

Titre: Prédiction de survie du cancer du poumon par imagerie 4DCT et réseaux génératifs lors d'un traitement par radiothérapie
Title: Prediction of lung cancer survival using 4DCT imaging and generative networks during radiotherapy treatment

Auteur: Emilie Ouraou
Author:

Date: 2024

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Ouraou, E. (2024). Prédiction de survie du cancer du poumon par imagerie 4DCT et réseaux génératifs lors d'un traitement par radiothérapie [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/62010/>

Document en libre accès dans PolyPublie Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/62010/>
PolyPublie URL:

Directeurs de recherche: Samuel Kadoury
Advisors:

Programme: Génie informatique
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Prédiction de survie du cancer du poumon par imagerie 4DCT et réseaux
génératifs lors d'un traitement par radiothérapie**

EMILIE OURAOU

Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie informatique

Septembre 2024

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Prédiction de survie du cancer du poumon par imagerie 4DCT et réseaux génératifs lors d'un traitement par radiothérapie

présenté par **Emilie OURAOU**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Tarek OULD-BACHIR, président

Samuel KADOURY, membre et directeur de recherche

Hervé LOMBAERT, membre

DÉDICACE

*À mon père,
qui m'a donné l'envie de travailler dans ce domaine... .*

REMERCIEMENTS

Je souhaiterais tout d'abord remercier mon directeur de recherche, Samuel Kadoury, de m'avoir donné l'opportunité de travailler sur ce projet de recherche, pour sa supervision et ses conseils.

Je remercie également les membres du jury, Tarek Ould-bachir et Hervé Lombaert, d'avoir accepté d'évaluer ce mémoire.

Je voudrais également remercier Houda Bahig et Marion Tonneau de m'avoir partagé leur expertise médicale.

Je tiens également à remercier tous les membres du laboratoire medICAL pour l'aide qu'ils m'ont apportée ainsi que pour les moments de détente que nous avons partagés.

Mon séjour au Québec n'aurait pas été le même sans la présence de mes amis et colocataires que j'y ai rencontrés et tous les bons moments que nous avons vécus ensemble. Merci également à Ron et Paloma dont la présence a égayé notre maison.

Finalement, j'aimerais remercier ma famille et mon copain de m'avoir soutenue et encouragée tout au long de ce mémoire malgré la distance.

RÉSUMÉ

La radiothérapie externe est une méthode de traitement privilégiée pour traiter les cas de cancer du poumon, car elle permet d'éviter l'opération de patients âgés ou ne souhaitant pas être opérés. Cette méthode nécessite cependant de prendre en compte le mouvement de la tumeur au cours de la respiration pour mieux la cibler durant le traitement, épargnant ainsi les tissus sains tout en détruisant les cellules cancéreuses.

Bien que le mouvement respiratoire soit de mieux en mieux intégré dans la planification du traitement, sa gestion lors du traitement est encore compliquée. Les méthodes actuelles consistent principalement à augmenter les marges de sécurité du faisceau, en prenant le risque de s'assurer de cibler la tumeur au prix d'une toxicité accrue pour les cellules saines. Une deuxième stratégie possible est de demander au patient de retenir sa respiration, ou utiliser des compressions abdominales pour limiter le mouvement respiratoire, ce qui restreint le confort du patient. Une autre méthode plus récente consiste à planter des marqueurs au sein du patient afin de suivre la tumeur en temps réel. Cette dernière méthode représente une avancée considérable, mais nécessite une chirurgie locale pour planter les marqueurs au sein du patient. Dernièrement, il est possible d'utiliser l'imagerie pour déterminer la position de la tumeur en temps réel et repositionner le faisceau de radiation en conséquence. Cependant les délais de traitement et de remplacement ne permettent pas réellement le temps réel. Pour réaliser un suivi en temps réel de la tumeur il est nécessaire de prédire ses positions à l'avance afin de disposer du temps nécessaire pour remplacer le faisceau. Les développements récents de l'intelligence artificielle permettent de répondre à cette problématique.

Nous pouvons également nous interroger sur l'impact que ce mouvement pourrait avoir sur l'efficacité du traitement. Il a été montré qu'un meilleur ciblage de la tumeur durant le traitement permet de mieux protéger les tissus sains, ce qui permet en retour d'augmenter la dose apportée à la tumeur. Cependant très peu d'études démontrent l'impact du mouvement de la tumeur sur la survie du patient. Nous supposons alors que le mouvement de la tumeur impacte l'efficacité du traitement, et par conséquent le risque de récidive, un mouvement tumoral plus élevé étant lié à davantage de risque de récurrence. Il a également été prouvé que les récurrences des cancers du poumon progressent vite, et que leur détection rapide améliore les chances de survie de patients. On voit alors l'intérêt de prédire au mieux les risques de récurrences à l'issue du traitement, afin d'assurer une meilleure prise en charge des patients.

Face à ce constat, nous cherchons tout d'abord à développer un modèle déterminant les

déformations du poumon au sein du cycle respiratoire, dans l'idée de permettre un meilleur suivi de la tumeur durant la radiothérapie. Nous étudions ensuite le risque de récidive ou décès des patients pour mettre en avant les patients à risque et nous montrons l'importance de l'information portant sur le mouvement de la tumeur pour réaliser cette prédition.

Ces deux études sont réalisées sur une base de données interne au Centre hospitalier de l'Université de Montréal regroupant des patients atteints de tumeurs du poumon traités par radiothérapie entre 2009 et 2019. Chaque étude a été réalisée sur un sous-ensemble de cette base de données. La prédition des déformations du poumon utilisait les scans tomodensitométriques en quatre dimensions de 165 patients. Chacun de ces scans est constitué de 10 volumes pulmonaires acquis au cours d'un cycle respiratoire. La prédition de l'issue du traitement s'est effectuée sur un ensemble de 381 patients, à partir des scans tomodensitométriques, acquis au début du cycle respiratoire, et des données cliniques récoltées par les médecins.

Le premier objectif de déterminer les déformations du poumon au cours de la respiration est réalisé à l'aide d'un modèle de type encodeur-décodeur combinant trois branches encodeuses. La prédition de la dernière phase d'un cycle respiratoire à l'aide des données des phases précédentes atteint une corrélation croisée de 77% entre les volumes réels et prédis sur le jeu de test, avec un temps de calcul moyen de 63 ms, compatible avec le suivi en temps réel.

Le second objectif de prédition de l'issue du traitement est ensuite réalisé, en considérant les récurrences loco-régionales, la survie sans progression et la survie globale des patients deux ans après leur traitement. La solution proposée compare différents modèles (forêts aléatoires, machines à support de vecteur, régression logistique et réseaux convolutifs) disposant en entrée des données cliniques, des données d'imagerie, de l'information sur l'amplitude de mouvement de la tumeur ou de combinaisons de ces différentes informations, dans le but d'évaluer l'influence du mouvement sur la qualité de la prédition. Un test statistique montre que les modèles disposant de l'amplitude de la tumeur ont significativement de meilleures performances que les autres, atteignant une aire sous la courbe de 71% lors de la prédition des risques de récurrences locorégionales 2 ans après la fin du traitement. Le mouvement de la tumeur semble donc impacter les performances du traitement.

Cette étude présente donc un fort potentiel pour l'étude clinique. Le premier modèle est capable de générer des scans qui ont pu être perdus, et ainsi permet d'éviter de devoir écarter des patients d'une étude. Étant donné le faible nombre de patients disponibles en général dans les études médicales, cette application peut s'avérer très utile. Notre second modèle, en démontrant l'importance du mouvement sur la survie du patient, pourrait mener à davantage de suivi pour les patients dont le mouvement n'a pas été corrigé au cours du

traitement, ce qui pourrait améliorer les chances de survie de ces patients.

ABSTRACT

External radiotherapy is one of the treatment of choice for treating lung cancer as it allows avoiding surgery on older or otherwise ineligible patients. However this method needs to take into account the tumour movement during respiration in order to improve tumour targeting during treatment, consequently avoiding unnecessary damages to healthy tissues while destroying cancerous cells.

Respiratory movement is increasingly better integrated into treatment planning, yet dealing with it during treatment delivery is still an issue. Current methods mainly consist of increasing safety margins in the radiation beam, which chooses to make sure to accurately target the tumour at the expense of an increased toxicity for healthy tissues. A second possible strategy is to ask the patient to hold their breath or use abdominal compression to limit their respiratory motion, which reduces the patient's comfort. Another more recent method consists of implanting fiducial markers in the patient in order to follow the tumour in real time. This last method represents a significant improvement but nonetheless requires to perform local surgery to implant the markers. Finally, medical imaging can be used to infer real-time position of the tumour and readjust the radiation beam adequately. However latencies in the calculation and replacement steps do not allow for a real-time following of the tumour. To do so, it is necessary to predict the tumour's future positions in advance in order to replace the beam in time. Recent developments of artificial intelligence allow to answer this issue.

This leads us to question the impact of this movement on the treatment's efficacy. Tumour tracking was proved to better protect healthy tissues, which in turn allows to increase the dose applied to the tumour. However, very few studies investigate the tumour's motion impact on a patient's survival. We hypothesize that this motion impacts treatment efficacy, and consequently the risk of recurrences, a greater tumour motion being linked to an increase in recurrence risk. It was also proved that lung cancer recurrences progress quickly, and therefore detecting them early increases patient survival. This shows the importance of predicting recurrence risks at the end of treatment, in order to assure better patient care.

Given those observations, we first seek to develop a model predicting future lung deformations during the respiratory cycle, in order to enable a better tumour tracking during radiation therapy. Afterwards, we focus on predicting the risks of relapses or deaths to flag at-risk patients and we point out the pertinence of the information regarding the tumour motion to accurately perform this prediction. Those two studies are performed on an internal database

from the *Centre hospitalier de l'Université de Montréal* consisting of patients treated for lung tumours using radiation therapy between 2009 and 2019. Each study was performed on a subset of this database. Lung deformation prediction used computed tomography four-dimensional scans of 165 patients. Each of these scans consists of 10 pulmonary volumes imaged during a breathing cycle. Treatment outcome prediction was performed on a subset of 381 patients. This subset contained computed tomography scans acquired at the beginning of the respiratory cycle and clinical information acquired by the doctors.

The first objective, determining lung deformations during respiration, was performed using an encoder-decoder architecture combining three encoding branches. Predicting the last phase of a respiratory cycle using previous phases' images obtains a cross-correlation of 77% on the test set, with an average computation time of 63 ms, which is compatible with a real-time following of the tumour.

The second objective, predicting the treatment outcome (loco-regional recurrence, progression-free survival and overall survival two years after their treatment) was then performed. The proposed solution compares various models (random forests, support vector machines, logistic regressions and convolutional neural networks) using as input data combinations of the clinical data, the imaging data and the amplitude of the tumour movement, in order to evaluate the influence of the movement on the quality of the prediction. A statistical test shows that the models using the tumour's movement amplitude have significantly better performances than those without access to this information, achieving an area under the curve of 71% when predicting loco-regional recurrence risks 2 years after the end of treatment. Consequently tumour movement seems to impact treatment performances.

This study offers a strong potential for clinical applications. Our first model is capable of generating lost scans, which avoids having to exclude patients from a study. Given the limited datasets generally available for medical studies, this application can be quite useful. Our second model, by demonstrating the importance of tumour motion for patients' survival, could lead to closer follow-up for patients whom tumour motion was not corrected during treatment. This could lead to better patient survival.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	viii
TABLE DES MATIÈRES	x
LISTE DES TABLEAUX	xiii
LISTE DES FIGURES	xiv
LISTE DES SIGLES ET ABRÉVIATIONS	xvii
CHAPITRE 1 INTRODUCTION	1
1.1 Définitions et concepts de base	1
1.2 Objectifs de recherche	2
1.3 Plan du mémoire	2
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 Les cancers	4
2.2 Cancer du poumon	4
2.3 Radiothérapie	5
2.3.1 Principe de fonctionnement	5
2.3.2 Défis	6
2.3.3 Exemple de machines de radiothérapie	6
2.3.4 Déroulement du traitement	8
2.3.5 Gestion de la respiration	9
2.4 Imagerie tomodensitométrique	10
2.5 Radiomiques	12
2.6 Apprentissage machine et apprentissage profond	15
2.6.1 Réseaux de neurones	16
2.6.2 Apprentissage machine	22
2.6.3 Stratégie d'entraînement	25

2.6.4	Métriques	25
2.7	Etat de l'art	30
2.7.1	Modélisation du mouvement du poumon	30
2.7.2	Prédiction de survie et récurrences des cancers du poumon non à petites cellules	34
2.8	Synthèse	38
CHAPITRE 3 MÉTHODOLOGIE DU TRAVAIL DE RECHERCHE		39
3.1	Jeux de données utilisés	40
3.2	Étape de prédiction des déformations	41
3.2.1	Explication de la démarche	41
3.2.2	Validation de la démarche	41
3.3	Étape de prédiction des issues du traitement	41
3.3.1	Explication de la démarche	41
3.3.2	Validation de la démarche	42
CHAPITRE 4 PRÉDICTION DE 4DCT EN RESPIRATION LIBRE DE DÉFORMATIONS DU POUMON EN UTILISANT DES AUTO-ENCODEURS PROBABILISTES		43
4.1	Etat de l'art	43
4.1.1	Modèles à patient unique	43
4.1.2	Modèles de population	43
4.2	Méthodes	44
4.2.1	Protocole d'acquisition	44
4.2.2	Prétraitement des données	44
4.2.3	Définition du modèle	45
4.2.4	Entraînement du modèle	47
4.3	Résultats	48
4.3.1	Génération des DVF	48
4.3.2	Modèle général	48
4.4	Discussion	49
4.5	Conclusion	51
CHAPITRE 5 ARTICLE 1 : PREDICTING EARLY STAGE LUNG CANCER RECURRENCE AND SURVIVAL FROM COMBINED TUMOR MOTION AMPLITUDE AND RADIOMICS FEATURES ON FREE-BREATHING 4D-CT		52
5.1	Introduction	53

5.2	Materials and Methods	55
5.2.1	Patient cohort	55
5.2.2	CT scanning and treatment protocol	56
5.2.3	Radiomic features	57
5.2.4	Random forest prediction model	58
5.2.5	Training strategy	58
5.3	Results	59
5.3.1	Acquired dataset	59
5.3.2	Statistical analysis	60
5.3.3	Clinical features and clinical outcomes	63
5.3.4	Reproducibility to the CK population	64
5.3.5	Predicting overall survival and loco-regional recurrences	65
5.4	Discussion	66
5.5	Conclusion	70
CHAPITRE 6 DISCUSSION GÉNÉRALE		71
6.1	Autour de la prédiction des déformations du poumon	71
6.2	Autour de la prédiction des issues du traitement et de l'importance du mouvement	72
6.3	Perspectives futures	73
6.4	Intégration pour l'utilisation en clinique	74
CHAPITRE 7 CONCLUSION		75
7.1	Synthèse des travaux	75
7.2	Limitations de la solution proposée	75
RÉFÉRENCES		77

LISTE DES TABLEAUX

Tableau 2.1	Exemple d'une matrice de confusion dans un cas de classification binaire.	26
Tableau 4.1	Performances numériques obtenues par le modèle de prédiction des déformations lors de la prédiction de la dernière phase du cycle respiratoire.	49
Tableau 5.1	Clinical and treatment characteristics of the eligible patients cohort after exclusion criteria. Numbers are presented as <i>frequency (percentage)</i> when indicated, <i>median (min–max)</i> otherwise. SBRT : stereotactic body radiation therapy. VMAT : volumetric modulated arc therapy. CK : CyberKnife. TOMO : helical tomotherapy. FEV1 : forced expiratory volume in 1 second. DLCO : diffusing capacity of the lungs for carbon monoxide. KPS : Karnofsky performance status. BED ₁₀ : biologically effective doses.	68
Tableau 5.2	The significant features found during multivariate analysis. For all of them, $p < 0.005$. HR : hazard ratio. CI : confidence interval. FEV1 : forced expiratory volume in 1 second. DLCO : diffusing capacity of the lungs for carbon monoxide.	69

LISTE DES FIGURES

Figure 2.1	Vue en 3D des poumons d'un patient atteint de NSCLC	5
Figure 2.2	Système de traitement LINAC.	7
Figure 2.3	Système de traitement CyberKnife	8
Figure 2.4	Exemple d'une tranche d'un CT scan des poumons présentant les structures d'intérêt (GTV, IGTV et PTV).	12
Figure 2.5	Exemple d'image. a) Image en niveaux de gris discréétisés. b) Représentation numérique de l'image a.	13
Figure 2.6	Exemple de matrices de co-occurrences. a) Représentation numérique de l'image d'exemple. b) Matrice de co-occurrences dans la direction horizontale (0°). c) Matrice de co-occurrences dans la direction verticale (90°).	14
Figure 2.7	Exemple de matrices GLRLM. a) Représentation numérique de l'image d'exemple. b) Matrice GLRLM dans la direction horizontale (0°). c) Matrice GLRLM dans la direction verticale (90°).	14
Figure 2.8	Exemple de matrices GLSZM. a) Représentation numérique de l'image d'exemple. b) Matrice GLSZM associée.	15
Figure 2.9	Schéma de l'architecture d'un réseau Perceptron multicouches	17
Figure 2.10	Schéma de l'architecture d'un réseau CNN.	18
Figure 2.11	Schéma de l'architecture d'un auto-encodeur	20
Figure 2.12	Schéma de l'architecture d'un Transformer.	22
Figure 2.13	Représentation schématique d'un arbre de décision fictif.	24
Figure 2.14	Schéma d'un exemple de validation croisée. Les données encadrées constituent le jeu de test de l'itération, pendant que les autres données forment le jeu d'apprentissage.	26
Figure 2.15	Exemple de courbe ROC comparant deux modèles différents.	28
Figure 3.1	Schéma du processus développé au cours de ce mémoire. Le premier modèle génère les volumes et phases respiratoires manquants à notre jeu de données. Le second modèle utilise les données cliniques, radiomiques et l'amplitude du mouvement de la tumeur pour prédire l'état de santé du patient deux ans après la fin de son traitement.	40
Figure 4.1	Architecture du modèle de prédiction des déformations.	45
Figure 4.2	Architecture du modèle de prédiction des déformations dans sa configuration de test.	47

Figure 4.3	Comparaison des volumes des phases 90% réelles et prédictes par le modèle pour différentes tranches et orientations pour deux patients différents.	50
Figure 4.4	Exemple de calcul d'erreur de recalage de la tumeur.	50
Figure 5.1	A sample of a 4D-CT scan on a coronal slice. The 4D-CT consists of 10 respiratory phases, with the first one (phase 0%) taken at the middle of the expiration phase on which we extracted the radiomic features. The PTV tumor segmentation is represented on each phase in yellow. The blue lines show the vertical motion amplitude of the tumor and the green point represents its center.	57
Figure 5.2	Global pipeline used for the prediction. We use 20x data augmentation on the CTs to have access to a more diverse database. Radiomics features are extracted from this augmented dataset. Clinical data are one-hot encoded for categorical data and numeric features are normalized. Recurrence or death cases are over-sampled in the training phase to better learn to flag at-risk cases. Clinical and radiomic features are selected by ANOVA and a classifier (SVM, LR or RF) predicts the patient outcome.	59
Figure 5.3	Schematic illustration of the dataset splitting strategy. We performed a nested cross-validation consisting of 6 outer folds and 6 inner folds.	60
Figure 5.4	Kaplan-Meier curves of the three most important clinical features (motion, FEV1 and gender) as well as the model prediction.	61
Figure 5.5	Mean AUC scores and AUC distribution for each classifier and each data configuration for PFS prediction.	62
Figure 5.6	Comparative performances of RF, SVM and LR models with input features from various combinations of clinical, radiomics and motion features for PFS prediction. The Cumming plots compare the four data configuration AUC on 20 random training/test splits for each classifier.	63
Figure 5.7	Importance of each feature for the outcome prediction. The importance was evaluated using the Gini importance criteria, indicating motion is the most important feature, highlighting its importance to be taken into consideration.	64

Figure 5.8	Feature permutation importance for the selected features in the outcome prediction. Feature permutation tests, evaluating the feature importance using a feature permutation process repeated 120 times. Motion is the second most important feature, highlighting its contribution to the risk classification.	65
Figure 5.9	Average AUC measures on the 6 test splits for the PFS, LRR and OS prediction models for all combinations of input data and classifiers.	66

LISTE DES SIGLES ET ABRÉVIATIONS

4D	4 Dimensions
3D	3 Dimensions
2D	2 Dimensions
1D	1 Dimension
NSCLC	Cancer du poumon non à petites cellules (<i>Non Small Cell Lung Cancer</i>)
SBRT	Radiothérapie stéréotaxique (<i>Stereotactic Body Radiation Therapy</i>)
LINAC	Accélérateur linéaire médical (<i>Medical Linear Accelerator</i>)
VMAT	Arcthérapie volumétrique modulée (<i>Volumetric Modulated Arc Therapy</i>)
TOMO	Tomothérapie
CK	CyberKnife
CT	Tomodensitométrie (<i>Computed Tomography</i>)
IRM	Imagerie par Résonance Magnétique
GTV	Volume tumoral macroscopique (<i>Gross Tumour Volume</i>)
IGTV	Volume tumoral macroscopique interne (<i>Internal Gross Tumour Volume</i>)
CTV	Volume cible anatomoclinique (<i>Clinical Target Volume</i>)
PTV	Volume cible prévisionnel (<i>Planning Target Volume</i>)
MLP	Perceptron multicouche (<i>Multi-Layer Perceptron</i>)
CNN	Réseau de neurones convolutifs (<i>Convolutional Neural Network</i>)
STN	Module de transformation spatiale (<i>Spatial Transformer Networks</i>)
RNN	Réseau de neurones récurrents (<i>Recurrent Neural Networks</i>)
LSTM	Réseau récurrent à mémoire court et long terme (<i>Long Short-Term Memory</i>)
ConvLSTM	Réseau convolutionnel à mémoire court et long terme (<i>Convolutional Long Short-Term Memory</i>)
GRU	Réseau unité récurrente fermée (<i>Gated Recurrent Unit</i>)
GAN	Réseau antagoniste génératif (<i>Generative Adversarial Network</i>)
cGAN	Réseau antagoniste génératif conditionnel (<i>Conditional Generative Adversarial Network</i>)
AE	Auto-encodeur
VAE	Auto-encodeur variationnel
ViT	Transformer de vision (<i>Vision Transformer</i>)

LR	Régression logistique (<i>Logistic Regression</i>)
SVM	Machine à vecteur de support (<i>Support Vector Machine</i>)
RF	Forêt aléatoire (<i>Random Forest</i>)
PCA	Analyse par composantes principales (<i>Principal Component Analysis</i>)
TP	Vrais Positifs
TN	Vrais négatifs
FN	Faux négatifs
FP	Faux positifs
ROC	Caractéristique de fonctionnement du récepteur (<i>Receiver Operating Characteristic</i>)
AUC	Aire sous la courbe (<i>Area Under the Curve</i>)
CC	Corrélation Croisée (<i>Cross Correlation</i>)
NCC	Corrélation Croisée Négative (<i>Negative Cross Correlation</i>)
MSE	Erreur Quadratique Moyenne (<i>Mean Square Error</i>)
RMSE	Écart Quadratique Moyen (<i>Root Mean Square Error</i>)
SSIM	Similarité Structurelle (<i>Structural Similarity</i>)
TRE	Erreur de recalage de cibles (<i>Target Registration Error</i>)
DVF	Champs de vecteur de déplacement (<i>Displacement Vector Field</i>)

CHAPITRE 1 INTRODUCTION

1.1 Définitions et concepts de base

Le cancer du poumon est l'un des types de cancer les plus diagnostiqués au monde, avec 2,21 millions de cas diagnostiqués mondialement en 2020, ainsi que le plus meurtrier (et de loin) avec 1,79 million de décès mondialement en 2020 [1].

Le traitement le plus courant contre ce type de cancer est la chirurgie, qui va viser à retirer la tumeur. Mais du fait de l'âge souvent avancé des patients, nombre d'entre eux sont inopérables. Dans ce cas la radiothérapie externe est le traitement privilégié par les oncologues [2, 3]. Le traitement par radiothérapie consiste à soumettre la tumeur à une dose d'irradiation dans le but de détruire les cellules tumorales. Ce traitement est toutefois compliqué par les mouvements respiratoires du patient, qui diminuent la précision de la focalisation du faisceau de radiation sur la tumeur. Ceci entraîne le risque de causer des dommages à des tissus sains ou au contraire de ne pas suffisamment affecter la tumeur, au risque de ne pas complètement traiter le cancer. Différentes méthodes peuvent être mises en place pour pallier ce problème. Il est possible d'augmenter les marges d'irradiation, mais souvent cela est peu souhaitable du fait des dommages que cela pourrait entraîner. Une autre possibilité est de réduire les mouvements respiratoires en limitant l'amplitude des mouvements de respiration ou en bloquant temporairement la respiration du patient. Cependant ces méthodes sont peu confortables pour le patient et ne sont pas toujours réalisables du fait de leur état de santé. Plus récemment, on peut aussi utiliser l'imagerie médicale pour suivre la position de la tumeur au cours du traitement. Cette stratégie nécessite de pouvoir réagir suffisamment vite pour repositionner au besoin le faisceau de radiation en cas de mauvaise correspondance avec la position de la tumeur. Une des façons possibles de réaliser cela est de prédire en avance les futures positions de la tumeur, afin de disposer de davantage de temps pour repositionner le faisceau au besoin. Le développement de l'apprentissage profond est une bonne piste pour développer ces méthodes de suivi.

Cependant, malgré les avancées des traitements, les cas de nouvelles occurrences du cancer ou de décès du patient sont encore communs. Il est estimé que deux ans après le traitement, environ 5% des patients sont atteints de récurrences locales (pour lesquelles la nouvelle tumeur se trouve dans le même lobe pulmonaire), 10% sont atteints de récurrences régionales (dans ce cas la nouvelle tumeur apparaît dans le même poumon) et 15% des patients souffrent de récurrences distantes (survenant dans le poumon controlatéral ou bien d'autres régions du corps). Le taux de décès est de 30% environ [3–5].

Lorsqu'on traite des cas de cancer, le temps de réaction peut être primordial, car plus la tumeur a eu le temps de se développer plus les risques encourus par le patient sont élevés. L'amélioration de la prise en charge des cancers nécessite donc de mieux évaluer les risques de chaque patient afin de prévenir au plus tôt les cas de récidives du cancer. Il est toutefois difficile de différencier les patients à risque étant donné que les facteurs cliniques associés à ces risques sont encore mal connus et liés à des relations internes complexes. Les prédictions de l'issue des cancers des poumons sont donc encore peu précises. Par exemple les médecins ont tendance à surestimer la durée de survie à court terme mais à sous-estimer le temps de survie à long terme [6].

On voit donc l'intérêt d'utiliser des solutions liées à l'intelligence artificielle, qui est de plus en plus utilisée pour la prédiction des effets à long terme de traitements. En effet les algorithmes d'apprentissage automatique permettent d'identifier des motifs récurrents cachés et d'en déduire un modèle permettant de prédire l'évolution de la maladie du patient [7].

Toutefois très peu de travaux se sont concentrés sur les conséquences du mouvement de la tumeur sur l'efficacité du traitement. La limitation ou la prise en compte du mouvement de la tumeur est pourtant une composante principale de la mise en place du traitement par radiothérapie.

1.2 Objectifs de recherche

L'objectif principal de ce mémoire est d'étudier l'influence du mouvement de la tumeur au cours de la respiration sur le succès du traitement chez un patient. Pour cela nous commencerons par étudier le mouvement respiratoire pulmonaire au cours d'un cycle respiratoire. L'objectif sera de prédire les déformations successives du poumon afin de pouvoir déterminer en temps réel les futures positions d'une tumeur au sein du cycle respiratoire. Ensuite, nous nous concentrerons sur la prédiction des chances de succès du traitement par radiothérapie. Un premier sous-objectif sera d'aboutir à un modèle permettant de classifier les patients à risque de récidive ou de décès après traitement. Un second sous-objectif consistera à évaluer l'influence de l'apport d'information concernant le mouvement tumoral sur la qualité de la prédiction, afin d'évaluer l'importance de prendre en compte le mouvement de la tumeur lors de la classification des patients à risque.

1.3 Plan du mémoire

Ce mémoire de maîtrise est organisé de la manière suivante. Dans le chapitre 2, une revue de la littérature portant sur les différents aspects théoriques de cette recherche est présen-

tée. Ce chapitre aborde les thèmes du cancer du poumon, de la radiothérapie, de l'imagerie tomodensitométrique, des radiomiques, et les principales métriques utilisées pour évaluer les performances de modèles d'apprentissage. On y abordera également diverses approches d'apprentissage automatique et d'apprentissage profond permettant l'estimation du mouvement ainsi que la prédiction des résultats du traitement. Dans le chapitre 3, la méthodologie de ce travail de recherche est exposée. Les chapitres 4 et 5 présentent les modèles développés au cours de ce travail de recherche. Dans le chapitre 4 on trouvera un modèle de génération de phases respiratoires manquantes dans un cycle respiratoire, et au chapitre 5 le modèle de prédiction des risques de récidives et décès 2 ans après le traitement par radiothérapie. Le chapitre 6 portera sur une discussion générale des solutions proposées et exposera leurs limites. Enfin le chapitre 7 sera la conclusion de ce travail.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 Les cancers

Le cancer désigne un regroupement de maladies nombreuses et diverses qui résultent d'un dysfonctionnement de certaines cellules de l'organisme. Ces cellules, initialement normales, développent des erreurs lors de la réPLICATION de leur matériel génétique, et au lieu de réparer ces mutations ou de s'autodétruire, elles continuent de se multiplier de manière anarchique. Selon l'Institut National du Cancer [8], on considère que ce phénomène apparaît lorsque plus d'une dizaine de mutations ont lieu. Ces cellules anormales prolifèrent d'abord localement pour former une tumeur, puis à plus grande distance, créant ainsi de nouvelles tumeurs à distance appelées métastases.

Les mutations à l'origine du cancer apparaissent de façon aléatoire, bien que des facteurs de risque existent. On y retrouve notamment le tabac, l'alcool, l'alimentation, l'exposition à la pollution ou bien l'hérédité. Du fait des modes de vie et du vieillissement de la population le nombre de nouveaux cas de cancer tend à augmenter [8, 9]. Les cancers étaient la première ou deuxième plus grande cause de décès avant l'âge de 70 ans pour 112 des 183 pays étudiés par l'Organisation mondiale de la Santé en 2019 [9].

