatre conditions : fracture, pneumothorax, opacité des espaces aériens, nodule/masse. Il existe 1 962 ID d'image uniques dans l'ensemble de test et 2 412 ID d'image uniques dans l'ensemble de validation, soit un total de 4 374 images pour lesquelles des libellés ont été attribués. Seules les valeurs "YES" et "NO" apparaissent dans les colonnes de libellés attribués. S'il manque une valeur dans une colonne, cela signifie que l'image n'a pas été incluse dans l'ensemble d'images pour lesquelles des libellés ont été attribués.

Lorsque on utilise ces libellés, nous devons inclure la citation suivante :

[2] Anna Majkowska, Sid Mittal, David F. Steiner, Joshua J. Reicher, Scott Mayer McKinney, Gavin E. Duggan, Krish Eswaran, PoHsuan Cameron Chen, Yun Liu, Sreenivasa Raju Kalidindi, Alexander Ding, Greg S. Corrado, Daniel Tse, Shravya Shetty, Chest Radiograph Interpretation Using Deep Learning Models: Assessment Using Radiologist Adjudicated Reference Standards and Population-Adjusted Evaluation, *Radiology*, 2019.

3.9. Les données

Le jeu de données NIH Chest X-ray est composé de 112 120 images radiographiques avec des étiquettes de maladie provenant de 30 805 patients uniques; chacune des images a une étiquette décrivant l'une des 14 lésions (atélectasie, consolidation, infiltrat, pneumothorax, œdème, emphysème, fibrose, épanchement, pneumonie, épaissement pleural, cardiomégalie, nodule, masse et hernie) ou l'absence de lésion. Plusieurs lésions peuvent être présentes sur une seule radiographie. Parmi les images avec des cadres de délimitation répertoriés, seules huit des maladies (atélectasie, cardiomégalie, épanchement, infiltrat, masse, nodule, pneumonie et pneumothorax) sont représentées. Un gros problème est que l'ensemble de données est assez déséquilibré, la plupart des images ne contenant aucune maladie et bien moins d'un millier d'exemples pour les classes les moins fréquentes. Étant donné que les images qui ont des boîtes englobantes dans l'ensemble de données ne représentent que huit classes, ce projet ne traite que de ces huit classes (atélectasie, cardiomégalie, épanchement, infiltrat, masse, nodule, pneumonie et pneumothorax). Ces huit catégories sont en fait les huit originales représentées dans l'ensemble de données, avant que des exemples comportant les six dernières catégories ne soient ajoutés ultérieurement à l'ensemble de données. Cela signifie également qu'une perte pondérée par classe est probablement optimale pour tenter d'atténuer les effets d'un déséquilibre de classe. La principale mise en garde de l'ensemble de données est que les étiquettes de classe ne sont pas enregistrées par un professionnel de la santé; ils sont plutôt glanés en utilisant des techniques de PNL à partir de leurs rapports de radiologie correspondants, qui ne sont pas publiés pour des raisons de confidentialité. Ainsi, les créateurs de jeux de données estiment que leurs étiquettes sont précises à

environ 95%. En outre, il existe des incohérences structurées dans l'ensemble de données qui produisent un bruit structuré, ce qui signifie que l'obtention de bonnes performances sur le viseur de test peut ne pas être cliniquement utile du tout. Cependant, Oakden-Rayner fait l'éloge du modèle CheX-Net5, affirmant qu'il prévoit probablement au niveau des experts, malgré les failles des données. Ainsi, des résultats cliniquement utiles sont possibles malgré le mauvais étiquetage fréquent des données. Au moins, les données de la boîte englobante semblent légitimes. Les 983 images avec encadrés ont été étiquetées par un expert; bien que ces boîtes puissent être inexactes, elles sont beaucoup plus susceptibles d'être correctes que l'étiquetage dépendant de la PNL. Toutes les images de la boîte englobante font partie de l'ensemble de test. Chaque image est étiquetée avec une maladie et une description (x, y, w, h) , où x et y désignent le coin supérieur gauche de la boîte, et w et h désignent la largeur et la hauteur de la boîte.

3.10. Statistiques descriptives de la Base de données

Nous allons présenter dans cette section les différentes statistiques descriptives que nous avons effectuées sur l'ensemble des images de la base NIH pour toutes les classes de maladies étudiées.

3.11. Pathologies selon le sexe

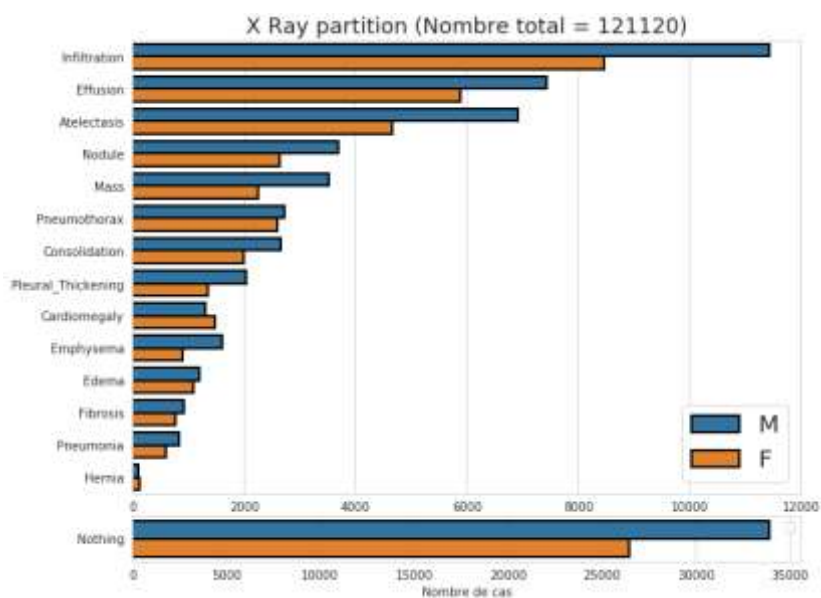


Figure 69: Répartition des pathologies selon le sexe

3.12. Répartition des pathologies simples et Multiples

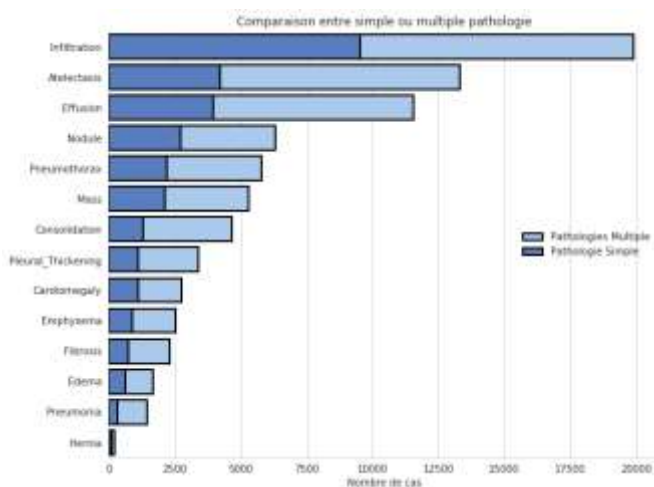


Figure 70: Répartition des pathologies simples et Multiples

3.13. l'atélectasie

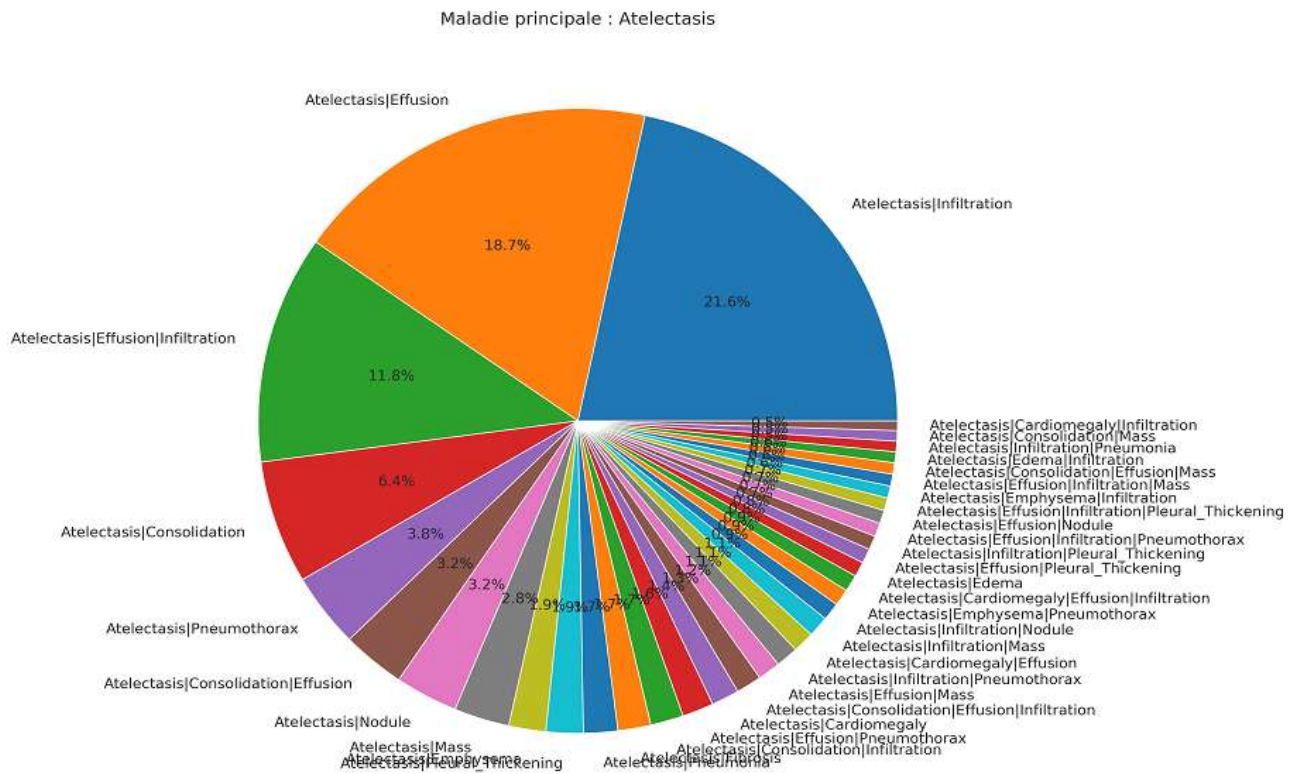


Figure 71: Lésions concomitantes les plus fréquentes avec l'atélectasie

3.14. Cardiomegalie

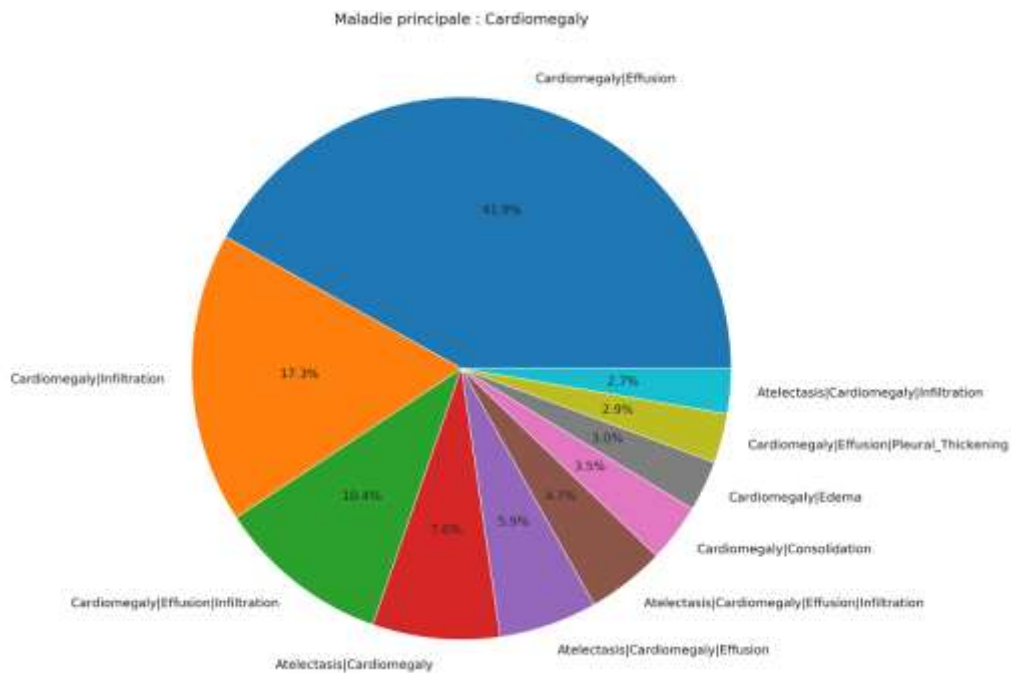


Figure 72: Lésions concomitantes les plus fréquentes avec la Cardiomegalie

3.15. Consolidation

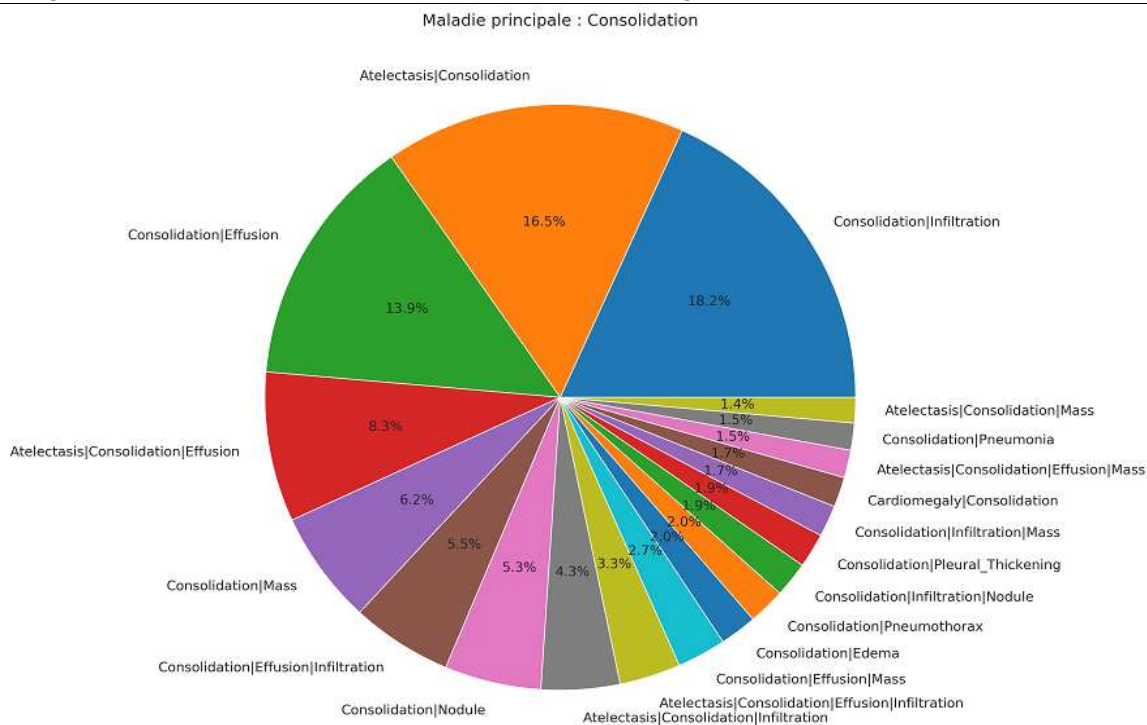


Figure 73: Lésions concomitantes les plus fréquentes avec « Consolidation »

3.16. Oedème

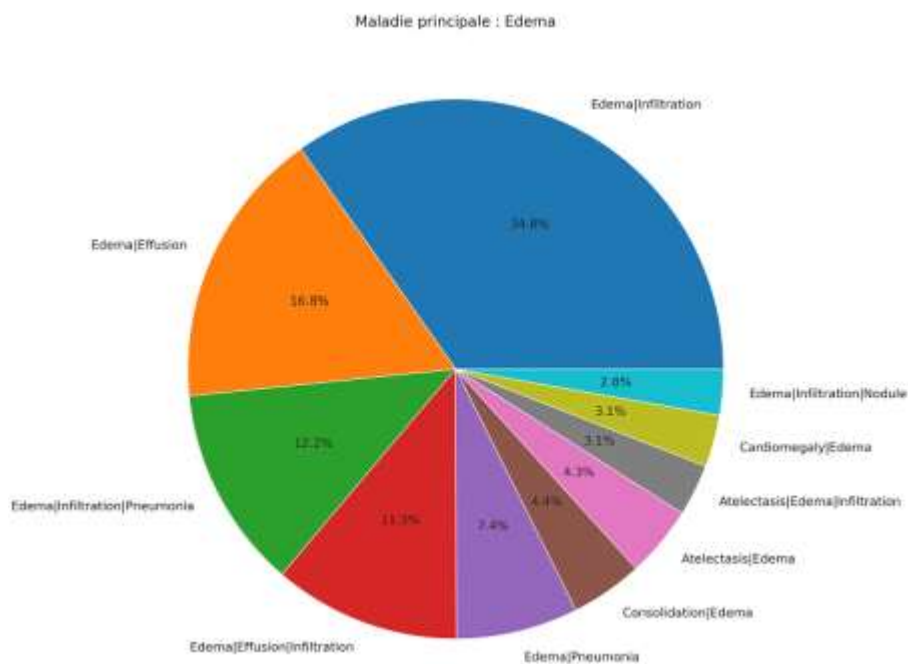


Figure 74: Lésions concomitantes les plus fréquentes avec l'œdème

3.17. Effusion

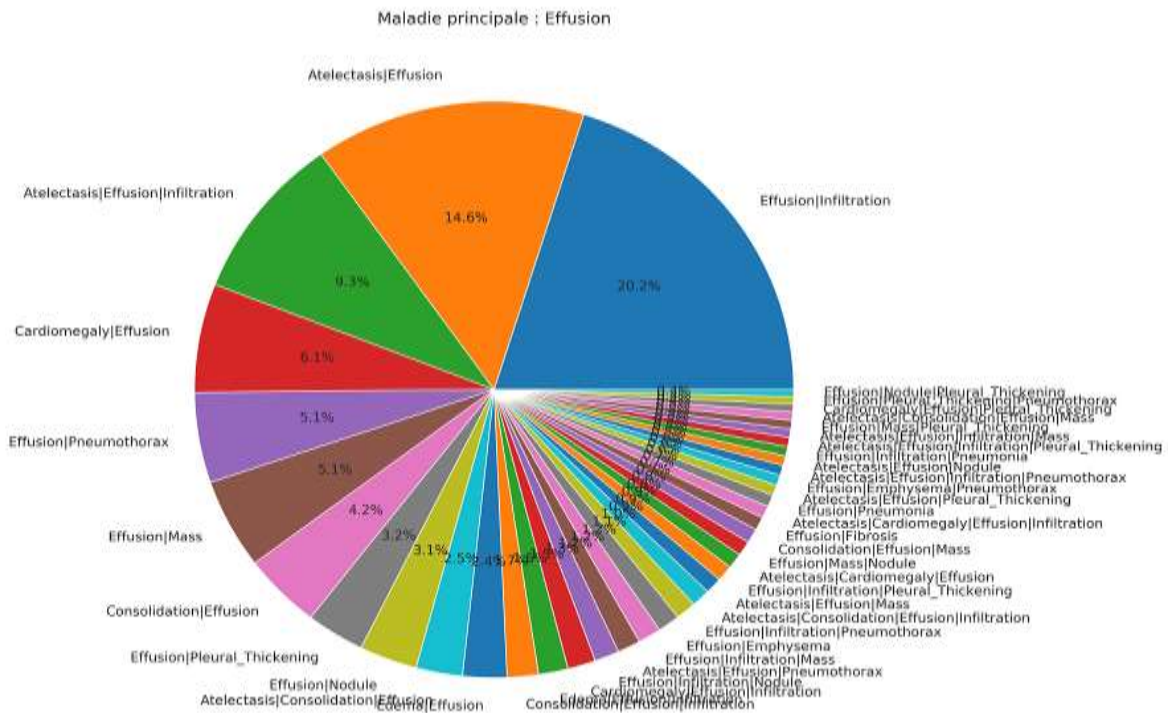


Figure 75 : Lésions concomitantes les plus fréquentes avec la pleurésie 'Effusion'

