



HAL
open science

**Analyses quantitatives de l'imagerie thoracique
tomodensitométrique dans des maladies bronchiques et
du parenchyme pulmonaire en relation avec des troubles
fonctionnels respiratoires : approches automatisées
basées sur l'intelligence artificielle**

Thi Trieu Nghi Hoang

► **To cite this version:**

Thi Trieu Nghi Hoang. Analyses quantitatives de l'imagerie thoracique tomodensitométrique dans des maladies bronchiques et du parenchyme pulmonaire en relation avec des troubles fonctionnels respiratoires : approches automatisées basées sur l'intelligence artificielle. Imagerie. Université Paris Cité, 2021. Français. NNT : 2021UNIP5044 . tel-03859167

HAL Id: tel-03859167

<https://theses.hal.science/tel-03859167>

Submitted on 18 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Paris

ED 563 Médicament, Toxicologie, Chimie, Imageries

UM3 - Institut Cochin

ANALYSES QUANTITATIVES DE L'IMAGERIE THORACIQUE TOMODENSITOMETRIQUE DANS DES MALADIES BRONCHIQUES ET DU PARENCHYME PULMONAIRE EN RELATION AVEC DES TROUBLES FONCTIONNELS RESPIRATOIRES : APPROCHES AUTOMATISEES BASEES SUR L'INTELLIGENCE ARTIFICIELLE

Par Thi Trieu Nghi HOANG

Thèse de doctorat d'Imageries

Dirigée par le Professeur Anh-Tuan DINH-XUAN

Présentée et soutenue publiquement le 14 décembre 2021

Devant un jury composé de :

Colas TCHERAKIAN, Maître de conférences HDR, Université Versailles Saint Quentin en Yvelines, rapporteur

Yurdagül UZUNHAN, Professeur, Université Paris XIII Nord, rapporteur

Marie-Pierre REVEL DUBOIS, Professeur, Université de Paris, examinatrice

Anh-Tuan DINH-XUAN, Professeur, Université de Paris, directeur de thèse



Except where otherwise noted, this is work licensed under
<https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>

Titre : Analyses quantitatives de l'imagerie thoracique tomodensitométrique dans des maladies bronchiques et du parenchyme pulmonaire en relation avec des troubles fonctionnels respiratoires : approches automatisées basées sur l'intelligence artificielle.

Résumé :

L'analyse quantitative des lésions pulmonaires en imagerie thoracique tomodensitométrique est désormais possible grâce aux outils d'aide au diagnostic (ou Computer-Aided Diagnosis - CAD). Cette approche permet également une meilleure compréhension des mécanismes physiopathologiques grâce aux outils d'intelligence artificielle et à l'approche radiomique. Dans le cadre de notre Thèse, nous avons pu appliquer les outils de CAD et d'intelligence artificielle (machine learning et deep-learning) pour construire des scores de sévérité dans la dyskinésie ciliaire primitive (DCP), prédire la survenue du dysfonctionnement chronique du greffon pulmonaire (DCGP) et la mortalité à court terme de la sclérodemie systémique (ScS). Nous avons également combiné les résultats de machine learning sur les deux types de filtre médiastinal et pulmonaire du scanner thoracique. Nous avons ainsi amélioré l'efficacité des algorithmes d'apprentissage profond sans avoir recours à la nouvelle acquisition de scanner irradié. En pratique, nous avons montré que la caractéristique radiomique de premier ordre autorise la quantification automatique des modifications des images et l'établissement d'un score objectif pour évaluer la gravité de la DCP. L'analyse préliminaire de la combinaison des caractéristiques de densité en histogramme et de texture de la radiomique des scanners inspiratoires pourrait également prédire les modifications fonctionnelles respiratoires de la DCGP 6 mois avant leur survenue. Mais les résultats de notre étude préliminaire, n'étant pas statistiquement significatifs, nécessitent des études à plus grande échelle passant par des collaborations multicentriques. Pour la ScS, l'algorithme de « deep-learning » AtlasNet permet la quantification des lésions du parenchyme pulmonaire et la prédiction de la mortalité à court terme. Dans le dernier chapitre de notre thèse, nous rapportons l'influence du choix de filtre de reconstruction en TDM sur l'entraînement des modèles de machine learning pour la segmentation des zones lésionnelles pulmonaires interstitielles dans la ScS et la COVID-19, permettant ainsi d'augmenter la performance des modèles combinant les deux filtres de reconstruction.

Mots clefs : imagerie thoracique tomodensitométrique, intelligence artificielle, computer aide diagnostic, machine learning, radiomique, dyskinésie ciliaire primitive, sclérodemie systémique, dysfonction chronique du greffon pulmonaire

Title: Quantitative Analyses of Chest CT Imaging in Bronchial and Pulmonary Parenchyma Diseases Related to Lung Function Disturbances: An Automated Approach Based on Artificial Intelligence.

Abstract :

Quantitative analyses of structural lesions in thoracic tomodensitometric imaging of airways and lung parenchyma are now possible with computer-aided diagnostic tools (CAD). Underlying physiopathological mechanisms can also be better deciphered using artificial intelligence tools and the radiomic approach. In this PhD thesis, we have applied CAD, machine learning and deep-learning tools to build severity scores for primary ciliary dyskinesia (PCD), and to predict the occurrence of chronic lung allograft dysfunction (CLAD) and short-term mortality rate of patients with systemic sclerosis (ScS). By combining the results of machine learning on the two types of mediastinal and pulmonary kernels of the chest scanner, we can improve the efficiency of machine learning algorithms without the need to acquire a new irradiated scanner. For example, we have shown that using first-order radiomic parameters can allow automatic quantification of changes in imaging and establishment of an objective score for the assessment of disease severity in patients with PCD. For CLAD, the preliminary analysis of the combination of the density characteristics in histogram and the texture of the radiomics of the inspiratory scans could predict 6 months ahead of the occurrence of lung function tests changes. Our preliminary results however did not reach statistical significance and require larger studies with multicentre collaborations. For ScS, the AtlasNet deep-learning algorithm allows quantification of lung parenchyma lesions to help predict short-term mortality. In the last chapter of this thesis, we report the influence of the choice of reconstruction kernels in CT on the training of machine learning models in the process of segmentation of pathologic areas of interstitial lung diseases of ScS and COVID-19. This will hopefully offer the possibility to improve the performance of models in combinations of two reconstruction kernels.

Keywords : CT chest imaging, artificial intelligence, computer aided diagnostic, machine learning, radiomics, primary ciliary dyskinesia, systemic sclerosis, chronic lung allograft dysfunction

REMERCIEMENTS

Beaucoup de temps s'est écoulé depuis que j'ai commencé mon programme de doctorat, et les choses ont radicalement changé. J'ai eu le privilège de rencontrer, de parler et d'échanger des idées avec de nombreuses personnes, à la fois directement et indirectement, ce qui a fait de moi une meilleure personne. Quelques paragraphes ne me suffisent pas pour exprimer toute la gratitude et les émotions que je ressens, mais je ferai de mon mieux pour les mettre en mots.

Tout d'abord, je tiens à remercier mon superviseur, **Professeur Anh-Tuan DINH-XUAN**. Cela fait 10 ans que je suis venu le rencontrer pour la première fois au cabinet de Cochin pour postuler en Master Physiologie, et depuis il ne m'a jamais rien refusé. Il est toujours celui qui anticipe et qui me montre la voie, m'enseignant des valeurs dans la vie. Je tiens à lui remercier pour son accompagnement, ses conseils et son consolation dans les moments difficiles.

Je remercie le **Professeur Marie-Pierre REVEL** qui est mon tuteur depuis 2015. Je me souviens encore du premier jour où nous nous sommes rencontrés, elle m'a fait visiter le département avec la joie dans les yeux. Merci pour ses câlins de confiance français (avant COVID-19 !). Bien qu'elle a souffert de mes pauvres compétences en présentation et de mes mauvais français et anglais, elle n'a pas refusé lorsque j'ai demandé à retourner son service de radiologie pour poursuivre mes études doctorales. J'aurais aimé qu'elle puisse imaginer combien de problèmes je lui causerais au cours des prochaines années avant de prendre sa décision. Cela aurait été mieux pour elle, mais j'ai eu de la chance car elle n'a pas eu cette vision. En conséquence, je suis devenu son élève, pour une fois et pour toujours^.

Je remercie **Dr Guillaume CHASSAGNON**, difficile d'imaginer comment ce travail pourrait vu le jour sans ses instructions. Je profite de ces quelques mots pour témoigner de mon admiration pour son travail acharné et son dévouement en tant que médecin et chercheur reconnue. En effet, c'est un médecin modèle de la nouvelle ère de la France, que moi, du côté vietnamien, je ne peux qu'admirer et suivre.

Merci au **Dr Huy Thong HUA**, pour votre patience et votre aide pour améliorer mes premiers écrits. Merci d'avoir partagé les repas dans la cafétéria animée de l'hôpital !

Aux membres de mon jury de thèse,

- Je tiens à remercier Monsieur le **Dr. Colas TCHERAKIAN**, pour son précieux révision de ma thèse d'avoir pris du temps sur vos horaires chargés pour corriger cette thèse et d'y apporter des suggestions indispensables pour l'améliorer.

- Je remercie également Madame le **Pr Yurdagül UZUNHAN** pour le temps consacré à la lecture de mon travail. C'est pour moi un grand honneur que de vous avoir dans mon jury.

- Je tiens également à remercier **Didier FRECHEVILLE**, **Corinne DUTHEIL**, secrétaires du Professeur DINH-XUAN, et **Isabelle PARRAIN**, secrétaire du Professeur REVEL, Madame **Elisabeth HOMBRADOS**, gestionnaire de l'ED 563-MTCI pour toute l'aide qu'ils m'ont apportée. Ils ont veillé à ce que les diverses procédures universitaires concernant mon programme de doctorat. Ils ont remarquablement bien organisé nos réunions de recherches, les cours de DU, de conférences hors l'hôpital et ont rapidement réglé tous les problèmes. Leurs sourires et leur hospitalité me font me sentir bien à chaque fois que je suis venue les voir.

- A l'équipe des Pneumologues et Radiologues de l'**hôpital Cochin**, et des Ingénieurs du laboratoire CentraleSupélec, pas de recherche en médecine si d'autres ne sont pas sur le pont à couvrir le service. Je mesure la chance d'avoir le soutien indéfectible à tous du **Dr Véronique BOUSSAUD**, du **Pr Pierre-Regis BURGEL**, du **Dr Nicolas CARLIER**, du **Dr Aëlle Le Gall**, du **Dr Alexis Régent**, du **Charlotte Martin**, du **Dr Gaël FRECHE**, du **Dr Séverine DANGÉARD**, du **Dr Souhail BENNANI**, du **Dr Claude GUINET**, de **Souad SISMAIL**, d'Enzo BATTISTELLA, de **Maria VAKALOPOULOU** et du **Pr Nikos PARAGIOS** !

- A mes collègues des équipes médicales et paramédicales de l'**hôpital Vinmec Central Park**, pour ces années passées et à venir de à travailler dans la bonne humeur au service des patients. En particulier, je voudrais remercier **Dr Hai Đang TRAN**, **Dr Su Quoc LE**, **Dr Thi Ngoc Hieu DO**, **Dr Manh Cuong DANG**, **Dr Cong Trinh TRAN**, **Dr Tra My TON NU**, **Dr Huu Hanh Nhi LE**, **Dr Thi Nga My LE** et les secrétaires médicales **Thi Phuong Thao LE**, **Thi Ngan Ha LE**, les manipulateurs **Ngoc Thanh PHAN**, **Thi Thanh Lien NGUYEN** pour l'ambiance qu'ils ont su créer et maintenir sein du Service d'Imagerie Diagnostique.

- Merci aux amis à mon côté derrière ce travail : **Dr Nhat Nam LE DONG**, **Ngoc HAI VU** ; les amis du groupe « Tarot » : **Phuc Hoang TA**, **Ngoc Bao DUONG**, **Thi Bich Hong NGUYEN**, **Linh Lan VO** ; et en particulier la famille de **Madame Huguette** et **Monsieur Jean-Pierre**. Le temps passé avec vous était si agréable.

- A la famille **DINH-XUAN** pour m'avoir si bien accueilli parmi vous depuis plus de 3 ans, pour votre gentillesse et votre délicatesse préparation pour des déjeuners complètes et chaleureuses du dimanche. Je garderai également les bons souvenirs que nous avons partagés lors de nos voyages de vacances.

- A la famille **TRAN** pour m'avoir parrainée depuis 10 ans, pour votre soutien, tous nos fous rires à la maison qui m'ont fait regagner confiance dans les moments difficiles de ma thèse.

- Merci à **VT Bui** de m'avoir montré le chemin et de m'avoir aidé avec persévérance à surmonter les difficultés dans le processus de rédaction de ma thèse.

- Surtout, merci à ma famille pour leur amour constant et leur soutien dans tous mes efforts. À ma famille, mon père **Tuong HOANG** et ma mère **Thi Hoa LUONG**, j'espère que vous êtes fier de moi. Je ne vais pas oublier mes deux frères, **Manh Chinh** et **Gia Cat**, ma sœur **Que Tran** qui ont pardonné les heures passés avec un ordinateur.

- Enfin, je terminerai par **DIEU**, merci pour sa beauté et la leçon à retenir après tous : « Car nous marchons par la foi, et non par la vue. » (2 Corinthiens 5 :7).

TABLE DES MATIERES

REMERCIEMENTS	4
LISTE DES FIGURES	9
LISTE DES TABLEAUX	10
LISTE DES PRINCIPALES ABREVIATIONS	11
1. INTRODUCTION	13
1.1 PATHOLOGIE ET MANIFESTATION EN IMAGERIE DE MALADIE PULMONAIRE ET L'IMPLICATION DE LA QUANTIFICATION DES LESIONS.....	15
1.1.1 PATHOLOGIE ET MANIFESTATION EN IMAGERIE DE LA DYSKINESIE CILIAIRE PRIMITIVE	15
1.1.2 PATHOLOGIE ET MANIFESTATION EN IMAGERIE DE LA DYSFONCTION CHRONIQUE DU GREFFON PULMONAIRE	16
1.1.3 PATHOLOGIE ET MANIFESTATION EN IMAGERIE DE LA PNEUMOPATHIE INTERSTITIELLE DIFFUSE DE LA SCLERODERMIE.....	20
1.1.4 ANALYSES QUANTITATIVES DES ATTEINTES DES VOIES AERIENNES EN IMAGERIE THORACIQUE EN RELATION AVEC LES TROUBLES FONCTIONNELS RESPIRATOIRES.....	22
1.1.5 ANALYSES QUANTITATIVES D'ATTEINTE DE PARENCHYME PULMONAIRE EN IMAGERIE EN RELATION AVEC LES TROUBLES FONCTIONNELS RESPIRATOIRES	23
1.2 INTRODUCTION A L'INTELLIGENCE ARTIFICIELLE	24
1.2.1 HISTOIRE DE L'INTELLIGENCE ARTIFICIELLE.....	25
1.2.2 APPRENTISSAGE AUTOMATIQUE	27
1.2.3 RÉSEAUX DE NEURONES ARTIFICIELS	28
1.2.4 APPRENTISSAGE PROFOND	29
1.2.5 AVANTAGES ET INCONVENIENCES DE L'INTELLIGENCE ARTIFICIELLE EN IMAGERIE THORACIQUE	33
1.3 CONTRIBUTIONS ET ORGANISATION DE THESE.....	36
1.3.1 PRINCIPAUX DÉFIS	36
1.3.2 CONTRIBUTIONS.....	37
1.3.3 ORGANISATION DE LA THESE	37
2 MATERIELS ET METHODES	40
2.1 DE BIG DATA A L'AGE DES RECHERCHES AXÉES SUR LES DONNÉES.....	40
2.1.1 BIG DATA	40
2.1.2 BIOMARQUEURS.....	42
2.1.3 SYSTEME D'AIDE AU DIAGNOSTIC PAR ORDINATEUR.....	42
2.1.4 LA SCIENCE AXÉE SUR LES DONNÉES.....	43
2.1.5 RADIOMIQUE.....	43
2.2 LES ALGORITHMES D'APPRENTISSAGE AUTOMATIQUE.....	46

2.2.1	TYPE DE L'ALGORITHME	46
2.2.2	BASE DE DONNÉES.....	47
2.2.3	SUR-APPRENTISSAGE ET SOUS-APPRENTISSAGE DES MODÈLES.....	47
2.2.4	STRATÉGIE D'OPTIMISATION	48
2.3	RADIOMIQUE EN PNEUMOLOGIE	49
2.3.1	DATA SELECTION ET STANDARDISATION DES PROTOCOLES D'IMAGERIE	49
2.3.2	ANNOTATION DES DONNÉES EN IMAGERIE THORACIQUE	50
2.3.3	PRÉ-TRAITEMENT DES DONNÉES	55
2.3.4	EXTRACTION DES PARAMETRES D'IMAGERIE.....	55
2.3.5	SELECTION DES PARAMETRES D'IMAGERIE :.....	60
2.3.6	CLASSIFICATION.....	61
2.3.7	VALIDATION.....	72
3	BIOMARQUEUR QUANTIFIANT LES LESIONS PULMONAIRES DANS LA DYSKINÉSIE CILLIAIRE PRIMITIVE.....	73
3.1	CONTEXTE ET OBJECTIFS.....	73
3.2	PUBLICATION	74
3.2.1	RESUME	75
3.2.2	ARTICLE.....	75
4	PREDICTION DE LA DYSFONCTION CHRONIQUE DU GREFFON PULMONAIRE (DCGP).....	83
4.1	CONTEXTE ET OBJECTIFS.....	83
4.2	PUBLICATION	85
4.2.1	RESUME	86
4.2.2	ARTICLE.....	86
4.3	APPLICATION DE LA RADIOMIQUE POUR LA PREDICTION DE LA DYSFONCTION CHRONIQUE DU GREFFON PULMONAIRE	98
4.3.1	METHODE	98
4.3.2	IMPLEMENTATION ET RESULTATS EXPERIMENTAUX.....	99
4.4	DISCUSSION.....	108
4.5	CONCLUSION.....	110
5	CONTRIBUTION DE L'EVALUATION AUTOMATIQUE DE L'ETENDUE DE LA PNEUMOPATHIE INTERSTITIELLE DIFFUSE (PID) EN TOMODENSITOMETRIE EN PREDICTION DE LA MORTALITE DANS LA SCLERODERMIE SYSTEMIQUE.....	111
5.1	CONTEXTE ET OBJECTIFS.....	111
5.2	PUBLICATION	111
5.2.1	RESUME	112
5.2.2	MANUSCRIT D'ARTICLE.....	112

6	ÉTUDE DE L'IMPACT DU FILTRE DE RECONSTRUCTION SUR LE DEVELOPEMENT ET L'UTILISATION D'ALGORITHMES DE DEEP LEARNING POUR LA SEGMENTATION DES MALADIES PULMONAIRES DIFFUSES.....	133
6.1	CONTEXTE ET OBJECTIFS.....	133
6.2	PUBLICATION	133
6.2.1	RESUME	134
6.2.2	MANUSCRIT D'ARTICLE.....	134
7	DISCUSSION GENERALE.....	140
7.1	PRINCIPALES PROJETS DE THESES	140
7.2	APPLICATION DE L'INTELLIGENCE ARTIFICIELLE A BASE DES OBSERVATIONS EN TDM .	141
7.3	LA RESTRICTION DES DONNEES DES MALADIES RARE ET LES PROMESSES DU DEEP-LEARNING.....	143
7.4	CHOIX DU NOYAU DE RECONSTRUCTION POUR D'DEEP-LEARNING DANS LE CADRE DES LESIONS PARENCHYMATEUSES PULMONAIRES	143
7.5	LIMITATIONS ET PERSPECTIVES.....	144
8	CONCLUSION GENERALE	145
9	REFERENCE BIBLIOGRAPHIQUE	147

LISTE DES FIGURES

Figure 1-1: Principaux projets de Thèse.	14
Figure 1-2: Les changements structuraux tels que l'épaississement de la paroi bronchique, les impactions d'arbre en bourgeon et mucoïde, affectent principalement les régions basales de la DCP.	16
Figure 1-3: BOS chez un patient de 42 ans(37).	17
Figure 1-4: Femme de 59 ans atteinte d'un syndrome restrictif d'allogreffe 2 ans après une transplantation pulmonaire bilatérale(36).....	18
Figure 1-5: DCGP mixte chez un patient de 30 ans(37).....	19
Figure 1-6: Tomodensitométrie thoracique à haute résolution de la maladie pulmonaire interstitielle associée à la sclérose systémique (67).....	21
Figure 1-7: L'interrelation des quatre termes liés à l'IA en radiologie.	25
Figure 1-8: Deux booms de l'IA dans le passé et le troisième boom actuel de l'IA. Les progrès de la recherche CAO sont profondément liés aux progrès de la technologie de l'IA.	26
Figure 1-9: Illustration pour un neurone artificiel avec n d'entrées x_1, \dots, x_n , et son poids synaptique correspondant w_1, \dots, w_n en supposition d'avoir un décalage, appelé bias, pour le neurone.	29
Figure 1-10: Exemple de classification des tumeurs bénigne (B) et maligne (M) sur les radiographies thoraciques par un ConvNet [61].	32
Figure 1-11: Organisation de thèse.	38
Figure 2-1 Le flux de travail de radiomique.....	44
Figure 2-2: Vue d'ensemble des processus méthodologiques du service de santé d'apprentissage rapide et l'intégration de flux de travail radiomique(215).....	45
Figure 2-3: Organigramme illustrant le flux de travail de la radiomique et l'application du RQS (215).	46
Figure 2-4: Architecture U-net (exemple pour 32x32 pixels).....	52
Figure 2-5: Architecture du framework AtlasNet. ReLU = unité linéaire rectifiée (239).....	54
Figure 2-6: Illustration schématique des paramètres de la tomodensitométrie quantitative. 3D, en trois dimensions (258).....	56
Figure 2-7: Matrice de cooccurrence des niveaux de gris: Explication du processus de matrice de cooccurrence de niveau de gris. (a) Du niveau d'intensité de l'échelle de gris (à gauche) à la création de la matrice de cooccurrence (à droite) (b) différentes directions de la Matrice de cooccurrence des niveaux de gris	57
Figure 2-8: : Exemple de calcul de matrice de longueur des plages (j) de niveau de gris (i) pour une image 4x4 avec 4 niveaux de gris en quatre directions principales ($0^\circ, 45^\circ, 90^\circ, 135^\circ$)	58
Figure 2-9: : Exemple de calcul de matrice de taille des zones homogènes pour une image 4x4 avec 4 niveaux de gris.....	59
Figure 2-10: Exemple de calcul de matrice de différence des niveaux de gris de voisinage (S(j)) pour une image 4x4 avec 4 niveaux de gris (j).....	59
Figure 2-11: Exemple de calcul de matrice de la dépendance en niveaux de gris $Q(k, s)$ pour l'image 5x5 $M(i, j)$	60
Figure 2-12: Partition des données dans la méthode de « validation croisée » (275).....	73
Figure 4-1: L'organigramme de la base de données pour l'étude de prédiction de DCGP par radiomique.	100

LISTE DES TABLEAUX

Tableau 4-1: Définition des différents phénotypes de DCGP (6).	83
Tableau 4-2: Caractéristiques de DCGP et témoins (N=100).....	100
Tableau 4-3: Les caractéristiques de la population de DCGP et de témoins (n=73).	101
Tableau 4-4: Les caractéristiques de la cohorte d'entraînement/validation et de test (n=73).	101
Tableau 4-5: Première étape de sélection des variables radiomiques.	102
Tableau 4-6: Deuxième étape : la sélection des méthodes de classification.....	106
Tableau 4-7: Troisième étape de sélection des variables : classification par ablation itérative des variables.	107
Tableau 4-8: Performance des modèles en group del'entraînement et en test.....	108

LISTE DES PRINCIPALES ABREVIATIONS

AA : Apprentissage Automatique
ANN : Artificial Neural Network
AP : Apprentissage profond
AWT : Airway Wall Thickness
BO : Bronchiolite Oblitérante
BPCO : Broncho-Pneumopathie Chronique Obstructive
CAD : Computer Aided Diagnosis
CF : Cystic fibrosis
CLAD : Chronic Lung Allograft Dysfunction
ConvNet : Réseau Neuronal Convolutif
CPT : Capacité pulmonaire totale
CVF : Capacité vitale forcée
DAO : diagnostic assisté par ordinateur
DCGP : Dysfonctions chroniques du greffon pulmonaire
DCP : Dyskinésie ciliaire primitive
DLCO : Diffusion du Monoxyde de Carbone
EFR : Explorations fonctionnelles respiratoires
FeNO : La fraction d'oxyde nitrique expiré
FPI : Fibrose Pulmonaire Idiopathique
HU : Hounsfield Unit
IA : Intelligence Artificielle
IRM : Imagerie par Résonance Magnétique
LAA : Low atténuation area
LTx : Lung transplantation
mIP : projection d'intensité minimale
ML : Machine Learning
ORL : Oto-Rhino-Laryngologie
PEC : Photophérèse Extra-Corporelle
PIC : Pneumonie Interstitielle Commune
PID : Pneumonie Interstitielle Diffuse
PINS : Pneumonie Interstitielle Non Spécifique
PPFE : Pleuro-Parenchymal Fibro-Elastosis
QCT : Quantitative Computer Tomography
reLU : Rectified Linear Unit
RNA : réseau neuronal artificiel
ROI : Region of interest
RQS : Radiomics Quality Score
ScS : Sclérose Systémique
SRA : Syndrome Restrictif d'Allogreffe

SVM : Support Vector Machine

TDM : Tomodensitométrie

VEMS : Volume Expiratoire Maximal à la Première Seconde

VOI : Volume d'Intérêt

ZFD : Zone à Faible Densité

ZHD : Les zones à Haute Densité

1. INTRODUCTION

Les maladies pulmonaires, également connues sous le nom de maladies ou affections respiratoires, sont des maladies des voies respiratoires et d'autres structures des poumons. Selon le Forum de la Société Internationale de la Respiration (1), environ 334 millions de personnes souffrent d'asthme et, chaque année, la tuberculose tue 1,4 millions de personnes, 1,6 millions des personnes meurent d'un cancer du poumon, tandis que la pneumonie tue également des millions de personnes. Il est clair que les maladies pulmonaires sont l'une des principales causes de décès et d'invalidité dans le monde. La détection précoce joue un rôle clé pour augmenter les chances de guérison et améliorer la survie à long terme (2–4). Traditionnellement, la maladie respiratoire peut être explorée par des examens cliniques, des tests biologiques (cutanés, crachats, sanguins, ou anatomopathologiques), un test d'exploration fonctionnelle respiratoire (EFR), un examen radiographique pulmonaire et une tomodensitométrie (TDM)(5–7). Récemment, l'émergence de nouveaux progrès techniques de l'intelligence artificielle (IA) a montré un grand potentiel lorsqu'elle est appliquée sur des images médicales, y compris les maladies pulmonaires, en particulier sur sa capacité de synthétiser d'informations concernées multiples d'examen cliniques ou paracliniques pour but de détection, de quantification, de classification et aussi prédiction de survie à long terme.

Le domaine d'IA évolue rapidement, cependant, il est encore jeune dans son âge. Peu d'essais cliniques prospectives ont testé sur la technologie dans le monde réel. "Au moment, nous sommes longs sur la promesse et le potentiel et court sur la preuve et la validation », souligne Eric Topol (Scripps Research Translational Institute, La Jolla, CA, USA), au conseil du Service Nationale de Santé du Royaume-Uni en son approche vis-à-vis l'IA (8). Ainsi, les chercheurs ne peuvent être intéressés que par l'innovation technique sans viser l'application clinique et la validation de logiciels est moins intéressant pour les chercheurs en sciences fondamentales en raison de la quantité de procédures d'administration nécessaire pour obtenir l'autorisation d'organismes de réglementation pour un tel dispositif médical. La validation du logiciel est donc principalement effectuée par les entreprises commerciales, mais toujours uniquement s'il existe une viabilité commerciale claire pour ce produit. Toutefois, la nature de la médecine implique des connaissances physiologiques et physiopathologiques comme la pierre angulaire des algorithmes de gestion des problèmes de santé au service du patient. Dans le cadre de ce tableau, une collaboration plus étroite entre les universités et l'industrie peut être nécessaire, ainsi qu'entre les chercheurs techniques et les cliniciens. Pour ces types de collaborations, des études interdisciplinaires de l'université sont essentielles afin de combler cette lacune (9).

D'un point de vue médical, les applications de l'IA, prévues d'exploser d'ici quelques années, devraient être orientées par les connaissances de base en physiologie et mettre le bénéfice du patient en avant toute mise en œuvre de logiciel ou d'algorithme. C'était pour ces raisons que nous avons choisi une base d'argumentation physiologique qui tourne au tour des changements morphologiques et fonctionnels observés dans les affections respiratoires avant tout approche de l'IA dans les recherches scientifiques. D'une part, l'investigation des manifestations fonctionnelles résultant des, d'autre part les changements morphologiques observant en TDM thoracique. Nous espérons à partir des corrélations du couple « fonction – morphologie », pouvoir clarifier ou vérifier les voies physiologiques en hypothèse, en particulier dans les maladies rares dont le mécanisme pathologique reste encore obscur à présent : la dyskinésie ciliaire primitive (DCP), la sclérose systémique (ScS), les dysfonctions chroniques du greffon pulmonaire (DCGP ou DCGP - Chronic Lung Allograft Dysfunction). Pour la DCP, l'application de la caractéristique radiomique de premier ordre (histogramme) autorise la quantification automatique des changements en imagerie et l'établissement un score objectif pour l'évaluation de la gravité de la maladie. Pour la ScS,

l'algorithme de « l'apprentissage profond » (AP) permet la quantification des lésions du parenchyme pulmonaire à viser la prédiction de mortalité de long terme des patients. Pour les DCGPs, la synthèse des méthodes de quantification automatise des lésions de la première partie du travail demeure la base scientifique et propose la direction de recherche ultérieurs. Bien que les résultats préliminaires de l'implémentation de la radiomique et d'AP s'avèrent non statistiquement significative dans la prédiction de la survenue des manifestations fonctionnelles respiratoires par effet de sur-apprentissage, ces travaux ouvrent une nouvelle voie de recherche pour résoudre le problème de diagnostic précoce de cette complication qui est insidieuse au début mais provoque des lésions irréversibles ultérieures. Dans la dernière partie, nous rapportons ainsi l'influence du choix de filtres de reconstruction en TDM sur l'entraînement des modèles d'apprentissage automatique dans le processus de segmentation des zones lésionnelles de la maladies pulmonaires interstitielles de la ScS et COVID-19. La Figure 1-1 résume les principaux projets de la thèse.

PROJETS DE THESE

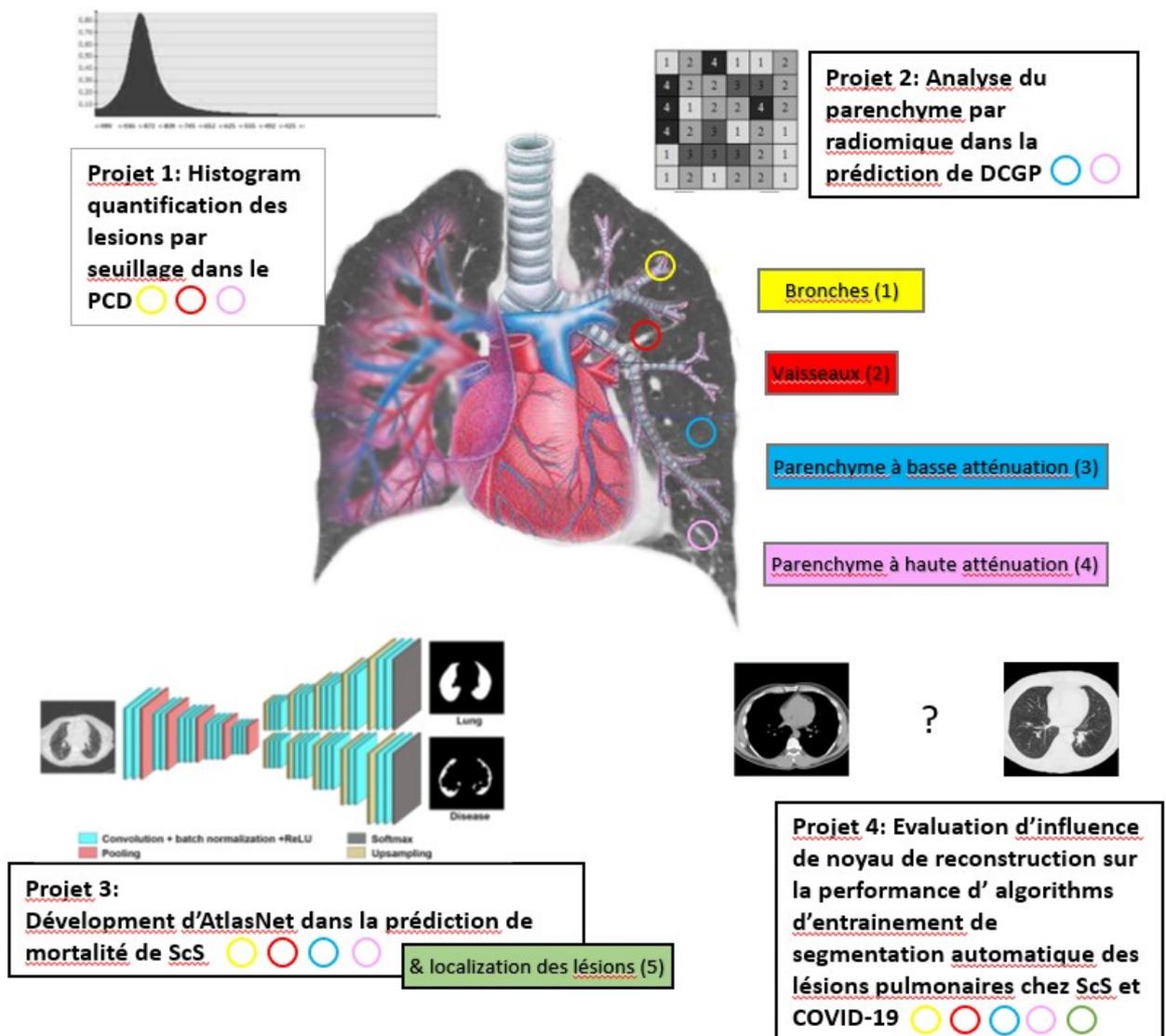


Figure 1-1: Principaux projets de Thèse.

1.1 PATHOLOGIE ET MANIFESTATION EN IMAGERIE DE MALADIE PULMONAIRE ET L'IMPLICATION DE LA QUANTIFICATION DES LESIONS

Le développement frénétique de la technologie d'imagerie médicale, tant en matériel qu'en logiciel, offre une puissance d'exception pour la TDM. Au cours des deux dernières décennies, la tomодensitométrie a été exploitée pour la caractérisation détaillée des structures du poumon et la description des maladies respiratoires. L'introduction de l'acquisition volumétrique a permis des analyses plus sophistiquées des données de TDM à travers des algorithmes informatisés, à savoir TDM quantitative (ou QCT - Quantitative Computer Tomography) (10). Ainsi, les images devenaient essentiellement des données numériques et les ordinateurs sont particulièrement bien adaptés à l'analyse de ce type de données. Des ordinateurs auraient la capacité de regarder au-delà de ce que l'œil humain peut voir et mettre à part de minuscules variations dans les images examinées, aider au diagnostic, détecter les premiers signes de la maladie voire prédire son évolution. Des centaines de milliers de TDM ont été analysés et plusieurs paramètres de QCT ont été développés et testés par rapport à des biomarqueurs cliniques, fonctionnels et pronostiques. Dans ce chapitre, nous résumons l'état de l'art de QCT dans l'analyse quantitative d'atteinte de voies aériennes et de parenchyme pulmonaire en général et ses applications dans les deux maladies DCP et DCGPs.

1.1.1 PATHOLOGIE ET MANIFESTATION EN IMAGERIE DE LA DYSKINESIE CILIAIRE PRIMITIVE

La dyskinésie ciliaire primitive (DCP) est une maladie autosomique récessive qui se caractérise par des défauts de la structure et de la fonction ciliaires, entraînant une clairance mucociliaire défectueuse et une maladie oto-sino-pulmonaire chronique. Sa prévalence est d'environ 1/15 000 dans la population caucasienne et comme les symptômes ne sont pas spécifiques, un diagnostic tardif est courant et le traitement est généralement instauré à un stade avancé de la maladie (11,12). Le pronostic est principalement lié à l'insuffisance respiratoire due à une obstruction des voies respiratoires par des sécrétions et à des infections ultérieures qui conduisent progressivement au développement d'une bronchectasie. À ce jour, la prise en charge recommandée des patients est de surveillance de l'appareil respiratoire et de l'appareil oto-rhino-laryngologie (ORL) tous les 6 à 12 mois (13,14).

Bien que des recherches approfondies aient été menées sur la DCP, la taille de la série d'imagerie rapportée est souvent petite et la prise en charge actuelle se limite principalement à un consensus d'experts extrapolé à partir de preuves de la mucoviscidose (15,16). Malgré la similitude de ces deux maladies, des rapports récents ont mis en évidence des différences significatives entre la DCP et la mucoviscidose en termes de physiopathologie, de paramètres cliniques, de résultats de la fonction pulmonaire et d'imagerie, en particulier celles de tomодensitométrie (TDM)[14, 17]. Contrairement à la mucoviscidose, dans laquelle les lobes supérieurs sont principalement affectés, les anomalies pulmonaires de la DCP sont principalement localisées dans les lobes moyens et inférieurs (Figure 1-2 **Error! Reference source not found.**).

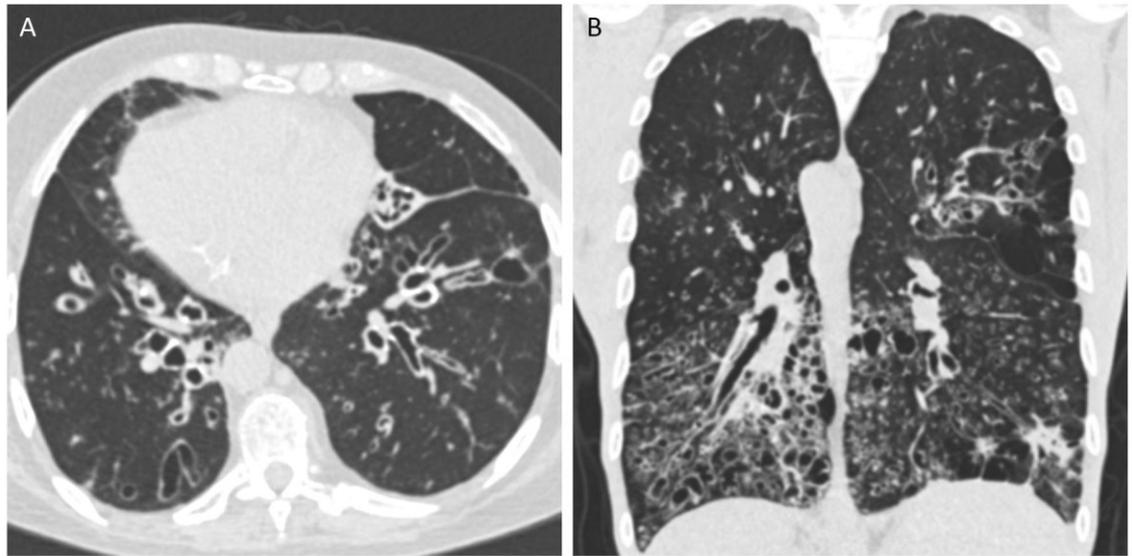


Figure 1-2: Les changements structurels tels que l'épaississement de la paroi bronchique, les impactions d'arbre en bourgeon et mucoïde, affectent principalement les régions basales de la DCP. Image TDM avec A : coupe axiale ; et B : coupe coronale.

Comme l'évaluation des changements fonctionnels et structurels est essentielle dans la prise en charge de la DCP, de nombreux chercheurs ont étudié la corrélation entre les deux entités avec diverses méthodologies telles que le suivi longitudinal ou des études cas-témoins, avec des résultats discordants. Un point commun entre ces investigations est que l'évaluation de la gravité de la TDM reste principalement descriptive ou semi-quantitative(18–21). Parmi les approches semi-quantitatives, la plus couramment utilisée est le score tomодensitométrique de Brody, qui a été principalement développé pour évaluer l'implication des voies respiratoires dans la mucoviscidose(21,22). L'utilisation de ce score souffre de plusieurs limitations. D'abord, elle nécessite une formation dédiée, elle coûte du temps et surtout, elle s'appuie sur une évaluation subjective des anomalies du scanner(22,23).

Une récente étude sur la tomодensitométrie thoracique dans la dyskinésie ciliaire primaire évalue des rapports de la diamètre des conduits aériens et la dimension des artères adjacentes (24). Leurs résultats étaient surprenants en disant que chez les enfants atteints de DCP, les voies respiratoires sont plus dilatées que chez les témoins et ne présentent pas d'épaississement de la paroi bronchique. Néanmoins, les observations préliminaires au TDM de la maladie de DCP, les principales anomalies comprennent la bronchectasie, l'épaississement des parois bronchiques, la présence d'impactions mucoïdes, de condensation et d'atélectasie pulmonaire(25) et Figure 1-2, qui sont toutes susceptibles d'augmenter la densité des tissus pulmonaires et donc de modifier les paramètres de densité de l'image.. La discordance des résultats de recherche montre que le mécanisme pathologique de cette maladie rare reste encore obscur et a besoin d'être exploré avec des études sur de plus grandes populations et l'utilisation d'outils d'investigation quantitative.

1.1.2 PATHOLOGIE ET MANIFESTATION EN IMAGERIE DE LA DYSFONCTION CHRONIQUE DU GREFFON PULMONAIRE

Le terme « Dysfonction Chronique du Greffon Pulmonaire » (DCGP ou CLAD : Chronic Lung Allograft Dysfunction) , introduit pour la première fois par Glanville, a été proposé comme terme général pour décrire les manifestations cliniques d'une gamme de processus pathologiques dans les voies respiratoires et les compartiments parenchymateux de l'allogreffe pulmonaire survenant plus de 3 mois après l'opération de transplantation qui conduisent à une détérioration de la fonction pulmonaire significative et persistante (26,27). La DCGP est la principale complication à long terme

après transplantation due à un rejet chronique entraînant un taux de mortalité de 60% après 5 ans chez les patients transplantés pulmonaires (LTx) (28). Comme l'allo-réactivité sous-clinique peut sérieusement compromettre l'intégrité de l'allogreffe, il est nécessaire de développer des tests physiologiques et d'imagerie capables de détecter une lésion précoce des voies respiratoires. D'après la définition clinique, la DCGP correspond à une baisse substantielle et persistante de la valeur mesurée du VEMS plus de 20 % par rapport à la valeur de référence. Cette dernière est calculée comme la moyenne des deux meilleures mesures du VEMS postopératoires (prises à plus de 3 semaines d'intervalle)(29,30). La classification phénotypique de DCGP la plus récente mise à jour par le groupe de Verleden et coll. inclut le syndrome obstructif (ou BO - bronchiolite oblitérante), qui représente 70% des dysfonctionnements d'allogreffe, les 30 % d'entités restantes étant le syndrome restrictif d'allogreffe (SRA), les phénotypes mixtes et indéfinis sur le modèle de déclinaison des EFRs et la manifestation en TDM (opacités) [77–79 et Tableau 4-1:Définition des différents phénotypes de DCGP (6)]. L'évolution du patient avec SRA est pire que celle du patient avec BO, avec une survie médiane de 6-18 mois après le diagnostic de SRA comparée à 3-5 ans après le diagnostic de BO (34,35).

Les opacités suggérées par les auteurs se réfèrent à des opacités parenchymateuses et/ou à un épaississement pleural compatible avec un diagnostic de fibrose pulmonaire et/ou pleurale et susceptibles de provoquer une physiologie restrictive. Depuis, compléter la mesure des EFRs (VEMS et CPT) avec TDM est obligatoire non seulement pour exclure d'autres causes de dégradation de la fonction pulmonaire mais aussi pour classer les phénotypes de DCGP (33,36). Brun et coll. ont mis à jour les caractéristiques TDM associées aux différents phénotypes de DCGP tels qu'ils ont été définis dans les récentes directives internationales (37).

Les caractéristiques tomodynamométriques de la BO comprennent l'atténuation mosaïque du parenchyme pulmonaire et le piégeage de l'air expiré, parfois associés à de grandes anomalies des voies respiratoires. L'atténuation mosaïque et le piégeage d'air sont des conséquences directes de la fibrose des petites voies aériennes. L'hypoventilation due à la bronchiolite oblitérante est responsable d'une vasoconstriction hypoxique et se manifeste par des zones de faible atténuation avec de petites et rares marques vasculaires sur le scanner. La tomodynamométrie expiratoire peut être utilisée pour différencier le piégeage d'air d'autres causes d'atténuation de la mosaïque, telles que les maladies pulmonaires vasculaires occlusives et infiltrantes (38–40) Figure 1-3. De grandes anomalies des voies aériennes, y compris un épaississement de la paroi bronchique et une légère bronchectasie cylindrique, sont souvent présentes (36). De petits nodules centro-lobulaires et un arbre en bourgeon peuvent également être observés, liés à une impaction bronchiolaire et/ou à un épaississement de la paroi bronchique.

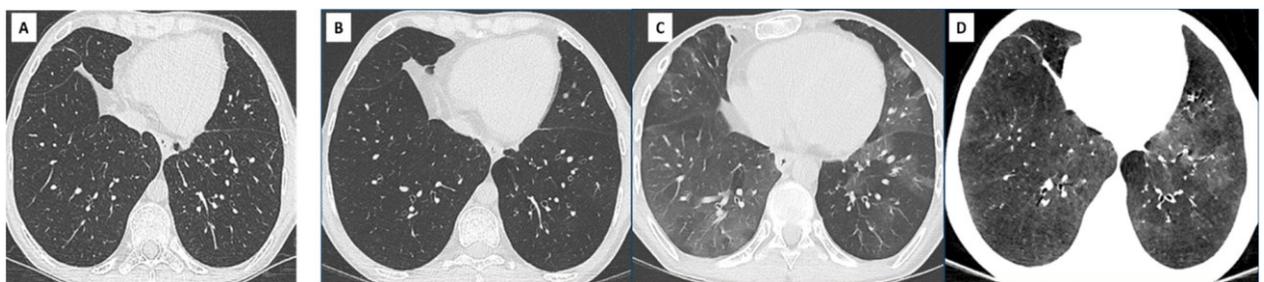


Figure 1-3: BOS chez un patient de 42 ans(37).

Figure 1-3 : TDM au diagnostic DCGP (A) et 2 ans plus tard (B–D) ; (A) Août 2018 : DCGP est diagnostiqué sur la spirométrie avec un profil obstructif. Le scanner ne montre pas d'anomalie significative, confirmant le phénotype BOS ; (B) Juin 2020 : tomodynamométrie inspiratoire. Bien que la densité du parenchyme pulmonaire semble assez homogène, l'hypoperfusion globale est évidente lorsque l'on compare ce scanner au scanner de base. Les vaisseaux sont plus petits et la

densité globale du parenchyme pulmonaire est plus faible. Il n'y a pas d'épaississement significatif de la paroi bronchique ni d'anomalie importante des voies aériennes ; (C) juin 2020 : l'analyse expiratoire révèle une atténuation mosaïque et un piégeage d'air hétérogène liés au BOS ; (D) Juin 2020 : la technique de projection d'intensité minimale (mIP) appliquée sur les images inspiratoires.

Les caractéristiques d'imagerie du SRA dépendent du stade auquel la maladie est découverte(41–43) . Au cours des stades précoces et des exacerbations aiguës, une opacification et une consolidation en verre dépoli sont souvent observées (41). Après résolution de ces lésions, une fibrose s'ensuit généralement, entraînant une bronchectasie par traction et de réticulation (41). Dans les stades ultérieurs, les patients ont des zones confluentes progressives de consolidation périphérique représentant une fibrose pleuro-parenchymateuse (44,45). Une nette démarcation du parenchyme pulmonaire normal et anormal est souvent observée. Les caractéristiques secondaires comprennent une perte de volume entraînant une rétraction hilare et une bronchectasie par traction et une bronchiolectasie.

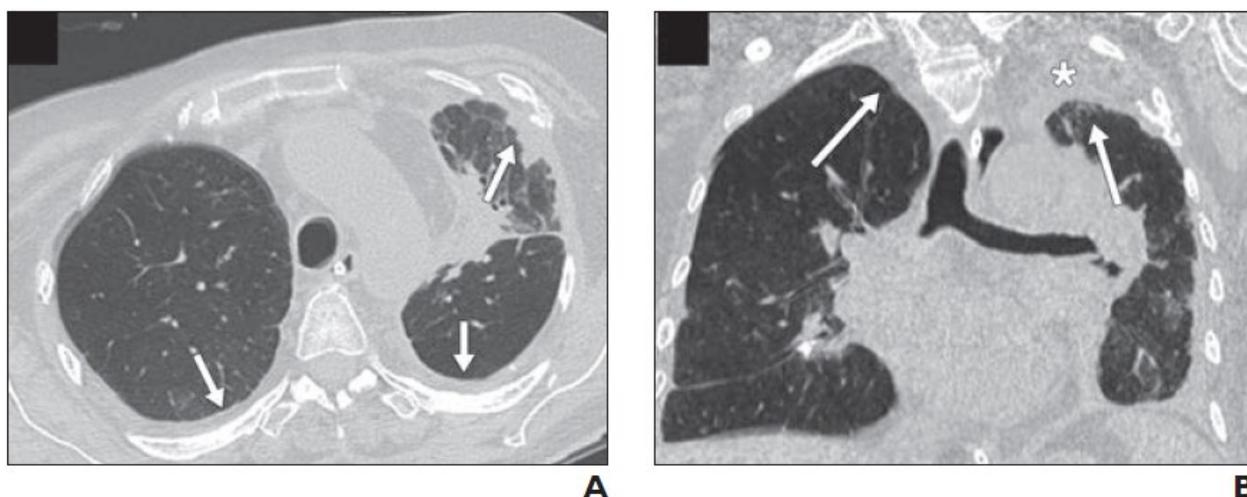


Figure 1-4: Femme de 59 ans atteinte d'un syndrome restrictive d'allogreffe 2 ans après une transplantation pulmonaire bilatérale(36).

Figure 1-4 : A. L'image tomодensitométrique axiale non rehaussée dans la fenêtre pulmonaire au niveau des apex montre la zone pulmonaire supérieure – une réticulation sous-apicale et une fibrose prédominante à gauche plus qu'à droite (flèches). B. Les images tomодensitométriques reformatées coronales montrent une perte de volume asymétrique marquée du poumon gauche avec une fibrose confluyente asymétrique (flèches) principalement à l'apex du poumon gauche (astérisque).

La combinaison d'obstruction, de restriction et d'opacités pulmonaires persistantes est généralement définie comme le phénotype mixte de DCGP (46). Ce phénotype mixte était associé à une prédominance apicale des opacités en TDM, à une fibroelastose pleuro-parenchymateuse (PPFE) en pathologie, à un VEMS et VEMS/CVF inférieurs aux valeurs théoriques par rapport au SRA ab initio, et à une survie similaire par rapport aux patients RAS ab initio.

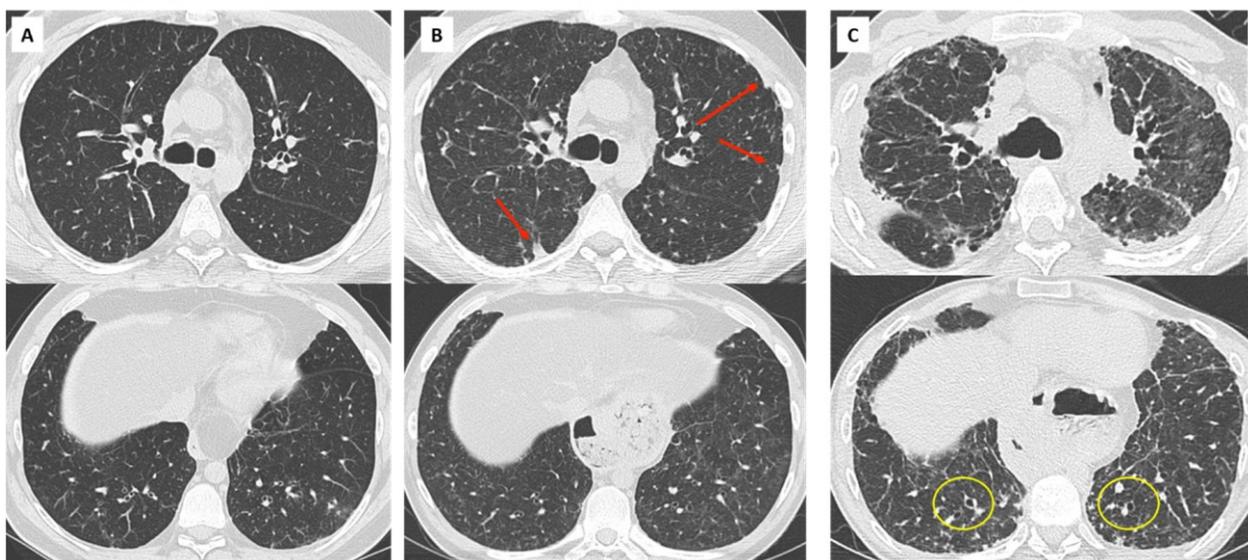


Figure 1-5: DCGP mixte chez un patient de 30 ans(37).

Figure 1-5 : (A) Diagnostic de DCGP en 2007 avec phénotype obstructif initial. Le scanner montre un léger épaissement de la paroi bronchique ; (B) déclin progressif de la CPT, suggérant une évolution vers un DCGP mixte (oct 2009). A noter l'apparition d'opacités sous-pleurales dans les lobes supérieurs (flèches rouges) et d'opacités en verre dépoli légères et diffuses ; (C) détérioration fonctionnelle et radiologique rapide, avec fibrose prédominante supérieure et consolidation sous-pleurale évoquant une PPF, bronchiectasie cylindrique légère dans les régions basales (cercles jaunes).

L'identification des différents phénotypes de DCGP peut être une première étape vers une thérapie individualisée. Idéalement, le DCGP peut être prévenu par un traitement prophylactique, bien que le traitement immunosuppresseur classique ne semble pas empêcher le développement du DCGP(47). D'autres traitements prophylactiques ont été étudiés récemment, l'azithromycine étant le seul médicament qui augmente la survie à long terme du greffon sans DCGP(48). Un essai randomisé contrôlé par placebo avec de la vitamine D (100 000 UI, une fois par mois) a conclu que cet agent n'a pas été efficace pour prévenir le DCGP, alors qu'il n'y avait pas non plus de différence dans l'un des critères d'évaluation secondaires tels que le rejet aigu, la fonction pulmonaire en test d'explorations fonctionnelles respiratoires (EFRs) et la distance minimale en test de marche de 6 minutes (49). Certains agents thérapeutiques ont montré un certain bénéfice pour ralentir la progression de la maladie dans le DCGP établi, et plus particulièrement dans la BO, comme l'azithromycine (50). Il a été démontré que le Montélukast, un antagoniste des récepteurs des leucotriènes, ralentit la progression de la maladie dans un essai prospectif randomisé contrôlé par placebo, principalement chez des patients atteints d'un BO d'apparition tardive et de bas grade (51). Ce résultat a été confirmé dans une vaste étude rétrospective de 153 patients DCGP, chez lesquels une atténuation de la baisse du VEMS a été observée dans un sous-ensemble de patients (52). Il existe de plus en plus de preuves de l'utilisation de la photophérèse extracorporelle (PEC), bien que les études prospectives et randomisées restent absentes. Le début de la thérapie PEC chez les patients BO a atténué la pente de la baisse du VEMS (53). Semblable à celui du Montélukast, le traitement par PEC à un stade précoce du BO conduit à un meilleur résultat (54). De plus, l'arrêt de la PEC a entraîné une baisse significative de la fonction pulmonaire et une augmentation de la mortalité des patients (55). Par conséquent, l'PEC semble être une thérapie prometteuse pour le DCGP (BO en particulier) ; Cependant, il n'est pas universellement disponible. Une approche plus expérimentale est l'administration de cellules souches mésenchymateuses, ce qui est faisable, sûr

et tolérable chez les patients atteints de BO (56) et peut ralentir la vitesse de déclin de la fonction pulmonaire (57).

Compte tenu du mauvais pronostic de la SRA, des stratégies de traitement pour prévenir la SRA ou ralentir sa progression sont désespérément nécessaires ; Cependant, les rapports de la littérature sont plutôt rares. Une série de cas récente de 11 patients a démontré que la pirféridone, un agent antifibrotique, est sans danger pour les patients transplantés pulmonaires et peut atténuer le déclin incessant de la fonction pulmonaire dans le SRA (58). Un autre rapport de cas d'un patient atteint de SRA a indiqué le bénéfice potentiel du traitement au nintédanib, un autre agent antifibrosant (59). La re-transplantation pulmonaire peut être un dernier recours pour traiter la DCGP en phase terminale et certains centres, pour la plupart très expérimentés, sont en mesure de montrer une survie similaire chez les patients ayant subi une re-transplantation pulmonaire pour la DCGP par rapport aux transplantations primaires (60,61), bien que l'évolution postopératoire soit beaucoup plus difficile et que la fonction pulmonaire maximal plus faible soit atteinte (62).

Bref, aucun traitement ne soit actuellement capable d'inverser le DCGP après le diagnostic, l'identification précoce du DCGP permettrait des stratégies proactives et ciblées pour ralentir la progression de la maladie avant que des dommages irréversibles de l'allogreffe ne se produisent(63). Les nouvelles directives ISHLT par rapport de la DCGP (33) peuvent fournir un cadre correct pour une dénomination et des définitions uniformes et, peuvent être un bon point de départ pour le lancement de grandes études de cohorte prospectives multicentriques, qui sont désespérément nécessaires sur le terrain. Dans ce contexte, il est raisonnable de choisir des tests physiologiques et d'imagerie permettant une détection précoce du DCGP, avec un rôle potentiel pour l'imagerie quantitative pour servir de biomarqueur, permettant l'identification précoce du DCGP (63).

1.1.3 PATHOLOGIE ET MANIFESTATION EN IMAGERIE DE LA PNEUMOPATHIE INTERSTITIELLE DIFFUSE DE LA SCLERODERMIE

La sclérodermie (ScS) est une maladie du tissu conjonctif caractérisée principalement par l'épaississement et le durcissement de la peau. La forme combinée « scléro » signifie « dur » en grec, et le mot « derme » signifie la peau. Il existe deux formes principales de sclérodermie : localisée et systémique (également appelée sclérodermie systémique). Dans la sclérodermie localisée, la maladie affecte principalement la peau et peut avoir un impact sur les muscles et le squelette. Dans la sclérodermie systémique, il existe une atteinte des organes internes, tels que le tube digestif, le cœur, les poumons et les reins, entre autres (64). Sa prévalence varie selon l'appartenance ethnique, le sexe et la zone géographique. En Europe, l'incidence estimée de la sclérose systémique était de 100 à 260 par million(65). L'atteinte pulmonaire est retrouvée à l'autopsie dans 70 à 100% des cas et représente la principale cause de morbidité et de mortalité (66). Deux types d'atteinte pulmonaire sont observés : la pneumonie interstitielle due à une fibrose tissulaire, et moins fréquemment une hypertension pulmonaire liée à une microvasculopathie et / ou une fibrose pulmonaire avancée. La plupart des phénotypes de pneumonie interstitielle diffuse (PID) liées à la ScS (76%) sont représentés par une pneumonie interstitielle non spécifique (PINS) qui se manifeste généralement au scanner sous forme de réticulations, d'opacités, de verre dépoli et de bronchectasie par traction avec prédominance basale et périphérique (67). La pneumonie interstitielle commune (PIC) s'y est observé de façon moins fréquente(11%) (68) (Figure 1-6).

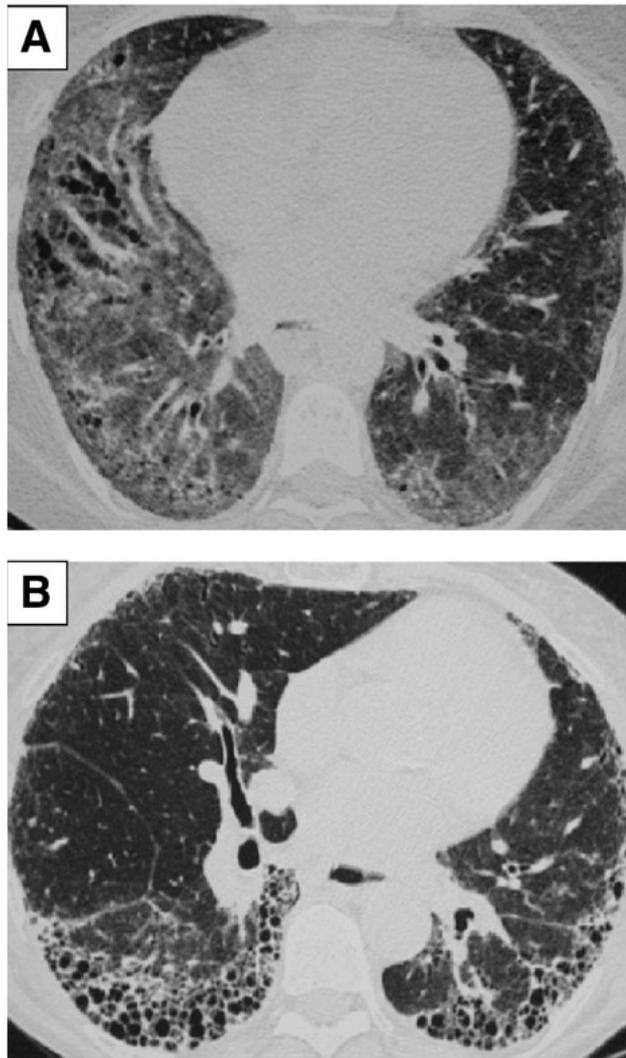


Figure 1-6: Tomodensitométrie thoracique à haute résolution de la maladie pulmonaire interstitielle associée à la sclérose systémique (67).

A. Verre dépoli avec bronchectasie par traction compatible avec la pneumonie interstitielle non spécifique. B. Cavités alvéolaires compatibles avec la pneumonie interstitielle habituelle

Pour la surveillance de la maladie, les indices de EFRs, en particulier la capacité vitale forcée sont les marqueurs prédominants. Selon les critères d'American Thoracic Society, de l'European Respiratory Society, de la Japanese Respiratory Society et de « Latin American Thoracic Association », dans le contexte de la fibrose pulmonaire idiopathique, la détérioration fonctionnelle est définie comme une baisse relative de 10% de la capacité vitale forcée absolue ou de 15% en capacité de diffusion absolue pour le monoxyde de carbone (DLco) (6,69,70). Cependant, certains auteurs considèrent les changements dans les EFRs comme un point final insensible (6) et suggèrent de combiner la mesure de la capacité vitale forcée avec une évaluation morphologique de la progression de la maladie comme la meilleure alternative dans les essais cliniques. Ainsi, dans le groupe européen de grand essai et de recherche sur la sclérodémie (« The European Scleroderma Trials and Research group »), la prévalence de la PID chez les patients atteints de ScS au scanner était de 53%, alors que la dyspnée n'était présente que dans 35% des cas(71). Néanmoins, dans le tableau de ScS, aucune recommandation claire n'existe concernant le rôle de la tomodensitométrie dans la surveillance de la PID. La tomodensitométrie est souvent réalisée dans des situations d'augmentation des symptômes pulmonaires ou de baisse des mesures de EFRs(72).

La pneumonie interstitielle non spécifique et la pneumonie interstitielle commune sont les principales formes de PID liée à la ScS, qui évolue d'un stade inflammatoire initial à la fibrose. Ainsi,

le suivi de la PID sur tomodensitométrie nécessite non seulement de comparer l'étendue de la maladie, mais également d'évaluer le rétrécissement pulmonaire lié au processus de fibrose. Jusqu'à nos jours, il existe de très peu d'évidence sur la valeur de l'évaluation des changements longitudinaux de la PID liée à la ScS au TDM, et toutes se sont concentrées sur l'étendue de la PID (73–75).

1.1.4 ANALYSES QUANTITATIVES DES ATTEINTES DES VOIES AERIENNES EN IMAGERIE THORACIQUE EN RELATION AVEC LES TROUBLES FONCTIONNELS RESPIRATOIRES

La voie aérienne est généralement mesurée en fonction de sa génération et la moyenne de la génération spécifique peut être calculée dans tout le poumon. Néanmoins, Gupta et coll. ont rapporté qu'une seule bronche pourrait être représentative(76). La caractérisation directe des voies respiratoires est faisable pour quelques générations bronchiques, ainsi, l'évaluation est largement acceptable jusqu'à la cinquième génération. La quantification directe des voies respiratoires comprend le diamètre et la surface de la lumière, l'épaisseur de la paroi, la surface absolue et la surface relative de la paroi (obtenue par le rapport entre la surface de la paroi et la surface totale de la bronche à une section donnée) et la densité de la paroi(77,78). Des voies aériennes plus petites peuvent être indirectement quantifiées au moyen d'une équation de régression linéaire entre la racine carrée de la surface totale et le périmètre interne d'une bronche « hypothétique » de 10 mm (environ 3 mm de diamètre intérieur), l'indice nommé « Airway wall thickness » (ou AWT-Pi10). AWT-Pi10 est considéré comme le plus cohérent parmi les paramètres dimensionnelles des voies respiratoires, bien que calculé uniquement sur un minimum de 12 bronchioles sous-segmentaires(79). De grands essais longitudinaux prospectifs tels que COPDGene et ECLIPSE ont inclus des milliers de patients avec et sans bronchopneumopathie chronique obstructive (BPCO) pour faire une analyse approfondie des tableaux cliniques et fonctionnels, de l'imagerie et de la génétique de ce syndrome (80,81). Ces essais majeurs représentent des exemples de l'investigation de QCT dans la BPCO et fournissent de grandes bases de données utiles à la pratique clinique. Les principaux résultats de ces études comprennent la progression régionale de la maladie(82,83), les paramètres de QCT associées significativement aux exacerbations de la BPCO(84). Un autre exemple, le groupe de Grenier et coll. a trouvé récemment chez les asthmatiques sévères, des changements morphologiques dans les grandes voies respiratoires, évalués quantitativement à l'aide de l'enregistrement 2D-3D des voies aériennes et de nouvelles algorithmes, caractérisés par l'épaississement de la paroi des voies respiratoires, un rétrécissement de la lumière et des sténoses bronchiques(85). Au-delà de la taille des composants bronchiques, la densité a été étudiée montrant une association potentielle avec l'infiltration des mastocytes chez des patients asthmatiques(86).

Cependant, la variabilité et les biais potentiels des mesures des voies respiratoires sont importants et comprennent le volume pulmonaire, l'âge et l'inflammation transitoire(87). Pour cette raison, l'application clinique du QCT des voies aériennes est encore à être validée.

D'un autre côté, parmi les variables physiologiques, la capacité vitale forcée (CVF) et le volume expiratoire forcé à la première seconde (VEMS) sont des paramètres spirométriques le plus souvent utilisés comme critère principal dans les essais cliniques sur les syndromes pulmonaires restrictif et obstructif, respectivement(78,88,89). Par ailleurs, la capacité de diffusion du monoxyde de carbone (DLCO) est également associée de façon significative avec le pourcentage de région de faible atténuation au TDM (Low atténuation area- LAA) et le AWT-Pi10 dans le groupe BPCO(90). La concentration fractionnaire expirée du monoxyde d'azote (FeNO) est aussi corrélée de façon significative à l'épaississement de la paroi bronchique chez les patients asthmatiques, notamment au niveau des bronches de sixième génération, la corrélation est indépendante de la surface

corporelle, l'âge, la durée de l'asthme, la dose de corticostéroïdes inhalés, l'éosinophilie et les IgE plasmatiques...(91).

1.1.5 ANALYSES QUANTITATIVES D'ATTEINTE DE PARENCHYME PULMONAIRE EN IMAGERIE EN RELATION AVEC LES TROUBLES FONCTIONNELS RESPIRATOIRES

Mesures de densité

Zone à faible densité. Les zones à faible densité (ZFD) sont un indice de la densité pulmonaire, définie comme le pourcentage de voxels du parenchyme en dessous d'un certain seuil d'atténuation au rayon- X. Bien que de nombreux seuils aient été étudiés, le seuil de -950 d'unité Hounsfield (HU) est le plus couramment utilisé (90,92). ZFD₉₅₀ a été validé par histologie et sert donc de nouveau marqueur de l'emphysème. Les évidences récentes suggérant que l'évaluation quantitative de l'emphysème est mieux corrélée avec l'extension morphologique que l'évaluation visuelle grâce à sa meilleure reproductibilité.

Zone de haute densité. Les zones à haute densité (ZHD) sont un indice de la densité pulmonaire, définie comme le pourcentage de voxels du parenchyme d'atténuation entre -600 et -250 HU. Il a été démontré que les ZHD sont associées au tabagisme (93) et à un risque accru d'atteintes interstitielles pulmonaires (94). Les ZHD a également été associé à des biomarqueurs de l'inflammation, du remodelage de la matrice extracellulaire, de la réduction de la fonction pulmonaire et d'un risque plus élevé de décès (95).

Récemment, la quantification automatique du parenchyme par seuillage de densité a été appliquée dans l'évaluation de la pneumonie due au COVID-19: l'emphysème pour un seuil de -1024 HU à -977 HU (96), le parenchyme sain résiduel de -977 HU à -703 HU (97,98), le verre dépoli de -703 HU à -368 HU et la consolidation de -100 HU à 5 HU (97,99). Le group de Roberto et coll. ont montré que la pneumonie du COVID-19 était reflétée avec une valeur médiane en pourcentage de verre dépoli de 19,50% et que le volume de verre dépoli seul présentait une différence significative entre les patients suspectés et ceux non suspectés pour l'infection au COVID-19.

Piégeage d'air. Le piégeage d'air représente le pourcentage de ZFD sur les images tomodynamométriques expiratoires, que l'on pense représenter une hyperinflation ou un « piégeage d'air », qui pourrait être calculé pour le poumon entier, pour un segment ou un sous-segment pulmonaire. Par définition, le piégeage d'air est la mesure de la densité pulmonaire sur les images TDM acquises à l'expiration complète, représente le pourcentage de voxels en dessous d'un seuil donné. Plusieurs seuils ont été étudiés, parmi lesquels, se trouvent le seuil de -856HU sur les images TDM en fin d'expiration (100), le -850 HU (101) et le -900 HU (102). Le seuil de -856 HU est le plus couramment utilisé. Les mesures de piégeage de l'air par TDM sont significativement corrélées avec les mesures de la fonction pulmonaire dans la BPCO (83,103) et l'asthme (100,102,104), et peuvent être utilisées pour évaluer la réponse au traitement (105).