2.2 Cancer du poumon

Parmi l'ensemble des cas de cancer, les cancers du poumon sont le deuxième type de cancer le plus couramment diagnostiqué, représentant 11.4% des cas de cancer [9]. Il s'agit également du cancer le plus meurtrier, étant donné qu'il représente 18% des cas de décès par cancer, soit environ 1.8 million de décès en 2020 [9]. Il est estimé qu'au moins les deux tiers des décès dus à un cancer du poumon sont attribuables au fait de fumer [9–11], avec un taux d'incidence et de mortalité environ 2 fois plus élevés chez les hommes que chez les femmes. La pollution de l'air augmente également le risque de cancer du poumon [9]. La majorité des patients atteints par ce type de cancer sont relativement âgés (entre 65 et 84 ans), avec un âge moyen au moment du diagnostic de 66 ans [10, 11]. Les symptômes les plus fréquents sont la toux, l'hémoptysie, des douleurs dans la poitrine et l'essoufflement [10]. Le taux de survie 5 ans après diagnostic varie entre 10 et 20% dans la plupart des pays [9]. Dans la majorité des cas, les cancers du poumon sont diagnostiqués à un stade avancé, ce qui impacte négativement les chances de survie [11, 12].

Les cancers du poumon sont répartis en deux formes : cancers du poumon à petites cellules et

cancers du poumon non à petites cellules (pour *Non Small Cell Lung Cancer*) (NSCLC), qui représentent environ 85% des cas [10] et sur lesquels nous nous concentrerons dans cette étude. Eux-mêmes sont divisés en trois types : adénocarcinomes (les plus répandus), carcinomes épidermoïdes et carcinomes à grandes cellules [10].

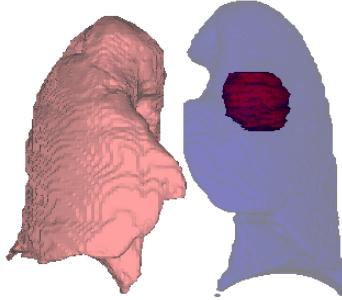


FIGURE 2.1 Vue en 3D des poumons d'un patient atteint de NSCLC. On y voit les poumons droit (en rose) et gauche (en bleu) ainsi que de la tumeur (en rouge). Il s'agit d'un adénocarcinome situé dans le lobe supérieur gauche.

Les patients atteints d'un cancer du poumon sont traités par chirurgie (ablation de la tumeur), chimiothérapie ou radiothérapie, ou bien en combinant plusieurs de ces traitements [10, 11]. Du fait de l'âge avancé des patients, il arrive régulièrement qu'ils ne soient pas opérables à cause de leur état de santé général. Dans ce cas le traitement de référence est la radiothérapie [2, 3]. Dans le cadre de cette étude nous nous limiterons aux cas de traitements par radiothérapie externe.

2.3 Radiothérapie

2.3.1 Principe de fonctionnement

Le traitement par radiothérapie consiste à envoyer de forts rayons ionisants de façon à abîmer le matériel génétique des cellules cancéreuses et ainsi empêcher leur multiplication et entraîner leur autodestruction [13, 14]. Dans environ 90% des cas, l'onde envoyée est constituée de rayons X, mais il peut aussi s'agir d'électrons, ou plus rarement encore de protons [14]. Cette radiation va interagir avec les électrons présents dans les cellules cibles. Elle peut introduire une ionisation directement en éjectant les électrons de l'atome, ou alors en excitant les électrons pour les faire monter à un autre niveau d'énergie ce qui leur permet d'exciter à leur tour d'autres atomes et ainsi créer de nouvelles ionisations. Ceci abîme le matériel

génétique de la cellule qui cherche à se répliquer et finira par entraîner sa destruction [15, 16].

2.3.2 Défis

Malgré la précision apportée à la focalisation sur la tumeur, des cellules saines sont aussi affectées par la radiation, ce qui entraîne des effets secondaires. Ceci est d'autant plus vrai dans le poumon, où la respiration entraîne un déplacement important de la tumeur pouvant aller jusqu'à 5 cm [17]. Ces effets secondaires peuvent par exemple entraîner de la fatigue, des problèmes de peau, des pertes de poils ou de cheveux. Dans le cadre particulier de la radiothérapie du thorax on peut aussi être confronté à de la toux, des maux de gorge, une difficulté à avaler, une inflammation du poumon, une fibrose pulmonaire ou des troubles cardiaques [13].

Cependant comme les cellules tumorales se divisent plus rapidement, elles sont plus sensibles aux effets de la radiothérapie [13]. L'objectif est donc d'être le plus précis possible en irradiant la tumeur pour trouver le meilleur compromis entre la maximisation de la destruction des cellules cancéreuses et la minimisation de l'effet sur les tissus sains et les organes alentour. L'application des doses se fait par fractions, c'est-à-dire en divisant la dose totale reçue par le patient en l'irradiant au cours de différentes séances où la dose appliquée est une fraction de la dose totale recommandée. Cette méthode permet un meilleur contrôle de la tumeur tout en limitant les effets secondaires, car le temps entre deux radiations permet aux tissus sains de se remettre [15].

2.3.3 Exemple de machines de radiothérapie

La radiothérapie externe utilisée dans notre étude consiste en l'application de rayons ionisants produits par une source externe au patient, au contraire de la radiothérapie interne où la source de radiation est à l'intérieur du patient pour être au plus proche de la zone tumorale. Les radiations doivent donc traverser la peau du patient. Plusieurs techniques de radiothérapie différentes coexistent, chacune avec ses avantages.

LINAC

La machine accélérateur linéaire médical ou *Medical Linear Accelerator* (LINAC) est un accélérateur linéaire de particules. Il s'agit de la machine la plus couramment utilisée en radiothérapie externe [18]. La Figure 2.2 présente un exemple de machine LINAC.

Différentes méthodes d'administration du traitement par cette machine existent. En particulier nous pouvons citer l'archérapie volumétrique modulée (VMAT). Cette méthode nécessite



FIGURE 2.2 Système de traitement LINAC. Image extraite de Wikimedia Commons.

une image précise en trois dimensions de la tumeur. Plusieurs faisceaux de radiation sont dirigés vers la tumeur en provenance de différentes directions afin de la traiter sous tous les angles. La zone de superposition de ces faisceaux reçoit une dose plus forte que les régions alentour, ce qui permet d'administrer une forte dose sur la tumeur tout en limitant l'irradiation des tissus normaux voisins. Ceci est également facilité par la modulation de l'intensité des faisceaux. Les émetteurs de radiation peuvent effectuer une rotation de 360 degrés, ce qui réduit la durée des séances quotidiennes de traitement [18].

Tomothérapie

La tomothérapie (TOMO) consiste à coupler un scanner avec un accélérateur de particules. À la manière d'une numérisation tomodensitométrique (CT), l'accélérateur de particules tourne autour du patient pendant que la table sur laquelle est installé le patient se déplace longitudinalement [14]. L'irradiation est donc fournie selon une forme hélicoïdale, ce qui aide à mieux limiter l'irradiation des tissus aux alentours de la tumeur. Le couplage avec le balayage CT permet également d'ajuster le plan de traitement à la position exacte de la tumeur au début de chaque séance de traitement.

CyberKnife

Le Cyberbistouri ou CyberKnife (CK) est un petit accélérateur linéaire compact tenu par un bras robotique capable de se déplacer dans toutes les directions possibles [14]. Ce système est également pourvu de caméras à rayons X permettant de détecter des déplacements de la tumeur, qui peuvent par exemple être dus à la respiration ou à un mouvement du patient, et

ainsi repositionner l'émetteur [18].

Radiothérapie Stéréotaxique

Dans notre étude les patients sont traités par radiothérapie stéréotaxique (SBRT), administrée par VMAT, TOMO ou CK. Cette méthode de radiothérapie adaptée aux petites tumeurs délivre de fortes doses de radiations sur un faible nombre de séances comparé aux radiothérapies plus classiques. Ceci nécessite des marges restreintes et une immobilisation du patient durant le traitement pour augmenter la dose reçue par la tumeur tout en limitant la dose reçue par les organes alentour. Plusieurs faisceaux de radiation de faible dose sont émis à partir de différents angles et se superposent au niveau de la tumeur, qui reçoit alors une forte dose totale [3, 18, 19].

2.3.4 Déroulement du traitement

Un traitement par radiothérapie se déroule en plusieurs étapes.

La première consiste en la planification du traitement. Il s'agit tout d'abord de déterminer la position qui sera adoptée par le patient tout au long du traitement, le plus souvent allongé sur le dos. Des points de repère peuvent être utilisés afin de garder la même position d'une séance à l'autre. Ensuite une image de la partie du corps à traiter est réalisée, souvent par tomodensitométrie ou imagerie par résonance magnétique (IRM). Cette image est utilisée pour repérer la position de la cible à irradier et celle des organes à risque à éviter [14, 18].

Ces informations permettent de concevoir le plan de traitement du patient. L'oncologue radiothérapeute détermine la technique d'irradiation à utiliser, la dose totale et le nombre



FIGURE 2.3 Système de traitement CyberKnife. Image extraite de Wikimedia Commons.

total de séances, ce qui détermine les doses reçues à chaque séance. Ces doses sont appelées fractions, elles correspondent à la dose totale divisée par le nombre de séances. Le plan de dose définit également la dose à attribuer à chaque organe. Il est déterminé par le type de cancer, son stade, sa sensibilité à la radiothérapie, la taille et l'emplacement de la tumeur, la capacité des tissus alentour à tolérer la radiation, l'état de santé global du patient, ainsi que les traitements précédents et prévus pour la suite du traitement [18].

Une fois le traitement planifié, les séances d'irradiation peuvent débuter. Ces séances durent quelques minutes. En général la procédure totale dure plusieurs semaines, avec des séances de radiothérapie 5 jours par semaine durant 5 à 8 semaines. Les autres jours permettent de laisser aux cellules saines le temps de récupérer [20, 21].

Enfin, une fois le traitement terminé des consultations régulières sont à planifier afin de surveiller l'évolution de la maladie du patient ainsi que le développement de potentiels effets secondaires à la radiothérapie [14].

2.3.5 Gestion de la respiration

La respiration du patient durant le traitement entraîne un mouvement des poumons et donc un déplacement de la tumeur, en particulier dans le cas de tumeurs localisées dans la partie inférieure du poumon. Il est estimé que les tumeurs du poumon peuvent avoir un mouvement d'une amplitude allant jusqu'à 5 cm [17]. Il faut alors prendre en compte ce déplacement pour conserver le faisceau de radiation sur la tumeur, au risque de limiter l'efficacité du traitement. Plusieurs solutions existent pour cela, on peut les regrouper en deux catégories : les approches adaptatives et les approches non adaptatives.

Méthodes non adaptatives

Citons d'abord les méthodes non adaptatives. La première consiste à limiter le mouvement respiratoire du patient en compressant son abdomen à l'aide d'une ceinture, ce qui limite par la même occasion le déplacement de la tumeur [3].

Une autre méthode consiste à augmenter les marges autour de la tumeur afin de traiter une plus grande surface. Cependant cette méthode augmente également les dommages causés aux cellules saines. Pour rester au plus proche des différentes positions de la tumeur on utilise en général un CT scan en 4 dimensions (4D) couvrant un cycle respiratoire du patient sur lequel on détermine le mouvement de la tumeur et on utilise cette information lors de la conception du plan de dose afin que la tumeur reçoive la dose prescrite quelle que soit sa position [2].

Une autre méthode non adaptive appelée *breath hold* (suspension de la respiration) consiste

à demander au patient de retenir sa respiration durant quelques dizaines de secondes à un moment précis de son inspiration et à n'allumer le faisceau de radiation que durant ce moment [2, 14]. Mais cette technique peut présenter des difficultés ou de l'inconfort pour certains patients qui peuvent avoir du mal à retenir leur respiration pendant aussi longtemps et de façon reproductible [2]. Elle augmente aussi le temps de traitement.

Une autre méthode appelée *gating* ne délivre la dose d'irradiation que lorsque la tumeur se trouve devant le faisceau [2, 14]. Cette méthode augmente elle aussi le temps de traitement. De plus, elle nécessite de connaître la position de la tumeur.

Méthodes adaptatives

Le second groupe regroupe les méthodes adaptatives. Elles consistent à suivre la tumeur avec le faisceau de radiation, en déplaçant l'émetteur en temps réel [14]. Mais cela nécessite de connaître la position de la tumeur en temps réel. Pour cela on utilise souvent des marqueurs dont il est plus facile de suivre la position que celle de la tumeur. Cela peut être accompli en introduisant les marqueurs autour de la tumeur, ce qui nécessite une opération et peut causer des risques de pneumothorax [22], ou bien en suivant des marqueurs externes qui ne sont pas toujours parfaitement représentatifs du mouvement de la tumeur [23].

Plus récemment, des réseaux de neurones sont développés afin de prédire les futures positions de la tumeur sans nécessiter de marqueurs, en se basant sur les images du cycle respiratoire normal du patient.

2.4 Imagerie tomodensitométrique

L'imagerie tomodensitométrique ou CT scan utilise les rayons X pour obtenir une image interne du patient.

Diamétralement opposés à la source de rayonnement se trouvent des détecteurs chargés de mesurer l'absorption des rayons par le patient. Cette absorption varie selon la densité et l'épaisseur des structures traversées par les rayons. Par exemple des structures denses comme les os ressortent en blanc sur l'image alors que les tissus mous apparaîtront en différentes teintes de gris. Pour faire ressortir davantage la zone d'intérêt on utilise parfois un agent de contraste contenant davantage de substances absorbant les rayons X, ce qui rend la zone d'intérêt plus claire donc plus visible sur l'image [24, 25].

Le patient est allongé sur un lit qui translate horizontalement pendant que l'émetteur à rayons X tourne autour du patient. Lorsque l'émetteur achève une rotation complète autour

du patient, on a enregistré une image en 2 dimensions (2D).

Ce phénomène est modélisé par la transformée de Radon de la carte des densités d'absorptions, appelée sinogramme. On projette une fonction f représentant les densités d'absorption à chaque angle selon un angle $\theta + \pi/2$. Notons δ l'impulsion de Dirac. La transformée de Radon s'exprime sous la forme :

$$\begin{aligned} g_\theta(R) &= \iint f(x, y) \delta(x \cos \theta + y \sin \theta - R) dx dy \\ &= \int_0^{2\pi} \int_0^\infty f(r, \phi) \delta[r \cos(\theta - \phi) - R] r dr d\phi \end{aligned} \quad (2.1)$$

En pratique on utilise un pas de rotation $\Delta\theta$.

La transformée de Radon permet d'obtenir un sinogramme composé de toutes les projections $g(R, \theta)$ [26]. On est ensuite capable de reconstruire l'image d'origine $f(x, y)$ par filtrage, rétroposition et interpolation.

Ensuite le lit translate légèrement pour permettre d'imager une nouvelle zone du patient située quelques millimètres plus loin. L'épaisseur de chaque image dépend de la machine, mais varie en général entre 1 et 10 mm. Une fois le nombre désiré d'images axiales obtenu, elles sont empilées pour produire une représentation en 3 dimensions (3D) [24, 25].

Avant de réaliser le traitement par radiothérapie, un CT scan 4D (4DCT) du patient est réalisé. Celui-ci enregistre un cycle respiratoire du patient, découpé en 10 phases composées d'un CT scan 3D. Ce 4DCT sert à réaliser la segmentation nécessaire au plan de dose, en déterminant la position des structures d'intérêt. Parmi celles-ci on distingue d'abord le Gross Tumour Volume (GTV) qui représente la tumeur en elle-même. Lorsqu'on utilise l'information 4D pour inclure toutes les positions de la tumeur au cours du cycle respiratoire on parle d'Internal Gross Tumour Volume (IGTV). On ajoute autour du GTV une marge correspondant à une propagation de la maladie non visible sur les images, afin d'être assuré de traiter l'entièreté de la tumeur. Cela forme le Clinical Target Volume (CTV). On ajoute encore à cela une nouvelle marge afin de compenser les incertitudes liées au système de traitement et au mouvement de la tumeur. Cette nouvelle segmentation est appelée Planning Target Volume (PTV) et forme la région cible, celle recevant le maximum de radiation lors du traitement. La Figure 2.4 présente une segmentation de ces différents volumes d'intérêt sur une tranche de CT scan de poumon.

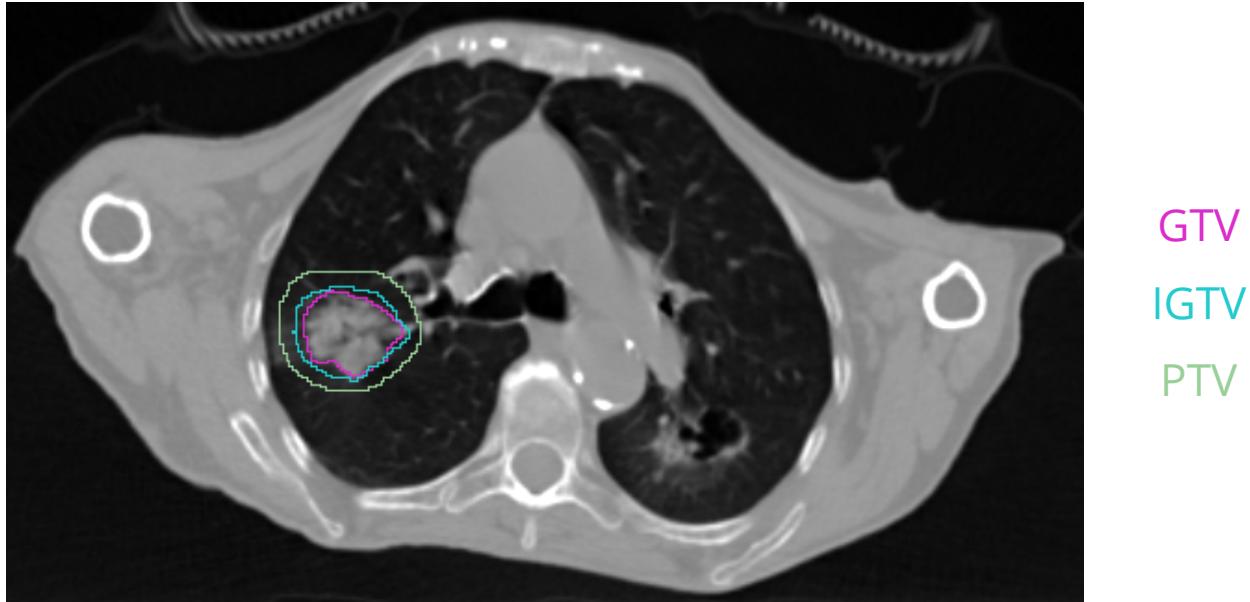


FIGURE 2.4 Exemple d'une tranche d'un CT scan des poumons présentant les structures d'intérêt (GTV, IGTV et PTV).

2.5 Radiomiques

Les radiomiques sont des statistiques calculées sur les images. Elles ont été développées face à l'intuition que l'analyse mathématique des images médicales permettra de faire apparaître des informations pertinentes contenues dans les images, mais non perceptibles directement par l'oeil humain [27]. Elles permettent également une analyse quantifiée des caractéristiques telles que l'intensité, la forme ou la texture, ce qui permet de pallier aux interprétations subjectives des images [28]. Elles présentent également l'avantage de convenir à toutes les modalités d'imagerie et d'être non invasives. Elles sont cependant sensibles aux paramètres de l'image et à son prétraitement [28].

L'utilisation de radiomiques suit plusieurs étapes. Tout d'abord l'acquisition de l'image médicale, puis la segmentation de la zone d'intérêt (telle que la tumeur). Le prétraitement de l'image est nécessaire, par exemple la réalisation d'une interpolation normalisant la taille des voxels permet la comparaison entre différents patients. Ce n'est qu'ensuite que les radiomiques peuvent être extraits. Cependant il existe un grand nombre de caractéristiques dont certaines sont redondantes. Utiliser toutes ces caractéristiques dans un même modèle entraîne un risque de surapprentissage, il faut donc sélectionner un nombre réduit de radiomiques à utiliser [29].

Les données radiomiques extraites de CT scans ont montré de bonnes performances lors de

la prédition de la survie ou de la survie sans récurrence de patients atteints de NSCLC [6].

Les caractéristiques radiomiques ont été regroupées en différents types. On trouve notamment les caractéristiques morphologiques décrivant les propriétés géométriques de la lésion (telles que son volume, sa forme, ses dimensions), de premier ordre (ou d'intensité) décrivant la distribution des intensités des niveaux de gris dans la région d'intérêt (comme l'intensité moyenne, sa variance, ou l'asymétrie de la distribution par exemple), de second ordre (ou de texture) exprimant l'organisation spatiale des pixels les uns par rapport aux autres et déterminant l'hétérogénéité des régions d'intérêt [27, 28, 30].

Les radiomiques de second ordre peuvent être sous-divisées selon l'organisation spatiale qu'elles étudient. Elles nécessitent de discréteriser les intensités des images étudiées. Pour expliquer les différentes sous-catégories de radiomiques de second ordre, nous prendrons l'exemple de l'image visible Figure 2.5.

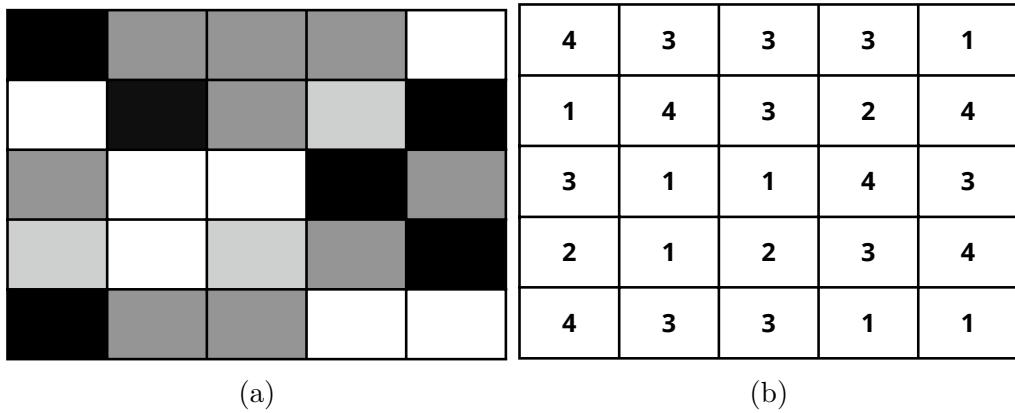


FIGURE 2.5 Exemple d'image. a) Image en niveaux de gris discréétisés. b) Représentation numérique de l'image a.

On y trouve en premier lieu les caractéristiques issues de matrices de co-occurrence des niveaux de gris (GLCM, pour *Gray-Level Cooccurrence Matrix*) étudiant les relations spatiales entre paires de voxels dans différentes directions. Ces matrices ont été introduites par Haralick et coll. [31] et expriment la façon dont les intensités voisines se combinent. On fixe une distance d'intérêt entre les voxels et on mesure les occurrences simultanées de niveaux d'intensité prédéfinis le long d'une direction. Dans les images 3D, 13 directions différentes sont étudiées puis les 13 matrices obtenues sont agrégées afin d'obtenir une meilleure invariance par rotation. La Figure 2.6 montre deux exemples de matrices de co-occurrence obtenues à partir de l'image d'exemple 2.5.

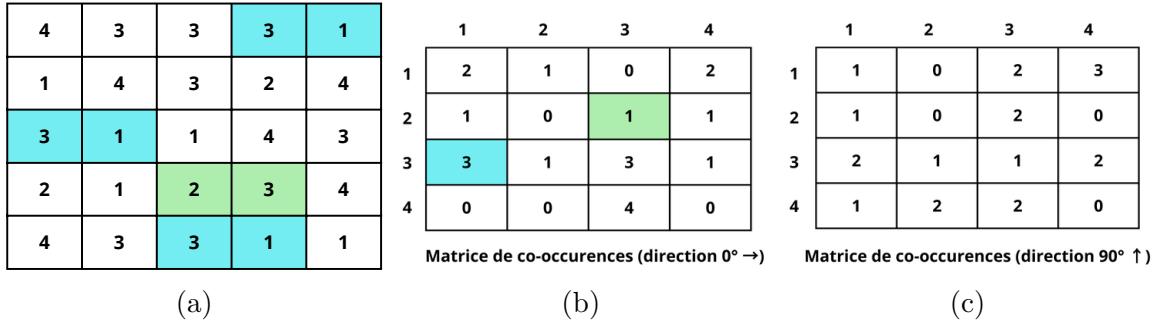


FIGURE 2.6 Exemple de matrices de co-occurrences. a) Représentation numérique de l'image d'exemple. b) Matrice de co-occurrences dans la direction horizontale (0°). c) Matrice de co-occurrences dans la direction verticale (90°).

Les radiomiques de second ordre comportent également les informations issues des matrices de GLRLM (pour *Gray-Level Run-length Matrix*). Celles-ci ont été proposées par Galloway [32]. Contrairement aux matrices de co-occurrences qui étudient les occurrences conjointes au sein de voxels adjacents, les GLRLM étudient la longueur des occurrences simultanées de pixels de même intensité le long d'une séquence consécutive le long d'une direction donnée. Là encore les mesures sont prises dans différentes directions (13 en 3D) et les différentes matrices résultantes sont agrégées. La Figure 2.7 présente l'exemple de matrices GLRLM pour deux directions différentes.

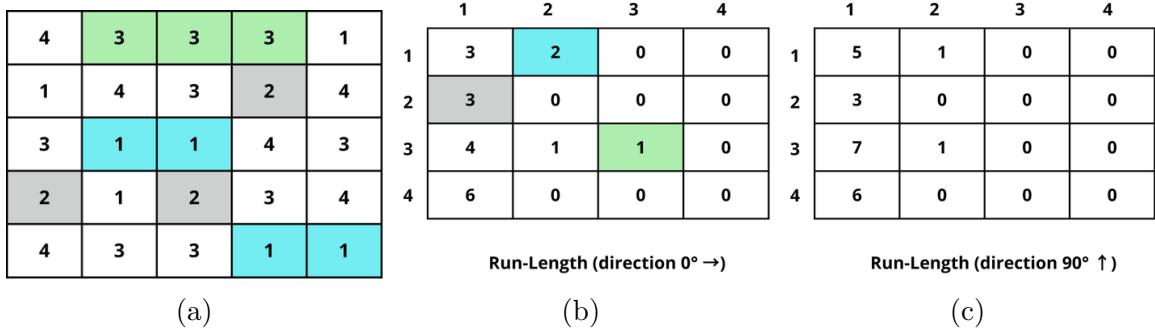


FIGURE 2.7 Exemple de matrices GLRLM. a) Représentation numérique de l'image d'exemple. b) Matrice GLRLM dans la direction horizontale (0°). c) Matrice GLRLM dans la direction verticale (90°).

Les matrices GLSZM (pour *Gray-Level Size Zone Matrix*) ont été introduites par Thibault et coll. [33]. Elles comptent le nombre de zones de voxels voisins de même intensité. La notion

de voisin est définie par la connectivité. En trois dimensions on considère qu'un voxel est lié à tous ceux parmi ses 26 voisins présentant le même niveau de gris que lui. La définition est similaire en 2 dimensions en considérant un ensemble de 8 voisins par voxel. La Figure 2.8 montre un exemple de matrice GLSZM.

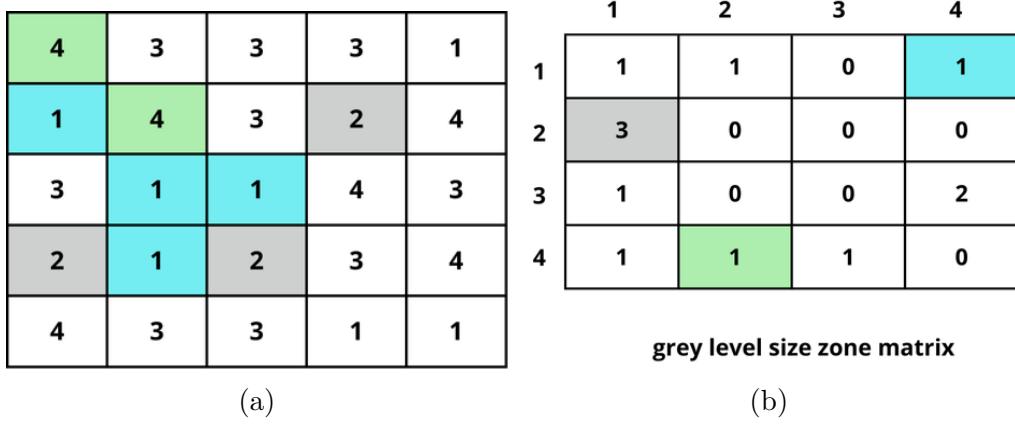


FIGURE 2.8 Exemple de matrices GLSZM. a) Représentation numérique de l'image d'exemple. b) Matrice GLSZM associée.

Enfin les matrices NGTDM (pour *Neighbourhood Gray-Tone Difference Matrix*) ont été proposées par Amadasun et King [34] et quantifient la somme des différences entre le niveau de gris d'un voxel et le niveau de gris moyen de ses voisins situés à l'intérieur d'une distance prédéfinie.

2.6 Apprentissage machine et apprentissage profond

Dans cette partie nous traitons des méthodes d'apprentissage artificiel. Nous présenterons d'abord différents modèles d'apprentissage profond (*deep learning*) constitués de réseaux de neurones, tels que les Perceptrons, les réseaux de neurones convolutifs, les modules de transformation spatiale, les encodeurs-décodeurs et les Transformers. Dans une seconde sous-partie nous étudierons différents modèles d'apprentissage machine. Nous étudierons en particulier les régressions logistiques, les machines à vecteur de support et les forêts aléatoires. Contrairement aux réseaux de neurones, ces modèles utilisent uniquement des données d'entrées dont les caractéristiques (*features*) ont été définies par les utilisateurs, là où les modèles d'apprentissage profond sont capables de traiter directement les images d'origine.

Parmi les modèles présentés ici, nous avons utilisé en particulier des réseaux de type auto-encodeurs variationnels et des modules de transformation spatiale pour prédire les déforma-

tions successives du poumon, et des modèles de type régression logistique, régression par machine à vecteur de support et des forêts aléatoires pour prédire les chances de succès du traitement.

2.6.1 Réseaux de neurones

Perceptron multicouches

Le Perceptrons multicouches (*Multilayer perceptron*) (MLP) représente le réseau de neurones de base. Comme son nom l'indique, il est composé de différentes couches de Perceptrons.