3.18. Emphysème

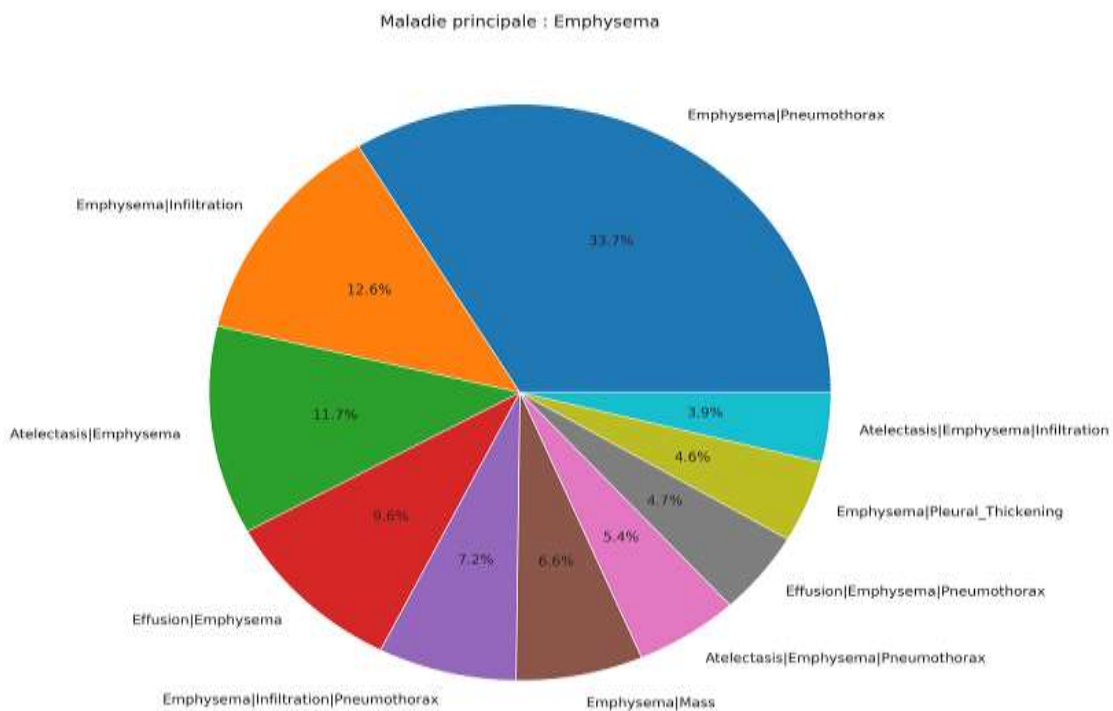


Figure 76 : Lésions concomitantes les plus fréquentes avec l'emphysème

3.19. Fibrose

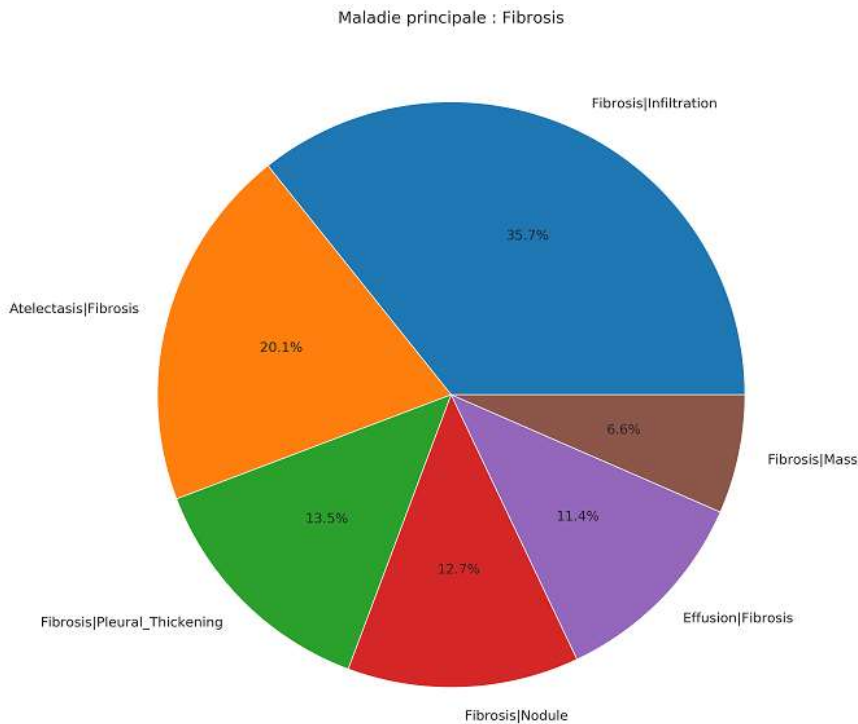


Figure 77 : Lésions concomitantes les plus fréquentes avec la Fibrose

3.20. Infiltration pulmonaire

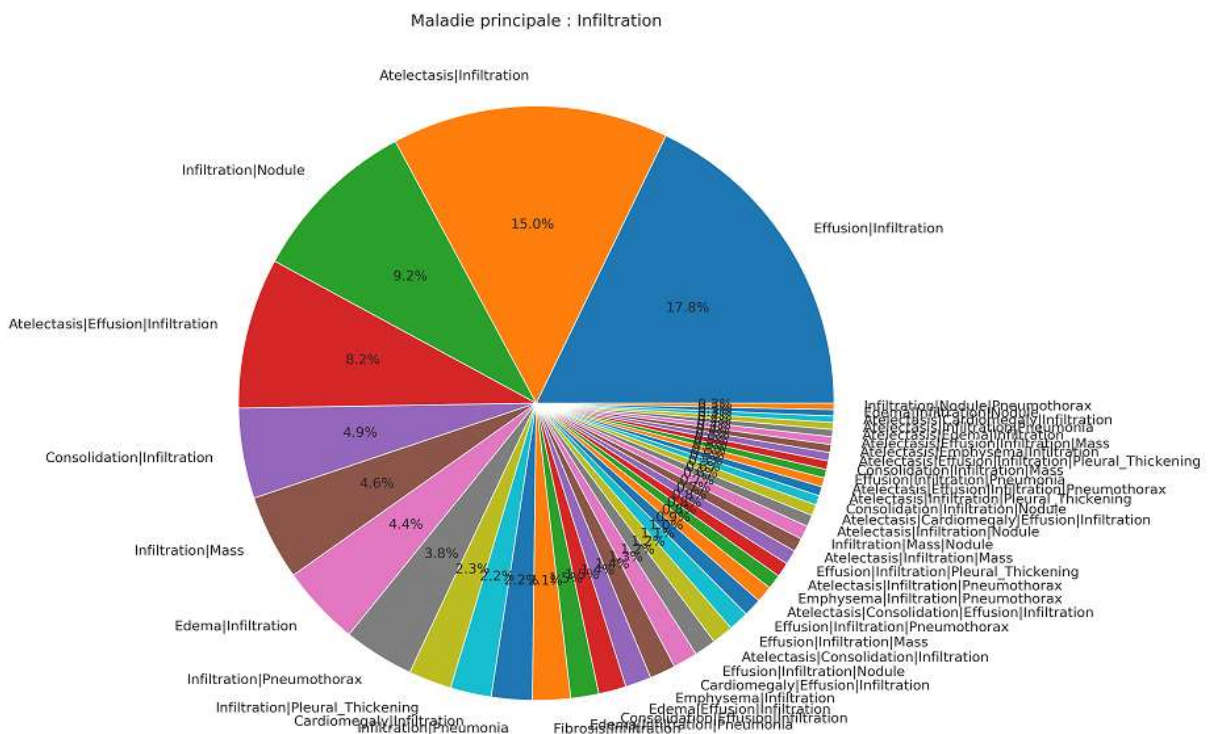


Figure 78 : Lésions concomitantes les plus fréquentes avec l'Infiltration pulmonaire

3.21. Masse

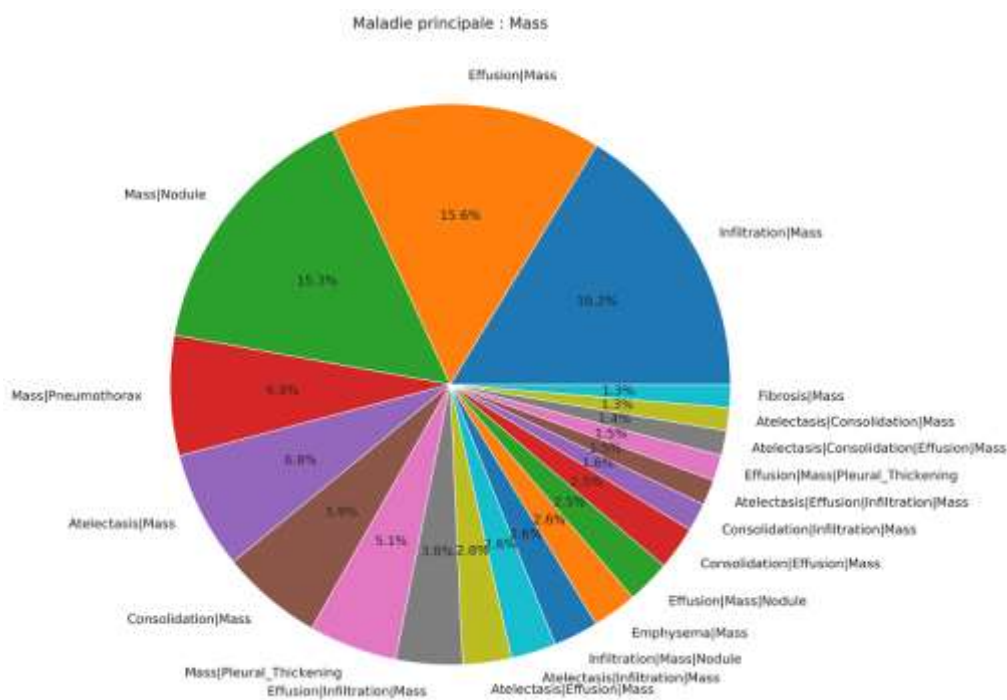


Figure 79 : Lésions concomitantes les plus fréquentes avec les tumeurs (nommées Masse)

3.22. Nodules

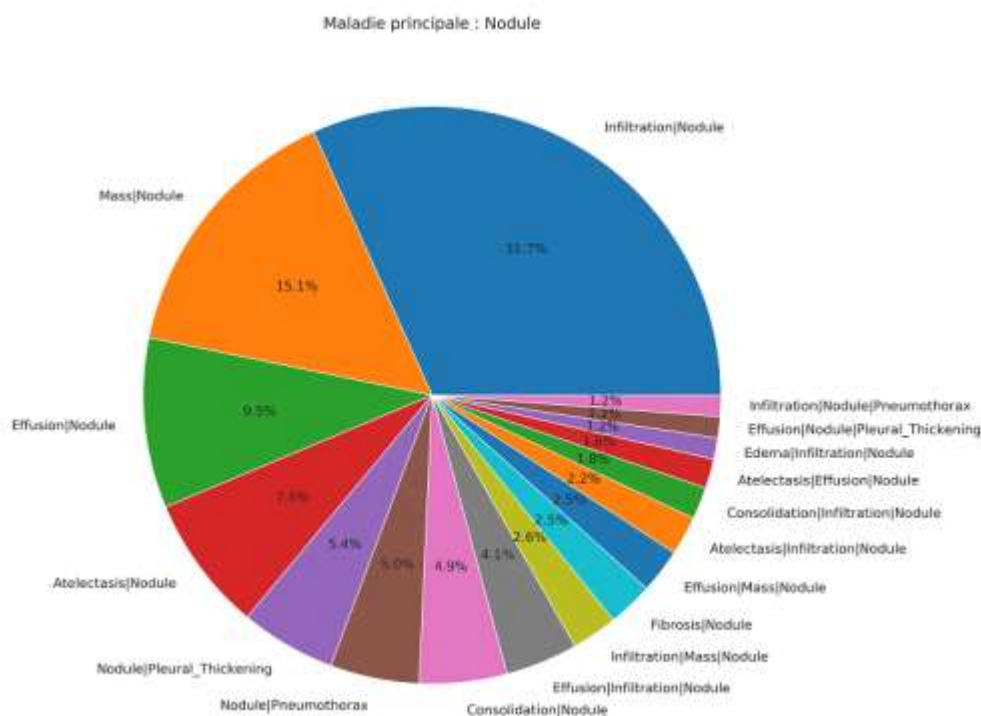


Figure 70 : Lésions concomitantes les plus fréquentes avec les Nodules

3.23.Épaississement

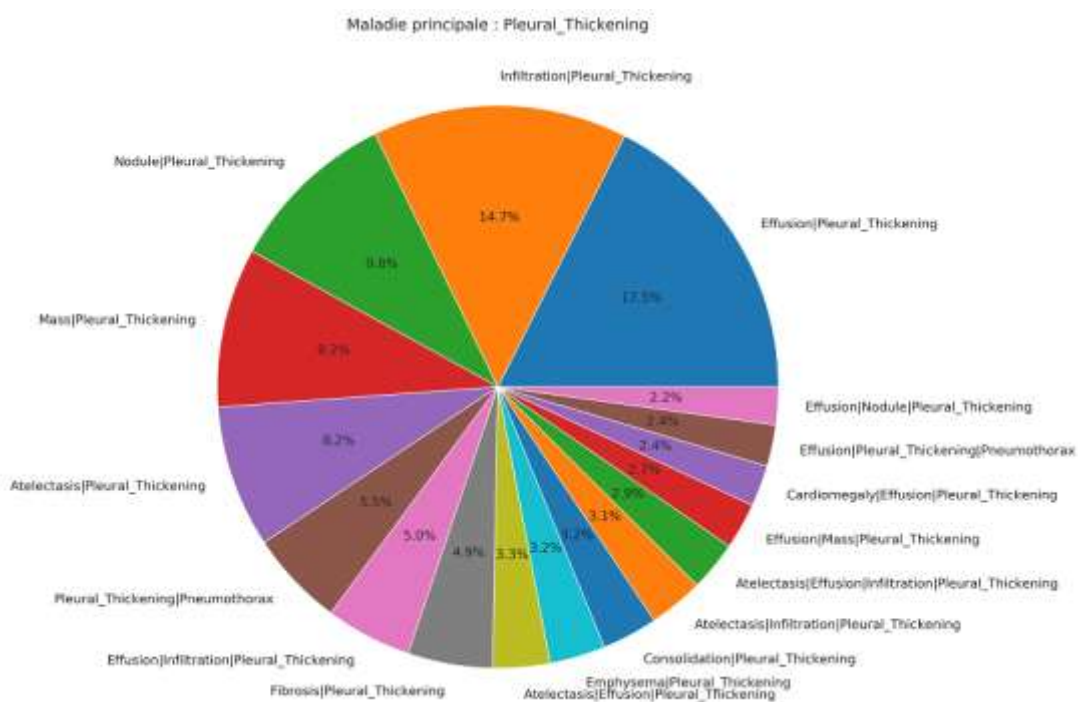


Figure 71 : Lésions concomitantes les plus fréquentes avec l'épaississement

3.24.Pneumonie

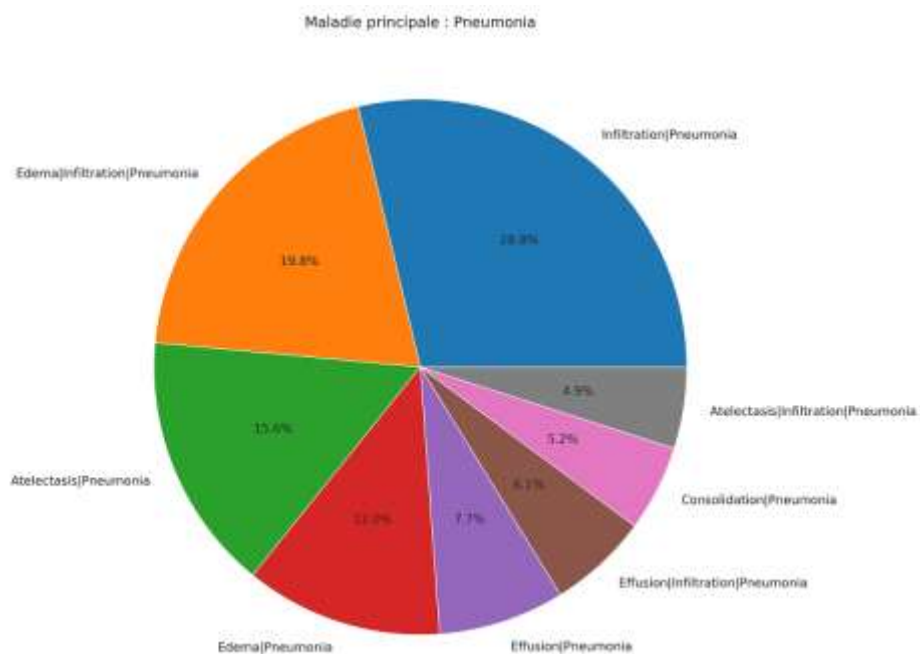


Figure 72 : Lésions concomitantes les plus fréquentes avec la Pneumonie

3.25. Pneumothorax

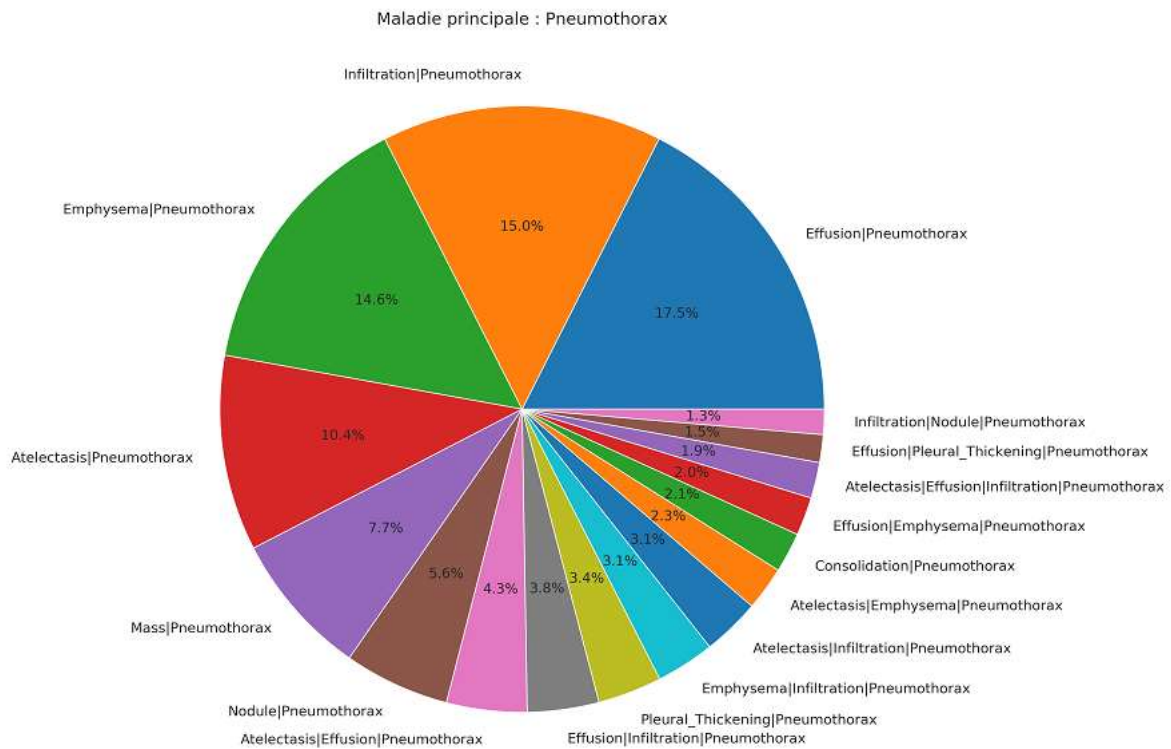


Figure 73 : Lésions concomitantes les plus fréquentes avec le Pneumothorax

CHAPITRE IV

RESULTATS ET DISCUSSION

Dans cette section, nous allons présenter le matériel informatique utilisé pour les simulations. Nous présenterons aussi les tailles et les pourcentages des trois ensembles d'images utilisés à savoir l'ensemble d'apprentissage, l'ensemble de validation et l'ensemble de test. Nous présenterons aussi les critères mesurés de validation de notre approche et nous finirons par présentation des résultats obtenus.