Analyse d'histogramme

Les paramètres de l'histogramme de densité global des images TDM – tels que l'asymétrie, le kurtosis et la densité pulmonaire moyenne - sont utiles pour estimer l'étendue de la pneumopathie interstitielle diffuse - PID.(66,106,107) Par exemple, dans la fibrose pulmonaire, le dépôt de collagène augmente la densité pulmonaire, provoquant un déplacement vers la droite de la fréquence de l'histogramme et en réduisant son pic, autrement dit, en augmentant l'asymétrie et le kurtosis (108). En outre, ces paramètres sont suffisamment reproductibles et ne sont pas affectés par la réduction de la dose de rayonnement chez les sujets atteints de PID (109–111).

Best et coll. ont montré que le kurtosis et le score visuel de l'étendue du profil fibrosant comme les seuls indices de prédiction de la mortalité dans une étude rétrospective de 167 sujets atteints de fibrose pulmonaire idiopathique – FPI - recrutés dans un essai clinique. Récemment, dans une étude portant sur 46 sujets atteints de FPI, les paramètres de l'histogramme étaient corrélées à la EFRs et étaient associées à la survie sans greffe pulmonaire de la même manière que l'évaluation visuelle effectuée par deux experts(112). De même, les paramètres de l'histogramme peuvent faire la distinction entre différentes catégories de risque de mortalité bien définies chez les sujets atteints de PID liée à la sclérose systémique(113).

Analyse de texture

Actuellement, il existe un certain nombre de logiciels disponibles pour l'analyse de la texture du parenchyme pulmonaire, chacun d'eux étant capable d'extrapoler et d'évaluer différents groupes de paramètres de "Radiomics". La classification du parenchyme est appliquée à l'unité de volume du voxel en utilisant l'analyse texturale, la compréhension des paramètres de la cartographie d'histogramme basée sur la vision par l'ordinateur et la morphologie tridimensionnelle. L'analyse texturale est basée sur des régions d'intérêt (ROI – Region of Interest) sélectionnés dans le poumon par des experts entraînés, selon un ensemble de motifs spécifiques (normal, réticulation, rayon de miel, etc...). L'histogramme et les paramètres texturaux de chaque ROI volumétrique sont extraits, et un algorithme d'apprentissage automatique est utilisé pour développer un modèle prédictif pour des "textures" spécifiques(114–117). La majorité des travaux sur la détection des modèles de fibrose pulmonaire était basée sur la classification d'images 2D en utilisant une approche basée sur les patches texturales. Cette approche consiste à diviser le poumon en de nombreux petits patches de même taille (par exemple, 32 × 32 pixels) et à les classer dans l'une des classes de motifs de fibrose. Ces classificateurs sont ensuite formés sur des bases de données comprenant des milliers de patches annotés, des représentants de chaque classe à identifier du verre dépoli, du rayon miel, de l'emphysème. Le logiciel CALIPER a été développé en utilisant l'approche basée sur les patches, pour la quantification de l'étendue de la maladie et évolution de la fibrose pulmonaire idiopathique (118,119). Cependant, ce type de d'approche d'analyse est limité d'une part par une subjectivité intrinsèque (par exemple, due à la formation préalable des experts), d'autre part par le risque de classer à tort des parties centrales des bronches ou périphériques près de la paroi thoracique ou à l'interface entre deux classes d'anomalies pour ne pas inclure l'information informations sur l'environnement pulmonaire dans les patches (120,121). En revanche, il existe d'autres approches comme la segmentation de toute l'étendue fibreuse sans quantifier chaque composante (122). Cela nécessite de contourner les zones fibreuses anormales sur chaque tranche anormale, ce qui prend du temps mais permet ensuite d'appliquer le modèle à l'ensemble du poumon.

1.2 INTRODUCTION A L'INTELLIGENCE ARTIFICIELLE

L'intelligence artificielle (IA) peut être globalement définie comme la capacité d'un système informatique pour effectuer une tâche qui nécessiterait habituellement ou traditionnellement l'intelligence de l'être humain (123). Le domaine scientifique a été fondé sur l'hypothèse que « l'intelligence humaine peut être décrite avec une telle précision qu'une machine peut être faite pour la simuler » (124). Cette intelligence démontrée par des machines, contrairement à l'intelligence naturelle affichée par les humains et les animaux, qui implique la conscience et l'émotivité.

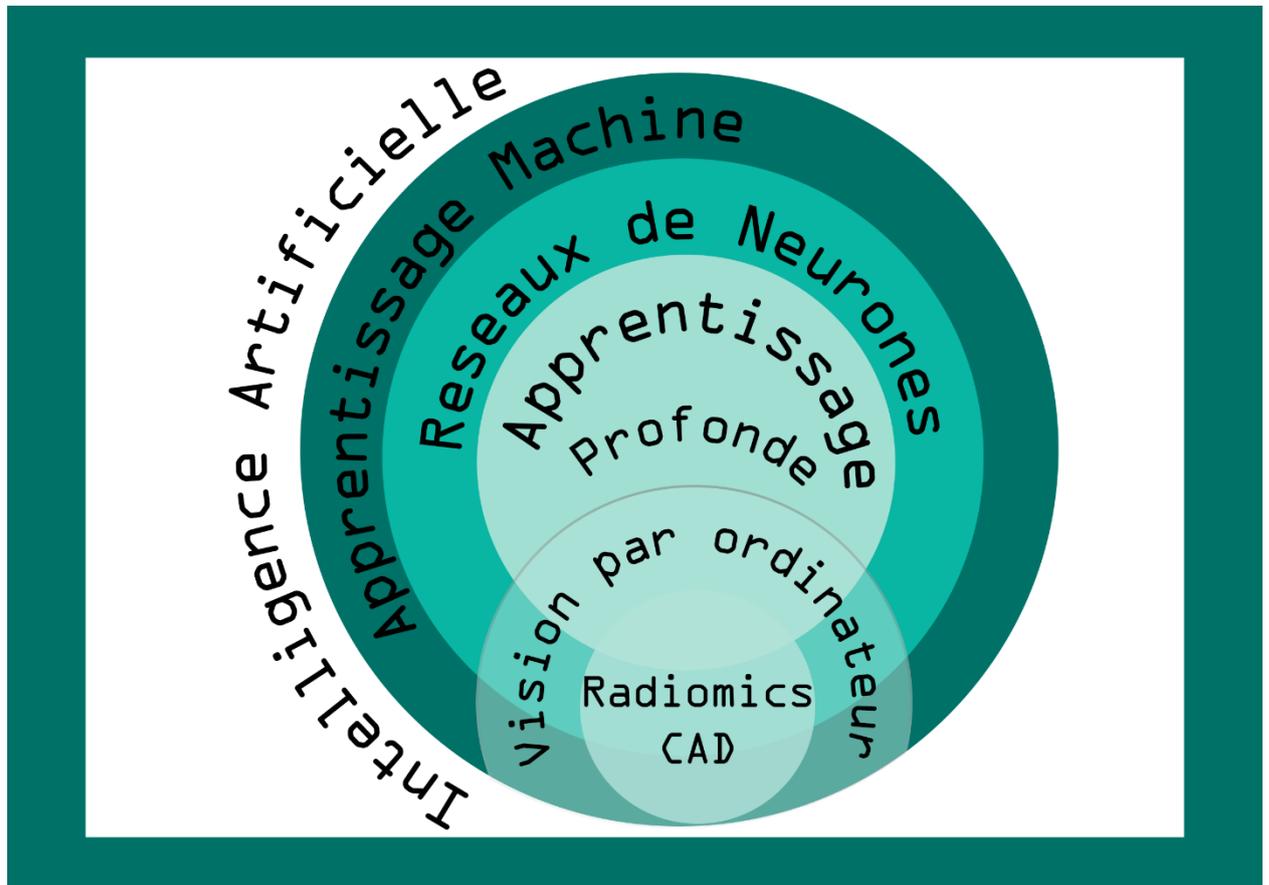


Figure 1-7: L'interrelation des quatre termes liés à l'IA en radiologie.

De nos jours, l'IA est un domaine de recherche qui programme les ordinateurs pour qu'ils aient l'intelligence. L'apprentissage automatique est l'une des techniques importantes de l'IA qui rend les machines (ordinateurs) plus intelligentes en leur offrant un apprentissage basé sur l'expérience. Un réseau neuronal artificiel (RNA) est une technologie d'apprentissage qui modélise le traitement de l'information dans le cerveau. L'apprentissage profond est à la tête du boom actuel de l'IA, entraîné par les méthodes d'apprentissage automatique, qui ont été développées en superposant des RNA conventionnelles et de nouvelles techniques d'apprentissage et en améliorant les performances de l'ordinateur. Si les méthodes d'IA imitent la manière dont les humains utilisent leurs organes sensoriels pour interpréter leur environnement, cela s'appelle « Perception de la machine ». Si ces sens sont visuels, on parle de « vision par ordinateur ». Dans le domaine de radiologie, l'objectif de cette vision est de déterminer les paramètres d'imagerie de complexité variable, invisibles à l'œil humain, afin d'établir des corrélations avec les résultats cliniques. Ce processus est intitulé « Radiomics » et ses produits appelées « outil de diagnostic assisté par ordinateur » (DAO) ou « Computer Aided Diagnosis » (CAD) (Figure 1-7).

1.2.1 HISTOIRE DE L'INTELLIGENCE ARTIFICIELLE

La conférence de Dartmouth en été 1956 était largement considérée comme l'événement fondateur de l'intelligence artificielle en tant que domaine scientifique (125). On attribue à la proposition de l'ouverture de cette réunion l'introduction du terme « l'intelligence artificielle » par John McCarthy 1955 (124) : « *Nous proposons qu'une étude de l'intelligence artificielle de 2 mois et 10 hommes soit réalisée au cours de l'été 1956 au Collège Dartmouth de Hanovre, New Hampshire. L'étude doit procéder sur la base de la conjecture que chaque aspect de l'apprentissage ou toute autre caractéristique de l'intelligence peut en principe être décrite si précisément qu'une machine peut être amenée à le simuler. Une tentative sera faite pour trouver comment faire en sorte que les machines utilisent le langage, forment des abstractions et des concepts, résolvent des types de*

problèmes désormais réservés aux humains et s'améliorent de façon automatique. Nous pensons qu'un progrès significatif peut être réalisé dans un ou plusieurs de ces problèmes si un groupe de scientifiques soigneusement sélectionnés y travaille ensemble pendant un été. » De telles manières, l'IA est née mais cette dernière n'était pas le fruit d'une seule journée, comme pour la majorité des événements historiques, c'était l'assimilation des initiatives précédentes d'une « machine qui pense ». Ainsi, le terme d'IA a été choisi en partie pour sa neutralité ; en évitant de se concentrer sur la théorie des automates étroits et en évitant la cybernétique qui était fortement axée sur la rétroaction analogique.

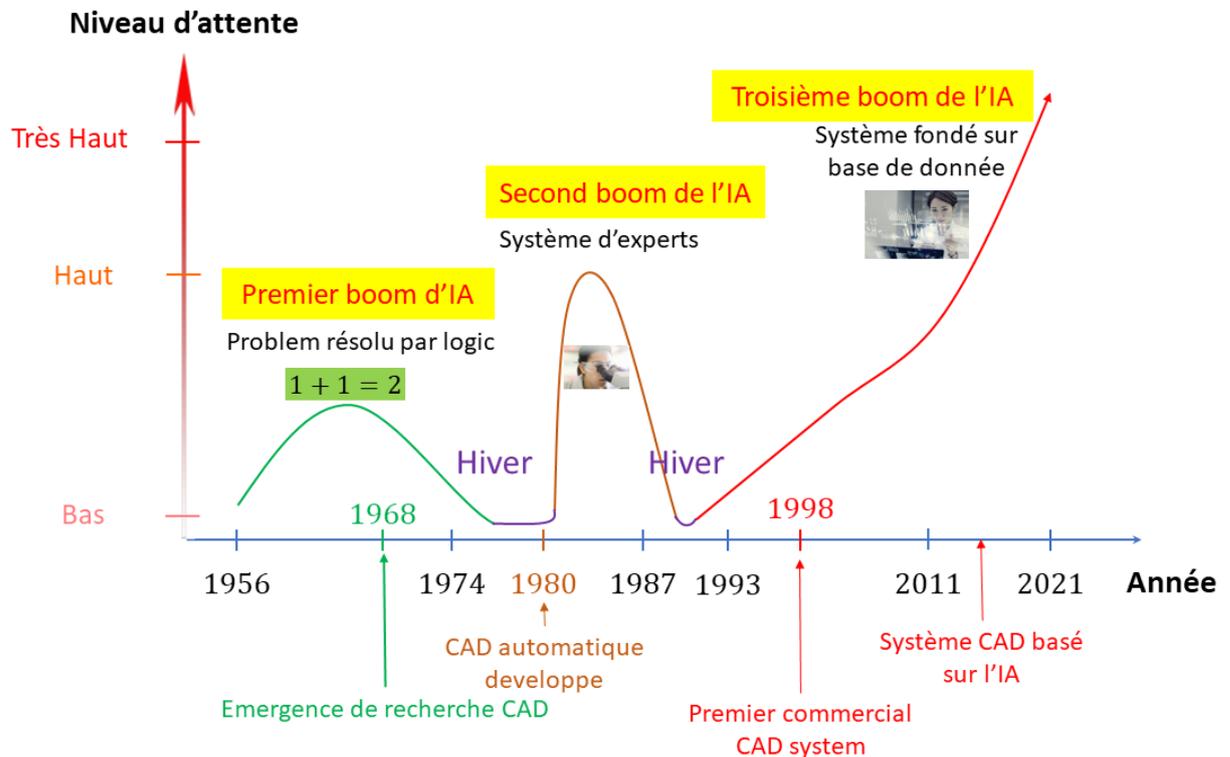


Figure 1-8: Deux booms de l'IA dans le passé et le troisième boom actuel de l'IA. Les progrès de la recherche CAO sont profondément liés aux progrès de la technologie de l'IA.

Au cours des années après cette inauguration, l'IA a connu plusieurs vagues d'optimisme, (126,127), suivies de déception et de perte de financement (128,129), de nouvelles approches, de succès et de financement renouvelé (Figure 1-8) et (130). La médecine n'était pas exclue de cette évolution, en particulier, dans le domaine de la détection et l'avènement des outils de diagnostic assisté par ordinateur ou CAD - « computer-aided-diagnostic ». Les progrès de la recherche sur les systèmes de CAD sont bien alignés sur les progrès de l'IA. La Figure 1-8 illustre cette relation dans laquelle l'axe x et l'axe y représentent respectivement la chronologie par année et le niveau d'attente sur l'IA. Depuis la mise en œuvre de l'IA en 1956 à la conférence de Dartmouth, l'IA a évolué au fil du temps avec l'augmentation des attentes au sein du développement rapide de la société.

Le premier boom de l'IA a duré de 1965 à 1974 avec un financement gouvernemental généreux pour résoudre des problèmes en utilisant des approches basées sur la logique. Pendant ce temps, le premier « chatbot » (131) et le robot intelligent (Shakey) (132) ont été introduits. Après des attentes trop élevées en matière d'IA, une forte réduction du financement et de l'intérêt pour la recherche sur l'IA de 1974 à 1980 a conduit à un net déclin des études dans ce domaine. Cette période s'appelle désormais le premier « hiver de l'IA ».

Le deuxième boom de l'IA, qui repose essentiellement sur la connaissance des experts et est appelé ainsi « systèmes experts », est survenu après le premier hiver de l'IA et a obtenu des réussites remarquables. Le jalon incarne ce boom est que de nombreux produits d'IA commercialisés étaient disponibles sur le marché (133), tels que les produits activés par la voix. Le deuxième « hiver de l'IA » commence avec l'effondrement de la Lisp Machine en 1987 (134) car il est trop coûteux de mettre à jour et de maintenir des systèmes.

Après une réduction passagère de l'intérêt et du financement, l'IA a fait un retour avec l'événement où DeepBlue d'IBM a battu le champion du monde des jeux d'échecs, Garry Kasparov, en 1997. En accord avec de nombreux observateurs (135), le troisième boom de l'IA a commencé depuis 1993 avec la stratégie centrale d'approche fondée par des bases de données. Voici quelques repères de ce troisième boom. En 2002, Amazon a utilisé des systèmes automatisés pour fournir des recommandations et IBM Watson bat deux champions humains au quiz télévisé « Jeopardy » en 2011 (136). Récemment, Google a introduit des voitures sans conducteur en 2012 (137) et Alpha Go en 2016 (138) pour battre un champion du monde à jeux de Go.

Soutenue par le développement de l'IA, les outils CAD en médecine a évolué au fil du temps depuis 1968 lorsque ses recherches ont officiellement vu le jour (139,140). La recherche en CAD à croissance rapide a commencé avec la CAD automatisée dans les années 1980 et a abouti à un premier système CAD commercial pour la mammographie, le système « Image Checker », approuvé par la Food and Drug Administration (FDA) des États-Unis en 1998 (141). Avec le troisième boom de l'IA, les CAD (130,142) devraient s'adapter puis améliorer ses performances en réponse aux problèmes cliniques et aux limites de la technologie contemporaine. Ainsi, au XXI^e siècle, les techniques d'IA ont connu une résurgence suite aux progrès simultanés de la puissance informatique, de grandes quantités de données et de la compréhension théorique ; et les techniques d'IA sont devenues une partie essentielle de l'industrie de la technologie, aidant à résoudre de nombreux problèmes difficiles en informatique, en génie logiciel et en recherche opérationnelle (133).

Ces succès n'étaient pas dus à un nouveau paradigme révolutionnaire, mais principalement à l'application fastidieuse des compétences en ingénierie et à l'augmentation considérable de la vitesse et de la capacité des ordinateurs dans les années 90. En fait, l'ordinateur de Deep Blue était 10 millions de fois plus rapide que le Ferranti Mark 1 que Christopher Strachey a appris à jouer aux échecs en 1951. Cette augmentation spectaculaire est mesurée par la loi de Moore, qui prédit que la vitesse et la capacité de mémoire des ordinateurs doublent tous les deux ans, du fait que le nombre de transistors métal-oxyde-semi-conducteur double tous les deux ans. Certaines personnes considèrent également l'IA comme un danger pour l'humanité si elle progresse sans relâche (143). D'autres pensent que l'IA, contrairement aux révolutions technologiques précédentes, créera un risque de chômage de masse (144).

1.2.2 APPRENTISSAGE AUTOMATIQUE

Dans le domaine de l'IA, il y a différentes sous-catégories (voir Figure 1-7). En fait, IA peut inclure des systèmes qui exécutent simplement des tâches qui n'exigeraient pas que le système apprenne à tout. En imagerie médicale, par exemple, l'IA peut être appliquée pour détecter les structures anatomiques en utilisant des algorithmes artisanaux qui mettent en œuvre l'idée d'un logiciel. La grande majorité des méthodes qui ont été développés au cours des dernières décennies cette catégorie. Un sous-groupe de techniques d'IA, appelé « apprentissage automatique » (AA) ou Machine Learning (ML), est capable d'apprendre automatiquement des données qu'il est présenté avec. L'apprentissage automatique est basé sur la définition d'Arthur Samuel en 1959, comme un "domaine d'étude qui donne aux ordinateurs la capacité pour apprendre sans être explicitement programmé » (145)

Les approches d'apprentissage automatique sont traditionnellement divisées en trois grandes catégories, en fonction de la nature du « signal » ou du « retour d'information » disponible pour le système d'apprentissage :

- Apprentissage supervisé : L'ordinateur est présenté avec des exemples d'entrées et leurs sorties souhaitées, données par un « expert », et le but est d'apprendre une règle générale qui instruit les entrées aux sorties.
- Apprentissage non supervisé : aucune étiquette n'est donnée à l'algorithme d'apprentissage, le laissant seul pour trouver une structure dans la base de données entrées. L'apprentissage non supervisé peut être un objectif en soi (découvrir des modèles cachés dans les données) ou un moyen d'atteindre un but (apprentissage des fonctionnalités).
- Apprentissage par renforcement : un programme informatique interagit avec un environnement dynamique dans lequel il doit réaliser un certain objectif (comme conduire un véhicule ou jouer à un jeu contre un adversaire). Au fur et à mesure qu'il navigue dans son espace de problèmes, le programme reçoit des commentaires analogues aux récompenses, qu'il essaie de maximiser sa performance(146).

D'autres approches ont été développées qui ne s'intègrent pas parfaitement dans cette triple catégorisation, et parfois plus d'une est utilisée par le même système d'apprentissage automatique. Par exemple, la modélisation de sujets, la réduction de la dimensionnalité ou le méta-apprentissage(147).

1.2.3 RÉSEAUX DE NEURONES ARTIFICIELS

La pierre angulaire d'un Réseau Neuronal Artificiel (RNA) ou « Artificial Neural Network » (ANN) est un modèle mathématique d'un neurone (voir Figure 1-9)(9). Un neurone artificiel peut apprendre à donner un signal de sortie uniquement si une certaine combinaison d'entrées se produit. Pour ce faire, il faut un certain nombre de valeurs d'entrée, x_1, \dots, x_n , multiplie chacun avec son poids synaptique, w_1, \dots, w_n , et les accumuler dans une *unité de sommation* pour une *fonction de transfert* correspondante f_i . Puisqu'un neurone lui-même a son propre poids (148), désigné comme «biais », le résultat de l'unité de sommation est la somme de toutes les unités d'association et de biais. Cette procédure entraîne une valeur, appelée a , qui est une entrée pour une unité de sortie. L'unité de sortie incarne l'activité du neurone. Cette unité de sortie peut être linéaire, discrète ou une autre fonction de distribution continue. La fonction utilisée sur la Figure 1-9 est une fonction binaire dans laquelle le neurone est inactivé si la valeur a est inférieure à un seuil t donné ou activé au cas contraire. Les poids synaptiques et les biais sont alors les paramètres à régler « appris » dans afin de produire l'activation appropriée.

Entrée Poids synaptique

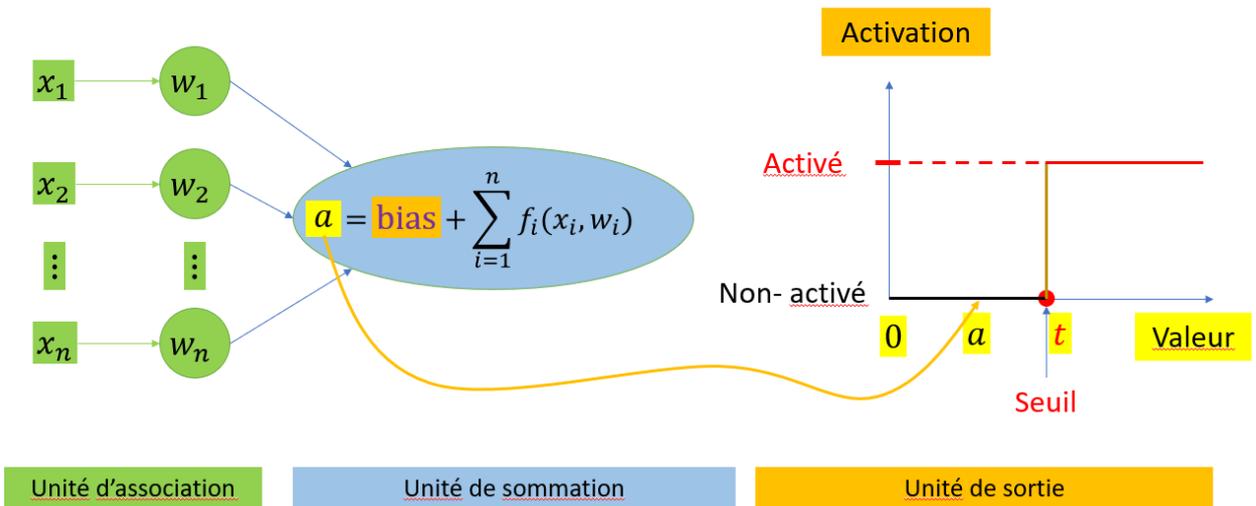


Figure 1-9: Illustration pour un neurone artificiel avec n d'entrées x_1, \dots, x_n , et son poids synaptique correspondant w_1, \dots, w_n en supposition d'avoir un décalage, appelé bias, pour le neurone.

1.2.4 APPRENTISSAGE PROFOND

Pour faire face aux problèmes plus complexes, un réseau des neurones peut être construit contenant de nombreuses couches de neurones, d'où le terme « d'apprentissage profond » ou de « deep-learning ». Dans ce cas-là, il y aura peut-être des millions de paramètres de poids synaptique et de biais qui ont besoin de formation. L'augmentation des couches améliore le pouvoir d'expression et la performance de ces méthodes et pourrait produire un niveau d'abstraction plus élevé (149). Le deep-learning représente actuellement l'état de l'art en apprentissage automatique pour une diversité de tâches et d'applications, en particulier pour les problèmes impliquant de grandes collections de données d'entraînement structurés, ce qui est le cas pour la classification des fibroses interstitielles pulmonaires (2). Puisque ce système est complété en ajustant des hyperparamètres avec une grande quantité de données d'apprentissage, on dirait que les CAD issues du deep-learning sont basés sur les données, et le résultat qui détermine sa performance dépend de la dimension de collection des données d'images. On s'attend alors à ce que ces produits de nouvelle génération d'IA montrent de meilleures performances que les CAD conventionnelles au fur et à mesure que les données d'image augmentent (130).

Ainsi, les techniques d'apprentissage automatique conventionnelles étaient limitées dans leur capacité à traiter les données naturelles sous leur forme brute. Pendant des décennies, construire un système de reconnaissance de formes ou d'apprentissage automatique requies l'ingénierie minutieuse et l'expertise considérable du domaine pour concevoir une fonctionnalité d'extracteur qui a transformé les données brutes (telles que les valeurs de pixels d'une image) dans une représentation interne appropriée ou un vecteur de paramètres à partir duquel le sous-système d'apprentissage, souvent un classificateur, pourrait détecter ou classer les modèles dans l'entrée.

L'apprentissage de la représentation est un ensemble de méthodes permettant à une machine d'être alimenté en données brutes et découvrir automatiquement les représentations nécessaires pour la détection ou la classification. Les méthodes du deep-learning sont des méthodes d'apprentissage de la représentation avec plusieurs niveaux de représentation, obtenu en composant des modules simples mais non linéaires, où chacun transformer la représentation à un niveau (en commençant par l'entrée brute) en une représentation à un niveau plus élevé,

légèrement plus abstrait. Avec la composition suffisante de telles transformations, des fonctions de haute complexité peuvent être apprises. Pour les tâches de classification, les couches de représentation supérieures amplifient les aspects de l'entrée qui sont importants pour la discrimination et suppriment les variations non pertinentes. Dans le cadre de la radiologie, une image digitale est sous la forme d'un tableau de valeurs de pixels. Les paramètres appris dans la première couche de représentation représentent généralement la présence ou l'absence de bords à des orientations et à des emplacements particuliers de l'image. La deuxième couche détecte généralement les motifs en repérant des arrangements particuliers de bords, quelles que soient les petites variations des positions de ces bords. La troisième couche peut assembler des motifs dans des combinaisons plus grandes qui correspondent à des parties d'objets familiers, et les couches suivantes détecteraient des objets sous forme de combinaisons de ces pièces. L'aspect clé de deep-learning est que ces couches de fonctionnalités ne sont pas conçues par des ingénieurs humains : elles sont apprises à partir de données à l'aide d'une procédure d'apprentissage générale automatique (149).

De telles manières, le deep-learning a fait des progrès majeurs dans la résolution de problèmes qui ont résisté aux meilleures tentatives de la communauté de l'intelligence artificielle pendant de nombreuses années. Nous pensons que l'apprentissage profond aura beaucoup plus de succès dans le futur proche car il nécessite très peu d'ingénierie humaine, donc il pourrait facilement profiter des augmentations de la quantité d'algorithmes et de collections de données disponibles. De nouveaux algorithmes et architectures d'apprentissage actuellement en cours de développement pour les réseaux de neurones profonds ne feraient qu'accélérer cette progression(149). De plus en plus de bases de données d'images médicales sur les maladies pulmonaires deviennent publiques, les chercheurs ont alors l'accès à davantage de données. Ainsi, des études plus approfondies ont été menées. La quantité totale d'investigation des maladies pulmonaires utilisant le deep-learning a augmenté régulièrement au fil des ans. La majorité d'elles portaient sur la détection de la tuberculose, suivie du COVID-19, cancer du poumon et pneumonie(3). La raison pour laquelle les travaux sur la tuberculose sont élevés est que la majorité des habitants infectés par la tuberculose venaient de régions pauvres en ressources avec de médiocres infrastructures de soin(4).

Vers 2020, les travaux sur la détection du COVID-19 ont émergé alors que les travaux menés sur d'autres maladies ont considérablement diminué. Cela signifie que l'utilisation de deep-learning pour l'investigation des maladies pulmonaires est toujours un domaine d'étude actif et que les chercheurs du monde entier tentent de réduire le fardeau de détections du COVID-19 (3).

ARCHITECTURES DE RÉSEAU NEURONAL MULTI-COUCHES

De nombreuses applications d'apprentissage profond utilisent le type de réseau de neurones artificiels acycliques (« feed-forward »), qui apprennent à mapper une base de données d'entrée de taille fixe (par exemple, une image) à une sortie de taille fixe. Pour passer d'une couche à la suivante, un ensemble d'unités calcule une somme pondérée de leurs entrées à partir de la couche précédente et passez au résultat via une fonction d'activation non linéaire. À présent, la fonction non linéaire la plus populaire est la fonction d'unité linéaire rectifiée (ReLU), qui est simplement le redresseur demi-onde $f(z) = \max(z, 0)$. Au cours des dernières décennies, les réseaux neuronaux utilisaient des non-linéarités plus lisses, telles que $\tanh(z)$ ou $1/(1 + \exp(-z))$, mais le ReLU apprend généralement beaucoup plus rapidement dans des réseaux à plusieurs couches, permettant la formation d'un réseau sans préformation non supervisée(150).

La première application majeure de cette approche de préformation a été la reconnaissance vocale, et cela a été rendu possible par l'avènement d'unités de processeur graphique (GPU) faciles à programmer et a permis aux chercheurs de former des réseaux 10 à 20 fois plus vite(151). Il y avait, cependant, un type particulier du réseau neuronal multi-couche qui est

beaucoup plus facile à former et généralisé beaucoup mieux que les réseaux avec une connectivité complète entre les couches adjacentes : le réseau neuronal convolutif (ConvNet). Ce dernier a réalisé de nombreux progrès pratiques à l'époque où les réseaux de neurones étaient en disgrâce et il a récemment été largement adopté par la communauté de vision par ordinateur(152).

RESEAUX NEURONAUX CONVOLUTIFS (ConvNet)

L'architecture d'un réseau neuronal convolutif (ConvNet) typique est structurée comme une succession d'étapes. Les premières étapes sont composées de deux types de couche : couches convolutives et couches de pooling. Les unités dans une couche convolutive sont organisées en cartes d'entités, dans lesquelles, chaque unité est connectée à des correctifs locaux dans les cartes des paramètres de la couche précédente à travers un ensemble de poids synaptiques appelé de strates de filtres de convolutions. Le résultat de la somme pondérée locale est ensuite passée à travers une fonction d'activation non-linéaire telle qu'une ReLU. Toutes les unités d'une carte d'entités partagent la même strate de filtres de convolutions. Différentes cartes d'entités dans une couche utilisent différentes strates de filtres de convolutions. Le rôle de la couche convolutive soit de détecter les conjonctions locales d'entités de la couche précédente, le rôle de la couche de pooling consiste à fusionner des entités sémantiquement similaires (Figure 1-10).

Deux ou trois étapes de convolution, de fonction non-linéarité et de pooling sont empilés, suivis de couches plus convolutifs et d'une couche entièrement connectée. Le processus de rétro-propagation de gradient via un ConvNet est aussi simple que via un réseau profond classique, permettant de former à toutes les poids synaptiques des strates de filtres de convolutions. Les couches convolutives et de pooling dans ConvNets sont directement inspiré des notions classiques de cellules simples et de cellules complexes du cortex visuel de l'animal(153–155). Les ConvNets sont alors conçus pour traiter les données qui se présentent sous la forme de plusieurs tableaux, par exemple une image couleur composée de trois tableaux 2D contenant des intensités de pixels dans les trois canaux de couleur. Premièrement, dans les données de tableau telles que les images, les groupes de valeurs locales sont souvent fortement corrélées, formant des motifs facilement détectables. Deuxièmement, les statistiques locales des images et les autres signaux sont invariants à l'emplacement. Autrement dit, si un motif peut apparaître dans une partie de l'image, il peut apparaître n'importe où, d'où l'idée d'unités à différents endroits partageant les mêmes poids et détecter le même motif dans différentes parties du tableau. Mathématiquement, l'opération de filtrage effectuée par une carte de paramètres est un élément discret de convolution.

Toutefois, les ConvNets ont été largement abandonnés par la communauté traditionnelle de vision par ordinateur et d'apprentissage automatique jusqu'au Concours ImageNet en 2012. Quand les réseaux neuronaux convolutifs ont été appliqués à une base de données d'environ un million d'images de l'internet qui contenait 1000 classes différentes, ils ont donné de grands résultats en divisant presque par deux les taux d'erreur des meilleures méthodes d'approche en concurrence(156). Ce succès résulte de l'utilisation efficace des GPU, des ReLUs, et d'une nouvelle technique de régularisation appelée "dropout"(157), et des techniques pour générer plus d'exemples de formation en déformant les données existantes. Ce succès a provoqué une révolution dans la vision par ordinateur; les ConvNets sont désormais l'approche dominante pour presque toutes les reconnaissances et les tâches de détection(158,159) et aborder la performance humaine sur certaines tâches.

Les architectures ConvNet récentes possèdent environ 10 à 20 couches de ReLU, des centaines de millions de poids synaptiques et des milliards de connexions entre unités neuronales. Alors que la formation de si grands réseaux aurait pu prendre des semaines il y a seulement deux ans, des progrès dans le matériel, les logiciels et l'algorithme de parallélisation a réduit les temps de formation à quelques heures(149).

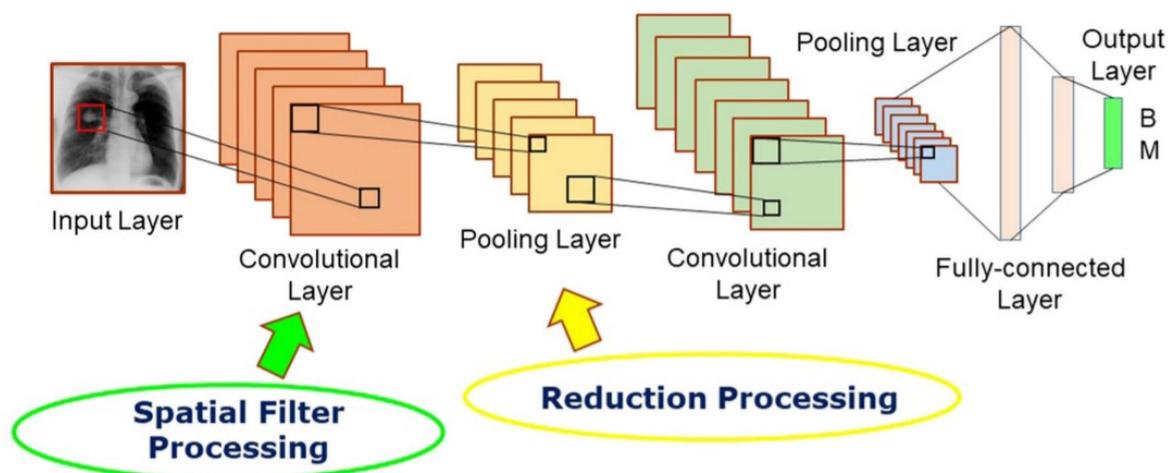


Figure 1-10: Exemple de classification des tumeurs bénigne (B) et maligne (M) sur les radiographies thoraciques par un ConvNet [61].

RESEAUX NEURONAUX RECURRENTS

Pour les tâches qui impliquent des données d'entrée séquentielles, telles que la parole et la langue, il est souvent préférable d'utiliser des réseaux neuronaux récurrents (RNRs) (Fig.5). Les RNNs traitent une séquence d'entrée un élément à un temps, en maintenant dans leurs unités cachées un « vecteur d'état » qui implicitement contient des informations sur l'histoire de tous les éléments passés de la séquence. Lorsque nous considérons les sorties des unités cachées à différents pas de temps discrets comme s'ils étaient les sorties de différents neurones dans un réseau profond de multicouche, il devient clair comment appliquer la rétropropagation pour former les RNRs. Les RNRs sont des systèmes dynamiques très puissants, mais leur formation s'est avérée problématique car les rétropropagations de gradients grandissent ou rétrécissent à chaque pas de temps, donc sur de nombreuses étapes, elles explosent ou disparaissent généralement.

En radiologie, RNN sont utilisés pour résoudre conjointement différents problèmes interdépendants, comme la détection et la caractérisation des nodules (160). Dans l'architecture de ReCTnet, les couches convolutives et récurrentes sont combinés pour apprendre des représentations d'images expressives en exploitant des dépendances spatiales entre les coupes axiales de TDM. Il était démontré que tirer parti des dépendances intra-coupe augmente considérablement la sensibilité pour détecter les nodules pulmonaires sans accroître le taux de faux positifs. Sur la base de données "LIDC / IDRI" (161) accessible au public composé de 1018 TDM annotés, ReCTnet atteint une sensibilité de détection de nodules pulmonaires de 90,5% avec une moyenne de 4,5 fausses positives détectées par examen (160). Comparaisons avec un réseau de neurones convolutif multicanal concurrent pour la segmentation multi-coupes et d'autres méthodologies publiées utilisant la même base de données prouve que ReCTnet offre des meilleures performances.

1.2.5 AVANTAGES ET INCONVENIENCES DE L'INTELLIGENCE ARTIFICIELLE EN IMAGERIE THORACIQUE

Le domaine de l'imagerie médicale réalise actuellement de grandes avancées de l'apprentissage automatique pour les soins de santé publique. En septembre 2019, l'organisation «FDA - Food and Drug Administration » des États-Unis a autorisé GE Healthcare à commercialiser un ensemble d'algorithmes conçus pour détecter le pneumothorax sur la radiographie du thorax (162). Quelques jours plus tôt, un résumé présenté à l'IASLC - La Conférence mondiale sur le cancer du poumon - a présenté les résultats d'un essai portant sur un système de CAD comme le premier lecteur de tomodensitométrie à faible dose dans le dépistage du cancer du poumon(163). Une autre étude a montré qu'un réseau de neurones créé par des scientifiques de Google était au moins aussi précis que les radiologues pour détecter les nodules pulmonaires malins (164). Une étude publiée en 2018 a révélé qu'un algorithme de deep-learning était à peu près équivalent à un groupe de radiologues thoraciques en classification des maladies pulmonaires interstitielles (165). Un modèle similaire s'est avéré capable de diagnostiquer la BPCO chez les fumeurs et prédire des épisodes de pneumopathies aigus et ainsi la mortalité(166).

Les études mentionnées ci-dessus impliquent une méthode de deep-learning appliquée aux TDM et la radiographie du thorax. Les images sont essentiellement des données numériques et les ordinateurs sont particulièrement bien adaptés pour analyser ce type de données. Des ordinateurs peuvent ainsi regarder au-delà de ce que l'œil humain puisse voir et taquiner à part de minuscules variations dans les images examinées, aider au diagnostic, détecter les signes préliminaires de la maladie, et prédire son évolution. Le meilleur usage de ces nouvelles technologies passionnantes donnera non seulement la meilleure utilisation du temps des médecins mais permettra aussi d'améliorer la précision, l'efficacité et la rapidité du diagnostic, et alors l'optimisation du flux de travail. Par exemple, en utilisant des paramètres basées sur la densité et la texture, le groupe de Kao et coll. ont développé un système de CAD pour déterminer automatiquement les examens thoraciques anormaux dans la liste de travail des radiologues, le délai d'exécution pour signaler une radiographie anormale a été réduit de 44 % (167). Les outils de CAD peuvent être utilisés pour les tâches de détection spécifiques sur la radiographie thoracique, telles que la détection de la tuberculose, de la pneumonie ou du nodule pulmonaire (3,168). D'autres tâches encore plus avancées telles que la détection de multiples maladies sont également en cours de développement. D'énormes perspectives sur l'impact de l'intelligence artificielle appliquée à l'imagerie thoracique pour les endroits où il y a une pénurie d'experts capables d'interpréter les radiographies thoraciques, même lorsque du matériel d'imagerie est disponible(3,169).

Toutefois, l'application d'algorithmes de deep-learning à l'analyse d'images médicales présente plusieurs défis. Premièrement, le manque de grandes bases de données de formation est souvent mentionné comme un obstacle. Dans le deep-learning, il est très important d'avoir une grande base de données de formation, comme la communauté a assumé qu'une plus grande quantité d'images peut aider à améliorer la précision de l'entraînement. Même un algorithme faible avec une grande quantité de données peut être plus précise qu'un algorithme fort avec une quantité modeste de données (170).

Un autre obstacle est le déséquilibre des classes. Lors de la formation à la classification binaire, si le nombre des échantillons d'une classe sont beaucoup plus élevés que l'autre classe, le modèle résultant serait biaisé(171). Les algorithmes de deep-learning fonctionnent de manière optimale lorsque la quantité d'échantillons dans chaque classe est égale ou équilibrée.

Une façon d'agrandir la base de données d'entraînement sans obtenir de nouvelles images consiste à utiliser les techniques d'augmentation d'image. L'augmentation d'image crée des variations des images originales. Ceci est réalisé en effectuant différentes méthodes de traitement,

telles que les rotations, les retournements, les traductions, les zooms et l'ajout de bruit sur l'image originale (172).

Une autre stratégie pour régler le problème de restriction de données est l'apprentissage par transfert qui est devenu une méthode populaire en vision par ordinateur (173). Avec l'apprentissage par transfert, un modèle appris à partir d'un domaine peut être réutilisé sur un domaine différent. L'apprentissage par transfert peut être effectué avec ou sans modèle pré-entraîné. Le modèle pré-entraîné est un modèle développé pour résoudre une tâche similaire. Au lieu de créer un modèle pour résoudre une nouvelle tâche similaire, le modèle formé sur un autre problème est utilisé comme point de départ. Même si le modèle pré-entraîné soit entraîné sur une tâche différente de la tâche actuelle, les fonctionnalités apprises, dans la plupart des cas, jugé utile pour la nouvelle tâche. Par exemple, de nombreuses architectures ConvNet. sont pré-entraînées sur ImageNet (156). Bush et coll. ont préformé un modèle ConvNet. sur un sous-groupe de données d'ImageNet pour classer les radiographies thoraciques comme positives ou négatives pour la présence de nodules pulmonaires avec une sensibilité de 92% et une spécificité de 86 % (174).

Actuellement, l'utilisation des systèmes PACS en radiologie est courante dans la plupart des hôpitaux occidentaux depuis au moins une décennie et ceux-ci sont remplis de millions d'images. Nous constatons également que de plus en plus des bases de données publiques sont mis à disposition des chercheurs. Par exemple, la base de données largement utilisé pour évaluer les algorithmes d'analyse d'images pour détecter les nodules dans la TDM pulmonaire de LIDC-IDRI (161), les données de la cohorte des patients testés pour COVID-19 à l'Université de SUNY (175). L'une des plus grandes bases de données est chestX-ray8, construite à partir du PACS clinique des hôpitaux affiliés à l'Institut national de la santé (« The National Institute of Health »). Cette base de données comprend 111220 de radiographies de vue frontale de 30805 patients et dans un premier temps les étiquettes d'image de 8 maladies, puis étendue à 14 maladies (thorax -ray14), y compris l'atélectasie, la consolidation, l'infiltration, le pneumothorax, l'œdème, l'emphysème, la fibrose, l'épanchement, la pneumonie, l'épaississement pleural, cardiomégalie, nodule, masse et hernie (176).

Un autre défi concerne l'acquisition d'annotations ou d'étiquetages pertinents pour les images médicales. Compte tenu de la complexité pour former des algorithmes en domaine de santé, les chercheurs ont besoin généralement l'avis des experts du domaine (par exemple, radiologues, pathologistes) de faire des annotations spécifiques à la tâche pour les données d'image. Ce type d'analyse supervisée est limité par une subjectivité intégrée, et sans un contrôle de qualité approprié, l'apprentissage automatique apprendra à copier les erreurs humaines. Ces raisons expliquent la tendance actuelle d'un intérêt renouvelé sur l'apprentissage non supervisé de la communauté de l'apprentissage automatique. Les méthodes non supervisées sont attrayantes, car elles permettent la formation (initiale) du réseau avec la richesse des données non étiquetées disponibles dans le monde (177).

L'utilisation de ConvNet pour les images tomodensitométriques est plus complexe que pour les images radiographiques thoraciques 2D, en raison de la nature 3D des images. L'étiquetage d'une base de données suffisamment volumineux peut prendre un temps considérable, ce qui pose de problème. Par exemple, pour former des systèmes de deep-learning pour la segmentation des images en 3D de TDM, des annotations tranche par tranche doivent être faites et cela prend beaucoup de temps en raison du nombre élevé de coupes. Pour surmonter ce problème, certaines études utilisent des ConvNet. 2D appliqués à chaque tranche, tandis que d'autres choisissent d'adopter une approche basée sur les patches ou de réduire la taille de l'image au prix - dans les deux cas - d'une perte d'information. Malgré ces difficultés techniques, les résultats sont prometteurs et les réseaux neuronaux ConvNets permettent généralement d'obtenir de meilleurs résultats que les méthodes traditionnelles d'apprentissage automatique. Par exemple, le logiciel

Caliper a été développé en utilisant l'approche basée sur les patches, pour la quantification de l'étendue de la maladie et de l'évolution de la fibrose pulmonaire idiopathique(118,119).

Une autre particularité de la TDM thoracique est que les images sont généralement reconstruites à l'aide de deux filtres de reconstruction différents, un noyau haute fréquence, le noyau pulmonaire (NP) pour l'évaluation du parenchyme pulmonaire et un noyau de reconstruction standard, le noyau médiastinal (NM) pour l'évaluation du médiastin. Des études récentes ont montré que les performances des outils basés sur l'AP pour la détection et la caractérisation des nodules pulmonaires sont influencées par les paramètres d'acquisition et de reconstruction, tels que le choix du noyau de reconstruction (178,179). Cependant, il n'existe actuellement aucune donnée évaluant l'influence des filtres de reconstruction sur la segmentation des maladies pulmonaires à l'aide de méthodes de recherches basées sur les données.

Dans l'analyse d'images médicales, les informations utiles ne sont pas juste contenu dans les images elles-mêmes. Les médecins exploitent souvent une mine de données sur les antécédents des patients, l'âge, la démographie et autres examens paracliniques pour prendre de meilleures décisions. Certains auteurs ont déjà étudié la combinaison de ces informations dans des réseaux de deep-learning de manière simple (180). Cependant, comme ces auteurs notent que les améliorations obtenues n'étaient pas aussi importantes que prévu. L'un des défis est d'équilibrer le nombre de fonctionnalités d'imagerie dans le réseau de deep-learning (généralement des milliers) avec le nombre de fonctionnalités cliniques (généralement une poignée seulement) pour éviter les paramètres cliniques d'être noyées.

Tout comme les humains peuvent parfois se fier à leur expérience ou à leur « instinct » sans pouvoir expliquer clairement leur décision, les réseaux de neurones peuvent également être considérés comme des « boîtes noires ». Surtout en médecine, où la responsabilité est importante et peut avoir de graves conséquences juridiques, il ne suffit souvent pas d'avoir un bon système de prédiction. Ce système doit également pouvoir s'articuler d'une certaine manière au public général. Plusieurs stratégies et campagnes sociales ont été développées pour comprendre à quoi répondent les couches intermédiaires des réseaux convolutifs, par exemple les réseaux de déconvolution (181), la rétropropagation guidée (182), ou la composition profonde de Taylor (183).

Enfin, semblable à l'éducation et à la certification d'un réseau de neurones du cerveau biologique (c'est-à-dire les médecins), un réseau de neurones artificiels doit également être formé et validé de manière approfondie avant de pouvoir être appliqué en clinique. Il existe cependant un déséquilibre entre la quantité d'innovations techniques publiées, comme décrit ci-dessus, et la quantité de techniques qui en font partie réellement la pratique clinique. En raison du biais de publication, une grande partie des innovations ne passe pas la validation externe ou ne fonctionne que sur des bases de données très spécifiques. Mais il semble également que la réalisation d'études de validation soit un facteur limitant lui-même. Les chercheurs peuvent uniquement s'intéresser à l'innovation purement technique sans viser une application clinique et la validation logicielle est moins intéressante pour les chercheurs de sciences fondamentales en raison de la quantité d'administration nécessaire pour obtenir l'autorisation d'organismes de réglementation pour un tel dispositif médical. La validation des logiciels est donc principalement effectuée par des sociétés industrielles, mais uniquement s'il existe une viabilité commerciale claire pour ce produit. Pour cette raison, une collaboration plus étroite entre les universités et l'industrie pourrait être nécessaire, de même entre les chercheurs techniques, biologiques et cliniques. Pour ces types de collaborations, les travaux de recherche interdisciplinaires en université sont essentiels pour combler davantage cette lacune(9). Il existe déjà plusieurs initiatives des différentes organisations professionnelles pour créer un nouveau programme de formation à l'utilisation de l'IA en imagerie médicale.

1.3 CONTRIBUTIONS ET ORGANISATION DE THESE

1.3.1 PRINCIPAUX DÉFIS

Dyskinésie ciliaire primitive

La demande clinique d'un score d'évaluation de la gravité de la maladie de DCP propre n'a pas été suffisamment pris en compte. Le mécanisme pathologique de cette maladie rare reste encore obscur et a besoin d'être exploré. L'imagerie TDM quantitative automatique par des outils récents de CADs pourrait d'une part éliminer la subjectivité des scores visuels, en ajoutant de la simplicité et de l'efficacité à l'usage clinique et d'autre part de dépasser la performance de l'œil humain à détecter les minuscules variations morphologiques.

Dysfonction chronique du greffon pulmonaire

Plusieurs paramètres de TDM quantitatifs ayant été impliqués dans la population post-transplantation pulmonaire, depuis la mesure des voies aériennes et de la vascularisation avec une différence remarquable entre les sujets atteints de BO, et l'approche du parenchyme essayant de prédire la survenue de BO avant les manifestations cliniques nécessitant d'avantage d'investigations méthodologiques expérimentales.

Ainsi, il n'existe actuellement aucun article explorant la valeur potentielle de la "Radiomique", la méthode qui extrait un grand nombre de paramètres d'images médicales avec une analyse ultérieure par des algorithmes de caractérisation des données (184,185), pour la prédiction de la DCGP. Le deep-learning basé sur les réseaux de neurones convolutifs représente une avancée majeure dans le domaine de l'intelligence artificielle appliquée à l'imagerie thoracique, mais nécessite une grande quantité de données (186,187). Rappelons que la plupart des séries sur l'imagerie quantitative pour l'évaluation DCGP après LTx sont basées sur un nombre limité de patients, car la transplantation pulmonaire n'est pas une procédure courante.

Sclérodémie

Plusieurs systèmes de notation visuelle ont été proposés pour la quantification de la gravité de la PID dans les travaux de recherches. Ces systèmes souffrent d'un accord inter-observateur de faible à modéré (74,188). Ainsi, les corrélations entre ces scores visuels et les indices d'EFR restent modérées (66,74). En effet, la large gamme de valeurs normales d'EFRs rend les changements modérés difficiles à interpréter. L'identification par TDM des patients susceptibles de s'aggraver peut également être importante pour des études pharmaceutiques ultérieures(67). Une meilleure prédiction du risque individuel de mortalité à court et long terme est un indicateur important pour la décision de traitement. La notation visuelle en TDM n'est pas réaliste en routine clinique, mais de nouveaux scores automatiques quantitatifs basés sur l'histogramme et des méthodes texturales ont été proposés, bien qu'ils ne soient pas actuellement largement disponibles(66,189).

D'un point de vue physiologique, l'approche automatique actuelle qui ne porte que sur l'étendue morphologique de l'atteinte pulmonaire ne reflète pas l'intégralité du processus pathologique de la maladie ScS qui inclut à la fois l'inflammation et la fibrose interstitielle. D'un autre côté, l'évaluation de la réduction de la taille pulmonaire est plus complexe et nécessite une comparaison côte à côte des images tomodensitométriques à chaque niveau.

Le noyau de reconstruction pour l'entraînement des modèles de deep-learning en tâche de segmentation des zones malades de la pneumopathie diffuse.

Dans différentes régions, divers hôpitaux ont des paramètres différents pour les mêmes éléments d'examen tomodensitométrique (notamment différence de doses de rayonnement de TDM, de filtres de convolution). Des études récentes ont montré que les performances des outils

basés sur l'IA pour la détection et la caractérisation des nodules pulmonaires sont influencées par les paramètres d'acquisition et de reconstruction, tels que le choix du noyau de reconstruction(178,179). Cependant, il n'existe actuellement aucune donnée évaluant l'influence des filtres de reconstruction sur la segmentation des maladies pulmonaires diffuses à l'aide de méthodes de recherches basées sur ces données.

1.3.2 CONTRIBUTIONS

Nos contributions sont classées en deux volets : le premier volet concerne l'application de l'intelligence artificielle dans le cadre des maladies de DCP, de DCGP et de ScS et le deuxième volet, l'influence du choix de noyau de convolution pour l'application de de deep-learning dans l'évaluation des maladies pulmonaires interstitielles.

Pour la DCP, nous développons une méthode de quantification automatisée pour l'évaluation des lésions structurelles du poumon par l'examen de tomодensitométrie dans la corrélation avec les indices fonctionnels par test d'explorations fonctionnelles respiratoires (EFR).

Pour la maladie de DCGP, les travaux sont partagés en deux étapes. À la première étape, nous effectuons la synthèse de la littérature des méthodes de quantification des changements morphologiques en imagerie auprès de la survenue de pathologie. Deuxièmement, nous implémentons des travaux expérimentaux fondés sur la radiomique – technique qui consiste à identifier à partir de caractéristiques d'image, non perceptibles à l'œil humain à la recherche possible biomarqueurs, pour permettre un diagnostic précoce de la DCGP.

Dans le cadre de la ScS, nous avons démontré une corrélation linéaire entre l'étendue de l'ILD sur le scanner et la mortalité. Cette association pourrait inciter les cliniciens à renforcer la surveillance clinique des stades précoces de la SSc-ILD et à discuter des mesures thérapeutiques précoces, au lieu d'attendre une aggravation significative. La mesure automatisée de l'ILD sur la tomодensitométrie nous a permis d'identifier un groupe de patients présentant un risque de mortalité à court terme plus élevé.

La dernière partie consiste à évaluer l'influence des filtres de reconstruction sur la segmentation des maladies pulmonaires diffuses à l'aide de deep-learning. Nous avons procédé à des investigations dans le cadre des deux bases de données ScS-ILD et Covid-19 par une architecture de deep-learning populaire, et avons rapporté un impact significatif du choix du noyau de reconstruction sur les performances des modèles et que les algorithmes d'entraînement et qu'une combinaison de filtres pulmonaires et médiastinaux améliore considérablement les performances.

1.3.3 ORGANISATION DE LA THESE

L'organisation de cette thèse est illustrée à la Figure 1-11. Nous commençons par les notions préliminaires sur la méthodologie de l'imagerie quantitative et de l'apprentissage automatique en recherche du domaine médical du chapitre 2, la partie de matériels et méthodes. Comme abordé dans la section 1.1.1, au chapitre 3, nous appliquerons une approche d'imagerie quantitative par CAD pour évaluer la gravité de l'atteinte pulmonaire dans une cohorte de patients adultes avec une dyskinésie ciliaire primitive. Les paramètres d'histogramme de l'images obtenue par la TDM sont exploités pour la sélection du seuil de densité qui obtient la meilleure corrélation avec des indices d'EFR. Ce seuil détermine le pourcentage de zones à haute densité, et représente des anomalies telles que l'épaississement des parois bronchiques, des impactions mucoïdes, de condensation et d'atélectasie fréquemment observées dans la DCP. Nous proposons un biomarqueur d'imagerie automatisée dédiée pour cette population de malades. Ce travail a été publié dans le journal «BMC Pulmonary Medicine » (190).

ORGANISATION DE THESE

- 1) Introduction: PCD, ScS, DCGP et IA
 2) Méthodologie: Mégadonnées, Radiomique

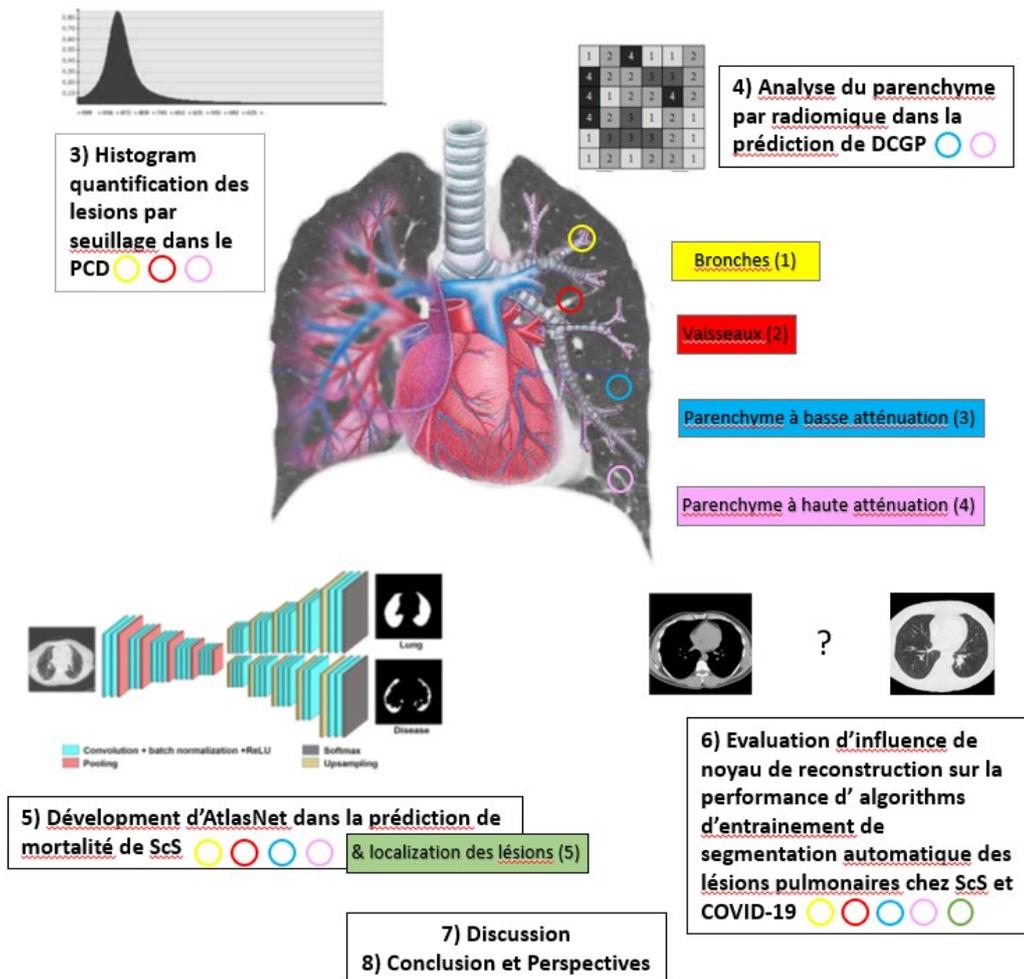


Figure 1-11: Organisation de thèse.

Dans le chapitre 4, nous essayons de résoudre le problème de diagnostic et de prédiction la dysfonction chronique du greffon pulmonaire (DCGP) dans une cohorte de 73 patients (41 avec DCGP et 32 témoins sans DCGP appariés en âge, sexe et temps de suivi post-LTx) au délai de 6 mois avant la survenue de la dégradation sans récupération de la fonction pulmonaire. La radiomique est utilisée, l'extraction et le choix des caractéristiques radiomiques était réalisé automatiquement par la boîte d'outil « PyRadiomics » disponible en ligne. Les variables radiomiques de forme, de l'histogramme, de texture ont été analysés. Ensuite, en divisant en sous-jeux de données d'entraînement, validation et de test, nous avons évalué la performance des différents modèles de classification pour la prédiction de DCGP. Notre classificateur d'ensemble rapporte des résultats préliminaires statistiquement non-significatifs dans le groupe de test puisqu'ils sont bien supérieurs aux résultats en jeu de validation (Tableau 4-8). On a noté également un effet de sur-apprentissage étant donné que les résultats d'entraînement sont bien plus élevés que pour le test (Tableau 4-6) Cependant, les résultats de nos travaux suggèrent la faisabilité d'application de la radiomique pour l'entraînement des modèles de deep-learning en tâche de prédiction de la survenue de DCGP et soulignent la nécessité d'élaborer des investigations de plus grande dimension avec une collaboration multicentrique des établissements disposant du greffon pulmonaire dans le futur. Ce travail ouvre une nouvelle direction pour une prise en charge plus efficace des patients en

exploitant des informations de l'imagerie permettant de prédire la DCGP avant les manifestations fonctionnelles respiratoires.

Dans le chapitre 5, nous avons construit un modèle de prédiction du risque incluant âge avancé, sexe masculin, score de dyspnée NYHA, les caractéristiques de la ScS (type de maladie et auto-anticorps identifiés) et étendue de l'ILD au scanner à l'issue de l'AtlasNet- un algorithme basé sur le deep-learning. Ce modèle a montré une bonne capacité à prédire la mortalité à 10 ans (c Index 0,789 et D Index 1,83). L'ajout de la mesure automatisée de l'étendue de l'ILD sur la tomodensitométrie à ce modèle de base a légèrement amélioré la capacité de prédire la mortalité à 10 ans, mais n'a pas atteint la signification statistique. En considérant la mortalité précoce, la contribution de la tomodensitométrie a augmenté et nous a permis d'améliorer significativement la capacité de prédire la mortalité à 2 ans.

Le travail dans le chapitre 6 concerne l'évaluation de l'influence et ainsi la valeur contributive de la combinaison des deux filtres de convolution médiastinal et pulmonaire des données de TDM thoraciques d'entrée dans l'entraînement des modèles de la radiomique dans une tâche de segmentation de la maladie pulmonaire diffuse de la ScS et du COVID-19. En utilisant deux ensembles de données de différentes maladies et une architecture de deep-learning populaire, l'U-net, nous avons démontré que le choix du noyau de reconstruction a un impact significatif sur les performances des modèles et que les algorithmes d'entraînement sur une combinaison de filtres pulmonaires et médiastinaux améliorent considérablement les performances.

Enfin, nous discutons dans le chapitre 7 les questions relatives au développement de l'IA en imagerie médicale.

2 MATERIELS ET METHODES

2.1 DE BIG DATA A L'AGE DES RECHERCHES AXÉES SUR LES DONNÉES

2.1.1 BIG DATA

Traditionnellement, les professionnels de la santé ont largement utilisé l'expérience et l'observation pour guider les diagnostics et les décisions de traitement. Cependant, au cours des dernières décennies, les données à haut débit de la biologie moléculaire et des dossiers de santé électroniques, associées à des capacités informatiques croissantes allant des téraoctets aux pétaoctets, ont propulsé la médecine axée sur les données au premier plan (191,192). Il est maintenant généralement admis que la médecine fondée sur les preuves - et plus récemment encore, la médecine véritablement personnalisée et de précision - devrait constituer l'étalon-or des soins. Pourtant, cette nouvelle norme nécessite la collecte et l'analyse de données à une échelle sans précédent - en d'autres termes, des « Big data » ou « Mégadonnées ». Une définition pratique des « Big data » est « des ensembles de données si volumineux ou complexes que les méthodes traditionnelles de traitement des données sont inadéquates » (193). Alors que cette définition capture la réalité de la mégadonnée, la définition proposée à l'origine par Gartner, qui décrit le big data par son volume, sa variété et sa vitesse (les 3V), a été adoptée par de nombreux auteurs (194). Les V supplémentaires utilisés par d'autres pour décrire le Big Data incluent la véracité, la valeur et la variabilité.

- Volume : Les volumes d'ensembles de données disponibles augmentent de façon exponentielle, les études modernes produisant des téraoctets, des pétaoctets ou des exaoctets de données (195). Dans un effort pour développer une définition « précise, bien formée et sans ambiguïté » des mégadonnées dans les soins de santé, Baro et al. ont effectué une recherche dans la littérature pour le terme « mégadonnées » et identifié 196 articles directement liés à la santé humaine. Parmi ceux-ci, 48 comprenaient des ensembles de données et ont été exploités pour le nombre d'individus statistiques (n , qui peut être supérieur au nombre d'individus physiques), et le nombre de variables (p) mesurées pour chaque individu (196). Ces auteurs ont proposé un seuil quantitatif pour les mégadonnées basé uniquement sur le nombre total de points de données dans l'ensemble de données, définissant les « grands ensembles de données » comme ceux dans lesquels $\log(n \times p) \geq 7$.

- Variété : Les mégadonnées dans les soins de santé proviennent de nombreuses sources distinctives à de nombreux niveaux, de la molécule à la cellule, de l'organe entier, de l'individu (c'est-à-dire des mesures « cliniques »), de l'environnement et de la population, avec une variété de mesures possibles effectuées à chaque niveau. Les données recueillies à partir de ces sources peuvent être structurées ou non structurées (194,197). Les données structurées ont un haut degré d'organisation, ce qui les rend susceptibles d'être représentées dans les lignes et dans les colonnes d'une matrice de données. Les données structurées sont souvent stockées dans des bases de données relationnelles. Contrairement aux données structurées, les données non structurées n'ont pas d'organisation prédéfinie et, bien qu'elles puissent avoir leur propre structure interne, elles ne sont pas conformes aux lignes et aux colonnes et ne sont pas facilement stockées dans les bases de données. Imagerie médicale [Imagerie par Résonance Magnétique (IRM), Radiographies, Tomodensitométrie (TDM)] présentent des exemples de données non structurées. On estime qu'environ 80% des informations dans le secteur des soins de santé humaine ne sont pas structurées(198,199).

- Vitesse - La vitesse à laquelle les « Big Data » s'accumulent fait référence à la fréquence de mise à jour des données ou à la croissance des données au fil du temps(200). Les données mises à jour en temps réel ont une valeur ajoutée car elles peuvent aider les chercheurs et les cliniciens à prendre des décisions qui offrent des avantages stratégiques. Par exemple, dans la modélisation, l'impact spécifique des décisions de prévention, de traitement ou de gestion.

- Variabilité - La variabilité des mégadonnées fait référence à l'exhaustivité des données et à la manière dont les données peuvent ou non évoluer au fil du temps, saisit également la complexité des données biomédicales, même dans le cas où les données proviennent d'une seule source (192,193,201). Par exemple, les données d'expression génique peuvent être différentes dans différents tissus et en réponse à différents stimuli environnementaux. Cet aspect des mégadonnées rend l'interprétation dans le bon contexte particulièrement cruciale et peut affecter la capacité d'extrapoler les résultats dans une nouvelle condition (192).

- Véracité – C'est la partie de « Big Data » à laquelle nous pouvons croire. Le contrôle de la véracité dans un ensemble de données nécessite une attention particulière de « nettoyage » des données et de contrôles d'intégrité rigoureux avant d'effectuer une analyse supplémentaire des données (201).

- Valeur – La raison pour laquelle nous devrions nous soucier du « Big Data ». Alors que les mégadonnées ont le potentiel d'élargir considérablement les connaissances pour de nombreuses conditions cliniques dans lesquelles la collecte de données prospectives et structurées est prohibitive en termes de temps et de coût. De cet aspect, un accord au sein des communautés de recherche biomédicale concernant les meilleures pratiques en matière de collecte, de stockage, de contrôle de qualité et d'analyse des données peut améliorer la valeur des grands ensembles de données. Les méthodologies standardisées permettent d'utiliser les données à plusieurs reprises pour répondre à différentes questions, ainsi que de faire des comparaisons plus directes entre les résultats de différentes cohortes.

En fait, l'utilisation des mégadonnées dans la recherche translationnelle présente plusieurs avantages qui complètent l'expérimentation directe traditionnelle, fondée sur la théorie :

- Les approches de Mégadonnées ont la capacité de générer de nouvelles connaissances plus rapidement que le paradigme traditionnel de la découverte scientifique (202). Ainsi, les technologies modernes à haut débit peuvent générer des milliers à des millions de points de données à partir de milliers d'individus, en quelques heures à quelques jours.
- Les approches de Mégadonnées sont souvent non influencées par des connaissances préalables (203).
- Les approches de Mégadonnées sont holistiques. Les mégadonnées ne se limitent pas à une seule voie, cellule, tissu, individu ou population - elles prennent en compte les maladies à travers les molécules, les cellules, les tissus, les individus, les populations et les expositions environnementales (204).

Les récompenses potentielles de la découverte scientifique axée sur le Big Data se sont traduites par l'enthousiasme et l'application des approches Big Data dans les soins de santé humaine, et son utilisation est à l'origine des révolutions scientifiques. Une meilleure récapitulation de la physiopathologie conduit à une meilleure compréhension de l'étiologie et de la progression de la maladie, ce qui se traduit par une précision améliorée et un diagnostic plus précoce, ainsi que par l'application d'une thérapie personnalisée en fonction de la combinaison unique de gènes, du risque environnemental et du phénotype précis de la maladie. Une identification plus rapide des patients à haut risque se traduit par des décisions de traitement plus rapides. La fusion de mégadonnées entre les institutions peut identifier des maladies rares et des réactions ou des interactions médicamenteuses. Cependant, les « Big Data » ne remplaceront jamais complètement les découvertes scientifiques traditionnelles fondées sur la théorie. Ces approches seront toujours importantes pour la validation des mécanismes biologiques. En plus, la révolution du big data en est encore à son début, avec des défis importants à relever avant de revendiquer pleinement ses avantages(199). Par l'exemple, l'imagerie médicale joue un rôle central dans le diagnostic des maladies en médecine humaine, et les progrès des techniques d'imagerie au cours des dernières décennies ont eu un impact important sur les capacités de diagnostic. Cependant, toutes images

sont un type de données non structurées et, bien que les ordinateurs puissent facilement gérer les données d'image brutes, ils ne peuvent pas facilement reconnaître le type d'image, les annotations sur l'image ou tirer des conclusions à partir des données (205). Ainsi, les défis de l'informatique d'imagerie sont d'acquérir des données d'imagerie avec une fidélité haute qui représentent précisément l'image et de traiter ces données non structurées en un contenu interprétable.