Le Perceptron multicouches peut être vu métaphoriquement comme un réseau de neurones biologique où un Perceptron représente un neurone du cerveau. Comme dans le cerveau, l'information est transmise de neurone en neurone [35].

Dans le cas du Perceptron multicouches, celui-ci comporte un ou plusieurs noeuds d'entrée et un noeud de sortie unique relié aux couches d'entrée par des couches intermédiaires (ou couches cachées) de Perceptrons. La Figure 2.9 représente une schématisation d'un réseau Perceptron composé d'une unique couche cachée. L'erreur en sortie entre le résultat proposé par le modèle et le résultat attendu est rétropropagée vers les entrées pour mettre à jour les poids. L'apprentissage du réseau consiste à ajuster les poids des neurones pour obtenir la sortie désirée [36, 37].

Considérons un vecteur d'entrée x , un vecteur de poids W , un biais b , et une fonction d'activation f . L'équation de fonction de sortie est :

$$y = f(W^T x + b) \quad (2.2)$$

Les poids sont mis à jour via rétropropagation de l'erreur, en minimisant une fonction de coût (ou fonction d'écart) L qui est définie de façon à sélectionner la meilleure solution possible en fonction du problème. Considérons un taux d'apprentissage α . À chaque itération, les poids w_i sont mis à jour et deviennent w'_i selon l'équation :

$$w'_i = w_i + \alpha \frac{\partial L}{\partial w_i} \quad (2.3)$$

Réseaux de neurones convolutifs

Lorsqu'on utilise des images, la structure de l'image en elle-même est porteuse d'information, et pas uniquement la valeur de chaque pixel [37]. Leur traitement nécessite donc un type de

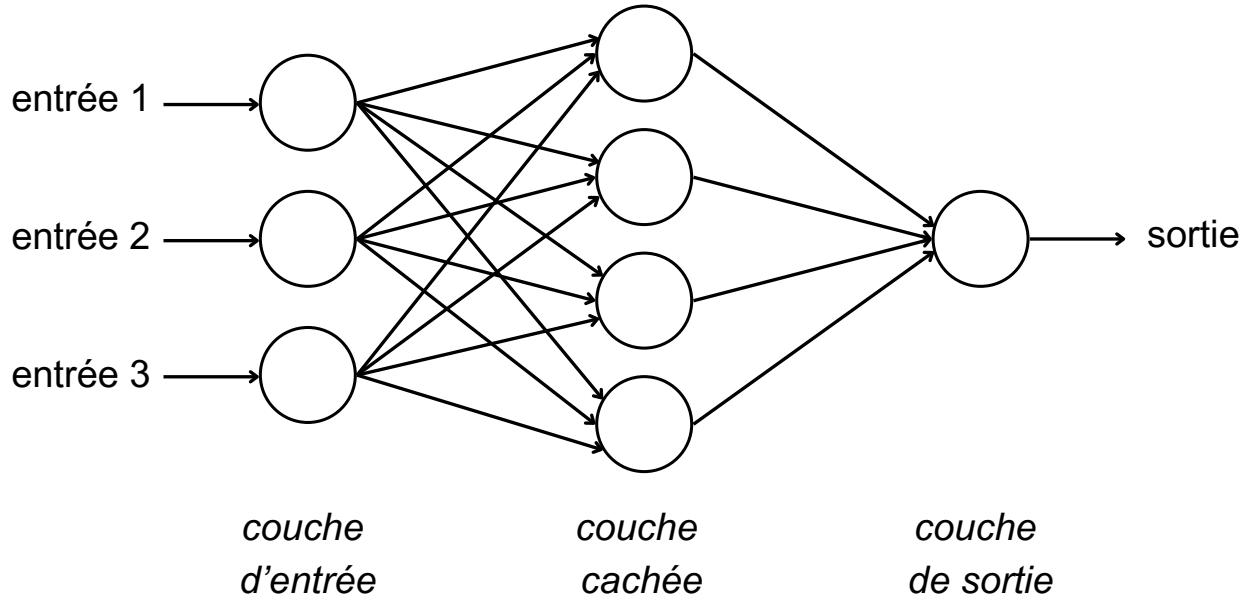


FIGURE 2.9 Schéma de l'architecture d'un réseau Perceptron multicouches

modèle particulier capable d'apprendre la structure de l'image. Les réseaux de neurones convolutifs (CNN) (pour *Convolutional Neural Networks*) se sont révélés très performants dans des tâches de vision par ordinateur. Ils utilisent un filtre de convolution à la place de l'opération matricielle des Perceptrons. Une couche d'un CNN est un ensemble de filtres. Lors de la propagation avant chaque filtre est appliqué à l'entrée selon :

$$y = f(W * x + b), \text{ avec } * \text{ l'opération de convolution} \quad (2.4)$$

Une fois l'image filtrée par ces différents filtres, les résultats sont agrégés (*pooling*) sur une petite région en utilisant la valeur maximale ou moyenne. Cette couche permet de réduire progressivement le nombre de paramètres du modèle. Ceci permet de réduire le temps de calcul et permettre une invariance spatiale [37].

Les éléments d'un niveau de convolution calculent leur activation à partir des éléments du niveau précédent, ce qui permet de réduire le nombre de paramètres et d'augmenter la taille des champs réceptifs à chaque couche. Les couches de plus haut niveau arrivent ainsi à détecter des caractéristiques de plus en plus abstraites et complexes, comme des formes puis des textures et ensuite des parties d'objet [37–39].

L'une des forces des CNN est leur capacité à partager les poids, ce qui permet de réduire le nombre de paramètres [37]. Ce type de réseau est également invariant aux faibles translations [37]. Par exemple si le modèle sait reconnaître un motif dans le coin droit, il le reconnaîtra

partout. Grâce à cela les CNN ont besoin de moins d'exemples d'entraînement pour apprendre des représentations capables de généralisation [39].

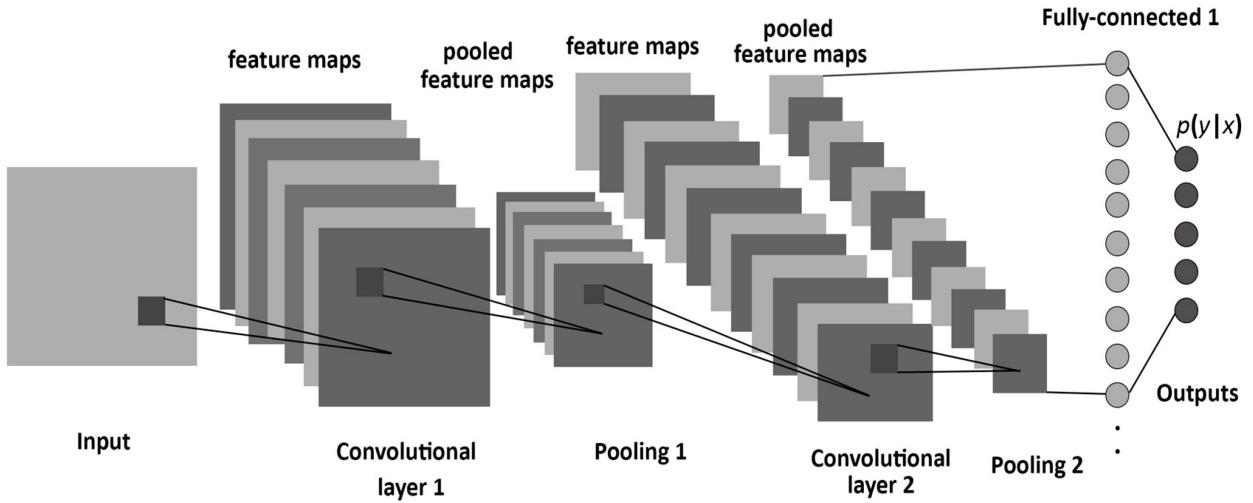


FIGURE 2.10 Schéma de l'architecture d'un réseau CNN. Image provenant de [40].

Module de transformation spatiale

Les modules de transformation spatiale (STN), proposés en 2015 par Jaderberg et coll. [41], permettent de rendre les CNN plus robustes en améliorant leur capacité d'invariance spatiale. En effet les modules d'agrégation (via *MaxPooling*) des CNN les ont rendus invariants aux variations spatiales de leurs images d'entrée. Cependant les fenêtres utilisées sont souvent de petite taille (par exemple 2x2 pixels), ce qui fait que cette invariance ne s'applique pas aux couches intermédiaires.

Ces modules STN peuvent être insérés au sein d'architectures déjà existantes et permettent aux CNN de réaliser une manipulation spatiale des données générées par les couches intermédiaires.

Les STN se composent de trois parties. Tout d'abord le réseau prend en entrée un volume U et, à l'aide d'un certain nombre de couches cachées, propose en sortie de cette première étape les paramètres θ d'une transformation spatiale T_θ à appliquer à cette image U pour la rapprocher de l'ensemble de données. Ensuite, le générateur de grille va permettre de réaliser les translations, recadrages, rotations, mises à l'échelle et distorsions nécessaires. Enfin l'échantillonneur va prendre en entrée l'image U et la grille pour produire une image transformée selon les paramètres désirés.

Encodeur-Décodeur

Les modèles de type Encodeur-Décodeur sont constitués de deux parties. La première partie (l'encodeur) traduit l'entrée vers un format de plus faible dimension, et la seconde partie (le décodeur) traduit l'information compressée vers l'espace de sortie.

La fonction objectif vise à minimiser l'erreur entre la sortie attendue y et la sortie du décodeur. Notons f l'opération réalisée par l'encodeur et g celle réalisée par le décodeur, avec x l'entrée fournie au modèle. La fonction objectif d'un modèle encodeur-décodeur sera :

$$\arg \min_{f,g} \|y - (g \circ f)x\|^2 \quad (2.5)$$

L'avantage de ce type de modèle est que les entrées et sorties peuvent être de longueur variable, ce qui est utile par exemple pour le traitement du langage. Les modèles encodeur-décodeurs sont par conséquent régulièrement utilisés lors de la traduction d'une langue vers une autre.

Auto-encodeurs

Les modèles de type Auto-encodeur (AE) cherchent à minimiser les différences entre leurs entrées et leurs prédictions de sortie. Pour cela l'encodeur réalise l'encodage de l'entrée x vers le code z tel que $z = f(x)$, et le décodeur se charge du décodage du code z vers la sortie reconstruite r , définie comme $r = g(z)$, et qui est idéalement proche de l'entrée. L'utilité de ce modèle réside dans son espace latent z représentant uniquement les attributs utiles à la reconstruction des objets. L'utilisateur définit des contraintes sur la sortie de l'encodeur z en fonction des caractéristiques qu'il souhaite apprendre. En général il limite sa taille de façon à ce que le modèle apprenne à compresser cette information [37, 39]. Pour cette raison ce modèle peut être vu comme une généralisation non linéaire de l'analyse par composantes principales (PCA) [37].

La fonction objectif d'un modèle auto-encodeur peut être définie par :

$$\arg \min_{\theta, \omega} L(x, g_{\omega}(f_{\theta}(x))) \quad (2.6)$$

avec θ et ω les paramètres de l'encodeur et du décodeur, et L une mesure de similarité.

Cependant ces AE ne permettent pas une compression très efficace et ne possèdent pas d'espace latent très structuré non plus [39], ce qui n'est pas très utile en pratique. Au contraire les auto-encodeurs variationnels (VAE) apprennent des espaces latents très structurés et se

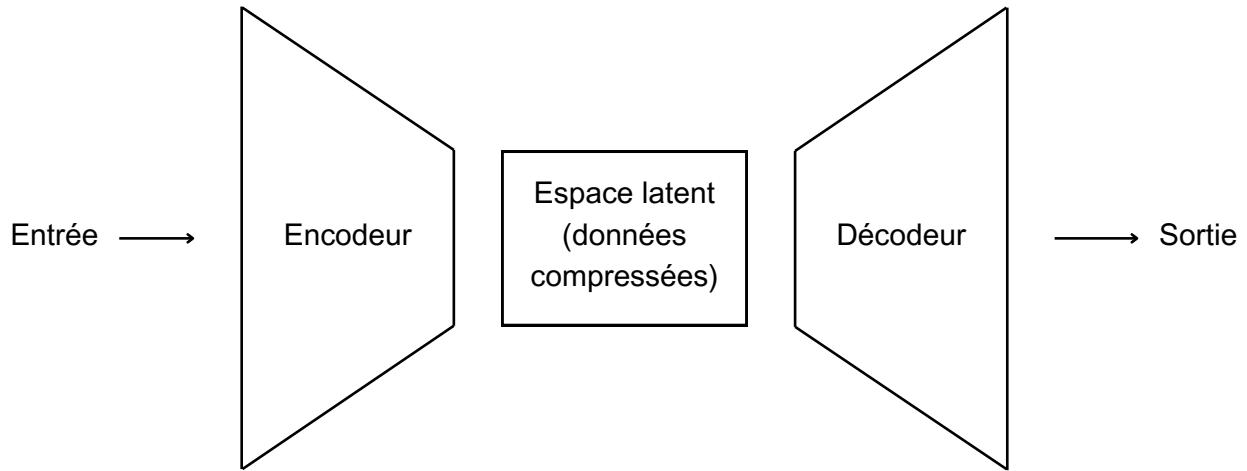


FIGURE 2.11 Schéma de l'architecture d'un auto-encodeur

sont révélés être très efficaces dans des tâches de génération d'images [39]. Leur spécificité consiste à considérer les images d'entrée comme représentantes d'une distribution statistique dont l'espace latent apprend la moyenne et la variance. Ainsi l'espace latent prend désormais la forme d'une distribution gaussienne. Cet aspect aléatoire du modèle augmente sa robustesse et permet de l'utiliser comme modèle génératif étant donné qu'il peut désormais générer des exemples qui n'étaient pas présents à l'origine dans le jeu d'entraînement.

Modèles Transformer

Les modèles Transformer [42] sont basés sur une architecture de type encodeur-décodeur utilisant le mécanisme d'attention développé en vue du traitement du langage naturel [43]. Ce mécanisme permet d'établir des liens relationnels entre différents jetons sémantiques qui composent une phrase, et l'objectif de ces réseaux de neurones est d'apprendre à établir ces liens logiques par les paramètres du modèle. Chaque séquence d'entrée est ordonnée par un encodage de position puis passe dans chaque encodeur. Ceux-ci sont composés d'une couche d'auto-attention, dont les sorties sont acheminées vers un réseau de neurones à propagation avant. Avant et après chacune de ces couches, on propage l'information par des connexions résiduelles et on la régularise avec une couche de normalisation par niveau [43, 44]. La Figure 2.12 présente une représentation de l'architecture d'un Transformer. En fonctionnant comme une pile d'encodeurs et de décodeurs, le Transformer permet une mise à échelle et un approfondissement des représentations sur un large nombre de paramètres.

Contrairement aux réseaux de neurones récurrents (RNN) et même s'ils sont conçus pour traiter des données séquentielles, les modèles Transformers sont capables d'analyser simul-

tanément l'ensemble des données d'entrée grâce aux encodages de position et au mécanisme d'auto-attention, sans nécessairement devoir traiter la séquence dans l'ordre d'apparition des mots. Cela permet de réduire le temps d'entraînement avec des calculs plus simplement parallélisables, dont la complexité d'auto-attention en $O(n^2 \times d)$, avec n la taille de séquence et d la dimension, permet de se mettre à l'échelle des données et de la taille de modèle voulue. L'utilisation du mécanisme d'attention aide à résoudre le problème de disparition du gradient et d'établir des relations à plus grande distance. Il permet aussi d'augmenter l'interprétabilité du modèle [42].

Dosovitskiy et coll. [45] ont proposé un modèle de Transformer appelé *Vision Transformer* (ViT) capable de traiter des images en les séparant en parcelles de taille 16 x 16 qui sont traitées comme des mots par le Transformer. L'encodeur du Transformer prend donc en entrée les plongements associés à chacune de parcelles 16x16 appris par un modèle linéaire ainsi que les encodages de position de chaque morceau d'image.

Cependant tant les réseaux Transformer que ViT nécessitent un jeu de données d'apprentissage très important. En conséquence leur utilisation nécessite presque obligatoirement de passer par un *finetuning* à partir de modèle pré-entraînés sur de gros volumes de données.

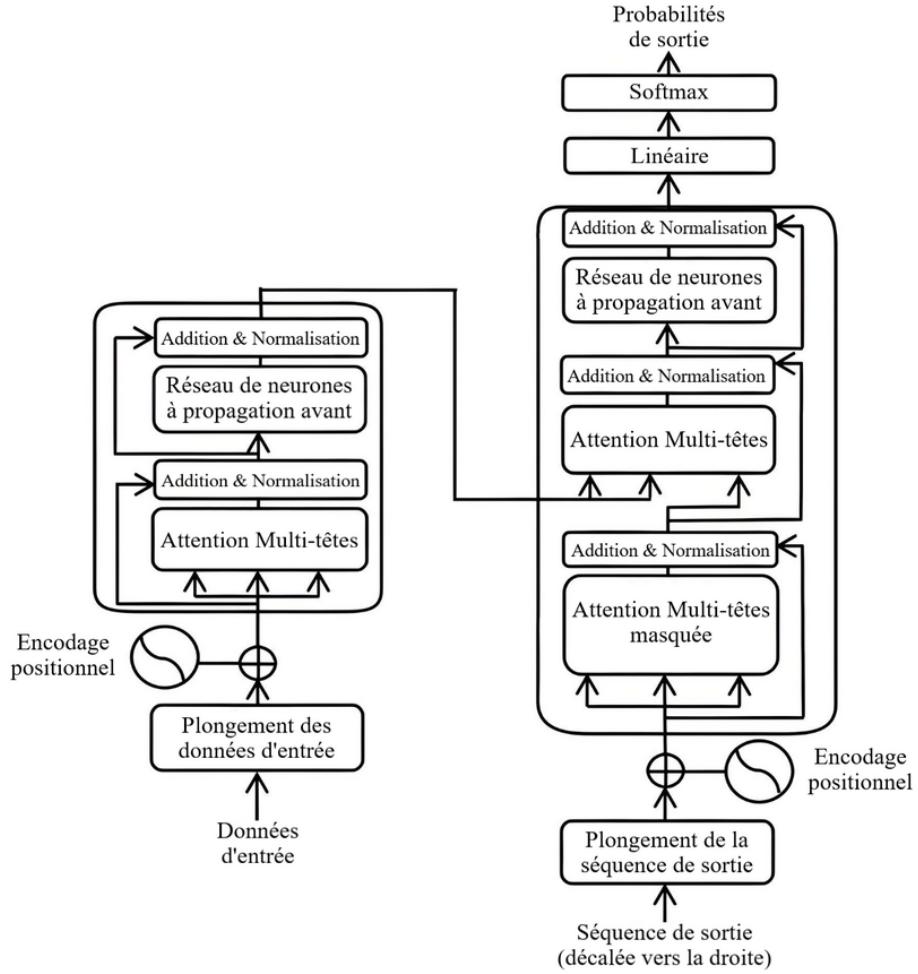


FIGURE 2.12 Schéma de l'architecture d'un Transformer. Image extraite de Wikimedia Commons.

2.6.2 Apprentissage machine

Régression logistique

La régression logistique, ou *Logistic Regression* (LR), applique une fonction sigmoïde à une combinaison linéaire de données d'entrées. Elle s'exprime sous la forme :

$$\hat{y} = \sigma(w^T x + b), \text{ avec } \sigma \text{ la fonction sigmoïde telle que } \sigma(z) = \frac{1}{e^{-z}} \quad (2.7)$$

Régression par machine à vecteur de support

La machine à vecteur de support, ou *Support Vector Machine* (SVM), vise à créer un hyperplan de séparation permettant de classifier les points du jeu de données. Pour cela son objectif sera de générer un hyperplan maximisant la distance avec les points d'entraînement. Ainsi au lieu de minimiser la fonction de perte sur le jeu d'apprentissage cette stratégie vise à minimiser la fonction de perte en généralisation attendue, en choisissant le séparateur situé le plus loin possible des exemples observés. Cette stratégie permet une meilleure généralisation [39, 46].

Les SVM permettent également de classifier des données non linéairement séparables grâce à la méthode des noyaux. Cette méthode permet de transformer l'espace des données en un espace de dimension supérieure dans lequel il est possible de trouver un hyperplan permettant de catégoriser les exemples d'entraînement. Cet hyperplan linéaire dans un espace de dimension supérieure n'est pas nécessairement linéaire dans l'espace d'origine [39, 46].

Les SVM présentent l'avantage d'être facilement interprétables et assez résistants au surapprentissage [39, 46]. Cependant ils peuvent présenter des difficultés d'apprentissage sur des jeux de données constitués de nombreux exemples [39].

Modèle de forêt aléatoire

Les forêts aléatoires, ou *Random Forest* (RF), se basent sur des arbres de décision.

Un arbre de décision est une structure similaire à un diagramme permettant de classifier une variable objectif en fonction des valeurs successives d'attributs en entrée. Ils se basent sur un partitionnement récursif des données jusqu'à atteindre une séparation en classes satisfaisante [39, 47]. Chaque noeud de l'arbre correspond au test de la valeur d'un des attributs et les branches issues de ce noeud correspondent aux valeurs possibles de cet attribut. Les feuilles de l'arbre indiquent les différentes valeurs possibles à retourner par le modèle.

La figure 2.13 présente un exemple d'arbre de décision fictif. Les feuilles de l'arbre sont colorées selon les prédictions réalisées par l'arbre. Dans notre exemple fictif, tout fumeur de plus de 70 ans serait considéré comme à risque, alors qu'une femme non fumeuse de plus de 70 ans ne serait pas considérée comme à risque.

Les arbres de décision ont l'avantage d'être visuels et ainsi de permettre, comme leur nom l'indique, d'aider à la prise de décisions. En revanche ils sont très instables : changer un seul exemple dans le jeu d'apprentissage peut impacter l'ensemble des prédictions [47]. Un unique individu peut donc grandement affecter les performances du modèle, ce qui peut susciter de la méfiance sur la pertinence d'un arbre précis.

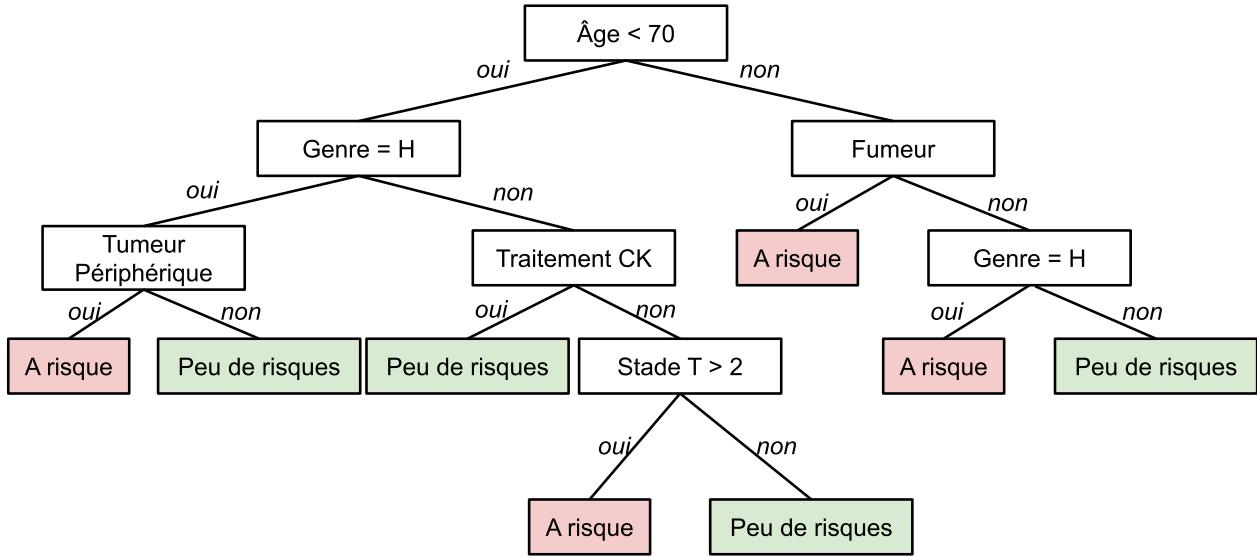


FIGURE 2.13 Représentation schématique d'un arbre de décision fictif.

Pour éviter leur surapprentissage on utilise la stratégie d'élagage (*prunning*) permettant de supprimer les noeuds non pertinents. Cette méthode consiste à d'abord générer un arbre puis tester l'intérêt de chaque noeud menant à des feuilles. Cet intérêt est vérifié à l'aide d'un test statistique sous l'hypothèse nulle que le noeud ne fait qu'une partition aléatoire des données sans motif sous-jacent. Si cette hypothèse ne peut être infirmée alors on peut estimer que ce noeud ne détecte en fait que du bruit et remplacer ce noeud par un noeud feuille. On procède ensuite ainsi pour chaque noeud feuille, jusqu'à ce qu'ils soient tous acceptés ou tous élagués. Ceci permet d'obtenir des arbres plus performants car disposant de meilleures capacités de généralisation, mais également plus petits donc plus simples à comprendre [46].

Les RF consistent à apprendre un grand nombre d'arbres de décision, ce qui permet de réduire la variance des résultats. Les forêts aléatoires se basent sur un ré-échantillonnage du jeu d'apprentissage, en supprimant certains exemples et en doublant d'autres. On construit ainsi un grand nombre d'arbres spécialisés dont on agrège les résultats via un vote majoritaire [39, 47]. Ces modèles fonctionnent pour un grand nombre de tâches, en particulier celles se basant sur des données tabulaires.

2.6.3 Stratégie d'entraînement

La séparation des données en jeu d'apprentissage et jeu de validation peut être rendue difficile par un jeu de données trop limité. En effet dans ce cas la taille du jeu d'entraînement pourrait être trop petite pour rencontrer suffisamment de cas différents, ou le jeu de test pourrait être trop petit pour être représentatif de la population réelle. Cette difficulté est notamment présente dans les applications médicales où l'accès aux données protégées des patients est limité, ce qui réduit le nombre de patients disponibles.

Une solution à ce problème est d'utiliser une stratégie de validation croisée. Ceci consiste à réaliser de multiples évaluations du jeu de donnée basées sur différents découpages de l'ensemble de données. Une validation croisée à k blocs partitionne aléatoirement le jeu de données en k partitions égales. Pour chacune des k itérations on utilise une des partitions comme jeu de test et les $k-1$ autres comme jeu d'entraînement, comme on peut le voir représenté sur la Figure 2.14. On moyenne ensuite les performances de chaque bloc. Cette stratégie permet d'obtenir une bonne approximation des performances globales du modèle en dehors de ce jeu de données [48].

Dans le cas de jeux de données déséquilibrés on utilise la validation croisée stratifiée, où l'on s'assure d'avoir environ la même proportion d'événements dans chaque bloc afin de ne pas biaiser les estimations de performance par les variations de proportion de chaque classe.

2.6.4 Métriques

Différentes métriques peuvent servir à évaluer les modèles.

Matrice de confusion

Une matrice de confusion, ou table de contingence, permet de visualiser les résultats d'un modèle, en mettant en avant les cas correctement classés par le modèle ainsi que les erreurs réalisées classe par classe. Les lignes de la matrice indiquent les classes réelles, et les colonnes contiennent les éléments prédis par le classificateur. Le tableau 2.1 présente un exemple de matrice de confusion. Les exemples correctement classés sont dans la diagonale de la matrice, les autres éléments sont mal classés. On y voit par exemple le nombre de Vrais Positifs (TP), qui sont les exemples prédis comme appartenant à la classe Oui par le modèle et dont la classe réelle est effectivement la classe Oui, ainsi que le nombre de Faux négatifs (FN), qui sont les éléments classés comme appartenant à la classe Non par le modèle, mais dont la classe réelle est la classe Oui.

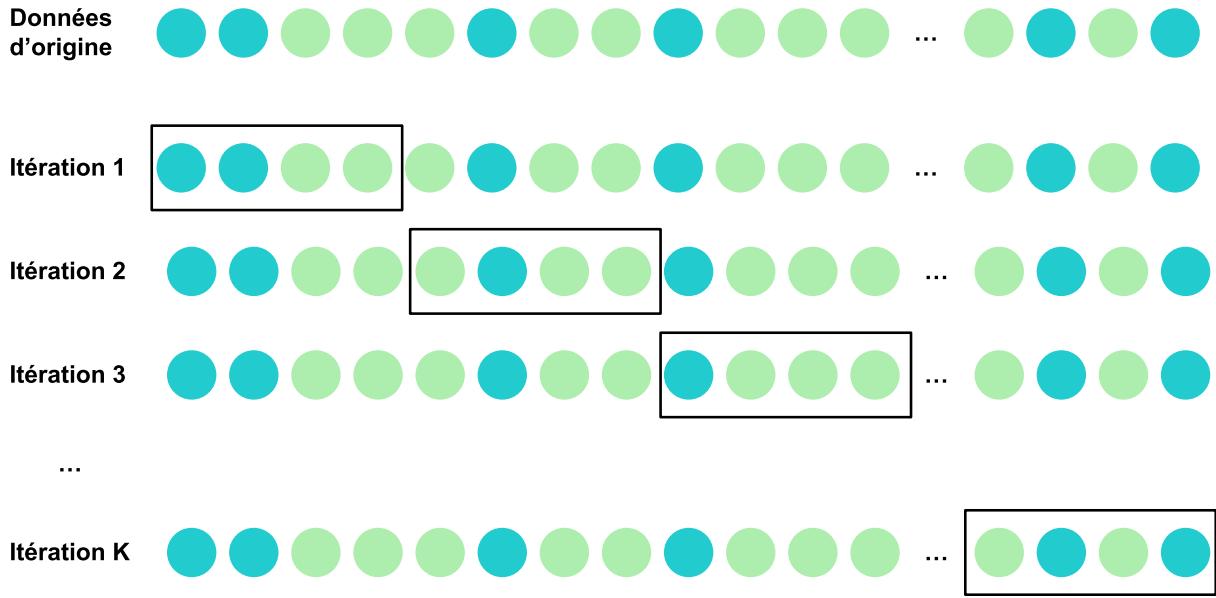


FIGURE 2.14 Schéma d'un exemple de validation croisée. Les données encadrées constituent le jeu de test de l'itération, pendant que les autres données forment le jeu d'apprentissage.