1. MATÉRIEL UTILISÉ

Nous avons implémenté notre projet en utilisant Python 3.7 qui est un langage de programmation interprété, de haut niveau et polyvalent. Créée par Guido van Rossum et lancée pour la première fois en 1991, la philosophie de conception de Python met l'accent sur la lisibilité du code grâce à son utilisation notable d'espaces blancs importants. Le langage Python est largement utilisé dans le domaine de la science des données. Nos modèles ont été développés à l'aide de TensorFlow-GPU qui est une bibliothèque logicielle gratuite et open-source pour le flux de données utilisé pour les applications d'apprentissage automatique telles que les réseaux de neurones, nous l'utilisons donc avec une bibliothèque de wrapper Keras en Python, les expériences ont été réalisées sur Lenovo ThinkPad P51 avec Intel® Core™ i7-7820HQ, 2,90 GHz, NVIDIA Quadro M2200 8 Go GPU et 8 Go de RAM, avec l'utilisation du GPU, nous avons pu réduire le temps de calcul. En plus de cela nous utilisons le noyau de Kaggle qui se caractérise par la présence d'un GPU NVIDIA TESLA P100 (figure 74) et de 16G de RAM.



Figure 74: Carte graphique NVIDIA Quadro M2200 utilisée pour l'apprentissage

2. ENSEMBLE D'IMAGES POUR L'APPRENTISSAGE ET LA VALIDATION DU MODÈLE PROPOSÉ

Afin de pouvoir suivre la prédiction de notre système en fonction du nombre de classe présentes dans la base du NIH nous avons construit trois ensembles d'images. Un ensemble à 3 classes, un ensemble à 4 classes et un ensemble à 14 classes. Nous avons ainsi mené une première expérience de reconnaissance automatique d'image sur le premier ensemble constitué de 3 classes : Nodule, Masse, non défini. Ensuite nous avons mené une deuxième expérience de reconnaissance sur le deuxième ensemble formé de quatre classes : Nodule, Masse, Effusion, non défini. Enfin une troisième expérience a été menée sur l'ensemble contenant les 14 classes. Pour chaque expérience nous avons mesuré plusieurs critères de performance pour caractériser le pouvoir de reconnaissance de notre modèle proposé.

Pour construire le premier ensemble nous avons collecté d'une façon aléatoire de la grande base NIH près de 16 578 images annotées de chaque type de classe : 5539 images de type Masse, 5505 images de type Nodule et 5534 images de type Non Défini. Afin de pouvoir booster la phase de l'apprentissage nous avons procédé à une augmentation du nombre d'image de cet ensemble en utilisant le procédé

d'augmentation présenté dans le paragraphe suivant. Nous avons ensuite appliqué le procédé de pré traitement présenté lui aussi dans le paragraphe suivant.

Le deuxième ensemble contient 4 classes d'images : 5539 images de type Masse, 5505 images de type Nodule, 5505 image de type Effusion et 5534 images de type Non Défini. Nous avons ainsi obtenu 22083 images. Pour augmenter encore la taille de cette base nous avons procéder à la technique d'augmentation de base. Nous avons ensuite appliqué le procédé de pré traitement présenté lui aussi dans le paragraphe suivant.

Le troisième ensemble contient 14 classes d'images. Pour chaque type de classe nous avons collecté d'une façon aléatoire un certain nombre d'images annotées de la base de données NIH. Pour augmenter encore la taille de cette base nous avons procéder à la technique d'augmentation de base. Nous avons ensuite appliqué le procédé de pré traitement présenté lui aussi dans le paragraphe suivant.

Pour la collecte et l'annotation des images nous avons utilisé un fichier de la base qui contient la méthode de fractionnement que nous prendrons en considération dans l'étape d'apprentissage automatique. La méthode de fractionnement dépend de la colonne patient fondée dans le fichier csv «Data_Entry_2017.csv».

3. ENSEMBLE D'IMAGE POUR LE TEST

Afin de tester la prédiction de notre modèle nous avons sélectionné aléatoirement des images de la base de données. Ces images ont été pré traitées avant de procéder au test de prédiction.

4. PRÉTRAITEMENT ET AUGMENTATION DES DONNÉES

Le prétraitement des ensembles de données est une étape cruciale avant la formation des modèles, nous prenons en compte trois facteurs qui améliorent la précision de la classification.

Premièrement, il y a des images de mauvaise qualité, des images de rotation, des points de vue différents et des couleurs inversées dans certaines images, et des images radiographiques pulmonaires d'enfants (figure 75). La suppression de ces images a amélioré le résultat de la classification.

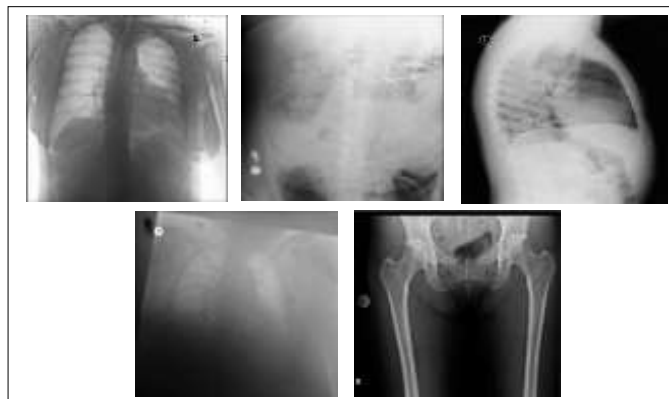


Figure 75: Exemple d'images radiographiques à pré traiter .

Deuxièmement, nous avons recadré et enregistré les images à partir du centre à 75% pour mieux se concentrer sur la zone pulmonaire en utilisant un script pour résoudre ce problème. Nous avons aussi supprimé la colonne vertébrale supérieure, les fils médicaux et un peu d'écriture sur les images originales. Le résultat de l'exécution de notre script est montré sur la figure 76:

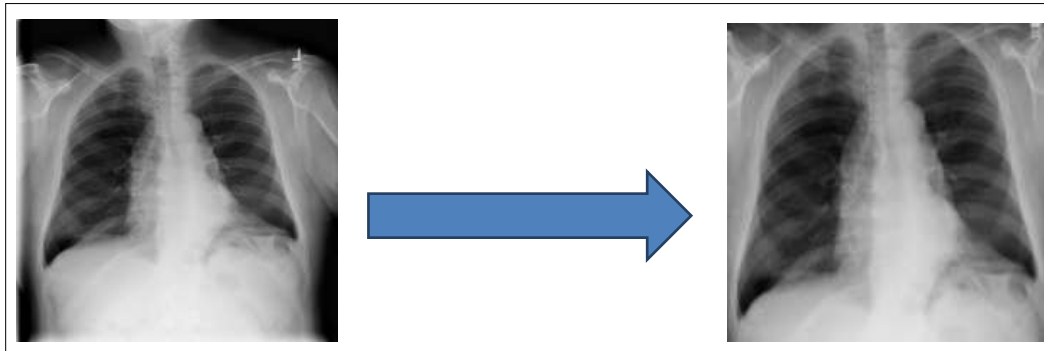


Figure 76: Résultat de l'exécution de notre scripte de prétraitement sur une image

Troisièmement, nous allons changer la taille de chaque image au format 224 * 224 pour une meilleure précision. Les grandes dimensions d'image devraient donner plus de détails malgré un temps de formation plus long.

En plus de ces trois opérations de prétraitements, nous allons utiliser une technique d'augmentation des données pour mieux généraliser les modèles et éviter le sur-ajustement. Parmi ces techniques il y a le retournement horizontal pour faire face aux symptômes de la pneumonie de chaque côté de la poitrine, la variation de l'intensité de la luminosité et les rotations des images.

5. MESURE DES PERFORMANCES DE NOTRE MODÈLE

Pour pouvoir suivre et assurer la convergence de la phase d'apprentissage de notre modèle nous avons mesuré pour chaque itération la fonction de perte (Loss function).

Afin de pouvoir évaluer le modèle proposé vous allons mesurer les critères suivants :

1. Pour l'algorithme d'apprentissage en profondeur mon va mesurer ASC (Aire Sous la Courbe) tout en calculant l'Intervalle de Confiance (IC).
2. En utilisant la probabilité de seuil d'indice de Youden maximisée, nous allons aussi mesurer la sensibilité, la spécificité, la précision, les valeurs de

prédiction négatives, les indices Youden et les scores F1 pour les ensembles de données.

3. Pour mesurer la cohérence du Modèle CNN proposé avec les résultats des tests pathogènes nous allons mesurer la valeur du Critère Kappa.

5.1. Fonction de perte

Pour bien représenter l'ensemble des classes, nous choisissons un vecteur d'étiquette à C dimensions : $y = [y_1, y_2, \dots, y_C]$ où $y_c = 1$ indique la représentation de la pathologie c et $y_c = 0$ sinon. Cela signifie que tous les vecteurs nuls $y = [0, 0, \dots, 0]$ représentent un statut «normal» (aucune pathologie de la portée de 14 maladies n'est trouvée dans cet échantillon). Ce paramètre transforme le problème de classification multi-étiquettes en C problèmes de classification binaire indépendants.

En utilisant la définition ci-dessus, la fonction de perte est égale à la perte d'entropie croisée binaire totale pour chaque pathologie possible. Et comme les étiquettes d'image sont plutôt modestes (c'est-à-dire plus de «0» que de «1»), le facteur d'équilibrage positif / négatif β_0, β_1 est introduit pour imposer l'apprentissage des étiquettes positives.

$$L(p, y) = \beta_p \sum_{c=1, y_c=1}^{c=14} -\ln(p) + \beta_N \sum_{c=1, y_c=0}^{c=14} -\ln(1-p)$$

Où:

p : la probabilité de $y=1$, de la fonction de sortie sigmoïde.

$\beta_p = \frac{|P|+|N|}{|P|}$ est le facteur d'équilibrage positif, $\beta_N = \frac{|P|+|N|}{|N|}$ est le facteur

d'équilibrage négatif où $|P|$ et $|N|$ est le nombre total de 1 et de 0 dans un lot d'étiquettes d'image.

5.2. Courbe ROC

Une courbe ROC est construite en traçant le taux de vrais positifs (TPR) par rapport au taux de faux positifs (FPR). Le vrai taux positif est la proportion d'observations qui ont été correctement prédites comme positives parmi toutes les observations positives ($TP / (TP + FN)$). De même, le taux de faux positifs est la proportion d'observations dont on prédit à tort qu'elles sont positives parmi toutes les observations négatives ($FP / (TN + FP)$).

6. 1^{ERE} EXPERIENCE : RESULTATS POUR L'ENSEMBLE A 3 CLASSES

6.1. Taille des échantillons

Le premier ensemble de données contient 3 classes : Masse, Nodule et Non Défini.

Chaque classe est formée de :

Masse : 5539 images

Nodule: 5505 images

Non défini : 5534 images

6.2. Exemple d'images de chaque classe













| Images de type Masse (a) | Images de type Nodule (b) | Images de type Non Défini (c) |
|---|---|---|
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Figure 77 : Exemple de radiographie thoracique due la ere expérience : (a) Masse, (b) Nodule et (c) Non Défini

6.3. Pourcentage des images considérées dans les phases d'apprentissage, de validation et de test

Nous avons utilisé un total de 16578 images pour l'apprentissage, la validation, et le test. La répartition de ces images est donnée dans le tableau 2.

Tableau 2: Répartition des données images de l'expérience 1.

| | Mass | Nodule | Non défini | Total |
|---------------|------|--------|------------|-------|
| apprentissage | 4024 | 3970 | 3983 | 11977 |
| validation | 699 | 705 | 710 | 2114 |
| test | 816 | 830 | 841 | 2487 |

6.4. Configuration de notre modèle proposé

Le processus d'exécution de notre algorithme profond contient les étapes suivantes :

- a. Prétraitement des images CXR,
- b. Augmentation des données, Apprentissage par transfert en utilisant 5 réseaux CNN déjà pré-entraînés sur la bse ImageNET qui sont : MobileNet, DenseNet121, InceptionV3, ResNet50 et Xception,
- c. Extraction des caractéristiques,
- d. Classification

Nous avons remplacé la dernière couche entièrement connectée de chaque modèle par une couche de dimension (3×1).

La configuration de l'architecture de chaque réseau CNN de notre algorithme est donnée par le tableau 3. Ce dernier Le tableau 3 présente les cinq modèles dont le taux d'apprentissage initial est égal à 0,001 et le nombre d'époques à 30. De plus, la taille minimale du lot est fixée à 32. Nous avons programmé notre algorithme de telle sorte qu'il réduise le taux d'apprentissage après 5 époques si la perte a cessé de diminuer.

En ce qui concerne la technique d'optimisation, nous avons choisi le classifieur Adam.

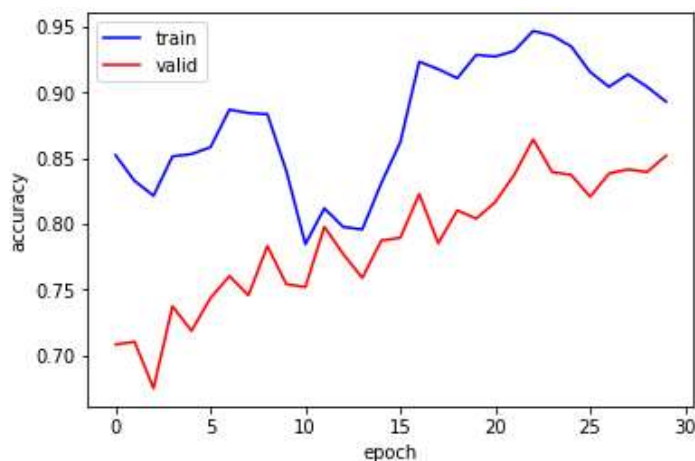
Tableau 3: Configuration de l'architecture de chaque réseau CNN utilisé dans le modèle proposé

| Type de Réseau | Nb. de couches | Taille des Lots | Nb. Époques | Taux d'apprentis | Optimiseur |
|----------------|----------------|-----------------|-------------|------------------|------------|
| MobilNet | 28 | 32 | 30 | 0.001 | Adam |
| DenseNet | 121 | 32 | 30 | 0.001 | Adam |
| InceptionV3 | 48 | 32 | 30 | 0.001 | Adam |
| ResNet50 | 50 | 32 | 30 | 0.001 | Adam |
| Xception | 71 | 32 | 30 | 0.001 | Adam |

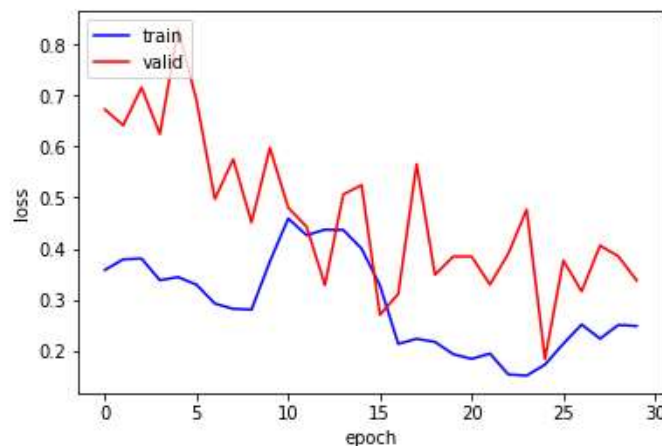
6.5. Courbes de perte d'apprentissage et de précision pour les phases d'apprentissage et de validation

Afin de pouvoir suivre la convergence de la phase d'apprentissage de notre modèle nous avons mesuré pour chaque époque la fonction de perte d'apprentissage de chaque modèle CNN.

De même pour suivre la précision de la phase de validation de notre modèle, nous avons mesuré la précision de chaque réseau CNN de notre modèle pour chaque époque. Les résultats de ces mesures sont portés sur les figures suivantes (79–82).

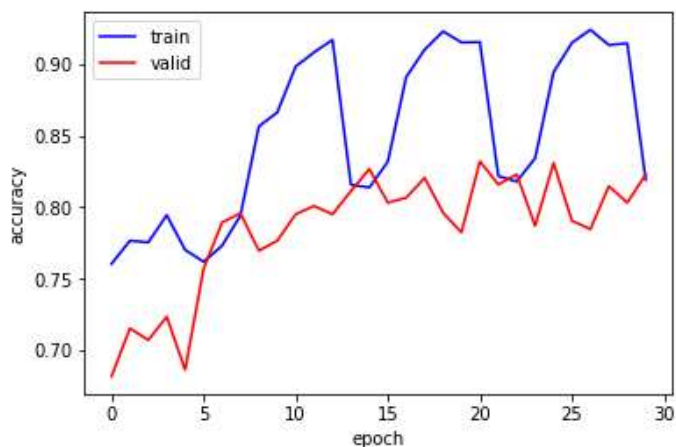


(.a.)

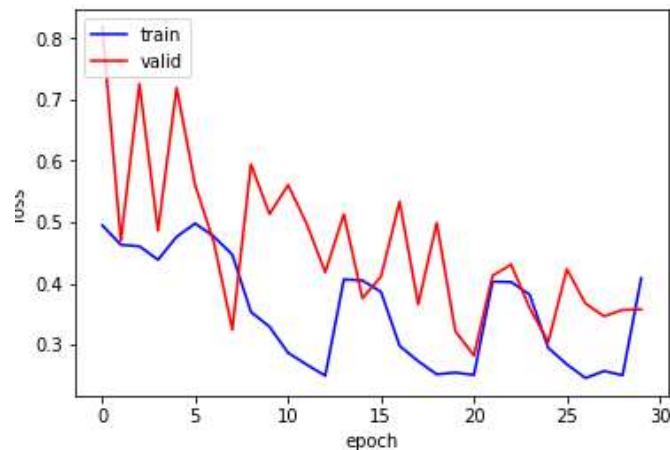


(b)

Figure 78: Courbes de perte d'apprentissage et de précision du réseau CNN Xception:
(a) courbe de précision (b) courbe de perte.