2.1.2 BIOMARQUEURS

Les biomarqueurs sont un élément clé de la médecine de précision et sont utilisés à plusieurs fins, notamment l'évaluation des risques, le dépistage d'une maladie dans sa phase préclinique ou subclinique, le diagnostic de la maladie clinique, la stratification de la maladie, la sélection et/ou le suivi de la réponse au traitement et la prévision de la progression et du pronostic de la maladie (206,207). Par définition, un biomarqueur est une substance ou un processus qui peut être mesuré dans un échantillon biologique dont la fiabilité est corrélée à l'état pathologique et / ou à la réponse clinique d'un patient (208). Les biomarqueurs sont des molécules couramment importantes, telles que l'ADN, l'ARN, les métabolites ou les protéines/peptides qui circulent dans le sang ou dans les tissus. Cependant, les biomarqueurs peuvent également être basés sur d'autres modalités de diagnostic telles que l'imagerie structurale (206). Avant l'apparition de la maladie, l'évaluation des risques avec des biomarqueurs est effectuée dans le but de stratifier les individus en groupes pour identifier les individus les plus susceptibles de bénéficier d'une intervention précoce, de stratégies de prévention ou d'un dépistage diagnostique supplémentaire (207). Les biomarqueurs utilisés pour le dépistage doivent être très sensibles et avoir une spécificité et des valeurs prédictives raisonnables pour être utiles dans la pratique clinique. De plus, les avantages d'une intervention précoce (y compris de meilleurs résultats pour la maladie) devraient l'emporter sur les coûts (et les risques) de la réalisation du test de dépistage (207). Après l'apparition des signes cliniques, des biomarqueurs diagnostiques sont utilisés pour identifier définitivement la cause de la maladie. Ils sont également souvent utilisés pour déterminer la gravité de la maladie sur la base de mesures quantitatives. Au fil du temps, les biomarqueurs peuvent être utilisés pour déterminer la progression de la maladie et, finalement, prédire le pronostic dans une population clinique définie (207).

2.1.3 SYSTEME D'AIDE AU DIAGNOSTIC PAR ORDINATEUR

La vision par ordinateur, domaine scientifique présentant un intérêt particulier pour les radiologues, partage un certain nombre d'objectifs avec l'apprentissage automatique. L'objectif est de permettre aux machines d'analyser, de traiter et de comprendre le numérique pour automatiser les tâches que le système visuel humain peut effectuer(186).

Le diagnostic assisté par ordinateur (DAO) ou CAD fait référence aux outils qui ont été développés pour aider à la lecture d'images. Les systèmes CAD sont généralement des systèmes autonomes qui sont désignés par la Food and Drug Administration pour l'utilisation soit dans la détection, soit dans le diagnostic d'une maladie (209). Les premiers succès de CAD ont été les plus importants en imagerie du cancer du sein (210,211). Pour l'imagerie thoracique, l'application la plus importante concerne le diagnostic des nodules pulmonaires. Cela inclut la CAO pour la détection, appelée CADe, et la CAO pour la caractérisation, appelée CADx, utilisée pour évaluer la probabilité de malignité. Certains CADs combinent les deux tâches.

Le système de CAD vise traditionnellement à fournir une réponse unique (c'est-à-dire la présence d'une lésion ou d'un cancer). Bien que la radiomique soit une extension naturelle des systèmes de CAD, elle en est très différente. Cette dernière est explicitement un processus conçu pour extraire un grand nombre de paramètres quantitatifs d'images numériques, placer ces données dans des bases de données partagées, puis extraire les données pour la génération

d'hypothèses, les tests ou les deux. La radiomique est conçue pour développer des outils d'aide à la décision. Par conséquent, il s'agit de combiner des données radiomiques avec d'autres paramètres des patients, le cas échéant, pour augmenter la puissance des modèles d'aide à la décision.

2.1.4 LA SCIENCE AXÉE SUR LES DONNÉES

La science axée sur les Big Data est quantitative et souvent non-influencée par des connaissances antérieures - les données parlent d'elles-mêmes. La biologie axée sur les données considère le système dans son ensemble. Ces approches reconnaissent que la biologie ne se produit pas dans un système contrôlé expérimentalement, mais qu'elle est désordonnée et compliquée. Contrairement à ce que soutiennent de nombreux scientifiques fondés sur des hypothèses (203); La recherche basée sur les données n'est pas entièrement sans hypothèses. Les approches fondées sur les données partent souvent d'une hypothèse large et collectent ou génèrent de grands volumes de données pertinentes pour cette hypothèse. Ces données sont ensuite explorées pour générer des hypothèses plus spécifiques et mécanistiques par induction éliminatoire ou raisonnement abductif. La science axée sur les données supprime l'hypothèse selon laquelle le scientifique dispose d'une connaissance biologique adéquate pour générer les meilleures hypothèses spécifiques et vérifiable. Au lieu de cela, il suppose que la compréhension du scientifique des systèmes biologiques complexes est rudimentaire et, par conséquent, utilise des données pour générer des hypothèses spécifiques. Les recherches fondées sur les données et sur les hypothèses commencent par identifier un problème de recherche intéressant et le processus intellectuel de développement de théories générales concernant le problème sur la base des connaissances antérieures et des efforts de recherche précédents. Dans les deux cas, le chercheur s'emploie à identifier les questions scientifiques auxquelles il faut répondre et formule des hypothèses générales concernant le problème de la recherche. La science axée sur les hypothèses et les données divergent à l'étape suivante où les scientifiques axés sur des hypothèses effectuent une expérimentation exploratoire (par le biais d'expériences de laboratoire ou de déduction logique), puis formulent des hypothèses spécifiques. En revanche, dans la science axée sur les données, les données sont générées (par expérimentation) ou collectées (données d'observation), ce qui représente toutes les configurations possibles de données qui peuvent être pertinentes pour le problème de recherche (203). Pour satisfaire ces exigences, les données sont généralement de grande dimension (impliquant de nombreux paramètres) et accumulent un grand nombre d'observations (212). La collection de données est une étape critique dans laquelle le scientifique doit être en mesure de garantir que la quantité appropriée de données qui saisit les variables pertinentes susceptibles d'affecter le ou les résultats d'intérêt est incluse (212,212). Il est également essentiel que les données soient collectées de manière à minimiser les sources de biais et les facteurs de confusion potentiels.

2.1.5 RADIOMIQUE

La conversion d'images médicales numériques en données exploitables de haute dimension, un processus connu sous le nom de « radiomique », est motivée par le concept selon lequel les images biomédicales contiennent des informations qui reflètent la physiopathologie sous-jacente et que ces relations peuvent être révélées via des analyses quantitatives d'images. Le mot « radiomique », calqué sur la génomique, est apparu pour la première fois dans la littérature en 2012 dans un article publié par Lambin (185). La radiomique est explicitement un processus d'extraire un grand nombre de paramètres quantitatifs d'images numériques, placer ces données dans des bases de données partagées, puis extraire les données pour la génération d'hypothèses, les tests, ou les deux. La radiomique est conçue pour développer des outils d'aide à la décision ; il s'agit ainsi de combiner des données radiomiques avec d'autres paramètres des patients, selon les disponibilités, pour augmenter la puissance des modèles d'aide à la décision (184). Bien que simple d'un point de vue conceptuel, la pratique de la radiomique implique des étapes distinctes qui comprennent : (a)

l'acquisition des images, (b) l'identification des régions/des volumes d'intérêt, la segmentation des volumes, c) le prétraitement de données, (d) extraire et qualifier les paramètres descriptifs ou quantitatifs de l'image, (e) les utiliser pour alimenter une base de données interrogeable, et (f) extraire ces données pour développer des modèles de classifications pour prédire les résultats soit seuls, soit en combinaison avec des informations supplémentaires, telles que démographiques, cliniques, comorbidité, ou des données génomiques(184,213).

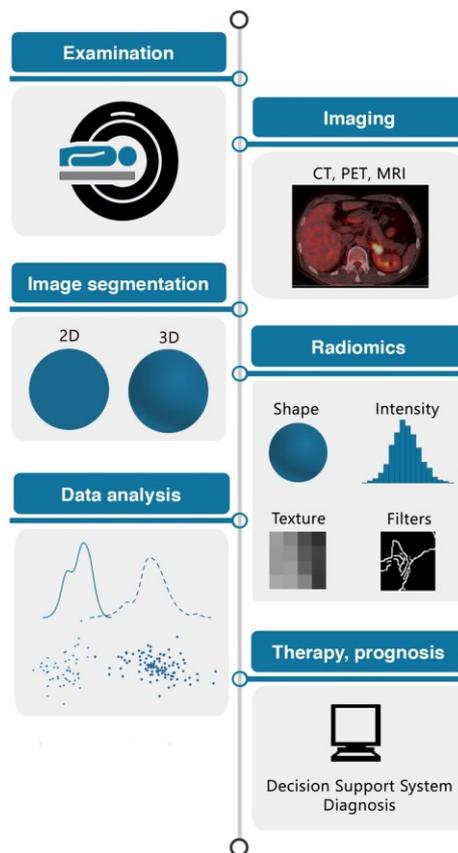


Figure 2-1 Le flux de travail de radiomique.

Figure 2-1 : Illustration schématique du parcours du patient comprenant l'acquisition d'images, l'analyse utilisant la radiomique par une modélisation statistique de haut niveau pour la classification des maladies, le regroupement des patients pour les stratégies thérapeutiques et la stratification des risques individuels en prognostique.

De par sa définition, la radiomique est une méthode d'étude scientifique basée sur les données (ou « data-driven ») dont le but était d'extraire un grand nombre de paramètres, parmi lesquels les plus intéressants statistiquement étaient sélectionnés pour ensuite en déduire une hypothèse scientifique. Cette approche est différente de l'approche scientifique classique basée sur l'hypothèse et qui tend à confirmer ou infirmer une hypothèse préalablement établie (214). Néanmoins, en réalité les approches fondées sur les données partent souvent d'une hypothèse large et collectent ou génèrent de grands volumes de données pertinentes pour cette hypothèse. Ces données sont ensuite explorées pour générer des hypothèses plus spécifiques et mécanistes par induction éliminatoire ou raisonnement abductif. Ainsi, « Big Data » par le pont de radiomique ne remplaceront jamais complètement les découvertes scientifiques traditionnelles fondées sur la théorie ; Ces dernières s'avèrent toujours importantes même plus d'important en jouant le rôle d'orientation initiale devant des mégadonnées. L'article de Lambin (215) ont décrit l'implication de la radiomique dans le système de support de décision cliniques moderne comprenant : la découverte, la sélection, la collection et la préparation des données, le développement, la validation et mise en œuvre du (des) modèle (s), l'évaluation de l'utilité clinique et finalement

raffinement grâce à la répétition continue du processus (des protocoles de contrôle et d'assurance qualité sont nécessaires tout au long du processus).

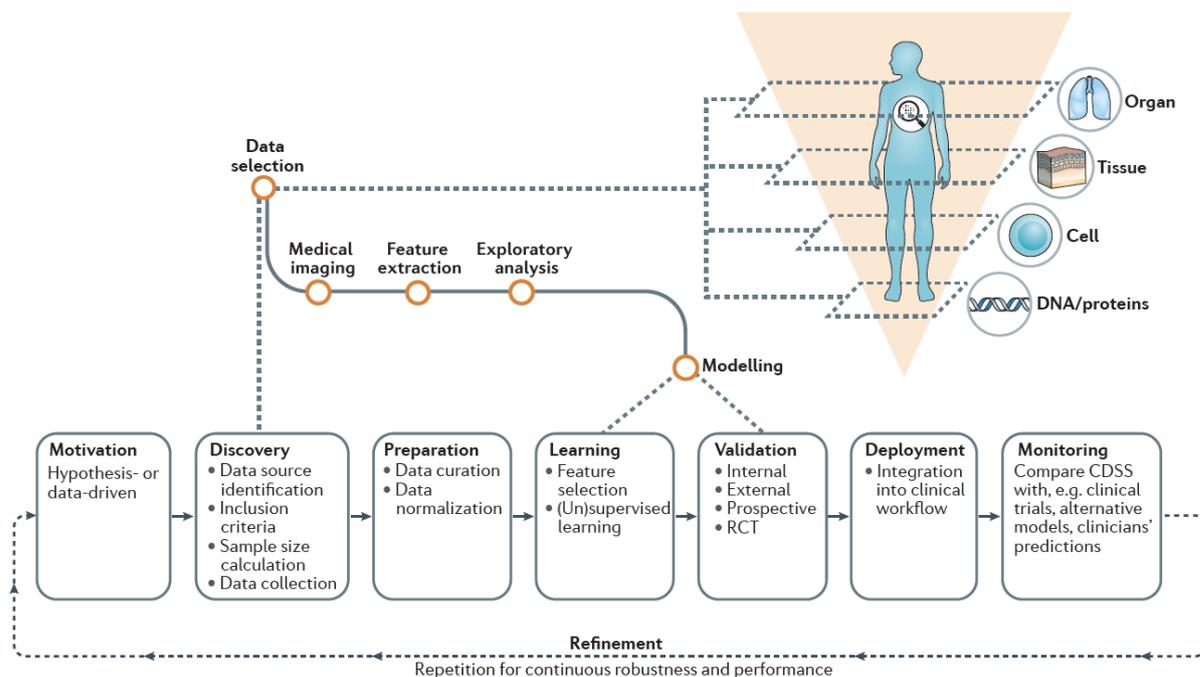


Figure 2-2: Vue d'ensemble des processus méthodologiques du service de santé d'apprentissage rapide et l'intégration de flux de travail radiomique(215).

Le cœur de la radiomique est l'extraction de données de paramètres de haute dimension pour décrire quantitativement les attributs des volumes d'intérêt. En pratique, les paramètres « sémantiques » et « agnostiques » sont les deux types de paramètres extraites en radiomique(184). Paramètres sémantiques sont celles qui sont couramment utilisées par les radiologues pour décrire les lésions telles que la forme, le volume et le diamètre. Avec la prescience que les paramètres sémantiques ont une valeur pronostique, les premières recherches en radiomique ont été développées à partir des lexiques de radiologie, tout comme BI-RADS, PI-RADS et Lung-RADS tentent de le faire. Les paramètres radiomiques agnostiques sur une image sont des descripteurs quantitatifs extraits mathématiquement, qui ne font généralement pas partie du lexique des radiologues. Ceux-ci peuvent être divisés en produits descripteurs de premier, deuxième ordre ou ordre supérieur, la forme. Les descripteurs de premier ordre décrivent la distribution des valeurs des voxels individuels sans se soucier des relations spatiales. Ce sont généralement des méthodes basées sur l'histogramme et réduisent une région d'intérêt à des valeurs uniques pour la moyenne, la médiane, le maximum, le minimum et l'uniformité ou le caractère aléatoire (entropie) des intensités sur l'image, ainsi que l'asymétrie (asymétrie) et l'aplatissement de l'histogramme des valeurs. Les descripteurs statistiques du second ordre sont généralement décrits comme des paramètres de « texture » ; Ils décrivent les interrelations statistiques entre les voxels avec des valeurs de contraste similaires ou différentes.

Les paramètres sont extraits et analysés pour être utilisés à des fins de quantification (184), du pronostic de réponse au traitement, ou de la corrélation avec d'autres biomarqueurs cliniques ou génomique. Il existe de nombreuses applications de la radiomique en oncologie thoracique, telles que la discrimination de l'adénocarcinome du carcinome épidermoïde (216), la prédiction de l'invasivité de l'adénocarcinome pulmonaire (217) ou la mutation du récepteur du facteur de croissance épithélial (EGFR) (218), reliant le «phénotype radiomique» de la tumeur et le génotype de la tumeur(219), ou prédiction de la réponse au traitement (220). Cependant, une étude récente évaluant 77 articles a rapporté un score moyen de qualité radiomique, une métrique évaluant la validité et l'exhaustivité des études radiomiques, de seulement 26,1% et a conclu que la qualité

scientifique globale des études radiomiques était insuffisante (221). Pour évaluer la qualité des études radiomiques, Lambin et al. propose le score de qualité radiomique (RQS) par 36 points qui recouvrent l'intégralité des étapes de cette approche.

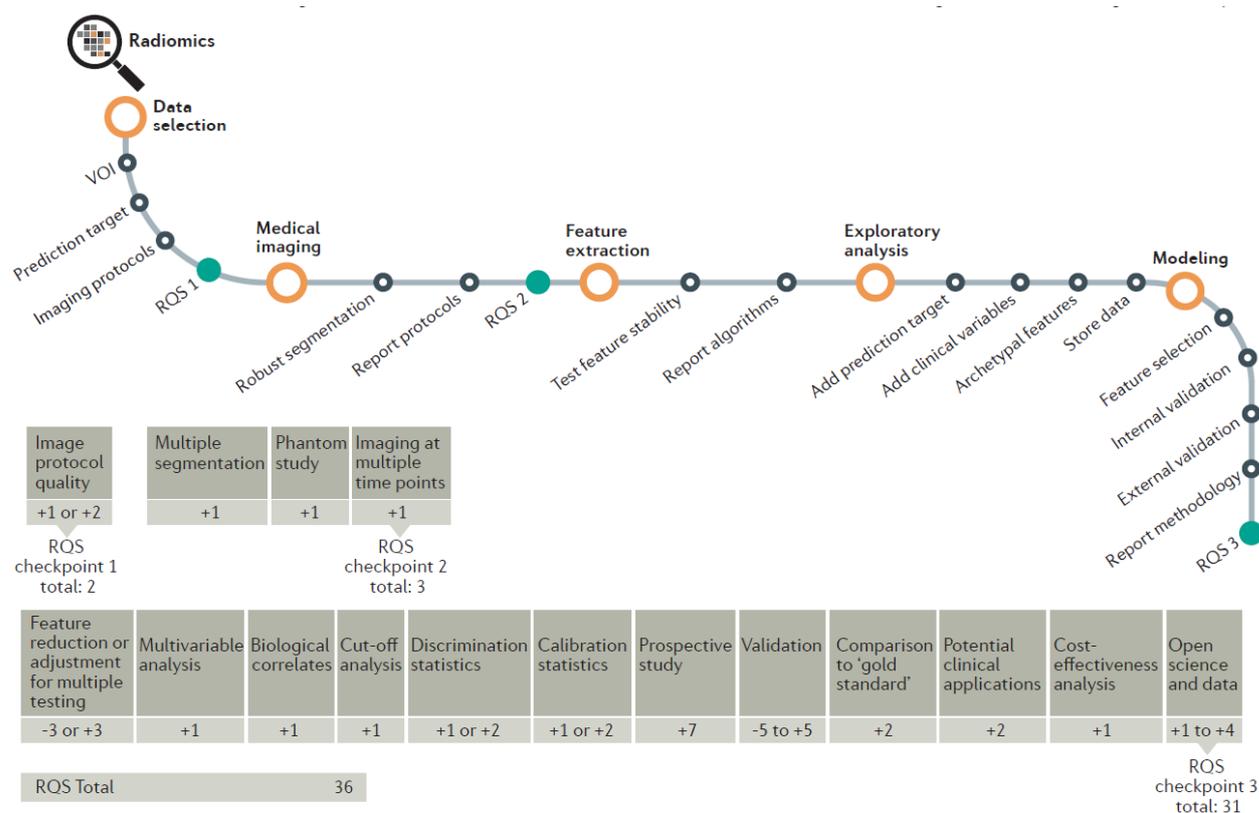


Figure 2-3: Organigramme illustrant le flux de travail de la radiomique et l'application du RQS (215). Le flux de travail comprend les étapes nécessaires dans une analyse radiomique. Le RQS récompense et pénalise à la fois la méthodologie et les analyses d'une étude, encourageant ainsi les meilleures pratiques scientifiques. RSQ, score de qualité radiomique; VOI, volume d'intérêt.

2.2 LES ALGORITHMES D'APPRENTISSAGE AUTOMATIQUE

2.2.1 TYPE DE L'ALGORITHME

Les algorithmes d'apprentissage automatique peuvent être classés en trois groupes principaux : les algorithmes supervisés, semi-supervisés ou non supervisés. Les algorithmes basés sur la supervision reposent sur des échantillons avec des annotations fournies par des experts cliniques, qui seront utilisés pour la formation. Les algorithmes d'apprentissage supervisé peuvent être formés pour des tâches de classification, telles que la présence ou l'absence de maladie ou d'anomalie, ou pour des tâches de régression, par exemple pour fournir un score de gravité ou un pronostic.

Récemment, des méthodes semi-supervisées qui combinent des données annotées et non annotées ont vu le jour. Dans ce cadre, les algorithmes apprennent progressivement grâce à une meilleure exploitation des données non annotées. L'apprentissage par renforcement est un exemple typique d'apprentissage semi-supervisé(146).

A l'inverse, les algorithmes basés sur un apprentissage non supervisé n'impliquent pas d'intervention humaine. Le regroupement est l'exemple le plus représentatif, où l'objectif est de regrouper les échantillons en sous-populations homogènes, comme par exemple, pour identifier différents phénotypes de la maladie pulmonaire obstructive chronique (BPCO)(222). Les performances des algorithmes non supervisés sont souvent inférieures à celles obtenues avec des

techniques supervisées. Toutefois, l'analyse supervisée est limitée par une subjectivité intégrée, et sans un contrôle de qualité approprié, l'apprentissage automatique apprendra à copier les erreurs humaines. Ces raisons expliquent la tendance actuelle d'un intérêt renouvelé sur l'apprentissage non supervisé de la communauté de l'apprentissage automatique. Les algorithmes non-supervisés sont attrayantes car elles permettent la formation du réseau avec la richesse des données non étiquetées disponibles dans le monde(149).

2.2.2 BASE DE DONNÉES

Pour les méthodes d'apprentissage automatique, y compris le deep-learning, la qualité des données est essentielle et pourrait être encore plus importante que l'algorithme d'apprentissage lui-même. Il garantit la capacité du modèle à fonctionner aussi bien sur des cas non vus pendant l'entraînement. Pour obtenir un modèle généralisable, il est important de disposer d'un ensemble de données représentatif de la maladie et également représentatif des différentes techniques d'acquisition. En radiologie, les ensembles de données doivent inclure les différents protocoles d'acquisition et les différentes formes de la maladie évaluée et inclure également des examens de sujets indemnes de la maladie. Un modèle de détection de la fibrose pulmonaire doit être formé à l'aide d'un ensemble de données reflétant l'hétérogénéité des profils de fibrose pulmonaire, mais comprenant également des TDM normaux et des images tomodensitométriques acquises sur divers équipements de tomodensitométrie. Si l'ensemble de données d'entraînement ne contient qu'un modèle de fibrose unique ou des acquisitions toutes réalisées sur la même unité de tomodensitométrie avec le même protocole de reconstruction, le risque que le modèle soit mal généralisable est élevé.

L'ensemble de données est généralement divisé en trois sous-catégories : entraînement, validation et test. L'ensemble d'entraînement - correspondant généralement à 60% de la base de données - est utilisé pour entraîner des variantes de versions du modèle avec des conditions d'initialisation et des hyperparamètres différents (ceux-ci seront définis dans la section suivante). Une fois les modèles entraînés, leurs performances sont évaluées à l'aide d'autre 20% des données, constituant ce que l'on appelle le jeu de données de validation. Le modèle avec les meilleures performances sur l'ensemble de données de validation est sélectionné. Ce modèle est finalement évalué à l'aide de 20% restant d'échantillons, qui n'ont jamais été utilisés auparavant et compose le jeu de données de test.

Une alternative au fractionnement de l'ensemble de données entre l'entraînement et la validation est la méthode de validation croisée k-fold qui permet l'entraînement et la validation en utilisant l'ensemble de données. Il consiste à diviser les ensembles d'entraînement et de validation en plusieurs divisions aléatoires, puis à utiliser différentes combinaisons pour l'apprentissage et la validation. La performance moyenne du modèle sur ces divisions est prise en compte pour juger de la performance et de l'acceptabilité du modèle. Cette approche est particulièrement utile lorsque le nombre de cas est limité.

2.2.3 SUR-APPRENTISSAGE ET SOUS-APPRENTISSAGE DES MODÈLES

Le sur-apprentissage est la situation dans laquelle le modèle entraîné fonctionne très bien sur l'ensemble de données d'entraînement mais échoue sur l'ensemble de test. Le sur-apprentissage se produit lorsque la performance du modèle continue de s'améliorer dans la cohorte d'entraînement mais diminue dans la cohorte de validation. En d'autres termes, le modèle génère des prédictions précises sur l'ensemble d'entraînement, mais ne parvient pas à les reproduire sur d'autres nouveaux cas. Cela peut être observé lorsque l'ensemble d'entraînement n'est pas bien équilibré ou lorsque le nombre d'échantillons n'est pas suffisant. Dans cette situation, il peut arriver que l'algorithme trouve une association d'entités et la considère comme pertinente pour le résultat,

alors que ce n'est que le résultat de combinaisons d'entités fortuites apprises à partir d'un ensemble de données non représentatif. Cette association disparaîtrait lors de l'utilisation d'un échantillon plus grand ou différent. Les problèmes de sur-apprentissage sont courants avec les algorithmes du deep-learning contenant de nombreuses couches générant de nombreuses variables à apprendre (de plusieurs centaines à plusieurs millions) à partir de petits ensembles d'entraînement.

Un autre problème qui peut être observé pendant l'entraînement est le sous-apprentissage. Cela se produit lorsque le modèle ne parvient pas à s'adapter aux ensembles d'entraînement et de validation. Les raisons du sous-apprentissage peuvent être multiples. En présence de plusieurs sous-populations au sein de l'ensemble d'entraînement, les modèles avec un nombre insuffisant de paramètres ne parviendront pas à englober la population entière. Une autre explication possible du sous-apprentissage concerne la nature du modèle qui a été choisi pour la prédiction. Par exemple, lorsque la notion de temps est critique pour le diagnostic (temps de rehaussement de l'agent de contraste), un modèle qui ignore cette information échouera le plus inévitablement même si le nombre de paramètres est suffisant.

2.2.4 STRATÉGIE D'OPTIMISATION

Le terme hyperparamètre fait référence à tous les paramètres qui sont définis avant l'entraînement de l'algorithme, par opposition à ceux qui dériveront de l'apprentissage. Le nombre et le type varient en fonction de l'architecture de deep-learning. Les hyperparamètres incluent le nombre de couches, le taux d'apprentissage qui dépend de la fonction de perte et de la stratégie d'optimisation.

La fonction de perte est un concept important à comprendre. Il correspond aux métriques utilisées par l'algorithme lors de l'entraînement pour tester ses performances. Il quantifie l'écart entre la prédiction par l'algorithme et la vérité-terrain donnée par l'annotation / étiquette expert. L'objectif de tout algorithme de deep-learning est toujours de minimiser sa fonction de perte, jusqu'à ce que l'écart entre la prédiction par le réseau et la vérité-terrain disparaisse. La fonction de perte varie en fonction de la tâche qui est adressée, comme la perte d'indice de similarité de Dice (223) pour les tâches de segmentation ou de LogLoss pour les tâches de détection et de classification.

Plusieurs stratégies peuvent être utilisées pour optimiser la fonction de perte pendant l'entraînement. Le plus couramment utilisé est la descente de gradient stochastique (224). Les méthodes de descente de gradient reposent sur un processus itératif où chaque itération permet de se rapprocher du modèle optimal. La descente de gradient stochastique permet une perturbation aléatoire du modèle qui pourrait instantanément dégrader ses performances mais finalement converger vers une meilleure solution.

Le taux d'apprentissage et le nombre d'époques sont également des paramètres d'optimisation importants. Le taux d'apprentissage contrôle le degré d'amélioration du réseau entre deux itérations. De faibles taux d'apprentissage garantissent des améliorations, mais d'une importance marginale. Les taux d'apprentissage élevés font référence à des modèles plus instables où l'amélioration d'une itération à la suivante pourrait être plus significative, associée au risque de dégradation de la performance globale. L'époque est un concept différent, qui fait référence au nombre de fois où l'ensemble de l'entraînement a été revisité pour mettre à jour les paramètres du modèle. Un plus grand nombre d'époques garantissent de meilleures performances sur l'ensemble d'apprentissage, au prix d'une complexité croissante de calcul ainsi que du risque de sur-apprentissage, en sélectionnant des fonctionnalités qui ne sont spécifiques qu'à l'ensemble de données d'apprentissage et sont mal généralisables.

2.3 RADIOMIQUE EN PNEUMOLOGIE

2.3.1 DATA SELECTION ET STANDARDISATION DES PROTOCOLES D'IMAGERIE

Les analyses radiomiques commencent par le choix d'une cible de prédiction - l'événement que l'on souhaite prédire et d'un protocole d'imagerie (215). Ainsi le traitement des mégadonnées nécessite une large hypothèse avant que grands volumes de données offrent à leur tour plus précise hypothèse par des méthodes d'induction éliminative ou de raisonnement abductif (225). Souvent on part de la recherche d'un biomarqueur pour un problème cliniquement important. Avant l'apparition de la maladie, le but de biomarqueur est de stratifier les individus en groupes pour identifier les individus les plus susceptibles de bénéficier d'une intervention précoce, de stratégies de prévention ou d'un dépistage diagnostique supplémentaire(207). Après l'apparition des signes cliniques, des biomarqueurs diagnostiques sont utilisés pour identifier définitivement la cause de la maladie ou pour déterminer la gravité de la maladie sur la base de mesures quantitatives(207). À partir de ce biomarqueur et en fonction de ses critères pour son application clinique, la modalité d'imagerie serait choisie. En imagerie thoracique, la tomодensitométrie joue un rôle central dans le diagnostic des maladies pulmonaire et grâce aux progrès des techniques d'imagerie au cours des dernières décennies, la TDM a eu un impact important sur les capacités de diagnostic en augmentant à la fois les détails dans lesquels le corps peut être imagé et en ajoutant récemment des indices fonctionnels en comparant avec la radiographie conventionnelle. Pour ces raisons, nous avons choisi TDM comme la modalité de premier choix pour l'investigation des changements structurels pulmonaires.

L'importance d'utiliser des protocoles d'imagerie normalisés pour éliminer la variabilité de confusion inutile dans les études radiomiques est reconnue (226); Cependant, les protocoles d'imagerie non standardisés sont courants. Par exemple, Galavis et al (2010) ont évalué la variabilité de 50 caractéristiques radiomiques TEP en raison de différents modes d'acquisition, tailles de matrice, largeurs de post-filtrage, algorithmes de reconstruction et nombres d'itérations (227). Parmi ces caractéristiques, quarante se sont avérées avoir une variabilité substantielle avec une différence relative de 30%. Seules quatre caractéristiques, y compris l'entropie, l'énergie dérivées de l'histogramme d'intensité (caractéristique de premier ordre) ; le coefficient de corrélation maximal (GLCM) (caractéristique de deuxième ordre) et « low-gray level run emphasis » (caractéristique d'ordre supérieur) se sont avérées avoir une variabilité de moins de 5%. Par conséquent, la reproductibilité et la comparabilité des études de radiomiques ne peuvent être obtenues que par une normalisation large des protocoles d'imagerie intra- et inter-institution.

Premièrement, une diminution de l'épaisseur de la tranche réduit les statistiques de photons dans une tranche (à moins que mAs ou kVp ne soient augmentés en conséquence), augmentant ainsi le bruit de l'image (228). Le champ de vision axial et la taille de la matrice de reconstruction déterminent la taille des pixels et donc l'échantillonnage spatial dans le plan axial, ce qui a un impact sur la description de l'hétérogénéité. La réduction de la taille des pixels augmente le bruit de l'image (lorsque les autres paramètres restent inchangés), mais augmente la résolution spatiale. Zhao et coll. ont évalué l'effet combiné de 3 épaisseurs de coupe de TDM différentes et de 2 algorithmes de reconstruction de TDM différents. De nombreuses fonctionnalités qui avaient été répétables dans des conditions de test-retest sont devenues irréproductibles en ce qui concerne l'épaisseur de coupe modifiée et les paramètres de reconstruction d'image, les caractéristiques de premier ordre et les métriques de forme étant moins sensibles que les caractéristiques texturales en analyse de radiomique(229).

Deuxièmement, lorsqu'on considère l'acquisition de TDM en spirale, le pas est une variable qui influence le bruit de l'image, ce qui rend difficile la comparaison entre différents scanners et fournisseurs. Par conséquent, des acquisitions non spirales (axiales) sont nécessaires pour ces

comparaisons. De même, des conditions cliniques, telles que la présence d'artefacts dus à des prothèses métalliques, peuvent affecter la qualité de l'image et nuire à l'analyse quantitative (230). De plus, la quantification électronique de la densité exprimée en unités Hounsfield peut varier avec l'algorithme de reconstruction ou l'étalonnage du scanner (231).

Ensuite, Fave et al ont testé l'effet de différents filtres de prétraitement d'image, tels que le rééchantillonnage de la profondeur de bits et les filtres, sur les caractéristiques radiomiques en TDM. La corrélation avec le volume de la tumeur et l'utilisation de filtres de prétraitement ont augmenté le risque que les caractéristiques soient significatives sur une analyse univariée par rapport au résultat, mais on ne sait pas si celles-ci conservent leur valeur prédictive dans un ensemble de validation indépendante (232). Trois études se sont concentrées exclusivement sur la répétabilité en utilisant les mêmes images de test-retest (233–235). Ils ont toujours trouvé que les métriques de forme et les caractéristiques de premier ordre étaient hautement répétables, mais il n'y avait pas de consensus sur les caractéristiques texturales répétables.

Ainsi, pour étudier en détail les effets des paramètres d'acquisition et des algorithmes de reconstruction sur les caractéristiques radiomiques, des études sophistiquées sont nécessaires.

2.3.2 ANNOTATION DES DONNÉES EN IMAGERIE THORACIQUE

Méthodes d'annotation :

Après acquisition des images, l'étape suivante consiste à définir dans quelle région l'extraction de données pour lesquelles l'analyse va être réalisée. Il existe différents types d'annotations en fonction de la tâche que l'algorithme cherche à traiter.

Les tâches de segmentation manuelle nécessitent le plus haut niveau d'annotation qui consiste à contourner / délimiter les anomalies sur chaque image. La segmentation manuelle par délimitation, en généralement faite en utilisant un logiciel comme ITK-SNAP, permet de se rapprocher de la pratique radiologique quotidienne. Ce type d'annotation permet de construire des algorithmes plus précis mais est fastidieux et prend du temps. Ces ensembles de données sont généralement plus petits et plus difficiles à générer. Il est bien reconnu que la variabilité interopérateur des tumeurs segmentés manuellement est élevée(236,237). Ainsi, l'analyse des lésions cancéreuses nécessite souvent l'intervention de l'opérateur en raison de l'hétérogénéité morphologique et de contraste inter- et intra-sujet lors de l'examen initial.

La segmentation automatique fait appel à des algorithmes qui détectent automatiquement les contours des lésions, et sont particulièrement performants dans le poumon. La segmentation des structures normales, telles que les éléments squelettiques et les organes, peut désormais être réalisée avec une automatisation complète. Étant donné que les annotations de grands ensembles de données sont générées par des extractions automatisées à partir des rapports, il est important de tester l'exactitude de l'étiquetage automatisé, ce qui est généralement effectué par comparaison avec les annotations des radiologues sur un sous-ensemble de radiographies(238). Une partie de nos travaux précédents consiste à développer des architectures de réseaux de deep-learning pour l'évaluation automatisée de l'étendue de la maladie pulmonaire interstitielle chez les patients atteints de sclérose systémique (239).

En intermédiaire entre les deux méthodes, la segmentation semi-automatique repose une délimitation en premier temps par un logiciel ; qui est ensuite ajusté par l'utilisateur(240,241). Parmar et coll. ont montré une meilleure reproductibilité des paramètres de texture obtenue par la méthode de segmentation semi-automatique par rapport à la segmentation manuelle seule (241). La segmentation pourra être réalisée plusieurs fois par un même observateur et par plusieurs observateurs afin d'évaluer la reproductibilité intra- et inter-observateur des paramètres extraits.

Application du deep-learning en tâche d'annotation d'image

Récemment, l'apprentissage automatique a attiré beaucoup d'attention avec des rapports encourageants sur une variété de problèmes ouverts tels que la détection d'objets dans les images (242), la super résolution d'images (243), la détection d'action dans la vidéo, le traitement du langage naturel (244). À la suite de ces avancées très prometteuses dans les domaines susmentionnés, des progrès lents et confiants ont également été réalisés dans les domaines de la médecine assistée par ordinateur.

En particulier, plusieurs architectures de deep-learning ont été évaluées ou développées spécifiquement pour la segmentation et/ou la classification sur des images médicales. Ces architectures sont constituées d'une séquence de couches principalement optimisées sur la tâche qu'elles doivent adresser. Architectures qui traitent de la classification sont généralement constituées de couches entièrement convolutives formant des architectures comme AlexNet (156), DenseNet (245) ou de nombreuses autres architectures proposées récemment dans la littérature. D'un autre côté, les architectures qui traitent des problèmes de segmentation sont généralement entièrement convolutives en adaptant une stratégie d'encodeur-décodeur tel que U-Net (246) pour la segmentation en 2D et V-Net (247) pour la segmentation des volumes en 3D. De plus, récemment, les réseaux de neurones récurrents (RNR) ont également été évalués par plusieurs auteurs (177).

La littérature existante peut être classée en deux catégories distinctes, les méthodes non basées sur modèle et les méthodes basées sur modèle. Les méthodes basées sur modèle supposent que la distribution géométrique préalable de la structure à segmenter peut-être apprise et modélisée avec des approches sous spatiales (par exemple des formes actives). Des modèles probabilistes ou graphiques et des approches basées sur des atlas étant représentants de cette catégorie(248). Ces méthodes fonctionnent bien avec les structures anatomiques normales où la notion de répétabilité a du sens mais ne parviennent pas à capturer la variation de la maladie qui est en général aléatoire. D'autre part, les approches non basées sur modèle reposent uniquement sur l'espace d'observation combinant des vraisemblances d'images avec différentes techniques de classification. Les ConvNets fournissent actuellement des résultats de pointe sur la segmentation d'images de l'analyse d'images médicales. Ces réseaux peuvent adapter une approche basée sur les patches, où le réseau utilise un patch extrait autour du voxel qui doit être segmenté ou une prédiction dense au niveau des pixels où le réseau produit la segmentation de chaque voxel produisant directement le masque de segmentation.

L'approche basée sur les patches peut également être utilisée avec l'apprentissage automatique classique et cela a été fait avec des résultats divers (189,249,250). Anthimopoulos et al. utilisant ConvNet pour la segmentation ILD avec une approche basée sur les patches, ont rapporté une précision de 86 % pour la classification des voxels pulmonaires en 7 classes(251). Pour obtenir ces résultats, ils ont utilisé un algorithme inspiré de l'architecture VGG mais avec une taille de noyau encore plus réduite et seulement 5 couches convolutives. Cela a été encore amélioré par Walsh et al. qui a utilisé un ConvNet composé de quatre couches de convolution et de deux couches entièrement connectées pour la segmentation ILD basée sur les patches (165). Les auteurs ont rapporté une précision allant de 85% à 96% selon leur répartition des données et ont montré la supériorité de ConvNet sur SVM qui offrait une précision de 77 à 90%. Le principal inconvénient de l'approche basée sur les patches est de ne pas prendre en compte les informations contextuelles telles que la localisation sous-pleurale et la prédominance basale habituellement observées dans l'ILD.

La prédiction dense au niveau des pixels basée sur l'image entière a été popularisée depuis l'introduction de réseaux entièrement convolutifs (fCNN-full ConvNet)(252). La principale limitation

de l'utilisation de ConvNet pour la prédiction dense était les couches de « pooling » de l'architecture de l'encodeur qui permettent de prendre en compte le contexte mais diminuent progressivement la résolution de l'image. En utilisant des connexions par saut (« skip connections » en anglais) pendant le processus de suréchantillonnage (« up-sampling process » en anglais), fCNN permet de récupérer la résolution spatiale pour obtenir une prédiction à pleine résolution malgré l'utilisation de plusieurs étapes de « pooling ». Le fCNN a été l'une des premières architectures encodeur-décodeur. Les connexions entre l'encodeur et les parties décodeurs ont été encore améliorées, par exemple par le transfert d'indices maxpooling de l'encodeur aux parties décodeurs dans SegNet (253) ou par une connexion directe entre les couches convolutives opposées et les couches convolutives en expansion utilisant la concaténation dans U-Net (246) (Figure 2-4). U-net, qui surpassait les précédentes architectures d'encodeur-décodeur, est devenu très populaire pour l'analyse d'images médicales. Une variante 3D du U-net, appelée V-net a été proposée par Cicek et coll (247).

Une alternative à l'architecture codeur-décodeur est l'utilisation de convolutions auriculaires (également appelées convolutions dilatées) dans les architectures Deeplab. Cette architecture a été implémentée avec succès par Anthimopoulos et coll. pour les segmentations des ILD et elle leur a permis d'améliorer la précision de validation croisée équilibrée de 72 % (251) lors de l'utilisation de leur précédente approche basée sur les patches à 82 % (254), tout en réduisant le temps d'inférence de 237 à 58 millisecondes.

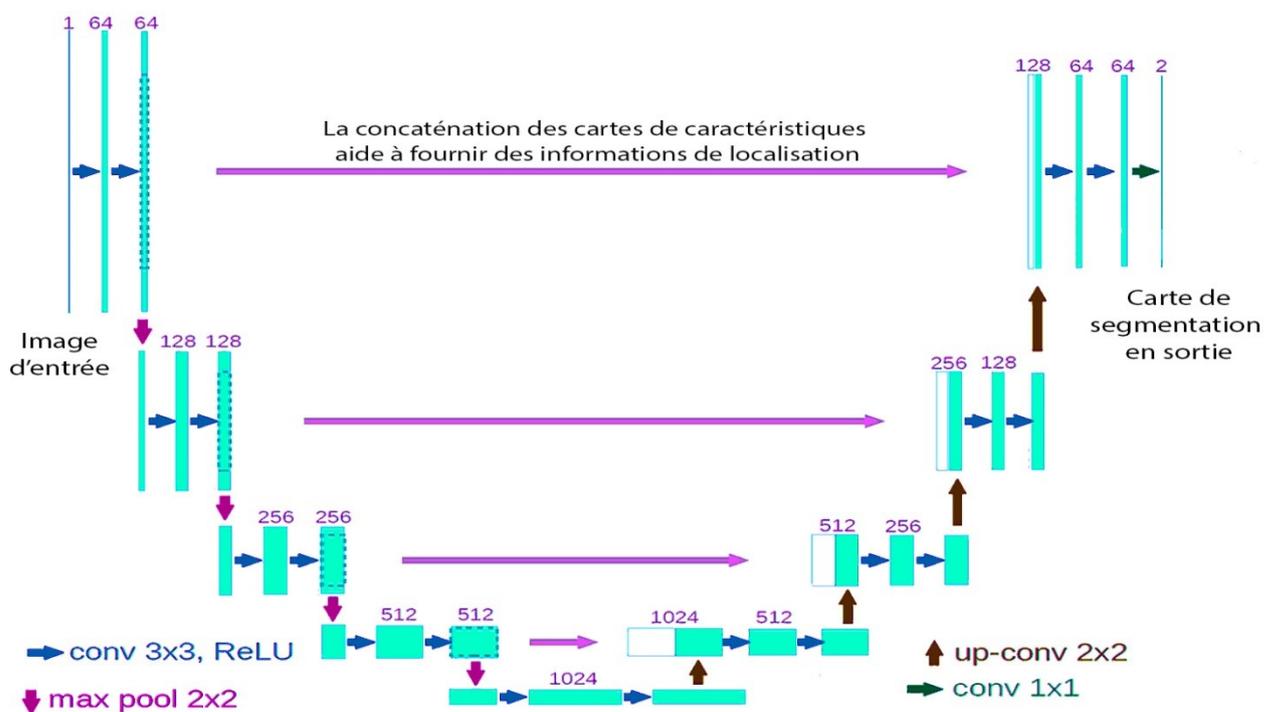


Figure 2-4: Architecture U-net (exemple pour 32x32 pixels).

Figure 2-4 : Chaque boîte verte clair correspond à une carte des caractéristiques multicanale. Le nombre de canaux est indiqué sur le dessus de la boîte. Les cases blanches représentent les cartes de caractéristiques en concaténation. Les flèches indiquent les opérations différentes.

Plus récemment, les travaux de Chassagnon et coll. ont proposé un algorithme de deep-learning « AtlasNet » pour l'évaluation automatique de l'étendue de la maladie pulmonaire interstitielle (ILD) liée à la sclérose systémique (ScS) sur des images de tomodensitométrie thoracique (239). En bref, l'architecture AtlasNet forme un certain nombre de réseaux de neurones convolutifs profonds (n) et chacun d'eux utilise une anatomie prédéfinie sur laquelle tous les cas de

patients en formation sont mappés via un enregistrement élastique, ce qui entraîne une augmentation naturelle des données. Dans cette étude, les auteurs ont augmenté le réseau AtlasNet grâce à une double architecture d'autodécodage. Ce choix d'architecture a été motivé par la nécessité de garantir des résultats de segmentation de la maladie anatomiquement plausibles, ce qui a été obtenu grâce à l'introduction d'interdépendances spatiales entre les poumons et les classes de maladie, prédites par deux parties différentes du décodeur. Leur architecture proposée pour ce problème est basée sur un réseau similaire à SegNet (Computer Vision and Robotics Group, Université de Cambridge, Cambridge, Angleterre) (253), qui est composé d'une partie d'encodage et de deux branches de décodage : une pour la segmentation pulmonaire et l'autre pour la segmentation de la maladie. Cinq blocs convolutifs sont inclus, chacun contenant deux successions de couches convolutives, normalisées par lots et rectifiées d'unités linéaires. Le premier bloc augmente le nombre de canaux d'entrée à 64, tandis que le reste des blocs augmente les canaux à deux fois leur taille, résultant en 1024 plans à la fin de l'encodeur. Des couches de mise en commun maximale sont également distribuées à la fin de chaque bloc convolutif, à l'exception du dernier, ramenant les volumes d'entrée à un quart de leur résolution d'origine. Ensuite, les caractéristiques produites sont données en entrée à deux branches de décodage identiques, l'une chargée de détecter la catégorie de maladie et l'autre chargée de détecter la catégorie de poumon. Semblable à l'encodeur, chaque partie de décodage implique des blocs convolutifs, mais cette fois, des couches de suréchantillonnage sont utilisées à la place des couches de mise en commun maximale pour ramener les vecteurs de caractéristiques à leurs dimensions d'origine. L'architecture est illustrée à la Figure 2-5 . Pour l'entraînement d'AtlasNet, six réseaux de neurones à convolution profonde différents avec une architecture d'autoencodeur Seg-Net sont utilisés. Après l'entraînement des six réseaux de neurones convolutifs, AtlasNet a combiné leurs prédictions en utilisant une stratégie d'ensemble et a appliqué un principe de vote à la majorité simple.

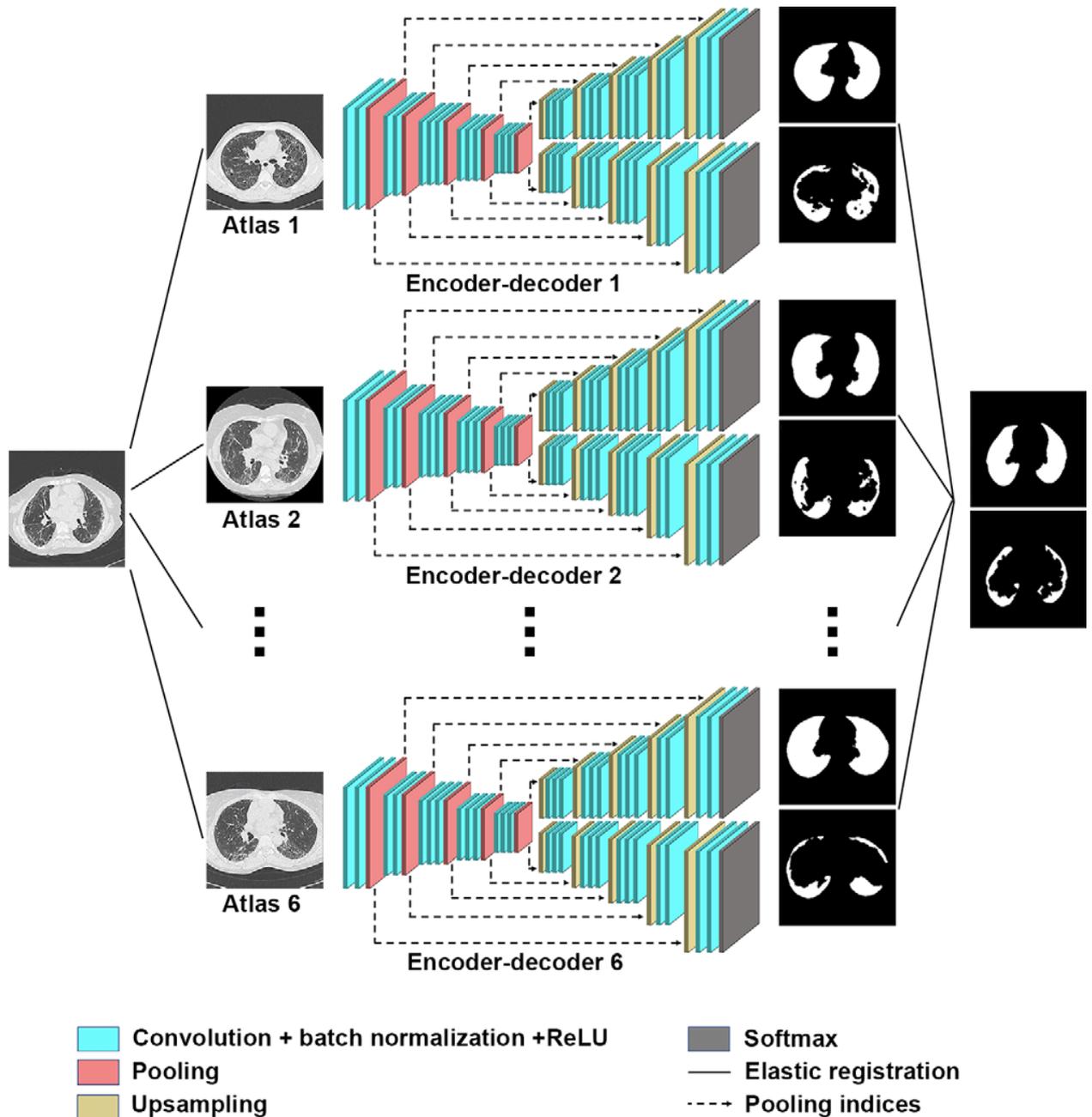


Figure 2-5: Architecture du framework AtlasNet. ReLU = unité linéaire rectifiée (239).

Dans ces travaux, le coefficient de similarité Dice (DSC) a été calculé dans l'ensemble de données de test pour évaluer l'accord entre les contours des radiologues et entre chaque radiologue et les contours générés par les deux algorithmes de deep-learning différents (AtlasNet et U-Net). Le DSC est une statistique utilisée pour comparer la similarité de deux segmentations. Il est couramment utilisé en segmentation d'images, en particulier pour des comparaisons avec des masques de référence dans des applications médicales. Il est calculé selon la formule suivante :

$$DSC = \frac{2 \times |S_1 \cap S_2|}{|S_1| + |S_2|}$$

Où S_1 et S_2 sont les aires de la première et de la deuxième segmentation, respectivement (223) . Un DSC de 1 correspond à une correspondance parfaite entre deux segmentations, alors qu'un DSC de 0 signifie aucun chevauchement.

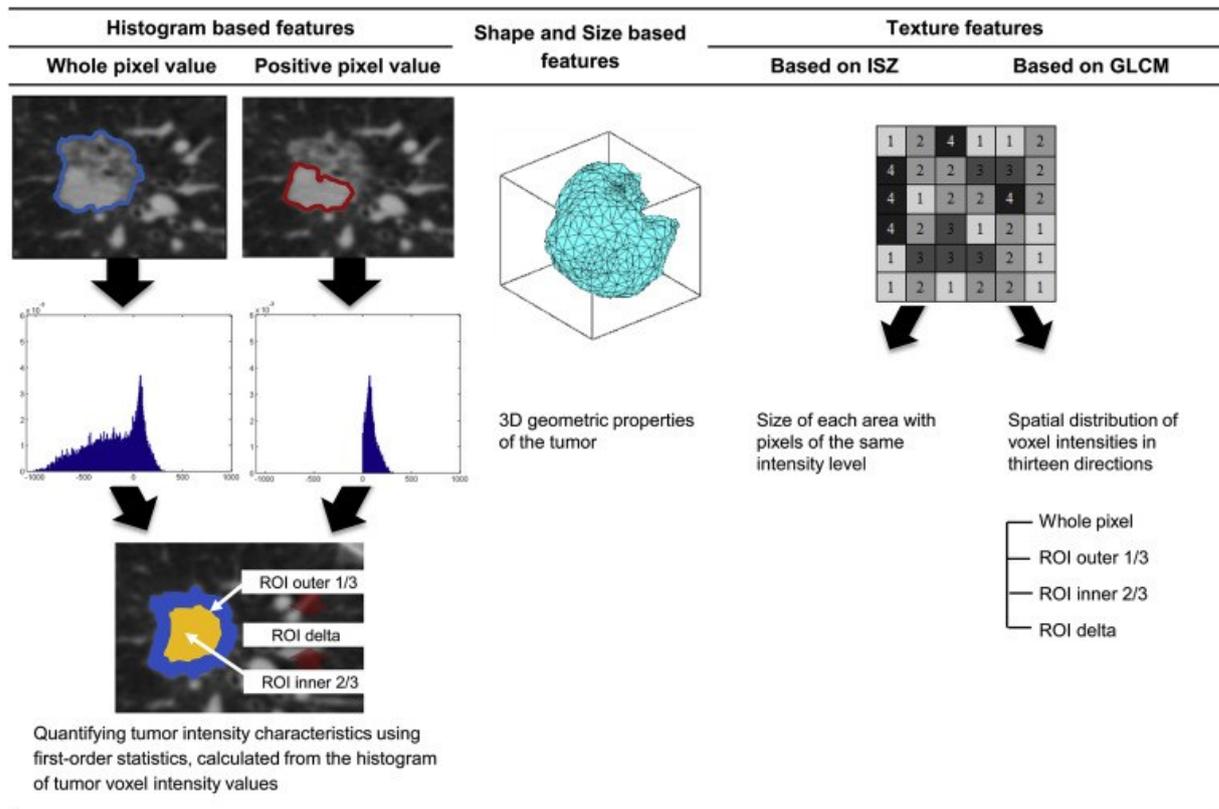
Pour principaux résultats, les DSC médianes entre les lecteurs et les contours ILD d'apprentissage en profondeur allaient de 0,74 à 0,75, tandis que les DSC médianes entre les contours des radiologues allaient de 0,68 à 0,71. L'étendue de la maladie obtenue à partir de l'algorithme, en analysant l'ensemble de la TDM, était corrélée à la capacité pulmonaire de diffusion du monoxyde de carbone, à la capacité pulmonaire totale et à la capacité vitale forcée ($r = -0,76$, $-0,70$ et $-0,62$ respectivement ; $P, .001$ pour tous) dans l'ensemble de données pour la corrélation avec les résultats PFT. L'étendue de la maladie en corrélation avec la capacité pulmonaire de diffusion pour le monoxyde de carbone, la capacité pulmonaire totale et la capacité vitale forcée étaient de $r = -0,65$, $-0,70$ et $-0,57$, respectivement, dans l'ensemble de données de validation externe ($P, 0,001$ pour tous) (239).

2.3.3 PRÉ-TRAITEMENT DES DONNÉES

Si dans la vraie vie les radiologues doivent analyser des images acquises avec différentes techniques (machines, dose, épaisseur de tranche, taille de pixel, algorithmes de reconstruction) et néanmoins faire des comparaisons utiles, il est souhaitable de « normaliser » les images avant de les alimenter dans l'algorithme de deep-learning. Différents degrés de prétraitement peuvent être appliqués, tels que la normalisation de la résolution physique impliquant l'épaisseur de coupe et la taille du voxel et / ou la normalisation des distributions de niveaux de gris pour suivre la distribution prédéfinie et / ou le débruitage d'image(255). Le prétraitement est absolument essentiel pour les études radiomiques impliquant la sélection de paramètres morphologiques, qui est soit sélectionnée par des humains avec des techniques d'apprentissage automatique traditionnelles, soit automatiquement identifiée lors de l'utilisation de deep-learning. Dans le premier cas, l'étape de prétraitement implique également le choix des paramètres de l'image, parmi les 3 catégories précédemment décrites, pour alimenter l'algorithme.

2.3.4 EXTRACTION DES PARAMETRES D'IMAGERIE

L'extraction des paramètres (ou caractéristiques) dans nos travaux de prédiction de la Dysfonction chronique du greffon pulmonaire (chapitre 4) a été réalisée avec la boîte à outils PyRadiomics 3.0.1 (256) disponible en ligne, implémentée en python 3.7.3. (257), avec une bibliothèque d'une totalité de 107 paramètres. Ces paramètres comprennent les paramètres de premier ordre, les paramètres de second ordre, de l'ordre supérieur et les paramètres de la forme.



ISZ, Intensity variance and size zone variance value; GLCM, Gray-level co-occurrence matrix; ROI, Region of interest

Figure 2-6: Illustration schématique des paramètres de la tomодensitométrie quantitative. 3D, en trois dimensions (258).

Les paramètres de forme et taille : Inclus des descripteurs du contour, de la taille et de la forme bi- ou tri- dimensionnelles de la région d'intérêt (*region of interest- ROI*). Ces caractéristiques sont indépendantes de la distribution d'intensité des niveaux de gris dans le ROI mais dépendent fortement de la segmentation. Par exemple : surface de la région d'intérêt (*region of interest, ROI*) en millimètres carrés, plus grand axe, sphéricité, compacité, allongement, etc.

Les paramètres de premier ordre : ces descripteurs décrivent la distribution de l'intensité des niveaux de gris des pixels ou voxels, basée sur l'histogramme comprenaient par exemple la densité pulmonaire moyenne, l'écart type, le mode, l'aplatissement, l'asymétrie (**Error! Reference source not found.**), l'entropie et l'énergie... Ces éléments fournissent des informations globales sur la distribution des intensités de niveaux de gris mais ne prennent pas en compte leur interrelation spatiale.

Les paramètres texturales de second ordre et de l'ordre supérieur: L'analyse de texture des images a été introduite pour la première fois en 1973 par Haralick et al (259). Les méthodes statistiques d'ordre supérieur imposent des grilles de filtre sur l'image pour extraire des motifs répétitifs ou non répétitifs. Celles-ci comprennent des analyses fractales, dans lesquelles des motifs sont imposés à l'image et le nombre d'éléments de grille contenant des voxels d'une valeur spécifiée est calculé (260); les fonctions de Minkowski, qui évaluent les modèles de voxels dont l'intensité est supérieure à un seuil (261); les ondelettes, qui sont des transformées de filtre qui multiplient une image par une matrice «d'ondes» linéaires ou radiales complexes; et transformées « la placiennes de filtres passe-bande gaussiens » qui peuvent extraire de l'image des zones avec des motifs de texture de plus en plus grossiers. La texture décrit différents aspects de l'information contextuelle, tels que la distribution locale d'intensité, le degré d'uniformité ou la dispersion d'amplitude, la situation des pixels les uns par rapport aux autres (organisation spatiale). La boîte à outils PyRadiomics offre des paramètres à partir de plusieurs méthodes de description telles que :

- La matrice de co-occurrence des niveaux de gris (grey-level cooccurrence matrix, GLCM) (259): décrit le changement de niveaux de gris par translation (Figure 2-7). GLCM est une technique d'analyse de texture populaire dans l'analyse d'images. L'image d'origine est d'abord convertie en une image bitmap de niveau de gris. Ensuite, pour chaque pixel et pour chaque pixel voisin (direction), leur niveau d'intensité respectif est mesuré et l'entrée de co-occurrence correspondante dans une matrice de niveau bidimensionnelle est incrémentée [Figure 2-7 -a]. La taille de la matrice dépend du nombre de niveaux de gris considérés (pour les 256 niveaux standard, la matrice résultante a une taille de 256×256). Habituellement, 4 directions ($\theta = 0^\circ, 45^\circ, 90^\circ$ et 135°) sont considérées et un total de 4 matrices de co-occurrence sont générées [Figure 2-7 - b]. Enfin, à partir de toutes les matrices de cooccurrence, un ensemble de mesures sont calculées, appelées entités Haralick.

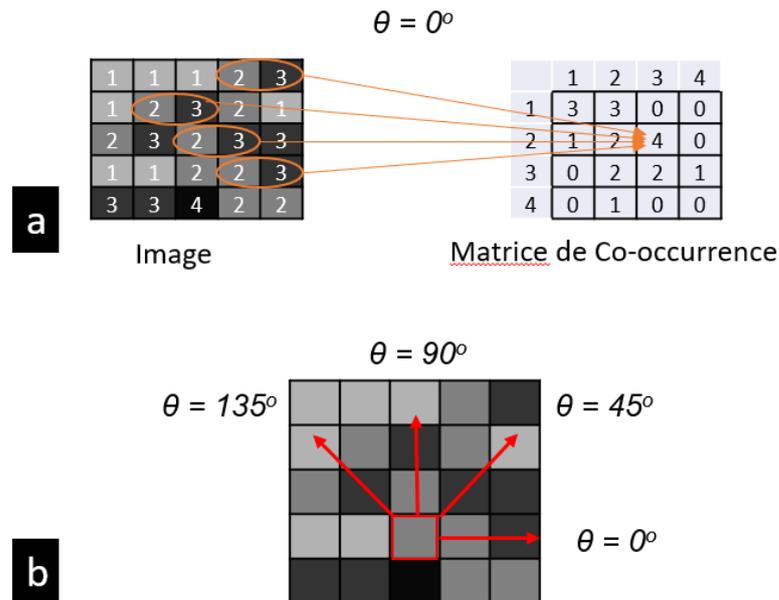


Figure 2-7: Matrice de cooccurrence des niveaux de gris: Explication du processus de matrice de cooccurrence de niveau de gris. (a) Du niveau d'intensité de l'échelle de gris (à gauche) à la création de la matrice de cooccurrence (à droite) (b) différentes directions de la Matrice de cooccurrence des niveaux de gris

- La longueur de plages de niveau de gris (grey-level run length matrix, GLRLM)(262) : nombre de lignes de niveaux de gris équivalents de longueur et de direction variables. Le parcours de niveau de gris est un ensemble de pixels colinéaires consécutifs ayant la même valeur de niveau de gris. La longueur de la plage est le nombre de pixels dans la course. Pour une image donnée, nous pouvons calculer une matrice de longueur de plage de niveau de gris pour les plages ayant une direction donnée. L'élément de matrice (i, j) spécifie le nombre de fois que l'image contient une plage de longueur j, dans la direction donnée, constituée de pixels de niveau de gris i. L'exemple de la Figure 2-8 montre une image 4 x 4 ayant quatre niveaux de gris (0-3) et les matrices de longueur de plage de niveaux de gris résultants pour les quatre directions principales. Considérez le nombre de fois que l'image contient une longueur de plage de j dans la direction 0° , consistant en un pixel ayant un niveau de gris i : 4 instances d'une longueur de plages de j = 1 constituée de pixels avec i = 0 (carré rouge) ; 0 instance d'une course - longueur de j = 3 constituée de pixels avec i = 2 (carré bleu).

Image:

0	1	2	3
0	2	3	3
2	1	1	1
3	0	3	0

Longueur de plages:

		j									
		0°	1	2	3	4	45°	1	2	3	4
N i v e a u x	i	0	4	0	0	0	0	4	0	0	0
	1	1	1	0	1	0	1	4	0	0	0
	2	3	3	0	0	0	2	0	0	1	0
	3	3	3	1	0	0	3	3	1	0	0
d e g r e s		90°	1	2	3	4	135°	1	2	3	4
	0	2	1	0	0	0	4	0	0	0	
	1	4	0	0	0	1	4	0	0	0	
	2	3	0	0	0	2	3	0	0	0	
3	3	3	1	0	0	3	5	0	0	0	

Figure 2-8: Exemple de calcul de matrice de longueur des plages (j) de niveau de gris (i) pour une image 4x4 avec 4 niveaux de gris en quatre directions principales (0°,45°,90°,135°)

-La matrice de taille des zones homogènes de niveaux de gris (grey-level size zone matrix, GLSZM)(263) : une texture homogène est composée de grandes zones de même intensité, et non de petits groupes de pixels ou de segments dans une direction donnée. Pour prendre ce fait en considération, il a fallu prendre en compte, dans une matrice, la taille de chaque zone avec des pixels de même niveau d'intensité. Cette matrice a été calculée selon le principe de plages de niveau de gris (Run Length Matrix): la valeur du carré (j, i) de la matrice est égale au nombre de zones de taille j et de niveau de gris i. La Figure 2-9 montre un exemple de calcul d'une telle matrice, baptisée taille des zones homogènes de niveaux de gris (Size Zone Matrix). La matrice résultante a un nombre fixe de lignes égal au nombre de niveaux de gris et un nombre dynamique de colonnes, déterminé par la taille de la plus grande zone. Plus la texture est homogène, plus la matrice sera large et plate. Cette matrice présente l'avantage de ne pas nécessiter de calculs dans plusieurs directions qui sont remplacés par le marquage de différentes zones. Cependant, la spécification du nombre de niveaux de gris est toujours nécessaire, mais cela rend les calculs robustes au bruit.

Image:

0	1	2	3
0	2	3	3
2	1	1	1
3	0	3	0

**Taille des zones homogènes (j)
de niveaux de gris (i):**

		j			
	0°	1	2	3	4
i	0	2	1	0	0
	1	1	0	1	0
	2	0	0	1	0
	3	2	0	1	0

Figure 2-9: Exemple de calcul de matrice de taille des zones homogènes pour une image 4x4 avec 4 niveaux de gris.

- *La différence des niveaux de gris de voisinage (neighborhood grey-tone difference matrix, NGTDM)* : Le but est de mesurer la différence entre les valeurs de niveaux de gris de chaque pixel d'image et la moyenne des valeurs tonales de ses voisins. Le j-ème élément du NGTDM représente la différence entre les pixels qui montrent la valeur de niveaux de gris j et la moyenne des valeurs de niveaux de gris de ses pixels voisins dans une image de l lignes et c colonnes, à une distance d. En d'autres termes, chaque élément NGTDM dans une matrice avec l lignes et c colonnes n'est calculé que pour les pixels aux positions (i / j), où $d < i < l - d$ et $d < j < c - d$. (Amadasun & King, 1989)(264,265). Figure 2-10 illustre une image représentée dans une matrice 4x4 et son NGTDM respectif pour $d = 1$, où j est le ton en niveaux de gris et $S(j)$ est la somme de la différence entre les j tons et la moyenne de leur environnement. Concernant la matrice illustrée dans l'image, les tons de gris en gras sont considérés pour évaluer le voisinage pour la construction du NGTDM, représenté sur le côté gauche. Par exemple, pour $j = 1$, le j-ème élément du NGTDM est donné par $S(1) = |1 - (0 + 0 + 2 + 0 + 3 + 2 + 3 + 3) / 8| = 0,625$.

Image:

0	0	2	3
0	1	3	1
2	3	3	0
1	1	1	0

Matrice de NGTDM:

j	S(j)
0	0,000
1	0,625
2	0,000
3	4,625

Figure 2-10: Exemple de calcul de matrice de différence des niveaux de gris de voisinage (S(j)) pour une image 4x4 avec 4 niveaux de gris (j).

- *La dépendance en niveaux de gris (grey-level dependence matrix, GLDM)*(266) : Une matrice de dépendance de niveau de gris (GLDM) quantifie les dépendances de niveau de gris dans une image. Une dépendance de niveau de gris est définie comme le nombre de voxels connectés à la distance δ

qui dépend du voxel central. Un voxel voisin de niveau de gris y est considéré comme dépendant du voxel central de niveau de gris x si $|y - x| \leq \alpha$. Dans une matrice de dépendance de niveau de gris $Q(k, s)$, l'élément (k, s) ème décrit le nombre de fois qu'un voxel de niveau de gris k avec s voxels dépendants dans son voisinage apparaît en image. À titre d'exemple dans bidimensionnel, considérons l'image 5×5 $M(i, j)$ dans la Figure 2-11, avec 5 niveaux de gris discrets. Avec le tableau comme indiqué sur la figure, nous construisons la matrice Q , où $Q(k, s)$ est le nombre total des entrées dans M qui ont le niveau de gris k et le nombre NGLDM s . Exemple, $Q(1,2) = 2$, car il y a deux entrées avec le niveau de gris de 1 et le nombre de GLMD de 2 se trouve à la position $M(2,4)$ (orange) et $M(3,4)$ (bleu). Par conséquent, les entrées de la matrice Q peuvent être considérées comme des comptages de fréquences de la variation de gris d'une image traitée. Il a une signification similaire à l'histogramme d'une image.

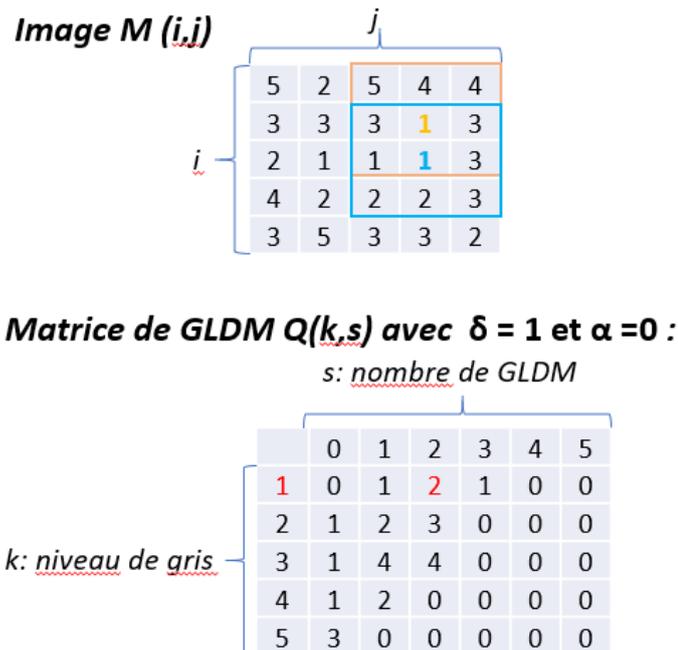


Figure 2-11: Exemple de calcul de matrice de la dépendance en niveaux de gris $Q(k, s)$ pour l'image 5×5 $M(i, j)$.

2.3.5 SELECTION DES PARAMETRES D'IMAGERIE :

Devant une centaine de paramètres extraits par les logiciels d'analyse d'image, le plus souvent, il est nécessaire de réaliser une réduction du nombre de paramètres pour éviter le phénomène de surapprentissage (*overfitting*) lié à la taille restreinte de la population étudiée. Un modèle sur ajusté est un modèle statistique qui contient plus de paramètres que ne peut le justifier le nombre d'observations, résultant en de « fausses découvertes » liées au hasard statistique. De multiples méthodes de réduction du nombre de paramètres existent et aucune recommandation actuelle ne permet de trancher pour l'une ou pour l'autre mais toutes doivent permettre d'obtenir idéalement dix fois moins de paramètres analysés que d'événements positifs. Par exemple, dans une étude analysant des paramètres quantitatifs extraits de TDM thoracique présentant une cohorte de 100 patientes, dont 50 atteintes de la fibrose, la sélection de paramètres devrait de ne conserver dans le modèle final que cinq paramètres ou moins. Ces méthodes de sélection pourraient être décrites en trois grandes catégories suivantes :

- **Méthodes de filtrage** : Elles permettent de sélectionner les paramètres en fonction de certaines de leurs caractéristiques, par exemple : La robustesse/répétabilité d'une mesure peut être étudiée par la stabilité lors de trois segmentations différentes, les paramètres non reproductibles ou non répétables étant éliminés ; La non-redondance : lorsque plusieurs

paramètres sont identifiés comme redondants, il est possible de ne conserver qu'un seul de ces paramètres.