	Classe Oui	Classe Non
Classe Oui	Vrais positifs (TP)	Faux négatifs (FN)
Classe Non	Faux positifs (FP)	Vrais négatifs (TN)

TABLEAU 2.1 Exemple d'une matrice de confusion dans un cas de classification binaire.

Sensibilité et spécificité

La sensibilité d'un modèle désigne sa capacité à détecter les cas positifs (TP). Elle est définie comme le nombre de cas positifs correctement classés par rapport au nombre de cas positifs au total, ce qui s'exprime par :

$$sensibilité = \frac{TP}{TP + FN} \quad (2.8)$$

La sensibilité est maximale lorsque tous les cas positifs ont été classifiés comme tels par le modèle. Cependant un tel modèle risque de tout considérer comme positif, et ainsi de ne pas être utile en pratique, car sa spécificité serait mauvaise.

La spécificité d'un modèle désigne elle la capacité du modèle à ne pas considérer à tort un

cas comme positif. Elle est définie comme :

$$specificite = \frac{TN}{TN + FP} \quad (2.9)$$

Cette métrique est maximale lorsque le modèle ne prédit aucun faux positif. Mais un tel modèle risquerait de ne trouver aucun des cas positifs, de peur de se tromper.

Ces deux métriques sont donc complémentaires, il faut veiller à ne pas augmenter l'une au détriment de l'autre.

Il est possible de combiner ces deux mesures en utilisant la précision (*accuracy*) qui mesure la proportion d'échantillons bien classés par rapport à l'ensemble du jeu de test, c'est-à-dire le nombre de prédictions correctes divisé par le nombre total de prédictions effectuées. Elle s'exprime par :

$$Precision = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.10)$$

Courbe ROC et AUC

La courbe *Receiver Operating Characteristic* (ROC) permet de visualiser la qualité d'un modèle en fonction des taux de Faux Positifs et Faux Négatifs. Il s'agit d'une courbe traçant les taux de vrais positifs ($TPR = \frac{TP}{TP+FN}$) en fonction des taux de vrais négatifs ($TNR = \frac{TN}{TN+FP}$). On fait varier ces taux en considérant différentes valeurs seuils de probabilité à partir de laquelle le modèle considère l'élément comme devant être prédit comme une classe positive.

Plus l'aire sous la courbe *Area Under the Curve* (AUC) est importante plus le modèle est intéressant. Au contraire un modèle aléatoire aurait plutôt une courbe ressemblant à une diagonale. Cette métrique présente l'intérêt de ne pas dépendre de la valeur seuil choisie, contrairement à la sensibilité et à la spécificité par exemple.

Métriques utilisées dans l'analyse des images prédictes

Pour évaluer la ressemblance entre des images prédictes et des images attendues, on sort du cadre de la classification binaire, étant donné que l'on compare des images constituées de nombreux pixels. Les métriques d'évaluation seront elles aussi différentes et visent à calculer la ressemblance entre ces images. Nous utiliserons dans ce mémoire l'erreur quadratique moyenne (MSE), la similarité structurelle (SSIM) et la Corrélation croisée (CC).

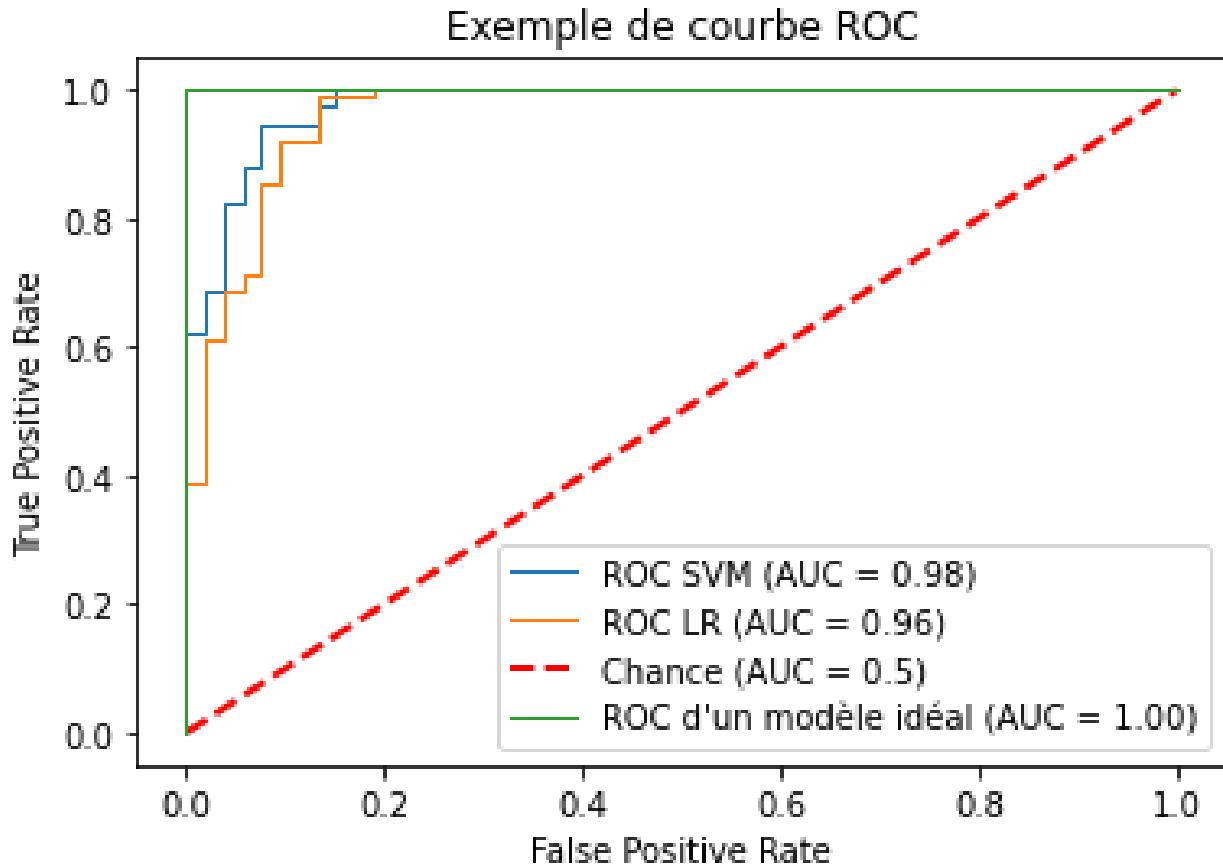


FIGURE 2.15 Exemple de courbe ROC comparant deux modèles différents. La courbe verte présente le cas idéal d'un modèle classant correctement tous les cas du jeu de test avec une certitude absolue. La ligne rouge quant à elle présente ce qu'obtiendrait un modèle prédisant aléatoirement une classe ou l'autre. Les courbes oranges et bleues présentent les résultats obtenus par deux modèles différents (une régression logistique et un SVM), chacun ayant de bons résultats sur le jeu de test. Dans cet exemple l'AUC du SVM est meilleure que celle de LR.

Erreur quadratique moyenne (MSE) La MSE mesure le carré de la différence pixel à pixel entre l'image générée et l'image cible.

Soit deux images à comparer I et J . La MSE entre ces deux images s'exprime sous la forme :

$$MSE(I, J) = \frac{1}{HL} \sum_{i=1}^H \sum_{j=1}^L [I(i, j) - J(i, j)]^2 \quad (2.11)$$

où H représente la hauteur de l'image et L sa largeur, et (i, j) définit la position du pixel considéré.

Cette métrique s'exprime également couramment sous la forme de l'écart quadratique moyen (RMSE). Celle-ci s'exprime sous la forme :

$$RMSE(I, J) = \sqrt{MSE(I, J)} \quad (2.12)$$

Cette nouvelle écriture de la métrique permet de la ramener à la même unité que celle de l'image d'origine, ce qui peut faciliter son interprétation.

Similarité structurelle (SSIM) La SSIM est une mesure proposée par Wang et coll. en 2004 [49]. Cette mesure évalue la similarité de structure et non pixel à pixel entre deux images. Leur idée directrice est que les pixels d'une image sont fortement corrélés, et ce d'autant plus qu'ils sont spatialement proches. Leurs variations permettent donc de déterminer les changements de structure.

La SSIM se calcule à l'aide de fenêtres de taille $N \times N$. Notons I et J les deux images à comparer, et x et y les fenêtres utilisées. Soit μ_x la moyenne de x , μ_y celle de y , σ_x^2 la variance de x , σ_y^2 celle de y et cov_{xy} la covariance de x et y . La similarité structurelle de x et y s'exprime selon l'équation :

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_x\sigma_y + c_2)(cov_{xy} + c_3)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)(\sigma_x\sigma_y + c_3)} \quad (2.13)$$

Pour stabiliser l'équation en cas de faibles dénominateurs, on introduit les variables $c_1 = (K_1 L)^2$ et $c_2 = (K_2 L)^2$, avec L la plage de valeurs des voxels et K_1, K_2, c_3 des constantes de faible valeur (par exemple $K_1 = 0.01$, $K_2 = 0.03$ et $c_3 = \frac{c_2}{2}$).

L'équation 2.13 peut être divisée en trois termes comparant les similarités en matière de luminance, contraste et structure selon l'équation :

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (2.14)$$

Dans cette équation le terme l représente la comparaison de luminance, exprimée par :

$$l(x, y) = \frac{(2\mu_x\mu_y + c_1)}{(\mu_x^2 + \mu_y^2 + c_1)} \quad (2.15)$$

c compare les images en termes de contraste selon :

$$c(x, y) = \frac{(2\sigma_x\sigma_y + c_2)}{(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2.16)$$

Enfin le terme s mesure la similarité des images en ce qui concerne les structures :

$$s(x, y) = \frac{(cov_{xy} + c_3)}{(\sigma_x \sigma_y + c_3)} \quad (2.17)$$

Il est possible d'adapter l'importance relative accordée à chaque comparaison en jouant sur les paramètres $\alpha > 0$, $\beta > 0$ et $\gamma > 0$, par défaut fixés égaux à 1.

La qualité de l'image totale est déterminée en moyennant les mesures de similarité de chaque fenêtre.

Corrélation croisée (CC) La CC mesure la similarité en termes de covariance des images I et J . Sous sa forme normalisée NCC elle s'exprime selon l'équation suivante :

$$NCC(I, J) = \frac{Cov(I, J)}{\sqrt{Var(I)Var(J)}} \quad (2.18)$$

2.7 État de l'art

Dans cette section nous détaillerons l'état de l'art pour les deux objectifs proposés.

2.7.1 Modélisation du mouvement du poumon

La modélisation du mouvement respiratoire se concentre souvent sur la modélisation du déplacement de la tumeur au cours de la respiration, étant donné que l'objectif est de suivre sa position avec le faisceau de radiation.

Pour cela, une première approche consiste à insérer des marqueurs à proximité de la tumeur. Ceux-ci sont facilement repérables sur les images médicales et permettent ainsi de suivre plus efficacement le déplacement de la tumeur. Les réseaux de type RNN sont couramment utilisés dans ce cas de figure. Par exemple Jiang et coll. [50] proposent un modèle personnalisé au patient utilisant des marqueurs internes implantés à proximité de la tumeur et pouvant être suivis par fluoroscopie. Les positions des marqueurs sont enregistrées sous la forme de signaux à 1 dimension (1D) en fonction du temps dans les trois dimensions spatiales. Trois modèles RNN prédisent chacun le déplacement de la tumeur dans une des directions spatiales en utilisant ses positions au cours du cycle précédent (les 4 dernières secondes) pour prédire ses positions futures. Ce modèle basé sur des RNN a été comparé à un ensemble de trois réseaux autorégressifs non linéaires à entrées exogènes, ou NARX (*nonlinear autoregressive exogenous model*), chacun prédisant le déplacement de la tumeur dans une direction. Ces réseaux NARX

sont une extension des réseaux RNN qui enregistrent les valeurs des 10 prédictions et mesures passées et s'en servent afin de minimiser l'erreur des prédictions futures. Ces deux méthodes ont été comparées sur sept patients différents et le modèle NARX a obtenu de meilleures performances, avec une RMSE en 3D de 82.32 mm pour une prédition sur un horizon de 600 ms.

Cependant l'utilisation de marqueurs nécessite souvent l'opération du patient et risque donc des complications ou n'est pas adaptée à tous les patients [51]. Il est donc préférable de s'en passer.

Certaines études se concentrent sur la prédition du mouvement de la tumeur dans la direction supérieur-inférieur, étant donné qu'il s'agit de la direction concentrant la plus grande amplitude de déplacement. Les méthodes d'apprentissage permettant de prédire l'amplitude de déplacement de la tumeur sont variées.

Ainsi Sun et coll. [52] utilisent un modèle constitué de deux MLP, où lorsque l'un des deux réseaux réalise une prédition, l'autre est entraîné sur les données de la période de respiration actuelle. Leur jeu de données se compose de l'enregistrement de l'amplitude de déplacement de la tumeur durant 2 minutes pour 202 patients où la première minute sert à l'entraînement des modèles et la deuxième à leur validation. La prédition s'effectue sur un temps glissant de 500 ms et leur modèle atteint une RMSE de 0.113 cm.

Lin et coll. [53] quant à eux proposent un modèle basé sur un réseau *Long Short-Term Memory* (LSTM). Leur jeu de données comporte 1 703 ensembles de positions de tumeurs provenant de 3 institutions différentes. Leurs résultats sont similaires à ceux de Sun et coll. avec une RMSE moyenne de 0.139 sur une fenêtre glissante de 500 ms.

Face aux nombreux types de réseaux possibles, Samadi Miandoab et coll. [54] ont comparé les performances de modèles RNN, LSTM, *Gated Recurrent Unit* (GRU), RNN bidirectionnel, LSTM bidirectionnel, GRU bidirectionnel et CNN-LSTM (comprenant un CNN suivi d'un LSTM) sur un jeu de données de 800 signaux respiratoires provenant de 30 patients différents. Ils ont trouvé que le modèle GRU obtenait les meilleures performances avec une RMSE de 0.108 mm.

Les modèles de type RNN sont malgré tout soumis au problème de la disparition du gradient, qui limite leurs performances lors de prédictions sur de longues séquences, par exemple à partir de 600 ms [54, 55]. Face à ce constat, Tan et coll. [55] ont développé un modèle appelé *Long Short-term Transformer* (LSTformer) en ajoutant un module destiné à limiter la dégradation de performances à long terme en combinant une cellule LSTM à un MLP. Les

auteurs ont développé leur modèle sur un jeu de données de 304 mouvements respiratoires issus de 38 patients différents et acquis sur une durée variant entre 6.5 et 132 minutes. Ils obtiennent une performance de RMSE à 600 ms de 0.501 mm, avec un temps de calcul de moins de 100 ms.

Les quatre solutions présentées ci-dessus ne s'appliquent qu'à des données en une dimension, donc limitées à l'amplitude de déplacement de la tumeur dans une unique direction.

Le modèle développé par Mafi et coll. [56] résoud ce problème en utilisant un signal composé de trois signaux 1D représentant le déplacement de la tumeur dans les directions gauche-droite *GD*, antérieur-postérieur *AP* et supérieur-inférieur *SI* et en regroupant ces trois informations en une seule à l'aide du déplacement 3D de la tumeur grâce à la formule $R = \sqrt{GD^2 + AP^2 + SI^2}$. La base de données comprend 143 signaux issus du traitement de 42 patients traités pour des cancers abdominaux ou du thorax. Les auteurs ont choisi d'utiliser un réseau de neurones à trois couches statiques accompagnées d'une couche récurrente. Ce modèle atteint une RMSE de 0.57 mm sur un horizon de 665 ms.

Les modèles présentés précédemment permettent un suivi de la tumeur en temps réel, mais ne permettent pas d'estimer les déformations de l'ensemble du poumon au cours de la respiration ni de visualiser ces déformations.

Les modèles de type PCA permettent de répondre à ce besoin.

Ainsi Bao et coll. [57] utilisent un jeu de données de 4D CT du foie de six patients pour prédire des champs de vecteur de déplacement (*Displacement Vector Field*) (DVF) de déformation du poumon avec une erreur de prédiction du mouvement moyenne de 0.57 mm. Pour obtenir ceci, ils commencent par définir un volume de référence 3D pour chaque patient (la phase d'expiration maximale) et utilisent un recalage pour calculer les DVF entre ce volume de référence et les volumes suivants. Ils utilisent ensuite un modèle de PCA sur la matrice constituée de l'ensemble des DVF et la combinent avec une estimation bayésienne pour maximiser la précision de la prédiction. Ils déduisent ainsi un DVF entre le volume de référence et le volume à prédire.

Wang et coll. [58] utilisent eux aussi un modèle de PCA, mais étudient également le lien entre les déformations externes de la peau de la poitrine du patient et les mouvements des poumons. En effet pour éviter de surexposer le patient à des radiations, les auteurs ont choisi de n'utiliser que deux CT scans 3D par patients (la phase d'inspiration et celle d'expiration) lors de leur phase de test et non une numérisation 4D. Leur jeu de donnée d'entraînement se

compose des CT scans 4D de cinq patients différents. Les auteurs commencent par construire artificiellement un 4D CT à l'aide des deux volumes dont ils disposent. Pour cela ils calculent le DVF entre les deux volumes et utilisent une régression de ridge à l'aide des volumes 4D présents dans le jeu d'entraînement. Ce 4D CT reconstruit permet d'obtenir le mouvement de la peau ainsi que le DVF au cours du cycle respiratoire. Une PCA en déduit la modélisation de la respiration associée à ce patient. Cette méthode permet d'obtenir une erreur moyenne de 0.51 mm.

D'autres études tirent parti du développement des méthodes d'intelligence artificielle plus récentes, telles que les CNN par exemple. Ainsi Terpstra et coll. [59] ont recours à des IRM fortement sous-échantillonnées et utilisent un CNN pour reconstruire le mouvement. En effet le délai entre le déplacement de la tumeur et la délivrance de la dose devrait être inférieur à 500 ms afin d'être assuré de viser correctement cette cible, mais les longs temps d'acquisition des IRM, sans compter le temps de calcul des DVF, compliquent la réalisation de cet objectif. Les auteurs utilisent alors des ciné-IRM 3D, qui permettent d'acquérir rapidement des IRM au prix d'une perte de résolution spatiale, et utilisent un CNN pour produire les DVF observés au cours de la respiration. Cette méthode est mise en place sur un jeu de données de 27 patients et obtient une erreur de registration de la cible de 1.87 mm sur un horizon de prédiction de 200 ms, avec un temps de calcul de 205 ms.

Pohl et coll. [60] ont choisi d'utiliser un modèle RNN prédisant le déplacement de trois points de repère définis proches de la tumeur. Les auteurs font l'hypothèse que le mouvement de chaque voxel est lié au déplacement de ces points de repère par une relation linéaire, ce qui permet d'en déduire un DVF de la déformation globale du poumon. Sur leur jeu de données de quatre patients, la corrélation croisée moyenne entre les images prédites et réelles était de 0.955, et le RNN mettait 119 ms à prédire la position des trois marqueurs à chaque étape.

Dans leur revue de la littérature, Lombardo et coll. [61] ont trouvé que les modèles *generative adversarial network* (GAN) sont sujets à davantage de latence que les modèles CNN, ce qui rend les GAN impropre à la localisation de tumeurs en temps réel.

Néanmoins Giger et coll. [62] ont développé un modèle basé sur des réseaux antagonistes génératifs conditionnels (cGAN) apprenant une régression entre images ultrasonores 2D et IRM 2D puis en déduisant un DVF permettant de prédire une IRM du patient en 3D capable de temps de calcul de moins de 200 ms. Ce modèle a été mis en oeuvre sur trois jeux de données provenant de deux patients sains différents et a atteint une erreur médiane de moins de 1 mm sur un horizon de prédiction de 400 ms.

2.7.2 Prédiction de survie et récurrences des cancers du poumon non à petites cellules

Il n'existe pas actuellement de modèle considéré dans la littérature comme une référence de comparaison de performances pour l'objectif de prédiction des risques de décès et récidive du patient après un traitement par SBRT [27]. Nous pouvons cependant citer plusieurs modèles répondant à cette problématique.

Différentes approches sont utilisées dans la littérature. La première utilise seulement les données cliniques récoltées lors du diagnostic du patient. Il s'agit le plus souvent de son âge et son sexe, s'il a déjà fumé auxquels sont parfois ajoutés des informations sur la pathologie du patient comme l'avancement du cancer ou le type de tumeur détectée par exemple. Nous pouvons citer l'exemple de l'étude de Kim et coll. [63] prédisant la survie à deux ans de 324 patients après leur traitement par afatinib (un traitement pris oralement luttant contre les métastases de NSCLC). Cette étude a comparé les performances de différents modèles d'apprentissage machine : LR, RF, SVM, classifieur bayésien naïf et arbre de décision, ainsi qu'un modèle Perceptron multicouche. Les meilleures performances ont été obtenues par un arbre de décision, capable de prédire la survie avec une AUC de 0.644. Cependant les auteurs n'ont pas trouvé de différences statistiques significatives entre les performances des différents modèles. Similairement, Hindocha et coll. [64] obtiennent des performances comparables lors de la prédiction de la survie à deux ans de 657 patients provenant de cinq hôpitaux différents. Leurs patients ont été traités par radiothérapie avec pour certain addition de chimiothérapie. Dans une méthodologie semblable à celle de Kim et coll., Hindocha et coll. comparent les performances de différents modèles d'apprentissage machine : LR, RF, SVM, méthode des k plus proches voisins, amplification de gradient, analyse discriminante de mélange et classifieur bayésien naïf, ainsi qu'un Perceptron. Ils obtiennent une performance en AUC de 0.717 lors de la prédiction de la survie à deux ans, et de 0.681 pour la survie sans récurrence en utilisant un modèle ensembliste englobant une analyse discriminante de mélange, une amplification de gradient et un Perceptron pour la prédiction de survie globale et un modèle des k plus proches voisins pour la survie sans récurrence. Les données semblent être les plus importantes pour la classification sont la taille de la tumeur, son stade, les modalités de traitement et le statut de fumeur du patient.

L'apprentissage profond peut également être utilisé pour ce genre de tâche, que ce soit lors de la récolte des données comme le font Yuan et coll. [65] ou lors de la classification, où l'on peut citer Lorenc et coll. [66]. Yuan et coll. [65] traitent directement avec les données médicales enregistrées textuellement dans les dossiers médicaux des patients et n'ont pas besoin de données précompilées dans des tableaux de données. Ils utilisent pour cela un modèle de traî-

tement du langage naturel afin de récupérer automatiquement les données d'intérêt et ainsi sauver un temps précieux. Ils réalisent ensuite la classification de survie globale à deux ans à l'aide d'un modèle de Cox multivariables (*multivariate Cox proportional hazards analysis*) construisant un nomogramme permettant de calculer les risques de décès des patients. Cette méthode leur permet d'atteindre une AUC de 0.825 sur un jeu de données de 14 655 patients atteints de NSCLC. Lorenc et coll. [66] quant à eux utilisent un Perceptron multicouche pour prédire la survie sans progression de 2 296 patients. Ils utilisent les informations sur le sexe du patient, son traitement, s'il a déjà fumé, ses antécédents familiaux ainsi que la présence ou non de symptômes et atteignent une performance de 0.920 d'AUC.

Nous pouvons ainsi observer l'intérêt des données cliniques dans la prédiction de survie. Elles sont également directement récupérées lors de la prise en charge du patient. Cependant les informations accessibles via ces données peuvent être limitées. En effet bien que l'on puisse inclure des données telles que le stade de la tumeur ou son volume total, les images dont se servent les radiologues pour faire leur diagnostic sont bien plus riches en informations. On voit ainsi l'intérêt des radiomiques, qui permettent d'inclure une grande quantité d'informations quantitatives décrivant l'aspect de la tumeur.

Par exemple Franceschini et coll. [67] utilisent 45 caractéristiques radiomiques extraites du volume tumoral pour prédire la survie sans progression de 102 patients traités par SBRT. Leur modèle de Cox multivariable permet d'atteindre une AUC de 0.88. Tang et coll. [68] ont comparé les performances de modèles d'apprentissage machine et d'apprentissage profond lors de l'utilisation de caractéristiques radiomiques pour la prédiction de la survie globale à trois ans. Leur jeu de données comporte 394 patients atteints de NSCLC et traités par radiothérapie. Les auteurs ont conclu que les modèles d'apprentissage profond ne présentent pas d'avantage pour la prédiction de survie, les meilleures performances étant atteintes par un modèle de forêt aléatoire avec une AUC de 0.974.

Lee et coll. [69] utilisent à la fois les caractéristiques radiomiques extraites de la tumeur (69 caractéristiques), mais aussi celles extraites de la zone périphérique à la tumeur, sur une distance allant jusqu'à 3 cm du bord de la tumeur (58 caractéristiques). Ils utilisent un ensemble de 217 patients traités par chirurgie et comparent des modèles SVM et RF pour prédire la survie sans récurrence à deux ans et atteignent une performance en AUC. de 0.71. Les auteurs ont montré que la combinaison des caractéristiques intra- et péri- tumorales améliore les performances de prédictions. Ding et coll. [70] ont étudié l'évolution des radiomiques au cours du traitement à l'aide de delta-radiomiques. Ces dernières évaluent non plus les valeurs absolues des caractéristiques radiomiques à un instant donné (en général avant le début du

traitement), mais suivent au cours du temps les variations des valeurs des différentes caractéristiques [27]. Par exemple Ding et coll. évaluent les différences radiomiques entre les CT scans pris avant le début du traitement par chimiothérapie, puis après les deuxième et quatrième cycles de traitement, sur un jeu de données comportant 132 patients atteints de NSCLC. Ils obtiennent une AUC de 0.834 pour la survie globale et 0.737 pour la prédiction du risque de récurrence locorégionale et ont montré que les modèles basés sur des delta-radiomiques obtiennent de meilleures performances que les modèles basés sur des radiomiques acquises à un instant donné.

Les données radiomiques permettent ainsi d'intégrer des informations provenant directement des scans du patient. Néanmoins elles nécessitent de définir manuellement les caractéristiques de l'image à intégrer dans le modèle. Face à cela les réseaux convolutionnels sont capables d'analyser directement les images médicales des patients. Par exemple Le et coll. [71] ont développé un modèle basé sur des CNN et disposant d'un module préprocesseur composé d'un réseau entièrement connecté (FCN) réalisant une normalisation image à image des données d'entrées. Le modèle traite des données CT, des scans de tomographie par émission de positons (PET scan) et des données cliniques pour prédire la survie globale, la récurrence locorégionale et les métastases à distance. Ce modèle est entraîné sur 298 patients traités par chimiothérapie et/ou radiothérapie dans quatre institutions différentes et est validé sur jeu de test externe de 371 patients où il atteint une AUC de 69% sur les trois résultats cliniques.

Xu et coll. [72] ont développé un modèle semblable aux delta-radiomiques, utilisant des modules CNN et RNN afin de traiter des CT scans pré et post traitement (réalisés à 1, 3 et 6 mois après le traitement) pour en déduire la survie globale deux ans après le traitement. Chaque scan est associé à un CNN pré-entraîné sur ImageNet et la sortie de ce module est intégrée à un GRU prédisant la survie du patient. L'utilisation de ce type de réseau RNN permet d'étudier la composante temporelle de l'information tout en étant capable de gérer les cas où l'information n'est pas disponible. Leur jeu de données comporte 179 patients traités par chimiothérapie et radiothérapie et leur modèle atteint une performance d'AUC de 0.74.

On peut également considérer les variables latentes reliant les images médicales aux issues cliniques comme des caractéristiques radiomiques potentielles. Celles-ci sont appelées radiomiques profondes (*deep radiomic features*).

Par exemple Yang et coll. [73] ont développé un modèle combinant radiomiques classiques, deep-radiomiques et données cliniques pour la prédiction de récurrences régionales sur un jeu de données de 84 patients traités par SBRT. Leurs caractéristiques radiomiques profondes

ont été définies à l'aide d'un encodeur U-Net pré-entraîné. Une validation croisée « tous sauf un » (*leave-one-out*) leur a permis d'obtenir une AUC de 0.843 en utilisant seulement les informations deep-radiomiques et de 0.913 en combinant les trois sources de données.

Afshar et coll. [74] se servent des radiomiques profondes pour éviter de devoir segmenter la tumeur. Pour réaliser cela, ils combinent des scans CT et PET ainsi que des données cliniques. Les données CT et PET sont encodées chacune par leur propre CNN, qui sont pré-entraînés à l'aide d'auto-encodeurs. Les sorties de ces deux CNNs représentant les données deep-radiomiques et les données cliniques sont ensuite données en entrée à un modèle de Cox. Cette méthodologie leur permet d'atteindre une AUC de 0.68 pour la prédiction de survie, 0.64 lors de la prédiction de la survie sans progression et de 0.63 pour la prédiction du risque de métastases à distance sur leur jeu de données de 132 patients. Les caractéristiques deep-radiomiques proposées par les réseaux CNNs présentent l'avantage d'être acquises sans segmentation manuelle de la tumeur, qui est une étape chronophage et sujette à l'interprétation du radiologue, mais sont cependant difficiles à interpréter, là où les radiomiques sont facilement interprétables.

Les modèles d'apprentissage profond présentent donc l'avantage d'être davantage automatisés étant donné qu'il n'est pas nécessaire de *feature engineering*. De plus la segmentation de tumeur améliore souvent les performances mais n'est pas obligatoire, contrairement aux radiomiques. Cependant ils perdent ainsi en interprétabilité. De plus ils nécessitent un accès à un plus grand nombre de patients.

Braghetto et coll. [75] ont comparé les performances de modèles utilisant des caractéristiques cliniques, radiomiques, deep-radiomiques et des réseaux de convolution lors de la prédiction de la survie à deux ans de 411 patients traités par radiothérapie ou par combinaison de radiothérapie et chimiothérapie. Les données cliniques, radiomiques et deep-radiomiques ainsi que la combinaison de différentes données d'entrées ont été traitées par différents sélecteurs de caractéristiques (comme ANOVA et PCA) ainsi que différents classifieurs (tels que SVM et RF par exemple). Enfin un CNN traitait directement avec les images 2D ou une combinaison de cinq slices 2D. Ils obtiennent une AUC de 0.59 sur les données cliniques seules, 0.61 sur les données deep-radiomiques, 0.64 avec les réseaux CNN et 0.67 pour les données radiomiques utilisant une sélection de caractéristiques par ANOVA et une RF. La combinaison des données radiomiques cliniques et deep-radiomiques atteignait également une AUC de 0.67. Braghetto et coll. en concluent que les radiomiques classiques portent l'essentiel de l'information.