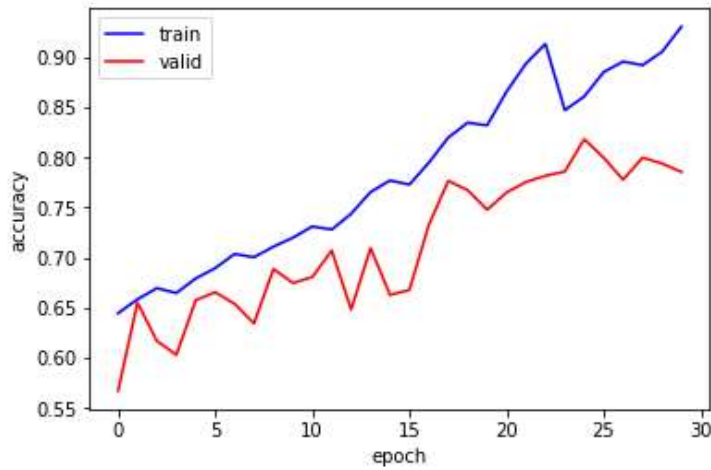


(a)

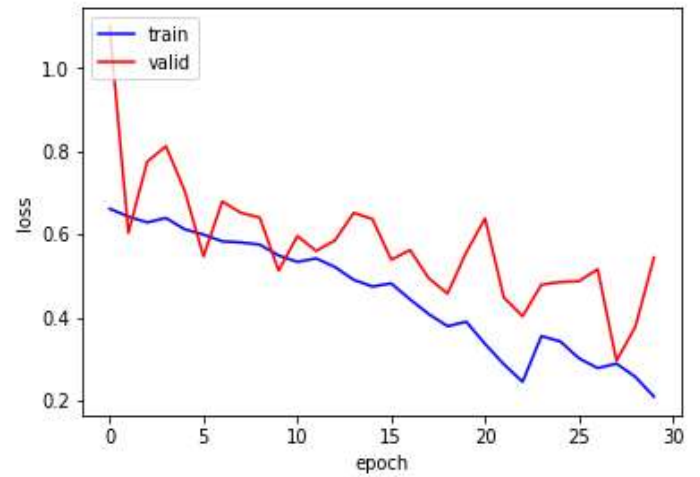


(b)

Figure 79: Courbes de perte d'apprentissage et de précision du réseau CNN
DenseNet121: (a) courbe de perte (b) courbe de précision



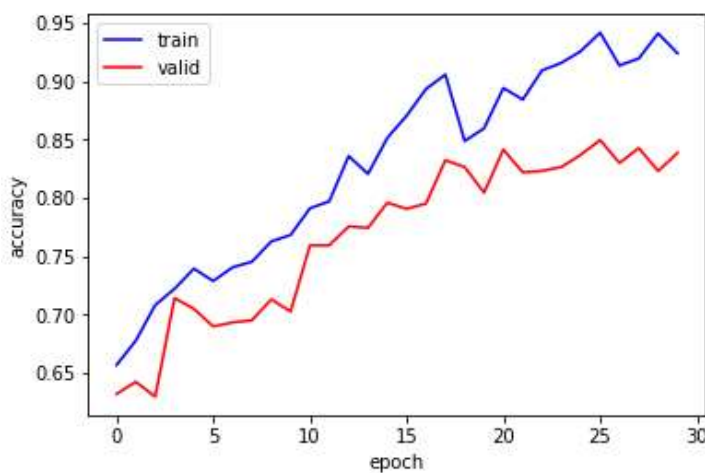
(.a.)



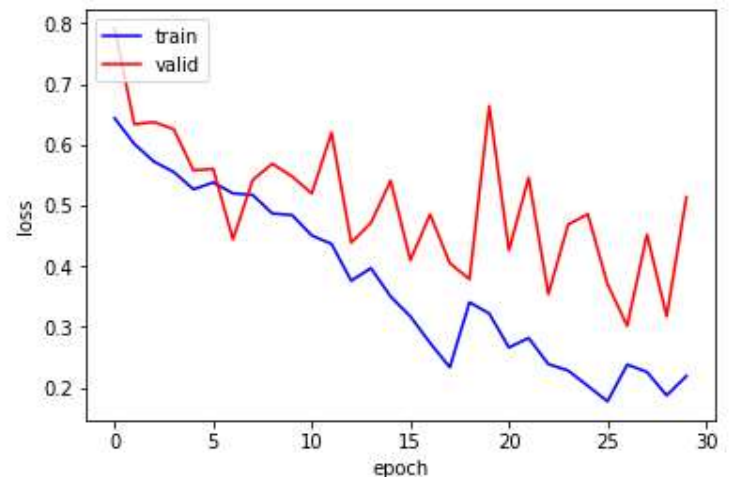
(b.)

Figure 80: Courbes de perte d'apprentissage et de précision du réseau CNN

ResNet50: (a) courbe de perte (b) courbe de précision.



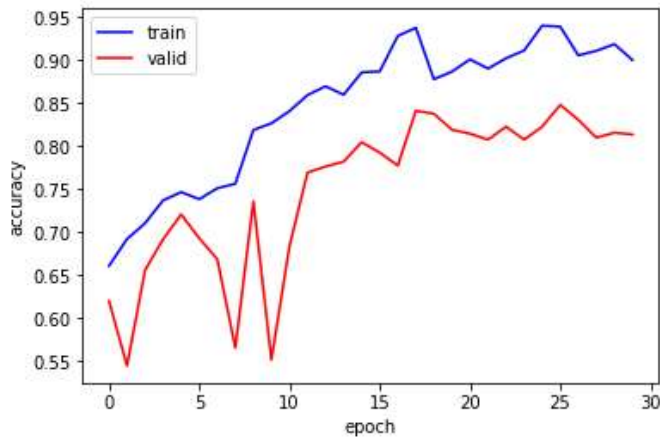
(.a.)



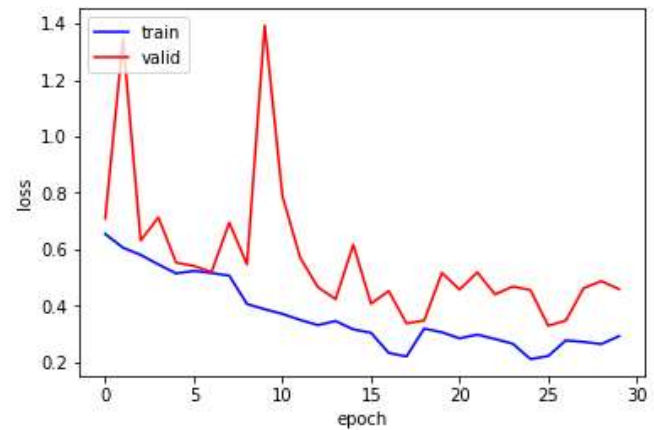
(b.)

Figure 81: Courbes de perte d'apprentissage et de précision du réseau CNN

InceptionV3: (a) courbe de perte (b) courbe de précision.



(a)



(b)

Figure 82: Courbes de perte d'apprentissage et de précision du réseau CNN

MobileNet: (a) courbe de perte (b) courbe de précision.

6.6. Courbe du ROC pour la phase du test

Pour suivre et assurer la convergence de la phase de test de notre modèle nous avons effectué une analyse ROC en utilisant l'aire sous la courbe (AUC) pour les trois classes de pathologies. Les courbes obtenues sur les figures 83–88 montrent bien la convergence de notre modèle pour les trois classes de pathologie qui constituent le jeu de données de notre ensemble considéré.

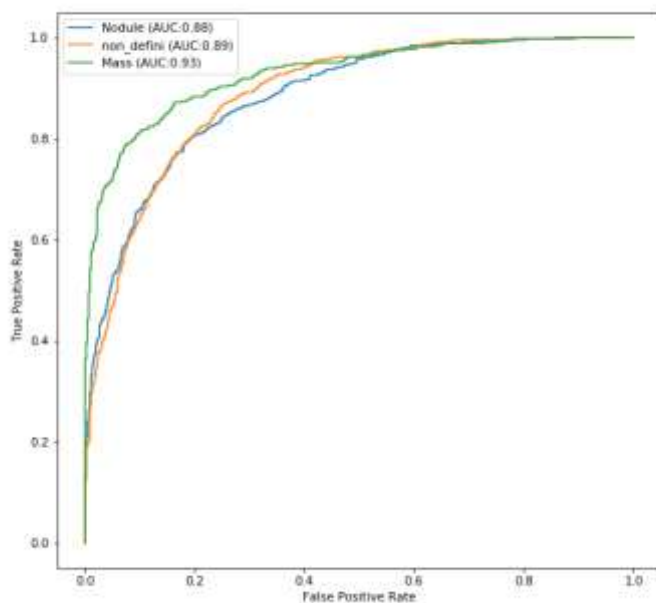


Figure 83 : Courbes ROC obtenues par le réseau Xception pour les classes : Masse, Nodule et Non défini

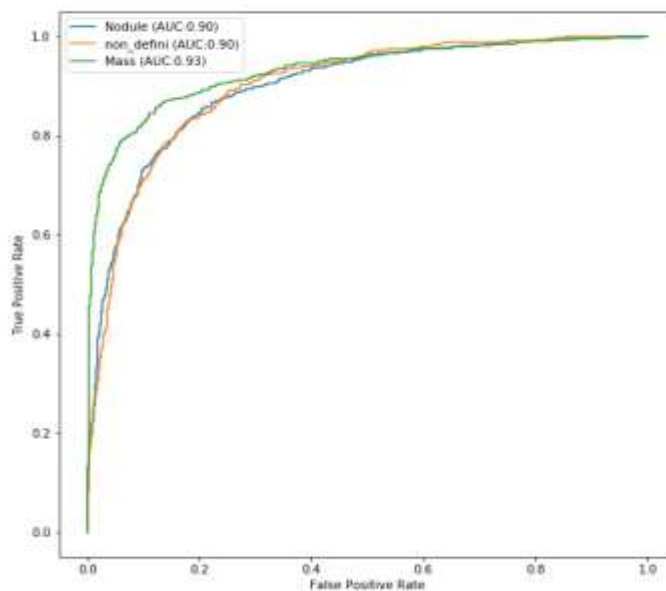


Figure 85: Courbes ROC obtenues par le réseau DENSENet21 pour les classes : Masse, Nodule et Non défini

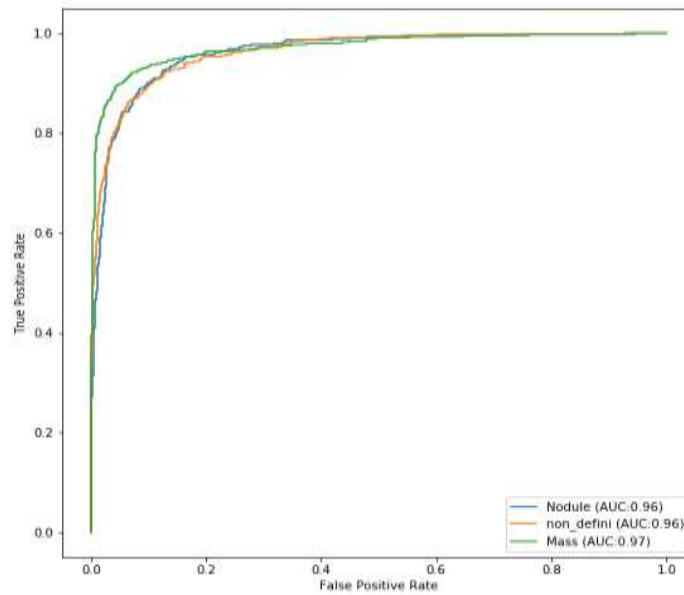


Figure 86: Courbes ROC obtenues par le réseau ResNet50 pour les classes : Masse, Nodule et Non défini

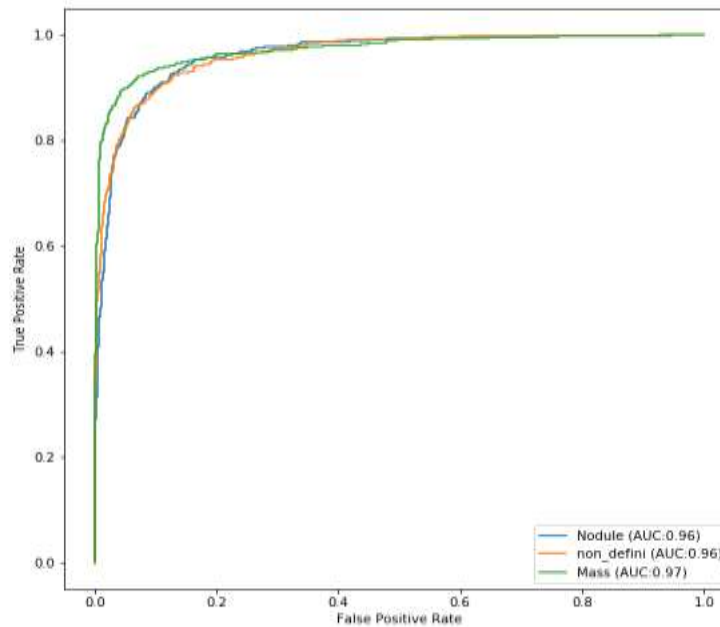


Figure 87: Courbes ROC obtenues par le réseau InceptionV3 pour les classes : Masse, Nodule et Non défini

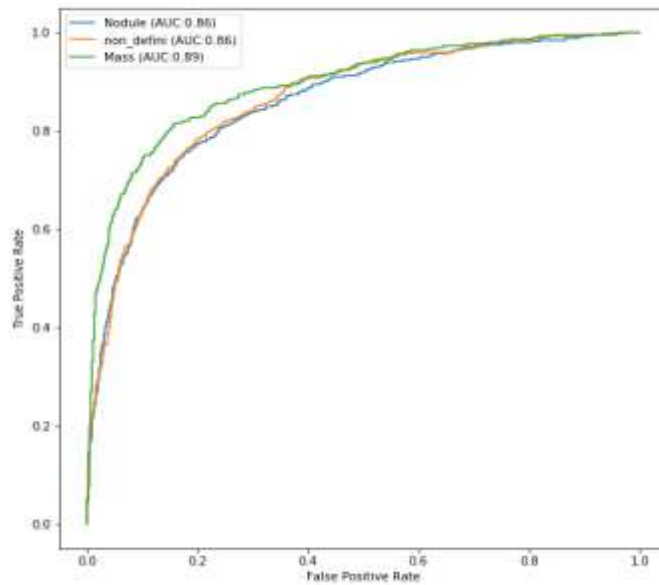


Figure 88: Courbes ROC obtenues par le réseau MobileNet pour les classes : Masse, Nodule et Non défini

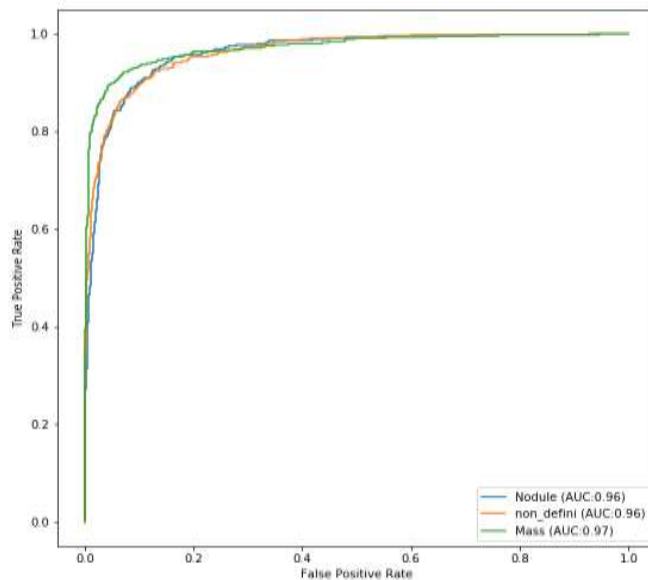


Figure 89: Courbes ROC obtenues par le réseau Ensemble Model pour les classes : Masse, Nodule et Non défini

6.7. Performances mesurées de notre modèle proposé

Nous avons rassemblé dans le tableau 4 les mesures obtenues des performances de notre modèle proposé. Nous remarquons que les résultats obtenus par l'ensemble Modèle proposé (stacking Model) dépasse ceux obtenus pour chaque réseau CNN pris individuellement ce qui montre l'efficacité de l'approche proposé.

Tableau 4: Performance (%) de nos modèles sur l'ensemble de test

| CNN | Précision % | Précision (macro) % | F1_score (macro) % | Recall (macro) % | kappa % | Sensitivity % | Specificity % | AUC % |
|-------------------|--------------|---------------------|--------------------|------------------|--------------|---------------|---------------|--------------|
| Xception | 87,02 | 87,031 | 87,03 | 87,04 | 80,83 | 90,05 | 90,09 | 90,02 |
| DenseNet121 | 78,72 | 79,0 | 78,78 | 78,76 | 68,09 | 87,68 | 80,93 | 86,60 |
| ResNet50 | 70,28 | 70,70 | 70,42 | 70,29 | 56,40 | 73,91 | 79,78 | 86,25 |
| InceptionV3 | 83,70 | 83,89 | 83,83 | 83,82 | 75,69 | 88,30 | 85,96 | 88,27 |
| MobileNet | 82,30 | 82,39 | 82,30 | 82,35 | 73,46 | 90,09 | 84,48 | 87,18 |
| Ensemble-M | 87,41 | 87,65 | 87,47 | 87,44 | 81,12 | 91,86 | 86,86 | 93,91 |

7.2^{ERE} EXPERIENCE : RESULTATS POUR L'ENSEMBLE A 4 CLASSES

7.1. Taille des échantillons

Le premier ensemble de données contient 3 classes : Masse, Nodule et Non Défini. Chaque classe est formée de :

Masse : 5539 images

Nodule: 5505 images

Effusion : 5505 images

Non défini : 5534 images

7.2. Exemple d'images de chaque classe

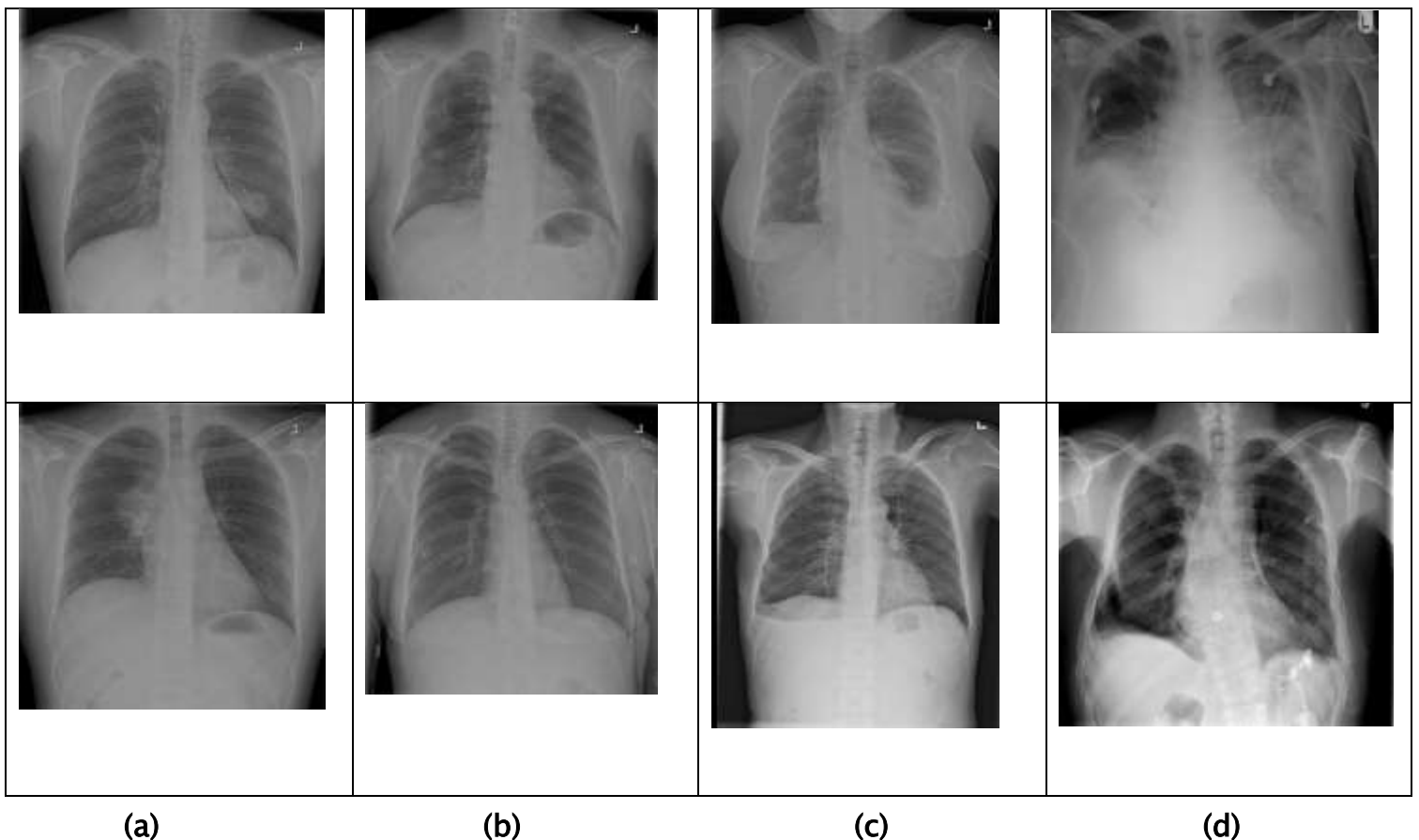


Figure 90: Exemple de radiographie thoracique due la 2^{ème} expérience : (a) Masse, (b) Nodule, (c) Effusion et (d) Non Défini

7.3. Pourcentage des images considérées dans les phases d'apprentissage, de validation et de test

Nous avons utilisé un total de 22133 images pour l'apprentissage, la validation, et le test. La répartition de ces images est donnée dans le tableau 2.