- *Méthodes de transformation des paramètres* : ces méthodes créent une nouvelle variable à partir d'une combinaison de plusieurs variables d'intérêt. Par exemple, l'analyse en composante principale crée de nouvelles variables composées de variables initiales.
- *Méthodes incluses dans le modèle de classification* : les paramètres sont sélectionnés directement selon leur capacité discriminative dans le modèle de classification choisi. Par exemple dans notre travail de prédiction de la survenue de DCGP, Nous avons appliqué 5 algorithmes de classification: Lasso, Arbre de décision, Gradient Boosting, Adaboost et nous avons retenu pour chacun toutes les variables auxquelles l'algorithme a attribué un poids, représentant la pertinence de cette variable pour cette méthode, supérieur à la moyenne.

2.3.6 CLASSIFICATION

- Les méthodes de classification permettant d'identifier l'appartenance ou non d'un individu à un groupe (sujet malade ou sain, lésion bénigne ou maligne, survi ou décès...). La plupart de temps, apprentissage automatique est basé sur les données supervisées pour classer les individus ; selon laquelle les sujets étudiés sont étiquetés (par exemple bénin/malin) ou attribués des paramètres mesurés. Par exemple, la base de données des rayons X thoraciques 14 contient environ 112 000 radiographies pulmonaires étiquetées comme contenant une ou plusieurs des 14 anomalies courantes telles que l'atélectasie, la cardiomégalie, l'épanchement, les infiltrats, la masse, les nodules, la pneumonie et le pneumothorax (176).

Il existe en parallèle les méthodes de classification dites non supervisées selon lesquelles les données ne sont pas étiquetées, autrement dit l'algorithme n'a pas de cible pour corrélérer les données d'entrée. Ces classificateurs devraient donc se contenter de trouver des comportements semblables dans les données et apprendre le modèle qui permet de les décrire au mieux. Une fois entraîné, l'algorithme sera capable de classer une nouvelle donnée dans l'un ou l'autre des groupes formés avec une bonne pertinence. Nos travaux ont impliqué les méthodes de classification supervisée suivantes pour la prédiction des patients malades :

- Le Lasso (opérateur de retrait et de sélection le moindre absolu ; également LASSO)(267) est une méthode d'analyse de régression qui effectue à la fois la sélection et la régularisation de variables afin d'améliorer la précision de prédiction et l'interprétabilité du modèle statistique résultant. Lasso produit un modèle linéaire avec l'équation suivante :

$$f(x) = \beta_0 + \sum_{i=1}^n \beta_i x_i$$

Où x est le paramètre vectoriel d'entrée, n est le nombre d'entités, β_i (coefficients bêta) est le poids attribué au paramètre i et β_0 (intercept) est un terme constant. La régression Lasso utilise la pénalité L1 qui est égale à la valeur absolue de la magnitude des paramètres. La fonction de perte de la régression LASSO est la suivante :

$$L(\hat{\beta}) = \sum_{i=1}^n (y_i - x_i^J \hat{\beta})^2 + \lambda \sum_{j=1}^m |\hat{\beta}_j|$$

La somme des valeurs absolues des coefficients du modèle a une valeur fixe t . L'effet de la pénalité L1 est d'effectuer un type de sélection de paramètre, où les paramètres avec de très petits coefficients ne sont pas utilisés pour le modèle. Lasso peut être utilisé pour fournir un modèle clairsemé qui se généralise généralement mieux et est plus facile à calculer. Il est généralement

utilisé lorsque le nombre de paramètres est supérieur au nombre de modèles d'observations mais que le Lasso ne peut pas sélectionner un nombre de paramètres supérieurs au nombre d'observations. Un autre inconvénient de la régression Lasso est qu'en cas de groupe de caractéristiques hautement corrélées, elle a tendance à ne sélectionner qu'une seule caractéristique du groupe et à ne pas se soucier de celle qui est incluse.

- Arbre de décision : Les Arbres de Décision (Decision trees - DTs) sont une méthode d'apprentissage supervisée non-paramétrique utilisée pour la classification et la rétrogradation (268,269). Le but est de créer un modèle qui prédit la valeur d'une variable prévue en apprenant des règles de décision simples déduites des paramètres de données. Considérons des vecteurs d'entraînement $x_i \in R^n$, $i=1, \dots, l$ et d'un vecteur d'étiquette $y \in R^l$, un arbre de décision récursivement partage l'espace de paramètres tel que les échantillons avec les mêmes étiquettes ou les valeurs prévues semblables sont groupés en branches.

Permettez aux données au noeud m d'être représentées par Q_m avec N_m échantillons. Pour chaque déchirure de candidat $\theta = (j, t_m)$ composant d'un paramètre j et d'un seuil t_m , partagez les données dans et les $Q_m^{gauche}(\theta)$ et $Q_m^{droit}(\theta)$ sous-ensembles:

$$Q_m^{gauche}(\theta) = \{(x, y) | x_j \leq t_m\}$$

$$Q_m^{droit}(\theta) = Q_m \setminus Q_m^{gauche}(\theta)$$

La qualité d'une déchirure d'un candidat de noeud m est alors calculée en utilisant une fonction d'impureté ou une fonction de perte $H()$, dont le choix basant sur la tâche en question (la classification ou la régression)

$$G(Q_m, \theta) = \frac{N_m^{gauche}}{N_m} H(Q_m^{gauche}(\theta)) + \frac{N_m^{droit}}{N_m} H(Q_m^{gauchedroit}(\theta)) +$$

Choisissez les paramètres qui minimise l'impureté:

$$\theta^* = \operatorname{argmin}_{\theta} G(Q_m, \theta)$$

reprise pour les sous-ensembles $Q_m^{gauche}(\theta^*)$ et $Q_m^{droit}(\theta^*)$ jusqu'à la profondeur admissible maximum est atteint, $N_m < \min_{\text{samples}}$ ou $N_m = 1$

Si la cible est un résultat de classification embauchant des valeurs 0,1, ..., k-1, pour le noeud m , laisser

$$p_{mk} = 1/N_m \sum_{y \in Q_m} I(y = k)$$

soyez la proportion de classe k les observations dans le noeud m . Si m est un noeud terminal, probabilité de prédiction pour cette région est mis à p_{mk} . Les mesures communes d'impureté sont le suivant:

Gini:

$$H(Q_m) = \sum_k p_{mk}(1 - p_{mk})$$

Entropy:

$$H(Q_m) = \sum_k p_{mk} \log(p_{mk})$$

Misclassification:

$$H(Q_m) = 1 - \max(p_{mk})$$

Les arbres de décisions ont pour l'avantage de produire des modèles clairsemés qui sont d'habitude facile de comprendre et d'interpréter. Pour limiter la grandeur du modèle, un minimum de paramètres d'entraînement pour chaque feuille et la profondeur maximum du modèle peut être implémenté au début. En addition, la taille d'arbre aide à enlever des branches de faible d'importance.

- AdaBoost (Adaptive Boosting) : est une technique de "boosting" très populaire qui vise à combiner plusieurs classificateurs faibles pour créer un classificateur puissant (270,271). L'article original d'AdaBoost a été publié par Yoav Freund et Robert Schapire. AdaBoost utilise les stumps de décision qui sont définis comme des arbres dans une forêt aléatoire, mais ne sont pas « complètement développées » puisqu'ils n'ont qu'un nœud et deux feuilles. Commençons par considérer une base de données avec N points, ou lignes, dans notre groupe de données.

$$x_i \in R^n, y_i \in \{-1, 1\}$$

Dans lequel, n est la dimension des nombres réels, ou le nombre d'attributs dans notre jeu de données ; x est l'ensemble des points de données ; y est la variable cible qui vaut -1 ou 1 car il s'agit d'un problème de classification binaire, désignant la première ou la deuxième classe. Nous calculons la pondération dans échantillons pour chaque point de données. AdaBoost attribue un poids à chaque modèle d'entraînement pour déterminer sa signification dans l'ensemble de données d'entraînement. Lorsque les pondérations attribuées sont élevées, cet ensemble de points de données d'entraînement est susceptible d'avoir un "plus grand mot" à dire dans l'ensemble d'apprentissage. De même, lorsque les poids attribués sont faibles, ils ont une influence minime sur l'ensemble de données d'apprentissage. Au départ, tous les points de données auront la même pondération w :

$$w = \frac{1}{N} \in [0,1]$$

où N est le nombre total de points de données. Les pondérations totalisent toujours 1, de sorte que la valeur de chaque poids individuel sera toujours comprise entre 0 et 1. Ensuite, nous calculons l'influence réelle de ce classificateur dans la classification des points de données à l'aide de la formule :

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - ErreurTotale}{ErreurTotale} \right)$$

Alpha est l'influence que ce classificateur aura dans le classement final. L'erreur totale n'est rien d'autre que le nombre total d'erreurs de classification pour cet ensemble d'entraînement divisé par la taille de l'ensemble d'apprentissage. Remarquez que lorsqu'un classificateur de décision fonctionne bien ou n'a pas de classification erronée (stump parfait !), cela se traduit par un taux d'erreur de 0 et une valeur alpha positive relativement élevée.

Après avoir branché les valeurs réelles d'Erreur Totale pour chaque stump , il est temps pour la mise à jour les poids de l'échantillon qui sont initialement pris comme 1 / N pour chaque point de données. Nous allons le faire en utilisant la formule suivante :

$$w_i = w_{i-1} * e^{\pm\alpha}$$

En d'autres termes, le nouveau poids de l'échantillon sera égal à l'ancien poids de l'échantillon multiplié par le nombre d'Euler, élevé à plus ou moins alpha (que nous venons de calculer à l'étape précédente).

AdaBoost a beaucoup d'avantages, principalement il est plus facile à utiliser avec moins de besoin de peaufiner les paramètres contrairement aux algorithmes comme SVM. En prime, vous pouvez également utiliser AdaBoost avec Support Vector Machine (SVM). Théoriquement, AdaBoost n'est pas sujet au surajustement bien qu'il n'y ait pas de preuves concrètes pour cela. Cela peut être dû au fait que les paramètres ne sont pas optimisés conjointement - l'estimation par étapes ralentit le processus d'apprentissage. AdaBoost peut être utilisé pour améliorer la précision de vos classificateurs faibles, ce qui le rend flexible. Il a maintenant été étendu au-delà de la classification binaire et a également trouvé des cas d'utilisation dans la classification de texte et d'images.

- Gradient boosting: "Boosting" est un type spécial d'ensemble Learning qui travaille en combinant plusieurs modèles de prédiction faibles (avec la pauvre exactitude) dans un modèle de prédiction meilleur (un modèle avec la forte exactitude)(272,273). Dans Gradient boosting, chaque modèle essaie d'améliorer son prédécesseur en réduisant les erreurs. Mais l'idée fascinante derrière Gradient Boosting est qu'au lieu de correspondre un modèle sur les données à chaque itération, il correspond le nouveau modèle aux erreurs restantes (residual errors) faites par le modèle précédent. Le procédé peut être exprimé mathématiquement suivant :

Entrée : Base de données $\{(x_i, y_i)\}_{i=1}^n$ et une fonction de perte $L(y_i, F(x))$

Où x_i : paramètres de contribution que nous nous introduisons dans notre modèle.

y_i - variable prévue que nous essayons de prédire.

Étape 1 : Initialisez le modèle avec une valeur constante :

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, \gamma)$$

Ici, y_i est les valeurs observées, L est la fonction de perte et γ est la valeur pour log(odds).

Nous calculons fonction de perte c'est-à-dire nous nous additionnons la Fonction de Perte pour chaque valeur observée. $\underset{\gamma}{\operatorname{argmin}}$ sur le γ signifie que nous devons trouver la valeur log(odds) qui minimise cette somme.

Étape 2 : pour le $m = 1$ au M :

$$\text{Calculez } r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} \quad \forall i = 1, 2, \dots, n$$

(B) Correspondre l' arbre de régression aux valeurs r_{im} restantes et créer des régions terminales R_{im} pour $j = 1, 2, \dots, J_m$

(C) Pour chaque feuille dans le nouvel arbre, $j = 1, 2, \dots, J_m$ nous calculons le γ qui est la valeur de production.

$$\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma)$$

(D) Mise à jour

$$F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$$

Étape 3: Sortie $F_M(x)$

Les avantages de l'amplification des « gradient boosting » sont les suivants : (1) Fournissent souvent une précision prédictive qui ne peut être surpassée ; (2) Beaucoup de flexibilité - peuvent optimiser différentes fonctions de perte et fournissent plusieurs options de réglages d'hyperparamètre qui rend la fonction d'ajustement très flexible ; (3) Aucun prétraitement des données requis - fonctionnent souvent très bien avec des valeurs catégorielles et numériques telles quelles ; (4) Gèrent les données manquantes - l'imputation n'est pas requise. Il existe également quelques inconvénients : (1) Les modèles d'amplification de gradient continueront de s'améliorer pour minimiser toutes les erreurs. Cela peut surestimer les valeurs aberrantes et provoquer un surajustement ; (2) Les calculs coûteux - nécessitent souvent de nombreux arbres (1000), ce qui peut être un temps et une mémoire épuisants ; (3) La grande flexibilité se traduit par de nombreux paramètres qui interagissent et influencent fortement le comportement de l'approche (nombre d'itérations, profondeur de l'arbre, paramètres de régularisation, etc.). Cela nécessite une grande recherche de grille lors du réglage. De nature moins interprétative, bien que cela soit facilement résolu avec divers outils.

- « Support Vector Machine » (SVM)(274,275) : est très populaire dans la communauté d'apprentissage automatique et utilisé pour diverses applications. SVM est une méthode d'apprentissage automatique supervisé séparant les données à l'aide d'hyperplans. Pour classer les données, il trouve pour chaque dimension l'hyperplan qui a la distance maximale (marge la plus grande) des données les plus proches (vecteurs de support) des classes à séparer. Pour un problème de classification binaire, et pour le SVM linéaire, le problème peut être :

$$\arg \min_{\omega \in \mathbb{R}, b \in \mathbb{R}} \frac{1}{2} \|\omega\|^2 \text{ s.t. } y^i (\langle \omega, x^i \rangle + b) - 1 \geq 0 \quad \forall i \in \{1, \dots, n\}$$

où y est la classe conçue et $y^i \in \{-1, +1\}$, x sont les caractéristiques de tous les n échantillons et ω, b sont les hyperparamètres de l'hyperplan.

Cette méthode peut être étendue pour la classification non linéaire en fonction du noyau (kernel) utilisé. Pour la classification non linéaire, des filtres à fonction de base radiale, polynomiale et sigmoïde peuvent être utilisés. Le terme de régularisation « C » est utilisé pour définir dans quelle mesure le point mal classé sera pénalisé. Dans cette méthode, il est important de sélectionner un noyau et de définir « C » pour éviter du sur-ajustement. SVM est efficace dans les espaces de grandes dimensions, et toujours efficace dans les cas où le nombre de dimensions est supérieur au nombre d'échantillons. Certains d'avantages des SVM sont qu'ils sont robustes pour un petit nombre de points de données tout en offrant une solution unique. Certains principaux inconvénients de cette méthode populaire sont que dans le cas où les points sur les limites ne sont pas informatifs, l'algorithme ne fonctionnera pas bien, alors qu'il peut également être coûteux en calcul. Ici encore, des méthodes d'optimisation similaires à celles utilisées pour la régression sont généralement appliquées afin d'obtenir les meilleures performances de l'algorithme. À partir d'eux, le gradient descendant stochastique est l'une des plus populaires.

- K nearest neighbors (KNN) (276,277) : KNN est un algorithme simple qui stocke tous les cas disponibles et classe les nouveaux cas en fonction d'une mesure de similarité. KNN a déjà été utilisé dans l'estimation statistique et la reconnaissance de patrons au début des années 70 en tant que technique non paramétrique. Un cas est classé par vote majoritaire de ses voisins, le cas étant attribué à la classe la plus courante parmi ses K voisins les plus proches mesurée par une fonction de distance. Si $K = 1$, alors le cas est simplement assigné à la classe de son voisin le plus proche. Pour les paramètres continus, les fonctions de distance sont les suivantes :

Euclidean :

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Manhattan :

$$\sum_{i=1}^k |x_i - y_i|$$

Minkowski :

$$\left(\sum_{i=1}^k (|x_i - y_i|)^q \right)^{\frac{1}{q}}$$

Dans le cas de variables catégorielles, la distance de Hamming doit être utilisée. Elle soulève également la question de la standardisation des variables numériques entre 0 et 1 lorsqu'il existe un mélange de variables numériques et catégorielles dans le jeu de données.

Hamming :

$$D_H = \sum_{i=1}^k |x_i - y_i|$$

La valeur optimale pour K est choisie en fonction des données. En général, une valeur K élevée est plus précise car elle réduit le bruit global mais il n'y a aucune garantie. La validation croisée est un autre moyen de déterminer rétrospectivement une bonne valeur K en utilisant une base de données indépendantes pour valider la valeur K. Historiquement, le K optimal pour la plupart des bases de données se situait entre 3 et 10.

La phase d'apprentissage de la classification K-plus proche voisin est beaucoup plus rapide par rapport aux autres algorithmes de classification. Il n'est pas nécessaire de former un modèle pour la généralisation, c'est pourquoi KNN est connu comme l'algorithme d'apprentissage simple et basé sur des instances. KNN peut être utile dans le cas de données non linéaires. Il peut être utilisé avec le problème de régression. La valeur de sortie de l'objet est calculée par la moyenne de la valeur des k voisins les plus proches.

En revanche, la phase de test de la classification K-plus proche voisin est plus lente et plus coûteuse en termes de temps et de mémoire. Il nécessite une grande mémoire pour stocker l'ensemble de données d'entraînement à des fins de prédiction. KNN nécessite une mise à l'échelle des données car KNN utilise la distance euclidienne entre deux points de données pour trouver les voisins les plus proches. La distance euclidienne est sensible aux magnitudes. Les entités à forte magnitude pèseront plus que les entités à faible magnitude. KNN ne convient pas non plus aux grandes données dimensionnelles.

- Modèles de réseaux neuronaux (275,278):

Comme nous l'avons dit précédemment, un réseau neuronal artificiel (ANN) est un modèle prédictif conçu pour fonctionner comme le fait un cerveau humain. En fait, les ANN sont au cœur même de deep-learning. Il y a trois types de couches (layers) à la structure d'un algorithme de réseau neuronal :

- La couche d'entrée (« input layer ») : ceci entre les valeurs de données passées dans la couche suivante.
- La couche cachée (« hidden layer ») : il s'agit d'un élément-clé d'un réseau neuronal. Il a des fonctions complexes qui créent des prédicteurs. Un ensemble de nœuds dans la couche cachée appelées neurones représente des fonctions mathématiques qui modifient les données d'entrée.
- La couche de sortie (« output layer »): les prédictions faites dans la couche cachée sont collectées pour produire la couche finale - qui est la prédiction du modèle.

En formulation mathématique, compte tenu d'un ensemble d'entraînement d'exemples :

$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, où $x_i \in R^n$ et $y_i \in \{0,1\}$

un réseau neuronal d'une couche cachée apprend la fonction:

$f(x) = W_2 g(W_1^T x + b_1) + b_2$ où $W_1 \in R^m$ et $W_2, b_1, b_2 \in R$ sont des paramètres du modèle. W_1, W_2 représentent les poids de la couche d'entrée et de la couche masquée, respectivement; Et b_1, b_2 représentent respectivement le biais ajouté à la couche masquée et à la couche sortie.

Pour la classification binaire, $f(x)$ passe par la fonction logistique (fonction d'activation) $g(z) = 1/(1 + e^{-z})$ pour obtenir des valeurs de sortie entre zéro et un. Un seuil, fixé à 0,5, attribuerait des échantillons de sorties plus grands ou égaux à 0,5 à la classe positive et le reste à la classe négative. S'il y a plus de deux classes, $f(x)$ lui-même serait un vecteur de taille ($n_classes$). Au lieu de passer par la fonction logistique, il passe par la fonction softmax qui s'écrit :

$softmax(z)_i = \frac{\exp(z_i)}{\sum_{l=1}^k \exp(z_l)}$ où z_i représente le i ème élément de l'entrée de softmax, qui correspond à la classe i , et k est le nombre de classes. Le résultat est un vecteur contenant les probabilités que l'échantillon x appartienne à chaque classe. La sortie est la classe avec la probabilité la plus élevée.

Le réseau neuronal peut utiliser différentes fonctions de perte en fonction du type de problème. La fonction de perte pour la classification est Cross-Entropy, qui dans le cas binaire est donnée comme suit:

$$Loss(\hat{y}, y, W) = -y \ln \hat{y} - (1 - y) \ln(1 - \hat{y}) + \alpha \|W\|_2^2$$

où $\alpha \|W\|_2^2$ est un terme de régularisation L2 (pénalité) qui pénalise les modèles complexes; Et $\alpha > 0$ est un hyperparamètre non négatif qui contrôle l'amplitude de la pénalité.

À partir des poids aléatoires initiaux, le réseau minimise la fonction de perte en mettant à jour à plusieurs reprises ces poids. Après avoir calculé la perte, une passe en arrière la propage de la couche de sortie aux couches précédentes, fournissant à chaque paramètre de poids une valeur de mise à jour destiné à diminuer la perte.

Pour la procédure de "gradient descent", le gradient $\nabla Loss_W$ de la perte par rapport aux poids est calculé et déduit de W . Plus formellement s'exprime par:

$$W^{i+1} = W^i - \epsilon \nabla Loss_W^i$$

où i est le pas d'itération, et ϵ est le taux d'apprentissage (« learning rate ») avec une valeur supérieure à 0. L'algorithme s'arrête lorsqu'il atteint un nombre maximum d'itérations prédéfini; Ou lorsque l'amélioration de la perte est inférieure à un certain nombre minime.

Les principaux avantages des réseaux de neurones sont suivants: (1) Les réseaux de neurones sont flexibles et peuvent être utilisés à la fois pour des problèmes de régression et de

classification. Toutes les données qui peuvent être rendues numériques peuvent être utilisées dans le modèle, car le réseau de neurones est un modèle mathématique avec des fonctions d'approximation; (2) Les réseaux de neurones sont bons à modéliser avec des données non linéaires avec un grand nombre d'entrées, dans ce cas, des caractéristiques d'images médicales. Il est fiable dans une approche de tâches impliquant de nombreuses fonctionnalités. Il fonctionne en divisant le problème de la classification en un réseau en couches d'éléments plus simples; (3) Une fois entraînées, les prédictions sont assez rapides. (4) Les réseaux de neurones peuvent être formés avec un nombre illimité d'entrées et de couches. Les réseaux de neurones fonctionnent mieux avec plus de points de données.

Les réseaux neuronaux possèdent néanmoins certains nombres des inconvénients: (1) les modèles de réseaux neuronaux pourraient être considérés comme des boîtes noires, ce qui signifie que nous ne pouvons pas savoir dans quelle mesure chaque variable indépendante influence les variables dépendantes; (2) Il est très coûteux et chronophage de s'entraîner avec des processeurs traditionnels; (3) Les réseaux de neurones dépendent beaucoup des données d'entraînement. Cela conduit au problème du sur-ajustement et de la généralisation.

- Gaussian Naive Bayes (275,275)

Les méthodes Naive Bayes sont un ensemble d'algorithmes d'apprentissage supervisé basés sur l'application du théorème de Bayes qui a été formulé par Thomas Bayes (1701-1761) et peut s'écrire:

$$\text{Posterior Probability} = \frac{\text{Conditional Probability} * \text{Prior Probability}}{\text{Predictor Prior Probability}}$$

ou :

$$P(A|B) = \left(\frac{P(A \cap B)}{P(B)} \right) = \frac{P(A) * P(B|A)}{P(B)}$$

Où,

$P(A)$ = la probabilité en priorité de se produire A

$P(B|A)$ = la probabilité de condition de B étant donné que A se produit

$P(A|B)$ = la probabilité de condition de A étant donné que B se produit

$P(B)$ = la probabilité de se produire B

Gaussian Naive Bayes :

Lorsqu'il s'agit de données continues, une hypothèse typique est que les valeurs continues associées à chaque classe sont distribuées selon une distribution normale (ou gaussienne).

Supposons que les données d'entraînement contiennent un attribut continu, x . Les données sont d'abord segmentées par la classe, puis la moyenne et la variance de x sont calculées dans chaque classe. Soit μ_k la moyenne des valeurs de x associées à la classe C_k , et soit σ_k^2 la variance corrigée de Bessel des valeurs de x associées à la classe C_k . Supposons que l'on ait collecté des valeurs d'observation v . Ensuite, la densité de probabilité de v donnée une classe C_k , $p(x = v|C_k)$ peut être calculée en branchant v dans l'équation d'une distribution normale paramétrée par μ_k et σ_k^2 . C'est-à-dire :

$$p(x = v | C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$$

Bernouilli Naïve Bayes :

Dans le modèle d'événement multivarié de Bernoulli, les entités sont des booléens indépendants (variables binaires) décrivant les entrées. Supposons les tâches de classification de documents avec lesquels ces modèles sont très populaires, où des caractéristiques d'occurrence de terme binaire sont utilisées plutôt que des fréquences de terme, si x_i est un booléen exprimant l'occurrence ou l'absence du i ème terme du vocabulaire, alors la vraisemblance (likelihood) d'un document ayant une classe C_k , est donnée par :

$$p(x|C_k) = \prod_{i=1}^n p_{ki}^{x_i} (1 - p_{ki})^{(1-x_i)}$$

Où p_{ki} est la probabilité que la classe C_k génère le terme x_i . Ce modèle d'événement est particulièrement populaire pour classer des textes courts.

Les avantages de la méthodologie Naïve Bayes sont : (1) L'algorithme fonctionne très rapidement et peut facilement prédire la classe d'un jeu de données de test. On pourrait l'utiliser pour résoudre des problèmes de prédiction multi-classes. Ce classificateur fonctionne mieux que les autres modèles avec moins de données d'entraînement si l'hypothèse d'indépendance des fonctionnalités est vérifiée. Si vous avez des variables d'entrée catégorielles, l'algorithme Naïve Bayes fonctionne exceptionnellement bien par rapport aux variables numériques. Au contraire, il reste beaucoup d'inconvénients : (1) Si votre ensemble de données de test contient une variable catégorielle d'une catégorie qui n'était pas présente dans l'ensemble de données d'entraînement, le modèle Naïve Bayes lui attribuera une probabilité nulle et ne pourra pas faire de prédictions à cet égard. Ce phénomène est appelé « fréquence zéro » et vous devrez utiliser une technique de lissage pour résoudre ce problème ; (2) Il suppose que toutes les fonctionnalités sont indépendantes. Bien que cela puisse sembler génial en théorie, dans la vraie vie, vous trouverez à peine un ensemble de fonctionnalités indépendantes.

- Analyse discriminante linéaire et quadratique(275,279,280) : L'analyse discriminante linéaire (Linear Discriminant Analysis - LDA) et l'analyse discriminante quadratique (Quadratic Discriminant Analysis - QDA) peuvent être considérées comme des méthodes de réduction de dimensionnalité, et peut également être considérée comme une méthode de classification, et également être appliquée aux deux simultanément, c'est-à-dire aux données de réduction de dimensionnalité pour la classification la plus efficace. La dimensionnalité des nouvelles données est inférieure au nombre des classes. Le mot « Discriminant » est interprété comme une information spécifique à chaque classe, ce qui ne le confond pas avec d'autres classes. LDA et QDA peuvent être dérivés de modèles probabilistes simples qui modélisent la distribution conditionnelle de classe des données $P(X|y = k)$ pour chaque classe k . Les prédictions peuvent ensuite être obtenues en utilisant la règle de Bayes, pour chaque échantillon d'entraînement $x \in R^d$.

$$P(y = k|x) = \frac{P(x|y = k)P(y = k)}{P(x)} = \frac{P(x|y = k)P(y = k)}{\sum_l P(x|y = l).P(y = l)}$$

Et on sélectionne la classe k qui maximise cette probabilité postérieure.

Plus spécifiquement, pour l'analyse discriminante linéaire et quadratique, $P(x|y)$ est modélisé comme une distribution gaussienne multivariée avec densité:

$$P(x|y = k) = \frac{1}{(2\pi)^{d/2}} \exp ()$$

Où d est le nombre de paramètre. Selon le modèle ci-dessus, le log postérieur de QDA est :

$$\begin{aligned} \log P(y = k|x) &= \log P(x|y = k) + \log P(y = k) + Cst \\ &= -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2} (x - \mu_k)^t \Sigma_k^{-1} (x - \mu_k) + \log P(y = k) + Cst \end{aligned}$$

Où le terme constant Cst correspond au dénominateur $P(x)$ en plus d'autres termes constants de la Gaussienne. La classe prédite est celle qui maximise ce log-postérieur. LDA suppose des données normalement distribuées et un vecteur moyen spécifique à la classe et une matrice de covariance commune. Donc, une matrice de covariance commune à toutes les classes d'un ensemble de données. Au contraire, dans la méthode de QDA, il n'y a pas d'hypothèses sur les matrices de covariance Σ_k des Gaussiens, conduisant à des surfaces de décision quadratiques. LDA est moins flexible que QDA car nous devons estimer moins de paramètres. Cela peut être bon lorsque nous n'avons que quelques observations dans notre ensemble de données d'entraînement afin de réduire la variance. D'un autre côté, lorsque les classes k ont des matrices de covariance très différentes, alors LDA souffre d'un biais élevé et QDA pourrait être un meilleur choix. Il s'agit donc du compromis biais-variance. Par conséquent, il est crucial de tester les hypothèses sous-jacentes de LDA et QDA sur l'ensemble de données, puis d'utiliser les deux méthodes pour décider laquelle est la plus appropriée.

- Select percentile (275): Il sélectionne les paramètres en fonction d'un centile des scores les plus élevés. Ces objets prennent en entrée une fonction de notation qui renvoie des scores. Les méthodes basées sur le test F estiment le degré de dépendance linéaire entre deux variables aléatoires. D'un autre côté, les méthodes d'information mutuelle peuvent capturer tout type de dépendance statistique, mais n'étant pas paramétriques, elles nécessitent plus d'échantillons pour une estimation précise.

- Chi-square : Il calcule les statistiques du chi carré entre chaque paramètre et classe non négatives. Ce score doit être utilisé pour évaluer les variables catégorielles dans une tâche de classification. Il compare la distribution observée des différentes classes de la cible Y parmi les différentes catégories de l'entité, avec la distribution attendue des classes cibles, quelles que soient les catégories d'entités.

- ANOVA f test : Il calcule la valeur f ANOVA pour l'échantillon fourni. Si les paramètres sont catégoriques, nous calculerons une statistique du chi carré entre chaque entité et le vecteur cible. Cependant, si les caractéristiques sont quantitatives, nous calculerons la valeur f ANOVA entre chaque caractéristique et le vecteur cible. Les scores de valeur f examinent si, lorsque nous regroupons la caractéristique numérique par le vecteur cible, les moyennes pour chaque groupe sont significativement différentes.

- Mutual information(281,282) : Le gain d'informations ou les informations mutuelles mesurent la quantité d'informations que la présence / l'absence d'une caractéristique contribue à faire la bonne prédiction sur la cible. L'information mutuelle (IM) entre deux variables aléatoires est une valeur non négative, qui mesure la dépendance entre les variables. Il est égal à zéro si et seulement si deux variables aléatoires sont indépendantes, et des valeurs plus élevées signifient une dépendance plus élevée. Cette fonction repose sur des méthodes non paramétriques basées sur l'estimation de l'entropie à partir des distances des k voisins les plus proches.

- Random forest (275,283,284) : L'algorithme de forêt aléatoire est basé sur la méthode de bagging. Il représente un concept de combinaison des modèles d'apprentissage pour augmenter les performances (précision plus élevée ou autre métrique). En un mot :

- N sous-ensembles sont constitués des ensembles de données d'origine
- N arbres de décision sont construits à partir des sous-ensembles
- Une prédiction est faite avec chaque arbre entraîné, et une prédiction finale est renvoyée sous forme de vote majoritaire

L'algorithme de formation pour les forêts aléatoires applique la technique générale d'agrégation bootstrap, ou ensachage, aux apprenants d'arbres. Étant donné un ensemble d'apprentissage $X = x_1, \dots, x_n$ avec des réponses $Y = y_1, \dots, y_n$, ensachage à plusieurs reprises (B fois) sélectionne un échantillon aléatoire avec remplacement de l'ensemble d'apprentissage et ajuste les arbres à ces échantillons :

Pour $b = 1, \dots, B$:

- Échantillonnez, avec remplacement, n exemples de formation de X, Y ; Appelez-les X_b, Y_b .
- Former un arbre de classification ou de régression f_b sur X_b, Y_b .

Après l'entraînement, des prédictions pour les échantillons invisibles x' peuvent être faites en faisant la moyenne des prédictions de tous les arbres de régression individuels sur x' :

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x')$$

Ou en prenant le vote à la majorité dans le cas des arbres de classement.

Cette procédure d'amorçage conduit à de meilleures performances du modèle car elle diminue la variance du modèle, sans augmenter le biais. Cela signifie que si les prédictions d'un seul arbre sont très sensibles au bruit dans son ensemble d'apprentissage, la moyenne de nombreux arbres ne l'est pas, tant que les arbres ne sont pas corrélés. Entraîner simplement plusieurs arbres sur un seul ensemble d'apprentissage donnerait des arbres fortement corrélés (ou même le même arbre plusieurs fois, si l'algorithme d'apprentissage est déterministe) ; L'échantillonnage bootstrap est un moyen de décorréliser les arbres en leur montrant différents ensembles d'apprentissage. De plus, une estimation de l'incertitude de la prédiction peut être faite comme l'écart type des prédictions de tous les arbres de régression individuels sur x' :

$$\sigma = \sqrt{\frac{\sum_{b=1}^B (f_b(x') - \hat{f})^2}{B - 1}}$$

Le nombre d'échantillons/ d'arbres, B, est un paramètre libre. En règle générale, quelques centaines à plusieurs milliers d'arbres sont utilisées, selon la taille et la nature de l'ensemble d'entraînement. Un nombre optimal d'arbres B peut être trouvé en utilisant la validation croisée, ou en observant *l'erreur out-of-bag*: l'erreur de prédiction moyenne sur chaque échantillon d'apprentissage x_i , en utilisant uniquement les arbres qui n'avaient pas x_i dans leur échantillon bootstrap. L'erreur d'entraînement et de test a tendance à se stabiliser après qu'un certain nombre d'arbres ont été ajustés.

Comme pour tout algorithme, l'utilisation Random Forest présente des avantages et des inconvénients. Pour les avantages, l'algorithme de forêt aléatoire n'est pas biaisé, car il existe plusieurs arbres et chaque arbre est formé sur un sous-ensemble de données. Fondamentalement, l'algorithme de forêt aléatoire repose sur la puissance de «la foule» ; Par conséquent, le biais global de l'algorithme est réduit. D'autre côté, cet algorithme est très stable. Même si un nouveau point de données est introduit dans l'ensemble de données, l'algorithme global n'est pas beaucoup affecté car les nouvelles données peuvent avoir un impact sur un arbre, mais il est très difficile pour cela d'avoir un impact sur tous les arbres. L'algorithme de forêt aléatoire fonctionne bien lorsque vous avez à la fois des caractéristiques catégoriques et numériques. L'algorithme de forêt aléatoire fonctionne également bien lorsque les données ont des valeurs manquantes ou n'ont pas été correctement mises à l'échelle (bien que nous ayons effectué la mise à l'échelle des fonctionnalités dans cet article uniquement à des fins de démonstration).

Aux côtés des inconvénients, l'utilisation de Random Forest réside beaucoup de complexité. Ils nécessitent beaucoup plus de ressources de calcul, en raison du grand nombre d'arbres de décision réunis. En raison de leur complexité, ils nécessitent beaucoup plus de temps pour s'entraîner que d'autres algorithmes comparables.

- Méthodes d'ensemble (285–287): Le but des méthodes d'ensemble est de combiner les prédictions de plusieurs estimateurs de base construites avec un algorithme d'apprentissage donné afin d'améliorer la généralisabilité / robustesse sur un seul estimateur. On distingue généralement deux familles de méthodes d'ensemble : Dans les méthodes de moyennage, le principe directeur est de construire plusieurs estimateurs indépendamment, puis de faire la moyenne de leurs prédictions. En moyenne, l'estimateur combiné est généralement meilleur que n'importe quel estimateur de base unique parce que sa variance est réduite, par exemple, méthodes d'ensachage, forêts d'arbres aléatoires,ect... En revanche, dans les méthodes d'amplification, les estimateurs de base sont construits séquentiellement et on essaie de réduire le biais de l'estimateur combiné. La motivation est de combiner plusieurs modèles faibles pour produire un ensemble puissant, par exemple, AdaBoost, Gradient Tree Boosting, ect...

2.3.7 VALIDATION

Pour éviter le phénomène de sur-apprentissage, lors de la réalisation d'une expérience d'apprentissage automatique (supervisée) on réserve une partie des données disponibles comme ensemble de test. Pour une validation standardisée le groupe de test est séparé dès le début de l'expérience et présentant des caractéristiques comparables au groupe d'entraînement notamment en matière de prévalence de la maladie étudiée Dans la boîte d'outils « scikit-learn », une division aléatoire en ensemble d'entraînement et de test peut être rapidement calculée avec la fonction d'assistance `train_test_split` (275).

Lors de l'évaluation de différents paramètres (« hyperparamètres ») pour les estimateurs, il existe toujours un risque de surajustement sur l'ensemble de test car les paramètres peuvent être modifiés jusqu'à ce que l'estimateur fonctionne de manière optimale. De cette façon, les connaissances sur l'ensemble de test peuvent « s'infiltrer » dans le modèle et les métriques d'évaluation ne rendent plus compte des performances de généralisation. Pour résoudre ce problème, une autre partie de l'ensemble de données peut être présentée comme un soi-disant « ensemble de validation » : la formation se déroule sur l'ensemble d'apprentissage, après quoi l'évaluation est effectuée sur l'ensemble de validation, et lorsque l'expérience semble réussir, l'évaluation finale peut être effectuée sur l'ensemble de test(288).

En pratique, étant donné pour un grand nombre des cas qu'on ne dispose pas d'une population suffisante pour la partition en trois ensembles (d'entraînement, validation et test), une méthode de validation croisée (cross validation) peut être réalisée afin d'évaluer les performances de l'algorithme sans perdre en quantité de données d'entraînement (Figure 2-12). Dans l'approche de base, appelée « cross-validation k folds », l'ensemble d'entraînement est divisé en k ensembles plus petits. La procédure suivante est suivie pour chacun des k « plis » :

- Le modèle est entraîné en utilisant k-1 des plis comme données d'apprentissage ;
- Le modèle résultant est validé sur la partie restante des données (c'est-à-dire qu'il est utilisé comme ensemble de test pour calculer une mesure de performance telle que la précision).

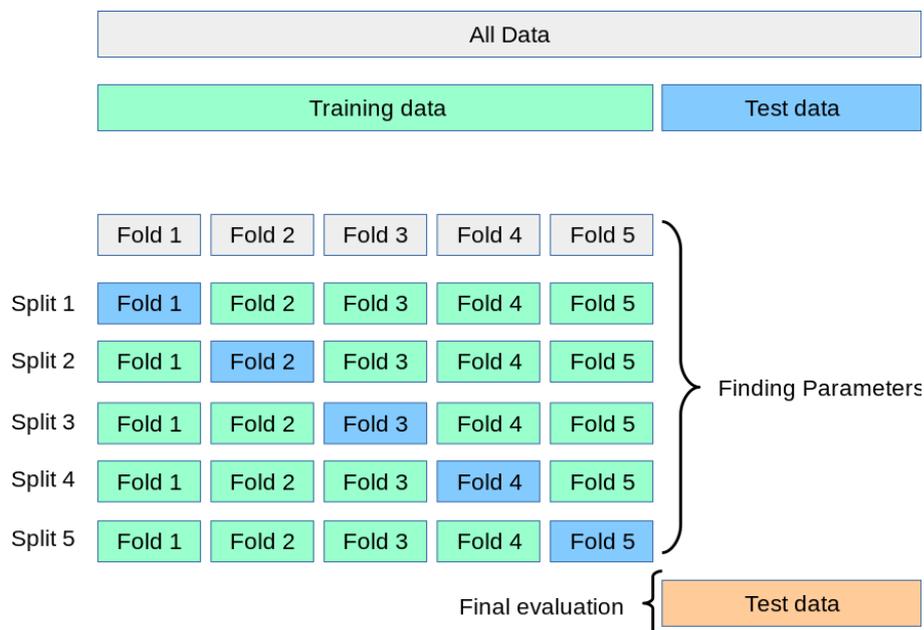


Figure 2-12: Partition des données dans la méthode de « validation croisée » (275).

La mesure de performance rapportée par la validation croisée de k fois est alors la moyenne des valeurs calculées dans la boucle. Cette approche peut être coûteuse en calcul, mais ne gaspille pas trop de données (comme c'est le cas lors de la fixation d'un ensemble de validation arbitraire), ce qui est un avantage majeur dans des problèmes tels que l'inférence inverse où le nombre d'échantillons est très petit.

Un modèle de classification utilisant des paramètres quantitatifs extraits de l'image doit avoir démontré ses performances diagnostiques sur un ou plusieurs groupes de test issus de centres différents et si possible de machines différentes pour prétendre à une application clinique.

3 BIOMARQUEUR QUANTIFIANT LES LESIONS PULMONAIRES DANS LA DYSKINÉSIE CILLIAIRE PRIMITIVE

3.1 CONTEXTE ET OBJECTIFS

La dyskinésie ciliaire primitive (DCP) est une maladie génétique rare caractérisée par une anomalie de la fonction ciliaire, entraînant une clairance mucociliaire inadéquate, une maladie oto-sino-pulmonaire chronique (289–291), et des infections bactériennes récurrentes et chroniques des voies respiratoires inférieures et une bronchectasie (290). La prévalence de la maladie est d'environ 1/20000 - 1/15 000 des naissances dans la population caucasienne et comme les symptômes ne sont pas spécifiques, le diagnostic tardif est courant et le traitement est généralement instauré à un stade avancé de la maladie (291). La prise en charge de la DCP repose sur la prévention des infections, le traitement des exacerbations et la réhabilitation respiratoire. Le pronostic est principalement lié à l'insuffisance respiratoire due à une obstruction des voies respiratoires par des sécrétions et à des infections qui conduisent progressivement au développement de bronchectasie (13). À ce jour, la prise en charge recommandée des patients est une surveillance pulmonaire et des oreilles, du nez et de la gorge (ORL) tous les 6 à 12 mois (14). La tomodensitométrie est la méthode de référence pour le diagnostic de la bronchectasie, mais son utilité pour la surveillance de la DCP n'est pas encore établie (15,22,292). Les corrélations entre les modifications structurelles de la tomodensitométrie et la gravité de la maladie (par l'évaluation de la fonction respiratoire) ont rarement été étudiées dans la DCP, en particulier chez les adultes (12,21,22,25,292–295).

La plupart des auteurs qui ont tenté de quantifier la maladie bronchique chez les patients atteints de DCP ont utilisé des méthodes de notation visuelle initialement conçues pour évaluer les

changements structurels pulmonaires chez les patients atteints de mucoviscidose (Cystic fibrosis - CF), mais la corrélation entre ces scores visuels et le volume expiratoire forcé en 1 seconde (VEMS) reste controversée chez les patients atteints de DCP (12,17,21,22,293–295). Par exemple, Boon et coll. ont rapporté une bonne corrélation négative entre un score TDM visuel et FEV1 ($R = -0,63$, $P < 0,001$), tandis que Cohen-Cymbarknoh et coll. n'en ont trouvé aucune corrélation ($R = -0,36$, $P = 0,61$)(17,21). Cependant, ces auteurs ont également signalé des différences concernant l'indice de masse corporelle (IMC), les germes infectieux et la pathogenèse des lésions pulmonaires dans les deux populations.

Un point commun entre ces investigations est que l'évaluation de la sévérité du scanner reste principalement descriptive ou semi-quantitative (18–21,296,296,297). Parmi les approches semi-quantitatives, la plus couramment utilisée est le score TDM de Brody et Bhalla, qui a été principalement développé pour évaluer l'implication des voies respiratoires dans la mucoviscidose (22). L'utilisation de ce score souffre de plusieurs limitations, elle nécessite une formation dédiée, elle prend du temps et surtout, elle repose sur une évaluation subjective des anomalies du scanner (22,23,298).

Bien que dans la bronchectasie, l'épaississement de la paroi bronchique, les impactions du mucus et la perfusion en mosaïque soient présents à la fois dans la DCP et la mucoviscidose, leur prédominance relative diffère entre les deux maladies. Contrairement à la mucoviscidose, dans laquelle les lobes supérieurs sont principalement affectés, les anomalies pulmonaires de la DCP sont principalement localisées dans les lobes moyens et inférieurs (Figure 1-2). La perfusion en mosaïque et les impactions du mucus des petites voies respiratoires prédominent dans la DCP, ce qui signifie que leurs poids respectifs dans le score global ne doivent pas être les mêmes que dans la mucoviscidose (17,292).

Ces raisons pourraient expliquer pourquoi certains auteurs n'ont pas trouvé de corrélation entre les scores visuels et la spirométrie chez les patients DCP. La plupart des changements structurels pulmonaires de la DCP, en particulier l'épaississement de la paroi bronchique, les impactions du mucus, les consolidations et l'atélectasie, sont susceptibles d'augmenter l'atténuation pulmonaire et de modifier les paramètres de l'histogramme de densité, qui peuvent être extraites des images de TDM. D'ailleurs, les structures à forte atténuation peuvent être quantifiées en utilisant une approche de seuillage similaire à celle utilisée pour mesurer l'emphysème en TDM, sauf que cette dernière est basée sur la quantification des zones pulmonaires à faible atténuation, avec des valeurs d'atténuation inférieures à moins 950 unités Hounsfield. Il a été rapporté que la quantification des changements des structures de haute atténuation dans les poumons, en utilisant également une approche de seuillage, montre une bonne corrélation avec le VEMS chez les patients atteints de mucoviscidose. Pour ces raisons, nous avons émis en hypothèse que la quantification des tissus lésionnels par seuillage en histogramme de densité des images pourrait évaluer la gravité de l'atteinte pulmonaire chez les patients de DCP.

3.2 PUBLICATION

Automated computed tomographic scoring of lung disease in adults with primary ciliary dyskinesia.

Trieu-Nghi Hoang-Thi, Marie-Pierre Revel, Pierre-Régis Burgel, Laurence Bassinet, Isabelle Honoré, Thong Hua-Huy, Charlotte Martin, Bernard Maitre and Guillaume Chassagnon.

BMC Pulmonary Medicine (2018) 18:194

3.2.1 RESUME

Contexte : Ces travaux de recherche visaient à développer un score de tomодensitométrie (TDM) automatisé basé sur la quantification tomодensitométrique des structures pulmonaires à forte atténuation, afin de fournir une évaluation quantitative des anomalies structurelles pulmonaires chez les patients atteints de dyskinésie ciliaire primitive (DCP).

Méthodes : Les patients DCP adultes (≥ 18 ans) qui ont réalisé à la fois une TDM thoracique et une spirométrie sur une période de 6 mois ont été rétrospectivement inclus dans l'étude. Un logiciel de segmentation pulmonaire disponible dans le commerce a été utilisé pour isoler les poumons du médiastin et de la paroi thoracique et obtenir des histogrammes de la densité pulmonaire. Les scores de densité TDM ont été calculés à l'aide de seuils fixes et adaptés basés sur diverses combinaisons de caractéristiques d'histogramme, telles que la densité pulmonaire moyenne (DPM), l'asymétrie et l'écart type (ET). De plus, une notation visuelle utilisant le score de Bhalla a été réalisée par 2 radiologues indépendants. Les corrélations entre les scores TDM, le volume expiratoire forcé en 1 s (VEMS) et la capacité vitale forcée (CVF) ont été évaluées.

Résultats : Soixante-deux patients adultes atteints de DCP ont été inclus. Parmi toutes les caractéristiques de l'histogramme, celles montrant de bonnes corrélations positives ou négatives avec le VEMS et la CVF étaient SD ($R = -0,63$ et $-0,67$; $p < 0,001$) et Skewness ($R = 0,67$ et $0,67$; $p < 0,001$). Parmi tous les seuils évalués, le score de densité basé sur DPM + 1ET a fourni la meilleure corrélation négative à la fois avec le VEMS ($R = -0,68$; $p < 0,001$) et la CVF ($R = -0,71$; $p < 0,001$), proche des corrélations du score visuel ($R = -0,60$; $p < 0,001$ pour le VEMS et $R = -0,62$; $p < 0,001$, pour la CVF).

Conclusions : Le score automatisé en TDM des anomalies structurelles pulmonaires pulmonaires dans la dyskinésie ciliaire primitive est faisable et peut s'avérer utile pour l'évaluation de la gravité de la maladie en clinique et dans les essais cliniques.

3.2.2 ARTICLE

RESEARCH ARTICLE

Open Access



Automated computed tomographic scoring of lung disease in adults with primary ciliary dyskinesia

Trieu-Nghi Hoang-Thi^{1,6}, Marie-Pierre Revel¹, Pierre-Régis Burgel², Laurence Bassinet³, Isabelle Honoré², Thong Hua-Huy⁴, Charlotte Martin¹, Bernard Maitre³ and Guillaume Chassagnon^{1,5*} 

Abstract

Background: The present study aimed to develop an automated computed tomography (CT) score based on the CT quantification of high-attenuating lung structures, in order to provide a quantitative assessment of lung structural abnormalities in patients with Primary Ciliary Dyskinesia (PCD).

Methods: Adult (≥ 18 years) PCD patients who underwent both chest CT and spirometry within a 6-month period were retrospectively included. Commercially available lung segmentation software was used to isolate the lungs from the mediastinum and chest wall and obtain histograms of lung density. CT-density scores were calculated using fixed and adapted thresholds based on various combinations of histogram characteristics, such as mean lung density (MLD), skewness, and standard deviation (SD). Additionally, visual scoring using the Bhalla score was performed by 2 independent radiologists. Correlations between CT scores, forced expiratory volume in 1 s (FEV₁) and forced vital capacity (FVC) were evaluated.

Results: Sixty-two adult patients with PCD were included. Of all histogram characteristics, those showing good positive or negative correlations to both FEV₁ and FVC were SD ($R = -0.63$ and -0.67 ; $p < 0.001$) and Skewness ($R = 0.67$ and 0.67 ; $p < 0.001$). Among all evaluated thresholds, the CT-density score based on MLD + 1SD provided the best negative correlation with both FEV₁ ($R = -0.68$; $p < 0.001$) and FVC ($R = -0.71$; $p < 0.001$), close to the correlations of the visual score ($R = -0.60$; $p < 0.001$ for FEV₁ and $R = -0.62$; $p < 0.001$, for FVC).

Conclusions: Automated CT scoring of lung structural abnormalities lung in primary ciliary dyskinesia is feasible and may prove useful for evaluation of disease severity in the clinic and in clinical trials.

Keywords: Computed tomography, Primary ciliary dyskinesia, Kartagener syndrome, Bronchiectasis

Background

Primary ciliary dyskinesia (PCD) is a rare genetic disorder characterized by defective ciliary structure and/or function, leading to inadequate mucociliary clearance and chronic oto-sino-pulmonary disease [1–3]. Organ laterality is also affected in almost half the patients [2]. Defective mucociliary airway clearance leads to recurrent and chronic bacterial infections of the lower respiratory tract, and to bronchiectasis [2].

Computed tomography (CT) is the gold standard method for the diagnosis of bronchiectasis, but its utility for monitoring PCD is not yet established [4, 5]. Correlations between CT structural changes and disease severity (lung function) have rarely been studied in PCD, especially in adults [5–13]. However, a large retrospective study recently suggested that a larger disease burden on CT may predict lung function decline in adults with PCD, indicating that CT assessment of lung structural abnormalities might be of value [5].

Most authors who have attempted to quantify bronchial disease in patients with PCD have used visual scoring methods initially designed to assess lung structural changes in patients with cystic fibrosis (CF), but the

* Correspondence: gchassagnon@gmail.com

¹Radiology Department, Groupe Hospitalier Cochin-Hôtel Dieu, AP-HP, Université Paris Descartes - Sorbonne Paris Cité, Paris, France

⁵Center for Visual Computing, CentraleSupélec, Gif-sur-Yvette, France

Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

correlation between these visual scores and forced expiratory volume in 1 s (FEV_1) remains controversial in patients with PCD [7–14]. For example, Boon et al. reported a good negative correlation between a visual CT score and FEV_1 ($R = -0.63$, $P < 0.001$), whereas Cohen-Cymerknoh et al. found no correlation at all ($R = -0.36$, $P = 0.61$) [11, 12].

Although bronchiectasis, bronchial wall thickening, mucus plugging and mosaic perfusion are present in both PCD and CF, their relative predominance differs between the two diseases. Mosaic perfusion and small-airway mucus plugging predominate in PCD, meaning that their respective weight in the overall CT score should not be the same as in CF [5, 11]. This may be why some authors failed to find a correlation between visual scores and spirometry in PCD patients. Furthermore, visual scores suffer from several limitations, including the need for dedicated training and subjectivity in the assessment of CT changes [15].

Most lung structural changes in PCD, especially bronchial wall thickening, mucus plugging, consolidations and atelectasis are likely to increase lung attenuation and to modify the density histogram characteristics, which can be extracted from the CT images. On the density histogram, the mode corresponds to the most highly represented attenuation value; and skewness describes the asymmetry of the density curve, which is shifted to the right when there is an increase of lung density or to the left in case of decrease. Additionally, high-attenuating structures can be quantified by using a thresholding approach similar to that used to measure emphysema on CT, except that the latter is based on the quantification of low-attenuating lung areas, with attenuation values below minus 950 Hounsfield units (HU) [16]. The quantification of high-attenuating structural changes in the lungs, also using a thresholding approach has been reported to show good correlation with FEV_1 in patients with CF [17].

We postulated that disease severity in PCD might also be assessed by quantifying high-attenuating lung structures and by analysing changes in the lung density distribution. We therefore developed an automated CT scoring method based on histogram characteristic analysis and threshold-based quantification of high-attenuating lung structures in patients with PCD.

Methods

Patients

This retrospective study, performed in two accredited PCD reference centres, was approved by the Institutional Review Board of Société Pneumologie de Langue Française. The need for informed consent was waived, in accordance with French rules for retrospective observational studies.

All adult outpatients, with a diagnosis of PCD according to the ERS guidelines [18] were eligible if they had both chest CT exams of the whole thorax performed between November 2009 and July 2016 and spirometric measurements, both performed within a 6-month period. Exclusion criteria were the unavailability of CT images with a slice thickness ≤ 2 mm, reconstructed with a soft kernel, or the administration of iodinated contrast medium during the CT acquisition.

CT examinations

All CT examinations had been performed in the supine position at full inspiration, with usual acquisition parameters, allowing obtaining high resolution CT images of the whole thorax during a single breath hold. Five different 16-to-64 multislice CT devices from two different vendors (Somatom Sensation 16 and Somatom Definition DS, Siemens Healthcare, Erlangen, Germany; Lightspeed plus, Bright Speed 16 and Optima CT 660, GE Healthcare, Milwaukee, WI) had been used, depending on the site and date of the CT examinations, all performed with equivalent acquisition parameters. The radiation dose resulting from each CT acquisition was evaluated by collecting the mean dose-length product (DLP) value from the dose reports.

Image analysis

Pulmonary situs type was identified as *solitus*, *inversus* or heterotaxy, based on the relationship between the upper-lobe bronchus and the ipsilateral pulmonary artery, and the morphology of the tracheobronchial tree [19]. CT images were also checked for prior lobectomy.

Lung structural changes were assessed by visual scoring and also by histogram analysis and thresholding of high attenuating lung structures.

Visual CT scoring was performed as follow

All the images were scored by one thoracic radiologist (CM) using the Bhalla score [20]. Twenty randomly selected examinations were also independently scored by a second radiologist (GC) to assess interobserver repeatability.

Automated CT scoring was performed as follows

First, the lungs were isolated from the mediastinum and chest wall using a commercially available, automated lung segmentation software (Myrian XP lung software version 1.19.1, Intrasense, Montpellier, France).

This allowed obtaining isolated whole lung volumes, for further density histogram analysis.

We also obtained separate volumes of the upper (right upper lobe and upper part of the left upper lung) and lower lungs (middle lobe, lingula, and lower lobes), after manual contouring of the fissures. This was only done

Table 1 Characteristics of the patients ($n = 62$)

Male/Female	32 / 30
Age, yr	39 (15)
Bronchial situs	
- solitus	37 (60)
- inversus	24 (39)
- left isomerism	1 (1)
Body mass index, kg/m ²	23.71 (4)
Percentage of predicted FEV ₁	67 (20)
Percentage of predicted FVC	80 (18)
Median interval between spirometry and CT, months [interquartile range]	0 [0–16]

For quantitative variables, data are mean with standard deviation in parentheses. For qualitative variables, data are numbers of patients, and numbers in parentheses are percentages.

Definition of abbreviations: CT computed tomography, FEV₁ forced expiratory volume in 1 s, FVC forced vital capacity

in 2 patients (3%), and the middle lobe plus the lingula in the remaining patient (2%).

CT examinations

The median interval between CT and spirometry was 0 days [interquartile range: 0–29], 41 of the 62 CT scans being performed on the same day as spirometry. The mean DLP per CT scan was 200.3 ± 100.4 mGy.cm.

Visual CT score

The interobserver repeatability for the visual score was excellent (ICC = 0.84). Visual CT score, performed for all CT scans by one of the 2 radiologists, showed good correlation with FEV₁ ($R = -0.60$; $p < 0.001$) and FVC ($R = -0.62$; $p < 0.001$).

Histogram characteristics

Two histogram characteristics – kurtosis ($R = 0.56$; $p < 0.001$) and skewness ($R = 0.60$; $p < 0.001$)– showed a moderate to good correlation to FEV₁ (Table 2). The same two characteristics also correlated with FVC ($R = 0.65$; $p < 0.001$ and 0.67 ; $p < 0.001$, respectively). SD showed good negative correlation to both FEV₁ ($R = -0.63$; $p < 0.001$) and FVC ($R = -0.67$; $p < 0.001$). Overall the correlations with FVC were slightly stronger than the correlations with FEV₁. Examples of variations in histogram shape according to pulmonary function are shown in Fig. 2.

CT-density scores

All CT-density scores showed moderate to good negative correlations with FEV₁ ($R = -0.54$ to -0.68 ; $p < 0.001$) and FVC ($R = -0.62$ to -0.71 ; $p < 0.001$) (Table 2). Overall, CT scores based on fixed thresholds showed weaker negative correlations with FEV₁ ($R = -0.51$ to -0.54 ; $p < 0.001$) and FVC ($R = -0.59$ to -0.62 ; $p < 0.001$)

Table 2 Correlations between spirometry, histogram characteristics and CT-density scores

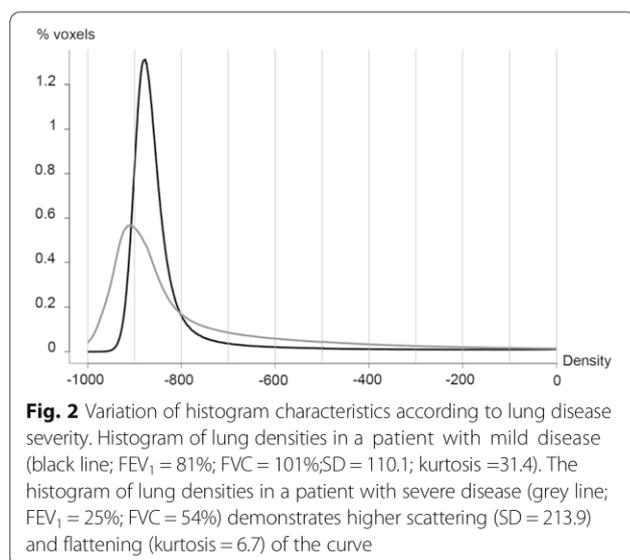
	FEV ₁		FVC	
	R	P value	R	P value
Histogram characteristics				
- MLD	-0.22	0.084	-0.31	0.012
- Mode	0.14	0.285	0.11	0.397
- SD	-0.63	< 0.001	-0.67	< 0.001
- Kurtosis	0.56	< 0.001	0.65	< 0.001
- Skewness	0.60	< 0.001	0.67	< 0.001
CT-density score with fixed thresholds				
- (-) 400 HU	-0.51	< 0.001	-0.59	< 0.001
- (-) 500 HU	-0.53	< 0.001	-0.61	< 0.001
- (-) 600 HU	-0.54	< 0.001	-0.61	< 0.001
CT-density score with adapted thresholds				
- MLD + 2 SD	-0.62	< 0.001	-0.67	< 0.001
- MLD + 1 SD	-0.68	< 0.001	-0.71	< 0.001
- Mode + 500 HU	-0.57	< 0.001	-0.64	< 0.001
- Mode + 400 HU	-0.60	< 0.001	-0.65	< 0.001
- Mode + 300 HU	-0.64	< 0.001	-0.70	< 0.001
- Mode + 3 SD	-0.54	< 0.001	-0.62	< 0.001
- Mode + 2 SD	-0.65	< 0.001	-0.70	< 0.001
- Mode + 1SD	-0.66	< 0.001	-0.68	< 0.001

Definition of abbreviations: CT computed tomography, FEV₁ forced expiratory volume in 1 s, FVC forced vital capacity, HU Hounsfield unit, MLD Mean Lung Density, SD Standard Deviation

than did CT-density scores based on adapted thresholds taking into account histogram characteristics. The strongest correlations were obtained using MLD + 1SD as threshold ($R = -0.68$; $p < 0.001$ for FEV₁ and $R = -0.71$ for FVC; $p < 0.001$) (Fig. 3). The correlations between CT score and PFTs were in the same range when considering each center separately: $R = -0.66$; $p < 0.001$ and -0.68 ; $p < 0.001$ for FEV₁ and $R = -0.68$; $p < 0.001$ and -0.70 ; $p < 0.001$ for FVC. With this threshold value (MLD + 1SD), the automated score (CT-density score) correlated well with the visual score ($R = 0.70$, $p < 0.001$). Within the Bhalla visual score, the automated CT-density score correlated well with air wall thickening and mucus plugging-related items ($R = 0.64$ and 0.61 , respectively; $p < 0.001$) and moderately with bronchiectasis-related items ($R = 0.58$; $p < 0.001$)”.

The automated score values were significantly higher in the lower lungs (median: 8.83; interquartile range: 7.61–10.06) than in the upper lungs (median: 6.25; interquartile range: 5.57–6.81) ($p < 0.001$).

Results of automated CT scoring in patients with different FEV₁ and FVC values are illustrated in Figs. 4 and 5.



Discussion

To the best of our knowledge, the present manuscript describes the first automated CT scoring method designed to quantify lung changes associated with primary ciliary dyskinesia, based on the measurement of high-attenuating structures and considering histogram characteristics on CT. This approach is close to the quantification of emphysema, based on the measurement of low-attenuating lung areas. The method presented here has been previously validated in CF patients [17].

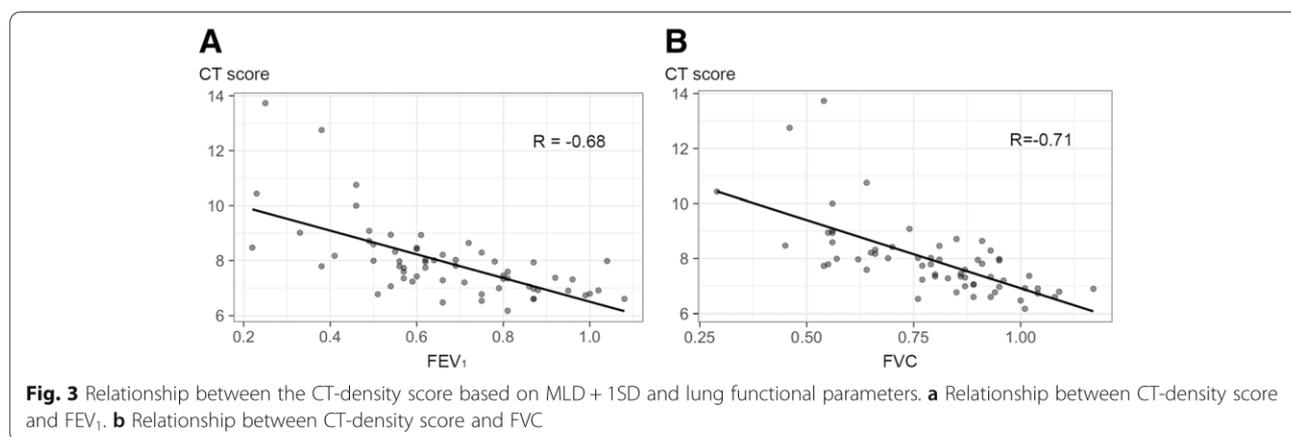
In the present cohort of 62 adults with PCD, the automated score correlated well with lung function (FEV₁ and FVC). Moreover, the value of the score was significantly different in the lower and upper portions of the lung, with higher score values in the lower part, consistent with the reported lower lung predominance of bronchial abnormalities in PCD patients [11].

Because the decline of lung function is slower in PCD than in CF, CT is less often performed. However, when

performed, lung structural changes on follow-up CT need to be compared to previous images, which is complex in view of the high number of CT images with the modern multidetector technology. Rather than a subjective and time-consuming assessment, automated scoring provides an objective quantitative evaluation.

Regarding the thresholding method, we found that adapted thresholds based on histogram characteristics correlated better than fixed thresholds with the spirometric parameters. Lung attenuation is known to be influenced by parameters such as the level of inspiration, the kilovoltage, and the patient’s position in the scan [21, 25, 26]. Expiration and, by extension, a lower level of inspiration, tend to flatten the CT density histogram, resulting in a higher SD and shifting of the curve towards higher density values, which increases Mode and MLD [26]. Instead of evaluating fixed thresholds alone, we postulated that inclusion of histogram characteristics in the threshold definition would compensate for variations not due to disease severity. Among the various thresholds tested here, MLD + 1SD gave the best results. This threshold is readily available, as most commercially available segmentation software programs provide both MLD and SD values.

One-third of our patients had previously undergone complete or partial lobectomy, even though bronchial abnormalities are not usually restricted to a single lobe in PCD and surgical resection is currently not considered an appropriate treatment for PCD [3]. This proportion is in line with the 41% prevalence reported by Kennedy et al. In this latter study, where visual CT scoring was performed, a maximal score was arbitrarily affected to the missing lung [14]. In our study, we only applied scoring to the existing lung and found good correlations of the visual score to FEV₁ and FVC, in the upper range of previously reported correlations (0.08 to -0.63 for FEV₁ and -0.38 to -0.60 for FVC) [8, 10, 12]. The correlation of the CT-density score to FEV₁ and



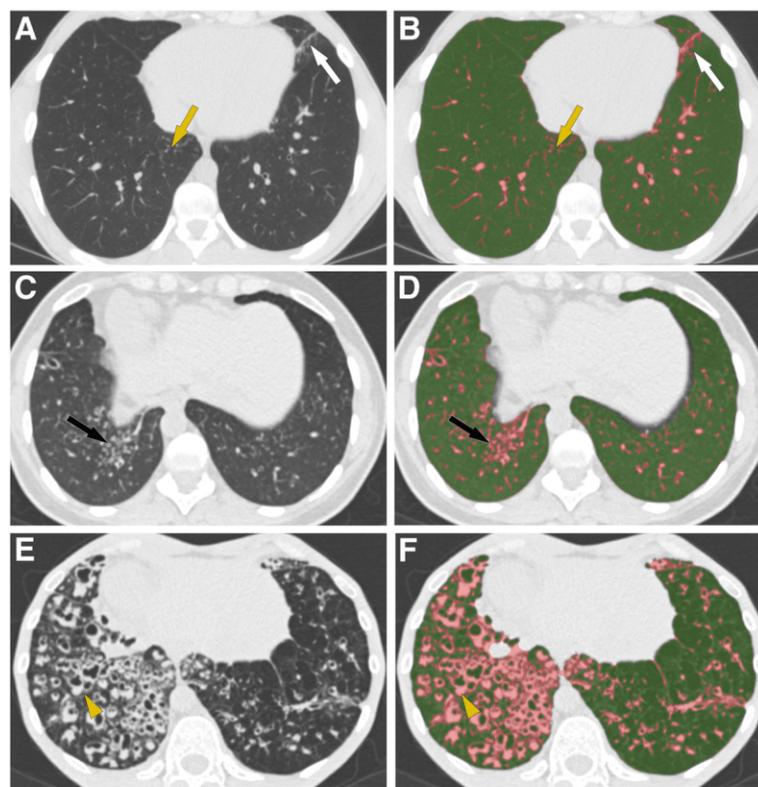


Fig. 4 CT-density scores in patients with increasing disease severity. Areas of high attenuation, using MLD + 1SD as threshold, are tagged in pink. **a, b** Patient with mild lung involvement ($FEV_1 = 80\%$, CT score = 6.01): **(a)** Small areas of tree-in-bud (yellow arrow) and subsegmental atelectasis (white arrow) are seen on the native axial image. **b** Post-processed image, illustrating that these PCD-related abnormalities appear in pink, along with pulmonary vessels. **c, d** Patient with moderate lung involvement ($FEV_1 = 60\%$, CT score = 8.37): areas of tree-in-bud (black arrow) seen on the native axial CT image **(c)** are more extensive and are pink-colored on the post-processed image **(d)** **(e, f)** Patient with severe lung involvement ($FEV_1 = 25\%$, CT score = 13.72): **(e)** Visually, bronchial abnormalities (yellow arrowhead) predominantly affect the right lung on the native axial CT image. **f** Post-processed CT image. The CT-density score is 17.59 for the right lung and 10.87 for the left lung

FVC was in the same range, with the advantage of an automated method for the CT-density score.

Because most structural changes (e.g., bronchial wall thickening, mucus plugging) in PCD increase lung density, we based our automated scoring on the identification

of high attenuating structures, leaving out empty bronchiectasis and mosaic perfusion. Thus, the automated score mainly quantifies potentially reversible, inflammatory changes such as mucoid impactions, bronchial wall thickening, bronchiolar nodules and consolidations,

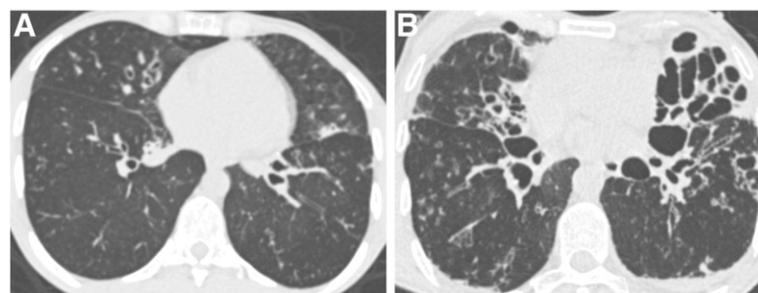


Fig. 5 Different imaging features in patients with similar FEV_1 values. **a** Patient with $FEV_1 = 38\%$ predicted and moderate bronchiectasis predominantly affecting the middle lobe. The CT-density score based on MLD + 1SD was 7.79. **b** Patient with $FEV_1 = 38\%$ predicted but much more severe bronchiectasis on visual assessment, especially in the lingula. The CT-density score based on MLD + 1SD was 12.75. These two examples show that CT imaging provides additional information to spirometry, especially regarding regional disease distribution (homogeneous versus heterogeneous), and the severity of bronchiectasis, which correspond to irreversible changes

whereas it does not consider irreversible changes (e.g., bronchiectasis). Based on these characteristics, we speculate variations in this score may prove useful in identifying worsening of bronchial disease (e.g., during pulmonary exacerbations) or improvement in bronchial patency (e.g., after recovery from a pulmonary exacerbation or due to the beneficial effect of therapy). Due to the unavailability of follow-up chest CT exams for most patients, we were not able to test this hypothesis and future longitudinal studies are needed.