2.8 Synthèse

L'étude de littérature réalisée dans ce chapitre a permis de montrer le rôle central du mouvement de la tumeur lors de la définition du traitement. Comme nous l'avons vu, l'amplitude du mouvement de la tumeur est prise en considération lors du choix du traitement. Au besoin, différentes méthodes de gestion de ce mouvement existent. Il peut s'agir de le limiter physiquement, mais de plus en plus les traitements se tournent vers le suivi de la tumeur en temps réel. Ceci nécessite de prédire le mouvement futur de la tumeur. De nombreuses études se concentrent sur le sujet, mais très peu d'entre elles étudient l'impact du mouvement sur le succès du traitement.

L'objectif de ce mémoire sera donc de développer un modèle capable de prédire le résultat du traitement de SBRT appliqué à des patients atteints de NSCLC et d'évaluer l'importance que prend le mouvement dans le succès de cette prédiction.

CHAPITRE 3 MÉTHODOLOGIE DU TRAVAIL DE RECHERCHE

L'étude de la littérature au chapitre précédent a montré l'utilité d'étudier l'impact du mouvement de la tumeur sur la survie du patient. L'objectif de ce projet sera de développer un modèle d'apprentissage capable de prédire le résultat du traitement en utilisant l'information du mouvement de la tumeur. La solution proposée consiste d'abord à évaluer le mouvement de la tumeur puis à prédire le succès du traitement.

Dans un premier temps nous nous sommes concentrés sur le mouvement respiratoire. En effet le jeu de données que nous avions à notre disposition comportait plusieurs patients pour lesquels certaines phases respiratoires du 4D CT étaient manquantes, pour des raisons de mémoire ou de temps d'acquisition. L'objectif était alors de prédire ces phases manquantes afin de reconstruire l'ensemble des séquences respiratoires. Le but était de simuler un traitement avec suivi de la tumeur en temps réel, où il est nécessaire de déterminer la position de la tumeur en avance afin de repositionner le faisceau de radiation au besoin. Nous voulions donc prédire les phases suivantes à partir d'une séquence de phases respiratoires. Pour cela nous nous sommes inspirés du modèle développé par Romaguera et coll. [23] et nous l'avons adapté à notre cas de figure. Ce modèle a été entraîné sur 165 patients. Pour tester notre modèle, nous avons prédit la dernière phase du cycle respiratoire.

Cette application a fait l'objet d'une publication d'un article à la conférence SPIE en février 2023. Le chapitre 4 détaille cette application.

Dans un second temps nous avons voulu déterminer si l'information du mouvement de la tumeur permettait de mieux anticiper le résultat du traitement. Pour cela nous avons calculé l'amplitude de déplacement des tumeurs au cours de la respiration. À cette information nous avons ajouté les informations extraites de la base de données cliniques fournie par le Centre Hospitalier de l'Université de Montréal ainsi que des informations radiomiques extraites des CT scans. Nous avons comparé différentes combinaisons de modèles prédictifs et configurations de données, en incluant ou non les informations sur l'amplitude du mouvement de la tumeur et en évaluant son impact sur les performances de prédiction. Un article présentant ces résultats a été soumis au journal *Medical Physics*. Cet article est présenté au chapitre 5.

La Figure 3.1 présente les deux modèles développés au cours de ce mémoire ainsi que leur interaction.

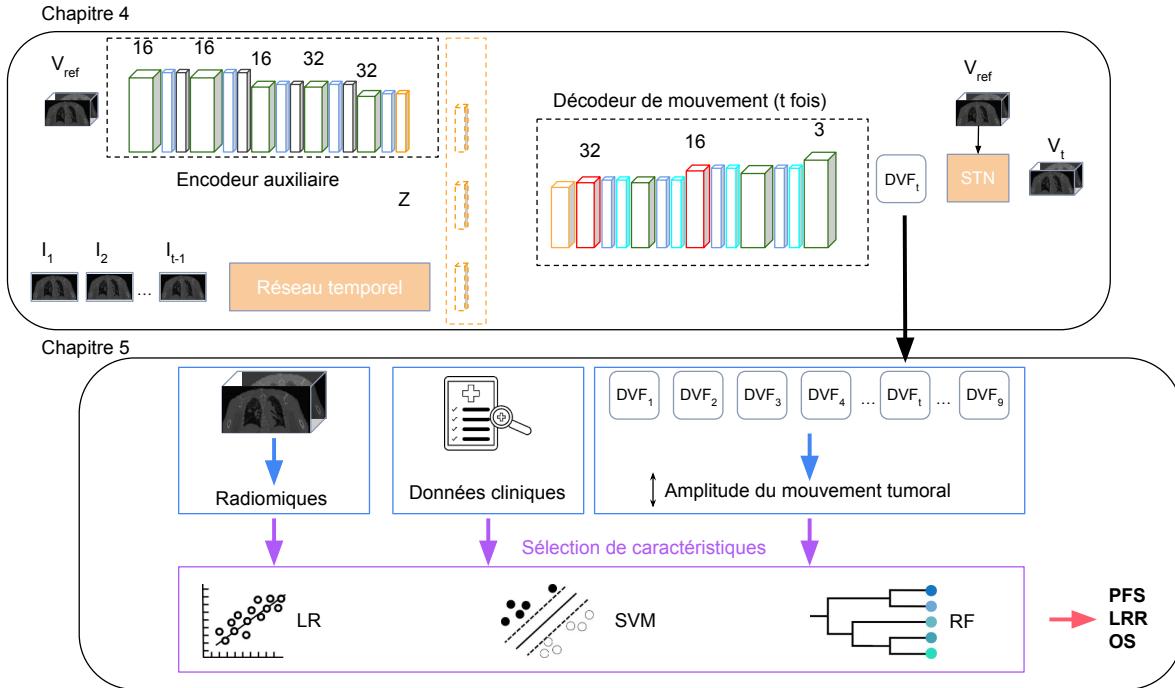


FIGURE 3.1 Schéma du processus développé au cours de ce mémoire. Le premier modèle génère les volumes et phases respiratoires manquantes à notre jeu de données. Le second modèle utilise les données cliniques, radiomiques et l'amplitude du mouvement de la tumeur pour prédire l'état de santé du patient deux ans après la fin de son traitement.

3.1 Jeux de données utilisés

Les données proviennent du Centre hospitalier de l'Université de Montréal. Il s'agit de leurs patients traités par SBRT pour des NSCLC entre 2009 et 2019. Les données utilisées étaient de deux types différents : des données d'imagerie (les CT scans en format DICOM) et l'information clinique.

Pour chacun des deux objectifs de recherche, un sous-ensemble différent du jeu de donnée a été utilisé.

Objectif 1 : Pour étudier les déformations successives du poumon au cours du cycle respiratoire, nous avons besoin d'avoir accès à l'ensemble des phases du cycle respiratoire, mais nous ne nécessitons pas d'informations cliniques. Or dans notre jeu de données, de nombreux patients ne disposaient pas de l'ensemble des 10 phases attendues. Nous avons donc limité notre jeu de donnée aux patients présentant un ensemble des phases respiratoires complet. Ceci concerne 165 patients.

Objectif 2 : Lorsque nous nous concentrons sur l'état du patient suite au traitement, nous avons désormais besoin des informations cliniques. En revanche nous n'utilisons pas l'ensemble des phases respiratoires du patient, ce qui nous permet de conserver davantage de patients. Dans cette partie nous disposons d'un ensemble de 381 patients.

3.2 Étape de prédiction des déformations

3.2.1 Explication de la démarche

La première étape de ce projet est de développer un modèle permettant d'anticiper les déformations du poumon liées à la respiration. Nous nous sommes basés sur un modèle développé par Romaguera et coll. [23] prédisant les déformations du foie au cours de la respiration à partir d'images IRM. Pour l'adapter au cas d'images CT des poumons, nous avons gardé la même configuration globale que celle proposée par Romaguera et coll., en adaptant principalement le nombre et la taille des couches des encodeurs et décodeurs.

Nous avons utilisé des CT scans 4D composés de 10 phases respiratoires. Suite à des problèmes de mémoire liés à la taille conséquente des données utilisées, nous avons dû modifier le préprocessing des données afin de redimensionner les scans et être capables de réaliser des opérations de calcul sur ces données, ce qui risque de réduire légèrement la précision de l'estimation des déformations du poumon.

3.2.2 Validation de la démarche

Nous avons utilisé un jeu de données constitué de 165 patients disposant de scans CT en 4 dimensions, chacun constitué des 10 phases respiratoires. Les données avaient une résolution de $512 \times 512 \times [105-340]$ que nous avons redimensionnés par découpe puis ré-échantillonnage à une taille de $128 \times 128 \times 64$. La validation du modèle a été effectuée sur un jeu de test constitué de 29 patients ainsi qu'à l'aide d'une validation croisée à 5 blocs effectuée sur le jeu d'apprentissage. La comparaison des résultats entre ces deux modes de tests permet une bonne approximation des performances de généralisation du modèle.

3.3 Étape de prédiction des issues du traitement

3.3.1 Explication de la démarche

La deuxième étape de ce projet est de développer un modèle capable de prédire l'issue du traitement (à savoir les cas de décès et de récurrences à diverses échelles). L'objectif de cette

étape était également de mettre en avant le rôle du mouvement de la tumeur dans l’issue du traitement.

De nombreux patients quittent l’étude en cours de suivi, voire directement après la fin du traitement. Pour garder une cohérence entre l’ensemble des patients nous avons choisi d’étudier l’issue du traitement à une durée fixe et non l’issue finale du traitement. Notre étude de la littérature nous a montré que les dates de deux, trois et cinq ans étaient les plus couramment utilisées, en particulier la limite de deux ans étant donné que la majorité des cas de récurrence surviennent en majorité (à 84%) dans les deux ans après la fin du traitement [76]. Ceci concordait avec la répartition de nos données, dont le temps de suivi moyen est de 3.5 ans. La limite à deux ans nous permet donc également d’éviter d’écarter trop de patients.

Nous avons tout d’abord essayé de mettre en place des modèles de type CNN. Cependant ces modèles n’étaient que peu performants, probablement du fait du nombre limité de patients à disposition. Ceci est également en cohérence avec l’étude de littérature réalisée au chapitre précédent. Nous avons donc décidé d’utiliser des données radiomiques. Suite à notre étude de littérature, nous avons décidé d’utiliser les PTV comme région d’intérêt plutôt que l’IGTV.

Nous avons comparé les performances de différents modèles d’apprentissage machine, tels que les SVM, LR et RF.

3.3.2 Validation de la démarche

Suite à notre décision d’étudier l’issue du traitement deux ans après la fin du traitement, ainsi que de la nécessité de disposer des segmentations des tumeurs, nous avons pu conserver un jeu de données constitué de 381 patients. Nous disposions des informations cliniques, des CT scans 4D et des segmentations (comprenant notamment les poumons, les GTV, IGTV et PTV).

Nous avons validé nos modèles à l’aide d’une validation croisée à 6 blocs, ou à 20 blocs dans le cadre des comparaisons statistiques. Les performances des modèles sont donc présentées sur 6 blocs afin de disposer de jeux de test de taille plus conséquente. Cependant pour la réalisation des tests statistiques nous souhaitions disposer de davantage d’effectifs, d’où notre choix d’utiliser 20 blocs plutôt que 6.

CHAPITRE 4 PRÉDICTION DE 4DCT EN RESPIRATION LIBRE DE DÉFORMATIONS DU POUMON EN UTILISANT DES AUTO-ENCODEURS PROBABILISTES

4.1 État de l'art

La prédiction des mouvements respiratoires durant la respiration libre s'effectue à l'aide de différentes méthodes. Les modèles de mouvement respiratoire utilisent des signaux analogues en lien avec le mouvement pour estimer ce dernier [77]. On les classe en deux groupes : les modèles spécifiques à un patient et les modèles de population.

4.1.1 Modèles à patient unique

Les modèles spécifiques à un patient utilisent des données provenant d'un unique patient pour entraîner le modèle de génération de mouvement. Parmi ce type de modèles on trouve d'abord la PCA. Par exemple Wang et coll. [58] ont présenté un modèle entraîné sur des données 4D et nécessitant uniquement deux CT scans 3D une fois entraîné, générant artificiellement les 4DCT à partir des deux CT 3D. Une analyse en composantes principales est ensuite utilisée pour estimer la relation existante entre les mouvements de la peau mesurés en surface et les déformations internes du poumon.

Bao et coll. [57] quant à eux utilisent les amplitudes du mouvement respiratoire du foie interne et externe pour calculer un DVF, et une PCA sélectionne le DVF du jeu d'entraînement le plus ressemblant à celui généré. Ce DVF similaire est ensuite appliqué au CT scan pour estimer le mouvement respiratoire propre à ce cas particulier.

D'autres modèles peuvent générer directement les données d'intérêt, comme par exemple celui de Giger et coll. [62], qui utilise des cGAN pour apprendre une régression entre des images 2D acquises par ultrasons et une IRM 4D.

Étant donné que les mouvements respiratoires peuvent varier fortement d'un patient à l'autre, les modèles à patient unique performent souvent mieux que les modèles de population [23, 78].

4.1.2 Modèles de population

Les modèles de population utilisent des données provenant de différents patients, ce qui leur permet d'être plus robustes que les modèles à patient unique en ayant eu accès à davantage de types de mouvements différents. Ils peuvent donc être appliqués à des patients inconnus,

là où les modèles à patients uniques ne sont construits que pour des patients spécifiques.

Les PCA peuvent aussi être appliquées à des modèles de population, tels que l'ont fait Fayad et coll. [79] pour prédire des phases respiratoires à partir de deux IRM 3D acquises lorsque le patient retient sa respiration ainsi que d'images IRM en 2D. Garau et coll. [80] ont développé un modèle basé sur des régions d'intérêt. Une PCA est entraînée sur un jeu de données de 4DCT pour apprendre les mouvements 3D du poumon, auxquels on ajoute des images CT 2D centrées sur la tumeur pour approfondir le mouvement en deux dimensions. L'inférence se fait sur des images IRM en 2D et permet d'estimer les volumes CT en trois dimensions au cours de la respiration.

Les modèles de type RNN sont également utilisés par les modèles de population. Par exemple Pohl et coll. [60] prédisent les déplacements de la tumeur à partir de trois repères implantés près de la tumeur.

4.2 Méthodes

Dans cette partie nous cherchons à prédire les déformations du poumon au cours d'un cycle respiratoire à l'aide de CT scans 4D. Nous utilisons un modèle inspiré de celui de Romaguera et coll. [23] qui prédit les déformations du foie au cours de la respiration à l'aide de données IRM et d'imagerie ultrasonore.

4.2.1 Protocole d'acquisition

Dans cette étude, nous avons utilisé 165 CT 4D enregistrés lors de la planification du traitement. Chacun de ces 4D CT consiste en 10 CT scans 3D, chacun représentant une phase respiratoire. Le début du cycle respiratoire a été fixé au milieu de l'expiration. Les CT scans 3D avaient une taille de 512 par 512 pixels, avec un nombre de slices variant entre 105 et 340. L'espace entre les pixels est compris entre $0.98 \times 0.98 \text{ mm}^2$ et $1.5 \times 1.5 \text{ mm}^2$, et l'épaisseur des slices varie entre 1.0 et 3.0 mm. Les structures d'intérêt pour cette étude étaient les poumons et l'IGTV, qui ont été segmentés par un oncologue.

4.2.2 Prétraitement des données

Les volumes d'origine étant très volumineux, ils ne pouvaient être utilisés en tant que tels à cause de limitations au niveau du GPU. Nous avons donc choisi de réduire leur taille et de les ré-échantillonner. Pour limiter au maximum la perte de qualité associée à ces opérations, nous avons d'abord coupé les volumes autour des poumons, de façon à nous concentrer

uniquement sur la région qui nous intéresse. Cependant chaque patient a une taille de poumon variable, et nous souhaitions disposer de volumes de taille constante. De plus, même après ce recadrage, les volumes restaient trop lourds pour être traités par notre modèle. Nous avons donc redimensionné chaque volume par ré-échantillonnage, en choisissant un volume de taille constante $128 \times 128 \times 64$. Finalement, chaque CT a été normalisé en utilisant un recentrage autour de la moyenne et en standardisant les valeurs par l'écart-type.

4.2.3 Définition du modèle

La figure 4.1 présente le modèle choisi choisi pour la prédiction des futures phases respiratoires.

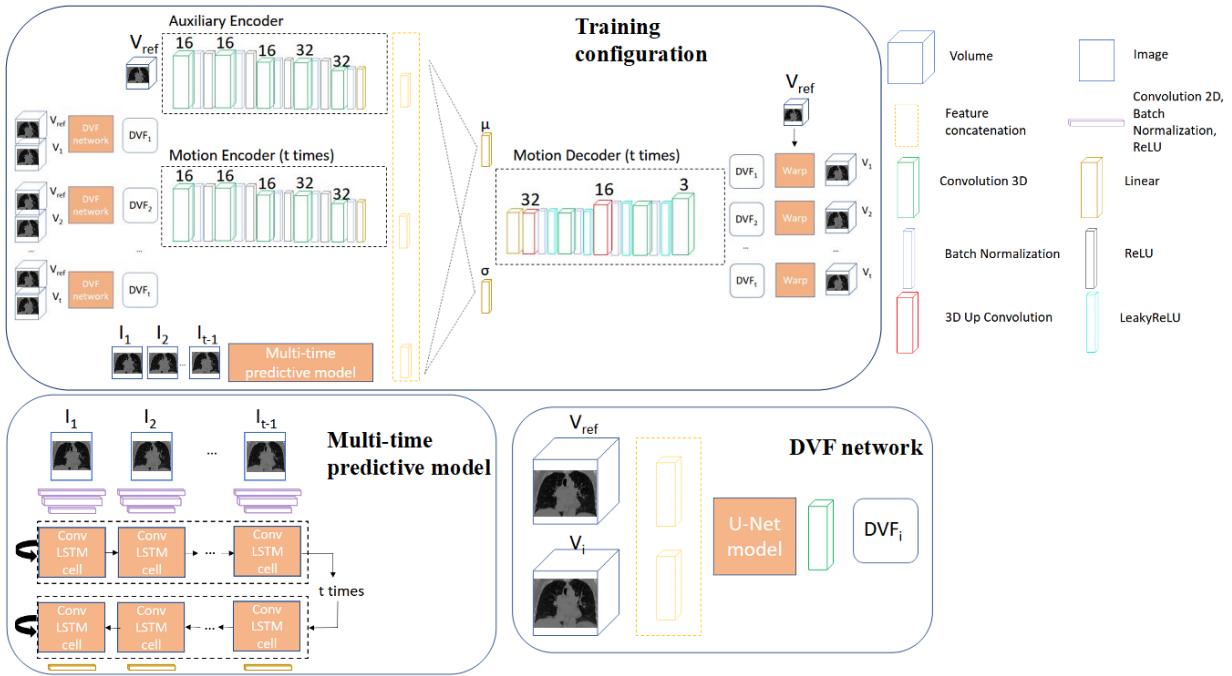


FIGURE 4.1 Architecture du modèle de prédiction des déformations. Dans sa configuration d'entraînement, le modèle utilise un auto-encodeur pour apprendre les mouvements internes au sein du poumon. Pour prédire les futurs volumes, un second module temporel génère les images suivantes. Il utilise un réseau prédisant les DVF pour déformer le volume de référence et générer ainsi le volume prédit.

Supposons que l'on veuille prédire la phase respiratoire V_t à l'instant t d'un patient p . Durant la phase d'apprentissage le modèle prend en entrée un volume de référence V_{ref} et des volumes cibles $\{V_1, V_2, \dots, V_t\}$. Le volume de référence $V_{ref} = V_0$ correspond à la première phase du cycle respiratoire, c'est-à-dire à la moitié de la phase d'expiration. Les volumes cibles

$\{V_1, V_2, \dots, V_t\}$ correspondent aux futures phases respiratoires du patient.

On utilise un réseau VoxelMorph [81] préentraîné pour estimer le DVF représentant la déformation ϕ_i entre le volume de référence V_{ref} et chacun des volumes cibles V_i . On obtient ainsi les DVF $\Phi = \{\phi_1, \phi_2, \dots, \phi_t\}$.

Le modèle de prédiction des phases est un réseau neuronal convolutionnel composé de trois branches : un encodeur du mouvement, un encodeur auxiliaire qui traite le volume de référence, et un réseau temporel. Ce réseau temporel prend en entrée une séquence d'images 2D extraites des phases respiratoires précédentes du patient (donc connues)

$I_{seq} = \{I_1, I_2, \dots, I_{t-1}\}$, et fournit en sortie des vecteurs de caractéristiques qui seront eux-même utilisés comme entrées d'un réseau encodeur-décodeur constitué de *Convolutional Long Short-Term Memory* (ConvLSTM).

La fonction objectif du réseau revient à maximiser la probabilité conditionnelle $p(\Phi|I_{seq}, V_{ref})$ d'obtenir la séquence de déformations réelles. Cette probabilité peut également s'exprimer comme :

$$p(\Phi|I_{seq}, V_{ref}) = \int_z p(\Phi|z, I_{seq}, V_{ref})p(z)dz \quad (4.1)$$

Résoudre cette équation demanderait de trop nombreux échantillons pour être réalisable. Nous nous concentrerons donc sur le postérieur $p(z|\Phi, I_{seq}, V_{ref})$. Celui-ci peut s'estimer à l'aide de la distribution $q_\Psi(\cdot)$ de paramètre Ψ . En pratique l'encodeur apprend à approcher l'estimation :

$$q_\Psi(z_i|I_{seq}, V_{ref}) = N(\mu(\phi_i, I_{seq}, V_{ref}), \sigma(\phi_i, I_{seq}, V_{ref})) \quad (4.2)$$

Le réseau apprend la moyenne μ et la covariance σ à partir des données. La différence entre les distributions p et q_Ψ est minimisée à l'aide de la divergence de Kullback-Leibler (ou entropie relative) notée KL , que l'on insère dans la fonction de perte suivante :

$$L = \frac{1}{t} \sum_{i=1}^t L_{rec}(V_{ref} \circ \phi_i, V_t) + KL(q_\Psi(z_i|I_{seq}, V_{ref})||p(z_i)) \quad (4.3)$$

Cette fonction de perte est minimisée en utilisant la descente stochastique du gradient. Le terme de reconstruction L_{rec} est calculé entre les volumes de référence modifiés de la déformation renvoyée par le décodeur ($V_{ref} \circ \phi_i$) et le volume V_t :

$$L_{rec} = L_{simi}(V_i, \hat{V}_i) + \alpha L_{smooth}(\hat{\phi}_i) \quad (4.4)$$

avec \hat{V}_i le résultat de la déformation du volume V_{ref} par l'estimation du déplacement ϕ_i , α un paramètre de régulation et L_{smooth} la fonction calculant la différence entre les positions

3D voisines p de la façon suivante :

$$L_{smooth}(\hat{\phi}_i) = \sum_{p \in \mathbf{R}^3} \|\nabla \phi(p)\|^2 \quad (4.5)$$

tandis que L_{simi} calcule la similarité entre les voxels prédicts et réels.

Les sorties de ces trois branches sont ensuite concaténées et associées pour former une distribution gaussienne. L'espace latent de cette distribution est ensuite concaténé avec la représentation du volume de référence obtenue en sortie de l'encodeur auxiliaire, ainsi qu'avec les plongements du module temporel. Le décodeur renvoie le vecteur concaténé dans l'espace d'origine et génère un DVF à partir de ce vecteur. Il ne reste plus qu'à un STN [82] de générer les volumes prédicts.

Lors de la phase d'inférence, on ne connaît pas les volumes cibles utilisés par VoxelMorph. On n'utilise donc plus la branche de l'encodeur traitant les déformations entre les volumes. À la place le décodeur a accès à la concaténation des résultats de l'encodeur chargé du volume de référence, de la sortie du module temporel et de la distribution gaussienne apprise durant la phase d'entraînement. La partie décodeur fonctionne de la même manière que durant la phase d'entraînement. La Figure 4.2 présente un schéma de cette configuration.

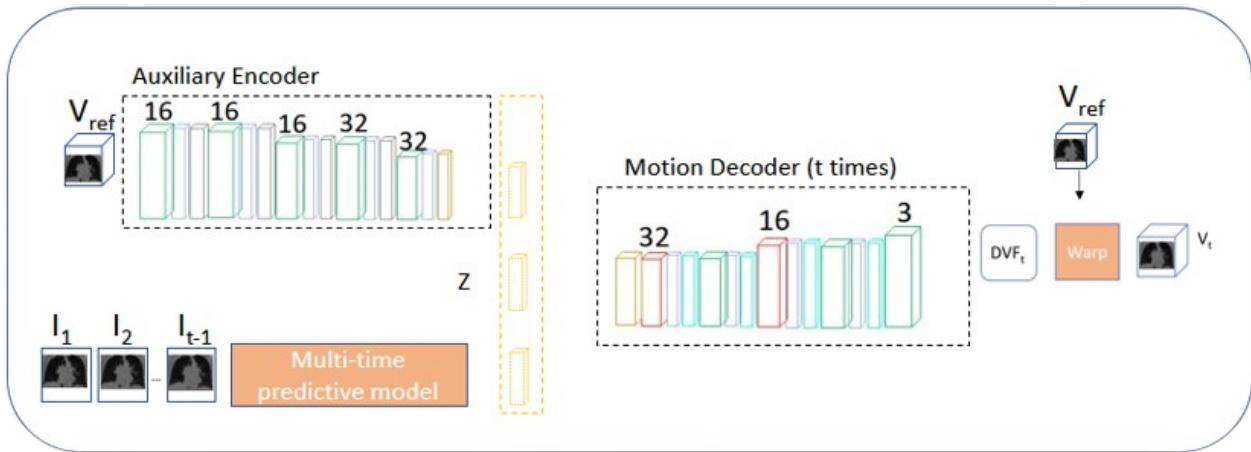


FIGURE 4.2 Architecture du modèle de prédiction des déformations dans sa configuration de test.

4.2.4 Entraînement du modèle

Une séquence de neuf phases respiratoires a été utilisée comme entrée afin de prédire la dernière phase du cycle respiratoire. L'algorithme d'optimisation Adam a été utilisé pour

optimiser les paramètres du modèle. Le taux d'apprentissage a été fixé à 10^{-4} à l'initialisation et était divisé par 10 lorsque la fonction de perte ne diminuait pas durant 4 époques consécutives. Un arrêt prématué a été utilisé pour interrompre l'apprentissage lorsque les performances ne s'amélioraient pas pendant 15 époques consécutives. La taille des lots a été fixée à 5, de façon à ne pas surcharger le serveur de calcul du fait de la taille importante des données utilisées. Le modèle est codé via PyTorch 1.11.0 [83]. L'entraînement a été réalisé à l'aide d'un GPU TITAN X Pascal disposant de 64 Gb de RAM.

4.3 Résultats

4.3.1 Génération des DVF

Les performances du module générant les DVF ont été évaluées à l'aide d'une validation croisée à 5 blocs, chaque bloc étant composé de 39 patients. Les métriques de Corrélation croisée négative (NCC), MSE et SSIM entre les volumes d'origine et ceux reconstruits par application des DVF ont servi à évaluer numériquement la précision du module. On obtient une performance de NCC égale à 0.87 (± 0.07), une MSE de 0.02 (± 0.04) et une SSIM valant 0.88 (± 0.08).

4.3.2 Modèle général

Une validation croisée à 5 blocs a été mise en place afin d'évaluer les performances du modèle lors de la prédiction du mouvement du poumon. La précision du modèle a été évaluée à l'aide des métriques de NCC, MSE et SSIM. Durant l'entraînement la NCC a servi de fonction de perte. Les performances ont également été évaluées sur un jeu de test séparé constitué de 29 patients. Les performances du modèle à la fois sur la validation croisée et sur le jeu de test sont présentées dans le Tableau 4.1.

Pour obtenir une évaluation visuelle des volumes générés par le modèle, la Figure 4.3 présente une comparaison entre les volumes réels et les volumes prédits. On représente plusieurs points de vue. Cette comparaison a été réalisée en utilisant la librairie VV [84], en superposant les deux volumes et en surlignant les différences entre les deux.

Nous avons également évalué l'erreur de recalage de cibles (*Target Registration Error*) (TRE) sur les tumeurs. La Figure 4.4 présente l'exemple des positions de la tumeur d'un patient sur les phases réelles et reconstruites. Le TRE moyen est de 1.1 ± 0.9 mm.

TABLEAU 4.1 Performances numériques obtenues par le modèle de prédiction des déformations lors de la prédiction de la dernière phase du cycle respiratoire. Les métriques NCC, MSE et SSIM ont été évaluées sur des sous-ensembles de 26 patients durant la phase de test de la validation croisée à 5 blocs, et sur le jeu de test séparé constitué de 29 patients durant la phase de test du modèle. Les flèches orientées vers le haut ou le bas indiquent si une meilleure performance est indiquée par une mesure plus élevée (\uparrow) ou plus faible (\downarrow).

	NCC (\uparrow)	MSE (en mm) (\downarrow)	SSIM (\uparrow)
Validation croisée	0.81 (± 0.07)	0.04 (± 0.03)	0.81 (± 0.07)
Jeu de test séparé	0.81 (± 0.06)	0.03 (± 0.01)	0.82 (± 0.06)

4.4 Discussion

La Figure 4.3 semble indiquer que les prédictions du modèle sont visuellement correctes. La majorité des erreurs de prédictions se trouvent dans la partie inférieure des poumons, qui est le lieu où les déplacements sont les plus marqués [78]. Les résultats présentés dans le Tableau 4.1 prouvent que notre modèle sait prédire les déformations du poumon. De plus nous obtenons des résultats supérieurs à ceux de Romaguera et coll. [23] dont le modèle a servi de source d'inspiration pour notre étude. Leur modèle atteignait des performances moyennes de 0.75 ± 0.12 en termes de SSIM, une MSE moyenne de 0.12 ± 0.07 mm et une NCC moyenne de 0.77 ± 0.11 . Nous avons donc obtenu une augmentation des performances de respectivement 9.3%, 75.0% et 5.2% pour la SSIM, MSE et NCC. Cependant il convient de noter que leurs résultats ont été obtenus sur un jeu de données constitué d'images IRM et concernant les déformations du foie. Dans notre jeu de données de CT scans du poumon, les amplitudes de respiration, les déformations des poumons ainsi que la régularité du mouvement de respiration pourraient être différentes et influer sur les résultats obtenus.