Tableau 5: Répartition des données images de l'expérience 1.

| | Masse | Nodule | Effusion | Non défini | Total |
|---------------|-------|--------|----------|------------|-------|
| apprentissage | 3993 | 3972 | 4003 | 4023 | 15991 |
| validation | 740 | 690 | 701 | 691 | 2822 |
| test | 806 | 843 | 830 | 841 | 3320 |

7.4. Configuration de notre modèle proposé

Le processus d'exécution de notre algorithme profond contient les étapes suivantes :

- Prétraitement des images CXR,
- Augmentation des données, Apprentissage par transfert en utilisant 5 réseaux CNN déjà pré-entraînés sur la base ImageNET qui sont : MobileNet, DenseNet121, InceptionV3, ResNet50 et Xception,
- Extraction des caractéristiques,
- Classification

Nous avons remplacé la dernière couche entièrement connectée de chaque modèle par une couche de dimension (4×1).

La configuration de l'architecture de chaque réseau CNN de notre algorithme est donnée par le tableau 3. Ce dernier Le tableau 3 présente les cinq modèles dont le taux d'apprentissage initial est égal à 0,001 et le nombre d'époques à 30. De plus, la taille minimale du lot est fixée à 32. Nous avons programmé notre algorithme de telle sorte qu'il réduise le taux d'apprentissage après 5 époques si la perte a cessé de diminuer.

En ce qui concerne la technique d'optimisation, nous avons choisi le classifieur Adam.

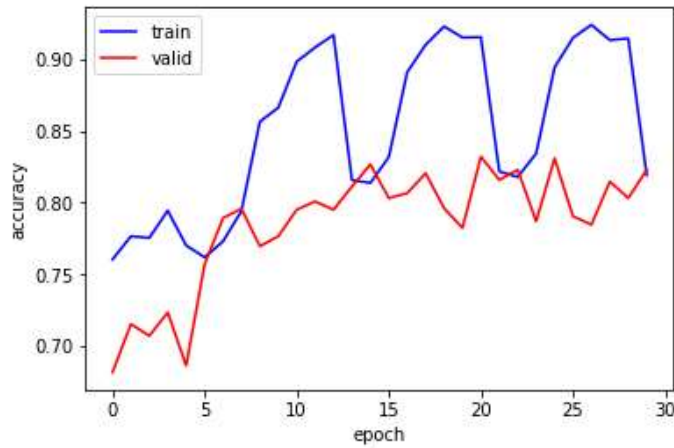
Tableau 6: Configuration des architecture de chaque réseau CNN utilisé dans le modèle proposé

| Type de Réseau | Nb. de couches | Taille des Lots | Nb. Époques | Taux d'apprentis | Optimiseur |
|----------------|----------------|-----------------|-------------|------------------|------------|
| MobilNet | 28 | 32 | 30 | 0.001 | Adam |
| DenseNet | 121 | 32 | 30 | 0.001 | Adam |
| InceptionV3 | 48 | 32 | 30 | 0.001 | Adam |
| ResNet50 | 50 | 32 | 30 | 0.001 | Adam |
| Xception | 71 | 32 | 30 | 0.001 | Adam |

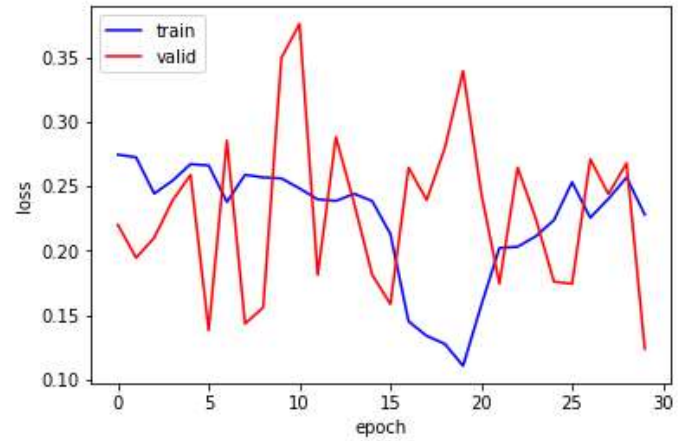
7.5. Courbes de perte d'apprentissage et de précision pour les phases d'apprentissage et de validation

Afin de pouvoir suivre la convergence de la phase d'apprentissage de notre modèle nous avons mesuré pour chaque époque la fonction de perte d'apprentissage de chaque modèle CNN.

De même pour suivre la précision de la phase de validation de notre modèle, nous avons mesuré la précision de chaque réseau CNN de notre modèle pour chaque époque. Les résultats de ces mesures sont portés sur les figures suivantes (91–97).



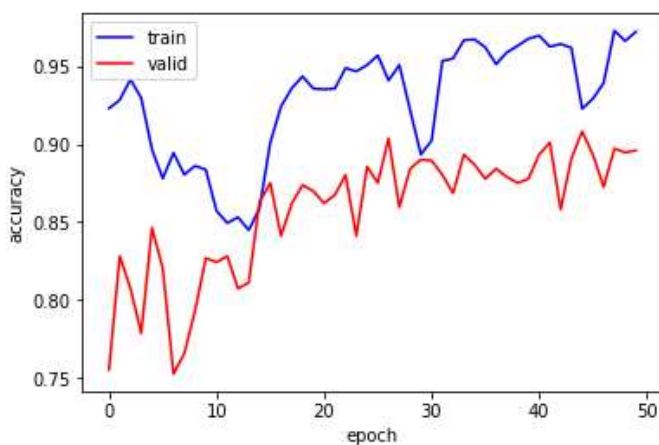
(.a.)



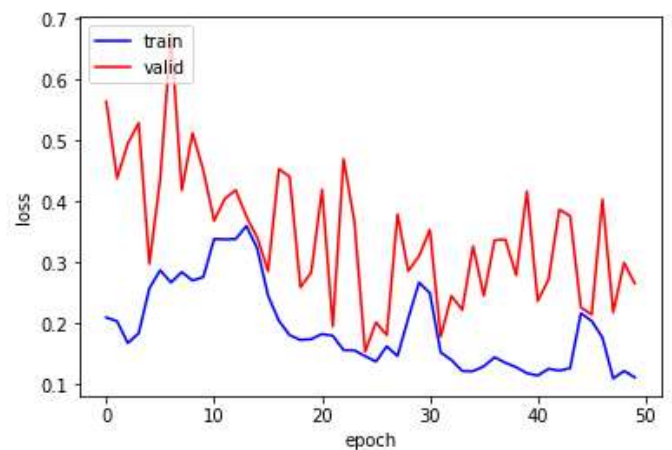
(b)

Figure 91: Courbes de perte d'apprentissage et de précision du réseau CNN Xception:

(a) courbe de précision (b) courbe de perte.



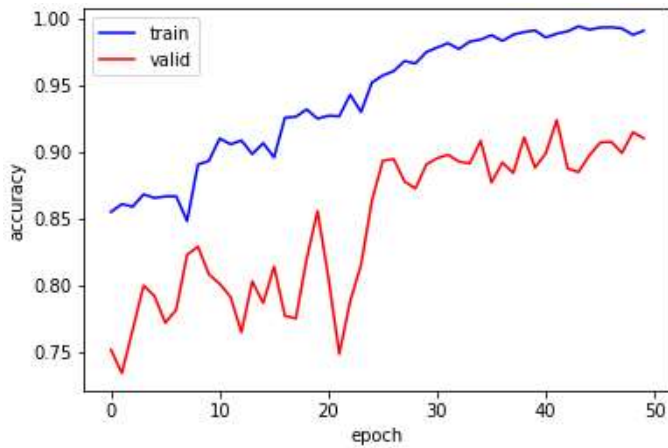
(a)



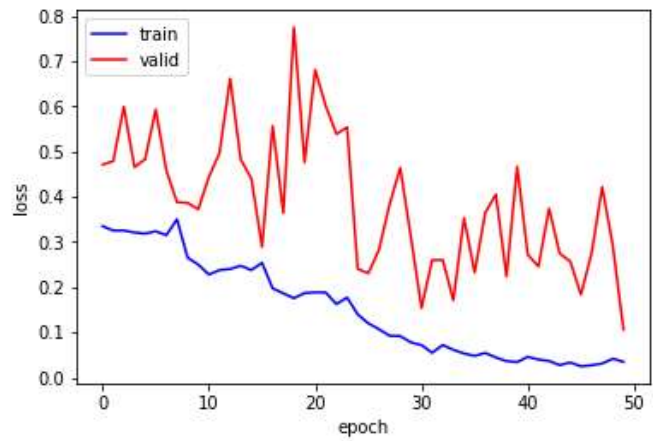
(b)

Figure 92: Courbes de perte d'apprentissage et de précision du réseau CNN

DenseNet121: (a) courbe de perte (b) courbe de précision



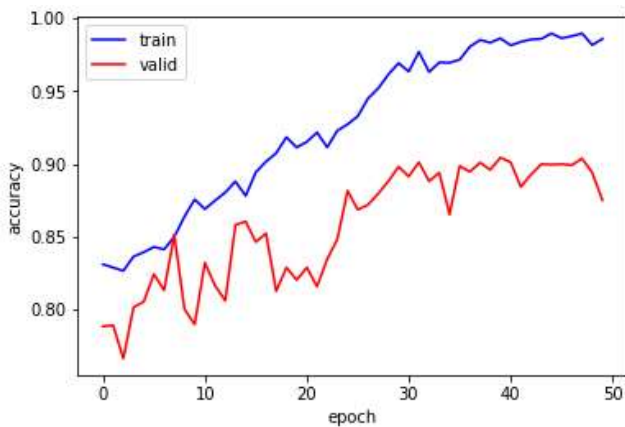
(.a.)



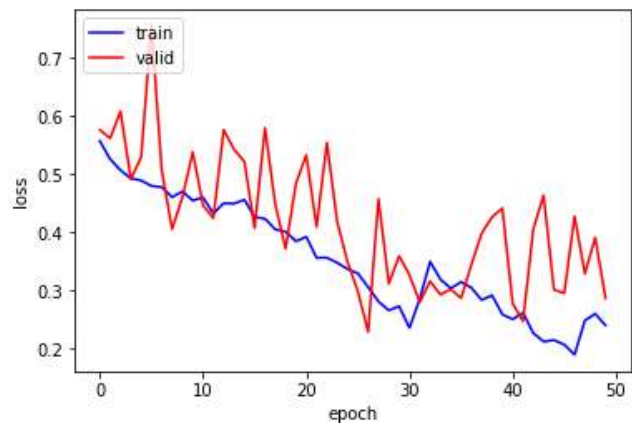
(b.)

Figure 93: Courbes de perte d'apprentissage et de précision du réseau CNN

ResNet50: (a) courbe de perte (b) courbe de précision.



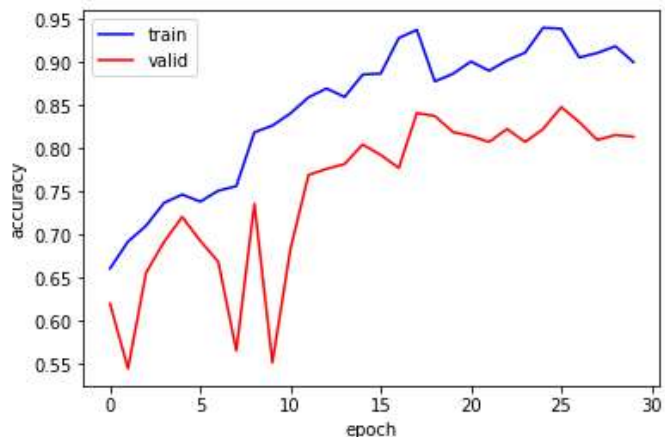
(.a.)



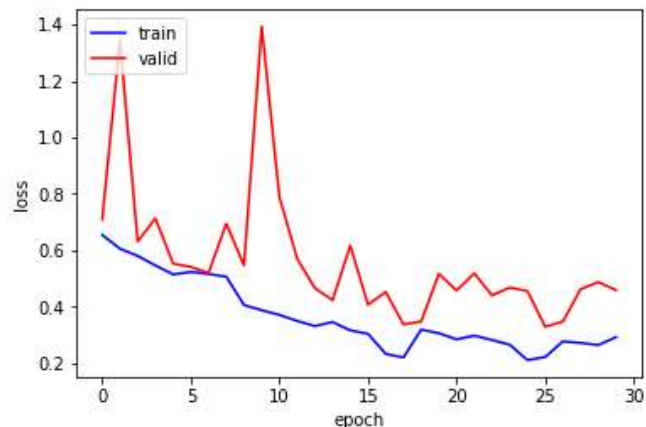
(b.)

Figure 94: Courbes de perte d'apprentissage et de précision du réseau CNN

InceptionV3: (a) courbe de perte (b) courbe de précision.



(a)



(b)

Figure 95: Courbes de perte d'apprentissage et de précision du réseau CNN

MobileNet: (a) courbe de perte (b) courbe de précision.

7.6. Courbe du ROC pour la phase du test

Pour suivre et assurer la convergence de la phase de test de notre modèle nous avons effectué une analyse ROC en utilisant l'aire sous la courbe (AUC) pour les trois classes de pathologies. Les courbes obtenues sur les figure 96–101 montrent bien la convergence de notre modèle pour les trois classes de pathologie qui constituent le jeu de données de notre ensemble considéré.

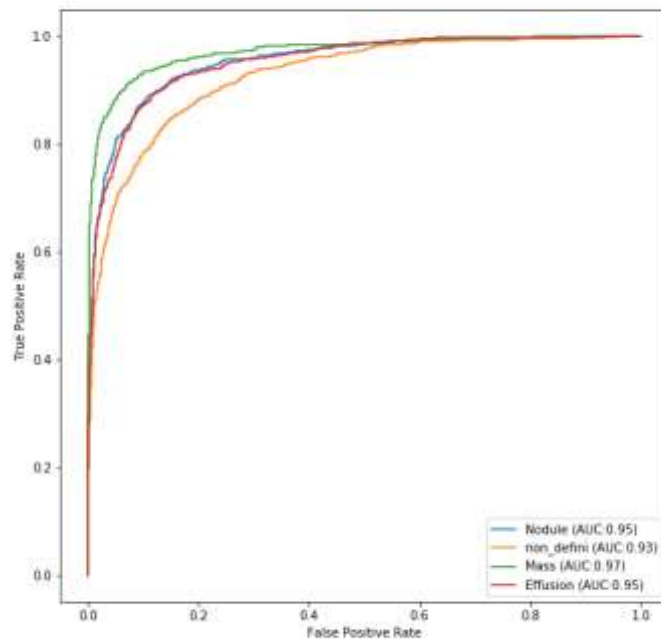


Figure 96: Courbes ROC obtenues par le réseau Xception pour les classes : Masse, Nodule, Effusion et Non défini

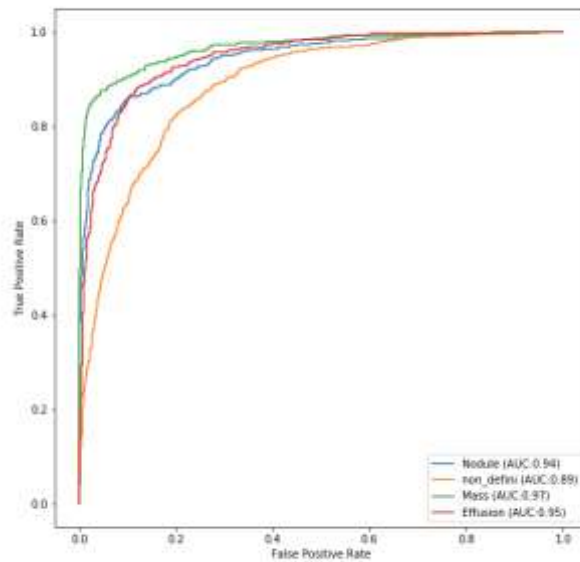


Figure 97: Courbes ROC obtenues par le réseau DenseNet121 pour les classes : Masse, Nodule, Effusion et Non défini

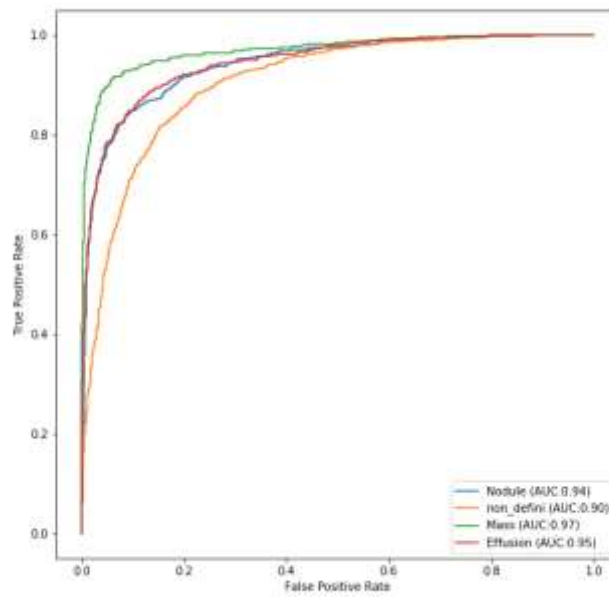


Figure 98: Courbes ROC obtenues par le réseau ResNet50 pour les classes : Masse, Nodule, Effusion et Non défini

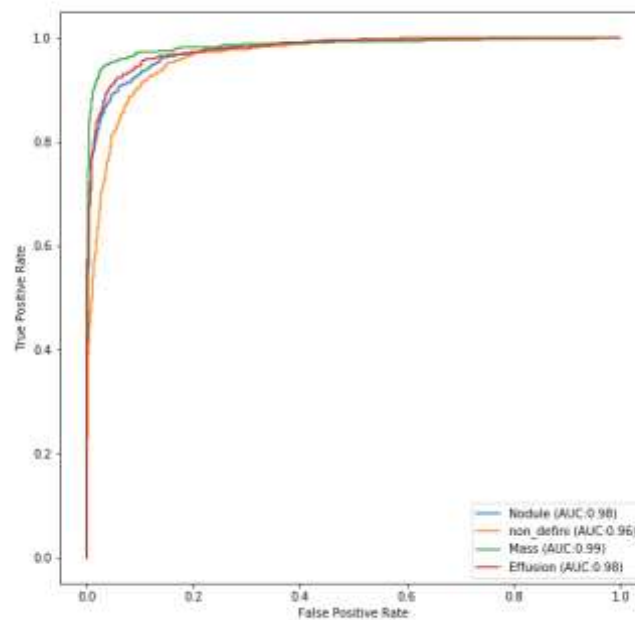


Figure 99 Courbes ROC obtenues par le réseau Inception V3 pour les classes : Masse, Nodule, Effusion et Non défini.