Using our thresholding method, one drawback is that the pulmonary vessels are incorporated into the high-attenuating lung volume, which does not therefore correspond only to the diseased lung. However, differences in pulmonary vessel volume among patients had probably little influence on score variations compared to those due to the bronchial disease, given the good correlation between our score and spirometric parameters. Software capable of pulmonary vessel volume segmentation is being developed [27] and could be used to exclude the pulmonary vessels and improve the performance of the score.

Even though the automated score correlated well with the evaluated functional parameters, we also found that patients with similar FEV₁ values could have quite different CT phenotypes. We believe that quantitative assessment of structural changes is of interest in addition to PFT measurements for both cross-sectional evaluation and disease monitoring. For example, disease progression in CF has long relied on assessment of lung function decline whereas CT scan analysis clearly shows that structural abnormalities may appear without significant changes in FEV₁ [28]. Thus, CT provides structural information which is complementary to spirometry in patients with CF and similar findings are likely occurring in patients with PCD. Calculating CT score does not imply additional procedures for the patients since it can be done from standard CT acquisitions performed as standard of care.

Our study has several limitations. Because this study was retrospective, the CT acquisition parameters were not standardized. This may have influenced the density thresholds. Standardized scanning protocols would probably improve the performance of the developed score. However, the fact that the scoring method can be applied to unstandardized CT examinations makes it suitable for daily clinical use. We did not perform longitudinal evaluation to determine whether changes in the automated CT score correlated with changes in pulmonary function, and whether, as previously suggested, CT-scored disease extent can predict the subsequent decline in pulmonary function. Indeed, our primary objective was to develop and validate an automated CT score by cross-sectional evaluation. Lastly, due to the relative

rarity of PCD, it was not possible to split our population into a development and validation cohort. Thus, the developed method should further be validated in an independent cohort of PCD patients.

Conclusion

In conclusion, automated density-based CT scoring, together with histogram characteristic analysis, is feasible in PCD patients and correlates well with FEV₁ and FVC. MLD + 1SD offered the best correlations with both FEV₁ and FVC. Quantitative analysis of structural abnormalities on CT scans may prove useful for objectively evaluating lung disease changes in PCD, which may prove useful both in daily clinical use and as an outcome in clinical trials.

Abbreviations

CF: Cystic fibrosis; CT: Computed tomography; DLP: Dose-length product; FEV₁: Forced expiratory volume in 1 s; HU: Hounsfield units; ICC: Intraclass correlation coefficients; MLD: Mean lung density; PCD: Primary ciliary dyskinesia; SD: Standard deviation

Acknowledgments

not applicable

Funding

None

Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

TTNH, MPR, PRB, BM and GC were involved in conception and design. TTNH, LB, IH, THH, CM and GC were involved in acquisition and analysis of the data. TTNH, MPR and GC were involved in drafting the manuscript. PRB, LB, IH, THH, CM and BM critically revised the work. All authors agreed to be accountable for all aspects of the work.

Ethics approval and consent to participate

Institutional Review Board of Société Pneumologie de Langue Française. The need for informed consent was waived.

Consent for publication

not applicable

Competing interests

MPR and GC have made a patent application for the image analysis method presented in this article. The other authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Radiology Department, Groupe Hospitalier Cochin-Hôtel Dieu, AP-HP, Université Paris Descartes - Sorbonne Paris Cité, Paris, France. ²Pulmonary Department, Groupe Hospitalier Cochin-Hôtel Dieu, AP-HP, Université Paris Descartes - Sorbonne Paris Cité, Paris, France. ³Service de Pneumologie et de Pathologie Professionnelle, DHU A-TVb, Centre Hospitalier Intercommunal de Créteil, Université Paris Est Créteil, Créteil, France. ⁴Physiology Department, Groupe Hospitalier Cochin-Hôtel Dieu, AP-HP, Université Paris Descartes - Sorbonne Paris Cité, Paris, France. ⁵Center for Visual Computing, CentraleSupélec, Gif-sur-Yvette, France. ⁶Department Diagnostic Imaging, Vinmec International Hospital - Central Park, Ho Chi Minh City, Vietnam.

Received: 28 August 2018 Accepted: 4 December 2018
Published online: 18 December 2018

References

- Afzelius BA. A human syndrome caused by immotile cilia. *Science*. 1976;193:317–9.
- Meeks M, Bush A. Primary ciliary dyskinesia (PCD). *Pediatr Pulmonol*. 2000;29:307–16.
- Honoré I, Burgel P-R. Primary ciliary dyskinesia in adults. *Rev Mal Respir*. 2016;33:165–89.
- Barbato A, Frischer T, Kuehni CE, Snijders D, Azevedo I, Baktai G, et al. Primary ciliary dyskinesia: a consensus statement on diagnostic and treatment approaches in children. *Eur Respir J*. 2009;34:1264–76.
- Shah A, Shoemark A, MacNeill SJ, Bhaludin B, Rogers A, Bilton D, et al. A longitudinal study characterising a large adult primary ciliary dyskinesia population. *Eur Respir J*. 2016;48:441–50.
- Kennedy MP, Noone PG, Leigh MW, Zariwala MA, Minnix SL, Knowles MR, et al. High-resolution CT of patients with primary ciliary dyskinesia. *AJR Am J Roentgenol*. 2007;188:1232–8.
- Santamaria F, Montella S, Tiddens HAWM, Guidi G, Casotti V, Maglione M, et al. Structural and functional lung disease in primary ciliary dyskinesia. *Chest*. 2008;134:351–7.
- Maglione M, Bush A, Montella S, Mollica C, Manna A, Esposito A, et al. Progression of lung disease in primary ciliary dyskinesia: is spirometry less accurate than CT? *Pediatr Pulmonol*. 2012;47:498–504.
- Magnin ML, Cros P, Beydon N, Mahloul M, Tamalet A, Escudier E, et al. Longitudinal lung function and structural changes in children with primary ciliary dyskinesia. *Pediatr Pulmonol*. 2012;47:816–25.
- Pifferi M, Bush A, Pioggia G, Caramella D, Tartarisco G, Di Cicco M, et al. Evaluation of pulmonary disease using static lung volumes in primary ciliary dyskinesia. *Thorax*. 2012;67:993–9.
- Cohen-Cymbarknoh M, Simanovsky N, Hiller N, Gileles Hillel A, Shoseyov D, Kerem E. Differences in disease expression between primary ciliary dyskinesia and cystic fibrosis with and without pancreatic insufficiency. *Chest*. 2014;145:738–44.
- Boon M, Vermeulen FL, Gysmans W, Proesmans M, Jorissen M, De Boeck K. Lung structure-function correlation in patients with primary ciliary dyskinesia. *Thorax*. 2015;70:339–45.
- Frija-Masson J, Bassinet L, Honoré I, Dufeu N, Housset B, Coste A, et al. Clinical characteristics, functional respiratory decline and follow-up in adult patients with primary ciliary dyskinesia. *Thorax*. 2017;72:154–60.
- Kennedy MP, Omran H, Leigh MW, Dell S, Morgan L, Molina PL, et al. Congenital heart disease and other heterotaxic defects in a large cohort of patients with primary ciliary dyskinesia. *Circulation*. 2007;115:2814–21.
- Calder AD, Bush A, Brody AS, Owens CM. Scoring of chest CT in children with cystic fibrosis: state of the art. *Pediatr Radiol*. 2014;44:1496–506.
- Gevenois PA, de Maertelaer V, De Vuyst P, Zanen J, Yernault JC. Comparison of computed density and macroscopic morphometry in pulmonary emphysema. *Am J Respir Crit Care Med*. 1995;152:653–7.
- Chassagnon G, Martin C, Burgel P-R, Hubert D, Fajac I, Paragios N, et al. An automated computed tomography score for the cystic fibrosis lung. *Eur Radiol*. 2018.
- Lucas JS, Barbato A, Collins SA, Goutaki M, Behan L, Caudri D, et al. European Respiratory Society guidelines for the diagnosis of primary ciliary dyskinesia. *Eur Respir J*. 2017;49:1601090.
- Chassagnon G, Morel B, Carpentier E, Ducou Le Pointe H, Sirinelli D. Tracheobronchial branching abnormalities: lobe-based classification scheme. *Radiographics*. 2016;36:358–73.
- Bhalla M, Turcios N, Aponte V, Jenkins M, Leitman BS, McCauley DI, et al. Cystic fibrosis: scoring system with thin-section CT. *Radiology*. 1991;179:783–8.
- Goris ML, Zhu HJ, Blankenberg F, Chan F, Robinson TE. An automated approach to quantitative air trapping measurements in mild cystic fibrosis. *Chest*. 2003;123:1655–63.
- Goris ML, Zhu HJ, Robinson TE. A critical discussion of computer analysis in medical imaging. *Proc Am Thorac Soc*. 2007;4:347–9.
- Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26:319–38.
- Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC. Lung volumes and forced ventilatory flows. Report working party standardization of

- lung function tests, European Community for steel and coal. Official statement of the European Respiratory Society. *Eur Respir J Suppl*. 1993;16:5–40.
- Robinson PJ, Kreef L. Pulmonary tissue attenuation with computed tomography: comparison of inspiration and expiration scans. *J Comput Assist Tomogr*. 1979;3:740–8.
 - Levi C, Gray JE, McCullough EC, Hattery RR. The unreliability of CT numbers as absolute values. *AJR Am J Roentgenol*. 1982;139:443–7.
 - Rudyanto RD, Kerkstra S, van Rikxoort EM, Fetita C, Brillet P-Y, Lefevre C, et al. Comparing algorithms for automated vessel segmentation in computed tomography scans of the lung: the VESSEL12 study. *Med Image Anal*. 2014;18:1217–32.
 - Loeve M, Krestin GP, Rosenfeld M, de Bruijne M, Stick SM, Tiddens HA. Chest computed tomography: a validated surrogate endpoint of cystic fibrosis lung disease? *Eur Respir J*. 2013;42:844–57.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



4 PREDICTION DE LA DYSFONCTION CHRONIQUE DU GREFFON PULMONAIRE (DCGP)

4.1 CONTEXTE ET OBJECTIFS

Par rapport à d'autres transplantations d'organes, la transplantation pulmonaire reste associée à un pronostic plus sombre, avec une survie médiane de seulement de six ans après transplantation (299). La principale cause de décès au-delà de la première année est la Dysfonction Chronique du Greffon Pulmonaire (DCGP) ou « Chronic Lung Allograft Dysfunction » (CLAD). Ce terme fait référence à la détérioration de la fonction pulmonaire à la suite d'un rejet chronique et ne se limite pas à un phénotype exclusivement obstructif, appelé la Bronchiolite Oblitérante (BO) ou « Bronchiolitis Obstructive Syndrome » (BOS), mais comprend également un phénotype restrictif de dysfonction chronique du greffon, appelé Syndrome Restrictif d'Allogreffe (SRA) ou « Restrictive Allograft Syndrome » (RAS). La classification des phénotypes de DCGP mise à jour en 2019 par la Société Internationale de Transplantation Cardiaque et Pulmonaire (ISHLT) comprend la bronchiolite obstructive (BO), représentant 70% des dysfonctionnements d'allogreffe, et trois autres entités, les phénotypes SRA, mixtes et indéfinis, représentant les 30% restants (33)(Tableau 4-1). Il est désormais conseillé de commencer les investigations chaque fois que le VEMS diminue de $\geq 10\%$ par rapport à la valeur de base (33). DCGP est « confirmée » si les anomalies physiologiques (baisse du VEMS de $\geq 20\%$) persistent pendant 3 mois après la prise de la première valeur (33). Pour la classification des sous-groupes, en bref, le BO est défini comme un syndrome fonctionnelle respiratoire purement obstructive (rapport VEMS/CVF $< 0,7$ et une capacité pulmonaire totale (CPT) normal) et aucune opacité pulmonaire à l'imagerie thoracique (en radiographie ou en TDM) (33). Le SRA se caractérise par des opacités persistantes à l'imagerie thoracique (radiographie thoracique et/ou TDM) et une baisse de la CPT de plus de 10 % par rapport à la meilleure CPT post-opératoire ou une baisse de la CVF de plus de 20 % par rapport à la meilleure CVF lorsque la CPT n'est pas disponible (300–302). Certains patients restent indéfinis après évaluation fonctionnelle et radiologique. Par exemple, un patient avec un défaut obstructif qui développe des opacités persistantes sur le scanner mais pas de baisse de CPT ou un patient avec un déclin combiné obstructif et restrictif de la fonction pulmonaire sans opacité sur scanner sont des exemples représentatifs d'une telle situation.

Tableau 4-1: Définition des différents phénotypes de DCGP (6).

Phénotypes	Manifestation physiologique	Manifestation radiologique
DCGP*	Diminution persistante $\geq 20\%$ du VEMS (sur la base de 2 valeurs VEMS à au moins 3 semaines d'intervalle) par rapport à la valeur initiale, définie comme la moyenne des 2 meilleures valeurs de VEMS postopératoires, en l'absence d'autres étiologies telles qu'infection ou rejet aigu	
BO**	Diminution persistante $\geq 20\%$ du VEMS par rapport à la valeur initiale (= définition DCGP) ET	

	anomalie ventilatoire obstructive (VEMS / capacité vitale forcée [CVF] <0,7)	
SRA***	Diminution persistante ≥ 20% du VEMS par rapport à la valeur initiale (= définition DCGP) ET diminution ≥ 10% de la Capacité pulmonaire totale (CPT) par rapport à la valeur initiale	Opacités persistantes sur l'imagerie thoracique
Phénotype mixte	Diminution persistante ≥ 20% du VEMS par rapport à la valeur initiale (= définition DCGP) ET combinaison d'anomalies ventilatoires obstructives et restrictives (VEMS / CVF <0,7 et une CPT ≤ 90% de la valeur initiale)	Opacités persistantes sur l'imagerie thoracique
Phénotype indéfini (1)	Diminution persistante ≥ 20% du VEMS par rapport à la valeur initiale (= définition DCGP) ET combinaison d'anomalies ventilatoires obstructives et restrictives (VEMS / CVF <0,7 et une CPT ≤ 90% de la valeur initiale)	
Phénotype indéfini (2)	Diminution persistante ≥ 20% du VEMS par rapport à la valeur initiale (= définition DCGP) ET anomalie ventilatoire obstructive (VEMS / CVF <0,7) et aucune diminution de la CPT	Opacités persistantes sur l'imagerie thoracique

Associer épreuves fonctionnelles respiratoires (EFRs) et imagerie tomodensitométrie est donc obligatoire non seulement pour exclure des causes de dégradation de la fonction pulmonaire autres que la DCGP, mais aussi pour classer les phénotypes de DCGP. Hota et coll. ont fait une description des différents phénotypes de DCGP en tomodensitométrie (TDM) haute résolution préalablement à la mise à jour de 2019 (36). Plus récemment, Brun et coll. ont décrit les caractéristiques tomodensitométriques associées aux différents phénotypes de DCGP tels que définis dans le document de consensus de 2019 (37). Les manifestations TDM de la BO comprennent l'atténuation mosaïque du parenchyme pulmonaire et le piégeage expiratoire, parfois associés à des anomalies des voies respiratoires proximales à type d'épaississement de la paroi bronchique. Des nodules centrolobulaires et des aspects d'arbres en bourgeon peuvent également être observés (37). La présentation TDM du SRA dépend du stade auquel la maladie est découverte(41–43) : Au stade précoce, il s'agit de consolidations et d'images en verre dépoli (41) ; lorsque la fibrose se développe, il s'agit de réticulations et de bronchectasies par traction (41) ; Les aspects tardifs sont proches de ceux de la fibro élastose pleuroparenchymateuse idiopathique, il s'agit d'épaississement pleural, de condensation périphérique et de bronchectasies de traction avec perte de volume, le tout ayant une prédominance supérieure. Les formes diffuses ou avec une prédominance basale des anomalies seraient de plus mauvais pronostic

La classification phénotypique est également importante pour le pronostic. Les patients ayant un phénotype restrictif de DCGP ayant une moins bonne survie (301). Lévy et coll. (303) ont démontré que les patients présentant des phénotypes DCGP non définis/non classés présentant des opacités de type SRA sur l'imagerie thoracique avaient une durée d'allogreffe significativement inférieure à celle des patients présentant le phénotype BO. Suhling et coll. (304) ont réalisé un score visuel des réticulations et des consolidations chez les patients transplantés pulmonaires (LTx) et ont noté une survie inférieure chez les patients présentant des réticulations sévères. Le suivi par imagerie est particulièrement important pour détecter la DCGP chez les patients ayant eu une transplantation monopulmonaire, car l'atteinte du poumon natif peut être confondante (305). Idéalement, la DCGP peut être prévenue par un traitement prophylactique, bien que le traitement immunosuppresseur classique ne semble pas empêcher son développement (47). D'autres traitements prophylactiques ont été étudiés récemment, l'azithromycine (48), la vitamine D (49,306), le montélukast (51), la photophérèse extracorporelle (PEC) (53,63), l'administration de cellules souches mésenchymateuses (56), les agents antifibrotiques tels que la pirfénidone (58), la nintédanib(59) Bien qu'aucun traitement ne soit actuellement disponible pour inverser le DCGP après le diagnostic, l'identification précoce du DCGP permettrait des stratégies proactives et ciblées pour ralentir la progression de la maladie avant que des dommages irréversibles de l'allogreffe ne se produisent(63). D'ailleurs, la retransplantation pulmonaire est un dernier recours en phase terminale. Certains centres très expérimentés, rapportent une survie similaire après retransplantation pour DCGP par rapport aux transplantations primaires (60,61), bien que la période postopératoire soit beaucoup plus difficile et que la restauration de la fonction pulmonaire soit moindre (62). Compte tenu du mauvais pronostic de la SRA, des stratégies de traitement préventif ou visant à ralentir sa progression sont absolument nécessaires. Comme le rejet chronique subclinique compromet l'intégrité de l'allogreffe, il est nécessaire de développer des tests physiologiques et d'imagerie permettant une détection précoce de la DCGP, avec un rôle potentiel pour l'imagerie quantitative pour servir de biomarqueur (63).

Dans une première partie, nous présenterons une synthèse des méthodes tomodensitométriques quantitatives rapportées pour la détection et la surveillance en imagerie TDM de la DCGP. La seconde partie rapportera notre expérience d'une approche fondée sur la radiomique – technique qui consiste à identifier à partir de caractéristiques d'image, non perceptibles à l'œil humain, de possible biomarqueurs, pour permettre un diagnostic précoce de la DCGP.

4.2 PUBLICATION

Chronic Lung Allograft Dysfunction Post Lung Transplantation: A Review of Computed Tomography Quantitative Methods for Detection and Follow-Up.

Trieu-Nghi Hoang-Thi, Guillaume Chassagnon, Thong Hua-Huy, Veronique Boussaud, Anh-Tuan Dinh-Xuan and Marie-Pierre Revel.

J. Clin. Med. **2021**, *10*, 1608.
<https://doi.org/10.3390/jcm10081608>

4.2.1 RESUME

La dysfonction chronique du greffon pulmonaire (DCGP) demeure la principale cause de morbidité et de mortalité après transplantation pulmonaire. Le terme englobe à la fois les phénotypes obstructifs et restrictifs, ainsi que les phénotypes mixtes et indéfinis. L'imagerie, à côté des tests de fonction pulmonaire, joue un rôle majeur dans l'identification du phénotype de DCGP et est essentielle pour le suivi après transplantation pulmonaire. L'imagerie quantitative permet d'effectuer une évaluation précise et indépendante de l'interprétation humaine des examens d'imagerie. Dans cet article de synthèse, nous discuterons du rôle des méthodes d'imagerie quantitative pour évaluer les voies respiratoires et le parenchyme pulmonaire sur des images de tomodensitométrie (TDM), pour une identification précoce de DCGP et pour une estimation de pronostic. Nous discuterons également de leurs limites et de la nécessité de nouvelles approches pour prédire, comprendre et identifier le DCGP dans ses débuts.

4.2.2 ARTICLE

Review

Chronic Lung Allograft Dysfunction Post Lung Transplantation: A Review of Computed Tomography Quantitative Methods for Detection and Follow-Up

Trieu-Nghi Hoang-Thi ^{1,2,3} , Guillaume Chassagnon ¹, Thong Hua-Huy ³, Veronique Boussaud ⁴, Anh-Tuan Dinh-Xuan ³  and Marie-Pierre Revel ^{1,*}

- ¹ AP-HP.Centre, Hôpital Cochin, Department of Radiology, Université de Paris, 75014 Paris, France; htrieunghi@yahoo.fr (T.-N.H.-T.); guillaume.chassagnon@aphp.fr (G.C.)
² Department of Diagnostic Imaging, Vinmec Central Park Hospital, Ho Chi Minh City 70000, Vietnam
³ AP-HP.Centre, Hôpital Cochin, Department of Respiratory Physiology, Université de Paris, 75014 Paris, France; huy-thong.hua@aphp.fr (T.H.-H.); anh-tuan.dinh-xuan@aphp.fr (A.-T.D.-X.)
⁴ AP-HP.Centre, Hôpital Cochin, Department of Pneumology, Université de Paris, 75014 Paris, France; veronique.boussaud@aphp.fr
* Correspondence: marie-pierre.revel@aphp.fr; Tel.: +33-1-5841-2471



Citation: Hoang-Thi, T.-N.; Chassagnon, G.; Hua-Huy, T.; Boussaud, V.; Dinh-Xuan, A.-T.; Revel, M.-P. Chronic Lung Allograft Dysfunction Post Lung Transplantation: A Review of Computed Tomography Quantitative Methods for Detection and Follow-Up. *J. Clin. Med.* **2021**, *10*, 1608. <https://doi.org/10.3390/jcm10081608>

Academic Editor: Takashi Ohtsuka

Received: 6 March 2021
Accepted: 8 April 2021
Published: 10 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Chronic lung allograft dysfunction (CLAD) remains the leading cause of morbidity and mortality after lung transplantation. The term encompasses both obstructive and restrictive phenotypes, as well as mixed and undefined phenotypes. Imaging, in addition to pulmonary function tests, plays a major role in identifying the CLAD phenotype and is essential for follow-up after lung transplantation. Quantitative imaging allows for the performing of reader-independent precise evaluation of CT examinations. In this review article, we will discuss the role of quantitative imaging methods for evaluating the airways and the lung parenchyma on computed tomography (CT) images, for an early identification of CLAD and for prognostic estimation. We will also discuss their limits and the need for novel approaches to predict, understand, and identify CLAD in its early stages.

Keywords: lung transplantation; image processing; computer-assisted; pulmonary disease; chronic obstructive; bronchiolitis obliterans

1. Introduction

Compared to other organ transplantation, lung transplantation remains associated with a poorer prognosis, with a median survival of only six years after transplantation [1]. The main cause of death beyond the first year is chronic lung allograft dysfunction (CLAD). This term refers to the deterioration of lung function as a result of chronic rejection and is not limited to an exclusively obstructive phenotype, referred to as bronchiolitis obstructive syndrome (BOS), but also includes a restrictive phenotype of chronic allograft dysfunction, termed restrictive allograft syndrome (RAS). The classification of CLAD phenotypes updated in 2019 by the International Society of Heart and Lung Transplantation (ISHLT) included BOS, accounting for 70% of allograft dysfunctions, and three other entities, RAS, mixed, and undefined phenotypes, representing the other 30% of phenotypes [2]. Their respective definitions are summarized in Table 1.

Hota et al. provided a comprehensive description of the different CLAD phenotypes on high-resolution CT [3]. RAS is associated with persistent parenchymal opacities and pleural thickening showing upper lung predominance, with pleuro parenchymal fibroelastosis on pathological examination, explaining the restrictive pattern [2,3]. Complementing pulmonary function tests (PFTs) with CT is thus mandatory not only to exclude other causes of lung function degradation but also to classify CLAD phenotypes. This classification is important for prognosis, with patients with a restrictive pattern of CLAD having worse survival [4]. Furthermore, Levy et al. [5] demonstrated that patients with

undefined/unclassified CLAD phenotypes who had RAS-like opacities on chest imaging had significantly worse allograft survival than patients with the BOS phenotype. Suhling et al. [6] performed visual scoring of reticulations and consolidations in lung transplanted (LTx) patients and found poorer survival in patients with severe reticular changes. Imaging follow-up is especially important to detect CLAD in patients with single lung transplantation, because native lung disease may confound the interpretation of physiology [7].

Table 1. Definition of different CLAD phenotypes.

Phenotypes	Physiological Changes	Radiological Changes
CLAD (Chronic lung allograft dysfunction)	Persistent $\geq 20\%$ decline in FEV1 (on the basis of 2 FEV1 values at least 3 weeks apart) compared with the baseline value, defined as the mean of the best 2 post-operative FEV1 measurement values, in the absence of other etiologies such as infection or acute rejection	
BOS (Bronchiolitis obliterans syndrome)	Persistent $\geq 20\%$ decline in FEV1 compared with the baseline value (=CLAD definition) AND obstructive ventilatory defect (FEV1/forced vital capacity [FVC] < 0.7)	
RAS (Restrictive allograft syndrome)	Persistent $\geq 20\%$ decline in FEV1 compared with the baseline value (=CLAD definition) AND $\geq 10\%$ decline in TLC relative to baseline	Persistent opacities on chest imaging
Mixed phenotype	Persistent $\geq 20\%$ decline in FEV1 compared with the baseline value (=CLAD definition) AND combination of obstructive and restrictive ventilatory defect (FEV1/FVC < 0.7 and a TLC $\leq 90\%$ of baseline)	Persistent opacities on chest imaging
Undefined phenotype (1)	Persistent $\geq 20\%$ decline in FEV1 compared with the baseline value (=CLAD definition) AND combination of obstructive and restrictive ventilatory defect (FEV1/FVC < 0.7 and a TLC $\leq 90\%$ of baseline)	
Undefined phenotype (2)	Persistent $\geq 20\%$ decline in FEV1 compared with the baseline value (=CLAD definition) AND obstructive ventilatory defect (FEV1/FVC < 0.7) and no decline in TLC	Persistent opacities on chest imaging

A switch from visual descriptive evaluation to quantitative automatic measurements has recently emerged in the field of medical imaging. Quantitative imaging has been mainly used for the evaluation of BOS, the obstructive phenotype of CLAD. The implementation of such imaging analysis tools for CLAD evaluation in LTx patients either directly confirms airway thickening and vascular remodeling in the allograft lung [8,9] or indirectly suggests their occurrence through air-trapping quantification [10,11]. Applying machine learning to imaging data analysis may further improve the management of CLAD [12]. In this review article, we will provide an overview of the CT quantification methods used to investigate CLAD of various phenotypes (see Figure 1 and Table 2), and discuss novel approaches to identify CLAD in its early stage. As sub-clinical chronic rejection compromises allograft integrity, there is a need to develop physiological and imaging tests allowing early detection of CLAD, with a potential role for quantitative imaging to serve as a biomarker, allowing early identification of CLAD [13].

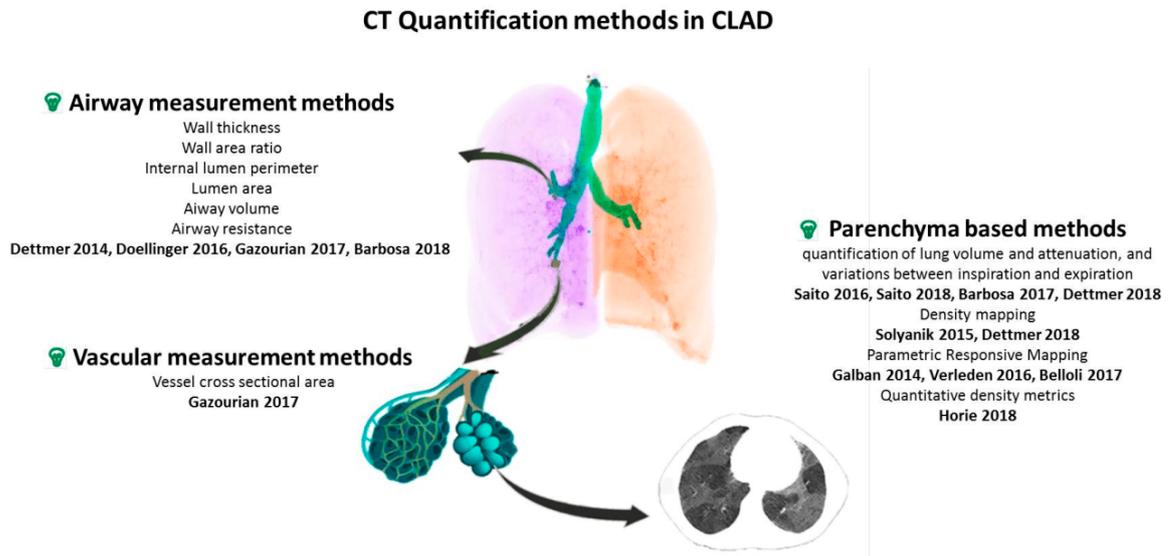


Figure 1. Overview of CT quantification methods for CLAD. CT: computer tomography, CLAD: chronic lung allograft dysfunction. Airway measurement methods [8,9,12,14], Vascular measurement methods [14], Parenchyma based methods [10,11,15–21].

Table 2. Overview of CT quantification methods for CLAD.

Author Year	Year	Study Design/Number of Patients	Time of CT Evaluation	Software	Main Quantification Parameters	Main Results
Airway measurement methods						
Dettmer [9]	2014	Prospective study 141 patients (25 BOS+)	6, 12, 24 months after LTx	MeVis Airway Examiner	WT, WA%, WTdiff between inspiration and expiration on two selected bronchi B01 and B10	Greater WA% on inspiration in BOS+
Doellinger [8]	2016	Retrospective study 26 patients (12 BOS+)	All available CT scans after LTx	YACTA module v.1.0.7.16	Δ WT and Δ WA%: temporal change of WT and WA%;	Temporal changes of WT and WA% showed significant differences between BOS+ and BOS–
Gazourian [14]	2017	Retrospective study 66 patients (20 controls, 22 BOS non progressors and 24 BOS progressors)	non-volumetric CT closest to baseline FEV ₁ and 2 follow-up CT scans	Airway Inspector (www.airwayinspector.org)	Internal lumen perimeter Lumen airway Airway vessel (A/V) ratio	Increase in the A/V ratio on follow-up CT scans for BOS progressors
Barbosa [12]	2018	Retrospective study 71 patients (41 BOS+)	2 CT scans (>3 months apart)	Mimics, TGrid 14.0 and Fluent 14.0	Airway volumes Airway resistances	Increase in central airway volume on expiratory CT in BOS+ Smaller airway volumes and airway surfaces and higher airway resistances at baseline in BOS developers
Vascular measurement method						
Gazourian [14]	2017	Retrospective study 22 patients (13 BOS+)	2 volumetric CT angiographies after LTx	Bronchi: Airway Inspector Vessels: Upper thresholding (cut-off level of –500 HU) and use of a connected components technique	Vessel cross sectional area (CSA) Airway lumen area Airway/Vascular ratio (A/V ratio)	Overtime decrease in CSA in BOS+ Overtime Increase in A/V ratio in BOS+
Parenchyma-based methods: quantification of air trapping						
Belloli [11]	2017	Retrospective study (22 BOS+ and controls matched by time from LTx; 52 BOS+)	Date of BOS	Lung segmentation: In-house algorithm Insp/Expiratory registration algorithm: Elastix	Parametric response mapping (PRM): Density-based quantification of air trapping (PRM ^{ISAD}) and parenchymal disease (PRM ^{PD})	FEV ₁ decline associated with higher PRM ^{ISAD} FEV ₁ and FVC decline associated with higher PRM ^{PD} PRM ^{ISAD} \geq 30% strongest predictor of death

Table 2. Cont.

Author Year	Year	Study Design/Number of Patients	Time of CT Evaluation	Software	Main Quantification Parameters	Main Results
Verleden [10]	2016	Retrospective study 40 patients (20 BOS+)	CT scans before, at the time of and after the diagnostic of BOS	Lung segmentation: In-house algorithm Insp/Expiratory registration algorithm: Elastix	Density-based quantification of air trapping (PRM^{fSAD}), parenchymal disease (PRM^{PD}) and normal lung (PRM^{Normal})	Increase in PRM^{fSAD} and decrease in PRM^{Normal} in BOS+ No difference in PRM^{fSAD} between BOS- and BOS+ before the diagnosis of BOS
Solyanik [15]	2015	Prospective study 147 patients (34 with air-trapping)	CT at 6 months after LTx	Not mentioned	Density-based quantification of air trapping (EXP. _{-790 HU to -950 HU} , E/I-MLD) Density mapping: voxel-to-voxel insp/expiration mapping	DM has the highest correlation to RV/TLC ($r = 0.663, p < 0.001$) DM and E/I-ratio MLD showed better correlation with RV/TLC than EXP. _{-790HU to -950HU}
Barbosa [16]	2017	Retrospective study 174 patients (98 BOS+)	CTs within 9 years after LTx	ANTs package	- Lung volume in inspiration and expiration - Lung volume difference between insp and expiration - Density-based quantification of air trapping (EXP. _{<.856 HU} , voxel volume with <75 HU increase on expiration)	Only 59% of qCT parameters associated with BOS+ BOS prediction model combining qCT and PFT parameters outperforms model based on PFTs alone in the unilateral LTx group
Dettmer [17]	2018	Prospective study 51 patients (17 BOS+)	Last CT within 1 year before BOS diagnostic First CT within 1 year after BOS diagnostic	Mevis Pulmo	Density-based quantification of air trapping (E/I-MLD ratio, E/I Volumes, density percentiles)	Significant increase in E/I-Volumes and decrease in E/I-MLD in BOS+ Changes more pronounced in the lower lobes Highest AUC for 10th percentile on expiration (0.903) and E/I-MLD ratio (AUC: 0.886)
Horie [18]	2018	Retrospective study 74 patients (23 RAS, 51 BOS)	CT performed ± 4 months from CLAD and/or RAS/BOS onset.	Lung segmentation on Vitrea workstation	Lung volume and MLD on inspiration Quantitative density metrics (QDM) defined as ratios of the right and left quantile weights of the density histogram on inspiratory CT	Significant difference of Lung volume and MLD in BOS and RAS patients Hazard ratio for death 3.2 times higher at the 75th percentile of QDM1 compared to the 25th percentile

Table 2. *Cont.*

Author Year	Year	Study Design/Number of Patients	Time of CT Evaluation	Software	Main Quantification Parameters	Main Results
Saito [20]	2016	Retrospective study 63 patients (19 RAS, 44 BOS)	CT performed at baseline and time of CLAD onset	Lung segmentation on Vitrea workstation	Lung volume on inspiration	Decrease in CT lung volume in RAS patients CT volumetry < 90% baseline had an accuracy of 0.937 for differentiating RAS from BOS
Saito [21]	2018	Retrospective study 58 patients, 14 CLAD	CT performed 3, 6, and 12 months after LTx and once yearly thereafter	Lung segmentation on Synapse Vincent workstation	Lung volume on inspiration and expiration Evaluation of Δ lung volume over time (difference between inspiration and expiration)	Δ lung volume onset/baseline significantly decreased in the CLAD group 0.80 cutoff had an AUC of 0.87

LTx: lung transplantation; WT: wall thickness; WA%: wall area percentage = ratio bronchial wall/total area (bronchial wall+ bronchial lumen); WTdiff: difference of wall thickness between expiration and inspiration; A/V ratio: airway vessel ratio; Δ WT and Δ WA%: temporal changes of WT and WA%; PRM^{SAD}: Parametric Response Mapping representing functional small airways disease; PRM^{PD}: Parametric Response Mapping representing parenchymal disease; PRM^{Normal}: Parametric Response Mapping representing normal pulmonary parenchyma; EXP._{-790 HU to -950 HU}: percentage of voxels with attenuation values from -790 HU to -950 HU on expiration; DM: density mapping; E/I-MLD: expiratory-to-inspiratory mean lung density ratio; E/I- Volumes: expiratory-to-inspiratory volume ratio; RV/TLC: residual volume/total lung capacity qCT: quantitative CT.

2. Airway Measurement Methods: Computer-Assisted Airway Morphometry

Since the most common phenotype of CLAD is that of BOS, quantitative methods allowing the direct assessment of the airways have been developed. BOS primarily affects the small airways, with diameters of less than 2 mm. Due to the limited spatial resolution of clinical CT scanners, new measurement methods had to be developed to allow the measurement of the density and wall thickness of the small airways [22]. Such a method made it possible to obtain values of the wall thickness of the small airways correlated with FEV1 in COPD patients [23]. Several software now available commercially make it possible to obtain an automatic segmentation of the bronchial tree and sections strictly perpendicular to the bronchial axis considered at different levels, as well as an automatic measurement of the bronchial wall thickness (WT) and the wall area ratio (WA%), the ratio of the airway wall area to the whole airway area.

Using the YACTA software [24], Doellinger et al. retrospectively examined a total of 2190 airway cross sections on CT scans performed within 11 years of follow-up of 26 LTx patients with at least one measurement in every lobe of the transplanted lung [8]. They showed significant differences between patients with and without BOS in term of changes over time of WA% and WT and concluded that bronchial wall thickening and luminal dilatation observed in lung transplant allograft rejection can be detected and quantified using computer-assisted airway morphometry. Gazourian et al., in a retrospective analysis of 22 patients, observed and increase in the airway internal lumen perimeter (PI), and airway lumen area (AI) in the 13 patients developing BOS [14]. They used Airway Inspector (now renamed Chest Imaging Platform) for evaluating the third and fourth generation B1 and B10 airways. They also evaluated the pulmonary vessels cross-sectional areas (CSA) after upper thresholding (cut-off level of -500 HU) and use of a connected components technique on volumetric CT pulmonary angiogram to isolate the vessel cross sections. These authors also measured the airway/vessel ratios (A/V ratio) on non-volumetric CT scans. They found a statistically significant decrease in the vessel CSA in BOS patients and a significant increase in the airway/vessel ratio in BOS progressors. This observation is in line with known physiology, namely vasoconstriction in poorly ventilated areas due to small airways disease. These two studies (Doellinger [8] and Gazourian [14]), although both based on a limited number, showed differences between BOS and controls, but automated quantification of airway dimension was limited by the need for manual adjustments. In addition, the methods used do not translate into early identification or prognostic tools for CLAD.

Dettmer [9] evaluated 25 patients with BOS and 116 controls using MeVis Airway Examiner. WA% on inspiration was significantly greater in patients with BOS, but the variability of bronchial wall measurements was high and the values for the WA% on inspiration in patients with and without BOS overlapped considerably, due to variable underlying lung volumes. Even though this limitation could be overcome by performing spirometry-controlled CT acquisitions, the authors concluded to a limited value of WA% for establishing a diagnosis of BOS in individual patients.

Barbosa et al. [12] conducted a retrospective analysis of LTx patients who had paired inspiratory and expiratory CTs, with an objective to detect BOS 0-p stage, as indicator of early disease ($\geq 10\%$ decline in FEV1 or $\geq 25\%$ decline in forced expiratory flow 25–75%). Baseline prediction of BOS was performed in a cohort of 41 patients, of which 15 developed BOS. They measured the airway volumes and airway resistances after 3D analysis of the airways and lung lobes using three different commercially available software, Mimics, TGrid 14.0 and Fluent 14.0, for computational fluid dynamics simulations. The authors found that BOS patients experienced an increase in central airway volume on expiratory CT, whereas patients with other cause of FEV1 decline had a decrease in the central airway volume on inspiratory CT. According to the authors, these findings reflected an increase in the extent of air trapping and peribronchiolar fibrosis which characterizes the BOS phenotype of CLAD. Image post-processing included segmentation of the tracheobronchial tree down to the level of airways with a diameter of 1–2 mm which required manual

correction and took 2 to 6 h per scan, which is the main limitation to consider the clinical use of this approach.

3. Lung Parenchyma Methods: Assessment of Lung Volume and Attenuation, and Their Variations between Inspiration and Expiration

Barbosa et al. [16], using the Advanced Normalization Tools (ANTs) software package, calculated the aerated lung volume in inspiration and expiration via exclusion of any voxel >-50 HU and then evaluated the volume change between inspiration and expiration as the difference between these two values, in a retrospective patient cohort excluding the RAS phenotype. In addition to volume change, two other quantitative parameters reflecting air trapping were evaluated: the volume of voxels with attenuation <-856 HU on expiration and the volume of voxels with an increase in attenuation <75 HU from inspiration to expiration following non-rigid registration between the inspiratory and expiratory imaging datasets. Only 59% of quantitative CT metrics were significantly correlated with BOS status in this study including 176 LTx patients, and none of the variables alone was a good predictor of BOS. However, a support vector machine (SVM) model based on the quantitative CT variables outperformed models using visual scoring of CT anomalies or PFT for BOS prediction after unilateral LTx.

Dettmer et al. [17] conducted a prospective evaluation with the objective to detect the BOS phenotype of CLAD in a cohort of 122 LTx patients. They performed CT acquisition with a spirometry-controlled technique at full inspiration and end of expiration and used Mevis Pulmo software, allowing volume measurement and histogram analysis (mean lung density/MLD, peak, percentiles) for the whole lung and separately for each lobe. They demonstrated that in patients with early-stage BOS, the lower lobe volume increased and the MLD decreased between the baseline and follow-up examinations, whereas the volume and MLD in the upper lobes remained nearly constant. The histogram parameters showing the highest accuracies for early BOS detection were the 10th percentile on expiration (AUC: 0.903) and the expiratory-to-inspiratory (E/I) MLD ratio (AUC: 0.886).

E/I MLD ratio is one of the multiple quantitative parameters used to evaluate the volume of air trapping in BOS. Solyanik et al. [15] evaluated two other automated methods to quantify air trapping, which had only slight differences with those previously mentioned in the study by Barbosa [16]: the lung volume having attenuation values ranging from -950 to -790 HU on expiratory CT scans (attenuation <-856 HU for Barbosa) and the lung volume with less than 80 HU change (<75 HU change for Barbosa) between inspiration and expiration. The latter requires non-rigid registration of inspiration-expiration CT-data and voxel-to-voxel mapping, which Solyanik described as “density mapping”. Among the three evaluated methods, density mapping showed the best correlation with the ratio of residual volume to total lung capacity (RV/TLC).

Parametric response mapping, similar to density mapping, involves assessing the attenuation of each voxel on inspiration and expiration after applying an elastic registration algorithm. Voxels with attenuation ≥ 950 HU and <-810 HU at inspiration and <-856 at expiration are considered to represent functional small airways disease (PRMfSAD). Belloli et al. [11] evaluating a retrospective cohort of 52 LTx patients reported that PRMfSAD values $\geq 30\%$ were the strongest predictor of survival in a multivariable model including BOS grade and baseline FEV1% predicted. Verleden et al. [10] used PRM to monitor BOS progression in a retrospective study including 20 BOS and 20 controls (no restrictive CLAD phenotype). They observed an increase at time of BOS diagnosis. They also reported that patients who died from BOS had significantly higher PRMfSAD than living patients. Galban et al. investigated the use of PRM as an imaging biomarker in the diagnosis of BOS. They found that PRMfSAD $> 28\%$ was highly indicative of BOS occurrence, whether a concurrent infection was present or not [19].

Other authors have developed quantitative methods only requiring inspiratory CT data. This was the case for Horie et al. [18], who evaluated CT lung density histograms on a single inspiratory CT in CLAD patients after lung segmentation. Indeed, lung fibrosis associated with the RAS phenotype of CLAD is likely to increase lung density, whereas

mosaic perfusion secondary to BOS has an invert effect, both being detectable when analyzing the lung density histogram: the right-sided tail reflects processes increasing lung density whereas the left-sided tail reflects processes decreasing lung density, such as mosaic attenuation in areas of BOS. Their objective was threefold, to evaluate prognosis after CLAD onset, distinguish between BOS and RAS phenotypes, and evaluate the prognosis within BOS and RAS patient groups separately. They evaluated the quantitative density metrics (QDM) defined as ratios of the right and left quantile weights of the histogram, reflecting the left vs. right asymmetry of the histogram. There was a statistically significant difference in QDMs between RAS and BOS patients. Higher QDM values were significantly associated with decreased survival. The authors concluded that this quantitative analysis of CT images was associated with survival after the onset of CLAD and was able to differentiate RAS and BOS phenotypes. QDM measurement has been used by the same authors [25] to predict the risk of subsequent CLAD in patients with CLAD-0p, defined as a drop in FEV1 to 81–90% of baseline. Higher QDM values were associated with a shorter time between CLAD-0p-CT and CLAD.

Fewer quantitative CT methods have been developed for the RAS phenotype of CLAD. Measurement of inspiratory and expiratory CT lung volumes can be easily performed and CT volumetry has been shown to differentiate RAS patients from BOS patients due to the strong positive correlation between inspiratory lung CT volumes and TLC [20]. Computed tomography is particularly useful for detecting CLAD affecting a single lung [7,21], because the unaffected contralateral lung compensates for the deficit in lung function on PFTs. A decrease over time in the difference between inspiratory and expiratory lung volumes has been reported for obstructive and restrictive CLAD phenotypes [21].

PRM is another method allowing quantitative evaluation of patients with the RAS phenotype of CLAD, who have persistent parenchymal opacities and pleural thickening. Voxels with attenuation values ≥ -810 HU at inspiration represent parenchymal disease (PRMPD). In the study by Belloli et al. [11] evaluating 22 LTx patients, those with concurrent declines in both FEV1 and FVC (e.g., RAS phenotype) were found to have significantly more PRMPD than their control group, even after adjusting for age, baseline FEV1% predicted, and baseline FVC % predicted. The prognostic value of PRMPD was not mentioned and should probably be evaluated in a larger patient group.

To be clinically useful, quantitative methods should not require multiple complex algorithms or manual steps to correct segmentation, which is the case with computer-assisted airway morphometry. In addition, these methods only assess the BOS phenotype of CLAD and have limited diagnostic value. Simpler methods based on CT volumetry are more accessible, with the assessment of lung volume and lobe attenuation on inspiration and expiration of interest for all CLAD phenotypes.

Regarding the processing of imaging data by artificial intelligence for CLAD diagnosis, the current literature is scarce. Classical machine learning methods such as the support vector machine (SVM) using quantitative CT metrics were of interest to diagnose BOS following unilateral LTx, as previously mentioned [16]. SVM was also used to evaluate multiple combinations of functional respiratory indexes (FRI) for BOS prediction at six months after transplantation [12]. A maximal accuracy of 85% was obtained by combining three baseline FRI features: the right middle lobe volume at total lung capacity (inspiratory CT scan), the right upper lobe airway resistance and the central airway surface at functional residual capacity (expiratory CT). Deep learning represents a major advance in the field of artificial intelligence applied to medical imaging, but requires a large amount of data [26,27]. As previously highlighted, most series on quantitative imaging for CLAD evaluation post LTx are based on a limited number of patients, because lung transplantation is not a common procedure. In 2015, only 14 centers reported performing 50 or more LTx per year [28]. Therefore, the standardization of imaging follow-up and data sharing through a common registry is crucial before considering the development of deep learning algorithms for the prediction and early diagnosis of CLAD.

4. Conclusions

Quantitative imaging methods offer the opportunity to perform reader-independent assessment of the airways and lung parenchyma in LTx patients. Although said to be automatic, most methods still require significant time-consuming manual corrections, and require the availability of segmentation and elastic registration algorithms, which limits their direct use in clinical routine. However, the results obtained in the studies published to date demonstrate the prognostic impact of methods such as parametric response mapping quantifying functional small airway disease (PRMfSAD) or quantitative density metrics (QDM). The performance for the prediction or early detection of CLAD needs to be strengthened, which could be an objective of deep learning-based methods. Developing prediction models is indeed important to improve outcomes for patients who are developing CLAD and better understand the underlining pathophysiology.

Author Contributions: Conceptualization, T.-N.H.-T. and M.-P.R.; methodology, G.C.; Literature research T.-N.H.-T. and T.H.-H.; writing—original draft preparation, T.-N.H.-T.; writing—review and editing, all authors.; supervision, M.-P.R. and A.-T.D.-X. All authors have read and agreed to the published version of the manuscript.

Funding: This review article received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Chambers, D.C.; Yusef, R.D.; Cherikh, W.S.; Goldfarb, S.B.; Kucheryavaya, A.Y.; Khusch, K.; Levvey, B.J.; Lund, L.H.; Meiser, B.; Rossano, J.W.; et al. The Registry of the International Society for Heart and Lung Transplantation: Thirty-Fourth Adult Lung and Heart-Lung Transplantation Report—2017; Focus Theme: Allograft Ischemic Time. *J. Heart Lung. Transplant.* **2017**, *36*, 1047–1059. [[CrossRef](#)]
- Verleden, G.M.; Glanville, A.R.; Lease, E.D.; Fisher, A.J.; Calabrese, F.; Corris, P.A.; Ensor, C.R.; Gottlieb, J.; Hachem, R.R.; Lama, V.; et al. Chronic Lung Allograft Dysfunction: Definition, Diagnostic Criteria, and Approaches to Treatment—A Consensus Report from the Pulmonary Council of the ISHLT. *J. Heart Lung. Transplant.* **2019**, *38*, 493–503. [[CrossRef](#)]
- Hota, P.; Dass, C.; Kumaran, M.; Simpson, S. High-Resolution CT Findings of Obstructive and Restrictive Phenotypes of Chronic Lung Allograft Dysfunction: More Than Just Bronchiolitis Obliterans Syndrome. *Am. J. Roentgenol.* **2018**, *211*, W13–W21. [[CrossRef](#)] [[PubMed](#)]
- DerHovanesian, A.; Todd, J.L.; Zhang, A.; Li, N.; Mayalall, A.; Finlen Copeland, C.A.; Shino, M.; Pavlisko, E.N.; Wallace, W.D.; Gregson, A.; et al. Validation and Refinement of Chronic Lung Allograft Dysfunction Phenotypes in Bilateral and Single Lung Recipients. *Ann. Am. Thorac. Soc.* **2016**, *13*, 627–635. [[CrossRef](#)] [[PubMed](#)]
- Levy, L.; Huszti, E.; Renaud-Picard, B.; Berra, G.; Kawashima, M.; Takahagi, A.; Fuchs, E.; Ghany, R.; Moshkelgosha, S.; Keshavjee, S.; et al. Risk Assessment of Chronic Lung Allograft Dysfunction Phenotypes: Validation and Proposed Refinement of the 2019 International Society for Heart and Lung Transplantation Classification System. *J. Heart Lung. Transplant.* **2020**, *39*, 761–770. [[CrossRef](#)]
- Suhling, H.; Dettmer, S.; Greer, M.; Fuehner, T.; Avsar, M.; Haverich, A.; Welte, T.; Gottlieb, J. Phenotyping Chronic Lung Allograft Dysfunction Using Body Plethysmography and Computed Tomography. *Am. J. Transplant.* **2016**, *16*, 3163–3170. [[CrossRef](#)] [[PubMed](#)]
- Philippot, Q.; Debray, M.-P.; Bun, R.; Frija-Masson, J.; Bunel, V.; Morer, L.; Roux, A.; Picard, C.; Jebrak, G.; Dauriat, G.; et al. Use of CT-SCAN Score and Volume Measures to Early Identify Restrictive Allograft Syndrome in Single Lung Transplant Recipients. *J. Heart Lung. Transplant.* **2020**, *39*, 125–133. [[CrossRef](#)]
- Doellinger, F.; Weinheimer, O.; Zwiener, I.; Mayer, E.; Buhl, R.; Fahlenkamp, U.L.; Dueber, C.; Achenbach, T. Differences of Airway Dimensions between Patients with and without Bronchiolitis Obliterans Syndrome after Lung Transplantation-Computer-Assisted Quantification of Computed Tomography. *Eur. J. Radiol.* **2016**, *85*, 1414–1420. [[CrossRef](#)]
- Dettmer, S.; Peters, L.; de Wall, C.; Schaefer-Prokop, C.; Schmidt, M.; Warnecke, G.; Gottlieb, J.; Wacker, F.; Shin, H. Bronchial Wall Measurements in Patients after Lung Transplantation: Evaluation of the Diagnostic Value for the Diagnosis of Bronchiolitis Obliterans Syndrome. *PLoS ONE* **2014**, *9*, e93783. [[CrossRef](#)]
- Verleden, S.E.; Vos, R.; Vandermeulen, E.; Ruttens, D.; Bellon, H.; Heigl, T.; Van Raemdonck, D.E.; Verleden, G.M.; Lama, V.; Ross, B.D.; et al. Parametric Response Mapping of Bronchiolitis Obliterans Syndrome Progression after Lung Transplantation. *Am. J. Transplant.* **2016**, *16*, 3262–3269. [[CrossRef](#)]

11. Belloli, E.A.; Degtiar, I.; Wang, X.; Yanik, G.A.; Stuckey, L.J.; Verleden, S.E.; Kazerooni, E.A.; Ross, B.D.; Murray, S.; Galbán, C.J.; et al. Parametric Response Mapping as an Imaging Biomarker in Lung Transplant Recipients. *Am. J. Respir. Crit. Care Med.* **2017**, *195*, 942–952. [[CrossRef](#)]
12. Barbosa, E.J.M.; Lanclus, M.; Vos, W.; Van Holsbeke, C.; De Backer, W.; De Backer, J.; Lee, J. Machine Learning Algorithms Utilizing Quantitative CT Features May Predict Eventual Onset of Bronchiolitis Obliterans Syndrome After Lung Transplantation. *Acad. Radiol.* **2018**, *25*, 1201–1212. [[CrossRef](#)]
13. Tissot, A.; Danger, R.; Claustre, J.; Magnan, A.; Brouard, S. Early Identification of Chronic Lung Allograft Dysfunction: The Need of Biomarkers. *Front. Immunol.* **2019**, *10*, 1681. [[CrossRef](#)]
14. Gazourian, L.; Ash, S.; Meserve, E.E.K.; Diaz, A.; Estepar, R.S.J.; El-Chemaly, S.Y.; Rosas, I.O.; Divo, M.; Fuhlbrigge, A.L.; Camp, P.C.; et al. Quantitative Computed Tomography Assessment of Bronchiolitis Obliterans Syndrome after Lung Transplantation. *Clin. Transplant.* **2017**, *31*. [[CrossRef](#)] [[PubMed](#)]
15. Solyanik, O.; Hollmann, P.; Dettmer, S.; Kaireit, T.; Schaefer-Prokop, C.; Wacker, F.; Vogel-Claussen, J.; Shin, H. Quantification of Pathologic Air Trapping in Lung Transplant Patients Using CT Density Mapping: Comparison with Other CT Air Trapping Measures. *PLoS ONE* **2015**, *10*, e0139102. [[CrossRef](#)] [[PubMed](#)]
16. Barbosa, E.M.; Simpson, S.; Lee, J.C.; Tustison, N.; Gee, J.; Shou, H. Multivariate Modeling Using Quantitative CT Metrics May Improve Accuracy of Diagnosis of Bronchiolitis Obliterans Syndrome after Lung Transplantation. *Comput. Biol. Med.* **2017**, *89*, 275–281. [[CrossRef](#)] [[PubMed](#)]
17. Dettmer, S.; Suhling, H.; Klingenberg, I.; Otten, O.; Kaireit, T.; Fuge, J.; Kuhnigk, J.M.; Gottlieb, J.; Haverich, A.; Welte, T.; et al. Lobe-Wise Assessment of Lung Volume and Density Distribution in Lung Transplant Patients and Value for Early Detection of Bronchiolitis Obliterans Syndrome. *Eur. J. Radiol.* **2018**, *106*, 137–144. [[CrossRef](#)] [[PubMed](#)]
18. Horie, M.; Salazar, P.; Saito, T.; Binnie, M.; Brock, K.; Yasufuku, K.; Azad, S.; Keshavjee, S.; Martinu, T.; Paul, N. Quantitative Chest CT for Subtyping Chronic Lung Allograft Dysfunction and Its Association with Survival. *Clin. Transplant.* **2018**, *32*, e13233. [[CrossRef](#)] [[PubMed](#)]
19. Galbán, C.J.; Boes, J.L.; Bule, M.; Kitko, C.L.; Couriel, D.R.; Johnson, T.D.; Lama, V.; Telenga, E.D.; van den Berge, M.; Rehemtulla, A.; et al. Parametric Response Mapping as an Indicator of Bronchiolitis Obliterans Syndrome after Hematopoietic Stem Cell Transplantation. *Biol. Blood Marrow Transplant.* **2014**, *20*, 1592–1598. [[CrossRef](#)] [[PubMed](#)]
20. Saito, T.; Horie, M.; Sato, M.; Nakajima, D.; Shoushtarizadeh, H.; Binnie, M.; Azad, S.; Hwang, D.M.; Machuca, T.N.; Waddell, T.K.; et al. Low-Dose Computed Tomography Volumetry for Subtyping Chronic Lung Allograft Dysfunction. *J. Heart Lung Transplant.* **2016**, *35*, 59–66. [[CrossRef](#)]
21. Saito, M.; Chen-Yoshikawa, T.F.; Nakamoto, Y.; Kayawake, H.; Tokuno, J.; Ueda, S.; Yamagishi, H.; Gochi, F.; Okabe, R.; Takahagi, A.; et al. Unilateral Chronic Lung Allograft Dysfunction Assessed by Biphasic Computed Tomographic Volumetry in Bilateral Living-Donor Lobar Lung Transplantation. *Transplant. Direct* **2018**, *4*, e398. [[CrossRef](#)] [[PubMed](#)]
22. Weinheimer, O.; Achenbach, T.; Bletz, C.; Duber, C.; Kauczor, H.U.; Heussel, C.P. About Objective 3-d Analysis of Airway Geometry in Computerized Tomography. *IEEE Trans. Med. Imaging* **2008**, *27*, 64–74. [[CrossRef](#)]
23. Achenbach, T.; Weinheimer, O.; Biedermann, A.; Schmitt, S.; Freudenstein, D.; Goutham, E.; Kunz, R.P.; Buhl, R.; Dueber, C.; Heussel, C.P. MDCT Assessment of Airway Wall Thickness in COPD Patients Using a New Method: Correlations with Pulmonary Function Tests. *Eur. Radiol.* **2008**, *18*, 2731–2738. [[CrossRef](#)] [[PubMed](#)]
24. Achenbach, T.; Weinheimer, O.; Buschsieweke, C.; Heussel, C.P.; Thelen, M.; Kauczor, H.U. Fully automatic detection and quantification of emphysema on thin section MD-CT of the chest by a new and dedicated software. *Rofo* **2004**, *176*, 1409–1415. [[CrossRef](#)] [[PubMed](#)]
25. Horie, M.; Levy, L.; Houbois, C.; Salazar, P.; Saito, T.; Pakkal, M.; O'Brien, C.; Sajja, S.; Brock, K.; Yasufuku, K.; et al. Lung Density Analysis Using Quantitative Chest CT for Early Prediction of Chronic Lung Allograft Dysfunction. *Transplantation* **2019**, *103*, 2645–2653. [[CrossRef](#)] [[PubMed](#)]
26. Chassagnon, G.; Vakalopoulou, M.; Paragios, N.; Revel, M.-P. Artificial Intelligence Applications for Thoracic Imaging. *Eur. J. Radiol.* **2020**, *123*, 108774. [[CrossRef](#)] [[PubMed](#)]
27. Chassagnon, G.; Vakalopoulou, M.; Paragios, N.; Revel, M.-P. Deep Learning: Definition and Perspectives for Thoracic Imaging. *Eur. Radiol.* **2020**, *30*, 2021–2030. [[CrossRef](#)]
28. Chambers, D.C.; Cherikh, W.S.; Harhay, M.O.; Hayes, D.; Hsich, E.; Khush, K.K.; Meiser, B.; Potena, L.; Rossano, J.W.; Toll, A.E.; et al. The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: Thirty-Sixth Adult Lung and Heart-Lung Transplantation Report-2019; Focus Theme: Donor and Recipient Size Match. *J. Heart Lung Transplant.* **2019**, *38*, 1042–1055. [[CrossRef](#)]

4.3 APPLICATION DE LA RADIOMIQUE POUR LA PREDICTION DE LA DYSFONCTION CHRONIQUE DU GREFFON PULMONAIRE

La greffe pulmonaire est la greffe d'organe ayant le moins bon pronostic, principalement du fait des rejets aigus et chronique. Les progrès chirurgicaux récents et la mise au point de nouveaux immunosuppresseurs ont permis l'amélioration de la survie post-opératoire et la diminution des rejets aigus (28). Le rejet chronique demeure le principal défi qui limite la survie à long terme des patients transplantés pulmonaires (47). La prise en charge optimale du rejet chronique repose sur un diagnostic précoce qui permet d'adapter rapidement l'immunosuppression et ainsi d'augmenter la durée de vie du greffon. Il est ainsi important de pouvoir détecter le plus tôt possible les signes de dysfonction. Cependant le diagnostic de rejet chronique est cependant compliqué car l'altération fonctionnelle respiratoire est insuffisamment sensible et spécifique, en particulier pour le diagnostic précoce. Les recommandations actuelles de caractérisation en sous-types de la DCGP incluent le scanner thoracique comme un examen essentiel (33).

La deuxième partie de cette thèse rapporte nos expériences sur l'utilisation de la radiomique pour tenter de prédire la survenue d'une DCGP sur le scanner thoracique avant la survenue d'un déclin irréversible des EFRs,

4.3.1 METHODE

Nous avons formulé l'hypothèse que des méthodes de radiomique pourraient aider à prédire la survenue du DCGP à partir des images des scanners réalisées au moins 6 mois avant. Nous avons donc construit un modèle basé sur une approche radiomique en utilisant des scanners réalisés avant la survenue d'une DCGP. Les caractéristiques de radiomique sont des paramètres descripteurs de l'image qui peuvent être combinés au sein de modèles en utilisant des méthodes de machine learning. Ces caractéristiques incluent généralement des statistiques de premier ordre qui sont calculées sur l'histogramme de l'image (telle que la densité moyenne des voxels), des paramètres de texture extraits des matrices de co-occurrence et des variables de forme.

Pour utiliser des caractéristiques radiomiques calculées uniquement sur les poumons, une étape préliminaire de segmentation pulmonaire a été nécessaire. Cette étape consiste à identifier les voxels de l'image correspondant au tissu pulmonaire. C'est ensuite seulement sur ces voxels qu'ont été calculés les paramètres radiomiques.

Après extraction des paramètres, la première étape de l'analyse radiomique consistait à sélectionner les paramètres pertinents. En effet, il est nécessaire de diminuer le nombre de paramètres radiomiques car un nombre trop important de paramètres peut générer des problèmes de sous-apprentissage (difficulté des modèles à converger) ou être source de surapprentissage. De plus certains paramètres ne sont pas corrélés à la cible que l'on souhaite prédire ou apportent des informations redondantes.

La deuxième étape de l'analyse radiomique consistait à créer une signature radiomique combinant les paramètres pertinents préalablement sélectionnés. Pour cela, plusieurs classificateurs ont été utilisés et combinés. Les 5 classificateurs avec les meilleures exactitudes pondérées sur la validation ont été utilisés pour la prochaine étape de raffinement de la signature radiomique et la prédiction finale grâce à une méthode d'ensemble par vote majoritaire.

Ensuite, une troisième étape de raffinement de la signature radiomique a été réalisée pour éliminer les paramètres redondants. Nous avons itérativement enlevé une variable de la signature et réalisé la classification. La variable qui donnait la plus basse exactitude pondérée dans le set de validation était retirée. Ce procédé était itéré jusqu'à ce qu'il ne reste plus de variable. La signature finale a été obtenue en déterminant un point d'inflexion et conservant toutes les variables après ce point.

Enfin, nous réalisons la classification de nouveau avec les paramètres conservés et évaluons la performance de l'algorithme de prédiction finale par la méthode d'ensemble par vote majoritaire en jeu de test.

4.3.2 IMPLEMENTATION ET RESULTATS EXPERIMENTAUX

Base de données :

Ce travail rétrospectif a été conduit sur la cohorte de patients ayant reçu une transplantation pulmonaire à l'hôpital Européen Georges-Pompidou (HEGP) entre 2005 et 2018 puis suivis à l'hôpital de Cochin. L'étude a été approuvée par le comité d'éthique de la recherche AP-HP.5 (CERAPHP.5, référence n° 2019-06-07) qui a renoncé au consentement des patients.

Les patients ayant une DCGP diagnostiquée selon les critères de l'ISHLT (33) ont été inclus et des cas témoins de patients transplantés sans DCPG ont été sélectionnés par appariement sur l'âge, le sexe et le temps de suivi. En raison de l'indisponibilité des valeurs de Capacité Pulmonaire Totale (CPT) pour la plupart des patients, la classification du phénotype de DCGP était basée sur les valeurs de VEMS et CVF en spirométrie. Selon la même méthodologie qu'adoptée par Gauthier et Verleden (30,307), les patients ayant seulement une baisse du VEMS $\leq 80\%$ de la valeur initiale (après exclusion des autres causes identifiées) étaient classés dans le groupe DCGP obstructive et ceux ayant un déclin concomitant de la FVC $\leq 80\%$ de la valeur de base étaient classés dans le groupe de DCGP restrictive.

Les critères d'exclusions étaient le décès dans l'année suivant la transplantation pulmonaire, et l'absence de données cliniques, biologique ou TDM lors du suivi. Pour le groupe témoin, les patients ayant un suivi de moins de 1 an après le point temporel considéré (point temporel sans DCGP) ou pas de scanner à ce point temporel ont également été exclus afin de s'assurer de l'absence de DCGP. L'organigramme de l'étude est présentée en Figure 4-1: L'organigramme de la base de données pour l'étude de prédiction de DCGP par radiomique.. Au total 50 patients ayant un DCGP et 50 patients témoins ont été inclus.

Le VEMS de référence a été défini, selon les recommandations de l'« International Society for Heart and Lung Transplantation (ISHLT) », comme la moyenne des deux valeurs de VEMS les plus élevées obtenues au moins 3 semaines après la transplantation (33). La valeur de référence de la CVF était définie de manière similaire comme la moyenne des deux valeurs de CVF les plus élevées obtenues au moins 3 semaines après la transplantation. La chute persistante des paramètres d'EFRs était définie comme la première baisse sans récupération du VEMS et / ou de la CVFs à plus de 90% de la valeur de référence post-transplantation. Les critères d'exclusion pour définir le déclin des EFRs étaient : l'infection pulmonaire aiguë, le rejet aigu, les pathologies tumorales pulmonaires, le dysfonctionnement anastomotique bronchique, le pneumothorax et l'épanchement pleural supérieur à 30% du volume pulmonaire.

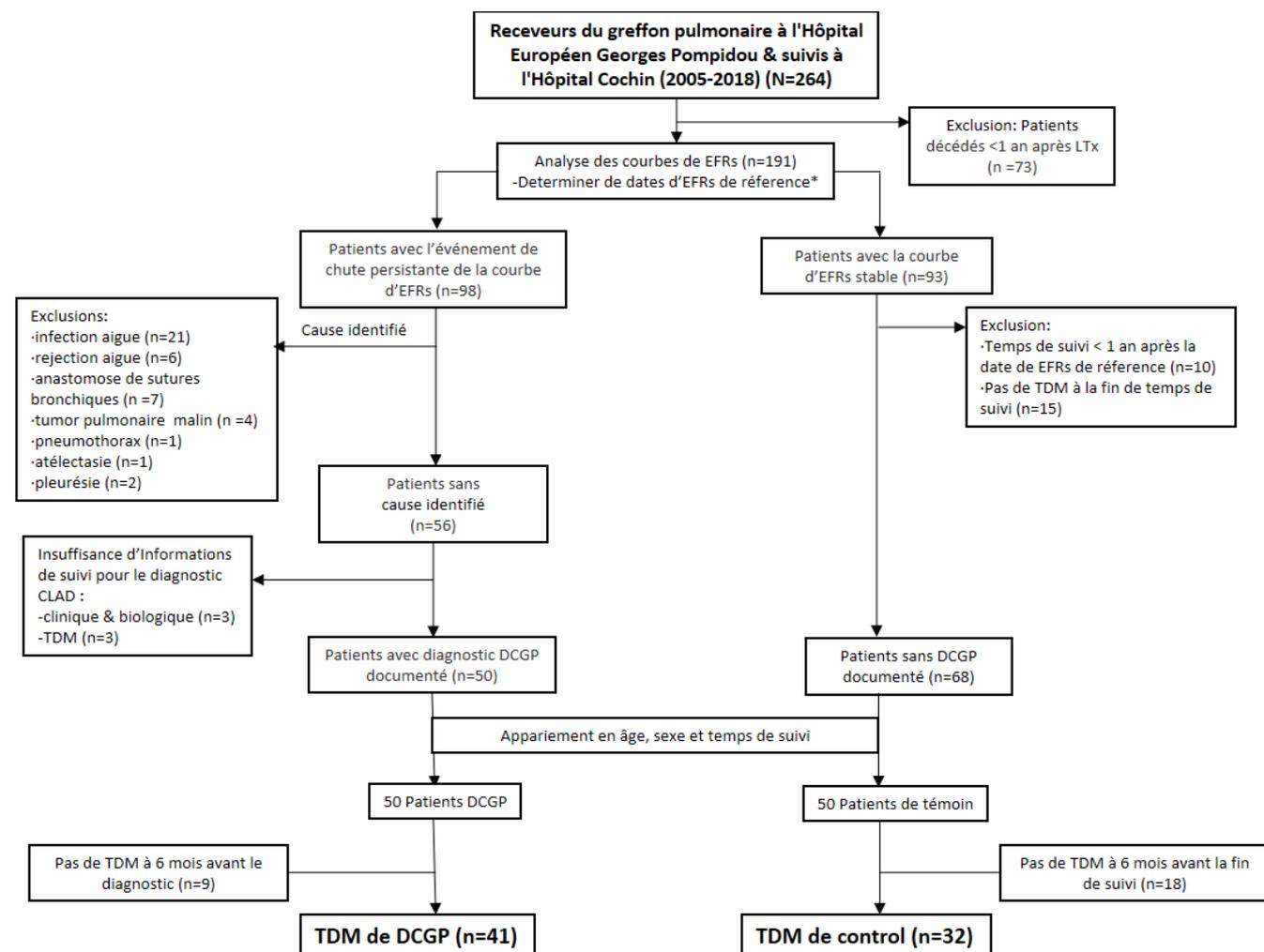


Figure 4-1: L'organigramme de la base de données pour l'étude de prédiction de DCGP par radiomique.