Au-delà de l'analyse visuelle, les résultats présentés dans le Tableau 4.1 montrent une grande similarité entre les performances obtenues sur les jeux de validation croisée et sur le jeu de test. Ceci semble indiquer que notre modèle est suffisamment robuste pour être appliqué à un ensemble varié de mouvements respiratoires et qu'il arrive à généraliser correctement aux cas inconnus. Notre modèle est un modèle de population et ainsi a été validé à l'aide de cas non déjà vus lors de l'entraînement. Mais il pourrait tout aussi bien être affiné (*fine-tuned*) sur un jeu de données spécifique à un patient qui contiendrait de nombreux cycles respiratoires acquis sur un même patient. Ceci lui permettrait d'apprendre spécifiquement les schémas respiratoires propres à ce patient et ainsi proposer des déformations respiratoires plus adaptées. Selon les résultats de Romaguera et coll. [23], l'application d'un modèle entraîné sur un jeu

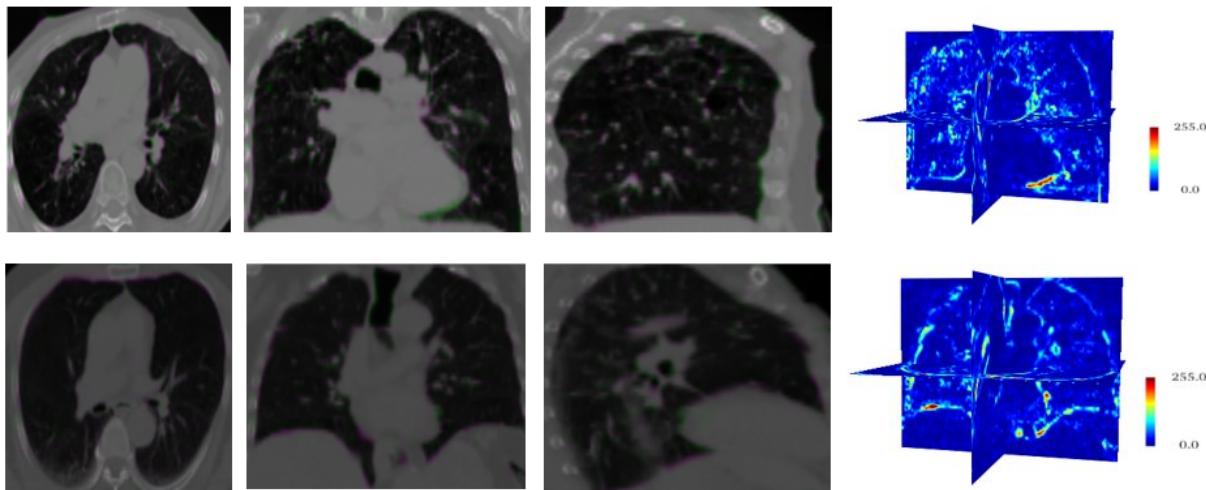


FIGURE 4.3 Comparaison des volumes des phases 90% réelles et prédites par le modèle pour différentes tranches et orientations pour deux patients différents. On remarque que les distributions d'intensité sont similaires entre les volumes prédits et attendus dans les deux cas. Les pixels colorés en vert représentent les endroits où les valeurs des volumes prédits étaient plus élevées que celles attendues, et les pixels violetts indiquent les pixels pour lesquels les valeurs prédites étaient plus faibles que celles du volume réel. La dernière colonne présente la différence en valeurs absolues et en 3 dimensions entre les volumes réels et les phases prédites.

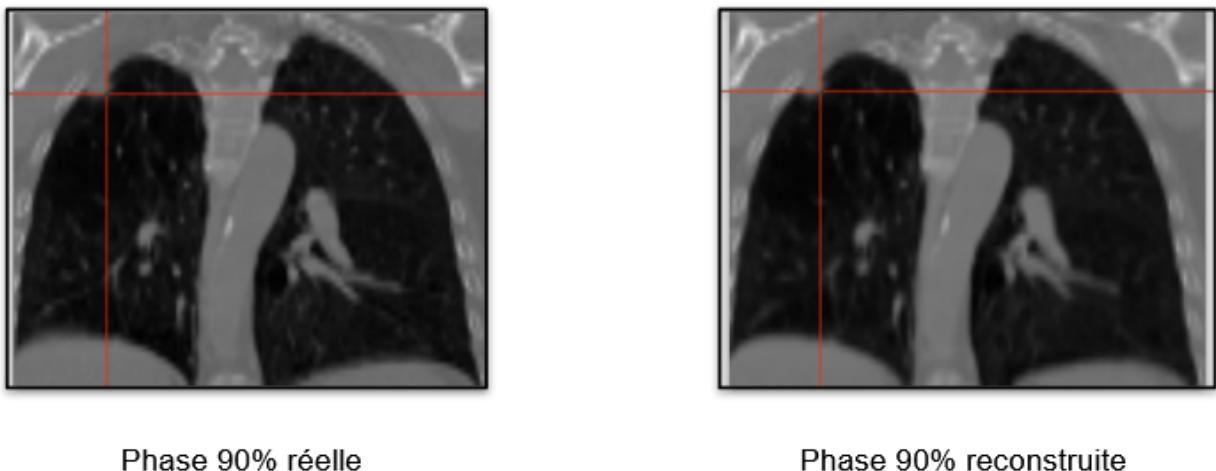


FIGURE 4.4 Exemple de calcul d'erreur de recalage de la tumeur. Les positions de la tumeur sont évaluées au niveau de la cible rouge. La distance entre ces deux points forme l'erreur de recalage.

de données spécifique à un patient donné permet d'améliorer grandement les performances de prédiction. Cependant dans notre cas cela nécessiterait de réaliser de nombreuses acquisitions via CT scans, ce qui augmenterait la dose de radiation reçue par le patient.

Étant donné que notre modèle commence par prédire les DVF entre les volumes de référence et les volumes cibles avant d'utiliser le STN pour générer le volume prédit, on peut suivre le mouvement de la tumeur, à condition que cette cible soit définie dans le volume de référence. Durant la phase d'inférence du volume, le temps de calcul moyen était de 63 ± 2.4 ms sur un NVIDIA Titan Pascal X GPU disposant de 64 Gb de RAM. Ce temps est inférieur au temps de latence des systèmes de radiation, qui sont en général de l'ordre de 100 à 400 ms [51]. Notre système est donc capable de prédictions en temps réel et pourrait ainsi permettre le suivi d'une tumeur en temps réel lors d'une radiothérapie.

Il convient néanmoins de garder à l'esprit que, à cause de limites computationnelles, nous avons été contraints de redimensionner nos scans, ce qui a entraîné une perte de qualité et a donc pu limiter l'information à disposition dans les images.

4.5 Conclusion

Au cours de ce chapitre nous avons présenté un modèle de mouvement probabiliste permettant de prédire les déformations 4D du poumon à partir de CT scans d'un cycle respiratoire, en utilisant un volume de référence de prétraitement et des coupes 2D acquises en temps réel. Ce modèle peut ainsi être utilisé pour anticiper les prochaines phases respiratoires, ou bien inférer une plusieurs phases manquantes dans un jeu de données 4D CT.

**CHAPITRE 5 ARTICLE 1 : PREDICTING EARLY STAGE LUNG
CANCER RECURRENCE AND SURVIVAL FROM COMBINED TUMOR
MOTION AMPLITUDE AND RADIOMICS FEATURES ON
FREE-BREATHING 4D-CT**

Cet article soumis au journal *Medical Physics* le 14 juillet 2024 présente une comparaison de différents modèles et données d'entrées utilisés pour déterminer les risques de récidives et décès pour des patients atteints de cancers du poumon. L'article inclus des expériences avec et sans données sur l'amplitude de déplacement de la tumeur au cours du cycle respiratoire.

Auteurs

Émilie Ouraou¹, Marion Tonneau², William T. Le¹, Houda Bahig², Samuel Kadoury¹

Affiliations

¹ Computer and Software Engineering Department, Polytechnique Montréal, Montréal, QC H3T 1J4, Canada

² Centre hospitalier de l'Université de Montréal (CHUM), Montréal, QC H2X 3E4, Canada

Background : Cancer control outcomes of lung cancer is hypothesized to be affected by several confounding factors, including tumor heterogeneity and patient history, which has been hypothesized to mitigate the dose delivery effectiveness when treated with radiation therapy. Providing an accurate predictive model to identify patients at risk would enable tailored follow-up strategies after treatment.

Purpose : This study introduces a predictive model for 2-year recurrence and survival in non-small-cell lung cancer patients treated with curative-intent stereotactic body radiation therapy, which combines patient clinical features, baseline CT radiomics, and tumor displacement amplitude as predictors.

Methods : A cohort of 381 patients treated for primary lung cancer with radiotherapy was collected for planning CT and clinical information. From this cohort, 101 patients (26.5%) experienced cancer progression (locoregional/distant metastasis) or death within 2 years of the end of treatment. Imaging data was analyzed for radiomics feature from the tumor segmented image, as well as tumor motion amplitude measured on 4D-CT. Three different automatic classification approaches — logistic regression, support vector machine and random forest — were evaluated to predict the overall outcomes.

Results : A 6-fold cross-validation study yielded an area under the receiver operating characteristic curve of 68% for progression-free survival when combining clinical data with radiomics features and tumor motion, the later showing significant improvement compared to standard clinical data. Model performances for loco-regional recurrence and overall survival sub-outcomes were established at 71% and 65% respectively.

Conclusions : Both tumor respiratory motion and planning CT radiomics features show promising predictive value for 2-year tumor control and survival, indicating potential need for improvement of motion management strategies in future studies using machine learning-based prognosis models.

5.1 Introduction

Lung cancer is the leading cause of cancer-related death, with non-small-cell carcinoma (NSCLC) being the most common form accounting for 85% according to the latest cross-sectional epidemiological US analysis [10, 85, 86]. The treatment of choice for inoperable early-stage NSCLC patients is stereotactic body radiotherapy (SBRT), a form of targeted high dose radiation therapy [2,3]. Despite increased adoption, short-term (2-year) recurrence risks remain, occurring in approximately 20% of cases, which negatively impacts survival ; the median overall survival (OS) being 40 months [3–5, 87]. Outside of descriptive population statistics, having accurate and personalized radiotherapy outcome prediction would allow patient risk

stratification and individualized surveillance strategy post-treatment and determine which patients may benefit from treatment intensification with immunotherapy.

Recently, multiple prognostic tools have been proposed to predict recurrence and survival risks for several types of cancer [88–90]. Many of these rely on patient clinical data, with the common ones being patient demographics (*e.g.* age, sex), known risk factors (*e.g.* smoking), biomarkers status, pathological characteristics (*e.g.* cancer stage), and therapeutic settings (*e.g.* treatment type). Yuan et al. [65] for instance obtained an area under the receiver-operating characteristic curve (AUC) of 0.825 when predicting the 2-year OS on a cohort of 14 655 NSCLC patients. Machine learning techniques such as logistic regression (LR), support vector machines (SVM) and random forest approaches (RF) meanwhile offer reduced computational overhead compared to deep learning methods. Kim et al. [63] and Hindocha et al. [64] observed similar performance at 0.644 for a 324 patients cohort and 0.717 for a 657 patients cohort, respectively. The later also developed a model for progression-free survival (PFS) with an AUC of 0.681. Still, using a neural network approach, Lorenc et al. [66] predicted PFS with 0.920 AUC on a lung cancer cohort of 2296 patients.

A subset of the recent work focuses on analyzing medical images for biomarker discovery via the field of radiomics to extract a large number of quantitative features from images in order to use structured data describing the qualitative features of the tumor [91–93]. In radiation therapy in particular, along with the baseline patient characteristics is the planning computed tomography (CT) image, which can be processed to extract these radiomic features [86, 94]. Lee et al. [69] compared the predictive power of radiomics feature computed on the whole tumor *vs* the intratumoral and peritumoral regions and recorded their highest performance (0.71 AUC on 217 patients cohort) with the whole tumor. Ding et al. [70] studied 132 NSCLC patients instead undergoing chemotherapy and obtained 0.834 OS AUC using delta radiomics between CT obtained pre- and post-treatment, and 0.737 AUC for locoregional recurrence (LRR) using the same method.

Finally, some previous works bypassed manual feature selection entirely by using deep learning approaches. Recent trends in medical image analysis inspired by recent progress in natural images using open source datasets such as ImageNet, has led multiple studies to adopt convolutional neural networks to analyze the entire CT scan directly. Xu et al. obtained 0.74 OS AUC with 179 locally advanced NSCLC patients undergoing combined chemoradiotherapy using the baseline CT. Afshar et al. trained an OS model on a cohort of 132 lung cancer patients leading to a 0.68 AUC, distant metastasis (DM) with 0.63 AUC and PFS with 0.64 AUC by combining both CT and positron emission tomography imaging data along with clinical data. Braghetto et al. [75] proposed to combine radiomic and imaging data, showing

that radiomic features were responsible for the most of the predictive value compared to the deep-learning extracted ones. The major limitations of these studies remains the low sample size, with respect to the high dimensionality of the inputs (2D or 3D volumes).

Lung cancer treatment via radiation therapy is particularly affected by respiratory motion, *i.e.* the cyclical contraction and expansion of the lungs. Effective dose delivery while avoiding the surrounding organs at risk was shown to correlate with tumor control, since the respiratory motion entails a displacement of the tumour, which can have an amplitude up to 10 cm [17, 78]. Usually this motion is controlled by using either abdominal compression, respiratory gating or tumor tracking with robotic radiotherapy, which helps to partially mitigate the motion amplitude of the tumor when it is more than 1cm. However for the majority of patients for whom the tumor motion amplitude is less than 1cm, the treatment is done during free-breathing and a 5mm margin is added to the internal target volume to account for the motion uncertainties. However very few studies report the impact of tumor motion on the cancer outcome [22]. Given that temporal planning CTs (4D-CT) are becoming more frequent in pre-treatment planning, capturing the trajectory of the tumor's motion, it is possible that the amplitude of this motion could serve as a predictor for recurrence and survival. The hypothesis being that larger motion patterns causing the tumor to receive less effective dose is correlated with recurrence outcomes.

This study investigates the predictive value of tumor motion amplitude measured on the 4D temporal planning CT of early-stage NSCLC patients undergoing curative-intent radiotherapy retrospectively collected from a single institution. Our model aims to predict 2-year LRR, OS and PFS using machine learning methods including LR, SVM and RF. Specifically, we compare a baseline model using only clinical factors vs models including either radiomics, tumor motion amplitude, or a combination of both. Experiments evaluate the effect motion modeling on the overall accuracy when using RF models which integrates motion for both OS and PFS outcomes.

5.2 Materials and Methods

5.2.1 Patient cohort

A cohort of lung cancer patients treated at the Centre Hospitalier de l'Université de Montréal (CHUM) in Québec, Canada is used in this study. All patients were treated between 2009 and 2019 for inoperable (or refusal of surgery for 54 patients) early-stage NSCLC using SBRT on one or more lesions. The inclusion criteria are : (1) aged 18 years or older at the time of diagnosis, (2) performance status ECOG (Eastern Cooperative Oncology Group Criteria) \leq

2, (3) TNM stage I or II [95,96], (4) ineligible for surgery due to advanced age, comorbidity, or refusal of operation, and (5) for patients treated using CyberKnife (CK), absence of fiducials. Clinical examination and CT scan of the thorax were performed pre-treatment, as well as a biopsy when possible. Patients with missing data, either CT scans or follow-ups, are excluded with the assumption of random non-correlated sources of absent information.

Study endpoints — LRR, OS and PFS — are defined as the patient at risk time from the end of treatment to the time of first recorded disease progression (recurrence or death), or which ever came first in the case of PFS. Recurrence is defined as : 1) local if a tumor appeared in the same pulmonary lobe, 2) regional if it appeared in lung or in the mediastinum, and 3) distant if it recurred anywhere else. For progression-free patients, a minimum follow-up of 2 years after the end of treatment is required for study inclusion — no minimum time is required for patients with recurrence or death outcomes. Follow-up exams occurred every 3 to 6 months during the first year after treatment, and every 6 to 9 months thereafter.

5.2.2 CT scanning and treatment protocol

This study was performed on a 4D-CT dataset of 381 patients acquired during SBRT treatment planning, each including the dosimetric plan. Patients were examined on a variety of scanners including : SIEMENS SOMATOM Definition Flash, Philips Brilliance Big Bore, Philips GeminiGXL 16 and GE LightSpeed16. For each 3D CT, the original volume size was 512×512 pixels, with a number of slices ranging from 105 to 340. Pixel spacing ranged between 0.98×0.98 mm 2 and 1.5×1.5 mm 2 while slice thickness ranged between 1.0 and 3.0mm. A total of 10 volumes was acquired per patient over the course of one respiratory cycle, with the initial volume starting at the middle of the exhalation (see Figure 5.1).

For each 4D scan, the region of interest (ROI) is the planning target volume (PTV), which is defined by adding an isotropic 5 mm margin to the internal gross tumor volume (IGTV) or internal target volume (ITV).

A patient-specific dose plan for radiation therapy was created by a medical physicist. Tumors moving less than 1cm were typically treated with volumetric modulated arc therapy (VMAT) (291 patients). Other treatment methods included 66 patients treated with CK and 24 patients treated with tomotherapy. The dose plans and number of fractions were also adapted to each patient. Doses ranged between 40 and 60 Gy, with a number of fractions between 3 and 8. The most common occurrences were 60 Gy in 3 fractions (55%) for peripheral tumors and 50 Gy in 5 fractions (19%) for central and ultra-central tumors.

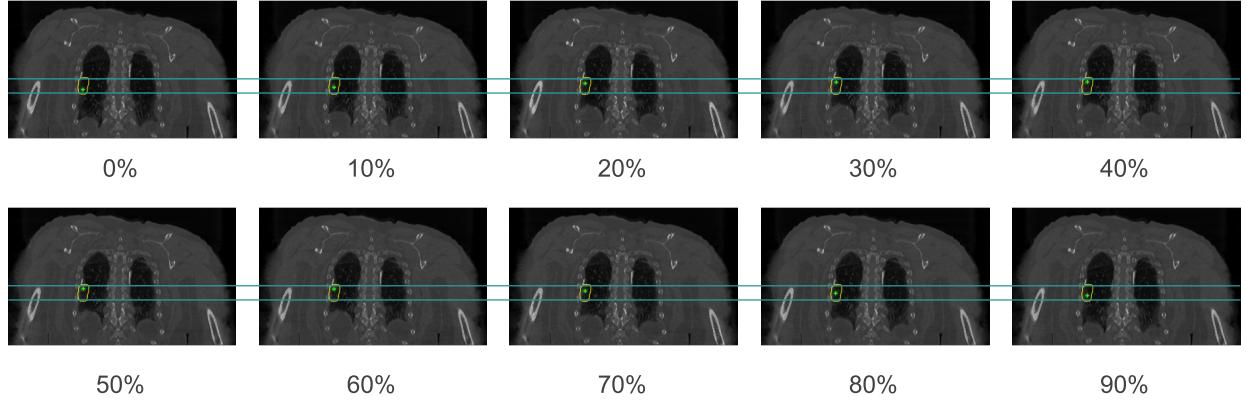


FIGURE 5.1 A sample of a 4D-CT scan on a coronal slice. The 4D-CT consists of 10 respiratory phases, with the first one (phase 0%) taken at the middle of the expiration phase on which we extracted the radiomic features. The PTV tumor segmentation is represented on each phase in yellow. The blue lines show the vertical motion amplitude of the tumor and the green point represents its center.

5.2.3 Radiomic features

In total, 107 radiomic features were extracted from the tumor ROI defined from the PTV. This includes 14 shape features, 18 first order features, 24 gray level co-occurrence matrix (GLCM) features, 16 grey level run length matrix (GLRLM) features, 14 grey level dependence matrix (GLDM) features, 16 grey level size zone matrix (GLSZM) features and 5 neighborhood grey tone difference matrix (NGTDM) features. This extraction was done using the PyRadiomics library [97] with 64 number of bins per ROI and a resampling to a pixel size of $3 \times 3 \times 3 \text{ mm}^3$ in order to standardize the inter-patient voxel spacing. While the treatment planning was done using a 4D-CT scan, a single 3D CT was used to extract the radiomics : the first respiratory phase (phase 0%). The other respiratory phases were used to calculate the tumor motion amplitude, using a lung motion model that infers the deformation vector fields (DVF) between sequential phases based on a deep learning deformation registration pipeline (VoxelMorph) [23]. Once the DVF are computed, the updated tumor position on each phase was calculated, from which the maximum amplitude based on the 3D position was extracted.

To increase the number of training samples as well as the diversity of the images, we performed offline image augmentation. Elastic deformations, scaling, rotations, gamma augmentation, mirror rotations, noise augmentation (by adding Gaussian noise, brightness and blurring), and contrast augmentation were used via the Batchgenerators library [98]. Each image was augmented 20 times and the radiomics were extracted from each of these augmented images.

5.2.4 Random forest prediction model

Table 5.1 presents the relevant clinical features considered in this study. Tumor localisation was defined as per RTOG 0813 protocol [99] for central lesions and per SUNSET protocol [100] for ultra-central lesions. Several tumor types (including adenocarcinomas and squamous cell types) were included in this study, as well as various SBRT techniques applied.

Categorical features were one-hot encoded, while numerical features were normalized in the range of 0–1. We considered that each patient from the augmented images had the same clinical information as their original counterpart. The clinical features were concatenated with the computed radiomic features.

To better deal with the imbalanced dataset (only 26.5% of the dataset encountered an event) we over-sampled the positive cases : each patient who faced a recurrence or death was further copied two times after data augmentation, in order to allow the model to be presented with more events.

To limit oversampling, only the most pertinent features from both clinical and radiomic information were selected to be seen by the prediction model during training. Feature selection was done using ANOVA from *SelectKBest* function from scikit-learn [101]. After feature selection, a RF model was used as a classification model. Figure 5.2 presents the global pipeline used for the outcome prediction. To handle the imbalanced dataset, we used a loss-weighted model. The optimum parameters were found using a 6-fold cross-validation in a grid-search.

To evaluate the predictive value of tumor motion on PFS prediction, for each presented analysis, a second model identical to the first (obtained after ANOVA feature selection) was trained with the inclusion of a single new variable : the maximum magnitude of the tumor displacement on the 4D planning CT.

5.2.5 Training strategy

Each compared model (LR, SVM and RF) was trained using a nested cross-validation (CV) strategy. Figure 5.3 illustrates the proposed evaluation strategy. Of the entire available cohort (381 patients), 304 were used for model development (hyperparameters optimization). A 6-fold CV was used with a grid-search on the features selection process and (for the RF model) the model hyperparameters. Once the inner CV reached maximal mean AUC across 6-fold, the models were retrained with the found hyperparameters and feature selection on the entire development set and evaluated on the held-out 77 patients test set. This entire process was repeated 20 times (shuffle splits) for statistical comparisons and 6 times for sub-outcomes

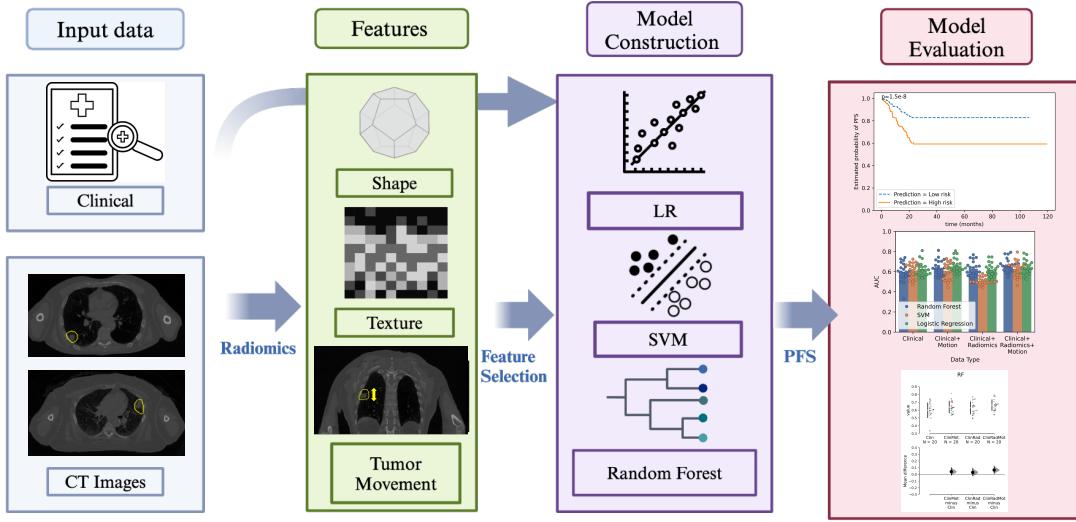


FIGURE 5.2 Global pipeline used for the prediction. We use 20x data augmentation on the CTs to have access to a more diverse database. Radiomics features are extracted from this augmented dataset. Clinical data are one-hot encoded for categorical data and numeric features are normalized. Recurrence or death cases are over-sampled in the training phase to better learn to flag at-risk cases. Clinical and radiomic features are selected by ANOVA and a classifier (SVM, LR or RF) predicts the patient outcome.

comparisons : results reported show the average mean AUC.

All implementations were done in Python version 3.9.10 using sklearn's v1.3.1 API [101]. Statistical comparison graphs (Cumming plot) were generated using the DABEST library [102].

5.3 Results

Here, we present the results which first determines the significant features used by the prediction model, using univariate and multivariate analyses. This is followed by results from the selection process, and finally a performance analysis of the prediction models, in comparison to established classification methods.

5.3.1 Acquired dataset

After applying the selection criteria, the dataset consisted of 381 eligible patients.

Patients characteristics are shown in Table 5.1. The median age is 75 years, and most patients

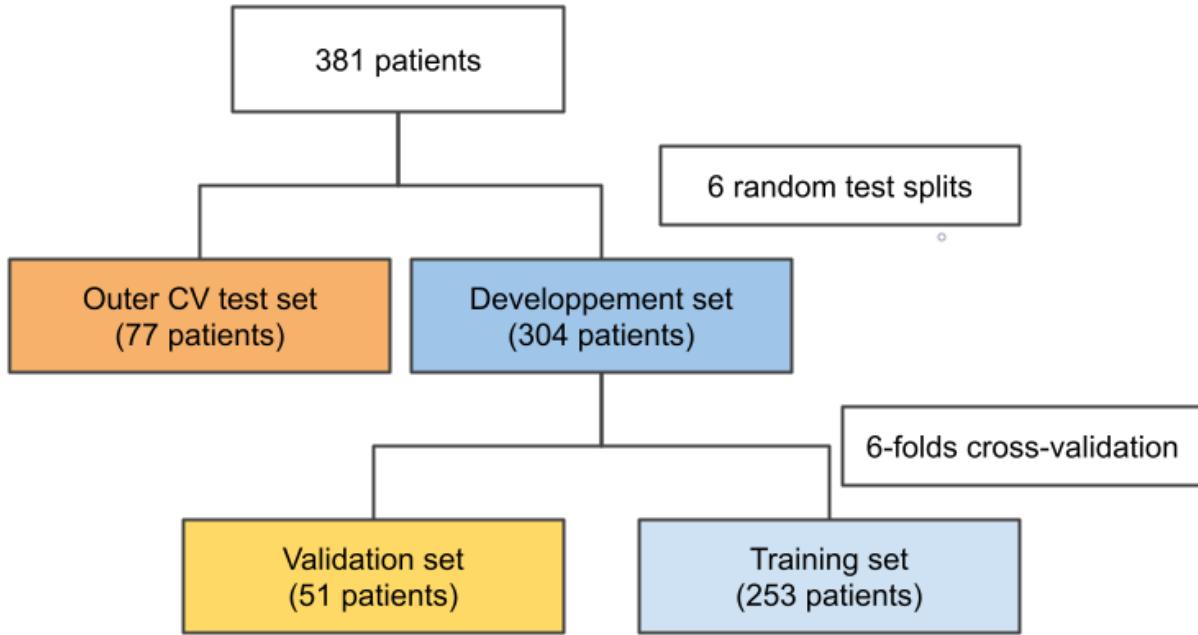


FIGURE 5.3 Schematic illustration of the dataset splitting strategy. We performed a nested cross-validation consisting of 6 outer folds and 6 inner folds.

are women (58%). The median follow-up time is 43 months (range : 0.6-119 months). Two years after the end of treatment, 17 patients (4.5%) have local recurrence, 37 (9.7%) regional and 53 (13.9%) distant metastasis, and 82 (21.5%) died (amongst whom 52 (63%) died directly from lung cancer, since outcomes can be multiple). A total of 280 (73.5%) of patients are progression-free.

5.3.2 Statistical analysis

Univariate and multivariate analysis

We performed a univariate analysis on the overall features using the Kaplan-Meier estimator. The features with statistically significant impact (with $p < 0.05$) for PFS prediction were : 1) the distance between 60 and 95% isodoses being less than 10mm (a higher distance meant lower survival likelihood), 2) the tumor motion being less than 5.9 mm (a higher tumor amplitude correlated with a improved PFS risk), 3) the treatment method used (patients treated with robotic radiotherapy or helical tomotherapy had better prognosis), 4) the gender of the patient (women survived longer), 5) the tumor localisation (larger tumors increase risk), 6) the smoking habits, 7) the Diffusing capacity of the Lungs for Carbon Monoxide

(DLCO), 8) the radiation dose and number of fractions used during treatment (higher doses in less fractions reduces outcome risk), and 9) the total biological dose received by the tumor.