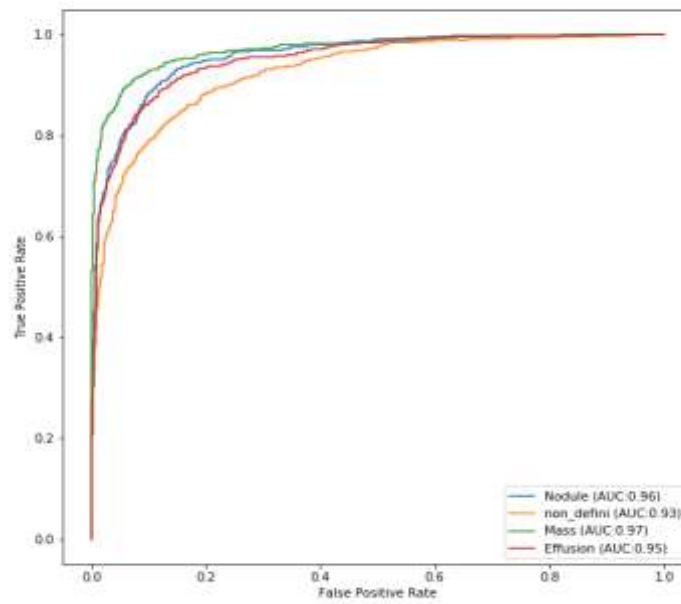


Figure 100: Courbes ROC obtenues par le réseau MobileNet pour les classes : Masse, Nodule, Effusion et Non défini

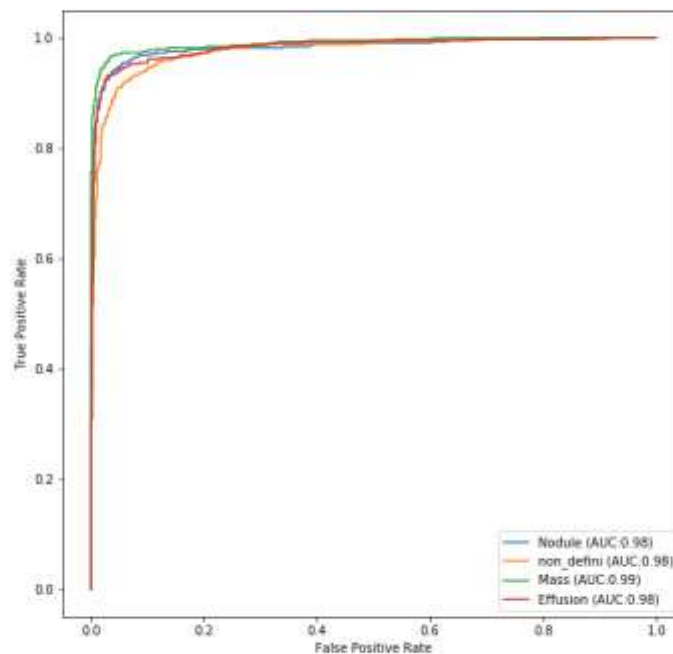


Figure 101: Courbes ROC obtenues par le réseau Ensemble Modèle pour les classes : Masse, Nodule, Effusion et Non défini

7.7. Performances mesurées de notre modèle proposé

Nous avons rassemblé dans le tableau 7 les mesures obtenues des performances de notre modèle proposé. Nous remarquons que les résultats obtenus par l'ensemble Modèle proposé (Stacking Model) dépasse ceux obtenus pour chaque réseau CNN pris individuellement ce qui montre la robustesse de l'approche proposé.

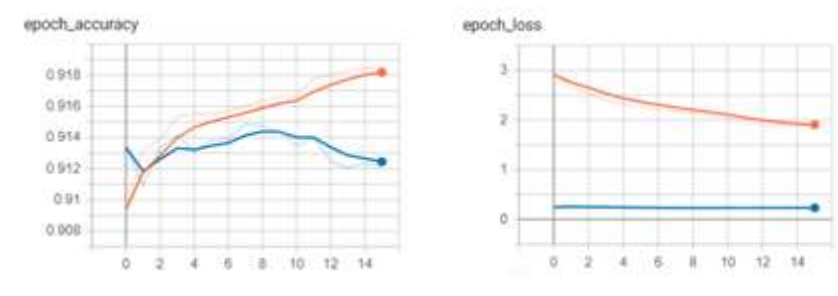
Tableau 7: Performance (%) de nos modèles sur l'ensemble de test

| CNN | Précision | Precision (macro) | F1_score (macro) | Recall (macro) | C. kappa | Sensitivity | Specificity | AUC |
|-------------------|---------------|----------------------|---------------------|-------------------|---------------|---------------|---------------|---------------|
| Xception | 83,94% | 84,01% | 83,98% | 83,96% | 78,59% | 91,76% | 89,42% | 96,18% |
| DenseNet12 1 | 84,75% | 84,79% | 84,65% | 84,79% | 79,68% | 92,19% | 92,46% | 96,38% |
| ResNet50 | 77,83% | 77,85% | 77,78% | 77,86% | 70,43% | 89,84% | 82,05% | 94,06% |
| InceptionV3 | 88,73% | 88,70% | 88,71% | 88,75% | 84,97% | 95,35% | 92,52% | 97,46% |
| MobileNet | 91,23% | 91,30% | 91,24% | 91,24% | 88,31% | 96,82% | 92,65% | 96,30% |
| Ensemble-M | 91,87% | 91,96% | 91,95% | 91,88% | 89,15% | 97,81% | 92,34% | 98,59% |

8.3 ^{EME} EXPERIENCE: RESULTATS POUR L'ENSEMBLE A 14 CLASSES

8.1. Résultats du réseau ResNet

Tout d'abord, nous affichons la courbe de précision et de perte pour les ensembles de formation et de validation.



(b)

Figure 102: la courbe de précision (en haut) et de perte (en bas) sur 25 époques pour ResNet50.

Comme nous le savons dans notre étude, nous avons un problème de classification multi-étiquettes déséquilibré, de sorte que la métrique de précision n'est pas la métrique fiable que nous pourrions utiliser pour évaluer les performances du modèle. Nous utilisons donc l'analyse ROC en utilisant l'aire sous la courbe (AUC) pour toutes les pathologies. La figure 103 résumait tous les résultats de chaque classe de pathologie.

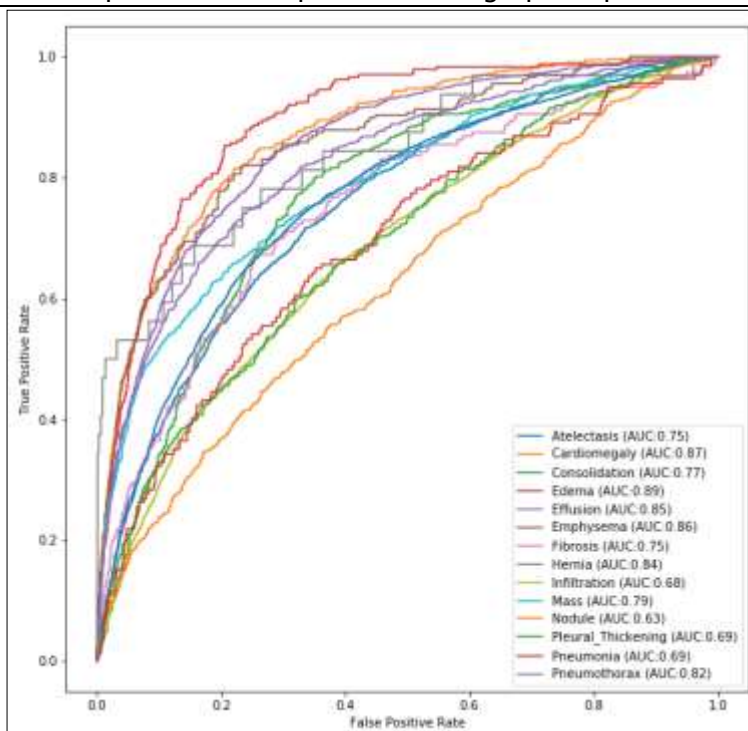


Figure 103: Courbe ROC avec résultat du score AUC du modèle ResNet50 sur les 14 pathologies

La courbe micro-moyenne ROC est de l'ordre de 0,83 (figure 104).

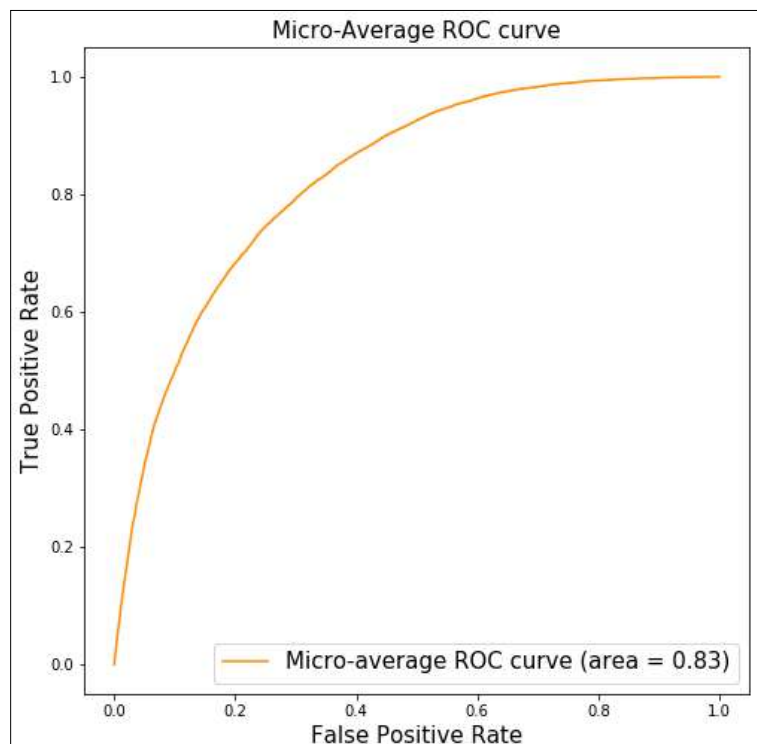


Figure 104: Courbe ROC micro-moyenne pour le modèle ResNet50.

8.2. Résultats du réseau ResNet101

Tout d'abord, nous affichons la courbe de précision et de perte pour les ensembles de formation et de validation (Figure 105).

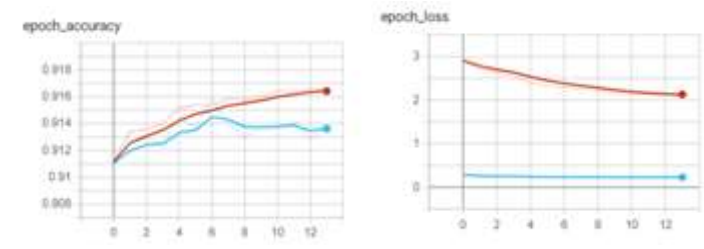


Figure 105: la courbe de précision (en haut) et de perte (en bas) sur 25 époques pour ResNet101.

La figure 106 résumait tous les résultats de chaque classe de pathologie.

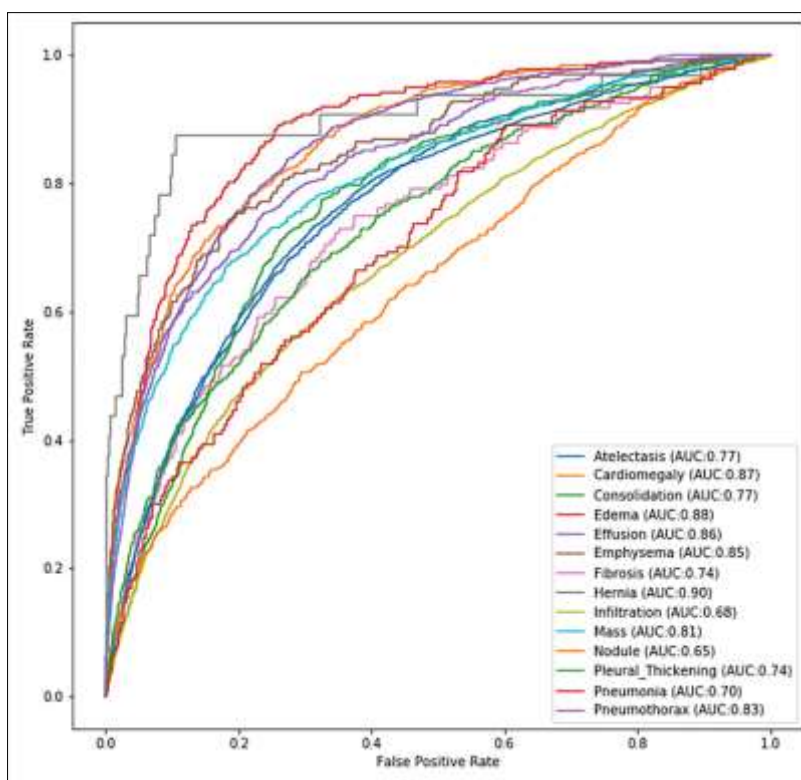


Figure 107: Courbe ROC avec résultat du score AUC du modèle ResNet101 sur les 14 pathologies

La courbe micro-moyenne ROC est de l'ordre de 0.8328 (figure 108).

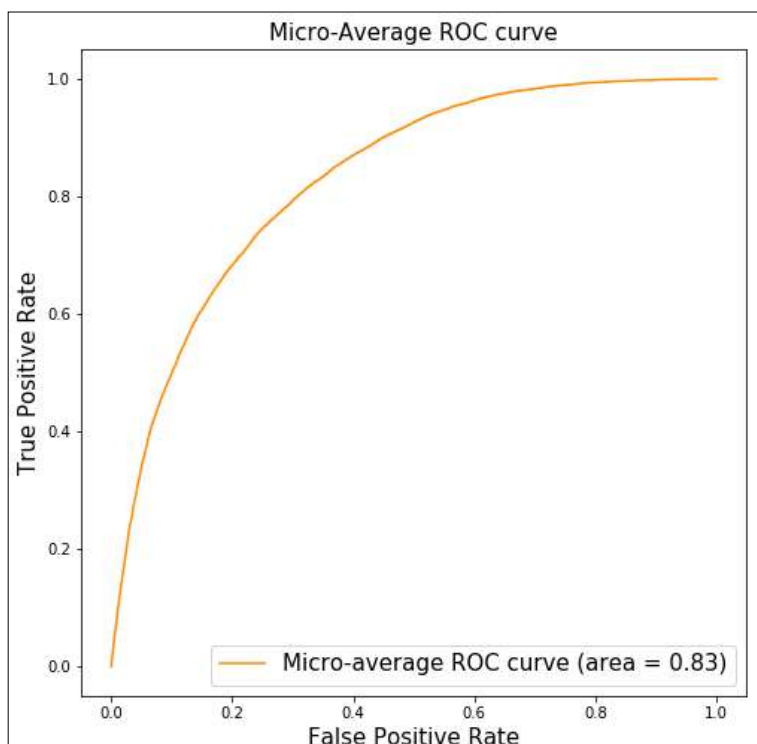


Figure 108: Courbe ROC micro-moyenne pour le modèle ResNet101.

8.3. Résultats du réseau DenseNet121

Nous traçons la courbe ROC pour évaluer les performances du modèle. La figure 109 montre la courbe ROC avec le score AUC.

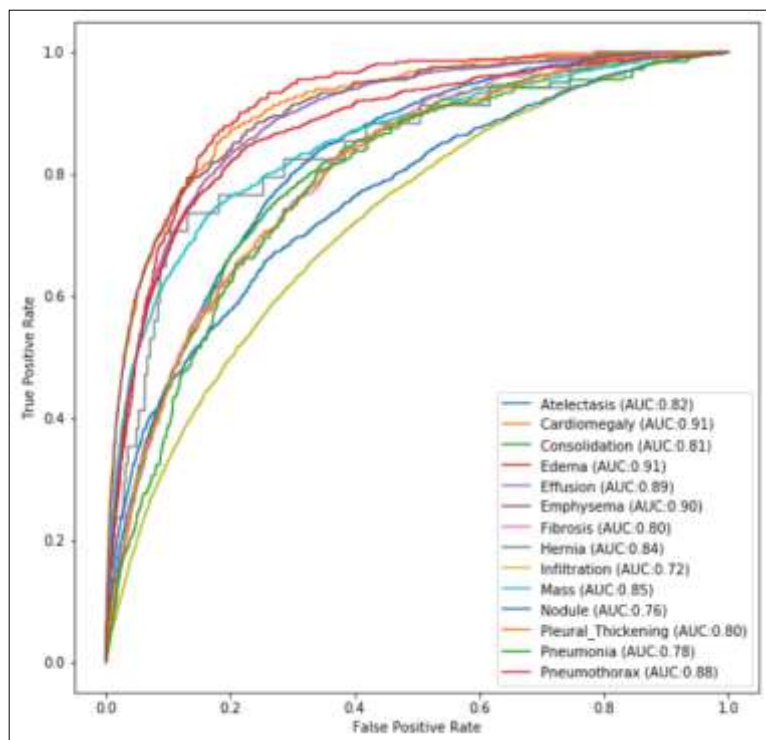


Figure 109: Courbe ROC avec résultat du score AUC du modèle DenseNet121 sur les 14 pathologies.

La courbe micro-moyenne ROC est de l'ordre de **0.84** (figure 110).

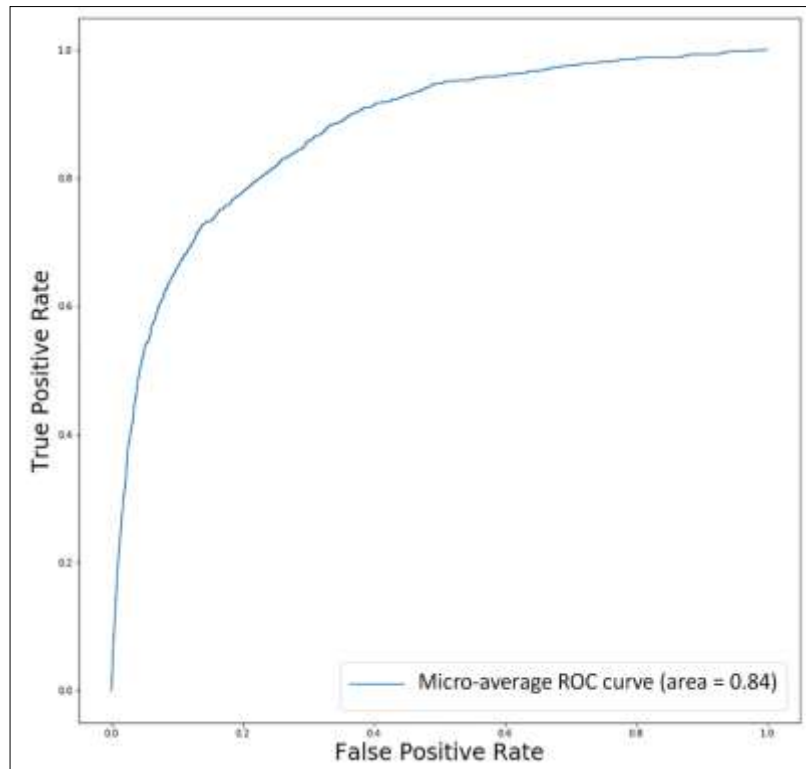


Figure 110: Courbe ROC micro-moyenne pour le modèle DenseNet121.