Les scanners thoraciques réalisés 6, 12 et 24 mois avant la survenue (patients DCGP) ou la non-survenue (patients témoins) de l'altération des EFRs ont été collectés s'ils remplissaient les critères suivants : (1) réalisé en décubitus dorsal, (2) sans injection de produit de contraste et (3) reconstruits avec une épaisseur de coupe de 0,625 à 1,25 mm et un filtre de reconstruction standard (I31f, mou ou médiastinal). Les examens ont été acquis sur 4 modèles différents de scanners (Somatom Sensation 16, Somatom DS et Somatom AS +; Siemens Healthineers, Erlangen, Allemagne; et Revolution HD; GE Healthcare, Milwaukee, Wis). Pour chaque examen, le produit dose-longueur (PDL) a été collecté afin d'évaluer la dose d'irradiation.

Les caractéristiques démographiques des patients et le nombre d'examen de la cohorte sont montrés dans le Tableau 4-2.

Tableau 4-2: Caractéristiques de DCGP et témoins (N=100).

	DCGP (n=50)	Témoins (n=50)	p
Homme : Femme	27 :23	27 :23	1
Age (an, ± Sd)	28 (±15)	27 (±16)	0.61
Temps de suivi (an)	4,43 (0,57-14,6)	4,19 (0,57-14,61)	0.72
TDM en inspiration (nombre d'examen)			

- Survenue de DCGP/fin du suivi	50	50	
- 6 mois avant	41	32	
- 12 mois avant	45	44	
- 24 mois avant	38	40	

Pour cette étude, nous avons utilisé les données d'une sous-population de patients pour qui un scanner était disponible 6 mois avant l'évènement « survenue » ou « non survenue » de DCGP. La base de données finale incluait donc 41 scanners des patients de DCGP (41/50) et 32 scanners des témoins sains (32/50) (Tableau 4-3).

Tableau 4-3: Les caractéristiques de la population de DCGP et de témoins (n=73).

		Tous	Témoins	DCGP	p
Nombre	Patients	73	32	41	
Age en année lors de l'opération de greffe	Median [10% quantile, 90% quantile] (année)	22 [13-59]	24.5 [13-58]	19 [13-58]	0.27
Sexe	Female	35	14	21	0.69
	Male	38	18	20	
IMC	Moyenne (±Et)	18.9 (±3.6)	20.7 (±3.7)	17.4 (±2.8)	<0.05
Type du greffon	Bi-pulmonaire	57	25	32	0.4
	Mono-pulmonaire	11	4	7	
	Combine avec greffon du Coeur/Foie	5	3	2	
Affection sous-jacente	Mucoviscidose	49	23	26	0.187
	Emphysème, BPCO	18	7	11	
	Autres (Lupus, LAM, HTAP...)	6	2	4	
Temps de suivi	Median [10% quantile, 90% quantile] (année)	2.15 [0.53-6.6]	2.06 [0.52-6.4]	2.64 [0.69-7.2]	0.45
Décalage TDM et la survenue de DCGP/ fin de suivi	Moyenne (±Et)	5.92 (±1.5)	5.97 (±1.2)	5.87 (±1.7)	0.78
EFRs de référence	VEMS (Litre, moyenne (±Et))	2.26 (±0.8)	2.37 (±0.8)	2,20 (±0.7)	0.29
	CVF(Litre, moyenne (±Et))	2.85 (±0.9)	3.05(±1)	2.67 (±0.8)	0.11
Grade de DCGP au début	0			17	
	1			9	
	2			8	
	3			7	
Sub-type de DCGP	Obstructive			24	
	Restrictive			17	

DCGP : Dysfonction Chronique du Greffon Pulmonaire ; IMC : Indice de Masse Corporelle ; TDM : Tomodensitométrie ; EFR : Exploration Fonctionnelle Respiratoire ; BPCO : Broncho-Pneumopathie Chronique Obstructive ; LAM : Lymphangioléiomyomatose; HTAP : hypertension artérielle pulmonaire Et : Ecart – type.

Pour tester des modèles, nous avons divisé la cohorte en jeu d'entraînement (80%, n=58) et de test (20%, n=15) par randomisation. Les caractéristiques de la cohorte constituant le jeu d'entraînement et de test sont décrites et Tableau 4-4.

Tableau 4-4: Les caractéristiques de la cohorte d'entraînement/validation et de test (n=73).

	Base de donnée d'entraînement (n=58)	Base de donnée de test (n=15)	p-value
Age	28.6 (±17.1)	25.7(±15.3)	0.52
Sexe	27/31	8/7	0.85
DCGP+	25/33	7/8	1
Kilovoltage (±Et)	106.7 (±8.2)	103.1(±9.6)	0.21
Epaisseur de coupe	0.735 (±0.28)	0.742 (±0.30)	0.93
Reconstruction kernel	49/53	13/15	1
Délai TDM par rapport de l'opération de greffe	3.16 (±2.7)	2.76(±1.97)	0.52

Décalage TDM et la survenue de DCGP/ fin de suivi (mois)	5.91(±1.4)	5.94(±1.8)	0.95
BMI	19.08 (±3.7)	18.17 (±2.9)	0.32

Segmentation pulmonaire et extraction des paramètres radiomique :

La segmentation pulmonaire a été réalisé pour l'ensemble des scanners après export, anonymisation et conversion des scanners en format *nifti*. La segmentation a été réalisée à l'aide d'un algorithme 3D de deep learning disponible en ligne (308). Toutes les segmentations ont été vérifiées et corrigées manuellement. Ces corrections ont été réalisées à l'aide du logiciel ITKsnap (<http://www.itksnap.org>).

L'extraction des paramètres radiomiques a été réalisée sur le logiciel Python à l'aide de la bibliothèque PyRadiomics (256). Les segmentations pulmonaires préalablement obtenues ont été utilisées pour limiter leur calcul des paramètres aux seules zones de poumon. L'extraction a été réalisée à partir des volumes 3D. Un total de 107 variables radiomiques a été extrait (Tableau 4-5) : 18 variables de premier ordre ; 73 caractéristiques de texture (24 variables calculées à partir de matrices de cooccurrence de niveaux de gris (GLCM), 14 variables calculées à partir de matrices de dépendance de niveau de gris (GLDM) ; 16 variables calculées à partir de matrices de longueur d'exécution de niveau de gris (GLRLM) ; 6 variables calculées à partir de matrices de différence de tons de gris voisins (NGTDM) et 16 variables calculées à partir de matrices de zone de taille de niveau de gris (GLSZM)) ; et 16 variables de forme.

Les résultats de la première étape la sélection des paramètres de radiomiques sont présentés en Tableau 4-5. Les 16 variables sélectionnées avec une prévalence totale $\geq 30\%$ ont été retenues. Il s'agissait de 3 variables de premier ordre (Robust Mean Absolute Deviation; Minimum; Interquartile Range; et 3 variables de forme (the Sphericity; Surface Volume Ratio; Maximum 2D Diameter Slice); 10 variables de *la matrice de taille des zones homogènes de niveaux de gris (grey-level size zone matrix, GLSZM tel que Zone Percentage, Low Gray Level Zone Emphasis Feature, Gray Level Variance ; Small Area Emphasis; Small Area High Gray Level Emphasis; Small Area Low Gray Level Emphasis; High Gray Level Zone Emphasis; Gray Level Non Uniformity; Gray Level Non Uniformity Normalized; Large Area High Gray Level Emphasis)*.

Tableau 4-5: Première étape de sélection des variables radiomiques.

Variable	Lasso	Arbre de décision	Line ar SVC	Gradi ent Boosting	AdaBoost	chi2	ANO VA F-value	Mutual Information	Total e	Corre latio n
original_glszm_ZonePercentage	3	2	3	3	4	2	2	0	19	0.13
original_shape_Sphericity	3	2	4	3	3	0	1	0	16	0.13
original_glszm_LowGrayLevelZoneEmphasis	3	1	5	0	0	4	3	0	16	-0.24
original_glszm_GrayLevelVariance	3	0	5	0	0	4	3	0	15	-0.24
original_glszm_SmallAreaEmphasis	3	2	4	4	2	0	0	0	15	0.02
original_shape_SurfaceVolumeRatio	3	1	3	4	2	1	0	0	14	-0.11
original_firstorder_RobustMeanAbsoluteDeviation	3	1	3	2	3	0	0	2	14	-0.10
original_glszm_SmallAreaHighGrayLevelEmphasis	3	1	4	3	1	1	0	0	13	0.07

original_glszm_SmallAreaLowGrayLevelEmphasis	3	1	4	2	1	1	1	0	13	-0.12
original_glszm_HighGrayLevelZoneEmphasis	3	1	5	1	0	0	3	0	13	0.24
original_glszm_GrayLevelNonUniformity	3	1	4	4	1	0	0	0	13	0.14
original_firstorder_Minimum	3	1	4	3	1	0	0	0	12	0.13
original_glszm_GrayLevelNonUniformityNormalized	3	0	5	0	1	0	3	0	12	0.24
original_shape_Maximum2DDiameterSlice	3	0	5	3	1	0	0	0	12	0.01
original_glszm_LargeAreaHighGrayLevelEmphasis	3	0	5	1	0	2	1	0	12	0.06
original_firstorder_InterquartileRange	3	1	3	4	1	0	0	0	12	-0.08
original_glszm_LargeAreaLowGrayLevelEmphasis	3	1	5	1	1	0	0	0	11	0.07
original_shape_Flatness	3	0	3	3	2	0	0	0	11	0.00
original_glszm_ZoneVariance	3	0	5	2	0	1	0	0	11	0.07
original_gldm_LargeDependenceHighGrayLevelEmphasis	3	1	3	3	0	1	0	0	11	-0.19
original_glszm_SizeZoneNonUniformity	3	0	3	0	2	1	1	0	10	0.14
original_glrlm_LongRunHighGrayLevelEmphasis	3	1	2	1	1	1	1	0	10	0.13
original_firstorder_10Percentile	3	1	3	2	1	0	0	0	10	-0.12
original_ngtdm_Strength	4	0	2	1	1	1	1	0	10	0.06
original_shape_Elongation	3	1	3	2	1	0	0	0	10	0.06
original_firstorder_Skewness	3	0	5	1	1	0	0	0	10	0.08
original_firstorder_RootMeanSquared	4	0	1	1	1	1	1	1	10	0.11
original_shape_LeastAxisLength	3	1	1	3	2	0	0	0	10	0.03
original_shape_Maximum2DDiameterRow	3	0	3	1	2	0	0	0	9	0.06
original_gldm_SmallDependenceHighGrayLevelEmphasis	3	0	5	1	0	0	0	0	9	0.06
original_glrlm_RunVariance	3	0	1	2	0	0	1	2	9	0.12
original_glrlm_ShortRunHighGrayLevelEmphasis	3	0	3	3	0	0	0	0	9	0.08
original_glrlm_GrayLevelNonUniformity	3	2	1	1	2	0	0	0	9	-0.02
original_firstorder_Maximum	3	0	3	2	1	0	0	0	9	0.02
original_glszm_LargeAreaEmphasis	3	0	5	1	0	0	0	0	9	0.07
original_shape_MajorAxisLength	3	0	3	2	1	0	0	0	9	0.03
original_firstorder_Energy	3	0	4	0	0	1	1	0	9	0.13
original_firstorder_TotalEnergy	3	0	4	0	0	1	1	0	9	0.13
original_glszm_SizeZoneNonUniformityNormalized	3	1	4	1	0	0	0	0	9	0.02
original_glrlm_HighGrayLevelRunEmphasis	3	0	2	1	0	0	1	1	8	-0.19
original_gldm_LargeDependenceEmphasis	3	0	0	1	1	0	0	3	8	0.06
original_glrlm_LowGrayLevelRunEmphasis	3	0	2	0	0	1	1	1	8	0.19
original_glrlm_ShortRunLowGrayLevelEmphasis	3	0	2	2	0	1	0	0	8	-0.10

original_ngtdm_Coarseness	3	0	2	1	0	1	1	0	8	0.07
original_glszm_ZoneEntropy	3	1	1	1	1	0	0	1	8	-0.05
original_gldm_SmallDependence Emphasis	3	0	5	0	0	0	0	0	8	0.05
original_firstorder_MeanAbsoluteDeviation	3	0	1	3	0	0	0	1	8	-0.12
original_glcm_Imc1	3	0	0	2	1	0	0	1	7	0.13
original_shape_MinorAxisLength	3	0	2	1	1	0	0	0	7	0.07
original_ngtdm_Busyness	3	0	2	1	1	0	0	0	7	-0.05
original_glcm_Imc2	3	0	1	2	0	0	0	1	7	-0.19
original_firstorder_Variance	3	1	2	0	1	0	0	0	7	-0.12
original_firstorder_Range	3	0	2	1	1	0	0	0	7	0.01
original_gldm_SmallDependence LowGrayLevelEmphasis	3	0	3	1	0	0	0	0	7	0.03
original_glrlm_RunLengthNonUniformity	3	0	2	1	1	0	0	0	7	-0.01
original_firstorder_Kurtosis	3	0	4	0	0	0	0	0	7	0.07
original_gldm_DependenceNonUniformityNormalized	3	0	1	2	0	0	0	1	7	0.08
original_shape_SurfaceArea	3	0	1	2	0	0	0	1	7	0.05
original_firstorder_Mean	3	0	1	1	0	0	0	1	6	-0.14
original_gldm_HighGrayLevelEmphasis	3	1	0	1	0	0	0	1	6	-0.13
original_firstorder_Median	3	0	2	0	0	0	0	1	6	-0.12
original_glcm_Contrast	3	0	1	2	0	0	0	0	6	-0.05
original_glcm_DifferenceAverage	3	1	1	1	0	0	0	0	6	-0.05
original_glrlm_LongRunLowGrayLevelEmphasis	3	0	1	1	0	1	0	0	6	0.13
original_glrlm_LongRunEmphasis	3	0	1	0	0	1	1	0	6	0.13
original_glrlm_GrayLevelNonUniformityNormalized	3	0	1	0	0	1	1	0	6	0.10
original_glrlm_GrayLevelVariance	3	0	1	0	1	0	1	0	6	-0.10
original_glrlm_RunPercentage	3	0	0	0	0	0	0	2	5	-0.05
original_shape_Maximum2DDiameterColumn	3	0	1	1	0	0	0	0	5	0.05
original_glcm_ClusterShade	3	0	1	1	0	0	0	0	5	-0.17
original_glcm_SumEntropy	3	0	0	0	0	0	0	2	5	-0.13
original_glcm_ClusterProminence	3	0	0	1	1	0	0	0	5	-0.17
original_ngtdm_Complexity	3	0	1	1	0	0	0	0	5	-0.05
original_glcm_Idmn	3	0	1	1	0	0	0	0	5	0.05
original_gldm_GrayLevelVariance	3	0	0	0	1	0	0	1	5	-0.14
original_glrlm_ShortRunEmphasis	3	0	0	1	1	0	0	0	5	-0.02
original_glrlm_RunEntropy	3	0	0	1	1	0	0	0	5	0.05
original_firstorder_Uniformity	3	0	0	1	0	0	0	1	5	0.14
original_shape_Maximum3DDiameter	3	0	1	0	1	0	0	0	5	0.06
original_gldm_LowGrayLevelEmphasis	3	0	0	1	0	0	0	1	5	0.13
original_gldm_DependenceEntropy	3	0	0	1	0	0	0	1	5	-0.09
original_glcm_Autocorrelation	3	0	0	0	0	1	0	0	4	-0.14
original_ngtdm_Contrast	3	0	1	0	0	0	0	0	4	-0.06

original_firstorder_Entropy	3	0	0	0	0	0	0	1	4	-0.14
original_shape_VoxelVolume	3	0	0	1	0	0	0	0	4	0.07
original_gldm_Correlation	3	0	1	0	0	0	0	0	4	-0.17
original_gldm_InverseVariance	3	0	1	0	0	0	0	0	4	-0.05
original_firstorder_90Percentile	3	0	0	1	0	0	0	0	4	-0.15
original_gldm_DependenceNonUniformity	3	0	0	0	1	0	0	0	4	0.09
original_gldm_SumSquares	3	0	0	1	0	0	0	0	4	-0.14
original_gldm_RunLengthNonUniformityNormalized	3	0	0	1	0	0	0	0	4	-0.04
original_gldm_JointEnergy	3	0	0	0	0	0	0	1	4	0.12
original_gldm_MCC	3	0	1	0	0	0	0	0	4	-0.17
original_gldm_Idn	3	0	1	0	0	0	0	0	4	0.05
original_gldm_Idm	3	0	1	0	0	0	0	0	4	0.05
original_gldm_Id	3	0	1	0	0	0	0	0	4	0.05
original_gldm_JointAverage	3	0	0	0	0	0	0	0	3	-0.13
original_gldm_JointEntropy	3	0	0	0	0	0	0	0	3	-0.12
original_gldm_ClusterTendency	3	0	0	0	0	0	0	0	3	-0.15
original_gldm_MaximumProbability	3	0	0	0	0	0	0	0	3	0.11
original_gldm_SumAverage	3	0	0	0	0	0	0	0	3	-0.13
original_shape_MeshVolume	3	0	0	0	0	0	0	0	3	0.07
original_gldm_DifferenceVariance	3	0	0	0	0	0	0	0	3	-0.06
original_gldm_DifferenceEntropy	3	0	0	0	0	0	0	0	3	-0.07
original_gldm_LargeDependenceLowGrayLevelEmphasis	3	0	0	0	0	0	0	0	3	0.10
original_gldm_DependenceVariance	3	0	0	0	0	0	0	0	3	-0.08
original_gldm_GrayLevelNonUniformity	3	0	0	0	0	0	0	0	3	0.09

Création et évaluation de la signature radiomique :

La première étape de la création du modèle consistait en la sélection des paramètres radiomiques pertinents. Pour cela, la base d'entraînement a été divisée en 5 sous-ensembles de même taille. Un total de 8 algorithmes de machine learning ont été utilisés, 5 algorithmes de classification (Lasso, Arbre de décision, Linear Support Vector Classification (SVC), Gradient Boosting et Adaboost) et 3 méthodes basées sur des métriques statistiques (Chi2, ANOVA F-value et la mutual information). Pour les algorithmes de classification, toutes les variables auxquelles l'algorithme a attribué un poids, représentant la pertinence de cette variable pour cette méthode, supérieur à la moyenne étaient retenues. Pour les algorithmes basés sur des métriques statistiques, les 5% de variables présentant la corrélation la plus élevée étaient conservées. Les 8 algorithmes de machine learning ont été appliqués sur chaque combinaison possible des 5 sous-ensembles de données. A chaque entraînement, l'algorithme était entraîné sur 4 sous-ensembles de données et évalué sur le 5ème. La prévalence d'une variable est définie comme le nombre de fois qu'elle est sélectionnée sur l'ensemble des entraînements. La sélection de variables était réalisée en déterminant un seuil sur la prévalence. Plusieurs seuils ont été comparés et le seuil permettant d'obtenir les meilleurs résultats était une prévalence de 30% de soit au moins 12 sélections sur les 40 entraînements (8 techniques * 5 sous-ensembles). Ainsi 16 paramètres sélectionnés avec une prévalence $\geq 12/40$ (30%) ont été retenus à la fin de cette étape (Tableau 4-5)

Afin d'éviter un surapprentissage sur des ensembles de données déséquilibrés, nous avons utilisé l'exactitude pondérée comme métrique. L'exactitude pondérée correspond à la macro-moyenne des scores de rappel par classe ou, de manière équivalente, à la précision brute où chaque échantillon est pondéré en fonction de la prévalence inverse de sa vraie classe. Si y_i est la vraie valeur du i -ième échantillon et w_i est le poids de l'échantillon correspondant, alors nous ajustons le poids de l'échantillon à :

$$\hat{w}_i = \frac{w_i}{\sum_j 1(y_j = y_i)w_j}$$

Où $1(x)$ est la fonction indicatrice. Étant donnée la prédiction \hat{y}_i pour l'échantillon i , l'exactitude pondérée est définie comme (309–312) :

$$exactitude - pondérée(y, \hat{y}, y_i) = \frac{1}{\sum \hat{w}_i} \sum_i 1(\hat{y}_i = y_i) \hat{w}_i$$

Pour l'étape de sélection des méthodes de classification, nous avons évalués 15 méthodes différentes de machine learning : Nearest Neighbor, Linear, Sigmoid, Radial Basis Function (RBF), Polynomial Kernel, Support Vecteur Machines (SVM), Gaussian Process, Décision Trees, Random Forests, Ada Boost, XGBoosting, Gaussian Naïve bayes, Bernoulli Naïve bayes, Multilayer Perceptron (MLP) & Quadratic Discriminant Analysis (QDA). Les résultats de la sélection des méthodes de classification sont présentés dans le Tableau 4-6. Nous avons observé un problème de sur-apprentissage car les performances obtenues lors de l'entraînement étaient souvent supérieures à celles obtenues sur le set de validation (Tableau 4-6). Les 5 méthodes donnant les meilleures exactitudes pondérées ont été retenues. Ces 5 méthodes (Neural Net, Gaussian Naïve Bayes, Bernouill Naïve Bayes, QDA et Gradient Boosting) ont combinées grâce à une méthode d'ensemble par vote majoritaire. L'exactitude pondérée était alors de 76% (± 4) lors de l'entraînement mais de seulement 60% (± 9) lors de la validation (Tableau 4-6).

Tableau 4-6: Deuxième étape : la sélection des méthodes de classification.

Classificateur	Exactitude Pondérée		Précision		Sensitivité		Spécificité	
	Entrainement	Validation	Entrainement	Validation	Entrainement	Validation	Entrainement	Validation
Nearest Neighbors	54%+/-4	4%+/-12	59%+/-3	51%+/-12	83%+/-9	60%+/-25	25%+/-15	20%+/-16
Linear SVM	59%+/-5	52%+/-9	61%+/-3	61%+/-5	98%+/-3	91%+/-13	20%+/-8	13%+/-12
poly SVM	59%+/-5	56%+/-13	62%+/-2	64%+/-10	77%+/-19	76%+/-19	41%+/-10	37%+/-19
sigmoid SVM	50%+/-0.0	5%+/-0.0	56%+/-0.0	60%+/-0.0	100%+/-0.0	100%+/-0.0	0.0+/-0.0	0.0+/-0.0
RBF SVM	68%+/-5	48%+/-7	72%+/-6	53%+/-11	74%+/-11	47%+/-33	61%+/-15	50%+/-24
Gaussian Process	86%+/-16	48%+/-4	85%+/-15	44%+/-25	97%+/-5	49%+/-36	75%+/-28	47%+/-31
Decision Tree	88%+/-6	49%+/-1	94%+/-8	58%+/-9	84%+/-11	44%+/-21	92%+/-12	53%+/-24
Random Forest	100%+/-1	42%+/-8	100%+/-0.0	51%+/-10	99%+/-2	47%+/-19	100%+/-0.0	37%+/-19
Neural Net	81%+/-11	58%+/-12	81%+/-11	64%+/-10	90%+/-6	69%+/-26	72%+/-18	47%+/-16
AdaBoost	100%+/-0.0	48%+/-4	100%+/-0.0	59%+/-3	100%+/-0.0	53%+/-8	100%+/-0.0	43%+/-13
Gaussian Naive Bayes	59%+/-3	62%+/-14	61%+/-2	67%+/-9	94%+/-2	87%+/-16	23%+/-8	37%+/-19
Bernouill Naive Bayes	57%+/-1	57%+/-8	60%+/-1	64%+/-5	100%+/-0.0	100%+/-0.0	15%+/-2	13%+/-16
QDA	64%+/-2	60%+/-10	64%+/-1	66%+/-6	97%+/-3	87%+/-13	31%+/-5	33%+/-15

Gradient Boosting	73%+/-7	58%+/-7	71%+/-6	65%+/-4	98%+/-2	80%+/-16	48%+/-13	37%+/-12
Top 5 des classificateurs de vote	76%+/-4	60%+/-9	74%+/-3	67%+/-7	92%+/-6	80%+/-15	59%+/-5	40%+/-23

Pour la troisième étape, de raffinement de la signature radiomique, nous avons utilisé une méthode d'ablation (Tableau 4-7). Ainsi, nous avons itérativement enlevé une variable de la signature radiomique et réalisée la prédiction sur la cohorte de validation. La variable qui donnait l'exactitude pondérée la plus basse était retirée. Ce procédé était répété jusqu'à ce qu'il ne reste plus de variable. La première colonne du

Tableau 4-7: Troisième étape de sélection des variables : classification par ablation itérative des variables représente la variable qui a été enlevée pour obtenir le résultat de la troisième colonne. La meilleure exactitude pondérée (60%) a été obtenue après l'ablation de la variance des niveaux de gris. Après ce point l'ablation de paramètres supplémentaires n'augmentait plus l'exactitude de l'algorithme. Ainsi 3 variables ont été conservées et sont montrées en gras dans le tableau : un paramètre de statistique de premier ordre, la valeur minimum d'histogramme, et les deux paramètres de texture, l'accentuation de petite zone à niveau de gris élevé et le non-uniformité des niveaux de gris normalisés.

Tableau 4-7: Troisième étape de sélection des variables : classification par ablation itérative des variables.

<i>Variable retiré</i>	<i>Nombre de variable</i>	<i>Exactitude pondérée de validation</i>
<i>Tous les 16 variables</i>	16	48%
<i>original_glszm_SmallAreaLowGrayLevelEmphasis</i>	15	49%
<i>original_glszm_ZonePercentage</i>	14	51%
<i>original_glszm_GrayLevelNonUniformity</i>	13	53%
<i>original_firstorder_InterquartileRange</i>	12	54%
<i>original_shape_Maximum2DDiameterSlice</i>	11	52%
<i>original_glszm_LargeAreaHighGrayLevelEmphasis</i>	10	54%
<i>original_glszm_HighGrayLevelZoneEmphasis</i>	9	54%
<i>original_firstorder_RobustMeanAbsoluteDeviation</i>	8	54%
<i>original_glszm_SmallAreaEmphasis</i>	7	58%
<i>original_shape_SurfaceVolumeRatio</i>	6	58%
<i>original_shape_Sphericity</i>	5	59%
<i>original_glszm_LowGrayLevelZoneEmphasis</i>	4	58%
<i>original_glszm_GrayLevelVariance</i>	3	60%
<i>original_firstorder_Minimum</i>	2	53%
<i>original_glszm_SmallAreaHighGrayLevelEmphasis</i>	1	50%
<i>original_glszm_GrayLevelNonUniformityNormalized</i>	0	45%

Les performances de la signature radiomique finale sont présentés dans le Tableau 4-8. L'exactitude pondérée de la signature radiomique pour la prédiction de la survenue d'une DCGP à 6 mois dans la cohorte de test était de 79% avec une sensibilité de 88% et une spécificité de 71%. Ces résultats étaient supérieurs à ceux observés dans la cohorte de validation où l'exactitude pondérée était de 76%, la sensibilité de 100% et la spécificité de 52%. A noter que les performances

individuelles de chacun des algorithmes étaient proches de l'aléatoire et deux d'entre eux (Gaussian Naive Bayes et Bernoulli Naive Bayes) avaient une spécificité de 0.

Tableau 4-8: Performance des modèles en group del'entraînement et en test.

Classificateur	Exactitude pondérée		Précision		Sensitivité		Spécificité	
	Entrainement	Test	Entrainement	Test	Entrainement	Test	Entrainement	Test
Neural Net	72%	58%	72%	58%	88%	88%	56%	29%
Gaussian Naive Bayes	50%	50%	57%	53%	100%	100%	0	0
Bernoulli Naive Bayes	50%	50%	57%	53%	100%	100%	0	0
QDA	64%	51%	65%	54%	100%	88%	28%	14%
Gradient Boosting	64%	51%	65%	54%	100%	88%	28%	14%
Top 5 des classificateurs de vote	76%	79%	73%	78%	100%	88%	52%	71%

4.4 DISCUSSION

À notre connaissance, cette étude est la première à avoir évalué l'utilisation d'une approche radiomique pour la prédiction de la survenue de la DCGP chez les patients transplantés pulmonaires et nos résultats montrent une exactitude pondérée de 79% pour la prédiction à 6 mois.

La principale manifestation scanographique de la DCGP est l'apparition d'un piégeage expiratoire traduisant une bronchiolite oblitérante post-transplantation. Plusieurs méthodes quantitatives ont été proposées pour l'évaluation du piégeage expiratoire dans la broncho-pneumopathie chronique obstructive (BPCO) puis transposées à l'étude de la DCGP. Ces méthodes sont basées sur l'étude des densités pulmonaires en inspiration et en expiration. Une première approche consiste à étudier la densité moyenne pulmonaire (mean lung density = MLD) en inspiration et en expiration et de calculer le rapport entre les 2, appelé E/I MLD. Dans une analyse de 122 sujets atteints de BPCO, E/I MLD avait des associations indépendantes avec le rapport volume résiduel sur la capacité pulmonaire totale (VR/CPT) ($\beta = 0,30$, $p = 0,003$), le syndrome obstructif, dans un modèle de régression combiné %LAA $_{-950}$, E/I MLD, Pi10, bronchectasie, VEMS % et les variables démographiques (âge, sexe, statut tabagique et l'IMC) (313). Dans le tableau de DCGP Dettmer (314), Saito (302) et coll ont étudié le rapport E/I MLD et le changement de volume pulmonaire entre l'inspiration et l'expiration. Dettmer et al. ont montré une augmentation significative d' E/I-Volumes et diminution d' E/I-MLD chez les patients présentant une obstruction bronchique avec des changements plus importants dans les lobes inférieurs(314). Le groupe de Saito rapporte que la différence de volume pulmonaire entre l'inspiration et l'expiration a significativement diminué dans le groupe CLAD au début de la maladie par rapport à la valeur de base (le seuil de 0,8 avait une AUC de 0,87) (315). Cette approche reste néanmoins globale et ne tient pas compte des modifications locales au sein du parenchyme pulmonaire. L'évaluation des variations de densité a nécessité un recalage, c'est-à-dire une mise en correspondance des voxels, des images acquises en inspiration et expiration. Une telle méthode a été proposée par Galban et coll. (83) pour l'évaluation de l'obstruction des petites voies aériennes dans la BPCO. L'image paramétrique obtenue est basée sur l'étude de la densité de chaque voxel poumon en inspiration et en expiration. Ainsi en utilisant les seuils de densité classique pour l'emphysème (-910 UH en inspiration) et le piégeage (-856 UH en expiration), il a été possible de générer une cartographie paramétrique où chaque voxel pulmonaire

est classé comme étant du poumon sain, de l'emphysème ou du piégeage. Solyanik et coll. ont évalué cette méthode chez les patients greffés pulmonaires et ont montré que les seuils de densité pour définir l'emphysème et le piégeage devaient être adaptés à -790 HU à -950 HU chez les greffés pulmonaires. Ils ont cependant trouvé que la quantification du piégeage réalisé de cette manière était plus faiblement corrélée à l'obstruction fonctionnelle (mesure du rapport VR/CPTen pléthysmographie) que ne l'étaient la densité pulmonaire moyenne et E/I MLD (316). La variabilité interindividuelle du niveau d'inspiration et d'expiration est une source importante de fluctuation des mesures lors de l'utilisation des seuils fixes de densité pulmonaire mais Solyanik et coll. avaient réalisé les acquisitions scanographiques avec synchronisation respiratoire à la spirométrie pour s'affranchir de cette problématique.

L'approche radiomique a l'avantage de permettre à la fois une étude globale et locale des poumons. Les caractéristiques de premier ordre calculés sur l'histogramme de densité de l'ensemble des poumons (tels que la densité pulmonaire moyenne) permettent une évaluation globale. Les caractéristiques de texture extraites des matrices de co-occurrence et les caractéristiques de forme permettent quant à elles une étude des changements locaux. L'utilisation de la radiomique n'est pas nouvelle en imagerie scanographique pulmonaire. Elle a déjà été évaluée pour prédire la malignité des nodules pulmonaires (229,240,258) ou la sévérité de maladies pulmonaires diffuses comme la mucoviscidose (317). Une originalité de notre travail est que nous avons essayé de prédire la survenue d'une DPCG 6 mois avant son diagnostic. La signature radiomique finale obtenue était basée sur 3 paramètres incluant une caractéristique de premier ordre et deux caractéristiques de texture. Les caractéristiques sélectionnées montrent l'importance de combiner des informations globales et des informations de textures reflétant l'hétérogénéité locale du poumon. De façon intéressante, la densité pulmonaire moyenne n'a pas été sélectionnée dans le modèle final. Nous avons obtenu une exactitude pondérée de 79%, correspondant à une sensibilité de 88% et une spécificité de 71%. Ceci est très encourageant car que notre méthode n'inclut pas les séries en expiration et est uniquement basée sur des scanners réalisés avant la survenue de la DCGP. Cela montre que le diagnostic de la DCGP pourrait être anticipé et que les séries additionnelles en expiration pourrait éventuellement être évitées ce qui permettrait de réduire la dose d'irradiation. Les images en expiration sont réalisées pour mettre en évidence le piégeage expiratoire mais celui-ci peut souvent être mis en évidence de façon indirecte sur la série en inspiration sous la forme d'une perfusion en mosaïque.

Ce travail présente plusieurs limites. La première est liée au caractère monocentrique de l'étude et à la petite taille de notre population. Bien que l'exactitude pondérée de la signature radiomique proposée soit de 79% dans la population de test on a observé des performances discordantes entre la validation et le test. Cet écart de performance montre une certaine instabilité du modèle. Cette instabilité résulte d'un sur-apprentissage probablement lié à trop petite taille de la cohorte d'apprentissage. Cette instabilité du modèle montre la nécessité d'une évaluation de sa généralisabilité sur une cohorte externe avant de pouvoir conclure sur ses performances réelles. Elle montre aussi la nécessité d'enrichir la cohorte d'apprentissage. La seconde limite de cette étude est liée au caractère rétrospectif et à la longue durée de l'étude qui expliquent l'absence de standardisation des acquisitions scanographiques. Ainsi, pas tous les patients n'avaient un scanner disponible 6 mois avant la survenue de la DCGP et les paramètres d'acquisitions étaient variables. Or plusieurs études ont montré l'importance de la standardisation des paramètres d'acquisition et de reconstruction des images pour la création de signatures radiomiques. De même, devant la proportion disponible insuffisante de scanners volumiques en expiration de l'ensemble des poumons, nous n'avons pas intégré les images en expiration dans la signature radiomique. Les images en

expiration sont pourtant un élément important de l'évaluation scanographique de la DCGP et auraient peut-être pu permettre d'améliorer les performances du modèle.

4.5 CONCLUSION

En conclusion, ce travail préliminaire montre qu'une approche radiomique à partir des scanners de suivi semble prometteuse pour anticiper la survenue d'une DCGP. Cependant, la mise au point d'un modèle fiable et suffisamment performant requiert une augmentation de la taille de la base de données d'apprentissage en passant par des collaborations multicentriques.

5 CONTRIBUTION DE L'ÉVALUATION AUTOMATIQUE DE L'ÉTENDUE DE LA PNEUMOPATHIE INTERSTITIELLE DIFFUSE (PID) EN TOMODENSITOMETRIE EN PREDICTION DE LA MORTALITE DANS LA SCLERODERMIE SYSTEMIQUE

5.1 CONTEXTE ET OBJECTIFS

La pneumopathie interstitielle diffuse (PID) est identifiée chez 32 à 52 % des patients atteints de sclérodermie systémique (ScS) et représente une cause majeure de morbidité et de mortalité chez ces patients (318,319). L'identification d'indicateurs de mauvais pronostic dans la SSc-ILD est un défi pour les cliniciens. De tels facteurs doivent être systématiquement recherchés lors de l'examen du traitement et de la surveillance des patients atteints de sclérodermie (320). Plusieurs scores ont été proposés pour prédire la mortalité dans la ScS (321–326). Dans la grande méta-analyse de Rubio-Rivas et coll., les principaux facteurs de risque de mauvais pronostic étaient la maladie pulmonaire interstitielle, l'hypertension pulmonaire, la sclérodermie cutanée diffuse, l'âge avancé, l'atteinte rénale, l'atteinte cardiaque, le sexe masculin et une vitesse de sédimentation élevée (327). La fibrose pulmonaire est une cause majeure de décès chez les patients atteints de ScS (326). L'intérêt pronostique de la mesure de l'étendue de la PID au scanner a déjà été prouvé par Goh et collaborateurs (328). Ces auteurs furent les premiers à évaluer l'apport combiné de l'imagerie et des explorations fonctionnelles respiratoires dans l'estimation pronostique, notamment la survie, de 215 patients porteurs de ScS. Ils ont ainsi construit un algorithme basé sur la sévérité de l'atteinte fonctionnelle respiratoire (avec des valeurs de la capacité vitale forcée < 70% de la théorique) et l'extension des signes de PID au scanner avec un seuil optimal à 20% (328). L'évaluation de la fibrose pulmonaire avec des méthodes automatisées assistées par ordinateur pourrait aider à stratifier le pronostic et le risque de décès des patients (113,239,329–331). Nos récents travaux ont rapporté une méthode basée sur le deep-learning pour évaluer l'étendue de la PID chez les patients atteints de ScS avec des résultats similaires à ceux des radiologues. L'avantage est que notre méthode permet une analyse quantitative de toutes les images CT, ce qui contraste avec la notation visuelle. L'étendue de la PID fournie par notre modèle avec l'architecture AtlasNet était bien corrélée avec les résultats des EFR, en particulier la capacité pulmonaire de diffusion du monoxyde de carbone (239). Pour ces raisons, nous implémentons cette étude rétrospective monocentrique pour déterminer si la nouvelle quantification automatisée de la fibrose pulmonaire était corrélée à la mortalité à court et à long terme dans la sclérose systémique.

5.2 PUBLICATION

Prognostic Interest of Automated Assessment of Interstitial Lung Disease on CT in Systemic Sclerosis

Aëlle Le Gall MD, Trieu-Nghi Hoang-Thi MD, Raphaël Porcher MD PhD, Bertrand Dunogué MD MSc, Berezné Alice MD, Loïc Guillevin MD PhD, Véronique Le Guern MD PhD, Pascal Cohen MD PhD, Benjamin Chaigne MD, Jonathan London MD, Matthieu Groh MD, Romain

Paule MD, Guillaume Chassagnon MD PhD, Maria Vakalopoulou PhD, Anh-Tuan Dinh-Xuan MD PhD, Marie Pierre Revel MD PhD, Luc Mouthon MD PhD, Alexis Régent MD PhD.

5.2.1 RESUME

Rationel et objectifs : La stratification du risque de décès dans la maladie pulmonaire interstitielle liée à la sclérodémie (ScS-ILD) est une question difficile. L'étendue de la fibrose pulmonaire sur la tomodynamométrie (TDM) est souvent évaluée par une méthode visuelle semi-quantitative qui manque de fiabilité. L'objectif de notre travail est de démontrer l'intérêt pronostique d'un nouvel algorithme basé sur le deep-learning, permettant une quantification automatisée et objective de la fibrose pulmonaire sur TDM dans la ScS.

Méthodes : Nous avons effectué une analyse de cohorte rétrospective de 318 patients atteints de ScS d'un centre de référence tertiaire. Une quantification automatisée de l'étendue de l'ILD sur TDM a été réalisée pour chaque patient à l'inclusion. Nous avons comparé deux modèles pronostiques, le premier incluant les prédicteurs de risque traditionnels dans la ScS (sexe, âge, type de ScS, profil d'auto-anticorps et score de dyspnée élevé) et le second ajoutant la mesure de l'étendue de l'ILD sur TDM.

Mesures/Principales résultats : le taux de mortalité était de 1,6 % (IC à 95 % 0,2-2,9) à 2 ans et de 26,3 % (IC à 95 % 19,3 à 32,6) à 10 ans. Jusqu'à 30 % du volume pulmonaire total, chaque élévation de 1 % de l'étendue de l'ILD était associée à une augmentation de 4 % du risque de décès (HR 1,04 IC à 95 % 1,01-1,07, p = 0,004). En plus des prédicteurs de mauvais pronostic bien connus dans la ScS, la quantification automatisée de la fibrose pulmonaire sur TDM a significativement amélioré la capacité à prédire la mortalité à court terme (indice de discrimination intégré à 2 ans 0,014, IC à 95 % 0,002-0,066, p=0,020, Différence dans l'aire sous la courbe des deux modèles 0,043, IC 95% 0,002-0,084, p=0,040).

Conclusion : la quantification assistée par ordinateur basée sur l'apprentissage en profondeur de l'étendue de l'ILD sur TDM fournit un outil complémentaire efficace pour la stratification du risque dans la ScS.

5.2.2 MANUSCRIT D'ARTICLE

Prognostic Interest of Automated Assessment of Interstitial Lung Disease on CT in Systemic Sclerosis

Aëlle Le Gall⁽¹⁾ MD, Trieu-Nghi Hoang-Thi⁽²⁾ MD, Raphaël Porcher⁽³⁾⁽⁴⁾ MD PhD,
Bertrand Dunogué⁽¹⁾ MD MSc, Berezné Alice⁽¹⁾ MD, Loïc Guillevin⁽¹⁾⁽³⁾ MD PhD,
Véronique Le Guern⁽¹⁾ MD PhD, Pascal Cohen⁽¹⁾ MD PhD, Benjamin Chaigne^{(1) (3)}
MD, Jonathan London⁽¹⁾ MD, Matthieu Groh⁽¹⁾ MD, Romain Paule⁽¹⁾ MD, Guillaume
Chassagnon⁽²⁾⁽³⁾ MD PhD, Maria Vakalopoulou⁽⁵⁾ PhD, Anh-Tuan Dinh-Xuan⁽⁶⁾ MD
PhD, Marie Pierre Revel⁽²⁾⁽³⁾ MD PhD, Luc Mouthon^{(1) (3)} MD PhD, Alexis Régent^{(1) (3)}
MD PhD

⁽¹⁾ Service de médecine interne, Centre de référence maladies auto-immunes et systémiques rares d'île de France, Hôpital Cochin, Assistance Publique-Hôpitaux de Paris, Paris, France

⁽²⁾ Service de radiologie, Hôpital Cochin, AP-HP, 27 rue du Faubourg Saint Jacques, 75014 Paris, France

⁽³⁾ APHP-CUP, Hôpital Cochin, F-75014 Paris, Université de Paris, 15 rue de l'école de médecine, 75006 Paris

⁽⁴⁾ Service d'épidémiologie clinique, Hôpital Hôtel Dieu, AP-HP, 1 place du Parvis de Notre Dame, Paris, France

⁽⁵⁾ Centre de Vision Numérique, Ecole CentraleSupélec, Gif-sur-Yvette, France

⁽⁶⁾ Service de Physiologie et Explorations fonctionnelles, Hôpital Cochin, AP-HP, Paris, France

Correspondance : Dr Alexis Régent, Service de médecine interne, Hôpital Cochin, 27 rue du Faubourg Saint Jacques, Paris, France. Téléphone + 33(0)158411455. Fax + 33(0)158411579

Mail: alexis.regent@aphp.fr

Each author's contribution to the study:

Conception and design: AR, LM, MPR, GC, ALG

Acquisition, analysis and interpretation: AR, RP, MPR, GC, TNHT, ATDX, ALG

Drafting the manuscript for important intellectual content: all authors

Final approval of the version to be published: all authors

Funding: this study received no financial support

Conflict of interest: none to declare

Word count (main text): 2463 words

Key words: Systemic sclerosis, Interstitial lung disease, Prognosis, High resolution computed tomography, Artificial intelligence

Abbreviations: AHSCT = Autologous hematopoietic stem cells transplantation, DLCO = Diffusing capacity of carbon monoxide, FEV = Forced expiratory volume, FVC = Forced vital capacity, GER = gastro-esophageal reflux, HR = Hazard ratio, HRCT = High resolution computed tomography, IC = Confidence interval, IDI = Integrated discrimination index, ILD = Interstitial lung disease, IQR = Interquartile range, KCO = Diffusion coefficient, NYHA = New York Heart Association, PAH = pulmonary arterial hypertension, PFTs = Pulmonary function tests, ROC = receiver operating characteristic, SSc = Systemic sclerosis, SSc-ILD = Scleroderma-related interstitial lung disease, TLC = Total lung capacity

Abstract:

Rationale: Stratifying the risk of death in scleroderma-related interstitial lung disease (SSc-ILD) is a challenging issue. The extent of lung fibrosis on high-resolution computed tomography (HRCT) is often assessed by a visual semi-quantitative method which lacks reliability.

Objectives: To demonstrate the prognostic interest of a novel deep-learning based algorithm, allowing an automated and objective quantification of lung fibrosis on HRCT in SSc.

Methods: We performed a retrospective cohort analysis of 318 SSc patients from a tertiary referral center. Automated quantification of ILD extent on HRCT was performed for each patient at inclusion. We compared two prognostic models, the first one including traditional risk predictors in SSc (gender, age, type of SSc, autoantibody profile and elevated dyspnea score) and the second one adding the measurement of ILD extent on HRCT.

Measurements/ Main results: The mortality rate was 1.6% (95%CI 0.2-2.9) at 2 years and 26.3% (95%CI 19.3-32.6) at 10 years. Up to 30% of the total lung volume, each elevation by 1% of ILD extent was associated to a 4% increase of the hazard of death (HR 1.04 95%CI 1.01-1.07, p=0.004). In addition to well-known poor prognosis predictors in SSc, the automated quantification of lung fibrosis on HRCT significantly improved the ability to predict short-term mortality (integrated discrimination index at 2 years 0.014, 95%CI 0.002-0.066, p=0.020, difference in the area under the curve of the two models 0.043, 95%CI 0.002-0.084, p=0.040).

Conclusion: The deep-learning-based, computer-aided quantification of ILD extent on HRCT provides an effective complementary tool for risk stratification in SSc.

Word count (abstract): 250 words

1 Introduction

Systemic sclerosis is a complex connective tissue disease characterized by fibrosis, vascular alteration and autoantibodies targeting cellular antigens leading to various internal organ involvement (1). Interstitial lung disease (ILD) is identified in 32 to 52% of patients with SSc and represents a major cause of mortality and morbidity in these patients (2,3). Scleroderma-related interstitial lung disease (SSc-ILD) can range from a non-progressive, asymptomatic disease, to a rapidly progressive fibrosis requiring treatment. In the EUSTAR cohort, pulmonary fibrosis was responsible for 16,8% of deaths and the main cause of SSc-related deaths (4). Identifying poor prognosis factors is a challenging issue for clinicians and it should be systematically sought when considering treatment and follow-up of patients with SSc. Indeed, treatment is decided on an individual basis and vary from abstention to autologous hematopoietic stem cell transplantation (AH SCT) in selected patients. In large meta-analysis, many risk factors for poor outcome were identified (5), but these parameters don't reflect the individual risk of death. Several scores have been proposed to predict mortality in SSc (6–9). Traditional approach to stratify ILD severity and patient's risk of death is based on pulmonary function tests (PFTs) and measurement of lung fibrosis extent on high-resolution computed tomography (HRCT) (10). Other parameters such as age, smoking status and 6-minute walking test are sometimes added in risk model for SSc and idiopathic pulmonary fibrosis (7–9). In the study from Goh et al. a threshold value of 20% was used to distinguish limited versus extensive pulmonary fibrosis. Combined with the value of forced vital capacity (FVC) on PFTs, the limited/extensive staging system was a strong predictor of death. However, this assessment relies on a visual, semi-quantitative evaluation of specific and non-specific lesions including ground-glass opacities, reticulations, traction bronchiectasis and honeycombing. It is a time-consuming method that requires an experienced radiologist and can be subject to inter-observer variability and subjectivity (11,12). The assessment of lung fibrosis with computer-aided automated methods could help stratify patient's prognosis and risk of death (13). Indeed, these methods allow an objective quantification of lung fibrosis, as compared to traditional visual semi-quantitative measurement. Furthermore, the development of a novel deep-learning method represents a promising technique to improve the accuracy of automated image analysis (14). We performed a retrospective, monocentric study to determine if an

algorithm-based automated quantification of lung fibrosis on HRCT was correlated to short-term and long-term mortality and may improve the risk stratification of patients with SSc.

2 Methods

2.1 Study population and data recording

We recruited patients from the database of the Centre de Référence des Maladies Auto-immunes Systémiques Rares d'Ile-De-France at Cochin Hospital, Paris, France. All consecutive patients with (1) a diagnosis of SSc according to the 2013 American College of Rheumatology (ACR)/European League Against Rheumatism Collaborative Initiative (EULAR) criteria and (2) who underwent a chest HRCT between September 2008 and March 2016 and PFTs within 3 months from CT-scan were eligible. The date of realization of the CT scan was considered as the inclusion date.

We collected clinical data from the patients' medical records in the Cochin Hospital medical database. Data collected at the time of inclusion were sex, age, smoking status, comorbidities, type of SSc, Rodnan skin score, New York Heart Association (NYHA) dyspnea score, Raynaud's phenomenon (RP), gastro-esophageal reflux. The date of apparition of the first non-RP symptom was considered as the disease onset. Additional detail on internal organ involvement definition and PFT recording is provided in supplementary file. Presence and target of autoantibodies (anti-nuclear antibodies) were recorded.

Cause and date of death were retrospectively determined with medical files. Pulmonary origin of death was considered when it was either the only or a major contributor to death according to medical files. Other dates of death were obtained through city-of-birth administrative files. In these cases, the cause was not obtained. Quantitative values are given as median and interquartile range (IQR).

2.2 Disease extent quantification on HRCT

The inclusion criteria regarding the CT examination characteristics were as follows:

- Acquisition performed in the supine position

- No use of contrast medium administration
- HRCT images reconstructed using a lung kernel, a 512x512 matrix and a slice thickness ≤ 2 mm.

Disease extent was automatically quantified using a previously reported deep-learning-based algorithm (15). This algorithm results from the training of a 2D convolutional neural network on models reflecting different thoracic morphologies (atlases). The developed solution is able to separate the lung from the surrounding tissue (chest wall and mediastinum) and the diseased lung from normal parenchyma to obtain the percentage of interstitial disease in the whole lung volume. The interstitial disease category encompasses ground glass opacities, reticulations, traction bronchiectasis and honeycombing, as a single category.

2.3 Outcomes

The main objective of our study was to determine if the extension of pulmonary fibrosis, measured with an automated algorithm-based method, was a predictive factor for short-term and long-term mortality. A secondary objective was to determine if this assessment could improve mortality risk stratification, in association with previously identified risk factors of poor prognosis, which were age, sex, type of SSc according to Leroy et Medsger (16), specific antibodies (anti-Scl-70, anti-centromer and anti-RNA-polymerase 3 antibodies), and dyspnea NYHA score. We also studied if the ILD extent as determined by HRCT was a predictive factor for pulmonary-related death.

As an exploratory analysis, we aimed to determine if our risk-stratification strategy could help to identify high mortality-risk patients who could be eligible for AHSCT. Further detail on eligibility criteria for AHSCT are given in supplementary file.

2.4 Statistical analysis

Overall survival curves were obtained using Kaplan-Meier estimator. Given the study follow-up, we censored survival at 10 years. The association of baseline characteristics with death was analyzed using Cox models. No variable selection was undertaken, and a set of potential prognostic variables was determined a priori. Those were sex, age, type of SSc, specific antibodies and dyspnea NYHA score. The measurement of ILD extent on HRCT was then added to that 'base' model.

We used the log-likelihood ratio as a global measure of model fit. The overall model discrimination was assessed by the concordance (*c*) index and Royston and Sauerbrei's *D* measure. To compare the improvement in risk prediction using CT-scan, we computed the difference in *c* index and in *D* between models without and with quantitative analysis of ILD extent, with their bootstrap 95% confidence interval (CI). We then used the extension of the integrated discrimination index (IDI) to survival data with bootstrap CI. Values were computed for short-term mortality (2 years) and long-term mortality (10 years). The area under the time-dependent ROC curve at 2 years was used as a measure of discrimination of early mortality.

2.5 Ethics

Patient information was given at the time of inclusion in the scleroderma database.

3 Results

3.1 Population description

Three hundred and forty-three patients were identified through the local data base of patients with systemic sclerosis. Among them, 25 patients did not meet inclusion criteria (6 patients did not have SSc according to ACR/EULAR criteria, and 19 patients did not have HRCT and PFTs performed). Three hundred and eighteen patients were included in the study. Patient characteristics at the time of inclusion are shown in Table 1.

Median disease duration at inclusion was 6 years, IQR 2-10. ILD was the most frequently reported internal organ involvement (196 patients, 61.6%). Muscular and myocardial involvement were reported in 44 (13.8%) and 40 (12.6%) patients each, respectively, and gastro-intestinal involvement (excluding gastro-esophageal reflux) was noticed in 34 patients (10.7%). PAH was reported in 24 cases (7.5%) and 16 patients had a past medical history of scleroderma renal crisis (5%).

Baseline PFTs were characterized by slightly impaired lung volumes, with a median FVC 93% predicted (IQR 75-105), contrasting with a significant decrease of DLCO (median corrected DLCO 64% predicted, IQR 49-77).

3.2 Main outcome

Median follow-up after HRCT was 94 months (IQR 73-111) corresponding to 26 406 person-months. Sixty-two patients died and the mortality rate was 1,6% (95% CI 0.2-2.9) at 2 years, 10,4% (95% CI 7.2-14.1) at 5 years and 26.3% (95% CI 19.3-32.6) at 10 years (Figure 1, panel A). Each percent of ILD extent's increase (up to 30% of the lung) was associated with a 4% increased risk in the hazard ratio (HR) of death (HR 1.04, 95% CI 1.01-1.07, $P=0.004$) (Figure 1, panel B). Only 39 patients had a pulmonary disease extent >30% and range of hazard ratio of predicted death was wide. We thus decided that patients with a disease extent>30% carried the same risk as 30%.

Univariate analysis enlightened four independent death predictor factors: older age, male sex, NYHA dyspnea score and ILD extent on HRCT. We constructed a risk predicting model including these factors and SSc characteristics (disease form and detected autoantibodies). This model showed a good ability to predict 10-years mortality (c Index 0,789 and D Index 1,83). Adding the automated measurement of ILD extent on HRCT to that base model slightly improved the ability to predict 10-years mortality but it did not reach statistical significance. When considering early mortality, ILD extent as determined with HRCT allowed to improve the ability to predict 2-years mortality (Table 2). Figure S1 shows the calibration plot of the model. FVC value at inclusion was missing for 8 patients, because the PFTs recording was incomplete. We then fitted the model in the sub-group of 310 patients with non-missing value and added FVC in the multivariable prediction model (Table S1). In this case, the assessment of ILD extent still allowed to improve very prognosis performance of the model (AUC 0.934 vs 0.913), although the difference did not reach statistical significance [$+0.021$ (-0.005 to 0.047)].

From this model, we divided the study population into five groups with an increasing risk of death (Figure 3). The patients' characteristics according to their predicted risk of short-term and long-term mortality are shown in Table 3. In the highest risk group, the overall survival was significantly poorer than in other groups: 92.2% (95% CI 85.8-99.0) at 2 years and 28.1 (95% CI 17.4-49.4) at 10 years.

Unfortunately, in most cases we were unable to determine the leading cause of death. Indeed, cause of death was missing or multifactorial. Therefore, it was not

possible to determine if the increased mortality rate observed in patients was related to pulmonary disease or not.

3.3 Secondary outcomes

Considering the whole cohort, 19 out of 318 patients fulfilled the criteria for AHSCT. Among the highest-risk group, 6 patients could theoretically fulfil the criteria for AHSCT. For 2 other patients, PFT results were lacking in order to assess their eligibility. For these 8 patients, the 2-year predicted risk of death overcome the procedure-related risk of death (predicted mortality of 7.8% at 2-years in this high-risk group). Thus, our risk assessment model helps to identify a subgroup of patients with a poor outcome who might benefit from more intensive therapy. For most patients in the group, AHSCT was contraindicated because they were more than 65 years old, or because they had PAH, cardiac and/or renal involvement. In addition, 36 patients in the high-risk group had a disease duration superior to 5 years at inclusion, consequence of our inclusion criteria.

4 Discussion

In this study we demonstrated a near-linear correlation between ILD extent on HRCT and 10-year mortality. This result is of paramount importance for clinicians who take care of patients. An automated measurement of ILD extent on HRCT as proposed in our study displays some advantages compared to visual assessment: its speed and its reproducibility. It gives an objective quantification of ILD extent, that can be repeated longitudinally to assess patient's evolution, an assessment that is easier and more accurate than the evaluation proposed by Goh et al. (10). In a previous study, Chassagnon et al. showed that there were fewer differences between the disease segmentation obtained by the algorithm and the one performed visually by a trained radiologist than among several radiologists (15). The near-linear correlation between ILD extent and mortality, per-se, might influence our therapeutic strategy in patients with SSc related ILD. Indeed, treatments of SSc-ILD were previously based on an aggravative course assessed by a decrease in FVC > 10% or a reduced DLCO of 15% or more (17,18). However, treatment with tyrosine kinase inhibitors (19) or tocilizumab (20) only brings modest effect on PFTs. Considering this modest impact on ILD, patients might benefit from an early introduction of these treatments in SSc-

related ILD. Each percent of lung fibrosis extension bearing its own risk of death, our result supports this wider use. However, we cannot demonstrate from our study that the reduced progression of ILD with new therapeutic approaches might improve the prognosis.

In our study, we analyzed a single HRCT for each patient. However, some study tried to identify an evolutive profile for each patient (21–23). Our automated method could be helpful in reaching this objective. In this view, it could help to identify the patients with a progressive disease who might have the greatest benefit of anti-fibrotic treatments.

Visual evaluation of fibrosis extent on HRCT is often proposed as a prognostic tool (10,21,22), but it is rough, time consuming and it requires a well-trained radiologist. Few prediction models have been proposed to help determine patients' individual risk of death. The modified Du Bois index showed a good ability to predict 1-year mortality in SSc-ILD (31). However, this model requires the 24-week change in 6-minute walking distance and is not a single-shot evaluation (24). The SADL model was developed in patients with SSc-related ILD and could accurately predict 3-year mortality (8). However, it cannot apply to unselected patients with SSc. The SCOpE score, calculated from 13 parameters, separated the cohort into 4 different risk groups and accurately predicted 3-year mortality in the EUSTAR SSc cohort (4). Last, in the score developed by Bryan et al. (6) and recently validated (7), 5-year predicted risk of death was estimated and the area under the ROC curve was 0.78. However, some of the parameters (eg. ESR and proteinuria) may vary according to clinical conditions. We propose here an easy-to-use mortality risk predictor derived from 6 parameters only. Our model can predict individual risk of short and long-term mortality up to 10 years, which might represent a decisive information when considering aggressive treatment regimens. In addition, the automated quantification of lung fibrosis helps to identify the patients at higher risk of death at year 2. In some selected patients, it might help to discuss early intensive therapeutic measures, instead of waiting for a significant worsening (25). AHSCT can be proposed to patients with a rapidly progressive SSc responsible for a high short-term risk of death, but the procedure is characterized by a 3% to 10% related mortality (26–28). Our model could help to reach this goal as the window of opportunity is narrow although eligibility criteria and contraindications of AHSCT should be sought carefully (29).

Our study presents some strengths: it was a large cohort of 318 patients, led in a referral center over a long time period. Each patient had an ascertained diagnosis of SSc and a complete screening for complications. The baseline characteristics of the study population were similar to those observed in previous SSc cohorts (30). The survival rate was comparable to recent data (7,31,32).

Our study also has some limitations: we were unable to determine the causes of death, and it was not possible to examine the specific weight of lung fibrosis in the risk of death. Our cohort was not large enough to analyze every variable correlated with mortality, so we had to narrow our model to 6 pre-defined parameters. Our prediction model needs to be validated in another cohort to confirm its results and performance.

5 Conclusion

Scientific knowledge on the subject: Clinicians lack reliable tools to predict individual risk of short-term and long-term mortality in SSc. ILD is a frequent and severe complication of SSc. Its assessment on HRCT often relies on visual semi-quantitative methods that carry some subjectivity. Artificial intelligence may respond to this issue.

What this study adds to the field: The extent of lung fibrosis is correlated to mortality in SSc, in a near-linear shape. We propose an algorithm-based automated quantification of ILD extent on HRCT as an effective complementary prognostic tool to help stratify the risk of death in SSc. Its use is fast, simple, and more accurate than visual assessment. In addition to traditional risk factors (gender, age, autoantibody profile, type of disease, dyspnea score), it might help clinicians in identifying patients with high risk of mortality and could be useful in routine practice.

6 References

1. Denton CP, Khanna D. Systemic sclerosis. *Lancet Lond Engl.* 7 oct 2017;390(10103):1685-99.
2. Bergamasco A, Hartmann N, Wallace L, Verpillat P. Epidemiology of systemic sclerosis and systemic sclerosis-associated interstitial lung disease. *Clin Epidemiol.* 2019;11:257-73.
3. Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972-2002. *Ann Rheum Dis.* juill 2007;66(7):940-4.
4. Elhai M, Meune C, Boubaya M, Avouac J, Hachulla E, Balbir-Gurman A, et al. Mapping and predicting mortality from systemic sclerosis. *Ann Rheum Dis.* 2017;76(11):1897-905.

5. Rubio-Rivas M, Royo C, Simeón CP, Corbella X, Fonollosa V. Mortality and survival in systemic sclerosis: Systematic review and meta-analysis. *Semin Arthritis Rheum.* oct 2014;44(2):208-19.
6. Bryan C, Knight C, Black CM, Silman AJ. Prediction of five-year survival following presentation with scleroderma: Development of a simple model using three disease factors at first visit. *Arthritis Rheum.* 1999;42(12):2660-5.
7. Franssen J, Popa-Diaconu D, Hesselstrand R, Carreira P, Valentini G, Beretta L, et al. Clinical prediction of 5-year survival in systemic sclerosis: validation of a simple prognostic model in EUSTAR centres. *Ann Rheum Dis.* oct 2011;70(10):1788-92.
8. Morisset J, Vittinghoff E, Elicker BM, Hu X, Le S, Ryu JH, et al. Mortality Risk Prediction in Scleroderma-Related Interstitial Lung Disease: The SADL Model. *Chest.* 2017;152(5):999-1007.
9. Mango RL, Matteson EL, Crowson CS, Ryu JH, Makol A. Assessing Mortality Models in Systemic Sclerosis-Related Interstitial Lung Disease. *Lung.* 2018;196(4):409-16.
10. Goh NSL, Desai SR, Veeraghavan S, Hansell DM, Copley SJ, Maher TM, et al. Interstitial lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med.* 1 juin 2008;177(11):1248-54.
11. Collins CD, Wells AU, Hansell DM, Morgan RA, MacSweeney JE, Du Bois RM, et al. Observer variation in pattern type and extent of disease in fibrosing alveolitis on thin section computed tomography and chest radiography. *Clin Radiol.* avr 1994;49(4):236-40.
12. Sverzellati N, Devaraj A, Desai SR, Quigley M, Wells AU, Hansell DM. Method for Minimizing Observer Variation for the Quantitation of High-Resolution Computed Tomographic Signs of Lung Disease: J Comput Assist Tomogr. sept 2011;35(5):596-601.
13. Ariani A, Silva M, Seletti V, Bravi E, Saracco M, Parisi S, et al. Quantitative chest computed tomography is associated with two prediction models of mortality in interstitial lung disease related to systemic sclerosis. *Rheumatol Oxf Engl.* 01 2017;56(6):922-7.
14. Vakalopoulou M, Chassagnon G, Bus N, Marini R, Zacharaki EI, Revel M-P, et al. AtlasNet: Multi-atlas Non-linear Deep Networks for Medical Image Segmentation. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* [Internet]. Springer, Cham; 2018 [cité 15 janv 2021]. p. 658-66. Disponible sur: https://link.springer.com/chapter/10.1007/978-3-030-00937-3_75
15. Chassagnon G, Vakalopoulou M, Régent A, Zacharaki EI, Aviram G, Martin C, et al. Deep Learning-based Approach for Automated Assessment of Interstitial Lung Disease in Systemic Sclerosis on CT Images. *Radiol Artif Intell.* 1 juill 2020;2(4):e190006.
16. LeRoy EC, Medsger TA. Criteria for the classification of early systemic sclerosis. *J Rheumatol.* juill 2001;28(7):1573-6.
17. Moore O, Proudman S, Goh N, Tj C, HR, OH, et al. Quantifying change in pulmonary function as a prognostic marker in systemic sclerosis-related interstitial lung disease. *Clin Exp Rheumatol* [Internet]. août 2015 [cité 14 juill 2021];33(4 Suppl 91). Disponible sur: <https://pubmed.ncbi.nlm.nih.gov/26243401/>
18. Khanna D, Mittoo S, Aggarwal R, Proudman SM, Dalbeth N, Matteson EL, et al. Connective Tissue Disease-associated Interstitial Lung Diseases (CTD-ILD) — Report from OMERACT CTD-ILD Working Group. *J Rheumatol.* nov 2015;42(11):2168-71.
19. Distler O, Highland KB, Gahlemann M, Azuma A, Fischer A, Mayes MD, et al. Nintedanib for Systemic Sclerosis-Associated Interstitial Lung Disease. *N Engl J Med.* 27 2019;380(26):2518-28.
20. Khanna D, Lin CJF, Furst DE, Goldin J, Kim G, Kuwana M, et al. Tocilizumab in systemic sclerosis: a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Respir Med.* oct 2020;8(10):963-74.
21. Forestier A, Le Gouellec N, Béhal H, Kramer G, Perez T, Sobanski V, et al. Evolution of high-resolution CT-scan in systemic sclerosis-associated interstitial lung disease: Description and prognosis factors. *Semin Arthritis Rheum.* déc 2020;50(6):1406-13.
22. Hoffmann-Vold A-M, Aaløkken TM, Lund MB, Garen T, Midtvedt Ø, Brunborg C, et al. Predictive Value of Serial High-Resolution Computed Tomography Analyses and Concurrent Lung Function Tests in Systemic Sclerosis: SERIAL LUNG ANALYSES IN SSc. *Arthritis Rheumatol.* mai 2015;67(8):2205-12.

23. Carnevale A, Silva M, Maietti E, Milanese G, Saracco M, Parisi S, et al. Longitudinal change during follow-up of systemic sclerosis: correlation between high-resolution computed tomography and pulmonary function tests. *Clin Rheumatol*. janv 2021;40(1):213-9.
24. du Bois RM, Albera C, Bradford WZ, Costabel U, Leff JA, Noble PW, et al. 6-minute walk distance is an independent predictor of mortality in patients with idiopathic pulmonary fibrosis. *Eur Respir J*. 1 mai 2014;43(5):1421-9.
25. Distler O, Volkman ER, Hoffmann-Vold AM, Maher TM. Current and future perspectives on management of systemic sclerosis-associated interstitial lung disease. *Expert Rev Clin Immunol*. oct 2019;15(10):1009-17.
26. Burt RK, Shah SJ, Dill K, Grant T, Gheorghiadu M, Schroeder J, et al. Autologous non-myeloablative haemopoietic stem-cell transplantation compared with pulse cyclophosphamide once per month for systemic sclerosis (ASSIST): an open-label, randomised phase 2 trial. *Lancet Lond Engl*. 6 août 2011;378(9790):498-506.
27. van Laar JM, Farge D, Sont JK, Naraghi K, Marjanovic Z, Larghero J, et al. Autologous hematopoietic stem cell transplantation vs intravenous pulse cyclophosphamide in diffuse cutaneous systemic sclerosis: a randomized clinical trial. *JAMA*. 25 juin 2014;311(24):2490-8.
28. Sullivan KM, Goldmuntz EA, Keyes-Elstein L, McSweeney PA, Pinckney A, Welch B, et al. Myeloablative Autologous Stem-Cell Transplantation for Severe Scleroderma. *N Engl J Med*. 04 2018;378(1):35-47.
29. Farge D, for the EBMT ADWP Scleroderma Working Group, on behalf of the Joint Accreditation Committee of the International Society for Cellular Therapy (JACIE) & European Society for Blood and Marrow Transplantation (EBMT), Burt RK, Oliveira M-C, Mousseaux E, et al. Cardiopulmonary assessment of patients with systemic sclerosis for hematopoietic stem cell transplantation: recommendations from the European Society for Blood and Marrow Transplantation Autoimmune Diseases Working Party and collaborating partners. *Bone Marrow Transplant*. nov 2017;52(11):1495-503.
30. Sobanski V, Giovannelli J, Allanore Y, Riemekasten G, Airò P, Vettori S, et al. Phenotypes Determined by Cluster Analysis and Their Survival in the Prospective European Scleroderma Trials and Research Cohort of Patients With Systemic Sclerosis. *Arthritis Rheumatol Hoboken NJ*. sept 2019;71(9):1553-70.
31. Hoffmann-Vold A-M, Fretheim H, Halse A-K, Seip M, Bitter H, Wallenius M, et al. Tracking Impact of Interstitial Lung Disease in Systemic Sclerosis in a Complete Nationwide Cohort. *Am J Respir Crit Care Med*. 15 nov 2019;200(10):1258-66.
32. Ferri C, Valentini G, Cozzi F, Sebastiani M, Michelassi C, La Montagna G, et al. Systemic Sclerosis: Demographic, Clinical, and Serologic Features and Survival in 1,012 Italian Patients. *Medicine (Baltimore)*. mars 2002;81(2):139-53.

Table 1. Main clinical and biological characteristics of patients at the time of inclusion.

Variable	N = 318
Age, mean (SD) y	52.7 (14.8)
Female, no. (%)	271 (85.2)
Smoking status, no. (%)	
Current smoker	46 (14.5)
Never smoker	197 (61.9)
Former smoker	75 (23.6)
Disease type, no. (%)	
Limited cutaneous SSc	171 (53.8)
Diffuse cutaneous SSc	129 (40.6)
SSc sine scleroderma	18 (5.7)
Rodnan score, median (IQR)	6 (2 to 14) [n=295]
Organ involvement, no. (%)	
Raynaud's phenomenon	309 (97.2)
Gastroesophageal reflux	297 (93.4)
Other gastrointestinal	34 (10.7)
Lung	196 (61.6)
Muscle	44 (13.8)
Kidney	16 (5.0)
Heart	40 (12.6)
Pulmonary hypertension	24 (7.5)
Antibodies	
Anti-Scl-70, no. (%)	122 (38.4)
Anti-centromer, no. (%)	90 (28.3)
Anti-SSA/Ro, no. (%)	38 (11.9)
Anti-U1-RNP, no. (%)	26 (8.2)
Anti-RNA polymerase III, no. (%)	15 (4.7)
Isolated antinuclear antibodies, no. (%)	39 (12.3)
NYHA score, no. (%)	
1	162 (50.9)
2	118 (37.1)
3	36 (11.3)
4	2 (0.6)
Pulmonary function tests	
TLC, median (IQR) % predicted	96 (82 to 108) [n=310]
FEV1, median (IQR) % predicted	90 (76 to 102) [n=227]
FVC, median (IQR) % predicted	93 (75 to 105) [n=310]
DLCOc, median (IQR) % predicted	64 (49 to 77) [n=289]
CT scan	
ILD extent, median (IQR) %	3.5 (0.3 to 15.9)

Table 2. Association of predefined variables with mortality up to 10 years.