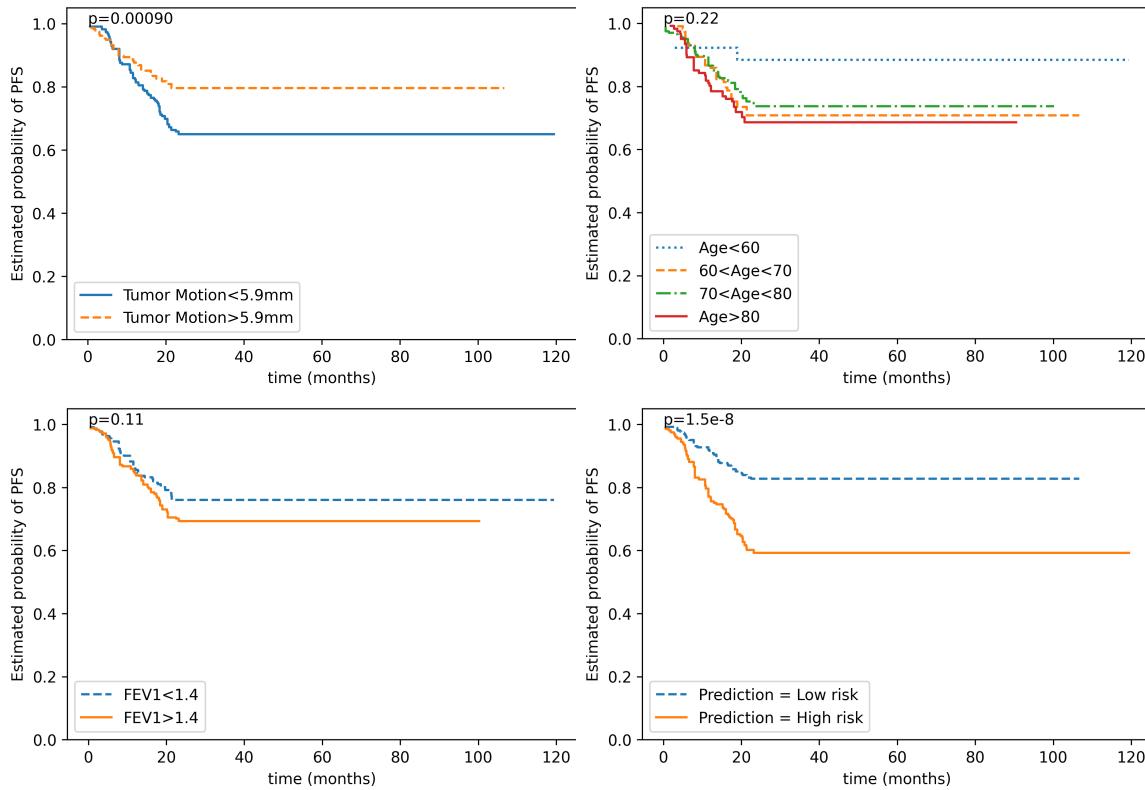


FIGURE 5.4 Kaplan-Meier curves of the three most important clinical features (motion, FEV1 and gender) as well as the model prediction.

Figure 5.4 shows the Kaplan-Meier curves for the three clinical features with the highest Gini importance, indicating how many times a feature was selected to split a node based on its predictive value, as well as the model outcome predictions considering the patients in the outer CV test sets.

A multivariate analysis was performed using a Cox model. Table 5.2 shows the features which correlated significantly with PFS.

Statistical importance of the motion information

We evaluated the statistical importance of each set of features : clinical information, radiomics features and the tumor motion amplitude. To do so, we used a nested cross-validation (CV) strategy. The outer CV was stratified and randomly shuffled with 80% train and 20% test split sizes that were repeated 20 times. The inner CV was also stratified along disease progression

events and consisted of 6 folds. The model performances were evaluated using AUC scores on the inner CV held-out validation split and on the outer CV test split. Each model was trained and tested with the same outer CV iterators and train/test data splits to allow a fair comparison. A two-sided Wilcoxon signed-rank test was used to assess whether a model performed significantly better than another using the 20 paired AUC scores from each model. The p-values were corrected for multiple comparisons using the Benjamini–Hochberg procedure to control the false discovery rate (FDR), and a threshold $FDR \leq 0.05$ was used to determine statistical significance.

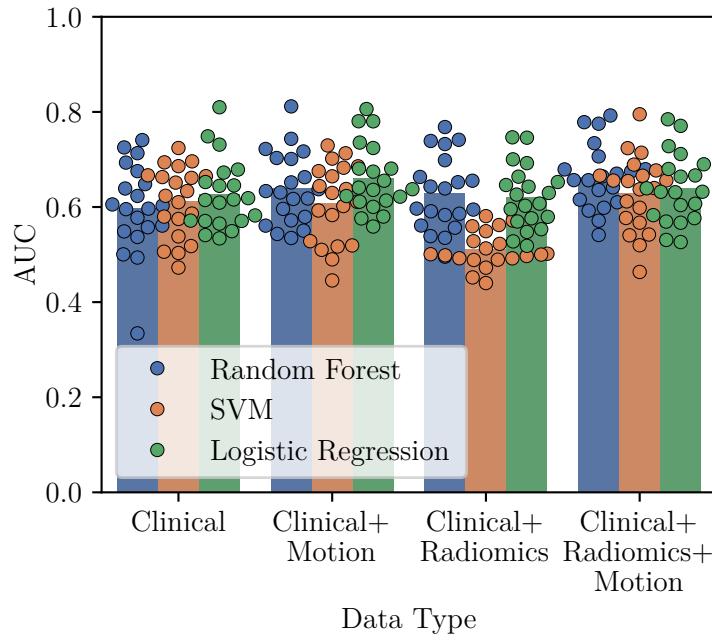


FIGURE 5.5 Mean AUC scores and AUC distribution for each classifier and each data configuration for PFS prediction.

Figures 5.5 and 5.6 show the results of the Wilcoxon statistical tests. For both RF and LR, the incorporation of motion information to the clinical data significantly improved the model performances. For the same two classifiers, radiomic features did not lead to a significant improvement of the AUC, and for SVM the incorporation of radiomic data seems to introduce ambiguity in the model as the clinical features only configuration outperformed it. However, the random forest performs better when adding both the radiomics feature and the tumor motion amplitude to the clinical data.

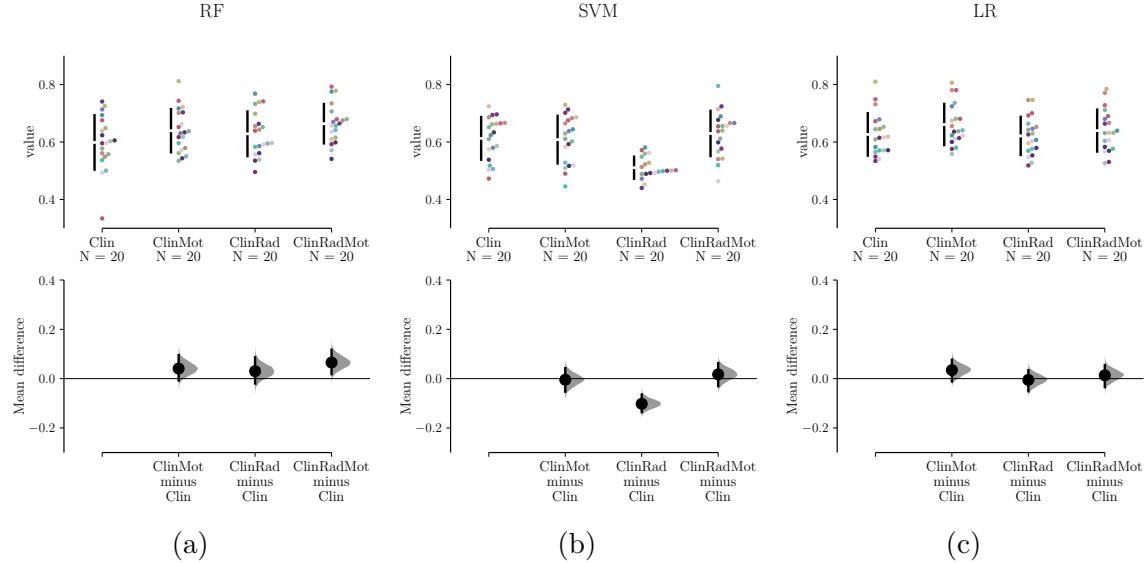


FIGURE 5.6 Comparative performances of RF, SVM and LR models with input features from various combinations of clinical, radiomics and motion features for PFS prediction. The Cumming plots compare the four data configuration AUC on 20 random training/test splits for each classifier. *Clin* : clinical data only, *ClinMot* : clinical and motion information, *ClinRad* : clinical and radiomic data, *ClinRadMot* : clinical, radiomic and motion information. Plots were generated using the DABEST library [102]. On the top part of each graphic, each point represents the AUC value on a fold and the vertical bar shows mean (gap) \pm standard deviation (vertical line). The lower part of each graph shows the effect size via mean differences (dot) with 95% confidence interval (vertical line) : intersection between this confidence interval and the horizontal null line suggests an absence of meaningful difference in effect. Comparisons in each sub-figure is made against the baseline clinical (*Clin*) model.

5.3.3 Clinical features and clinical outcomes

The best models selected between 30 and 37 parameters. Figure 5.7 shows that the shape of the tumor, its motion as well as the general state of the lungs are the main indicators of the risk of negative outcome.

The importance of the motion feature can also be seen by performing a feature permutation on the test set. Each feature was randomly permuted 20 times on each of the 6 outer CV test fold, corresponding to 120 feature permutations and we evaluate the associated decrease in accuracy score. Figure 5.8 shows the features with the most important decrease in performance. We can see that the gender is the feature which, when we swap its values across patients, entails the most important decrease in prediction accuracy, directly followed in importance by the motion.

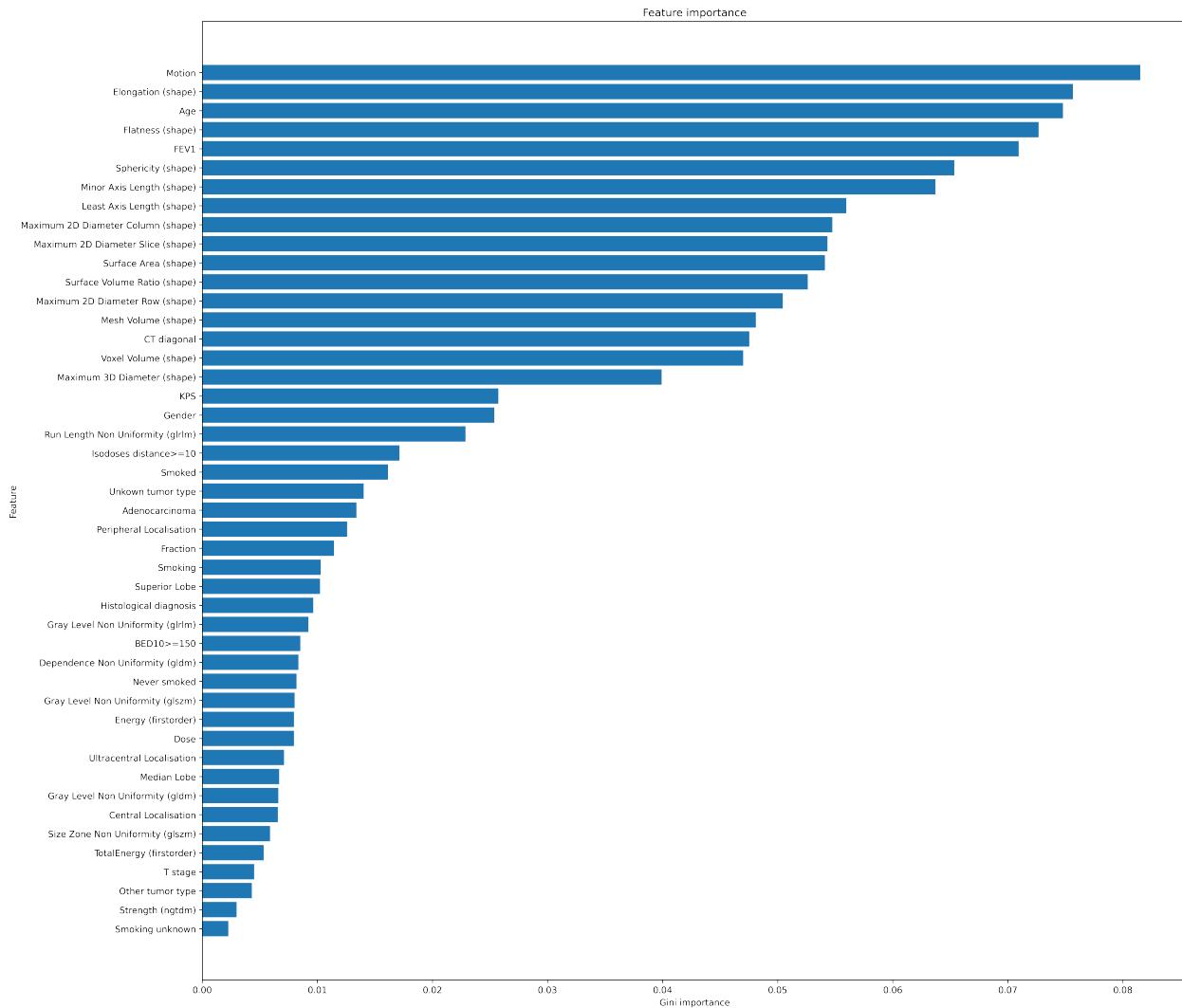


FIGURE 5.7 Importance of each feature for the outcome prediction. The importance was evaluated using the Gini importance criteria, indicating motion is the most important feature, highlighting its importance to be taken into consideration.

5.3.4 Reproducibility to the CK population

Since motion is managed completely differently on the CK, which uses tumor tracking, contrary to helicoidal therapy and VMAT which use wider margins guided by 4D CT, we wanted to assess the reproducibility of our results to the CK population. To do so we separated our dataset between CK and non-CK patients. We retrained our model using only the 315 patients treated with VMAT and helicoidal therapy. We obtained a 64% AUC on a six-folds cross-validation for PFS prediction. When tested on the patients treated with CK, we obtained an AUC of 66% and motion was the feature with the most importance.

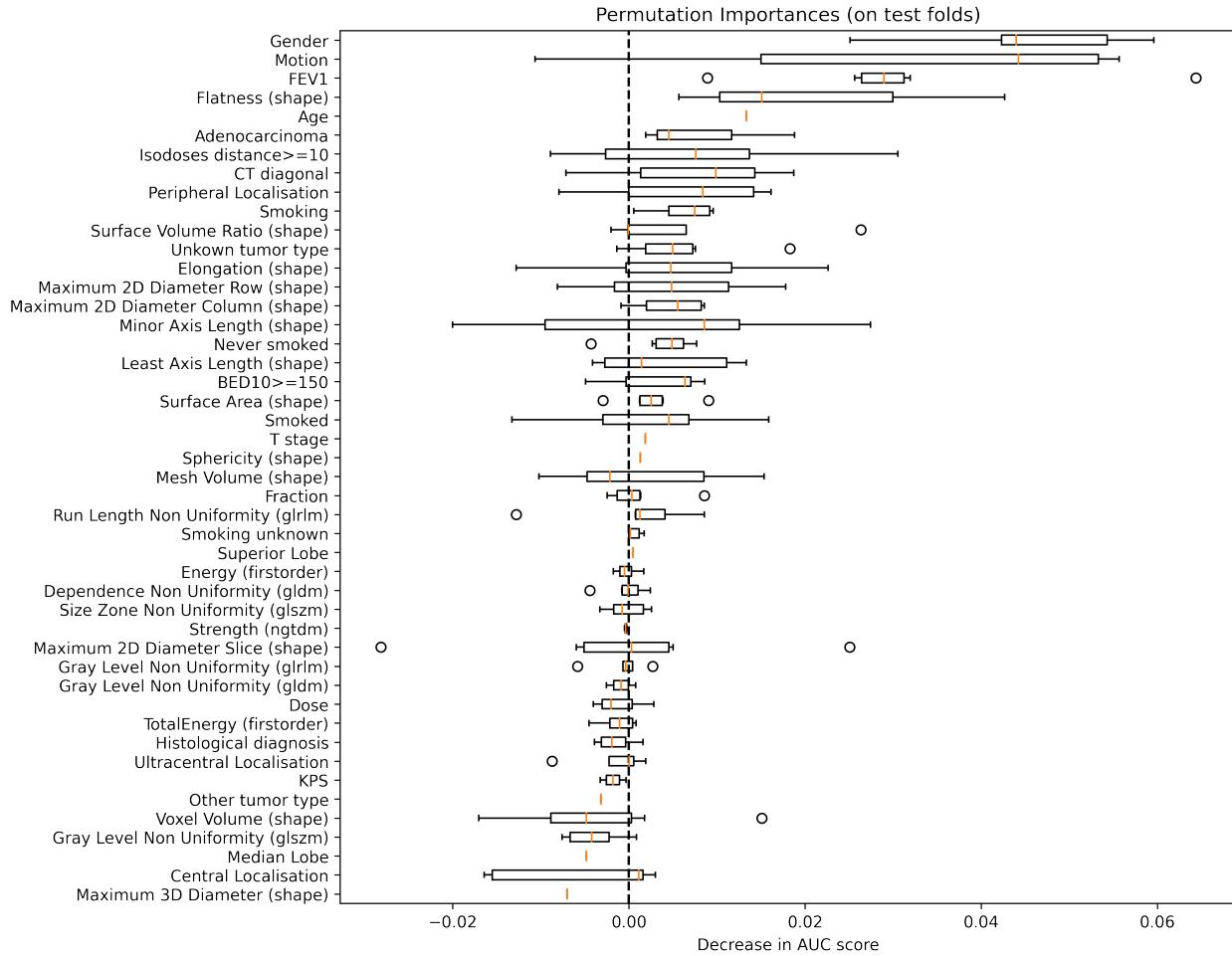


FIGURE 5.8 Feature permutation importance for the selected features in the outcome prediction. Feature permutation tests, evaluating the feature importance using a feature permutation process repeated 120 times. Motion is the second most important feature, highlighting its contribution to the risk classification.

Furthermore, as with the complete dataset, it was the second feature that led to the most accuracy decrease when randomly swapping its values across patients. This tends to confirm that although motion is managed completely differently on the CK, our results are still valid in this particular population.

5.3.5 Predicting overall survival and loco-regional recurrences

Figure 5.9 reports the average AUCs across the six shuffled splits when trying to predict the 2-years OS, LRR and PFS. The LRR classification seems to be the most accurate, reaching 71% mean AUC for both RF and LR models, with between 0–3% improvements when ad-

ding radiomics or tumor motion data. Conversely OS appears harder to predict, with the best performing model reaching 65% AUC. The combined PFS outcome achieved 63% AUC using the RF models with only clinical data, improving with each additional feature types (radiomics and tumor motion), reaching 68% with all 3.

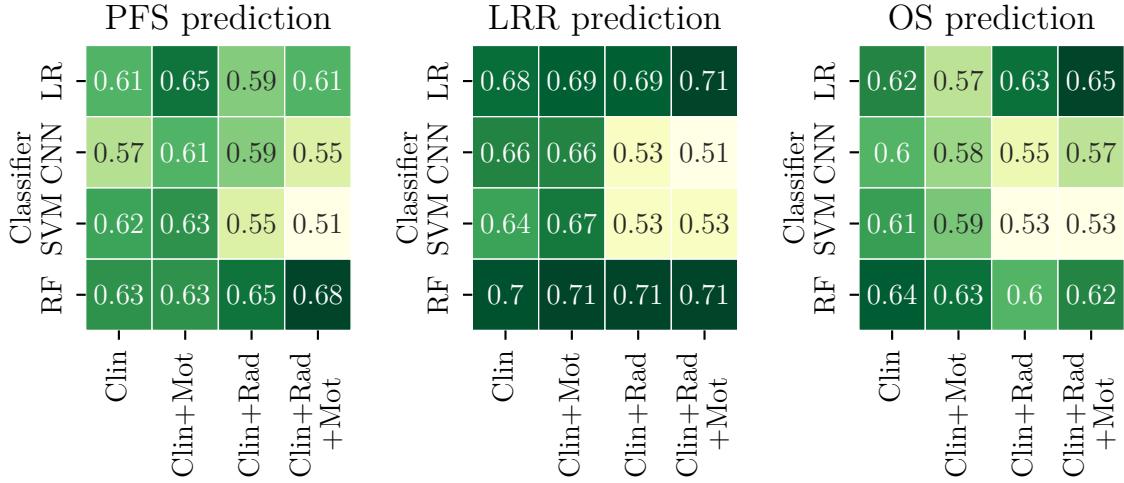


FIGURE 5.9 Average AUC measures on the 6 test splits for the PFS, LRR and OS prediction models for all combinations of input data and classifiers. *Clin* : Clinical information, *Rad* : Radiomics, *Mot* : Tumor motion

5.4 Discussion

The information associated with the amplitude of tumor motion remains under-utilized when anticipating patient response in NSCLC, despite its importance during treatment. Our results show that the integration of the tumor motion can improve the model predictions, given that the AUC increased by a statistically significant 3% when adding the displacement metric to the radiomics features and clinical characteristics. The importance of motion information was statistically significant, using the RF models based on the Wilcoxon test ($p \leq 0.05$).

Tonneau et al. [3] established that the tumor localisation, the Biologically Effective Dose (BED₁₀, using an α/β -ratio of 10 Gy) to the lesion, age, tumor T stage and histological cancer confirmation were the most pertinent features when predicting local or regional recurrences, DM or OS. In particular, central locations were associated with a higher risk of local and regional recurrence while ultra-central locations had a shorter OS and were more at risk of regional and distant recurrences. Figure 5.7 is coherent with these previous results. We can observe that for our model, peripheral localisation (the other possible localisation) was the

25th most important feature, and central localisation was also a selected feature, which is in accordance with their results : tumor localisation seems to have an impact on patient outcomes. We also found out that the age of the patient was an important decision factor, as it is our 3rd most important feature. Our model also selected the BED₁₀ (our 31nd most important feature), another one of the pertinent features according to Tonneau et al.

Regarding radiomics features, several features related to the tumor size are key outcome predictors (such as the volume, surface, axis lengths for instance). This data supports the observations of Li et al. [103].

Of the 3 investigated outcomes, OS had the lowest performances. This can probably be explained by the fact that although PFS outcome regroups more event cases, it is more diverse by combining both recurrences and deaths, yielding an AUC of 68%, a 7% improvement with respect to the baseline LR model with only clinical data. Conversely LRR outcomes are more similar and should present closer characteristics. On the other hand, death cases are both rarer and more diverse, all the more so that they can sometimes be unrelated to the disease. Indeed since we studied a population of patients deemed inoperable, they have a lot of comorbidities, and can often die from causes unrelated to their cancer, making OS harder to predict.

This study has some limitations. First, tumor delimitation and distance between the isodoses were obtained manually, which can introduce bias and a lack of repeatability as radiomics are known to be quite sensitive to segmentation methods and scanner settings. Using a semi-automatic tumor segmentation method could decrease inter-reader variability. Regarding scanner settings, we used CT scans from different machines in order to increase the robustness of the model. Second, all our patients were treated in the same center, which makes it hard to evaluate the generalisation potential of our model to patients from other institutions. Finally we performed a retrospective study, which will require further evaluation on a prospective study. Future work should aim at a larger patient cohorts with longer follow-ups beyond 2-year risk prediction.

TABLEAU 5.1 Clinical and treatment characteristics of the eligible patients cohort after exclusion criteria. Numbers are presented as *frequency (percentage)* when indicated, *median (min–max)* otherwise. SBRT : stereotactic body radiation therapy. VMAT : volumetric modulated arc therapy. CK : CyberKnife. TOMO : helical tomotherapy. FEV1 : forced expiratory volume in 1 second. DLCO : diffusing capacity of the lungs for carbon monoxide. KPS : Karnofsky performance status. BED₁₀ : biologically effective doses.

Characteristics	N=381
Age at diagnosis	47 (45–94)
Gender	
Female	222 (58.3%)
Male	159 (41.7%)
First case disease	
Yes	360 (94.5%)
No	21 (5.5%)
Smoking habit	
Current	117 (30.7%)
Former	240 (63%)
Never	17 (4.5%)
Pulmonary Lobe	
Inferior	118 (31.0%)
Superior	247 (64.8%)
Median	16 (4.2%)
Tumor localisation	
Peripheral	284 (74.5%)
Central	70 (18.4%)
Ultra-central	27 (7.1%)
Biopsy result	
Adenocarcinoma	155 (40.7%)
Squamous cell	82 (21.5%)
Other	20 (5.2%)
Histological diagnosis	
Yes	279 (73.2%)
No	102 (26.8%)
SBRT technique	
VMAT	291 (76.4%)
CK	66 (17.3%)
TOMO	24 (6.3%)
Charlson comorbidity index	
2-4	79 (20.7%)
5-7	256 (67.2%)
8-13	44 (11.5%)
T Stage (TNM 8ed [95, 96])	
T1	324 (85.0%)
T2	56 (14.7%)
T3	1 (0.3%)
FEV1	1.36 (0.4–3.22)
Size of the diagonal CT	20 (6–50)
DLCO proportion	56 (14–147)
KPS	80 (40–100)
Dose applied to the tumor	60 (40–60)
Fraction number	3 (3–8)
BED₁₀	180 (72–180)
60–95% isodoses distance (mm)	12.4 (4.3–45.0)
Tumor motion (mm)	5.92 (0–34.4)

TABLEAU 5.2 The significant features found during multivariate analysis. For all of them, $p < 0.005$. HR : hazard ratio. CI : confidence interval. FEV1 : forced expiratory volume in 1 second. DLCO : diffusing capacity of the lungs for carbon monoxide.

Feature	HR	95% CI
Female gender	1.80	1.62–2.00
Never smoked	0.21	0.13–0.33
Pathology		
Inferior lobe	1.66	1.22–2.26
Superior lobe	1.59	1.17–2.15
Peripheral tumor	0.77	0.64–0.92
Unknown tumor type	0.54	0.39–0.74
FEV1	2.33	1.78–3.06
DLCO	0.45	0.31–0.66
Motion	0.15	0.09–0.24
Distance isodoses	0.32	0.19–0.56
Shape		
Flatness	12.59	5.23–30.32
Sphericity	0.03	0.01–0.18
First order		
Total energy	1.0	1.0–1.0
Gray level (second order) [97]		
Cluster shade (GLCM)	1.0	1.0–1.0
SDHGLE (GLDM)	1.0	1.0–1.0
GLN (GLSZM)	1.0	1.0–1.0
HGLZE (GLSZM)	1.0	1.0–1.0

5.5 Conclusion

In conclusion, this study demonstrated that the integration of amplitude of tumor motion extracted from a motion model on 4D-CT, combined with radiomics features extracted from the PTV and patient clinical characteristics can accurately predict NSCLC patient outcomes. We find tumor size, shape, motion amplitude and the patient's gender, age and FEV1 to have the most important prognostic value to determine patient's PFS risks. With a more accurate evaluation of patient's risks of death or tumor recurrence, more frequent follow-up examinations combined with earlier interventions could help reduce patient mortality.

In future works, we plan to validate the model on an external validation cohort to further evaluate the prediction performance.

Acknowledgments

Financial support : Maurice-Tubiana Grant - Société Française de Radiothérapie Oncologique (SFRO), Nuovo-Soldati Foundation for Cancer Research, TransmedTech, and Astra Zeneca.

Conflict of Interest Statement

H Bahig and S Kadoury have received research funds from Varian Medical Systems and H Bahig is a consultant and ad board for Astra Zeneca, and EMD Serono.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted. The other authors have no relevant conflicts of interest to disclose.

CHAPITRE 6 DISCUSSION GÉNÉRALE

Les deux chapitres précédents ont présenté les solutions développées au cours de ce projet de maîtrise en vue de répondre à l'objectif développé au chapitre 3. Dans un premier temps, un modèle génératif a été développé dans le but de générer des phases de CT scans à partir d'un nombre limité de volumes des instants précédents. La seconde approche visait à développer un modèle de classification dans le but de déterminer les patients à risque élevé de progression de leur cancer. Dans ce chapitre nous discuterons de la méthodologie développée dans ce projet, ainsi que ses limitations et futures perspectives de développement.

6.1 Autour de la prédiction des déformations du poumon

Le modèle présenté dans le chapitre 4 a permis d'évaluer les capacités des auto-encodeurs variationnels à inférer un champ de déformations 3D à partir d'un volume de référence CT 3D et d'images 2D issues des phases respiratoires précédant le volume d'intérêt que l'on cherche à prédire.

Les limitations concernant ce modèle concernent d'abord les ressources de calcul nécessitées pour son entraînement. Étant donné la taille des CT scans ainsi que le nombre de paramètres du modèle, nous avons dû redimensionner les scans fournis au modèle. Ainsi le modèle ne traite pas directement avec les images d'origines. Ceci fait à la fois perdre du temps d'un point de vue de l'utilisation étant donné que cela ajoute une étape de prétraitement des données, mais fait également probablement perdre le modèle en précision, étant donné que certaines structures d'intérêt pour la prédiction du mouvement ont pu être légèrement affectées par le redimensionnement des images. Cependant une fois l'entraînement effectué, la prédiction sur les volumes redimensionnés peut s'effectuer en des temps compatibles avec le suivi en temps réel.

La taille de l'ensemble de données utilisé pour l'entraînement et l'évaluation pourrait également gagner à être augmentée. Ceci constitue un problème récurrent dans les applications médicales. Or les performances des réseaux de neurones sont liées à la quantité et à la diversité des données d'entraînement, qui leur permettent d'être confrontés à une représentation la plus réaliste possible des données qu'ils seront amenés à rencontrer dans des applications pratiques. Il est également important de disposer d'une diversité suffisante dans les données d'évaluation, car celle-ci permet d'estimer plus précisément les performances du modèle sur

des cas nouveaux. De plus nous ne disposons comme données d'entraînement que d'un seul cycle respiratoire par patient, or le mouvement respiratoire n'est pas toujours uniforme d'un cycle à l'autre.

D'autres types de modèles seraient également intéressants à appliquer et pourraient permettre de meilleures performances de prédiction. Par exemple les modèles ViT sont de plus en plus utilisés pour ce genre de tâches, en remplacement des LSTM utilisés par notre modèle. Les modèles de diffusion sont également de plus en plus populaires et pourraient présenter une alternative intéressante.

Enfin le modèle en tant que tel ne propose pas de mécanisme de suivi de la tumeur. Pour déterminer la position de la tumeur dans les phases prédictives, il faut d'abord indiquer la tumeur sur le volume de référence, puis les déformations prédictives servent à indiquer la position de la tumeur dans les phases d'intérêt. Cependant le modèle ne réalise pas par lui-même la détection de la tumeur.

6.2 Autour de la prédiction des issues du traitement et de l'importance du mouvement

Le modèle présenté dans le chapitre 5 a permis d'évaluer la pertinence d'utiliser les données de l'amplitude du mouvement de la tumeur pour la prédiction des risques de survie et récidives des patients, en plus des données cliniques et des radiomiques issues des volumes de références utilisés précédemment.

Les limitations de ce modèle concernent d'abord la méthode d'estimation des amplitudes de mouvement. Comme ces données n'étaient pas disponibles cliniquement pour tous les patients, nous avons choisi de les déterminer à l'aide des CT scans, et des DVF dans les cas de scans manquants, afin d'avoir une mesure cohérente entre tous les patients.