Pour résumer tous les résultats, nous créons le tableau 8 contenant le score AUC de chaque modèle.

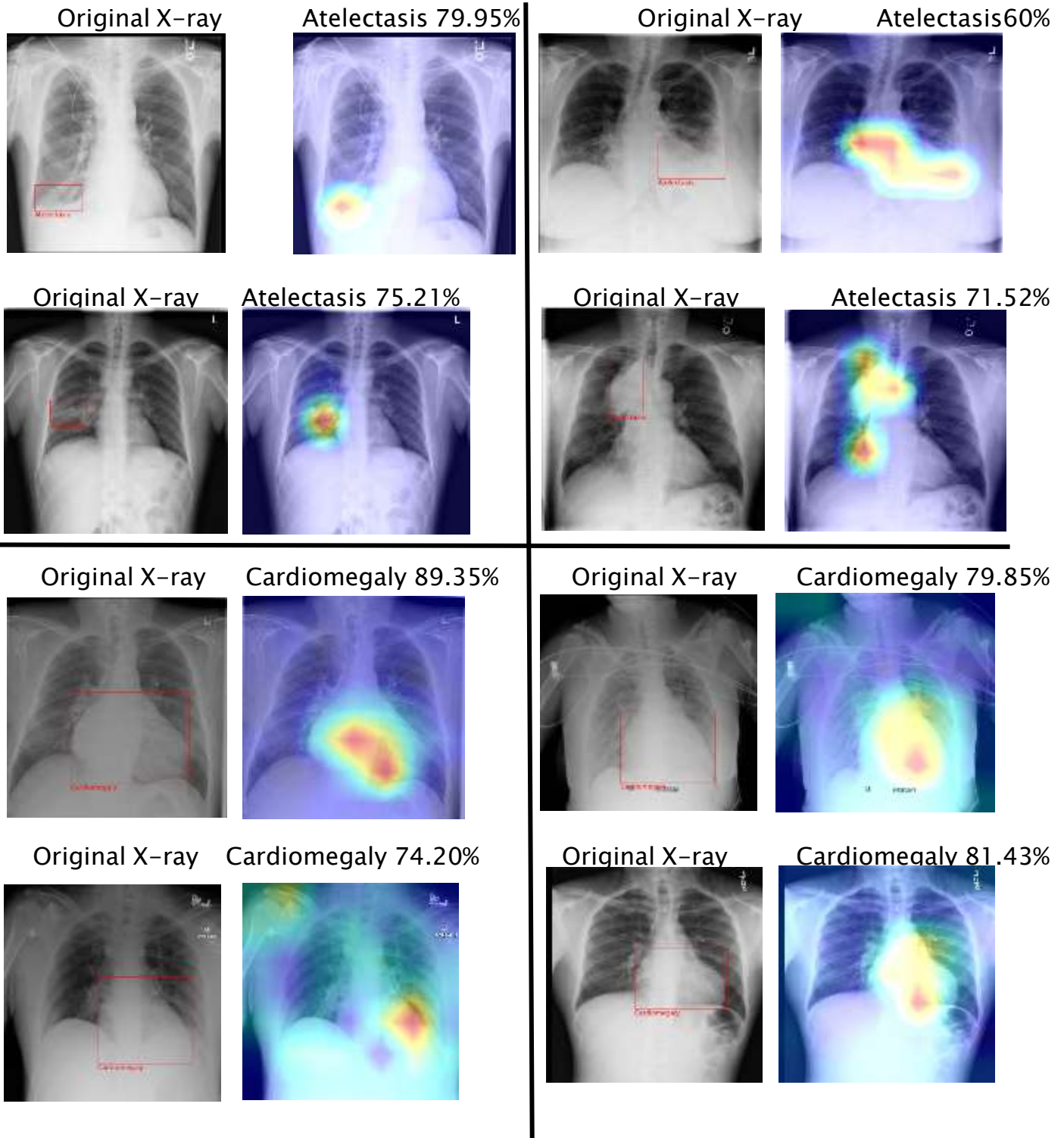
Tableau 8: Résultats AUC pour toutes nos expériences.

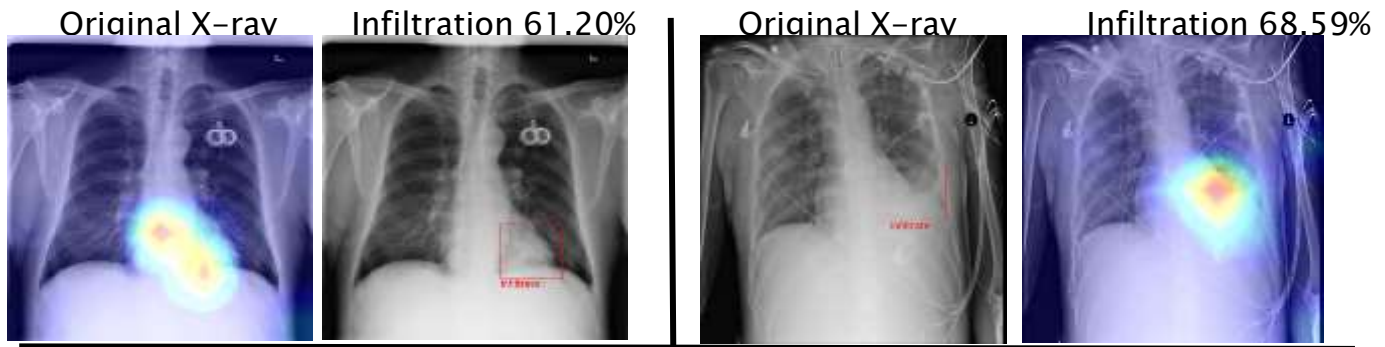
| Pathology | ResNet50 | ResNet101 | DenseNet121 |
|--------------------|----------|-------------|-------------|
| Atelectasis | 0.75 | 0.77 | 0.82 |
| Cardiomegaly | 0.87 | 0.87 | 0.91 |
| Consolidation | 0.77 | 0.77 | 0.81 |
| Edema | 0.89 | 0.88 | 0.91 |
| Effusion | 0.85 | 0.86 | 0.89 |
| Emphysema | 0.86 | 0.85 | 0.90 |
| Fibrosis | 0.75 | 0.74 | 0.80 |
| Hernia | 0.84 | 0.90 | 0.84 |
| Infiltration | 0.68 | 0.68 | 0.72 |
| Masse | 0.79 | 0.81 | 0.85 |
| Nodule | 0.63 | 0.65 | 0.76 |
| Pleural Thickening | 0.69 | 0.74 | 0.80 |
| Pneumonia | 0.69 | 0.70 | 0.78 |
| Pneumothorax | 0.82 | 0.83 | 0.88 |

9. CARTE THERMIQUE GRAD-CAM POUR UNE LOCALISATION GROSSIERE DES LÉSIONS PULMONAIRES AVEC LA PROBABILITE DE CLASSIFICATION

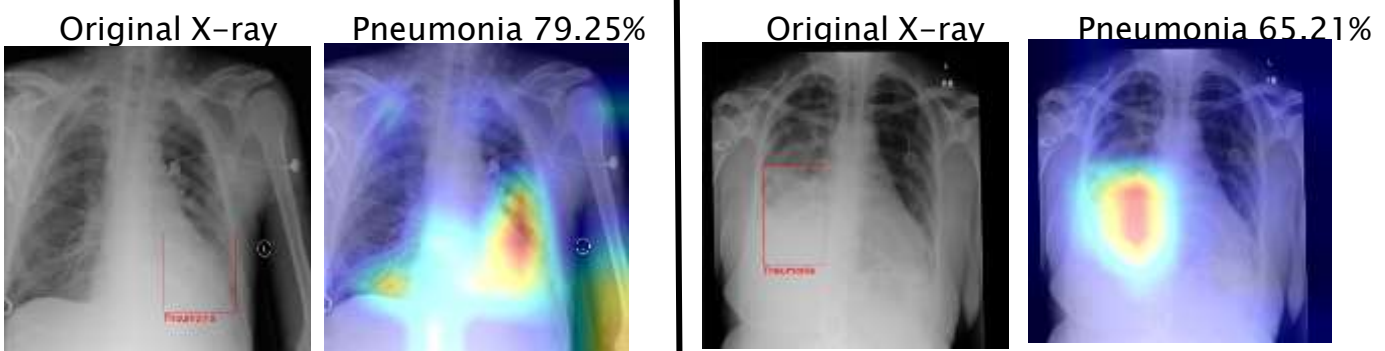
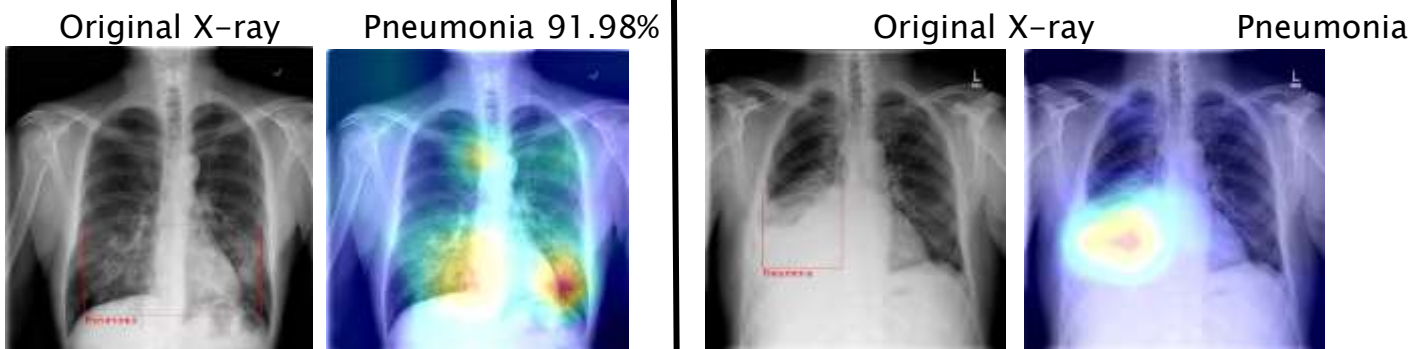
Dans cette section, nous mettrons en œuvre le concept de Grad-CAM qui nous permettra d'évaluer les prédictions de modèles CNN. La méthode met en évidence des régions importantes dans les images d'entrée pour un résultat de classification spécifique en utilisant le gradient de la couche convolutionnelle finale.

Grâce aux résultats de nos modèles, nous décidons d'utiliser le modèle DenseNet121 à cette fin en raison du bon résultat que nous avons obtenu avec l'architecture DenseNet121. Ainsi, dans la figure 81, nous allons présenter quelques images radiographiques avec leur carte thermique générée avec l'algorithme Grad-CAM pour produire une carte de **localisation grossière** mettant en évidence les régions importantes dans les images radiographiques pour prédire le concept, nous prendrons au hasard 2 à 4 images radiographiques de chaque pathologie. La figure 81 montre les résultats de ce processus en comparaison avec la boîte englobante de vérité terrain (couleur rouge) fournie par les radiologues (pour dessiner la boîte englobante des symptômes sur chaque image radiographique, nous utilisons le jeu de données « BBox_List_2017.csv»). Nous calculons la probabilité de la sortie faite par le modèle sur cette classe.

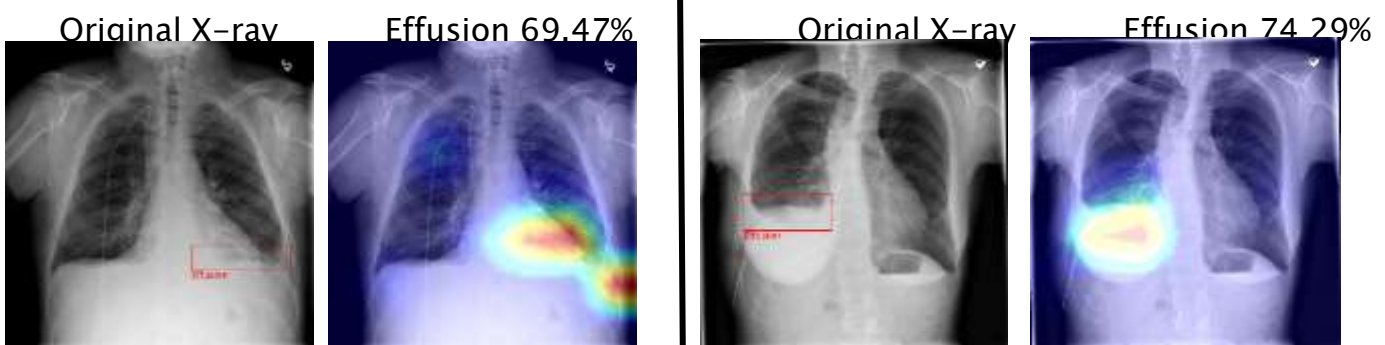




• Pneumonia



• Effusion

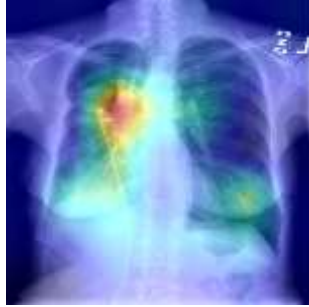


- Pneumothorax

Original X-ray



Pneumothorax 74.1%



Original X-ray



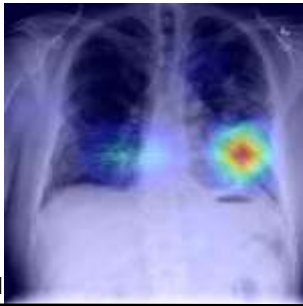
Pneumothorax 80.32%



Original X-ray



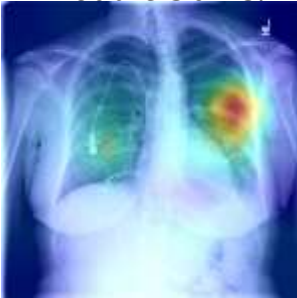
Pneumothorax 68.2%



Original X-ray



Nodule 50.13%



Original X-ray



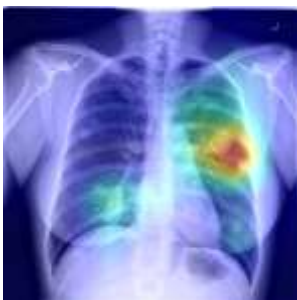
Nodule 52.85%



Original X-ray



Mass 46.2%



Original X-ray



Mass 41.43%

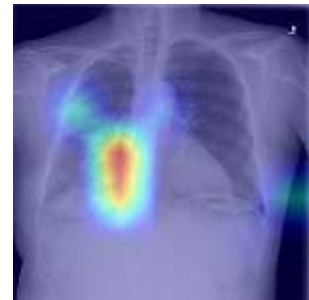
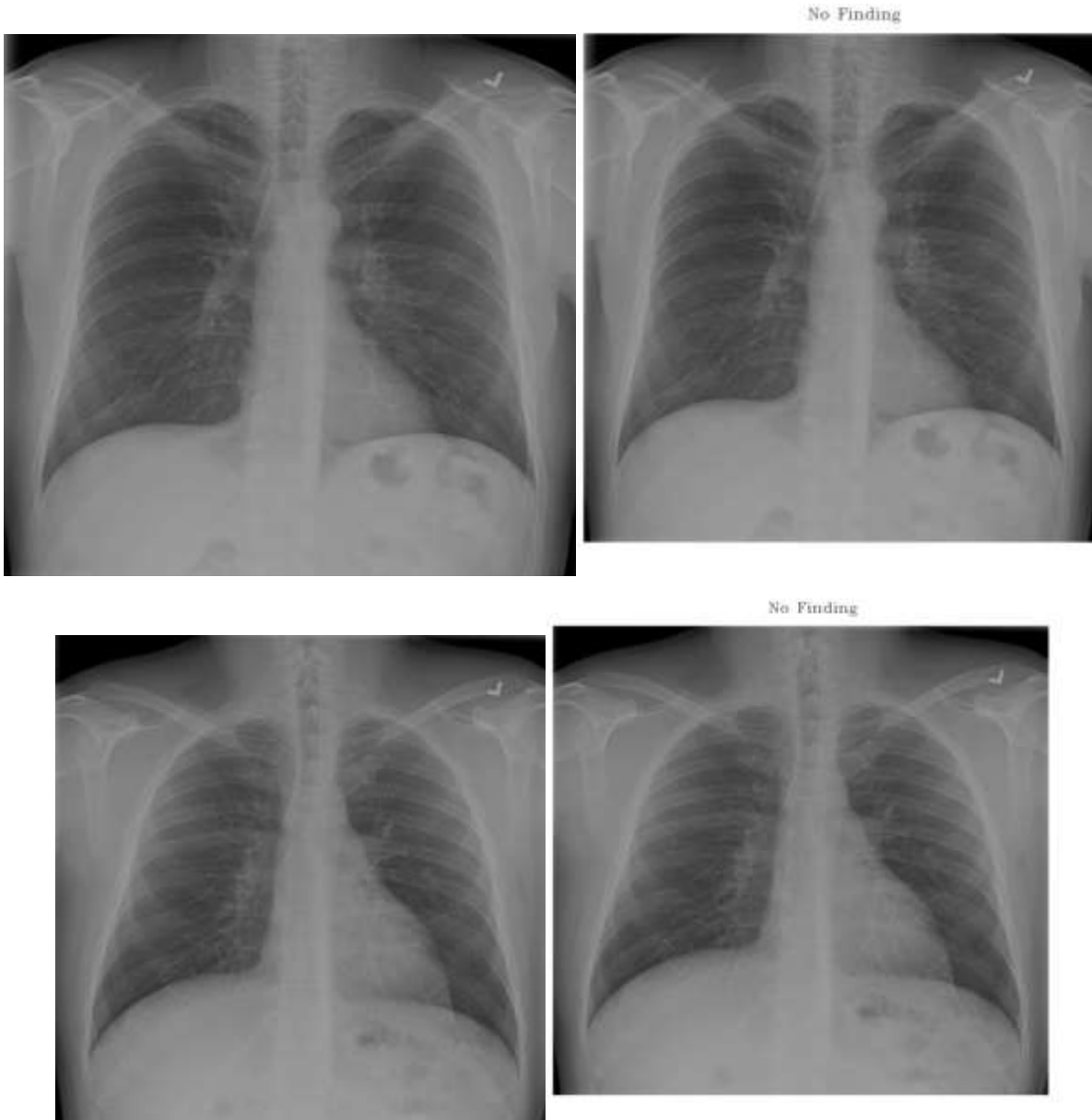


Figure 111 : Exemple de carte thermique Grad-CAM pour chaque classe de maladie pulmonaire

10. EXEMPLE D'IMAGES UTILISE POUR LE TEST

Nous présentons ici un échantillon des résultats de notre procédé. La radiographie de gauche est la radiographie initiale soumise à l'analyse. A sa droite nous retrouvons la même image avec le label déterminé par le logiciel.















11. DISCUSSION

Compte tenu de l'indisponibilité des images scanner ou IRM dans les zones rurales éloignées ainsi que de leur prix exorbitant, il était utile de mettre à disposition du personnel de santé une méthode basée sur l'intelligence artificielle pour prédire les maladies pulmonaires de manière rapide et précise avec de simples images radiographiques du thorax. Dans ce travail, nous avons proposé un système intelligent d'aide à la décision clinique pour la détection précoce de maladies pulmonaires à partir de radiographies thoraciques, qui sont moins chers, facilement accessibles aux populations rurales, et dont le dispositif d'acquisition se désinfecte facilement, se nettoie facilement et se maintient facilement.