Variable	HR (95% CI)	<i>P</i>	HR (95% CI)	<i>P</i>
Age (per 10 y)	1.25 (1.01 to 1.55)	0.040	1.31 (1.04 to 1.64)	0.022
Male sex	2.05 (1.16 to 3.62)	0.013	1.95 (1.10 to 3.44)	0.022
SSc type				
Limited cutaneous	1 (reference)	—	1 (reference)	—
Diffuse cutaneous	1.62 (0.87 to 3.03)	0.13	1.52 (0.80 to 2.86)	0.20
Sine scleroderma	0.61 (0.081 to 4.51)	0.62	0.70 (0.093 to 5.22)	0.73
Anti-Scl-70 antibodies	1.47 (0.76 to 2.86)	0.26	1.43 (0.73 to 2.81)	0.30
Anti-centromer antibodies	0.78 (0.34 to 1.80)	0.56	1.17 (0.47 to 2.89)	0.74
Anti-RNA-polymerase 3 antibodies	0.90 (0.27 to 3.00)	0.86	1.17 (0.35 to 3.91)	0.80
NYHA score*	2.61 (1.82 to 3.76)	< 0.0001	2.30 (1.57 to 3.37)	< 0.0001
ILD extent (per % until 30%)**	—	—	1.04 (1.01 to 1.07)	0.004
Model LR	69.38	—	77.92	0.007†
Model <i>c</i> index	0.789	—	0.795	—
Difference in <i>c</i> index	—	—	0.005 (-0.004 to 0.033)†	—
<i>D</i>	1.83	—	2.01	—
Difference in <i>D</i>	—	—	0.18 (-0.007 to 0.63)†	—
IDI at 2 years	—	—	0.014 (0.002 to 0.066)†	0.020†
IDI at 10 years	—	—	0.029 (-0.005 to 0.072)†	0.086†
AUC at 2 years	0.888	—	0.931	—
Difference in AUC at 2 years	—	—	0.043 (0.002 to 0.084)†	0.040†

* 2 (0.6%) score 4 grouped with score 3.

** 39 (12.2%) values above 30% were considered equal to 30% for association with death

† Comparison with the model without ILD extent.

Table 3. Characteristics at inclusion, predicted and observed survival according to fifths of predicted risk. The “highest risk” group is the fifth of individuals in the sample with highest predicted risk, and the “high risk” groups comprises individuals in the second fifth of highest risk.

Variable	Highest risk	High risk	Other fifths
No. patients	64	63	191
Age, mean (SD) y	58.9 (11.9)	56.9 (13.7)	49.2 (15.1)
Female, no. (%)	45 (70.3)	52 (82.5)	174 (91.1)
Smoking status, no. (%)			
Current smoker	4 (6.2)	5 (7.9)	37 (19.4)
Never smoker	41 (64.1)	41 (65.1)	115 (60.2)
Former smoker	19 (29.7)	17 (27.0)	39 (20.4)
Disease type, no. (%)			
Limited cutaneous SSc	17 (26.6)	28 (44.4)	126 (66.0)
Diffuse cutaneous SSc	46 (71.9)	34 (54.0)	49 (25.7)
SSc sine scleroderma	1 (1.6)	1 (1.6)	16 (8.4)
Time from diagnosis, median (IQR) y	6.0 (2.3 to 11.3)	5.9 (2.6 to 11.9)	5.8 (2.5 to 9.6)
No. (%) < 5y	28 (43.8)	26 (41.3)	84 (44.2)
Rodnan score, median (IQR)	10 (6 to 22)	9 (4 to 18)	4 (2 to 10)
Antibodies			
Anti-Scl-70, no. (%)	46 (71.9)	25 (39.7)	51 (26.7)
Anti-centromer, no. (%)	4 (6.2)	16 (25.4)	70 (36.6)
Anti-SSA/Ro, no. (%)	3 (4.7)	9 (14.3)	26 (13.6)
Anti-U1-RNP, no. (%)	1 (1.6)	7 (11.1)	18 (9.4)
Anti-RNA polymerase III, no. (%)	4 (6.2)	4 (6.3)	7 (3.7)
Isolated antinuclear antibodies, no. (%)	6 (9.4)	4 (6.3)	29 (15.2)
NYHA score, no. (%)			
1	1 (1.6)	9 (14.3)	152 (79.6)
2	33 (51.6)	47 (74.6)	38 (19.9)
3	29 (45.3)	6 (9.5)	1 (0.5)
4	1 (1.6)	1 (1.6)	0 (0)
Pulmonary function tests			
TLC, median (IQR) % predicted	77 (68 to 87)	92 (76 to 105)	102 (93 to 114)
FEV, median (IQR) % predicted	75 (66 to 89)	86 (74 to 96)	94 (86 to 108)
FVC, median (IQR) % predicted	73 (60 to 86)	86 (74 to 100)	99 (87 to 114)
DLCOc, median (IQR) % predicted	42 (36 to 51)	56 (45 to 69)	71 (61 to 80)
CT scan			
ILD extent, median (IQR) %	26.6 (16.9 to 35.1)	11.0 (1.5 to 20.1)	0.7 (0.2 to 4.9)
Predicted 10y survival*, median (IQR) %	56.0 (33.8 to 65.2)	79.4 (74.4 to 82.5)	94.0 (90.8 to 96.1)
Overall survival (95% CI) %			
2-year	92.2 (85.8 to 99.0)	100 (NE)	100 (NE)
10-year	28.1 (17.4 to 49.4)	68.9 (56.1 to 84.6)	91.3 (86.0 to 97.0)

NE: not estimated

*Individual risk score was calculated as follow :

Score = 0.0267 x age + 0.666 if sex is male + 0.417 if diffuse cutaneous SSc -0.359 if sine scleroderma + 0.359 if presence of anti-Scl-70 antibodies + 0.155 if presence of anti-centromere antibodies + 0.155 if presence of anti-RNA-polymerase 3 antibodies + 0.832 x NYHA score (a score of 4 being assigned 3 points) + 0.0382 x percent of ILD extent (values > 30% being considers equal to 30%)

Then, predicted mortality was calculated as follow :

Mortality = 1 - 0.8787^{exp(Score - 3.5276)}

Figure 1. Mortality over 10 years of follow-up. Panel A, Kaplan-Meier estimation of mortality over 10 years; the shaded region represents the pointwise 95% confidence interval. Panel B, smooth hazard ratio for ILD extension; the shaded region represents the pointwise 95% confidence interval. The yellow curve shows the hazard ratio when ILD extension > 30% carried the same risk as 30%. The grey histogram represents the distribution in the study sample.

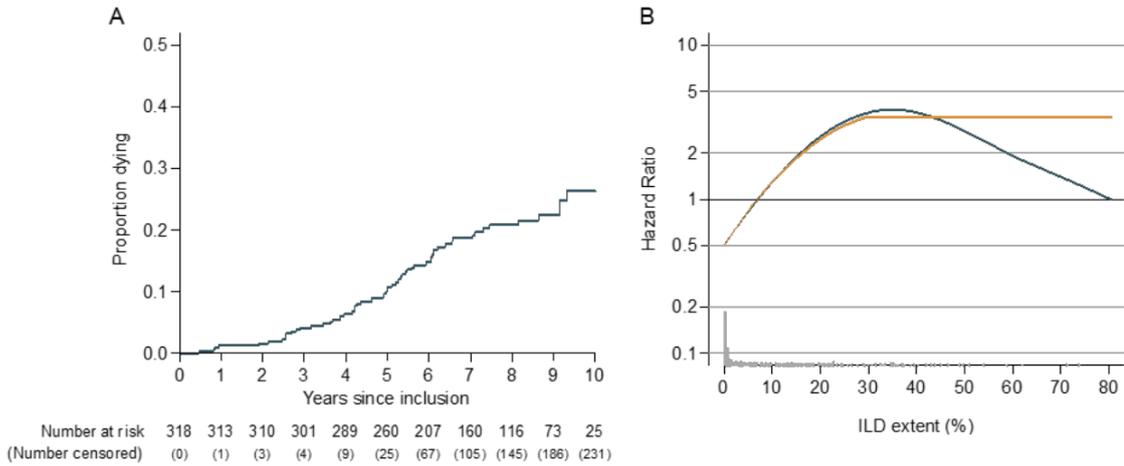
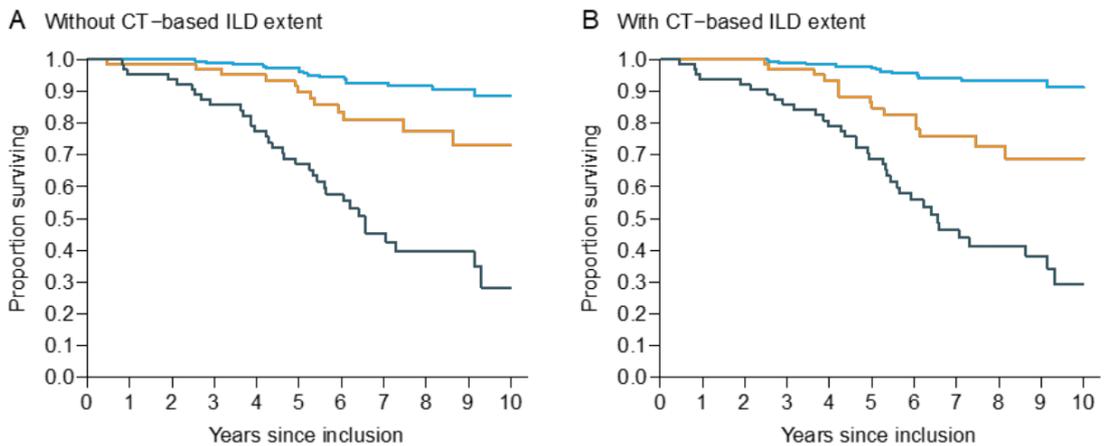


Figure 2. Survival according to predicted risk. The light blue curves represent individuals in the three lower fifths of predicted risk, the orange one the fourth fifth and the dark blue one the fifth with highest predicted risk.



Supplementary Material

Additional detail on the study methods:

Internal organ involvement was considered SSc-related unless specified otherwise, and was defined as follow:

- Myocardial involvement was defined by a decrease of ejection fraction $< 50\%$ or by a sustained elevation of troponine blood level > 2 Upper Limit of Normal (ULN) or by arrhythmias or conduction defects not attributable to other cardiac conditions and requiring treatment.
- Pulmonary arterial hypertension (PAH) was defined by a right heart catheterization showing a mean pulmonary arterial pressure > 20 mmHg with pulmonary vascular resistance ≥ 3 Wood Units, and pulmonary arterial wedge pressure ≤ 15 mmHg in order to exclude secondary PAH.
- Kidney involvement was defined by an history of scleroderma renal crisis, i.e association of acute renal failure and hypertension with no other plausible cause than SSc.
- Muscular involvement was defined by a sustained elevation of blood creatine kinase > 3 ULN or a muscular biopsy showing inflammatory lesions with no other cause than SSc.
- Gastro-intestinal tract involvement was defined by the use of artificial nutritional support, or sequential antibiotherapy to control microbial overgrowth, or dysmotility treatment (excluding simple prokinetic treatment of gastro-esophageal reflux).
- Lung involvement was defined by the presence of SSc compatible interstitial lung lesions on CT-scan considered as significant according to the referent clinician for each patient.

Each patient underwent PFTs within 3 months before or after the inclusion HRCT. When available in medical records, we collected the following data: FVC, Total Lung Capacity (TLC) and corrected Diffusing Capacity of Carbon Monoxide (DLCO). Measurements were given as percentage of the predicted value adjusted for age, sex, body height and origin.

Criteria for considering autologous hematopoietic stem cell transplantation were as follow:

- Age < 65 years.
- Diffuse systemic sclerosis with a less than 5 years disease duration.
- Mild pulmonary impairment (FVC > 65% and DLCOc > 40% of predicted values, respectively).
- No PAH, or kidney or cardiac involvement at the time of inclusion.

Active tobacco use may also be considered as a possible contraindication by some specialists in the field.

Table S1. Multivariable models with FVC.

The models were fitted in the subgroup of 310 individuals with non-missing values for FVC (8, 2.5% with missing values).

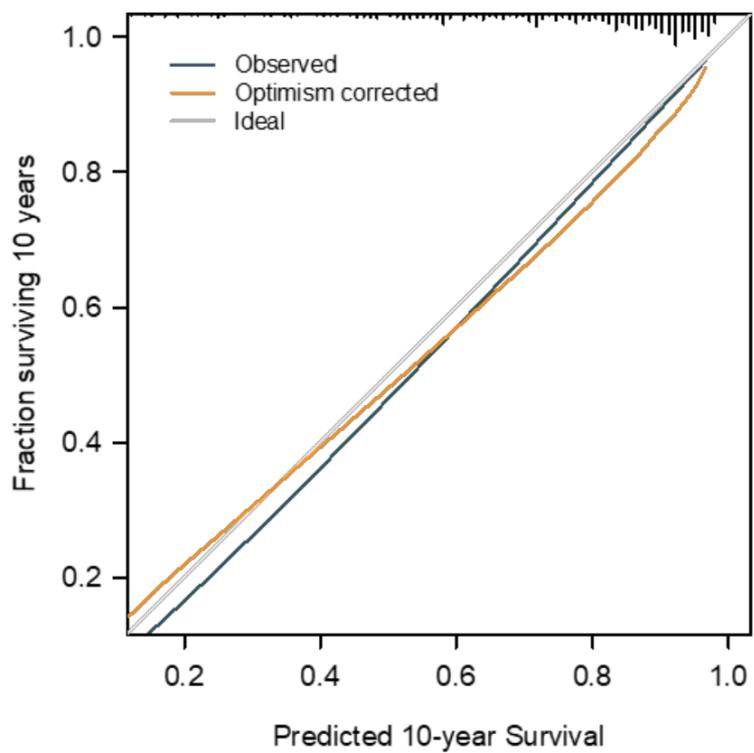
Variable	HR (95% CI)	<i>P</i>	HR (95% CI)	<i>P</i>
Age (per 10 y)	1.32 (1.06 to 1.66)	0.014	1.35 (1.06 to 1.70)	0.013
Male sex	1.75 (0.98 to 3.15)	0.061	1.72 (0.95 to 3.12)	0.073
SSc type				
Limited cutaneous	1 (reference)	—	1 (reference)	—
Diffuse cutaneous	1.54 (0.81 to 2.93)	0.19	1.51 (0.79 to 2.90)	0.21
Sine scleroderma	0.63 (0.084 to 4.69)	0.65	0.69 (0.092 to 5.14)	0.71
Anti-Scl-70 antibodies	1.30 (0.66 to 2.56)	0.44	1.32 (0.67 to 2.61)	0.43
Anti-centromer antibodies	0.88 (0.38 to 2.05)	0.77	1.18 (0.48 to 2.91)	0.72
Anti-RNA-polymerase 3 antibodies	1.17 (0.35 to 3.94)	0.80	1.27 (0.38 to 4.25)	0.70
NYHA score*	2.26 (1.53 to 3.33)	<0.0001	1.36 (1.03 to 1.79)	0.0002
FVC (per % predicted)	0.99 (0.97 to 1.00)	0.047	0.99 (0.98 to 1.01)	0.26
ILD extent (per % until 30%)**	—	—	1.03 (1.00 to 1.06)	0.028
Model LR	70.14	—	74.93	0.057†
Model <i>c</i> index	0.791	—	0.793	—
Difference in <i>c</i> index	—	—	0.002 (-0.008 to 0.023)†	—
<i>D</i>	1.85	—	2.00	—
Difference in <i>D</i>	—	—	0.15 (-0.03 to 0.49)†	—
IDI at 2 years	—	—	0.008 (0.000 to 0.034)†	0.053†
IDI at 10 years	—	—	0.02 (-0.008 to 0.073)†	0.17†
AUC at 2 years	0.913	—	0.934	—
Difference in AUC at 2 years	—	—	0.021 (-0.005 to 0.047)†	0.10†

* 2 (0.6%) score 4 grouped with score 3.

** 39 (12.6%) values above 30% were considered equal to 30% for association with death

† Comparison with the model without ILD extension.

Figure S1. Calibration plot of the model.



6 ÉTUDE DE L'IMPACT DU FILTRE DE RECONSTRUCTION SUR LE DEVELOPEMENT ET L'UTILISATION D'ALGORITHMES DE DEEP LEARNING POUR LA SEGMENTATION DES MALADIES PULMONAIRES DIFFUSES

6.1 CONTEXTE ET OBJECTIFS

Au cours de la dernière décennie, le deep learning, ou deep-learning, est devenu la méthode de choix pour l'automatisation de nombreuses applications d'analyse d'images tel que la segmentation (186). La segmentation consiste à identifier des structures dans l'image en classant chaque pixel de l'image comme appartenant ou n'appartenant pas à cette structure. Le gain de performances apportés par l'utilisation du deep learning permet désormais de segmenter avec précision des entités complexes telles que les lésions de pneumonie interstitielle (ILD) ou de pneumonie à coronavirus 2019 (COVID-19)(239,251,332,333).

Une particularité des scanners thoraciques est que les images sont reconstruites avec deux filtres de reconstruction différents, un filtre de reconstruction dit de haute fréquence ou filtre pulmonaire dédié à l'évaluation du parenchyme pulmonaire et un filtre de reconstruction dit de basse fréquence ou filtre médiastinal dédié à l'évaluation du médiastin. Des études récentes ont montré que les performances des outils de deep learning pour les tâches de détection et de caractérisation des nodules pulmonaires sont influencées par les paramètres d'acquisition et de reconstruction, tels que le choix du filtre de reconstruction (178,179). Cependant, il n'existe actuellement aucune donnée évaluant l'influence des filtres de reconstruction sur le développement et l'utilisation d'outils de deep learning pour la segmentation des maladies pulmonaires diffuses. Nous avons émis l'hypothèse que la combinaison des filtres de reconstruction pourrait améliorer les performances des modèles. Le but de ce travail était ainsi de déterminer l'impact du filtre de reconstruction sur le développement et l'utilisation de méthodes de deep learning et de déterminer si la combinaison d'images reconstruites avec les 2 filtres pourrait améliorer les résultats.

Dans ce but, nous avons utilisé une architecture de deep learning, un réseau de type U-net (246) dans 2 cas d'usage : la segmentation de la pneumopathie interstitielle dans la sclérodermie et la segmentation de la pneumonie COVID-19.

Nos résultats montrent l'importance du choix du filtre de reconstruction et son impact sur les méthodes de segmentation. Pour l'entraînement des modèles, la combinaison d'images reconstruites avec les 2 filtres de reconstruction améliore la qualité des segmentations comparativement à un entraînement sur des images toutes reconstruites avec le même filtre de reconstruction. Pour l'application des modèles, les performances sont meilleures lorsqu'on applique le modèle sur des images reconstruites avec le même filtre de reconstruction que celui utilisé pour son apprentissage. Lorsque le modèle a été reconstruit à partir d'images reconstruites avec les 2 filtres, il n'y a alors pas ou peu d'impact du choix du filtre de reconstruction pour les images sur lequel sera appliqué le modèle.

6.2 PUBLICATION

Deep learning for lung disease segmentation on CT: which reconstruction kernel should be used?

Trieu-Nghi HOANG-THI; Maria VAKALOPOULOU; Stergios CHRISTODOULIDIS; Nikos PARAGIOS; Marie-Pierre REVEL; Guillaume CHASSAGNON

Diagnostic and Interventional Imaging

Article accepted

6.2.1 RESUME

Objectif : Déterminer si un seul noyau de reconstruction ou les deux filtres de haute et basse fréquence doivent être utilisés pour former des modèles d'apprentissage en profondeur pour la segmentation de la maladie pulmonaire diffuse sur le scanner thoracique.

Matériels et méthodes : Deux jeux de données annotés sur la pneumonie COVID-19 (323 960 tranches) et la maladie pulmonaire interstitielle (ILD)(4 284 tranches) ont été utilisés. Des images TDM annotées ont été utilisées pour former une architecture UNet pour segmenter la maladie. Toutes les coupes CT avaient été reconstruites en utilisant à la fois un noyau pulmonaire et un noyau médiastinal. Trois entraînements différents résultant en 3 modèles différents ont été comparés pour chaque maladie : entraînement sur NP uniquement, NM uniquement ou images LK+MK. Les scores de similarité Dice (DSC) ont été comparés à l'aide du test de rang signé de Wilcoxon.

Résultats : Les modèles uniquement entraînés sur NP ont obtenu de meilleurs résultats sur les images LK que sur les images NM (médiane DSC = 0,62 (intervalle interquartile (IQR) = 0,54-0,69) vs 0,60 (IQR = 0,50-0,70), $P < 0,001$ pour COVID-19 et médiane DSC = 0,62 (IQR = 0,56-0,69) vs 0,50 (IQR = 0,43-0,57), $P < 0,001$ pour ILD. De même, les modèles uniquement entraînés sur NM ont de meilleurs résultats sur les images NM (médiane DSC = 0,62 (IQR = 0,53-0,68) vs 0,54 (IQR = 0,47-0,63), $P < 0,001$ pour COVID-19 et 0,69 (IQR=0,61-0,73) vs 0,63 (IQR=0,53-0,70), $p < 0,001$ pour ILD. Les modèles entraînés sur les deux filtres ont obtenu de meilleurs résultats ou de manière similaire que ceux entraînés sur un seul noyau. Pour COVID-19, le DSC médian était de 0,67 (IQR = 0,59-0,73) lorsqu'il était appliqué sur des images de NP et de 0,67 (IQR = 0,60-0,74) lorsqu'il était appliqué sur des images de NM, $P < 0,001$ pour les deux. Pour l'ILD, le DSC médian était de 0,69 (IQR = 0,63-0,73) lorsqu'il était appliqué sur des images de NP, $P = 0,006$ et 0,68 (IQR = 0,62-0,72) lorsqu'il était appliqué sur des images de NM, $P > 0,99$.

Conclusion : Les filtres de reconstruction ont un impact sur les performances des modèles basés sur le deep-learning pour la segmentation des maladies pulmonaires. L'entraînement sur les images de NP et NM améliore les performances.

6.2.2 MANUSCRIT D'ARTICLE



ELSEVIER



Original article

Deep learning for lung disease segmentation on CT: Which reconstruction kernel should be used?

Trieu-Nghi Hoang-Thi^{a,b}, Maria Vakalopoulou^c, Stergios Christodoulidis^c, Nikos Paragios^{c,d}, Marie-Pierre Revel^{a,b}, Guillaume Chassagnon^{a,b,*}

^a Université de Paris, Faculté de Médecine, 75006 Paris, France

^b Department of Radiology, Hôpital Cochin, AP-HP, centre, 75014 Paris, France

^c Université Paris-Saclay, CentraleSupélec, Mathématiques et Informatique pour la Complexité et les Systèmes, 3 91190 Gif-sur-Yvette, France

^d TheraPanacea, 75014 Paris, France

ARTICLE INFO

Article History:

Available online xxx

Keywords:

Deep learning

Multidetector computed tomography

Lung

ABSTRACT

Purpose: The purpose of this study was to determine whether a single reconstruction kernel or both high and low frequency kernels should be used for training deep learning models for the segmentation of diffuse lung disease on chest computed tomography (CT).

Materials and methods: Two annotated datasets of COVID-19 pneumonia (323,960 slices) and interstitial lung disease (ILD) (4,284 slices) were used. Annotated CT images were used to train a U-Net architecture to segment disease. All CT slices were reconstructed using both a lung kernel (LK) and a mediastinal kernel (MK). Three different trainings, resulting in three different models were compared for each disease: training on LK only, MK only or LK+MK images. Dice similarity scores (DSC) were compared using the Wilcoxon signed-rank test.

Results: Models only trained on LK images performed better on LK images than on MK images (median DSC = 0.62 [interquartile range (IQR): 0.54, 0.69] vs. 0.60 [IQR: 0.50, 0.70], $P < 0.001$ for COVID-19 and median DSC = 0.62 [IQR: 0.56, 0.69] vs. 0.50 [IQR: 0.43, 0.57], $P < 0.001$ for ILD). Similarly, models only trained on MK images performed better on MK images (median DSC = 0.62 [IQR: 0.53, 0.68] vs. 0.54 [IQR: 0.47, 0.63], $P < 0.001$ for COVID-19 and 0.69 [IQR: 0.61, 0.73] vs. 0.63 [IQR: 0.53, 0.70], $P < 0.001$ for ILD). Models trained on both kernels performed better or similarly than those trained on only one kernel. For COVID-19, median DSC was 0.67 (IQR: =0.59, 0.73) when applied on LK images and 0.67 (IQR: 0.60, 0.74) when applied on MK images ($P < 0.001$ for both). For ILD, median DSC was 0.69 (IQR: 0.63, 0.73) when applied on LK images ($P = 0.006$) and 0.68 (IQR: 0.62, 0.72) when applied on MK images ($P > 0.99$).

Conclusion: Reconstruction kernels impact the performance of deep learning-based models for lung disease segmentation. Training on both LK and MK images improves the performance.

© 2021 Société française de radiologie. Published by Elsevier Masson SAS. All rights reserved.

1. Introduction

Over the last decade, deep learning (DL) has become the method of choice for many applications in medical image analysis, especially for segmentation tasks [1–3]. The performance gain in image segmentation achieved with DL methods over traditional machine learning methods makes it now possible to accurately segment complex entities within the lung, such as interstitial lung disease (ILD) or coronavirus disease 2019 (COVID-19) pneumonia [4–7].

A particularity of chest computed tomography (CT) is that images are usually reconstructed using two different reconstruction kernels, a high frequency kernel, lung kernel (LK) for the evaluation of the lung parenchyma and a standard reconstruction kernel, mediastinal kernel (MK) for the evaluation of the mediastinum. Recent studies have shown that the performance of DL-based tools for lung nodule detection and characterization is influenced by the acquisition and reconstruction settings, such as the choice of the reconstruction kernel [8,9]. However, there are currently no data regarding the influence of the reconstruction kernels on lung disease segmentation using data-driven methods. We hypothesized that combining the reconstruction kernels could improve model performances by allowing the model to get rid of the variability related to the reconstruction parameters.

The purpose of this study was to determine whether a single reconstruction kernel or two reconstruction kernels should be used

Abbreviations: COVID-19, Coronavirus disease 2019; CT, Computed tomography; DL, Deep learning; DSC, Dice similarity coefficient; ILD, Interstitial lung disease; IQR, interquartile range; LK, Lung kernel; MK, Mediastinal kernel

* Corresponding author.

E-mail address: guillaume.chassagnon@aphp.fr (G. Chassagnon).

<https://doi.org/10.1016/j.diii.2021.10.001>

2211-5684/© 2021 Société française de radiologie. Published by Elsevier Masson SAS. All rights reserved.

Please cite this article as: T.-N. Hoang-Thi, M. Vakalopoulou, S. Christodoulidis et al., Deep learning for lung disease segmentation on CT: Which reconstruction kernel should be used?, Diagnostic and Interventional Imaging (2021), <https://doi.org/10.1016/j.diii.2021.10.001>

when training DL for segmentation of diffuse lung disease on chest CT. For this, a DL architecture [10] was evaluated on two different CT datasets of diffuse lung disease

2. Material and methods

2.1. Image datasets and annotation transfer

This retrospective multi-center study was performed in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

Data from two independent and previously published datasets were used to assess the impact of reconstruction kernels on DL-based segmentation. One dataset has been created to develop segmentation models for COVID-19 pneumonia (COVID dataset) [7] while the other was created to develop segmentation models ILD in systemic sclerosis (ILD dataset) [5]. Both datasets originally only included images reconstructed with a LK (LK images). For the purpose of this study, the corresponding images reconstructed with a MK (MK images) were retrieved. Annotations produced by the annotation of original LK images were projected on the MK images.

When the LK and MK images had the same slice thickness and reconstruction field of view, the mask of annotation was directly transferred to MK images. When the slice thickness and/or reconstruction field of view were different, the metadata were used to perform the required rescaling and cropping, so that the two acquired lung volumes could correspond.

2.2. COVID-19 dataset

This dataset comprised 180 unenhanced chest CT examinations from COVID-19 patients acquired at six university hospitals using four different CT models from three manufacturers. All images had been reconstructed using iterative reconstruction and a slice thickness ranging from 0.625- to 1 mm. Images had been annotated by 15 radiologists with one to seven years of experience in chest imaging. On each CT slice, all the COVID-19 related CT abnormalities (ground glass opacities, band consolidations, and reticulations) [11–13] had been segmented as a single

Fifty patients from three centers composed the training and validation dataset. All images of these 50 CT examinations (321,360 slices) had been annotated slice-by-slice. The 130 patients from the remaining three centers composed the test dataset. In this test dataset, only 20 slices per CT examination (2600 slices), equally spaced from the superior border of aortic arch to the lowest diaphragmatic dome had been annotated and each slide had been annotated by two different radiologists.

2.3. ILD dataset

This dataset comprised 31 unenhanced chest CT examinations from systemic sclerosis patients acquired at one university hospital using four different CT scanners from two manufacturers. Six patients out of the 37 patients from the original cohort were excluded because of the unavailability of MK images. Lung CT images were reconstructed with a slice thickness of 0.625- to 1.5 mm, using filter back projection or iterative reconstruction algorithms. ILD segmentation had been performed by four radiologists with one to four years of experience in chest imaging. All the ILD-related anomalies (ground glass opacities, reticulations, traction bronchiectasis and honeycombing) were segmented as a single class (ILD).

Chest CT examinations from 11 patients composed the training and validation dataset. All images of this dataset had been annotated by one radiologist (3884 slices). Chest CT examinations from the remaining 20 patients composed the test dataset. For each CT examination, 20 CT slices equally spaced from the lung apices to the right

diaphragmatic dome (20 slices) had been annotated by three different radiologists (400 images in total).

2.5. DL-based segmentation methods

For benchmarking the performance of the DL architecture, a standard U-Net architecture [10] consisting of five blocks with a down-sampling operation applied every two consequent Conv2D-BN-ReLU layers was used. Additionally, five decoding blocks were used for the decoding path, where at each block a transpose convolution was performed to up-sample the input. Skip connections were also employed between the encoding and decoding paths. Each two-dimensional slice extracted from the CT examination was given as an input to the network together with its annotation of the abnormal regions. Weighted cross entropy has been used as loss function for the training of the model with the weight of the pathological regions being inverse proportional to the appearance of each class, corresponding to 60 for the COVID dataset and to 65 for the ILD dataset. Adam optimizer was used with a learning rate of 0.0001 for the optimization. For fair comparison, the same number of iterations was kept for the training of all models.

Three different trainings, resulting in three different models were compared for each disease: training on LK images only, MK images only or LK+MK images. Thus, the same CT slices were used for each of the trainings but these CT slices had been reconstructed with only one kernel (LK or MK) or the two kernels (LK and MK).

2.6. Statistical analysis

Statistical analysis was performed using 'R' software (version 3.6.3, R Foundation, Vienna, Austria). The quality of the segmentations was assessed using the Dice similarity coefficient (DSC) [14]. As manual segmentations from several radiologists were available for the two test datasets, for each CT examination the average DSC between the model and the observers was calculated. Comparison between models was performed using the Wilcoxon signed-rank test. To compensate for multiple comparisons, a Bonferroni correction was applied. A P value < 0.05 was considered to indicate significant difference.

3. Results

3.1. Impact of the reconstruction kernel to build models

To evaluate the impact of the reconstruction kernel on models' generalizability, we first compared the performance of the segmentation models in test datasets combining LK and MK images from the same CT slices (Table 1, Fig. 1). For COVID-19 pneumonia segmentation, the model trained on LK images performed better than the one trained on MK images (median DSC = 0.61 [interquartile range (IQR): 0.52, 0.70] vs. 0.57 [IQR: 0.50, 0.67], respectively; $P < 0.001$). Conversely, for ILD segmentation, the model trained on MK images performed better than the one trained on LK images (median DSC = 0.65 [IQR: 0.58, 0.70] vs. 0.53 [IQR: 0.49, 0.63], respectively; $P = 0.007$). For both COVID-19 and ILD segmentation, training on a dataset combining the two reconstruction kernels (median DSC = 0.67 [IQR: 0.59, 0.72] for COVID-19 and 0.69 [IQR: 0.63, 0.73] for ILD) significantly improved the performance compared to models trained on LK images alone ($P < 0.001$ for COVID-19 and ILD segmentations) or MK images alone ($P < 0.001$ for COVID-19 and ILD segmentations).

When applying the model to a test dataset consisting solely of LK or MK images, models trained on both kernels performed as well as or better than models trained only on one of the two kernels. For COVID-19 pneumonia segmentation, the model trained on both kernels performed better than the one trained on LK images when applied on LK images only (median DSC = 0.67 [IQR: 0.59, 0.73] vs.

Table 1

Dice similarity scores for automated and manual segmentations depending on the reconstruction kernel used in the training and test datasets. The presented Dice similarity scores are the ones measured in the test cohorts.

	Applied on lung kernel		Applied on mediastinal kernel		Applied on both kernels	
	COVID-19	ILD	COVID-19	ILD	COVID-19	ILD
Trained on lung kernel	0.62 (0.54, 0.69) [0.27–0.90]	0.62 (0.56, 0.69) [0.38–0.87]	0.60 (0.50, 0.70) [0.24–0.90]	0.50 (0.43, 0.57) [0.31–0.83]	0.61 (0.52, 0.70) [0.26–0.90]	0.53 (0.49, 0.63) [0.37–0.85]
Trained on mediastinal kernel	0.54 (0.47, 0.63) [0.24–0.89]	0.63 (0.53, 0.70) [0.27–0.80]	0.62 (0.53, 0.68) [0.24–0.89]	0.69 (0.61, 0.73) [0.34–0.88]	0.57 (0.50, 0.67) [0.24–0.89]	0.65 (0.58, 0.70) [0.30–0.84]
Trained on both kernels	0.67 (0.59, 0.73) [0.30–0.92]	0.69 (0.63, 0.73) [0.32–0.88]	0.67 (0.60, 0.74) [0.29–0.92]	0.68 (0.62, 0.72) [0.29–0.88]	0.67 (0.59, 0.72) [0.30–0.92]	0.69 (0.63, 0.73) [0.30–0.88]

Note: Dice similarity scores are presented as median with interquartile ranges in parentheses and ranges in brackets; ILD = Interstitial lung disease.

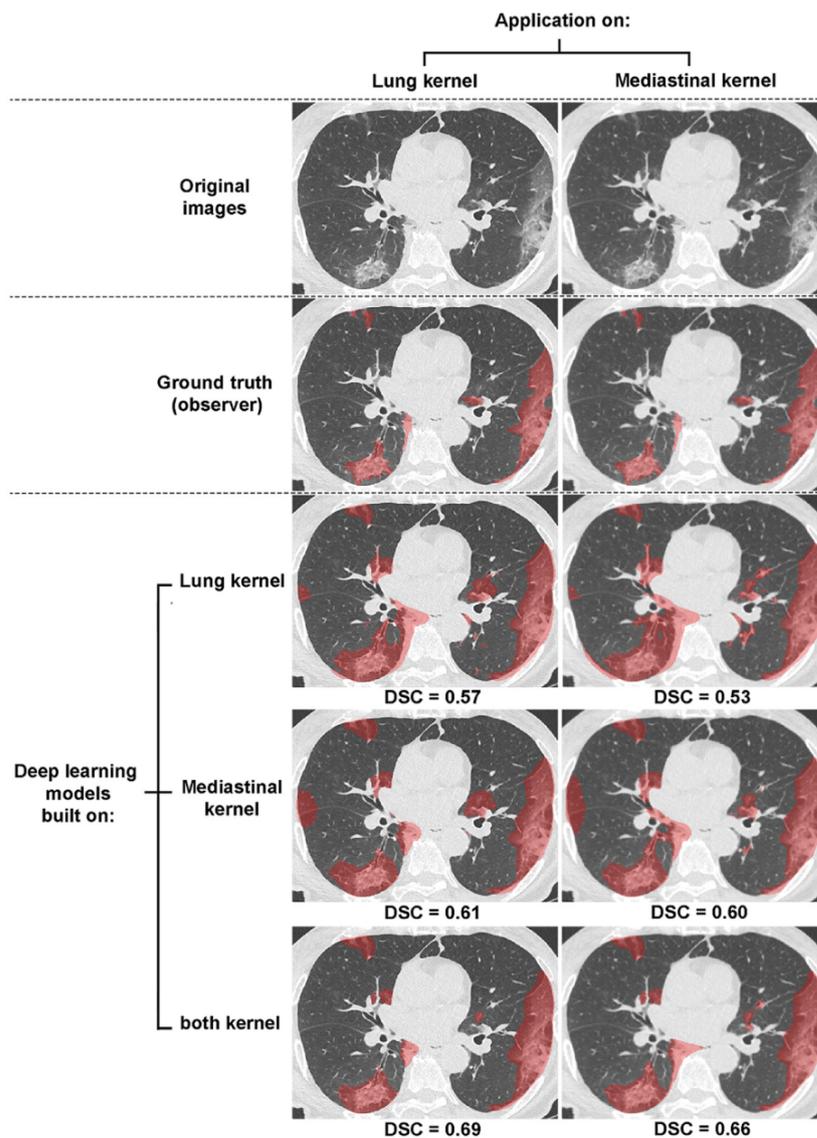


Fig. 1. Results of the COVID-19 segmentation models in a 42-year-old man with COVID-19 pneumonia. The deep learning model trained on the images reconstructed with the two kernels performs better than the models trained on the images reconstructed with only one of the kernels, whatever the kernel on which it is applied. The better performances are illustrated by the greater dice similarity score (DSC). Models trained on only one kernel are responsible for greater oversegmentation.

0.62 [IQR: 0.54, 0.69]; $P < 0.001$) and better than the model trained on MK images when applied on MK images only (median DSC = 0.67 [IQR: 0.60, 0.74] vs. 0.62 [IQR: 0.53, 0.68]; $P < 0.001$). For the ILD segmentation, the model trained on both kernels performed better than the one trained on LK images when applied on LK images (median DSC = 0.69 [IQR: 0.63, 0.73] vs. 0.62 [IQR: 0.56, 0.69]; $P = 0.006$) and as well as the model trained on MK images when applied on MK images (median DSC = 0.68 [IQR: 0.62, 0.72] vs. 0.69 [IQR: 0.61, 0.73]; $P > 0.99$).

3.2. Selection of the reconstruction kernel for application of trained models

To assess the impact of reconstruction kernel selection for model application, the performances of the models depending on the reconstruction kernel used in the test datasets were compared (Table 1, Fig. 1). Models trained only on LK images performed better on LK images than on MK images (median DSC = 0.62 [IQR: 0.54, 0.69] vs. 0.60 [IQR: 0.50, 0.70]; $P < 0.001$ for COVID-19 segmentation and 0.62 [IQR: 0.56, 0.69] vs. 0.50 [IQR: 0.43, 0.57]; $P < 0.001$ for ILD segmentation). Similarly, models trained only on MK images performed better on MK images (median DSC = 0.62 [IQR: 0.53, 0.68] vs. 0.54 [IQR: 0.47, 0.63]; $P < 0.001$ for COVID-19 segmentation and 0.69 [IQR: 0.61, 0.73] vs. 0.63 [IQR: 0.53, 0.70]; $P < 0.001$ for ILD segmentation).

When the segmentation models had been trained on a combination of the two kernels, the differences in performance according to the reconstruction kernel of the test dataset were smaller with a slight advantage to the LK. For COVID-19 segmentation, DSC was not significantly different whether applied to LK images (median DSC = 0.67 [IQR: 0.59, 0.73]) or MK images (median DSC = 0.67 [IQR: 0.60, 0.74]) ($P > 0.99$). On the opposite, the DSC was slightly greater when the ILD segmentation model was applied to LK images (median DSC = 0.69 [IQR: 0.63, 0.73] vs. 0.68 [IQR: 0.62, 0.72]; $P < 0.001$).

4. Discussion

To our knowledge, no studies have assessed the influence of the reconstruction kernel on image segmentation using DL. Using two different datasets and a popular DL architecture, we demonstrated that the choice of the reconstruction kernel significantly impacts the performance of the models and that training algorithms on a combination of lung and mediastinal kernels significantly improves the performance.

Regardless of the lung disease, our results show that the models performed better when reconstruction kernels were the same in the training and the test datasets. This demonstrates that while being more efficient than traditional machine learning methods, DL is still sensitive to image characteristics such as the reconstruction kernel. Decreasing the sharpness of the reconstruction kernel reduces the amount of noise in the image, which is already known to have a significant impact on emphysema quantification [15,16]. For systemic sclerosis-related ILD, Kim et al. demonstrated that by using image denoising methods, they increased the performance of their ILD segmentation model based on texture analysis with classical machine learning methods [17]. Although images reconstructed with a MK are less noisy, our study does not demonstrate the superiority of this kernel over LK in the setting of lung disease segmentation using DL. LK is usually the preferred kernel for visual analysis of lung parenchyma.

For pulmonary nodules, a few studies have also reported an impact of the reconstruction kernel on the performance of DL-based computer-aided detection tools and showed conflicting results regarding the best kernel to be used [8,9]. The selection of the reconstruction kernel is also known to impact the results of tools based on radiomics and texture analysis approaches, such as for pulmonary nodule characterization [18–22]. Recently, Choe et al. have suggested that the use of chest CT image conversion using CNN can reduce the

effect of the reconstruction kernels on radiomics parameters for lung nodule assessment [23].

An important result of our study is that we demonstrated that combining images reconstructed from the same CT slice using LK and MK improves learning performances. This combination is easy to perform as all chest CT acquisitions are usually reconstructed with both kernels and both sets of images are routinely stored. Moreover, the annotation time is not increased because manual segmentations performed on images reconstructed with one kernel can be easily applied on those reconstructed with the other kernel and has more influence when the dataset is smaller. Thus, the combination of the two kernels allows data augmentation without having to increase the number of patients or the annotation time, which are two factors limiting the size of medical imaging datasets. Combining the two kernels may allow the algorithm to free itself from the variability due to image noise. Interestingly, when the algorithm was trained on a dataset combining images reconstructed with both kernels, the choice of the kernel for test had almost no impact on the results and depended on the lung disease.

This study has several limitations. First, all the manual segmentations we used were performed on LK images. However, the fact that models trained on MK images performed better on the test dataset containing images also reconstructed with a MK suggests this had no significant impact. Second, only one DL architecture was used. Yet, the U-Net architecture is the most popular for segmentation tasks in medical imaging, while the small size of the ILD dataset did not allow the training of 3D deep learning architectures.

In conclusion, our work highlights the importance of the reconstruction kernel when working on a DL-based method to segment diffuse lung diseases. We have shown that learning on a combination of LK and MK images from the same CT slice should be done as it improves the performance and generalizability of the models. Importantly, this can be done without increasing the annotation time. The combination of two different kernels could also improve segmentation tasks for extra-thoracic applications, such as musculoskeletal imaging where images are commonly reconstructed with standard and bone reconstruction algorithms.

Credit author statement

T.-N. H.-T.: Study design; data collection; analysis and interpretation of the results; draft manuscript edition; manuscript final version approval

M.K.: Study design; data analysis; interpretation of the results manuscript final version approval

S.C.: Study design; data analysis; interpretation of the results; manuscript final version approval

N.P.: Study design; supervision; manuscript final version approval

M.-P. R.: Study design; supervision; manuscript final version approval

G.C.: Study design; data collection; analysis and interpretation of the results; supervision; manuscript final version approval

Declaration of Competing Interest

The authors have no conflict of interest in relation to this study and no relationships with any industry related to this work.

Human rights

This work has been performed in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

Funding

This work did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

All authors attest that they meet the current International Committee of Medical Journal Editors (ICMJE) criteria for Authorship.

References

- [1] Chassagnon G, Vakalopoulou M, Paragios N, Revel M-P. Artificial intelligence applications for thoracic imaging. *Eur J Radiol* 2020;123:108774.
- [2] Nakaura T, Higaki T, Awai K, Ikeda O, Yamashita Y. A primer for understanding radiology articles about machine learning and deep learning. *Diagn Interv Imaging* 2020;101:765–70.
- [3] Blanc D, Racine V, Khalil A, Deloche M, Broyelle J-A, Hammouamri I, et al. Artificial intelligence solution to classify pulmonary nodules on CT. *Diagn Interv Imaging* 2020;101:803–10.
- [4] Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mouggiakakou S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE Trans Med Imaging* 2016;35:1207–16.
- [5] Chassagnon G, Vakalopoulou M, Régent A, Zacharaki EI, Aviram G, Martin C, et al. Deep learning–based approach for automated assessment of interstitial lung disease in systemic sclerosis on CT images. *Radiol Artif Intell* 2020;2:e190006.
- [6] Li K, Fang Y, Li W, Pan C, Qin P, Zhong Y, et al. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). *Eur Radiol* 2020;30:4407–16.
- [7] Chassagnon G, Vakalopoulou M, Battistella E, Christodoulidis S, Hoang-Thi T-N, Dangeard S, et al. AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia. *Med Image Anal* 2021;67:101860.
- [8] Fu B, Wang G, Wu M, Li W, Zheng Y, Chu Z, et al. Influence of CT effective dose and convolution kernel on the detection of pulmonary nodules in different artificial intelligence software systems: a phantom study. *Eur J Radiol* 2020;126:108928.
- [9] Blazis SP, Dickerscheid DBM, Linsen PVM, Martins Jarnalo CO. Effect of CT reconstruction settings on the performance of a deep learning based lung nodule CAD system. *Eur J Radiol* 2021;136:109526.
- [10] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation editors. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Med image comput assist interv. MICCAI*; 2015. p. 234–41.
- [11] Chassagnon G, Regard L, Soyer P, Revel M-P. COVID-19 after 18 months: where do we stand? *Diagn Interv Imaging* 2021;102:491–2.
- [12] Li J, Long X, Wang X, Fang F, Lv X, Zhang D, et al. Radiology indispensable for tracking COVID-19. *Diagn Interv Imaging* 2021;102:69–75.
- [13] Jalaber C, Lapotre T, Morcet-Delattre T, Ribet F, Jouneau S, Lederlin M. Chest CT in COVID-19 pneumonia: a review of current knowledge. *Diagn Interv Imaging* 2020;101:431–7.
- [14] Dice LR. Measures of the amount of ecologic association between species. *Ecology* 1945;26:297–302.
- [15] Boedeker KL, McNitt-Gray MF, Rogers SR, Truong DA, Brown MS, Gjertson DW, et al. Emphysema: effect of reconstruction algorithm on CT imaging measures. *Radiology* 2004;232:295–301.
- [16] Gierada DS, Bierhals AJ, Choong CK, Bartel ST, Ritter JH, Das NA, et al. Effects of CT section thickness and reconstruction kernel on emphysema quantification relationship to the magnitude of the CT emphysema index. *Acad Radiol* 2010;17:146–56.
- [17] Kim HJ, Li G, Gjertson D, Elashoff R, Shah SK, Ochs R, et al. Classification of parenchymal abnormality in scleroderma lung using a novel approach to denoise images collected via a multicenter study. *Acad Radiol* 2008;15:1004–16.
- [18] Zhao W, Zhang W, Sun Y, Ye Y, Yang J, Chen W, et al. Convolution kernel and iterative reconstruction affect the diagnostic performance of radiomics and deep learning in lung adenocarcinoma pathological subtypes. *Thorac Cancer* 2019;10:1893–903.
- [19] He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep* 2016;6:34921.
- [20] Shafiq-ul-Hassan M, Zhang GG, Hunt DC, Latifi K, Ullah G, Gillies RJ, et al. Accounting for reconstruction kernel-induced variability in CT radiomic features using noise power spectra. *J Med Imaging* 2017;5:1.
- [21] Chen A, Karwoski RA, Gierada DS, Bartholmai BJ, Koo CW. Quantitative CT analysis of diffuse lung disease. *Radiographics* 2020;40:28–43.
- [22] Ligerio M, Jordi-Ollero O, Bernatowicz K, Garcia-Ruiz A, Delgado-Muñoz E, Leiva D, et al. Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis. *Eur Radiol* 2021;31:1460–70.
- [23] Choe J, Lee SM, Do K-H, Lee G, Lee J-G, Lee SM, et al. Deep learning–based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *Radiology* 2019;292:365–73.

7 DISCUSSION GENERALE

7.1 PRINCIPALES PROJETS DE THESES

L'imagerie médicale joue un rôle central dans le diagnostic des maladies en médecine et les progrès des techniques d'imagerie au cours des dernières décennies ont eu un impact important sur les capacités de diagnostic. Cependant, toutes images sont de type de données non structuré et, bien que les ordinateurs puissent facilement gérer les données d'image brutes, ils ne peuvent pas facilement reconnaître le type d'image, les annotations sur l'image ou tirer des conclusions à partir des données (205). La conversion d'images médicales numériques en données exploitables de haute dimension a permis un processus connu sous le nom de « radiomique ». Ce dernier est motivé par le concept que les images biomédicales contiennent des informations qui reflètent la physiopathologie sous-jacente et que ces relations peuvent être révélées via des analyses quantitatives d'images. Ainsi, les défis de l'informatique d'imagerie sont d'acquérir des données d'imagerie avec une fidélité haute qui représente précisément l'image et de traiter ces données non structurées en un contenu interprétable. Dans ce tableau, ces travaux de thèse mettent l'accent sur approches automatisées basées sur l'intelligence artificielle dans analyses quantitatives des lésions structurelles thoraciques tomodensitométriques dans diverses affections touchant les voies aériennes et le parenchyme pulmonaire en relation avec les troubles fonctionnels respiratoires : Pour la DCP, la caractéristique radiomique de premier ordre (histogramme) autorise la quantification automatique des changements en imagerie et l'établissement un score objective pour l'évaluation de la gravité de la maladie ; Pour la DCGP, l'analyse préliminaire de la combinaison des caractéristiques de densité en histogramme et de texture de la radiomique des scanners inspiratoires s'avère contributive dans la prédiction de la survenue des manifestations fonctionnelles respiratoires depuis un délai de 6 mois en avant malgré les résultats préliminaires ne sont pas encore statistiquement significatifs ; Pour la ScS, l'algorithme de « deep-learning » AtlasNet permet la quantification des lésions du parenchyme pulmonaire à viser la prédiction de mortalité à long et court terme ; Nous rapportons ainsi l'influence du choix de noyau de reconstruction en TDM sur l'entraînement des modèles d'apprentissage automatique dans le processus de segmentation des zones lésionnelles des maladies pulmonaires interstitielles de la ScS et COVID-19, ce qui ouvre désormais la possibilité d'augmenter la performance des modèles en combinaisons de deux filtres de reconstruction.

La conversion de l'image – donnée non-structurée, vient dans la forme d'un tableau des pixels à valeurs différentes – en donnée quantifiée interprétable, le pourcentage de tissu lésionnel, à partir de l'utilisation des caractéristiques radiomiques de premier ordre (histogramme) dans le DCP, la segmentation par l'algorithme d'AtlasNet dans la ScS et la combinaison des caractéristiques texturales de la radiomique pour la tâche de prédiction de la survenue de DCGP ont été réalisée de façon entièrement automatique en minimisant l'influence de de l'analyse supervisée humain qui est limitée par une subjectivité intégrée. Nous pensons que l'apprentissage automatique aura beaucoup plus de succès dans un avenir proche, car il nécessite très peu l'interprétation humaine, il peut donc facilement tirer parti de l'augmentation de la quantité de calculs et de données disponibles. Les nouveaux algorithmes et architectures d'apprentissage en cours de développement pour les réseaux de neurones profonds ne feront qu'accélérer ces progrès (149).

Le choix de paramètres de résultats tel que la mortalité dans la sclérodermie, la dégradation non-récupération des indices fonctionnels respiratoires sont clairement définies et robustes en comparaison à l'évaluation statique de la gravité de la maladie, ce qui va en bonne direction de

référence dans l'analyse des données longitudinales des études actuelles dans le domaine de deep-learning (9).

7.2 APPLICATION DE L'INTELLIGENCE ARTIFICIELLE A BASE DES OBSERVATIONS EN TDM

La recherche basée sur les données n'est pas entièrement sans hypothèses(215). Les approches fondées sur les données partent souvent d'une hypothèse large et collectent ou génèrent de grands volumes de données pertinentes pour cette hypothèse. Ces données sont ensuite explorées pour générer des hypothèses plus spécifiques et mécanistes par induction éliminatoire ou raisonnement abductif(203).

Ainsi, dans les travaux de DCP, nous partons des observations que la plupart des changements structurels pulmonaires de la DCP, en particulier l'épaississement de la paroi bronchique, les impactions du mucus, les consolidations et l'atélectasie, sont susceptibles d'augmenter l'atténuation pulmonaire et de modifier les paramètres de l'histogramme de densité, qui peuvent être extraites des images de TDM. Partant de l'hypothèse que la quantification des tissus lésionnels par seuillage en histogramme de densité des images pourrait évaluer la gravité de l'atteinte pulmonaire chez les patients de DCP. Plusieurs seuils de densité et ainsi les autres caractéristiques extraites de l'histogramme ont été implémenté pour trouver biomarqueur qui corresponde le mieux avec les indices fonctionnels respiratoires. Nous avons constaté que les seuils adaptés basés sur les paramètres de l'histogramme étaient mieux corrélés que les seuils fixes avec les paramètres spirométriques. Parmi les différents seuils testés, la densité pulmonaire moyenne + 1 écart type (DPM + 1 ET) a donné le meilleur résultat (190). Ce seuil est facilement accessible, car la plupart des logiciels de segmentation disponibles dans le commerce fournissent à la fois des valeurs DPM et ET.

Dans les travaux de DCGP, à base des observations que le processus pathologique de la maladie comprend à la fois la bronchiolite oblitérante et la fibrose de l'interstitium en différents niveaux de gravités formant deux sous-types de BO et SRA, qui induisent les changements à la fois des structures de haut et de bas niveau de densité. L'hypoperfusion en aval des bronchioles obstructives diminue l'atténuation relative du segment pulmonaire correspond par rapport à des segments voisins. Tandis que la fibrose de l'interstitium augmente la partie de haute atténuation du parenchyme par des patterns de réticulations ou les rayons miel, ou même l'épaississement de la plèvre viscérale ou la paroi bronchique... Devant ce tableau, nous avons réalisé des investigations sur l'ensemble des caractéristiques morphologiques. Les paramètres radiomiques de premier ordre, de deuxième ordre et de forme ont été extraits automatiquement depuis les données TDM de différents délais avant la manifestation en spirométrie de l'affection. Les trois variables retenues parmi les 107 paramètres analysés comprennent un paramètre de statistique de premier ordre, la valeur minimum d'histogramme, et les deux paramètres statistiques de deuxième ordre de texture, l'accentuation de petite zone à niveau de gris élevé et le non-uniformité des niveaux de gris normalisés reflètent le changement en matière de voxel à basse et haute intensité ainsi que de l'uniformité des niveaux de gris des TDM dans le cadre de la pathologie DCGP marquée par la perfusion en mosaïque et la fibrose interstitielle. Concernant les performances de notre méthode pour la prédiction de la survenue de DCGP à 6 mois en avant, notre classificateur d'ensemble rapporte des résultats préliminaires encourageants avec une exactitude pondérée de 79%, une précision pondérée de 78%, une sensibilité pondérée de 88% et une spécificité pondérée de 71% (Tableau 4-8) dans le groupe de test. Néanmoins, malgré de bons résultats en test, ceux-ci ne pourraient être statistiquement significatifs puisqu'ils sont bien supérieurs aux résultats en entraînement (Tableau 4-8). On peut noter également un effet de sur-apprentissage étant donné

que les résultats d'entraînement sont bien plus élevés qu'en validation (Tableau 4-7). Ces effets pourraient être dus à la faible taille du jeu de test et de la population étudiée. Cependant, les résultats de nos travaux suggèrent la possibilité d'améliorer la performance de prédiction du modèle par élargissement de la population au futur ou en collaboration multicentrique des établissements disposant le greffon pulmonaire.

Dans les travaux de ScS, le point de départ est l'hypothèse qu'un outil de segmentation basé sur le deep-learning, formé à partir d'une base de données d'images TDM contiguës entièrement annoté, permette la quantification automatisée de l'étendue de lésions interstitielles chez les patients atteints de la sclérose systémique (ScS-ILD). La nécessité de garantir la plausibilité anatomique des résultats de segmentation de la maladie est à l'origine de l'architecture AtlasNet, qui est augmenté en double auto-décodeur basée sur l'architecture SegNet, avec l'introduction d'interdépendances spatiales entre les poumons et les classes de maladies, prédites par deux parties différentes du décodeur (Figure 2-5). Les classes de maladie interstitielle englobent les opacités en verre dépoli, les réticulations, les bronchectasies par traction et du rayon-miel, en une seule catégorie. Pour résultats, dans l'ensemble, AtlasNet a obtenu de meilleurs résultats que U-Net, affichant une amélioration absolue de 1% des DSC pour la segmentation pulmonaire et de plus de 3% d'amélioration absolue des DSC pour la segmentation de la maladie. La corrélation avec les EFRs obtenue avec l'AtlasNet était dans la même plage lors de l'utilisation de l'architecture U-Net ($R = -0,75$ et $-0,76$ pour DLCO, $R = -0,69$ et $-0,70$ pour CPT, $R = -0,61$ à $-0,62$ pour CVF et $R = -0,53$ à $-0,54$ pour DLCO/VA ; $p < 0,001$ pour tous) (239). De ces résultats encourageants, nous avons implémenté dans ce travail de thèse une investigation sur la performance de ce biomarqueur morphologique en combinaison avec d'autres facteurs de risque pour la construction du modèle de prédiction de mortalité chez cette population de ScS. Nous avons démontré une corrélation linéaire entre l'étendue de l'ILD sur le scanner et la mortalité. Cette association pourrait inciter les cliniciens à renforcer la surveillance clinique des stades précoces de la SSc-ILD et à discuter des mesures thérapeutiques précoces, au lieu d'attendre une aggravation significative (334). La mesure automatisée de l'ILD sur la tomodensitométrie nous a permis d'identifier un groupe de patients présentant un risque de mortalité à court terme plus élevé. Ce biomarqueur morphologique, combiné avec des indicateurs péjoratifs bien connus dans la sclérose systémique tels que l'âge avancé, le sexe masculin, la sclérodermie cutanée diffuse, une CVF altérée, la présence d'auto-anticorps anti-Scl-70 ou anti-ARN-polymérase 3, pourrait aider les cliniciens à déterminer quand un traitement intensif doit être envisagé (326).

Alors, dans l'ensemble de nos projets, les hypothèses existent encore, à moins qu'ils sont « plus larges » que jamais. Au lieu de partir au début des travaux de façon traditionnelle par un signe ou un pattern spécifique en TDM des maladies, à l'aide des outils d'intelligences artificielles, nous avons pu mettre en question des problèmes plus large en matière données en disposition, et ainsi plus proche des besoins cliniques. Est-ce que les changements morphologiques identifiés au scanner sont corrélés aux indices d'explorations fonctionnelles respiratoires ? Est-ce que la quantification à base de segmentation automatique de parenchyme pulmonaire lésionnel pourrait-elle prédire la mortalité à long et court terme chez les ScS ? Les informations extraites des TDM pourraient-elles prédire la survenue de DCGP avant le début des manifestations fonctionnelles respiratoires ? C'était grâce aux progrès des outils d'aide au diagnostic, du deep-learning, des caractéristiques radiomiques extraites automatiquement que les tâches compliquées telles de segmentations des zones lésionnelles, de retrouver la différence minimale à l'œil humain en matière de densité et de texture étaient effectuées à l'ordre de quelques secondes lorsque le modèle est bien entraîné, et surtout avec des performances accentuées au fur et à mesure de l'accumulation

des données. Ces résultats ouvrent ainsi la possibilité d'augmenter la qualité des services médicaux et d'espérance de vie pour les maladies graves en phase terminales.

7.3 LA RESTRICTION DES DONNEES DES MALADIES RARE ET LES PROMESSES DU DEEP-LEARNING

Les données dans la science axée sur les données sont généralement de grandes dimensions - impliquant de nombreux paramètres- et accumulent un grand nombre d'observations (212). La collection de données est une étape critique dans laquelle le scientifique doit être en mesure de garantir que la quantité appropriée de données qui saisit les variables pertinentes susceptibles d'affecter le ou les résultats d'intérêt est incluse [211]. Il est également essentiel que les données soient collectées de manière à minimiser les sources de biais et les facteurs de confusion potentiels(215). Dans le projet de prédiction de la survenue de DCGPs, l'appariement en âge, sexe et temps de suivi post opération de LTx nous a coûté la diminution remarquable de patients éligibles pour l'entrée en modèle d'entraînement, en plus de ce fait, le TDM n'entre pas encore en surveillance de routine par préoccupation d'accumulation de radiation à long terme, seulement 41 TDM chez les patients acquis de DCGP et 32 TDM chez les patients sains en contrôle. D'ailleurs, il n'y existe pas encore un protocole d'acquisition standardisé pour l'exploration TDM chez cette population spécifique dans nos établissements, d'importants nombres de scanners n'ont pas pu entrer dans les données d'entraînement par manque de coupe fine, l'injection de produit de contraste... La restriction de taille de population d'entraînement et de test demeure une des principales causes du phénomène de sur-apprentissage du modèle entraîné. Etant donné que la population étudiée n'est pas suffisante pour la partition en trois ensembles (d'entraînement, validation et test), la méthode de validation croisée (cross validation) a été réalisée par les statisticiens afin d'évaluer les performances de l'algorithme sans perdre en quantité de données d'entraînement. Nous espérons alors que l'accumulation au fil du temps des données de greffons pulmonaires, la combinaison des données cliniques et laboratoires, la coopération multicentrique des établissements disposition de transplantation pulmonaire pourrait améliorer la performance du modèle « omics ».

7.4 CHOIX DU NOYAU DE RECONSTRUCTION POUR D'DEEP-LEARNING DANS LE CADRE DES LESIONS PARENCHYMATEUSES PULMONAIRES

Le potentiel de la radiomique pour améliorer les décisions cliniques est incontestable (335) et le domaine évolue rapidement. Le principal défi est la collecte et l'intégration optimales de diverses sources de données multimodales d'une manière quantitative qui fournit des prédictions cliniques sans ambiguïté avec précision et robustesse [341]. La particularité de la TDM thoracique par laquelle les images sont généralement reconstruites à l'aide de deux filtres de reconstruction différents, pulmonaire et médiastinal. Bien qu'il existe déjà des évidences d'influence du noyau de reconstruction et même de radiation dose sur la performance des modèles de segmentation des nodules ou masses pulmonaires (337), les preuves restent manquées pour les maladies pulmonaires diffuses. À base de ces particularités, notre équipe a sondé l'effet de ces deux filtres en tant que données entrées et testées sur la performance du modèle d'entraînement classique d'U-net, et a trouvé l'application du même noyau pour les données testées et entrées donnent la plupart de temps la meilleure performance qu'en cas de noyau différent des sous-jeux de données. À noter d'ailleurs, la combinaison de deux filtres pour l'entraînement qui ne requiert pas de nouveau d'acquisition d'image ni radiation ajoutée, augmente de façon pertinence de la performance du modèle peu importe le noyau appliqué en jeu de test. La combinaison des deux filtres peut

permettre à l'algorithme de s'affranchir de la variabilité due au bruit d'image, améliore ainsi les performances des modèles et la possibilité de les généraliser.

7.5 LIMITATIONS ET PERSPECTIVES

Nos études présentent plusieurs limites. Pour le DCP, la DCGP et ScS, les investigations étaient faites de façon rétrospective, les paramètres d'acquisition de TDM n'étaient pas standardisés. Cela peut avoir influencé les seuils de densité, les autres caractéristiques radiomiques, et l'algorithme de deep-learning. Or plusieurs études ont montré l'importance de la standardisation des paramètres d'acquisition et de reconstruction des images pour la création de signatures radiomiques et des algorithmes de deep-learning (221,226,232,338). Deuxièmement, nos cohortes n'étaient pas assez grandes pour analyser toutes les variables potentielles pour la recherche des biomarqueurs. Ainsi, il n'a pas été possible de diviser notre population de DCP en une cohorte de développement et de validation. Un écart de performance entre la performance des cohortes de tests et de validations de la DCGP montre une certaine instabilité du modèle. Dans la ScS, nous avons dû restreindre notre modèle à 6 paramètres prédéfinis. Par conséquent, nos modèles de prédiction, nos scores d'évaluation doivent être validés dans une autre cohorte pour confirmer leurs résultats et leurs performances.

L'implémentation des études prospectives avec contrôle des délais de recueil des données morphologiques et fonctionnelles respiratoires, la standardisation des protocoles de scanners pourrait améliorer les performances du score développé dans la DCP et la DCGP. Dans la DCGP en particulier, l'exploitation des scanners en inspiration seuls au lieu de scanner biphasiques promet de réduction de radiation doses pour la population des receveurs du greffon, ceux qui subissent souvent des scanners répétés à longue durée. Néanmoins, les scanners en phase d'expiration pourraient peut-être contribuer à la performance des modèles du deep-learning. Ainsi, les résultats de nos travaux sur l'influence du noyau de reconstructions pourraient être appliqué dans de futures investigations pour améliorer la performance des algorithmes d'apprentissage avec des bases de données disponibles sans l'obligation de réalisation de nouveau scanner.

8 CONCLUSION GENERALE

Dans cette thèse, nous proposons une approche holistique pour l'analyse quantitative de l'imagerie thoracique tomodensitométrique dans diverses affections touchant les voies aériennes et le parenchyme pulmonaire en relation avec les troubles fonctionnels respiratoires basée sur l'intelligence artificielle. Notre approche est de complexité croissante guidée par des observations des variations morphologiques liées à ces affections.

Pour le premier projet relatif à la DCP, nous avons utilisé la quantification automatique par simple seuillage d'histogramme de densité pour estimer la proportion du tissu ayant une forte atténuation par rapport au volume total du poumon. Ce score est basé sur les observations initiales des principales anomalies de l'atteinte pulmonaire de la DCP telles que l'épaississement bronchique, l'arbre en bourgeon, l'impaction mucoïde, la consolidation péribronchique, le collapsus lobaire et l'atélectasie, l'épanchement pleural, le dépôt de calcium... Les résultats correspondent aussi bien aux indices spirométriques que le score visuel (de Bhalla). Cette méthode de notation quantitative, très peu chronophage (environ 2 min par cas), ne nécessitant qu'une formation simple (pour l'utilisation de l'outil 3D sur Myrian) et donnant des résultats indépendants de l'opérateur, semble très prometteuse pour la pratique clinique et la recherche.

Le deuxième projet relatif au DCGP où le processus pathologique est complexe. Cette entité comprend la bronchiolite oblitérante, la perfusion en mosaïque et la fibrose de l'interstitium en différents niveaux de gravités. Ces processus induisent les changements à la fois des structures de haut et de bas niveau d'atténuation. Nous avons adopté une approche entièrement automatique pour l'extraction de l'ensemble des paramètres de radiomique depuis les données TDM de différents délais avant les modifications spirométriques. Les trois variables retenues parmi les 107 paramètres analysés reflètent les changements de voxel à basse densité ainsi que l'hétérogénéité de la disposition spatiale des niveaux de gris des TDM dans le cadre de la perfusion en mosaïque en DCGP. Les meilleures performances de notre méthode pour prédire la survenue de DCGP sont issues des données de TDM à 6 mois avant les modifications spirométriques avec une précision d'équilibre de 79%, une précision pondérée de 78%, une sensibilité pondérée de 88% et une spécificité pondérée de 71% (Tableau 4-8) dans le groupe testé. Ces résultats n'étaient cependant pas statistiquement significatifs étant supérieurs aux résultats d'entraînement (Tableau 4-8). On peut également noter un effet de sur-apprentissage étant donné que les résultats d'entraînement sont bien plus élevés que pour la validation (Tableau 4-7). Ces effets pourraient être dus à la faible taille du jeu de test et de la population étudiée. Cependant, les résultats de nos travaux suggèrent la possibilité d'améliorer la performance de prédiction du modèle en augmentant le nombre de sujets étudiés par des collaborations multicentriques des établissements disposant le greffon pulmonaire comme l'importance du problème clinique en regard.

Dans la troisième étude portant sur la ScS Nous avons démontré une corrélation linéaire entre l'étendue de l'ILD sur le scanner par nouvel algorithme d'AP- AtlasNet - et la mortalité. Cette association pourrait inciter les cliniciens à renforcer la surveillance clinique des stades précoces de la SSc-ILD et à discuter des mesures thérapeutiques précoces, et de ne pas attendre l'aggravation clinique. La mesure automatisée de l'ILD sur la tomodensitométrie permet l'identification d'un groupe de patients présentant un risque de mortalité à court terme plus élevé. Ce biomarqueur morphologique, combiné avec des indicateurs de mauvais pronostic dans la sclérose systémique telle que l'âge avancé, le sexe masculin, la sclérodermie cutanée diffuse, une CVF altérée, la

présence d'auto-anticorps anti-Scl-70 ou anti-ARN-polymérase 3, pourrait guider le clinicien dans sa stratégie thérapeutique.

Actuellement, le suivi des maladies pulmonaires chroniques dont les manifestations minimales au fil du temps sont au-delà de la capacité des examens cliniques repose principalement sur des données physiologiques telles que les explorations fonctionnelles pulmonaires. Les mesures spirométriques ne reflètent que la fonction pulmonaire, mais pas nécessairement l'activité de la maladie. De plus, de petits changements dans les paramètres dérivés de la spirométrie chez un individu sont difficiles à interpréter en raison de la variabilité de la mesure. L'utilisation de l'intelligence artificielle dans l'évaluation automatique et quantitative des lésions morphologiques en imagerie existe depuis des décennies mais n'a réussi à atteindre des performances adéquates que récemment, grâce à l'acquisition de grandes quantités de données annotées disponibles, des capacités de parallélisation massives des GPUs et d'un certain nombre de bibliothèques de contribution publique qui ont beaucoup facilité leur utilisation. Il existe plusieurs limites à l'utilisation des paramètres d'imagerie dans les essais cliniques, telles que le manque de standardisation des protocoles d'acquisition et de la dose d'irradiation consécutive à la tomographie par ordinateur, la dépendance du fabricant de tomographie par ordinateur des caractéristiques de l'image et par l'influence de variables physiologiques. Les études décrites dans cette Thèse apportent des données supplémentaires en faveur de l'impact des kernels de reconstruction de TDM sur l'entraînement des architectures de deep-learning, qui offre la possibilité d'augmenter leur performance avec des données disponibles (scanner à reconstruction pulmonaire et médiastinale) et extractibles à partir des archives PACS de la plupart des établissements de soins publics. Nous sommes conscients du fait que la radiomique n'est qu'à un stade balbutiant et que beaucoup de travail reste à faire pour le développement de nouveaux biomarqueurs en imagerie dans l'évaluation des lésions de voies aériennes et du parenchyme pulmonaire.

9 REFERENCE BIBLIOGRAPHIQUE

1. The Global Impact of Respiratory Disease, 2nd ed.; In Sheffield, UK: European Respiratory Society,; p. 5–42.
2. Levin DL. Deep learning and the evaluation of pulmonary fibrosis. *Lancet Respir Med.* 2018/09/16 ed. 2018 Nov;6(11):803–5.
3. Kieu STH, Bade A, Hijazi MHA, Kolivand H. A Survey of Deep Learning for Lung Disease Detection on Medical Images: State-of-the-Art, Taxonomy, Issues and Future Directions. *J Imaging.* 2020;6:131.
4. Melendez J, Sánchez CI, Philipsen RHHM, Maduskar P, Dawson R, Theron G, et al. An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Sci Rep.* 2016 Apr 29;6:25265–25265.
5. Setio AAA, Traverso A, de Bel T, Berens MSN, Bogaard C van den, Cerello P, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Med Image Anal.* 2017 Dec;42:1–13.
6. Wells AU. Forced vital capacity as a primary end point in idiopathic pulmonary fibrosis treatment trials: making a silk purse from a sow’s ear. *Thorax.* 2013 Apr;68(4):309–10.
7. Tai Joon A, Wan S, Chan Kwon P, Hyoung Kyu Y. The better explanation of COPD: the clinical role of FVC grouping. *Eur Respir J.* 2019 Sep 28;54(suppl 63):PA2640.
8. Burki TK. The role of AI in diagnosing lung diseases. *Lancet Respir Med.* 2019 Dec 1;7(12):1015–6.
9. Stoel B. Use of artificial intelligence in imaging in rheumatology – current status and future perspectives. *RMD Open.* 2020 Jan 1;6(1):e001063.
10. Silva M, Milanese G, Seletti V, Ariani A, Sverzellati N. Pulmonary quantitative CT imaging in focal and diffuse disease: current research and clinical applications. *Br J Radiol.* 2018 Feb;91(1083):20170644.
11. Marthin JK, Petersen N, Skovgaard LT, Nielsen KG. Lung function in patients with primary ciliary dyskinesia: a cross-sectional and 3-decade longitudinal study. *Am J Respir Crit Care Med.* 2010 Jun 1;181(11):1262–8.
12. Santamaria F, Montella S, Tiddens HAWM, Guidi G, Casotti V, Maglione M, et al. Structural and functional lung disease in primary ciliary dyskinesia. *Chest.* 2008 Aug;134(2):351–7.
13. O’Callaghan C, Chilvers M, Hogg C, Bush A, Lucas J. Diagnosing primary ciliary dyskinesia. *Thorax.* 2007 Aug 1;62(8):656.
14. Lucas JS, Burgess A, Mitchison HM, Moya E, Williamson M, Hogg C. Diagnosis and management of primary ciliary dyskinesia. *Arch Dis Child.* 2014 Sep;99(9):850–6.
15. Barbato A, Frischer T, Kuehni CE, Snijders D, Azevedo I, Baktai G, et al. Primary ciliary dyskinesia: a consensus statement on diagnostic and treatment approaches in children. *Eur Respir J.* 2009 Dec 1;34(6):1264.

16. Noone PG, Leigh MW, Sannuti A, Minnix SL, Carson JL, Hazucha M, et al. Primary Ciliary Dyskinesia. *Am J Respir Crit Care Med*. 2004 Feb 15;169(4):459–67.
17. Cohen-Cyberknoh M, Simanovsky N, Hiller N, Hillel AG, Shoseyov D, Kerem E. Differences in disease expression between primary ciliary dyskinesia and cystic fibrosis with and without pancreatic insufficiency. *Chest*. 2014 Apr;145(4):738–44.
18. Horsley AR, Davies JC, Gray RD, Macleod KA, Donovan J, Aziz ZA, et al. Changes in physiological, functional and structural markers of cystic fibrosis lung disease with treatment of a pulmonary exacerbation. *Thorax*. 2013 Jun;68(6):532–9.
19. Kongstad T, Buchvald FF, Green K, Lindblad A, Robinson TE, Nielsen KG. Improved air trapping evaluation in chest computed tomography in children with cystic fibrosis using real-time spirometric monitoring and biofeedback. *J Cyst Fibros Off J Eur Cyst Fibros Soc*. 2013 Dec;12(6):559–66.
20. Irving SJ, Ives A, Davies G, Donovan J, Edey AJ, Gill SS, et al. Lung Clearance Index and High-Resolution Computed Tomography Scores in Primary Ciliary Dyskinesia. *Am J Respir Crit Care Med*. 2013 Jul 1;188(5):545–9.
21. Boon M, Vermeulen FL, Gysemans W, Proesmans M, Jorissen M, De Boeck K. Lung structure–function correlation in patients with primary ciliary dyskinesia. *Thorax*. 2015 Apr 1;70(4):339.
22. Magnin ML, Cros P, Beydon N, Mahloul M, Tamalet A, Escudier E, et al. Longitudinal lung function and structural changes in children with primary ciliary dyskinesia. *Pediatr Pulmonol*. 2012 Aug;47(8):816–25.
23. Brody AS. Scoring Systems for CT in Cystic Fibrosis: Who Cares? *Radiology*. 2004 May 1;231(2):296–8.
24. Ferraro V, Andrinopoulou E-R, Sijbring AMM, Haarman EG, Tiddens HAWM, Pijnenburg MWH. Airway-artery quantitative assessment on chest computed tomography in paediatric primary ciliary dyskinesia. *ERJ Open Res*. 2020 Sep 14;6(3):00210–2019.
25. Kennedy MP, Noone PG, Leigh MW, Zariwala MA, Minnix SL, Knowles MR, et al. High-Resolution CT of Patients with Primary Ciliary Dyskinesia. *Am J Roentgenol*. 2007 May 1;188(5):1232–8.
26. Verleden SE, Vasilescu DM, McDonough JE, Ruttens D, Vos R, Vandermeulen E, et al. Linking clinical phenotypes of chronic lung allograft dysfunction to changes in lung structure. *Eur Respir J*. 2015;46(5):1430–9.
27. Glanville AR. Bronchoscopic monitoring after lung transplantation. *Semin Respir Crit Care Med*. 2010 Apr;31(2):208–21.
28. Kotsimpos T, Williams TJ, Anderson GP. Update on lung transplantation: programmes, patients and prospects. *Eur Respir Rev*. 2012;21(126):271–305.
29. Estenne M, Maurer JR, Boehler A, Egan JJ, Frost A, Hertz M, et al. Bronchiolitis obliterans syndrome 2001: an update of the diagnostic criteria. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2002 Mar;21(3):297–310.

30. Verleden GM, Raghu G, Meyer KC, Glanville AR, Corris P. A new classification system for chronic lung allograft dysfunction. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2014 Feb;33(2):127–33.
31. Verleden SE, Vandermeulen E, Ruttens D, Vos R, Vaneylen A, Dupont LJ, et al. Neutrophilic reversible allograft dysfunction (NRAD) and restrictive allograft syndrome (RAS). *Semin Respir Crit Care Med*. 2013 Jun;34(3):352–60.
32. Sato M, Waddell TK, Wagnetz U, Roberts HC, Hwang DM, Haroon A, et al. Restrictive allograft syndrome (RAS): a novel form of chronic lung allograft dysfunction. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2011 Jul;30(7):735–42.
33. Verleden GM, Glanville AR, Lease ED, Fisher AJ, Calabrese F, Corris PA, et al. Chronic lung allograft dysfunction: Definition, diagnostic criteria, and approaches to treatment—A consensus report from the Pulmonary Council of the ISHLT. *J Heart Lung Transplant*. 2019 May 1;38(5):493–503.
34. Verleden SE, Ruttens D, Vandermeulen E, Bellon H, Van Raemdonck DE, Dupont LJ, et al. Restrictive chronic lung allograft dysfunction: Where are we now? *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2015 May;34(5):625–30.
35. Verleden GM, Vos R, Vanaudenaerde B, Dupont L, Yserbyt J, Van Raemdonck D, et al. Current views on chronic rejection after lung transplantation. *Transpl Int Off J Eur Soc Organ Transplant*. 2015 Oct;28(10):1131–9.
36. Hota P, Dass C, Kumaran M, Simpson S. High-Resolution CT Findings of Obstructive and Restrictive Phenotypes of Chronic Lung Allograft Dysfunction: More Than Just Bronchiolitis Obliterans Syndrome. *Am J Roentgenol*. 2018 May 24;211(1):W13–21.
37. Brun A-L, Chabi M-L, Picard C, Mellot F, Grenier PA. Lung Transplantation: CT Assessment of Chronic Lung Allograft Dysfunction (CLAD). *Diagnostics*. 2021;11(5).
38. Bankier AA, Van Muylem A, Knoop C, Estenne M, Gevenois PA. Bronchiolitis Obliterans Syndrome in Heart-Lung Transplant Recipients: Diagnosis with Expiratory CT. *Radiology*. 2001 Feb 1;218(2):533–9.
39. de Jong PA, Dodd JD, Coxson HO, Storness-Bliss C, Paré PD, Mayo JR, et al. Bronchiolitis obliterans following lung transplantation: early detection using computed tomographic scanning. *Thorax*. 2006/05/02 ed. 2006 Sep;61(9):799–804.
40. Siegel MJ, Bhalla S, Gutierrez FR, Hildebolt C, Sweet S. Post-lung transplantation bronchiolitis obliterans syndrome: usefulness of expiratory thin-section CT for diagnosis. *Radiology*. 2001 Aug;220(2):455–62.
41. Sato M, Hwang DM, Waddell TK, Singer LG, Keshavjee S. Progression pattern of restrictive allograft syndrome after lung transplantation. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2013 Jan;32(1):23–30.
42. Ofek E, Sato M, Saito T, Wagnetz U, Roberts HC, Chaparro C, et al. Restrictive allograft syndrome post lung transplantation is characterized by pleuroparenchymal fibroelastosis. *Mod Pathol Off J U S Can Acad Pathol Inc*. 2013 Mar;26(3):350–6.

43. Dubbeldam A, Barthels C, Coolen J, Verschakelen JA, Verleden SE, Vos R, et al. Restrictive allograft syndrome after lung transplantation: new radiological insights. *Eur Radiol*. 2017 Jul;27(7):2810–7.
44. Verleden SE, de Jong PA, Ruttens D, Vandermeulen E, van Raemdonck DE, Verschakelen J, et al. Functional and computed tomographic evolution and survival of restrictive allograft syndrome after lung transplantation. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2014 Mar;33(3):270–7.
45. Sato M, Hwang DM, Ohmori-Matsuda K, Chaparro C, Waddell TK, Singer LG, et al. Revisiting the pathologic finding of diffuse alveolar damage after lung transplantation. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2012 Apr;31(4):354–63.
46. Verleden SE, Vos R, Verleden GM. Chronic lung allograft dysfunction: light at the end of the tunnel? *Curr Opin Organ Transplant*. 2019 Jun;24(3):318–23.
47. Benden C, Houghton M, Leonard S, Huber LC. Therapy options for chronic lung allograft dysfunction-bronchiolitis obliterans syndrome following first-line immunosuppressive strategies: A systematic review. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2017 Sep;36(9):921–33.
48. Vos R, Vanaudenaerde BM, Verleden SE, De Vleeschauwer SI, Willems-Widyastuti A, Van Raemdonck DE, et al. A randomised controlled trial of azithromycin to prevent chronic rejection after lung transplantation. *Eur Respir J*. 2011 Jan;37(1):164–72.
49. Vos R, Ruttens D, Verleden SE, Vandermeulen E, Bellon H, Van Herck A, et al. High-dose vitamin D after lung transplantation: A randomized trial. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2017 Aug;36(8):897–905.
50. Jain R, Hachem R, Morrell M, Trulock E, Chakinala M, Yusen R, et al. Azithromycin is associated with increased survival in lung transplant recipients with bronchiolitis obliterans syndrome. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2010 May 1;29:531–7.
51. Ruttens D, Verleden SE, Demeyer H, Van Raemdonck DE, Yserbyt J, Dupont LJ, et al. Montelukast for bronchiolitis obliterans syndrome after lung transplantation: A randomized controlled trial. *PloS One*. 2018 Apr 6;13(4):e0193564–e0193564.
52. Vos R, Eynde RV, Ruttens D, Verleden SE, Vanaudenaerde BM, Dupont LJ, et al. Montelukast in chronic lung allograft dysfunction after lung transplantation. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2019 May;38(5):516–27.
53. Moniodis A, Townsend K, Rabin A, Aloum O, Stempel J, Burkett P, et al. Comparison of extracorporeal photopheresis and alemtuzumab for the treatment of chronic lung allograft dysfunction. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2018 Mar;37(3):340–8.
54. Isenring B, Robinson C, Buergi U, Schuurmans MM, Kohler M, Huber LC, et al. Lung transplant recipients on long-term extracorporeal photopheresis. *Clin Transplant*. 2017 Oct;31(10).
55. Robinson CA, Huber L, Murer C, Schuurmans M, Kohler M, Hofbauer G, et al. Cessation of extracorporeal photopheresis in chronic lung allograft dysfunction: effects on clinical outcome in adults. *Swiss Med Wkly*. 2017;147:w14429.

56. Keller CA, Gonwa TA, Hodge DO, Hei DJ, Centanni JM, Zubair AC. Feasibility, Safety, and Tolerance of Mesenchymal Stem Cell Therapy for Obstructive Chronic Lung Allograft Dysfunction. *Stem Cells Transl Med.* 2018 Feb;7(2):161–7.
57. Chambers DC, Enever D, Lawrence S, Sturm MJ, Herrmann R, Yerkovich S, et al. Mesenchymal Stromal Cell Therapy for Chronic Lung Allograft Dysfunction: Results of a First-in-Man Study. *Stem Cells Transl Med.* 2017/02/01 ed. 2017 Apr;6(4):1152–7.
58. Vos R, Wuyts WA, Gheysens O, Goffin KE, Schaevers V, Verleden SE, et al. Pirfenidone in restrictive allograft syndrome after lung transplantation: A case series. *Am J Transplant Off J Am Soc Transplant Am Soc Transpl Surg.* 2018 Dec;18(12):3045–59.
59. Suhling H, Bollmann B, Gottlieb J. Nintedanib in restrictive chronic lung allograft dysfunction after lung transplantation. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant.* 2016 Jul;35(7):939–40.
60. Waseda R, Benazzo A, Hoetzenecker K, Jaksch P, Muraközy G, Gruber S, et al. The influence of retransplantation on survival for pediatric lung transplant recipients. *J Thorac Cardiovasc Surg.* 2018 Nov 1;156(5):2025-2034.e2.
61. Biswas Roy S, Panchanathan R, Walia R, Varsch KE, Kang P, Huang J, et al. Lung Retransplantation for Chronic Rejection: A Single-Center Experience. *Ann Thorac Surg.* 2018 Jan;105(1):221–7.
62. Halloran K, Aversa M, Tinckam K, Martinu T, Binnie M, Chaparro C, et al. Comprehensive outcomes after lung retransplantation: A single-center review. *Clin Transplant.* 2018 Jun;32(6):e13281.
63. Tissot A, Danger R, Claustre J, Magnan A, Brouard S. Early Identification of Chronic Lung Allograft Dysfunction: The Need of Biomarkers. *Front Immunol.* 2019 Jul 17;10:1681–1681.
64. Denton CP, Khanna D. Systemic sclerosis. *The Lancet.* 2017 Oct 7;390(10103):1685–99.
65. Chiffot H, Fautrel B, Sordet C, Chatelus E, Sibia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. *Semin Arthritis Rheum.* 2008 Feb;37(4):223–35.
66. Camiciottoli G, Orlandi I, Bartolucci M, Meoni E, Nacci F, Diciotti S, et al. Lung CT densitometry in systemic sclerosis: correlation with lung function, exercise testing, and quality of life. *Chest.* 2007 Mar;131(3):672–81.
67. Bussone G, Mouthon L. Interstitial lung disease in systemic sclerosis. *Autoimmun Rev.* 2011 Mar;10(5):248–55.
68. Bouros D, Wells AU, Nicholson AG, Colby TV, Polychronopoulos V, Pantelidis P, et al. Histopathologic subsets of fibrosing alveolitis in patients with systemic sclerosis and their relationship to outcome. *Am J Respir Crit Care Med.* 2002 Jun 15;165(12):1581–6.
69. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med.* 2011 Mar 15;183(6):788–824.

70. Le Gouellec N, Duhamel A, Perez T, Hachulla A-L, Sobanski V, Faivre J-B, et al. Predictors of lung function test severity and outcome in systemic sclerosis-associated interstitial lung disease. *PLOS ONE*. 2017 Aug 1;12(8):e0181692.
71. Meier FMP, Frommer KW, Dinser R, Walker UA, Czirjak L, Denton CP, et al. Update on the profile of the EUSTAR cohort: an analysis of the EULAR Scleroderma Trials and Research group database. *Ann Rheum Dis*. 2012 Aug;71(8):1355–60.
72. Wallace B, Vummidi D, Khanna D. Management of connective tissue diseases associated interstitial lung disease: a review of the published literature. *Curr Opin Rheumatol*. 2016 May;28(3):236–45.
73. Moore OA, Goh N, Corte T, Rouse H, Hennessy O, Thakkar V, et al. Extent of disease on high-resolution computed tomography lung is a predictor of decline and mortality in systemic sclerosis-related interstitial lung disease. *Rheumatol Oxf Engl*. 2013 Jan;52(1):155–60.
74. Goldin J, Elashoff R, Kim HJ, Yan X, Lynch D, Strollo D, et al. Treatment of scleroderma-interstitial lung disease with cyclophosphamide is associated with less progressive fibrosis on serial thoracic high-resolution CT scan than placebo: findings from the scleroderma lung study. *Chest*. 2009 Nov;136(5):1333–40.
75. Tashkin DP, Roth MD, Clements PJ, Furst DE, Khanna D, Kleerup EC, et al. Mycophenolate mofetil versus oral cyclophosphamide in scleroderma-related interstitial lung disease (SLS II): a randomised controlled, double-blind, parallel group trial. *Lancet Respir Med*. 2016/07/25 ed. 2016 Sep;4(9):708–19.
76. Gupta S, Hartley R, Khan UT, Singapuri A, Hargadon B, Monteiro W, et al. Quantitative computed tomography–derived clusters: Redefining airway remodeling in asthmatic patients. *J Allergy Clin Immunol*. 2014 Mar 1;133(3):729–738.e18.
77. Patyk M, Obojski A, Sokołowska-Dąbek D, Parkitna-Patyk M, Zaleska-Dorobisz U. Airway wall thickness and airflow limitations in asthma assessed in quantitative computed tomography. *Ther Adv Respir Dis*. 2020;14:1753466619898598–1753466619898598.
78. Ash SY, Diaz AA. The role of imaging in the assessment of severe asthma. *Curr Opin Pulm Med [Internet]*. 2017;23(1). Available from: https://journals.lww.com/copulmonarymedicine/Fulltext/2017/01000/The_role_of_imaging_in_the_assessment_of_severe.15.aspx
79. Hackx M, Gysels E, Severo Garcia T, De Meulder I, Alard S, Bruyneel M, et al. Chronic Obstructive Pulmonary Disease: CT Quantification of Airway Dimensions, Numbers of Airways to Measure, and Effect of Bronchodilation. *Radiology*. 2015 May 19;277(3):853–62.
80. Regan EA, Hokanson JE, Murphy JR, Make B, Lynch DA, Beaty TH, et al. Genetic Epidemiology of COPD (COPDGene) Study Design. *COPD J Chronic Obstr Pulm Dis*. 2011 Feb 1;7(1):32–43.
81. Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J*. 2008 Apr;31(4):869–73.
82. Bodduluri S, Kizhakke Puliyakote A, Nakhmani A, Charbonnier J-P, Reinhardt JM, Bhatt SP. Computed Tomography–based Airway Surface Area–to–Volume Ratio for Phenotyping

- Airway Remodeling in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med*. 2020 Aug 5;203(2):185–91.
83. Galbán CJ, Han MK, Boes JL, Chughtai KA, Meyer CR, Johnson TD, et al. Computed tomography–based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression. *Nat Med*. 2012 Nov 1;18(11):1711–5.
 84. Han MK, Quibrera PM, Carretta EE, Barr RG, Bleeker ER, Bowler RP, et al. Frequency of exacerbations in patients with chronic obstructive pulmonary disease: an analysis of the SPIROMICS cohort. *Lancet Respir Med*. 2017 Aug 1;5(8):619–26.
 85. Grenier PA, Fetita CI, Brillet P-Y. Quantitative computed tomography imaging of airway remodeling in severe asthma. 2016. 2016;6(1):76–83.
 86. Lederlin M, Laurent F, Portron Y, Ozier A, Cochet H, Berger P, et al. CT Attenuation of the Bronchial Wall in Patients With Asthma: Comparison With Geometric Parameters and Correlation With Function and Histologic Characteristics. *Am J Roentgenol*. 2012 Dec 1;199(6):1226–33.
 87. HACKX M, FRANCOTTE D, Garcia T, Muylem A, WALSDORFF M, GEVENOIS P. Effect of Total Lung Capacity, Gender and Height on CT airways measurements. *Br J Radiol*. 2017 Jun 14;90:20160898.
 88. Nambu A, Zach J, Schroeder J, Jin G, Kim SS, Kim Y-I, et al. Quantitative computed tomography measurements to evaluate airway disease in chronic obstructive pulmonary disease: Relationship to physiological measurements, clinical index and visual assessment of airway disease. *Eur J Radiol*. 2016 Nov;85(11):2144–51.
 89. Chen H, Chen G, Zeng Q, Xia T, Zhou J, Qiu S. Quantitative Assessment of Airway Pathology in Subjects With COPD Using Low-Dose High-Resolution Computed Tomography. *Respir Care*. 2017 Jul 1;62(7):953.
 90. Grydeland TB, Thorsen E, Dirksen A, Jensen R, Coxson HO, Pillai SG, et al. Quantitative CT measures of emphysema and airway wall thickness are related to DLCO. *Respir Med*. 2011 Mar 1;105(3):343–51.
 91. Nishimoto K, Karayama M, Inui N, Mori K, Kono M, Hozumi H, et al. Relationship between fraction of exhaled nitric oxide and airway morphology assessed by three-dimensional CT analysis in asthma. *Sci Rep*. 2017 Aug 31;7(1):10187–10187.
 92. Müller N, Staples C, Miller R, Abboud R. “Density mask”. An objective method to quantitate emphysema using computed tomography. *Chest*. 1988 Oct;94(4):782–7.
 93. Lederer DJ, Enright PL, Kawut SM, Hoffman EA, Hunninghake G, van Beek EJR, et al. Cigarette smoking is associated with subclinical parenchymal lung disease: the Multi-Ethnic Study of Atherosclerosis (MESA)-lung study. *Am J Respir Crit Care Med*. 2009 Sep 1;180(5):407–14.
 94. Kliment CR, Araki T, Doyle TJ, Gao W, Dupuis J, Latourelle JC, et al. A comparison of visual and quantitative methods to identify interstitial lung abnormalities. *BMC Pulm Med*. 2015 Oct 29;15:134.

95. Podolanczuk AJ, Oelsner EC, Barr RG, Hoffman EA, Armstrong HF, Austin JHM, et al. High attenuation areas on chest computed tomography in community-dwelling adults: the MESA study. *Eur Respir J*. 2016 Nov;48(5):1442–52.
96. Wang Z, Gu S, Leader JK, Kundu S, Tedrow JR, Scirba FC, et al. Optimal threshold in CT quantification of emphysema. *Eur Radiol*. 2013 Apr;23(4):975–84.
97. Gattinoni L, Caironi P, Cressoni M, Chiumello D, Ranieri VM, Quintel M, et al. Lung recruitment in patients with the acute respiratory distress syndrome. *N Engl J Med*. 2006 Apr 27;354(17):1775–86.
98. Ohkubo H, Kanemitsu Y, Uemura T, Takakuwa O, Takemura M, Maeno K, et al. Normal Lung Quantification in Usual Interstitial Pneumonia Pattern: The Impact of Threshold-based Volumetric CT Analysis for the Staging of Idiopathic Pulmonary Fibrosis. *PLoS One*. 2016;11(3):e0152505.
99. Belfiore MP, Urraro F, Grassi R, Giacobbe G, Patelli G, Cappabianca S, et al. Artificial intelligence to codify lung CT in Covid-19 patients. *Radiol Med (Torino)*. 2020/05/04 ed. 2020 May;125(5):500–4.
100. Jain N, Covar RA, Gleason MC, Newell JDJ, Gelfand EW, Spahn JD. Quantitative computed tomography detects peripheral airway disease in asthmatic children. *Pediatr Pulmonol*. 2005 Sep;40(3):211–8.
101. Busacker A, Newell JD Jr, Keefe T, Hoffman EA, Granroth JC, Castro M, et al. A multivariate analysis of risk factors for the air-trapping asthmatic phenotype as measured by quantitative CT analysis. *Chest*. 2008/08/08 ed. 2009 Jan;135(1):48–56.
102. Newman KB, Lynch DA, Newman LS, Ellegood D, Newell JD Jr. Quantitative Computed Tomography Detects Air Trapping due to Asthma. *CHEST*. 1994 Jul 1;106(1):105–9.
103. Schroeder JD, McKenzie AS, Zach JA, Wilson CG, Curran-Everett D, Stinson DS, et al. Relationships between airflow obstruction and quantitative CT measurements of emphysema, air trapping, and airways in subjects with and without chronic obstructive pulmonary disease. *AJR Am J Roentgenol*. 2013 Sep;201(3):W460-470.
104. Ueda T, Niimi A, Matsumoto H, Takemura M, Hirai T, Yamaguchi M, et al. Role of small airways in asthma: investigation using high-resolution computed tomography. *J Allergy Clin Immunol*. 2006 Nov;118(5):1019–25.
105. Tunon-de-Lara J-M, Laurent F, Giraud V, Perez T, Aguilaniu B, Meziane H, et al. Air trapping in mild and moderate asthma: effect of inhaled corticosteroids. *J Allergy Clin Immunol*. 2007 Mar;119(3):583–90.
106. Hartley PG, Galvin JR, Hunninghake GW, Merchant JA, Yagla SJ, Speakman SB, et al. High-resolution CT-derived measures of lung density are valid indexes of interstitial lung disease. *J Appl Physiol*. 1994 Jan 1;76(1):271–7.
107. Best AC, Lynch AM, Bozic CM, Miller D, Grunwald GK, Lynch DA. Quantitative CT Indexes in Idiopathic Pulmonary Fibrosis: Relationship with Physiologic Impairment. *Radiology*. 2003 Aug 1;228(2):407–14.

108. Walsh SLF, Nair A, Hansell DM. Post-processing applications in thoracic computed tomography. *Clin Radiol*. 2013 May;68(5):433–48.
109. Sverzellati N, Zompatori M, De Luca G, Chetta A, Bnà C, Ormitti F, et al. Evaluation of quantitative CT indexes in idiopathic interstitial pneumonitis using a low-dose technique. *Eur J Radiol*. 2005 Dec;56(3):370–5.
110. Iwasawa T, Kanauchi T, Hoshi T, Ogura T, Baba T, Gotoh T, et al. Multicenter study of quantitative computed tomography analysis using a computer-aided three-dimensional system in patients with idiopathic pulmonary fibrosis. *Jpn J Radiol*. 2016 Jan 1;34(1):16–27.
111. Marten K, Dicken V, Kneitz C, Höhmann M, Kenn W, Hahn D, et al. Interstitial lung disease associated with collagen vascular disorders: disease quantification using a computer-aided diagnosis tool. *Eur Radiol*. 2009 Feb 1;19(2):324–32.
112. Ash SY, Harmouche R, Vallejo DLL, Villalba JA, Ostridge K, Gunville R, et al. Densitometric and local histogram based analysis of computed tomography images in patients with idiopathic pulmonary fibrosis. *Respir Res*. 2017 Mar 7;18(1):45.
113. Ariani A, Silva M, Seletti V, Bravi E, Saracco M, Parisi S, et al. Quantitative chest computed tomography is associated with two prediction models of mortality in interstitial lung disease related to systemic sclerosis. *Rheumatology*. 2017 Jun 1;56(6):922–7.
114. Bartholmai BJ, Raghunath S, Karwoski RA, Moua T, Rajagopalan S, Maldonado F, et al. Quantitative Computed Tomography Imaging of Interstitial Lung Diseases. *J Thorac Imaging* [Internet]. 2013;28(5). Available from: https://journals.lww.com/thoracicimaging/Fulltext/2013/09000/Quantitative_Computed_Tomography_Imaging_of.8.aspx
115. Sverzellati N, Brillet P-Y. When Deep Blue first defeated Kasparov: is a machine stronger than a radiologist at predicting prognosis in idiopathic pulmonary fibrosis? *Eur Respir J*. 2017 Jan 1;49(1):1602144.
116. UPPALURI R, HOFFMAN EA, SONKA M, HUNNINGHAKE GW, McLENNAN G. Interstitial Lung Disease: a quantitative study using the adaptive multiple feature method. *Am J Respir Crit Care Med*. 1999 Feb 1;159(2):519–25.
117. Salisbury ML, Lynch DA, van Beek EJR, Kazerooni EA, Guo J, Xia M, et al. Idiopathic Pulmonary Fibrosis: The Association between the Adaptive Multiple Features Method and Fibrosis Outcomes. *Am J Respir Crit Care Med*. 2016 Oct 21;195(7):921–9.
118. Jacob J, Bartholmai BJ, Rajagopalan S, Kokosi M, Nair A, Karwoski R, et al. Automated Quantitative Computed Tomography Versus Visual Computed Tomography Scoring in Idiopathic Pulmonary Fibrosis: Validation Against Pulmonary Function. *J Thorac Imaging*. 2016 Sep;31(5):304–11.
119. Jacob J, Bartholmai BJ, Rajagopalan S, Kokosi M, Egashira R, Brun AL, et al. Serial automated quantitative CT analysis in idiopathic pulmonary fibrosis: functional correlations and comparison with changes in visual CT scores. *Eur Radiol*. 2018 Mar;28(3):1318–27.
120. D. Y. Chong, H. J. Kim, P. Lo, S. Young, M. F. McNitt-Gray, F. Abtin, et al. Robustness-Driven Feature Selection in Classification of Fibrotic Interstitial Lung Disease Patterns in Computed Tomography Using 3D Texture Features. *IEEE Trans Med Imaging*. 2016 Jan;35(1):144–57.

121. Hansell DM, Goldin JG, King TE Jr, Lynch DA, Richeldi L, Wells AU. CT staging and monitoring of fibrotic interstitial lung diseases in clinical practice and treatment trials: a Position Paper from the Fleischner society. *Lancet Respir Med*. 2015 Jun 1;3(6):483–96.
122. Vakalopoulou M, Chassagnon G, Bus N, Marini Silva R, Zacharaki EI, Revel M-P, et al. AtlasNet: Multi-atlas Non-linear Deep Networks for Medical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention* [Internet]. Granada, Spain; 2018. Available from: <https://hal.inria.fr/hal-01958236>
123. Copeland, B.. Artificial intelligence. In: *Encyclopedia Britannica* [Internet]. 2020. Available from: <https://www.britannica.com/technology/artificial-intelligence>.
124. McCarthy J, Minsky ML, Rochester N, Shannon CE. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Mag*. 2006 Dec 15;27(4):12.
125. R. Kline. Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence. *IEEE Ann Hist Comput*. 2011 Apr;33(4):5–16.
126. *The Tumultuous History of the Search for Artificial Intelligence*, Daniel Crevier. 1993. Basic Books, New York, NY. 432 pages. ISBN: 0-465-02997-3. \$27.50. *Bull Sci Technol Soc*. 1994;14(4):224–224.
127. McCorduck. Boom of the 1980s: rise of expert systems, Fifth Generation Project, Alvey, MCC, SCI. In 2004. p. 426–41.
128. James Lighthill. First AI Winter, Mansfield Amendment,. *Artif Intell Gen Surv*. 1993;115–7.
129. McCorduck, Pamela. Second AI winter. In: *Machines Who Think*. 2004. p. 430–5.
130. Fujita H. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. *Radiol Phys Technol*. 2020 Mar 1;13(1):6–19.
131. Weizenbaum J. ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine. *Commun ACM*. 1966 Jan;9(1):36–45.
132. Rosen CA, Nilsson NJ, Adam MB. A research and development program in applications of intelligent automata to reconnaissance-phase I. In: *Technical report*. Stanford Research Institute.; 1965.
133. Russell, Stuart J.; Norvig, Peter. *Artificial Intelligence: A Modern Approach*. 2nd ed. Upper Saddle River, New Jersey: Prentice Hall; 2003.
134. McCorduck P. *Machines who think : a personal inquiry into the history and prospects of artificial intelligence*. Natick, Mass.: A.K. Peters; 2004. 480–483 p.
135. World Intellectual Property Organization. *WIPO Technology Trends 2019 - Artificial Intelligence*.
136. Markoff J. Computer Wins on 'Jeopardy!': Trivial, It's Not. In *The New York Times*; 2011. Available from: <https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html>
137. Mark H. How Google's autonomous car passed the first US state self-driving test. *IEEE Spectr IEEE*. 2014;

138. Moyer C. How Google's AlphaGo Beat a Go World Champion - Inside a man-versus-machine showdown. *The Atlantic*. 2016 Mar 28;
139. Gorry GA, Barnett GO. Sequential diagnosis by computer. *JAMA*. 1968 Sep 16;205(12):849–54.
140. Gorry GA, Barnett GO. Experience with a model of sequential diagnosis. *Comput Biomed Res Int J*. 1968 May;1(5):490–507.
141. Premarket Approval (PMA) - M1000 IMAGECHECKER [Internet]. U.S. Food and Drug Administration. 1998. Available from: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P970058>
142. Yanase J, Triantaphyllou E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Syst Appl*. 2019 Dec 30;138:112821.
143. Lombardo P, Boehm I, Nairz K. RadioComics – Santa Claus and the future of radiology. *Eur J Radiol* [Internet]. 2020 Jan 1 [cited 2021 Jan 30];122. Available from: <https://doi.org/10.1016/j.ejrad.2019.108771>
144. Ford, Martin; Colvin, Geoff. Will robots create more jobs than they destroy? *The Guardian* [Internet]. 2015 Sep 6; Available from: <https://www.theguardian.com/technology/2015/sep/06/will-robots-create-destroy-jobs>
145. A. L. Samuel. Some Studies in Machine Learning Using the Game of Checkers. *IBM J Res Dev*. 1959 Jul;3(3):210–29.
146. Andrew AM. REINFORCEMENT LEARNING: AN INTRODUCTION by Richard S. Sutton and Andrew G. Barto, Adaptive Computation and Machine Learning series, MIT Press (Bradford Book), Cambridge, Mass., 1998, xviii + 322 pp, ISBN 0-262-19398-1, (hardback, £31.95). *Robotica*. 1999/03/01 ed. 1999;17(2):229–35.
147. Brazdil P, Carrier CG, Soares C, Vilalta R. *Metalearning: Applications to Data Mining* [Internet]. Springer Berlin Heidelberg; 2008. (Cognitive Technologies). Available from: <https://books.google.com.vn/books?id=DfZDAAAQBAJ>
148. Dawson CW, Wilby R. An artificial neural network approach to rainfall-runoff modelling. *Hydrol Sci J*. 1998 Feb 1;43(1):47–66.
149. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015 May 1;521(7553):436–44.
150. Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Neural Networks. Vol. 15, *Journal of Machine Learning Research*. 2010.
151. Raina R, Madhavan A, Ng AY. Large-Scale Deep Unsupervised Learning Using Graphics Processors. In: *Proceedings of the 26th Annual International Conference on Machine Learning* [Internet]. New York, NY, USA: Association for Computing Machinery; 2009. p. 873–80. (ICML '09). Available from: <https://doi.org/10.1145/1553374.1553486>
152. LeCun Y, Boser B, Denker J, Henderson D, Howard R, Hubbard W, et al. Handwritten Digit Recognition with a Back-Propagation Network. In: Touretzky D, editor. *Advances in Neural Information Processing Systems* [Internet]. Morgan-Kaufmann; 1990. p. 396–404. Available from:

- <https://proceedings.neurips.cc/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf>
153. HUBEL DH, WIESEL TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol.* 1962 Jan;160(1):106–54.
 154. Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex N Y N* 1991. 1991 Feb;1(1):1–47.
 155. Cadieu CF, Hong H, Yamins DLK, Pinto N, Ardila D, Solomon EA, et al. Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLOS Comput Biol.* 2014 Dec 18;10(12):e1003963.
 156. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ, editors. *Advances in Neural Information Processing Systems* [Internet]. Curran Associates, Inc.; 2012. p. 1097–105. Available from: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
 157. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J Mach Learn Res.* 2014;15(56):1929–58.
 158. Tompson J, Goroshin R, Jain A, LeCun Y, Bregler C. Efficient Object Localization Using Convolutional Networks. *ArXiv E-Prints.* 2014;arXiv:1411.4280.
 159. Taigman Y, Yang M, Ranzato M, Wolf L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition.* 2014. p. 1701–8.
 160. Ypsilantis P, Montana G. Recurrent Convolutional Networks for Pulmonary Nodule Detection in CT Imaging. 2016 Sep 28;
 161. Armato SG 3rd, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): a completed reference database of lung nodules on CT scans. *Med Phys.* 2011 Feb;38(2):915–31.
 162. GE Healthcare. GE Healthcare Receives FDA Clearance of First Artificial Intelligence Algorithms Embedded On-Device to Prioritize Critical Chest X-ray Review [Internet]. 2019. Available from: https://www.ge.com/news/press-releases/ge-healthcare-receives-fda-clearance-first-artificial-intelligence-algorithms#_ftnref1
 163. Yuan R, Mayo J, Streit I, Atkar-Khattra S, Myers R, Yee J, et al. MA10.06 Randomized Clinical Trial with Computer Assisted Diagnosis (CAD) Versus Radiologist as First Reader of Lung Screening LDCT. *J Thorac Oncol.* 2019 Oct 1;14(10):S287–8.
 164. Causey JL, Zhang J, Ma S, Jiang B, Qualls JA, Politte DG, et al. Highly accurate model for prediction of lung nodule malignancy with CT scans. *Sci Rep.* 2018 Jun 18;8(1):9286.
 165. Walsh SLF, Calandriello L, Silva M, Sverzellati N. Deep learning for classifying fibrotic lung disease on high-resolution computed tomography: a case-cohort study. *Lancet Respir Med.* 2018 Nov 1;6(11):837–45.

166. González G, Ash SY, Vegas-Sánchez-Ferrero G, Onieva Onieva J, Rahaghi FN, Ross JC, et al. Disease Staging and Prognosis in Smokers Using Deep Learning in Chest Computed Tomography. *Am J Respir Crit Care Med*. 2018 Jan 15;197(2):193–203.
167. Kao E-F, Liu G-C, Lee L-Y, Tsai H-Y, Jaw T-S. Computer-aided detection system for chest radiography: reducing report turnaround times of examinations with abnormalities. *Acta Radiol Stockh Swed* 1987. 2015 Jun;56(6):696–701.
168. Qin C, Yao D, Shi Y, Song Z. Computer-aided detection in chest radiography based on artificial intelligence: a survey. *Biomed Eng OnLine*. 2018 Aug 22;17(1):113.
169. Mollura DJ, Azene EM, Starikovskiy A, Thelwell A, Iosifescu S, Kimble C, et al. White Paper Report of the RAD-AID Conference on International Radiology for Developing Countries: Identifying Challenges, Opportunities, and Strategies for Imaging Services in the Developing World. *J Am Coll Radiol*. 2010 Jul 1;7(7):495–500.
170. Domingos P. A Few Useful Things to Know about Machine Learning. *Commun ACM*. 2012 Oct;55(10):78–87.
171. Hu Z, Tang J, Wang Z, Zhang K, Zhang L, Sun Q. Deep Learning for Image-based Cancer Detection and Diagnosis — A Survey. *Pattern Recognit*. 2018 May 1;83.
172. A. Mikołajczyk, M. Grochowski. Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW). 2018. p. 117–22.
173. X. Cao, D. Wipf, F. Wen, G. Duan, J. Sun. A Practical Transfer Learning Algorithm for Face Verification. In: 2013 IEEE International Conference on Computer Vision. 2013. p. 3208–15.
174. Bush I. Lung nodule detection and classification [Internet]. (Stanford Computer Science; vol. 2016 20:196–209). Available from: http://cs231n.stanford.edu/reports/2016/pdfs/313_Report.pdf
175. Harmon SA, Sanford TH, Xu S, Turkbey EB, Roth H, Xu Z, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun*. 2020 Aug 14;11(1):4080.
176. NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories [Internet]. Academic Torrents. [cited 2021 Mar 7]. Available from: <https://academictorrents.com/details/557481faacd824c83fbf57dcf7b6da9383b3235a>
177. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciampi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017 Dec 1;42:60–88.
178. Fu B, Wang G, Wu M, Li W, Zheng Y, Chu Z, et al. Influence of CT effective dose and convolution kernel on the detection of pulmonary nodules in different artificial intelligence software systems: A phantom study. *Eur J Radiol*. 2020 May;126:108928.
179. Blazis SP, Dickerscheid DBM, Linsen PVM, Martins Jarnalo CO. Effect of CT reconstruction settings on the performance of a deep learning based lung nodule CAD system. *Eur J Radiol*. 2021 Mar;136:109526.

180. Kooi T, van Ginneken B, Karssemeijer N, den Heeten A. Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. *Med Phys*. 2017 Mar;44(3):1017–27.
181. Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Computer Vision – ECCV 2014*. Cham: Springer International Publishing; 2014. p. 818–33.
182. Springenberg JT, Dosovitskiy A, Brox T, Riedmiller MA. Striving for Simplicity: The All Convolutional Net. *CoRR*. 2015;abs/1412.6806.
183. Montavon G, Lapuschkin S, Binder A, Samek W, Müller K-R. Explaining nonlinear classification decisions with deep Taylor decomposition. *Pattern Recognit*. 2017 May 1;65:211–22.
184. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016 Feb;278(2):563–77.
185. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer Oxf Engl* 1990. 2012/01/16 ed. 2012 Mar;48(4):441–6.
186. Chassagnon G, Vakalopoulou M, Paragios N, Revel M-P. Artificial intelligence applications for thoracic imaging. *Eur J Radiol [Internet]*. 2020 Feb 1 [cited 2021 Mar 7];123. Available from: <https://doi.org/10.1016/j.ejrad.2019.108774>
187. Chassagnon G, Vakalopolou M, Paragios N, Revel M-P. Deep learning: definition and perspectives for thoracic imaging. *Eur Radiol*. 2020 Apr 1;30(4):2021–30.
188. Warrick JH, Bhalla M, Schabel SI, Silver RM. High resolution computed tomography in early scleroderma lung disease. *J Rheumatol*. 1991 Oct;18(10):1520–8.
189. Kim HG, Tashkin DP, Clements PJ, Li G, Brown MS, Elashoff R, et al. A computer-aided diagnosis system for quantitative scoring of extent of lung fibrosis in scleroderma patients. *Clin Exp Rheumatol*. 2010/11/03 ed. 2010;28(5 Suppl 62):S26–35.
190. Hoang-Thi T-N, Revel M-P, Burgel P-R, Bassinet L, Honoré I, Hua-Huy T, et al. Automated computed tomographic scoring of lung disease in adults with primary ciliary dyskinesia. *BMC Pulm Med*. 2018 Dec 18;18(1):194.
191. Altaf-Ul-Amin Md, Afendi FM, Kiboi SK, Kanaya S. Systems Biology in the Context of Big Data and Networks. Das A, editor. *BioMed Res Int*. 2014 May 27;2014:428570.
192. Belle A, Thiagarajan R, Soroushmehr SMR, Navidi F, Beard DA, Najarian K. Big Data Analytics in Healthcare. *BioMed Res Int*. 2015;2015:370194.
193. Gligorijević V, Malod-Dognin N, Pržulj N. Integrative methods for analyzing big data in precision medicine. *Proteomics*. 2016 Mar;16(5):741–58.
194. S. Sagiroglu, D. Sinanc. Big data: A review. In: 2013 International Conference on Collaboration Technologies and Systems (CTS). 2013. p. 42–7.

195. Berger ML, Doban V. Big data, advanced analytics and the future of comparative effectiveness research. *J Comp Eff Res*. 2014 Mar;3(2):167–76.
196. Baro E, Degoul S, Beuscart R, Chazard E. Toward a Literature-Driven Definition of Big Data in Healthcare. Shirani S, editor. *BioMed Res Int*. 2015 Jun 2;2015:639021.
197. Chen M, Mao S, Liu Y. Big Data: A Survey. *Mob Netw Appl*. 2014 Apr 1;19(2):171–209.
198. Herland M, Khoshgoftaar TM, Wald R. A review of data mining using big data in health informatics. *J Big Data*. 2014 Jun 24;1(1):2.
199. Groves P, Kayyali B, Knott D, Kuiken S. The “big data” revolution in healthcare: Accelerating value and innovation. In 2016.
200. Jee K, Kim G-H. Potentiality of big data in the medical sector: focus on how to reshape the healthcare system. *Healthc Inform Res*. 2013 Jun;19(2):79–85.
201. Rodríguez-Mazahua L, Rodríguez-Enríquez C-A, Sánchez-Cervantes JL, Cervantes J, García-Alcaraz JL, Alor-Hernández G. A general perspective of Big Data: applications, tools, challenges and trends. *J Supercomput*. 2016 Aug 1;72(8):3073–113.
202. Andreu-Perez J, Poon CCY, Merrifield RD, Wong STC, Yang G-Z. Big data for health. *IEEE J Biomed Health Inform*. 2015 Jul;19(4):1193–208.
203. Pietsch W. Aspects of Theory-Ladenness in Data-Intensive Science. *Philos Sci*. 2015;82:905–16.
204. M. Viceconti, P. Hunter, R. Hose. Big Data, Big Knowledge: Big Data for Personalized Healthcare. *IEEE J Biomed Health Inform*. 2015 Jul;19(4):1209–15.
205. Rubin D.L., Greenspan H., Brinkley J.F. Biomedical Imaging Informatics. In: *Biomedical Informatics [Internet]*. In: Shortliffe E., Cimino J. London: Springer,; 2014. Available from: https://doi.org/10.1007/978-1-4471-4474-8_9
206. Hartwell L, Mankoff D, Paulovich A, Ramsey S, Swisher E. Cancer biomarkers: a systems approach. *Nat Biotechnol*. 2006 Aug;24(8):905–8.
207. Ziegler A, Koch A, Krockenberger K, Grosshennig A. Personalized medicine using DNA biomarkers: a review. *Hum Genet*. 2012 Oct;131(10):1627–38.
208. Trusheim MR, Berndt ER, Douglas FL. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nat Rev Drug Discov*. 2007 Apr 1;6(4):287–93.
209. Micheel CM, Nass SJ, Omenn GS, editors. *Evolution of Translational Omics: Lessons Learned and the Path Forward*. Washington (DC): National Academies Press (US); 2012.
210. Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Comput Med Imaging Graph Off J Comput Med Imaging Soc*. 2007/03/08 ed. 2007;31(4–5):198–211.
211. Giger ML. Update on the potential of computer-aided diagnosis for breast cancer. *Future Oncol Lond Engl*. 2010 Jan;6(1):1–4.

212. Pietsch W. The Causal Nature of Modeling with Big Data. *Philos Technol.* 2016 Jun 1;29(2):137–71.
213. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights Imaging.* 2020 Aug 12;11(1):91.
214. van Helden P. Data-driven hypotheses. *EMBO Rep.* 2012/12/21 ed. 2013 Feb;14(2):104–104.
215. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017 Dec;14(12):749–62.
216. Zhu X, Dong D, Chen Z, Fang M, Zhang L, Song J, et al. Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *Eur Radiol.* 2018 Jul 1;28(7):2772–8.
217. Fan L, Fang M, Li Z, Tu W, Wang S, Chen W, et al. Radiomics signature: a biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule. *Eur Radiol.* 2019 Feb 1;29(2):889–97.
218. Jia T-Y, Xiong J-F, Li X-Y, Yu W, Xu Z-Y, Cai X-W, et al. Identifying EGFR mutations in lung adenocarcinoma by noninvasive imaging using radiomics features and random forest modeling. *Eur Radiol.* 2019 Sep 1;29(9):4742–50.
219. Tu W, Sun G, Fan L, Wang Y, Xia Y, Guan Y, et al. Radiomics signature: A potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology. *Lung Cancer.* 2019 Jun 1;132:28–35.
220. Song J, Tian J, Zhang L, Qu X, Qian W, Zheng B, et al. Development and validation of a prognostic index for efficacy evaluation and prognosis of first-line chemotherapy in stage III–IV lung squamous cell carcinoma. *Eur Radiol.* 2019 May 1;29(5):2388–98.
221. Park JE, Kim D, Kim HS, Park SY, Kim JY, Cho SJ, et al. Quality of science and reporting of radiomics in oncologic studies: room for improvement according to radiomics quality score and TRIPOD statement. *Eur Radiol.* 2020 Jan 1;30(1):523–36.
222. Kim S, Lim M-N, Hong Y, Han S-S, Lee S-J, Kim WJ. A cluster analysis of chronic obstructive pulmonary disease in dusty areas cohort identified three subgroups. *BMC Pulm Med.* 2017 Dec 16;17(1):209–209.
223. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology.* 1945 Jul 1;26(3):297–302.
224. Bottou L. Large-Scale Machine Learning with Stochastic Gradient Descent. In: Lechevallier Y, Saporta G, editors. *Proceedings of COMPSTAT’2010.* Heidelberg: Physica-Verlag HD; 2010. p. 177–86.
225. McCue ME, McCoy AM. The Scope of Big Data in One Medicine: Unprecedented Opportunities and Challenges. *Front Vet Sci.* 2017;4:194.
226. Yip SSF, Aerts HJWL. Applications and limitations of radiomics. *Phys Med Biol.* 2016 Jul 7;61(13):R150-166.

227. Galavis PE, Hollensen C, Jallow N, Paliwal B, Jeraj R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol Stockh Swed.* 2010 Oct;49(7):1012–6.
228. McCollough C, Bakalyar DM, Bostani M, Brady S, Boedeker K, Boone JM, et al. Use of Water Equivalent Diameter for Calculating Patient Size and Size-Specific Dose Estimates (SSDE) in CT: The Report of AAPM Task Group 220. *AAPM Rep.* 2014 Sep;2014:6–23.
229. Zhao B, Tan Y, Tsai W-Y, Qi J, Xie C, Lu L, et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep.* 2016 Mar 24;6:23428–23428.
230. Dalal T, Kalra MK, Rizzo SMR, Schmidt B, Suess C, Flohr T, et al. Metallic prosthesis: technique to avoid increase in CT radiation dose with automatic tube current modulation in a phantom and patients. *Radiology.* 2005 Aug;236(2):671–5.
231. Rizzo SMR, Kalra MK, Schmidt B, Raupach R, Maher MM, Blake MA, et al. CT images of abdomen and pelvis: effect of nonlinear three-dimensional optimized reconstruction algorithm on image quality and lesion characteristics. *Radiology.* 2005 Oct;237(1):309–15.
232. Fave X, Zhang L, Yang J, Mackin D, Balter P, Gomez D, et al. Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer. *Transl Cancer Res [Internet].* 2016;5(4). Available from: <https://tcr.amegroups.com/article/view/8709>
233. Desseroit M-C, Tixier F, Weber WA, Siegel BA, Cheze Le Rest C, Visvikis D, et al. Reliability of PET/CT Shape and Heterogeneity Features in Functional and Morphologic Components of Non-Small Cell Lung Cancer Tumors: A Repeatability Analysis in a Prospective Multicenter Cohort. *J Nucl Med Off Publ Soc Nucl Med.* 2017 Mar;58(3):406–11.
234. Huynh E, Coroller TP, Narayan V, Agrawal V, Romano J, Franco I, et al. Associations of Radiomic Data Extracted from Static and Respiratory-Gated CT Scans with Disease Recurrence in Lung Cancer Patients Treated with SBRT. *PLOS ONE.* 2017 Jan 3;12(1):e0169172.
235. van Timmeren JE, Leijenaar RTH, van Elmpt W, Wang J, Zhang Z, Dekker A, et al. Test-Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific? *Tomogr Ann Arbor Mich.* 2016 Dec;2(4):361–5.
236. Rios Velazquez E, Aerts HJWL, Gu Y, Goldgof DB, De Ruyscher D, Dekker A, et al. A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* 2012 Nov;105(2):167–73.
237. van Dam IE, van Sörnsen de Koste JR, Hanna GG, Muirhead R, Slotman BJ, Senan S. Improving target delineation on 4-dimensional CT scans in stage I NSCLC using a deformable registration tool. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* 2010 Jul;96(1):67–72.
238. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilicus S, Chute C, et al. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. 2019.
239. Chassagnon G, Vakalopoulou M, Régent A, Zacharaki EI, Aviram G, Martin C, et al. Deep Learning-based Approach for Automated Assessment of Interstitial Lung Disease in Systemic Sclerosis on CT Images. *Radiol Artif Intell.* 2020 Jul;2(4):e190006.

240. Owens CA, Peterson CB, Tang C, Koay EJ, Yu W, Mackin DS, et al. Lung tumor segmentation methods: Impact on the uncertainty of radiomics features for non-small cell lung cancer. *PLoS One*. 2018;13(10):e0205003.
241. Parmar C, Rios Velazquez E, Leijenaar R, Jermoumi M, Carvalho S, Mak RH, et al. Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation. *PLOS ONE*. 2014 Jul 15;9(7):e102107.
242. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. 2016.
243. Dong C, Loy CC, He K, Tang X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans Pattern Anal Mach Intell*. 2016;38(2):295–307.
244. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and their Compositionality. 2013.
245. Iandola F, Moskewicz M, Karayev S, Girshick R, Darrell T, Keutzer K. DenseNet: Implementing Efficient ConvNet Descriptor Pyramids. 2014.
246. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.
247. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. 2016.
248. Iglesias JE, Sabuncu MR. Multi-atlas segmentation of biomedical images: A survey. *Med Image Anal*. 2015;24(1):205–19.
249. Depeursinge A, Van de Ville D, Platon A, Geissbuhler A, Poletti P-A, Muller H. Near-Affine-Invariant Texture Learning for Lung Tissue Analysis Using Isotropic Wavelet Frames. *IEEE Trans Inf Technol Biomed*. 2012;16(4):665–75.
250. Song Y, Cai W, Zhou Y, Feng DD. Feature-Based Image Patch Approximation for Lung Tissue Classification. *IEEE Trans Med Imaging*. 2013;32(4):797–808.
251. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Trans Med Imaging*. 2016;35(5):1207–16.
252. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. 2015.
253. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. 2016.
254. Anthimopoulos M, Christodoulidis S, Ebner L, Geiser T, Christe A, Mougiakakou S. Semantic Segmentation of Pathological Lung Tissue With Dilated Fully Convolutional Networks. *IEEE J Biomed Health Inform*. 2019 Mar;23(2):714–22.
255. Kim HJ, Li G, Gjertson D, Elashoff R, Shah SK, Ochs R, et al. Classification of Parenchymal Abnormality in Scleroderma Lung Using a Novel Approach to Denoise Images Collected via a Multicenter Study. *Acad Radiol*. 2008 Aug 1;15(8):1004–16.

256. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* 2017 Nov 1;77(21):e104.
257. Van Rossum G, Drake FL. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace; 2009.
258. Song SH, Park H, Lee G, Lee HY, Sohn I, Kim HS, et al. Imaging Phenotyping Using Radiomics to Predict Micropapillary Pattern within Lung Adenocarcinoma. *J Thorac Oncol.* 2017 Apr 1;12(4):624–32.
259. R. M. Haralick, K. Shanmugam, I. Dinstein. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern.* 1973 Nov;SMC-3(6):610–21.
260. Rose CJ, Mills SJ, O’Connor JPB, Buonaccorsi GA, Roberts C, Watson Y, et al. Quantifying spatial heterogeneity in dynamic contrast-enhanced MRI parameter maps. *Magn Reson Med.* 2009 Aug;62(2):488–99.
261. Larkin TJ, Canuto HC, Kettunen MI, Booth TC, Hu D-E, Krishnan AS, et al. Analysis of image heterogeneity using 2D Minkowski functionals detects tumor responses to treatment. *Magn Reson Med.* 2014 Jan;71(1):402–10.
262. Galloway MM. Texture analysis using gray level run lengths. *Comput Graph Image Process.* 1975 Jun 1;4(2):172–9.
263. Thibault G, FERTIL B, Navarro C, Pereira S, Lévy N, Sequeira J, et al. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification. 10th International Conference on Pattern Recognition and Information Processing. 2009.
264. M. Amadasun, R. King. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern.* 1989 Oct;19(5):1264–74.
265. Oliva JT, Lee HD, Spolaôr N, Coy CSR, Wu FC. Prototype system for feature extraction, classification and study of medical images. *Expert Syst Appl.* 2016 Nov 30;63:267–83.
266. Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. *Comput Vis Graph Image Process.* 1983 Sep 1;23(3):341–52.
267. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc Ser B Methodol.* 1996;58(1):267–88.
268. Hyafil L, Rivest RL. Constructing optimal binary decision trees is NP-complete. *Inf Process Lett.* 1976 May 1;5(1):15–7.
269. Rokach L, Maimon O. *Data mining with decision trees. Theory and applications.* Vol. 69, *Mach Percept Artif Intell.* 2008.
270. Kégl B. *The return of AdaBoost.MH: multi-class Hamming trees.* 2013.
271. Freund Y, Schapire RE. Experiments with a New Boosting Algorithm. In: *ICML.* 1996. p. 148–56.
272. Friedman J. Greedy Function Approximation: A Gradient Boosting Machine. *Ann Stat.* 2001 Oct 1;29:1189–232.

273. Mason L, Baxter J, Bartlett P, Frean M. Boosting Algorithms as Gradient Descent. In: Proceedings of the 12th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press; 1999. p. 512–8. (NIPS'99).
274. Cortes C, Vapnik V. Support-vector networks. *Mach Learn.* 1995 Sep 1;20(3):273–97.
275. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* 2011;12:2825–30.
276. Altman NS. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am Stat.* 1992 Aug 1;46(3):175–85.
277. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory.* 1967;13(1):21–7.
278. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature.* 1986 Oct 1;323(6088):533–6.
279. Rayens WS. Discriminant analysis and statistical pattern recognition. Taylor & Francis; 1993.
280. Tharwat A. Linear vs. quadratic discriminant analysis classifier: a tutorial. *Int J Appl Pattern Recognit.* 2016 Jan 1;3:145.
281. Wikipedia contributors. Mutual information — Wikipedia, The Free Encyclopedia [Internet]. 2021. Available from: https://en.wikipedia.org/w/index.php?title=Mutual_information&oldid=1023002590
282. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev E.* 2004;69(6):066138.
283. Wikipedia contributors. Random forest — Wikipedia, The Free Encyclopedia [Internet]. 2021. Available from: https://en.wikipedia.org/w/index.php?title=Random_forest&oldid=1021839899
284. Ho TK. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. 1995. p. 278–82 vol.1.
285. Opitz D, Maclin R. Popular Ensemble Methods: An Empirical Study. *J Artif Int Res.* 1999 Jul;11(1):169–98.
286. Polikar R. Ensemble based systems in decision making. *IEEE Circuits Syst Mag.* 2006;6(3):21–45.
287. Rokach L. Ensemble-based classifiers. *Artif Intell Rev.* 2010 Feb 1;33(1):1–39.
288. Stone M. Cross-Validatory Choice and Assessment of Statistical Predictions. *J R Stat Soc Ser B Methodol.* 1974;36(2):111–33.
289. Afzelius BA. A human syndrome caused by immotile cilia. *Science.* 1976 Jul 23;193(4250):317–9.
290. Meeks M, Bush A. Primary ciliary dyskinesia (PCD). *Pediatr Pulmonol.* 2000 Apr;29(4):307–16.

291. Kuehni CE, Frischer T, Strippoli M-PF, Maurer E, Bush A, Nielsen KG, et al. Factors influencing age at diagnosis of primary ciliary dyskinesia in European children. *Eur Respir J.* 2010 Dec;36(6):1248–58.
292. Shah A, Shoemark A, MacNeill SJ, Bhaludin B, Rogers A, Bilton D, et al. A longitudinal study characterising a large adult primary ciliary dyskinesia population. *Eur Respir J.* 2016;48(2):441–50.
293. Maglione M, Bush A, Montella S, Mollica C, Manna A, Esposito A, et al. Progression of lung disease in primary ciliary dyskinesia: is spirometry less accurate than CT? *Pediatr Pulmonol.* 2012 May;47(5):498–504.
294. Pifferi M, Bush A, Pioggia G, Caramella D, Tartarisco G, Di Cicco M, et al. Evaluation of pulmonary disease using static lung volumes in primary ciliary dyskinesia. *Thorax.* 2012 Nov;67(11):993–9.
295. Frija-Masson J, Bassinet L, Honoré I, Dufeu N, Housset B, Coste A, et al. Clinical characteristics, functional respiratory decline and follow-up in adult patients with primary ciliary dyskinesia. *Thorax.* 2017 Feb;72(2):154–60.
296. Ellemunter H, Fuchs SI, Unsinn KM, Freund MC, Waltner-Romen M, Steinkamp G, et al. Sensitivity of lung clearance index and chest computed tomography in early cf lung disease. *Respir Med.* 2010 Dec 1;104(12):1834–42.
297. Bommart S, Marin G, Bourdin A, Molinari N, Klein F, Hayot M, et al. Relationship between CT air trapping criteria and lung function in small airway impairment quantification. *BMC Pulm Med.* 2014 Feb 28;14(1):29.
298. Cantin L, Bankier AA, Eisenberg RL. Bronchiectasis. *Am J Roentgenol.* 2009 Sep 1;193(3):W158–71.
299. Lund LH, Khush KK, Cherikh WS, Goldfarb S, Kucheryavaya AY, Levvey BJ, et al. The Registry of the International Society for Heart and Lung Transplantation: Thirty-fourth Adult Heart Transplantation Report-2017; Focus Theme: Allograft ischemic time. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant.* 2017 Oct;36(10):1037–46.
300. Van Herck A, Verleden SE, Sacreas A, Heigl T, Vanaudenaerde BM, Dupont LJ, et al. Validation of a post-transplant chronic lung allograft dysfunction classification system. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant.* 2019 Feb;38(2):166–73.
301. DerHovanessian A, Todd JL, Zhang A, Li N, Mayalall A, Finlen Copeland CA, et al. Validation and Refinement of Chronic Lung Allograft Dysfunction Phenotypes in Bilateral and Single Lung Recipients. *Ann Am Thorac Soc.* 2016 May;13(5):627–35.
302. Saito T, Horie M, Sato M, Nakajima D, Shoushtarizadeh H, Binnie M, et al. Low-dose computed tomography volumetry for subtyping chronic lung allograft dysfunction. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant.* 2016 Jan;35(1):59–66.
303. Levy L, Huszti E, Renaud-Picard B, Berra G, Kawashima M, Takahagi A, et al. Risk assessment of chronic lung allograft dysfunction phenotypes: Validation and proposed refinement of the 2019 International Society for Heart and Lung Transplantation classification system. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant.* 2020 Aug;39(8):761–70.

304. Suhling H, Dettmer S, Greer M, Fuehner T, Avsar M, Haverich A, et al. Phenotyping Chronic Lung Allograft Dysfunction Using Body Plethysmography and Computed Tomography. *Am J Transplant Off J Am Soc Transplant Am Soc Transpl Surg*. 2016 Nov;16(11):3163–70.
305. Philippot Q, Debray M-P, Bun R, Frija-Masson J, Bunel V, Morer L, et al. Use of CT-SCAN score and volume measures to early identify restrictive allograft syndrome in single lung transplant recipients. *J Heart Lung Transplant Off Publ Int Soc Heart Transplant*. 2020 Feb;39(2):125–33.
306. Lowery EM, Cascino T, Forsythe SM, Alex C, Camacho P. Effect Of Vitamin D Deficiency On Episodes Of Rejection And Survival After Lung Transplantation. In: A24 IMPROVING OUTCOMES FOLLOWING LUNG TRANSPLANTATION [Internet]. American Thoracic Society; 2011 [cited 2021 Jul 24]. p. A1142–A1142. (American Thoracic Society International Conference Abstracts). Available from: https://doi.org/10.1164/ajrccm-conference.2011.183.1_MeetingAbstracts.A1142
307. Gauthier JM, Hachem RR, Kreisel D. Update on Chronic Lung Allograft Dysfunction. *Curr Transplant Rep*. 2016 Sep 1;3(3):185–91.
308. Hofmanninger J, Prayer F, Pan J, Röhrich S, Prosch H, Langs G. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *Eur Radiol Exp [Internet]*. 2020 Aug;4(1). Available from: <http://dx.doi.org/10.1186/s41747-020-00173-2>
309. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The Balanced Accuracy and Its Posterior Distribution. In: Proceedings of the 2010 20th International Conference on Pattern Recognition [Internet]. USA: IEEE Computer Society; 2010. p. 3121–4. (ICPR '10). Available from: <https://doi.org/10.1109/ICPR.2010.764>
310. John D. Kelleher; Brian Mac Namee ;Aoife D’Arcy; Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. 2015.
311. Mosley L. A balanced approach to the multi-class imbalance problem. In 2013.
312. Urbanowicz RJ, Moore JH. ExSTraCS 2.0: Description and Evaluation of a Scalable Learning Classifier System. *Evol Intell*. 2015/04/03 ed. 2015 Sep;8(2):89–116.
313. Ostridge K, Williams NP, Kim V, Harden S, Bourne S, Clarke SC, et al. Relationship of CT-quantified emphysema, small airways disease and bronchial wall dimensions with physiological, inflammatory and infective measures in COPD. *Respir Res*. 2018 Feb 20;19(1):31.
314. Dettmer S, Suhling H, Klingenberg I, Otten O, Kaireit T, Fuge J, et al. Lobe-wise assessment of lung volume and density distribution in lung transplant patients and value for early detection of bronchiolitis obliterans syndrome. *Eur J Radiol*. 2018 Sep;106:137–44.
315. Saito M, Chen-Yoshikawa TF, Nakamoto Y, Kayawake H, Tokuno J, Ueda S, et al. Unilateral Chronic Lung Allograft Dysfunction Assessed by Biphasic Computed Tomographic Volumetry in Bilateral Living-donor Lobar Lung Transplantation. *Transplant Direct*. 2018 Nov;4(11):e398.
316. Solyanik O, Hollmann P, Dettmer S, Kaireit T, Schaefer-Prokop C, Wacker F, et al. Quantification of Pathologic Air Trapping in Lung Transplant Patients Using CT Density Mapping: Comparison with Other CT Air Trapping Measures. *PLOS ONE*. 2015 Oct 2;10(10):e0139102.

317. Chassagnon G, Zacharaki EI, Bommart S, Burgel P-R, Chiron R, Dangeard S, et al. Quantification of Cystic Fibrosis Lung Disease with Radiomics-based CT Scores. *Radiol Cardiothorac Imaging*. 2020 Dec 17;2(6):e200022–e200022.
318. Bergamasco A, Hartmann N, Wallace L, Verpillat P. Epidemiology of systemic sclerosis and systemic sclerosis-associated interstitial lung disease. *Clin Epidemiol*. 2019;11:257–73.
319. Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972-2002. *Ann Rheum Dis*. 2007/02/28 ed. 2007 Jul;66(7):940–4.
320. Fischer A, Patel NM, Volkmann ER. Interstitial Lung Disease in Systemic Sclerosis: Focus on Early Detection and Intervention. *Open Access Rheumatol Res Rev*. 2019;11:283–307.
321. Bryan C, Knight C, Black CM, Silman AJ. Prediction of five-year survival following presentation with scleroderma: Development of a simple model using three disease factors at first visit. *Arthritis Rheum*. 1999 Dec 1;42(12):2660–5.
322. Fransen J, Popa-Diaconu D, Hesselstrand R, Carreira P, Valentini G, Beretta L, et al. Clinical prediction of 5-year survival in systemic sclerosis: validation of a simple prognostic model in EUSTAR centres. *Ann Rheum Dis*. 2011;70:1788–92.
323. Volkmann ER, Tashkin DP, Sim M, Li N, Goldmuntz E, Keyes-Elstein L, et al. Short-term progression of interstitial lung disease in systemic sclerosis predicts long-term survival in two independent clinical trial cohorts. *Ann Rheum Dis*. 2019 Jan;78(1):122–30.
324. Morisset J, Vittinghoff E, Elicker BM, Hu X, Le S, Ryu JH, et al. Mortality Risk Prediction in Scleroderma-Related Interstitial Lung Disease: The SADL Model. *Chest*. 2017 Nov;152(5):999–1007.
325. Mango RL, Matteson EL, Crowson CS, Ryu JH, Makol A. Assessing Mortality Models in Systemic Sclerosis-Related Interstitial Lung Disease. *Lung*. 2018 Aug;196(4):409–16.
326. Elhai M, Meune C, Boubaya M, Avouac J, Hachulla E, Balbir-Gurman A, et al. Mapping and predicting mortality from systemic sclerosis. *Ann Rheum Dis*. 2017 Nov;76(11):1897–905.
327. Rubio-Rivas M, Royo C, Simeón CP, Corbella X, Fonollosa V. Mortality and survival in systemic sclerosis: systematic review and meta-analysis. *Semin Arthritis Rheum*. 2014 Oct;44(2):208–19.
328. Goh NSL, Desai SR, Veeraraghavan S, Hansell DM, Copley SJ, Maher TM, et al. Interstitial lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med*. 2008 Jun 1;177(11):1248–54.
329. Mackintosh JA, Stainer A, Barnett JL, Renzoni EA. Systemic Sclerosis Associated Interstitial Lung Disease: A Comprehensive Overview. *Semin Respir Crit Care Med*. 2019 Apr;40(2):208–26.
330. Khanna D, Nagaraja V, Tseng C-H, Abtin F, Suh R, Kim G, et al. Predictors of lung function decline in scleroderma-related interstitial lung disease based on high-resolution computed tomography: implications for cohort enrichment in systemic sclerosis-associated interstitial lung disease trials. *Arthritis Res Ther*. 2015 Dec 23;17:372.

331. Bocchino M, Bruzzese D, Argiento P, Borgia A, Capaccio A, Romeo E, et al. Performance of a new quantitative computed tomography index for interstitial lung disease assessment in systemic sclerosis. *Sci Rep.* 2019 Jul 1;9.
332. Li K, Fang Y, Li W, Pan C, Qin P, Zhong Y, et al. CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19). *Eur Radiol.* 2020 Aug 1;30(8):4407–16.
333. Chassagnon G, Vakalopoulou M, Battistella E, Christodoulidis S, Hoang-Thi T-N, Dangeard S, et al. AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia. *Med Image Anal.* 2021 Jan 1;67:101860.
334. Distler O, Volkman ER, Hoffmann-Vold AM, Maher TM. Current and future perspectives on management of systemic sclerosis-associated interstitial lung disease. *Expert Rev Clin Immunol.* 2019 Oct;15(10):1009–17.
335. Aerts HJWL. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. *JAMA Oncol.* 2016;2(12):1636–42.
336. Lambin P, Zindler J, Vanneste BGL, De Voorde LV, Eekers D, Compter I, et al. Decision support systems for personalized and participative radiation oncology. *Adv Drug Deliv Rev.* 2017 Jan 15;109:131–53.
337. Choe J, Lee SM, Do K-H, Lee G, Lee J-G, Lee SM, et al. Deep Learning-based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses. *Radiology.* 2019 Aug;292(2):365–73.
338. He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep.* 2016 Oct 10;6:34921.