Notre modèle utilise l'apprentissage en profondeur grâce à sept réseaux de neurones à convolution extraits de la base de données ImageNet déjà pré-entraînés. Chaque réseau prédira une classe comme image d'entrée. L'avantage de notre modèle est qu'il attribuera un score à chaque prédiction de classe et n'enverra que la prédiction qui obtient le score le plus élevé. De cette manière, nous avons pu obtenir une précision globale moyenne dépassant 83% au moment où les autres modèles pris individuellement donnent une précision inférieure à celle-ci. Avec ces résultats, il peut être confirmé que le modèle proposé peut être utilisé pour détecter d'une façon précoce une maladie pulmonaire avec précision et rapidité, d'autant plus qu'il parvient à discriminer une maladie d'une autre.

La base de donnée n'est pas aussi parfaite qu'on l'imagine étant donné que l'annotation des images a été faite par une machine utilisant la technique de l'analyse du langage naturel (Naturel Langage Processing NLP). Sur le jeu de données NIH Chest X-ray qui est composé de 112 120 images radiographiques seules 983 images ont été annotées par un professionnel de la santé. De plus ces images ne participent pas à la

phase d'apprentissage du réseau mais elles sont utilisées dans la phase du teste. De plus parmi ces images annotées seules huit des maladies (atélectasie, cardiomégalie, épanchement, effusion, masse, nodule, pneumonie et pneumothorax) sont représentées.

Ceci nous a poussé à limiter le nombre de classe constituant l'ensemble de données utilisé en apprentissage. Pour cela nous avons commencé par un ensemble à 3 classes : Masse, Nodule et non défini. Nous avons obtenu une précision moyenne de **prédiction 0.87** en utilisant notre modèle proposé dépassant les résultats déjà publiés en littérature.

Tableau 9: Comparaison entre nos travaux et ceux de la littérature

| | ChestX- Ray8 [6] | Yao et al [29] | Li et al. [30] | DNetLoc [31] | CheXNet [32] |
|----------|---------------------|-------------------|-------------------|-----------------|-----------------|
| Masse | 0.706 | 0.72 | 0.83 | 0.82 | 0.83 |
| Nodule | 0.67 | 0.78 | 0.75 | 0.76 | 0.78 |
| Effusion | 0.78 | 0.80 | 0.87 | 0.83 | 0.83 |

Nous avons aussi mené une expérience en prenant un ensemble à 4 classes : Masse, Nodule, Effusion, Non défini. Nous avons ajouté la classe effusion qui contenait plus d'images bien annotées que les autres classes. De plus la plupart des images de cette classe ne présentaient qu'une seule classe de lésion contrairement aux autres images des autres classes qui présentaient souvent plusieurs maladies sur la même image. Conformément à ce que nous attendions nous avons obtenu une précision moyenne de **0.92** dépassant aussi les résultats publiés en littérature et les résultats trouvés dans l'expérience précédente.

Ceci montre que la performance du résultat dépend étroitement de la nature de la base d'apprentissage et de test utilisée.

Nous pouvons ainsi conclure que si notre logiciel peut déterminer avec précision quel est le type de la lésion en question sur la radiographie thoracique quand celle-ci est unique, il trouve quelques difficultés à interpréter les radiographies compliquées avec plusieurs lésions différentes, qui par ailleurs nécessitent la plupart des cas la réalisation d'un scanner.

CONCLUSION GENERALE

Dans cette thèse nous avons utilisé le Deep Learning pour extraire les caractéristiques graphiques des différentes lésions à partir des images radiographiques thoraciques. Ces types d'images sont plus accessibles pour les patients que les images tomendensitogrammes (scanners) en particulier dans les zones rurales et isolées.

L'architecture que nous avons proposé a été basée sur l'apprentissage profond (Deep Learning) par des réseaux de neurones à convolutions que nous avons exploités pour extraire les caractéristiques graphiques de la maladie à partir des images radiographiques thoraciques. Ces caractéristiques extraites (descripteurs) ont été utilisées pour entraîner notre modèle proposé afin de fournir un diagnostic clinique économisant ainsi un temps critique pour le contrôle des maladies. Pour cela nous avons utilisé l'apprentissage sur une base d'images thoracique en utilisant plusieurs architectures de réseaux neuronaux à convolution tels que le réseau à blocs résiduels (ResNet), le réseau à blocs denses (DenseNet) et le réseau du groupe de recherche VGG (Visual Geometry Group).

Pour augmenter la précision de notre modèle proposé, nous avons choisi dans un premier temps des réseaux déjà entraînés sur la base d'images ImageNet constituées par des millions d'images. Dans un deuxième temps nous avons utilisé la notion de transfert d'apprentissage pour construire une nouvelle architecture appelée ensemble modèle (Stacking Model) dans laquelle nous avons transmis les réseaux déjà entraînés où seule la dernière couche de chaque réseau a été utilisée pour un apprentissage sur notre base d'images radiographiques. De plus le modèle proposé effectuait un tri des différentes prédictions issues de chaque réseau de telle sorte que seule la prédiction qui a obtenu le plus grand score sera prise en considérée comme prédiction finale. De cette façon nous avons pu augmenter la prédiction de notre modèle par rapport à la prédiction de chaque modèle.

Les données radiologiques utilisées ont été collectées à partir de la base de

données du NIH. Au terme de la conception et de l'apprentissage de ce réseau profond, nous avons évalué son efficacité et nous avons pu démontrer l'efficacité et l'utilité d'un tel système.

RÉSUMÉS

Résumé

Les radiographies standards sont un outil d'imagerie utilisé depuis plusieurs décennies en complément de l'examen clinique, elles permettent de visualiser les lésions pathologiques (fractures, masses, épanchement...) et de poser le diagnostic des lésions anatomiques. Elles utilisent les rayons X qui ne sont pas coûteux à produire et restent un moyen de diagnostic non invasif.

L'interprétation des radiographies du thorax n'est pas des plus simples, raison pour laquelle il existe actuellement un large panel d'outils informatiques utilisé pour la détection des lésions sur les images radiographiques. Ces outils sont en développement depuis quelques décennies, mais leur efficacité n'est toujours pas suffisante et c'est pourquoi ils ne constituent aujourd'hui qu'un élément d'aide au diagnostic.

De nos jours, il existe plusieurs bases de données d'images radiologiques, normales et pathologiques, qui ont permis de mettre en place un procédé d'apprentissage pour les instruments informatiques vis-à-vis de la détection de certaines pathologies, connu sous le nom de Deep learning, dépassant ainsi les capacités humaines en termes de détection et de caractérisation.

L'objectif de notre travail est de développer un système intelligent d'aide à la décision clinique pour le diagnostic des lésions pulmonaires à partir de radiographies pulmonaires qui sont plus accessibles pour les populations des zones rurales et isolées, sous forme d'un logiciel informatique de détection automatique des différentes lésions en utilisant des algorithmes de reconnaissance d'images qui ont déjà fait leurs preuves dans d'autres domaines.

Dans ce but, nous avons entraîné une intelligence artificielle sur une base de données internationale disponible en ligne, à savoir « ChestX-ray14 », qui comprend

108 948 radiographies thoraciques vues de face de 32 717 patients avec leur diagnostic, ou label, extraits à partir des rapports radiologiques associés. Les labels sont listés comme suit : Infiltration, Pleurésie, Pneumothorax, Pneumonie, Atélectasie, Nodule, Masse, Cardiomégalie, Emphysème, Œdème et Hernie.

Les résultats préliminaires montrent que notre application était assez précise pour reconnaître les diagnostics avec un score ROC moyen de 86% de précision.

Abstract

Standard x-rays are an imaging tool that has been used for several decades as a complement to clinical examination, they make it possible to visualize pathological lesions (fractures, masses, effusion, etc.) and to diagnose anatomical lesions. They use X-rays which are inexpensive to produce and remain a non-invasive diagnostic tool.

Interpreting chest x-rays is not straightforward, which is why there is currently a wide range of computer tools used for detecting lesions on x-ray images. These tools have been in development for several decades, but their effectiveness is still not sufficient and that is why they only constitute a diagnostic aid element today.

Nowadays, there are several databases of radiological images, normal and pathological, which have made it possible to set up a learning process for computer instruments vis-à-vis the detection of certain pathologies, known as name of Deep learning, thus exceeding human capacities in terms of detection and characterization.

The objective of our work is to develop an intelligent clinical decision support system for the diagnosis of lung lesions from chest x-rays which are more accessible for populations in rural and isolated areas, in the form of computer software. automatic detection of different lesions using image recognition algorithms which have already proven their worth in other fields.

To this end, we trained an artificial intelligence on an international database available online, namely "ChestX-ray14", which includes 108,948 chest x-rays seen from the front of 32,717 patients with their diagnosis, or label, extracted from from the associated radiological reports. The labels are listed as follows: Infiltration, Pleurisy, Pneumothorax, Pneumonia, Atelectasis, Nodule, Mass, Cardiomegaly, Emphysema, Fibrosis, Edema and Hernia.

Preliminary results show that our app was accurate enough to recognize diagnoses with an average ROC score of 86% accuracy.

ملخص

تعد الأشعة السينية القياسية أداة تصوير تم استخدامها لعدة عقود كمكمل للفحص السريري ، فهي تتيح تصور الآفات المرضية (الكسور والكتل والانصباب وما إلى ذلك) وتشخيص الآفات التشريحية. يستخدمون أشعة سينية غير مكلفة لإنتاج وتظل أداة تشخيص غير جراحية.

تفسير صور الأشعة السينية للصدر ليس بالأمر السهل ، ولهذا يوجد حالياً مجموعة كبيرة من أدوات الكمبيوتر المستخدمة للكشف عن الآفات في صور الأشعة السينية. ظلت هذه الأدوات قيد التطوير لعدة عقود ، لكن فعاليتها لا تزال غير كافية ولهذا السبب فهي تشكل عنصراً مساعداً للتشخيص اليوم فقط.

في الوقت الحاضر ، هناك العديد من قواعد البيانات للصور الإشعاعية ، العادية والمرضية ، والتي جعلت من الممكن إعداد عملية تعلم لأدوات الكمبيوتر مقابل اكتشاف بعض الأمراض ، والمعروفة باسم اسم التعلم العميق ، وبالتالي تجاوز القدرات البشرية من حيث الكشف والتوصيف.

الهدف من عملنا هو تطوير نظام ذكي لدعم القرار السريري لتشخيص آفات الرئة من الأشعة السينية للصدر والتي يمكن الوصول إليها بشكل أكبر للسكان في المناطق الريفية والمعزولة ، في شكل برامج كمبيوتر. الكشف التلقائي عن الآفات المختلفة باستخدام خوارزميات التعرف على الصور التي أثبتت بالفعل قيمتها في مجالات أخرى.

تحقيقاً لهذه الغاية ، قمنا بتدريب ذكاء اصطناعي على قاعدة بيانات دولية متاحة على الإنترنت ، وهي "CHESTX-RAY14" ، والتي تشمل 108.948 صورة شعاعية للصدر شوهدت من أمام 32717 مريضاً مع تشخيصهم أو ملصقهم المستخرج من من التقارير الإشعاعية ذات الصلة. يتم سرد الملصقات على النحو التالي: التسلل ، التهاب الجنبة ، استرواح الصدر ، الالتهاب الرئوي ، انخماص الرئة ، العقيدات ، الكتلة ، تضخم القلب ، انتفاخ الرئة ، التليف ، الوذمة والفتق.

تُظهر النتائج الأولية أن تطبيقنا كان دقيقاً بما يكفي للتعرف على التشخيصات بمتوسط درجة

ROC تبلغ 86% .

REFERENCES

-
- [1] Manu Goyal; Thomas Knackstedt; Shaofeng Yan; Saeed Hassanpour, 'Artificial Intelligence–Based Image Classification for Diagnosis of Skin Cancer: Challenges and Opportunities', *Computers in Biology and Medicine*, DOI: 10.1016/j.compbiomed.2020.104065, 2020–10–27.
- [2] Kallianos K., Mongan J., Antani S., Henry T., Taylor A., Abuya J., Kohli M. How far have we come? Artificial intelligence for chest radiograph interpretation. *Clin. Radiol.* 2019;74:338–345. doi: 10.1016/j.crad.2018.12.015.
- [3] Hicham Amakdouf, Amal Zouhri, Mostafa El Mallahi, Ahmed Tahiri, Driss Chenouni and Hassan Qjidaa, 'Artificial intelligent classification of biomedical color image using quaternion discrete radial Tchebichef moments' *Multimedia Tools and Applications* September 2020
- [4] Choe J, Lee SM, Do KH, Lee G, Lee JG, Lee SM, et al., ' Deep Learning–based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses. *Radiology.*', 2019; 292(2): 365–73.
- [5] Negassi M, Suarez–Ibarrola R, Hein S, Miernik A, Reiterer A. ' Application of artificial neural networks for automated analysis of cystoscopic images: a review of the current status and future prospects', *World Journal of Urology.* 2020.
- [6] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald Summers, ChestX–ray8: Hospital–scale Chest X–ray Database and Benchmarks on Weakly–Supervised Classification and Localization of Common Thorax Diseases, *IEEE CVPR*, pp. 3462–3471, 2017
- [7] Qjidaa M., Y. Mechbal; A. Ben–fares; H. Amakdouf, Maaroufi M., Alami,

- B., Qjidaa, H., ' Early detection of COVID19 by deep learning transfer Model for populations in isolated rural areas', IEEE, International Conference on Intelligent Systems and Computer Vision, ISCV 9–11 June 2020, DOI:10.1109/ISCV49265.2020.9204099
- [8] Qjidaa, M., A. Ben-fares; Y. Mechbal; H. Amakdouf, Maaroufi M., Alami, B., Qjidaa, H., ' Development of a clinical decision support system for the early detection of COVID–19 using deep learning based on chest radiographic images', IEEE, International Conference on Intelligent Systems and Computer Vision, ISCV 9–11 June 2020, DOI: 10.1109/ISCV49265.2020.9204282
- [9] Taylor, A.G.; Mielke, C.; Mongan, J. (2018). Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study. PLoS Med. 2018, 15, e1002697.
- [10] Abderrahim Mesbah, Aissam Berrahou, Hicham Hammouchi, Hassan Berbia, Hassan Qjidaa, Mohamed Daoudi,. (2019). Lip reading with Hahn Convolutional Neural Networks', Image and Vision Computing Volume 88, August 2019, Pages 76–83 , Elsevier.
- [11] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, Ronald Summers, ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases, IEEE CVPR, pp. 3462–3471, 2017
- [12] Altman, Naomi S. "An introduction to kernel and nearest-neighbor nonparametric regression" (PDF). The American Statistician. 46 (3): 175–185. (1992). doi:10.1080/00031305.1992.10475879
- [13] Breiman, L. Using adaptive bagging to debias regressions, Technical Report 547, Statistics Dept. UCB [1999]

-
- [14] Boser, B., Guyon, I., Vapnik, V.: A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, Pittsburgh (1992).
- [15] Hastie, Trevor. Tibshirani, Robert. Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, NY, 2009
- [16] Rina Dechter. *Learning while searching in constraint-satisfaction problems*. University of California, Computer Science Department, Cognitive Systems Laboratory, 1986.
- [17] Aizenberg, Igor, Aizenberg, Naum N., Vandewalle, Joos P.L., 'Multi-Valued and Universal Binary Neurons Theory, Learning and Applications', Book, Springer, ISBN 978-1-4757-3115-6, 2000
- [18] LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; Jackel, L. D. "Backpropagation Applied to Handwritten Zip Code Recognition". *Neural Computation*. 1 (4): 541-551. (December 1989).doi:10.1162/neco.1989.1.4.541. ISSN 0899-7667.
- [19] Krizhevsky, Alex; Sutskever, Ilya; Hinton, Geoffrey E.. "ImageNet classification with deep convolutional neural networks". *Communications of the ACM*. 60 (6): 84-90. (2017-05-24) doi:10.1145/3065386. ISSN 0001-0782
- [20] Matthew D Zeiler, Rob Fergus, 'Visualizing and Understanding Convolutional Networks' *Computer Vision and Pattern Recognition*, arXiv:1311.2901 [cs.CV] , 12 Nov 2013
- [21] Christian Szegedy et al., 'Going Deeper with Convolutions', *Computer Vision and Pattern Recognition*, arXiv:1409.4842 [cs.CV], 17 Sep 2014.
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, 'Deep Residual

-
- Learning for Image Recognition' Computer Vision and Pattern Recognition, arXiv:1512.03385 [cs.CV], 10 Dec 2015.
- [23] Dauphin et al. , ' Identifying and attacking the saddle point problem in high-dimensional non-convex optimization', Conference, Advances in Neural Information Processing Systems, January 2014
- [24] Jeffrey Dean, Greg S. Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Quoc V. Le, Mark Z. Mao, Marc Aurelio Ranzato, Andrew Senior, Paul Tucker, Ke Yang, and Andrew Y. Ng. Large Scale Distributed Deep Networks. NIPS 2012: Neural Information Processing Systems, pages 1-11, 2012.
- [25] Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global Vectors for Word Representation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 1532-1543. <http://doi.org/10.3115/v1/D14-1162>
- [26] Diederik P. Kingma, Jimmy Ba, 'Adam: A Method for Stochastic Optimization', conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015
- [27] Paul R et al. Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma. Tomography 2, 388-395 (2016).
- [28] Yifan Peng et al., 'Transfer Learning in Biomedical Natural Language Processing: An Evaluation of BERT and ELMo on Ten Benchmarking Datasets', Proceedings of the 18th BioNLP Workshop and Shared Task, January 2019, DOI: 10.18653/v1/W19-5006
- [29] Li Yao, Eric Poblenz, Dmitry Dagunts, Ben Covington, Devon Bernard, and Kevin Lyman. 2017. Learning to diagnose from scratch by exploiting dependencies among labels. arXiv preprint arXiv:1710.10501 (2017).

- [30] Zhe Li, Chong Wang, Mei Han, Yuan Xue, Wei Wei, Li-Jia Li, and Fei-Fei Li. 2017. Thoracic Disease Identification and Localization with Limited Supervision. arXiv preprint arXiv:1711.06373 (2017).
- [31] Sebastian Guendel, Sasa Grbic, Bogdan Georgescu, Kevin Zhou, Ludwig Ritschl, Andreas Meier, and Dorin Comaniciu. 2018. Learning to recognize abnormalities, in chest x-rays with location-aware dense networks. arXiv preprint, arXiv:1803.04565 (2018).
- [32] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. 2017. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv preprint arXiv:1711.05225 (2017).

تشخيص إصابة الرئة من الأشعة الرئوية (SADC) إنشاء وتقييم نظام دعم القرار السريري الذكي

الأطروحة

قدمت و نوقشت علانية يوم 2020/12/22

من طرف

السيد مامون اقجيدع
المزداد في 05 غشت 1993 بفاس

لنيل شهادة الدكتوراه في الطب

الكلمات المفتاحية

تصوير الصدر بالأشعة السينية - نظام المساعدة التشخيصية - الذكاء الاصطناعي - التعلم العميق - الشبكات العصبية

اللجنة

| | |
|-----------|--|
| الرئيس | السيد سعيد بجراف أستاذ في علم الفيزياء الإحيائية |
| المشرف | السيد مصطفى المعروفي أستاذ في علم الأشعة |
| أعضاء | السيد مولاي يوسف العلوي الامراني أستاذ مبرز في علم الأشعة |
| | السيد بدرالدين علمي أستاذ مبرز في علم الأشعة |
| عضو مشارك | السيد اقجيدع حسن أستاذ في الإعلاميات |