De plus nous avons choisi de nous concentrer sur l'amplitude du mouvement de la tumeur, qui représente la composante principale du mouvement de la tumeur. Cependant il pourrait être intéressant d'utiliser l'ensemble de l'information sur le déplacement de la tumeur au cours du cycle respiratoire. Il serait également pertinent d'utiliser comme information sur le mouvement des données plus complètes qu'une simple mesure chiffrée du déplacement de la tumeur. L'inclusion de l'entièreté des DVF sur le cycle respiratoire (tels que ceux reçus en sortie du modèle de prédiction de déformation des poumons) pourrait donner davantage d'informations au modèle en permettant une analyse plus fine du déplacement de la tumeur, mais aussi en liant ce déplacement avec le déplacement total des poumons.

Il est aussi connu que les données radiomiques sont dépendantes du prétraitement réalisé sur

les images. Les choix de prétraitement pris dans notre étude peuvent donc avoir un impact sur les résultats présentés dans ce mémoire. De plus les segmentations utilisées ont été acquises manuellement par des radiologues, ce qui peut entraîner un biais de subjectivité et donc un manque de reproductibilité. L'utilisation d'un algorithme de segmentation automatique pourrait permettre de limiter ces biais.

Enfin, là encore, la taille du jeu de données pourrait gagner à être augmentée. Cheung et coll. [27] ont relevé dans leur revue de littérature que parmi les 25 études utilisant des données radiomiques qu'ils ont étudiées, les tailles de jeux de données étaient très faibles (entre 13 et 573 patients, avec une taille médiane de 87), ce qui place notre taille de jeu de données dans la moyenne haute, mais reste très faible par rapport aux tailles de données observées dans d'autres domaines d'application de l'intelligence artificielle. De la même façon, notre jeu de validation provient du même hôpital que nos données d'apprentissage, ce qui ne nous permet pas de conclure sur les performances de généralisation du modèle à d'autres environnements disposant de protocoles potentiellement différents. Cependant notre jeu de données dispose de CT scans provenant de différentes machines, ce qui peut être de bon augure quant aux capacités de généralisation du modèle à d'autres machines provenant d'autres hôpitaux, mais ceci devrait être démontré expérimentalement.

6.3 Perspectives futures

La principale limitation dans les deux axes de ce mémoire reste la taille du jeu de données. Dans de futurs travaux, il serait donc intéressant d'intégrer de nouveaux patients à cette étude, idéalement issus d'autres centres hospitaliers.

La gestion du suivi du mouvement de la tumeur pourrait aussi être améliorée, par exemple en utilisant l'ensemble de l'information de déformation provenant des DVF plutôt que de se limiter à l'amplitude du déplacement. Nous disposerions ainsi du déplacement de la tumeur en 3D et non seulement sa composante principale en 2D comme c'est le cas actuellement.

Il serait également intéressant de développer une approche delta-radiomique, en utilisant les scans réalisés lors du suivi des patients. Cette approche pourrait également être mise en place pour étudier les différences de radiomiques entre chacune des phases des CT scans 4D afin d'intégrer la composante temporelle des données. Une approche de type RNN ou Transformer pourrait également être intéressante, mais nécessitera un jeu de données bien plus conséquent, étant donné que les approches CNN pâtissaient déjà du manque de données à notre disposition.

Cette étude pourrait également être étendue à d'autres organes eux aussi affectés par les

mouvements respiratoires, tels que le foie par exemple.

6.4 Intégration pour l'utilisation en clinique

Les performances de prédiction des décès et récurrences restent faibles pour envisager d'être applicables directement en situation réelle. Davantage de vérifications doivent être effectuées pour confirmer ses performances, notamment en menant une étude prospective sur des patients issus de plusieurs centres hospitaliers. Néanmoins l'importance de l'apport de l'information sur l'amplitude du mouvement de la tumeur dans la qualité de la prédiction pourrait dès à présent être prise en compte et les futures études devraient intégrer cette composante dans leurs données. De plus certaines tumeurs ne présentant qu'un faible déplacement ne bénéficient pas de solutions de compensations du mouvement. Notre étude semble pourtant montrer que ces patients présentent une survie plus faible, ils gagneraient donc à être davantage suivis ou à disposer de solutions de compensations malgré le faible mouvement de leur tumeur.

CHAPITRE 7 CONCLUSION

Ce mémoire a débuté par l'étude du contexte de la problématique de ce projet de maîtrise. L'étude de littérature détaillée a permis de mettre en avant les méthodes courantes et leurs limites lors de la prédiction de l'issue du traitement contre le cancer du poumon. Cela nous a permis de déterminer l'objectif de ce projet, qui était de valider l'importance du mouvement de la tumeur pour l'aide à la prédiction de la survie sans progression de patients atteints de NSCLC.

7.1 Synthèse des travaux

Dans un premier temps, une étude du mouvement respiratoire du poumon a été réalisée. Un modèle a été développé afin de suivre les déformations du poumon au cours de la respiration et ainsi permettre un suivi de la tumeur en temps réel. Dans un second temps, l'impact de ce mouvement sur le résultat du traitement a été évalué. Nous avons entraîné différents modèles à classifier le patient selon son risque de mauvaise réponse au traitement. Le mouvement a été déterminé comme étant une caractéristique utile à cette classification.

7.2 Limitations de la solution proposée

La solution proposée présente cependant plusieurs limites. Tout d'abord concernant le nombre de patients conservés pour ces études. La base de données originellement à notre disposition était conséquente, mais du fait de la nécessité d'avoir accès à une réponse du patient au traitement (ce qui nécessite un temps de délai pour évaluer sa réponse) ainsi que d'accès aux scans et données cliniques nous n'avons pu conserver que 381 patients pour la prédiction de réponse. Cette taille de jeu de données est légèrement supérieure à la moyenne des tailles de jeux de données utilisés dans des études comparables, mais reste faible au regard de la quantité de données utilisées dans de nombreuses autres applications d'apprentissage automatique.

Les PTV ont été acquis manuellement par des radiologues et les amplitudes des mouvements des tumeurs ont été calculées manuellement également. Ceci pose la question de la reproductibilité des résultats à d'autres jeux de données. Il pourrait être intéressant d'utiliser une méthode permettant de segmenter automatiquement les régions d'intérêt et de calculer automatiquement l'amplitude maximale de la tumeur.

La capacité de généralisation du modèle pourrait également être davantage évaluée. Nous avons réalisé une validation croisée à 6 blocs afin d'obtenir une estimation des capacités de

généralisation du modèle. Cependant un jeu de test externe pourrait présenter davantage de difficultés.

RÉFÉRENCES

- [1] J. Ferlay, M. Colombet, I. Soerjomataram, D. M. Parkin, M. Piñeros, A. Znaor et F. Bray, “Cancer statistics for the year 2020 : An overview,” *International Journal of Cancer*, vol. 149, n°. 4, p. 778–789, 2021.
- [2] S. K. Vinod et E. Hau, “Radiotherapy treatment for lung cancer : Current status and future directions,” *Respirology*, vol. 25, n°. S2, p. 61–71, nov. 2020.
- [3] M. Tonneau, C. Richard, B. Routy, M.-P. Campeau, T. Vu, E. Filion, D. Roberge, D. Mathieu, R. Doucet, D. Beliveau-Nadeau et H. Bahig, “A competing risk analysis of the patterns and risk factors of recurrence in early-stage non-small cell lung cancer treated with stereotactic ablative radiotherapy,” *Radiotherapy and Oncology*, vol. 185, p. 109697, août 2023.
- [4] P. Ramesh, A. Jain, R. Karuppasamy et S. Veerappapillai, “Exploring Machine Learning Models for Recurrence Prediction in Lung Cancer Patients,” *Indian Journal of Pharmaceutical Education and Research*, vol. 56, n°. 3s, p. s398–s406, sept. 2022.
- [5] C. J. Hobbs, S. J. Ko, N. N. Paryani, J. M. Accurso, K. R. Olivier, Y. I. Garces, S. S. Park, C. L. Hallemeier, S. E. Schild, S. A. Vora, J. B. Ashman, W. G. Rule, J. R. Bowers, M. G. Heckman, N. N. Diehl et R. C. Miller, “Stereotactic Body Radiotherapy for Medically Inoperable Stage I-II Non-Small Cell Lung Cancer : The Mayo Clinic Experience,” *Mayo Clinic Proceedings : Innovations, Quality & Outcomes*, vol. 2, n°. 1, p. 40–48, mars 2018.
- [6] W. Xing, W. Gao, X. Lv, Z. Zhao, X. Xu, Z. Wu, G. Mao et J. Chen, “Artificial intelligence predicts lung cancer radiotherapy response : A meta-analysis,” *Artificial Intelligence in Medicine*, vol. 142, p. 102585, août 2023.
- [7] M. K. Thirunavukkarasu et R. Karuppasamy, “Forecasting determinants of recurrence in lung cancer patients exploiting various machine learning models,” *Journal of Biopharmaceutical Statistics*, vol. 33, n°. 3, p. 257–271, mai 2023.
- [8] Institut national du cancer. (2023) Qu'est-ce qu'un cancer ? [En ligne]. Disponible : <https://www.e-cancer.fr/Comprendre-prevenir-depister/Qu-est-ce-qu-un-cancer>
- [9] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal et F. Bray, “Global Cancer Statistics 2020 : GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA : A Cancer Journal for Clinicians*, vol. 71, n°. 3, p. 209–249, mai 2021.

- [10] N. Duma, R. Santana-Davila et J. R. Molina, “Non-Small Cell Lung Cancer : Epidemiology, Screening, Diagnosis, and Treatment,” *Mayo Clinic Proceedings*, vol. 94, n°. 8, p. 1623–1640, août 2019.
- [11] J. Khalifa, D. Lerouge, C. Le Péchoux, N. Pourel, J. Darréon, F. Mornex et P. Giraud, “Radiotherapy for primary lung cancer,” *Cancer/Radiothérapie*, vol. 26, n°. 1-2, p. 231–243, févr. 2022.
- [12] M.-S. Tsao, A. G. Nicholson, J. J. Maleszewski, A. Marx et W. D. Travis, “Introduction to 2021 WHO Classification of Thoracic Tumors,” *Journal of Thoracic Oncology*, vol. 17, n°. 1, p. e1–e4, janv. 2022.
- [13] Société canadienne du cancer. (na) Radiothérapie. [En ligne]. Disponible : <https://cancer.ca/fr/treatments/treatment-types/radiation-therapy>
- [14] Institut national du cancer. (2023) Qu'est-ce que la radiothérapie ? [En ligne]. Disponible : <https://www.e-cancer.fr/Patients-et-proches/Se-faire-soigner/Traitements/Radiotherapie>
- [15] E. Dormand, P. E. Banwell et T. E. Goodacre, “Radiotherapy and wound healing,” *International Wound Journal*, vol. 2, n°. 2, p. 112–127, juin 2005.
- [16] E. Alonso Ortiz, “GBM8378 - LINAC pour la radiothérapie,” 13 avril 2022, communication personnelle.
- [17] Q.-S. Chen, M. S. Weinhous, F. C. Deibel, J. P. Ciezki et R. M. Macklis, “Fluoroscopic study of tumor motion due to breathing : Facilitating precise radiation therapy for lung cancer patients,” *Medical Physics*, vol. 28, n°. 9, p. 1850–1856, sept. 2001.
- [18] Société canadienne du cancer. (na) Radiothérapie externe. [En ligne]. Disponible : <https://cancer.ca/fr/treatments/treatment-types/radiation-therapy/external-radiation-therapy>
- [19] Standford Medicine. (na) Stereotactic ablative radiotherapy (SABR/SBRT). [En ligne]. Disponible : <https://stanfordhealthcare.org/medical-treatments/s/stereotactic-body-radiation-therapy.html>
- [20] American Cancer Society. (2023) Getting external beam radiation therapy. [En ligne]. Disponible : <https://www.cancer.org/cancer/managing-cancer/treatment-types/radiation/external-beam-radiation-therapy.html#how-long-does-external-radiation-treatment-take>
- [21] Institut national du cancer. (na) Séance d'irradiation. [En ligne]. Disponible : <https://www.e-cancer.fr/Patients-et-proches/Se-faire-soigner/Traitements/Radiotherapie/Deroulement/Seance-d-irradiation>

- [22] A. Botticella, A. Levy, G. Auzac, I. Chabert, C. Berthold et C. Le Pechoux, “Tumour motion management in lung cancer : a narrative review,” *Translational Lung Cancer Research*, vol. 10, n°. 4, p. 2011–2017, avr. 2021.
- [23] L. V. Romaguera, T. Mezheritsky, R. Mansour, J.-F. Carrier et S. Kadoury, “Probabilistic 4D predictive model from in-room surrogates using conditional generative networks for image-guided radiotherapy,” *Medical Image Analysis*, vol. 74, p. 102250, déc. 2021.
- [24] Institut national du cancer. (2021) Le scanner ou tomodensitométrie (TDM). [En ligne]. Disponible : <https://www.e-cancer.fr/Patients-et-proches/Se-faire-soigner/Traitements/Examens/Scanner-et-cancer>
- [25] National Institute of Health. (2022) Computed tomography (CT). [En ligne]. Disponible : <https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct>
- [26] J. Cohen-Adad, “GBM8378 - imagerie radiographique et tomographie calculée,” 03 février 2022, communication personnelle.
- [27] B. M. F. Cheung, “Radiomics in stereotactic body radiotherapy for non-small cell lung cancer : a systematic review and radiomic quality score study,” *Radiation Oncology Journal*, vol. 42, n°. 1, p. 4–16, mars 2024.
- [28] J. E. van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi et B. Baessler, “Radiomics in medical imaging—“how-to” guide and critical reflection,” *Insights into Imaging*, vol. 11, n°. 1, p. 1–16, déc. 2020.
- [29] S. Vande Perre, L. Duron, A. Milon, S. Nougaret, L. Fournier et I. Thomassin-Naggara, “Radiomique : mode d’emploi. Méthodologie et exemples d’application en imagerie de la femme,” *Imagerie de la Femme*, vol. 29, n°. 1, p. 25–33, mars 2019.
- [30] J. D. Shur, S. J. Doran, S. Kumar, D. ap Dafydd, K. Downey, J. P. B. O’Connor, N. Papanikolaou, C. Messiou, D.-M. Koh et M. R. Orton, “Radiomics in Oncology : A Practical Guide,” *RadioGraphics*, vol. 41, n°. 6, p. 1717–1732, oct. 2021.
- [31] R. M. Haralick, K. Shanmugam et I. Dinstein, “Textural Features for Image Classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, n°. 6, p. 610–621, nov. 1973.
- [32] M. M. Galloway, “Texture analysis using gray level run lengths,” *Computer Graphics and Image Processing*, vol. 4, n°. 2, p. 172–179, juin 1975.
- [33] G. Thibault, J. Angulo et F. Meyer, “Advanced Statistical Matrices for Texture Characterization : Application to Cell Classification,” *IEEE Transactions on Biomedical Engineering*, vol. 61, n°. 3, p. 630–637, mars 2014.

- [34] M. Amadasun et R. King, "Textural features corresponding to textural properties," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, n°. 5, p. 1264–1274, oct. 1989.
- [35] Q. Cappart, "INF8215 - réseaux de neurones et apprentissage profond," 18 novembre 2021, communication personnelle.
- [36] A. Zouaq, "INF8460 - introduction aux réseaux de neurones," 19 septembre 2021, communication personnelle.
- [37] S. Kevin Zhou, Hayit Greenspan et Dinggang Shen, édit., *Deep Learning for Medical Image Analysis*. Elsevier, 2017.
- [38] C. Pal, "INF8225," 3 février 2022, communication personnelle.
- [39] F. Chollet, *Deep learning with Python*. Manning, 2018.
- [40] S. Albelwi et A. Mahmood, "A Framework for Designing the Architectures of Deep Convolutional Neural Networks," *Entropy*, vol. 19, n°. 6, p. 242, mai 2017.
- [41] M. Jaderberg, K. Simonyan, A. Zisserman et k. kavukcuoglu, "Spatial transformer networks," dans *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama et R. Garnett, édit., vol. 28. Curran Associates, Inc., 2015.
- [42] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser et I. Polosukhin, "Attention is All you Need," dans *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan et R. Garnett, édit., vol. 30. Curran Associates, Inc., 2017.
- [43] A. Zouaq, "INF8460 - bidirectional encoder representation from transformers (BERT)," 01 novembre 2021, communication personnelle.
- [44] C. Pal, "INF8225," 24 février 2022, communication personnelle.
- [45] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit et N. Houlsby, "An image is worth 16x16 words : Transformers for image recognition at scale," dans *International Conference on Learning Representations*, juin 2021.
- [46] S. J. Russell et P. Norvig, *Artificial intelligence : a modern approach*, 3^e éd., ser. Prentice Hall series in artificial intelligence. Pearson, 2016.
- [47] N. Sutton-Charani, "Decision trees," 22 janvier 2021, communication personnelle.
- [48] S. Harispe, "Introduction to machine learning - supervised learning : Generalities & linear models," 13 février 2021, communication personnelle.

- [49] Z. Wang, A. Bovik, H. Sheikh et E. Simoncelli, “Image Quality Assessment : From Error Visibility to Structural Similarity,” *IEEE Transactions on Image Processing*, vol. 13, n°. 4, p. 600–612, avr. 2004.
- [50] K. Jiang, F. Fujii et T. Shiinoki, “Prediction of lung tumor motion using nonlinear autoregressive model with exogenous input,” *Physics in Medicine & Biology*, vol. 64, n°. 21, p. 21NT02, oct. 2019.
- [51] A. Mylonas, J. Booth et D. T. Nguyen, “A review of artificial intelligence applications for motion tracking in radiotherapy,” *Journal of Medical Imaging and Radiation Oncology*, vol. 65, n°. 5, p. 596–611, août 2021.
- [52] W. Sun, Q. Wei, L. Ren, J. Dang et F.-F. Yin, “Adaptive respiratory signal prediction using dual multi-layer perceptron neural networks,” *Physics in Medicine & Biology*, vol. 65, n°. 18, p. 185005, sept. 2020.
- [53] H. Lin, C. Shi, B. Wang, M. F. Chan, X. Tang et W. Ji, “Towards real-time respiratory motion prediction based on long short-term memory neural networks,” *Physics in Medicine & Biology*, vol. 64, n°. 8, p. 085010, avr. 2019.
- [54] P. Samadi Miandoab, S. Saramad et S. Setayeshi, “Respiratory motion prediction based on deep artificial neural networks in CyberKnife system : A comparative study,” *Journal of Applied Clinical Medical Physics*, vol. 24, n°. 3, p. e13854, mars 2023.
- [55] M. Tan, H. Peng, X. Liang, Y. Xie, Z. Xia et J. Xiong, “LSTformer : Long Short-Term Transformer for Real Time Respiratory Prediction,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, n°. 10, p. 5247–5257, oct. 2022.
- [56] M. Mafi et S. M. Moghadam, “Real-time prediction of tumor motion using a dynamic neural network,” *Medical & Biological Engineering & Computing*, vol. 58, n°. 3, p. 529–539, mars 2020.
- [57] X. Bao, D. Xiao, B. He, W. Gao, J. Wang et F. Jia, “Prediction of Liver Respiratory Motion Based on Machine Learning,” dans *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. Dali, China : IEEE, déc. 2019, p. 1228–1233.
- [58] T. Wang, G. Xia, H. Li, C. Qi et H. Wang, “Patient specific respiratory motion model using two static CT images,” dans *Proceedings of the 2nd International Symposium on Artificial Intelligence for Medicine Sciences*. Beijing China : ACM, oct. 2021, p. 488–492.
- [59] M. L. Terpstra, M. Maspero, T. Bruijnen, J. J. Verhoeff, J. J. Lagendijk et C. A. Van Den Berg, “Real-time 3D motion estimation from undersampled MRI using multi-resolution neural networks,” *Medical Physics*, vol. 48, n°. 11, p. 6597–6613, nov. 2021.

- [60] M. Pohl, M. Uesaka, K. Demachi et R. Bhusal Chhatkuli, “Prediction of the motion of chest internal points using a recurrent neural network trained with real-time recurrent learning for latency compensation in lung cancer radiotherapy,” *Computerized Medical Imaging and Graphics*, vol. 91, p. 101941, 2021.
- [61] E. Lombardo, J. Dhont, D. Page, C. Garibaldi, L. A. Künzel, C. Hurkmans, R. H. Tijsen, C. Paganelli, P. Z. Liu, P. J. Keall, M. Riboldi, C. Kurz, G. Landry, D. Cusumano, M. Fusella et L. Placidi, “Real-time motion management in MRI-guided radiotherapy : Current status and AI-enabled prospects,” *Radiotherapy and Oncology*, vol. 190, p. 109970, janv. 2024.
- [62] A. Giger, R. Sandkühler, C. Jud, G. Bauman, O. Bieri, R. Salomir et P. C. Cattin, “Respiratory Motion Modelling Using cGANs,” dans *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López et G. Fichtinger, édit. Cham : Springer International Publishing, 2018, vol. 11073, p. 81–88.
- [63] T. Kim, S. J. Lee et T.-W. Jang, “Application of several machine learning algorithms for the prediction of afatinib treatment outcome in advanced-stage egfr-mutated non-small-cell lung cancer,” *Thoracic Cancer*, vol. 13, n°. 23, p. 3353–3361, 2022.
- [64] S. Hindocha, T. G. Charlton, K. Linton-Reid, B. Hunter, C. Chan, M. Ahmed, E. J. Robinson, M. Orton, S. Ahmad, F. McDonald et al., “A comparison of machine learning methods for predicting recurrence and death after curative-intent radiotherapy for non-small cell lung cancer : development and validation of multivariable clinical prediction models,” *EBioMedicine*, vol. 77, 2022.
- [65] Q. Yuan, T. Cai, C. Hong, M. Du, B. E. Johnson, M. Lanuti, T. Cai et D. C. Christiani, “Performance of a machine learning algorithm using electronic health record data to identify and estimate survival in a longitudinal cohort of patients with lung cancer,” *JAMA Network Open*, vol. 4, n°. 7, p. e2114723–e2114723, 2021.
- [66] A. Lorenc, A. Romaszko-Wojtowicz, Ł. Jaśkiewicz, A. Doboszyńska et A. Buciński, “Exploring the efficacy of artificial neural networks in predicting lung cancer recurrence : a retrospective study based on patient records,” *Translational Lung Cancer Research*, vol. 12, n°. 10, p. 2083, 2023.
- [67] D. Franceschini, L. Cozzi, F. De Rose, P. Navarria, A. Fogliata, C. Franzese, D. Pezzulla, S. Tomatis, G. Reggiori et M. Scorsetti, “A radiomic approach to predicting nodal relapse and disease-specific survival in patients treated with stereotactic body radiation therapy for early-stage non-small cell lung cancer.” *Strahlentherapie und Onkologie*, vol. 196, n°. 10, p. 922–931, oct. 2020.

- [68] F.-h. Tang, C. Xue, M. Y. Law, C.-y. Wong, T.-h. Cho et C.-k. Lai, “Prognostic Prediction of Cancer Based on Radiomics Features of Diagnostic Imaging : The Performance of Machine Learning Strategies,” *Journal of Digital Imaging*, vol. 36, n°. 3, p. 1081–1090, févr. 2023.
- [69] S. Lee, J. Jung, H. Hong et B.-S. Kim, “Prediction of two-year recurrence-free survival in operable NSCLC patients using radiomic features from intra-and size-variant peritumoral regions on chest ct images,” *Diagnostics*, vol. 12, n°. 6, p. 1313, 2022.
- [70] Z. Ding, C. Zhang, Q. Yao, Q. Liu, L. Lv et S. Shi, “Delta radiomics model for the prediction of overall survival and local recurrence in small cell lung cancer patients after chemotherapy,” *Academic Radiology*, vol. 31, n°. 3, p. 1168–1179, 2024.
- [71] W. T. Le, E. Vorontsov, F. P. Romero, L. Seddik, M. M. Elsharief, P. F. Nguyen-Tan, D. Roberge, H. Bahig et S. Kadoury, “Cross-institutional outcome prediction for head and neck cancer patients using self-attention neural networks,” *Scientific Reports*, vol. 12, n°. 1, p. 3183, févr. 2022.
- [72] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, R. H. Mak et H. J. Aerts, “Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging,” *Clinical Cancer Research*, vol. 25, n°. 11, p. 3266–3275, juin 2019.
- [73] Z. Yang, C. Wang, Y. Wang, K. J. Lafata, H. Zhang, B. G. Ackerson, C. Kelsey, B. Tong et F.-F. Yin, “Development of a multi-feature-combined model : proof-of-concept with application to local failure prediction of post-SBRT or surgery early-stage NSCLC patients,” *Frontiers in Oncology*, vol. 13, p. 1185771, sept. 2023.
- [74] P. Afshar, A. Mohammadi, P. N. Tyrrell, P. Cheung, A. Sigiuk, K. N. Plataniotis, E. T. Nguyen et A. Oikonomou, “DRTOP : deep learning-based radiomics for the time-to-event outcome prediction in lung cancer,” *Scientific Reports*, vol. 10, n°. 1, p. 12366, juill. 2020.
- [75] A. Braghetto, F. Marturano, M. Paiusco, M. Baiesi et A. Bettinelli, “Radiomics and deep learning methods for the prediction of 2-year overall survival in LUNG1 dataset,” *Scientific Reports*, vol. 12, n°. 1, p. 14132, août 2022.
- [76] D. E. Spratt, A. J. Wu, V. Adeseye, S. U. Din, F. Shaikh, K. M. Woo, Z. Zhang, A. Foster, K. E. Rosenzweig, R. Gewanter, J. Huang et A. Rimner, “Recurrence Patterns and Second Primary Lung Cancers After Stereotactic Body Radiation Therapy for Early-Stage Non-Small-Cell Lung Cancer : Implications for Surveillance,” *Clinical Lung Cancer*, vol. 17, n°. 3, p. 177–183.e2, mai 2016.
- [77] J. McClelland, D. Hawkes, T. Schaeffter et A. King, “Respiratory motion models : A review,” *Medical Image Analysis*, vol. 17, n°. 1, p. 19–42, janv. 2013.

- [78] P. J. Keall, G. S. Mageras, J. M. Balter, R. S. Emery, K. M. Forster, S. B. Jiang, J. M. Kapatoes, D. A. Low, M. J. Murphy, B. R. Murray, C. R. Ramsey, M. B. Van Herk, S. S. Vedam, J. W. Wong et E. Yorke, “The management of respiratory motion in radiation oncology report of AAPM Task Group 76a),” *Medical Physics*, vol. 33, n°. 10, p. 3874–3900, 2006.
- [79] H. J. Fayad, C. Buerger, C. Tsoumpas, C. Cheze-Le-Rest et D. Visvikis, “A generic respiratory motion model based on 4D MRI imaging and 2D image navigators,” dans *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*. Anaheim, CA, USA : IEEE, oct. 2012, p. 4058–4061.
- [80] N. Garau, R. Via, G. Meschini, D. Lee, P. Keall, M. Riboldi, G. Baroni et C. Paganelli, “A ROI-based global motion model established on 4DCT and 2D cine-MRI data for MRI-guidance in radiation therapy,” *Physics in Medicine & Biology*, vol. 64, n°. 4, p. 045002, févr. 2019.
- [81] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. V. Guttag et A. V. Dalca, “Voxelmorph : A learning framework for deformable medical image registration,” *IEEE Transactions on Medical Imaging*, vol. 38, p. 1788–1800, 2018.
- [82] M. Jaderberg, K. Simonyan, A. Zisserman et k. kavukcuoglu, “Spatial Transformer Networks,” dans *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama et R. Garnett, édit., vol. 28. Curran Associates, Inc., 2015.
- [83] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai et S. Chintala, “Pytorch : An imperative style, high-performance deep learning library,” dans *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, p. 8024–8035.
- [84] P. Seroul, D. Sarrut et D. Sarrut, “VV : a viewer for the evaluation of 4D image registration,” *The MIDAS Journal*, juill. 2008.
- [85] A. K. Ganti, A. B. Klein, I. Cotarla, B. Seal et E. Chou, “Update of incidence, prevalence, survival, and initial treatment in patients with non–small cell lung cancer in the us,” *JAMA oncology*, vol. 7, n°. 12, p. 1824–1832, 2021.
- [86] T. P. Coroller, P. Grossmann, Y. Hou, E. Rios Velazquez, R. T. Leijenaar, G. Hermann, P. Lambin, B. Haibe-Kains, R. H. Mak et H. J. Aerts, “CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma,” *Radiotherapy and Oncology*, vol. 114, n°. 3, p. 345–350, mars 2015.
- [87] S. Senthil, F. J. Lagerwaard, C. J. Haasbeek, B. J. Slotman et S. Senan, “Patterns of disease recurrence after stereotactic ablative radiotherapy for early stage non-small-cell

lung cancer : a retrospective analysis,” *The Lancet Oncology*, vol. 13, n°. 8, p. 802–809, août 2012.

[88] F. J. Candido dos Reis, G. C. Wishart, E. M. Dicks, D. Greenberg, J. Rashbass, M. K. Schmidt, A. J. van den Broek, I. O. Ellis, A. Green, E. Rakha *et al.*, “An updated predict breast cancer prognostication and treatment benefit prediction model with independent validation,” *Breast Cancer Research*, vol. 19, p. 1–13, 2017.

[89] V. Völkel, T. A. Huetting, T. Draeger, M. C. van Maaren, L. de Munck, L. J. Strobbe, G. S. Sonke, M. K. Schmidt, M. van Hezewijk, C. G. Groothuis-Oudshoorn *et al.*, “Improved risk estimation of locoregional recurrence, secondary contralateral tumors and distant metastases in early breast cancer : the influence 2.0 model,” *Breast cancer research and treatment*, vol. 189, p. 817–826, 2021.

[90] W. G. van der Plas-Krijgsman, D. Giardiello, H. Putter, E. W. Steyerberg, E. Bastiaannet, A. M. Stiggelbout, S. P. Mooijaart, J. R. Kroep, J. E. Portielje, G.-J. Liefers *et al.*, “Development and validation of the portret tool to predict recurrence, overall survival, and other-cause mortality in older patients with breast cancer in the netherlands : a population-based study,” *The Lancet Healthy Longevity*, vol. 2, n°. 11, p. e704–e711, 2021.

[91] K. J. Lafata, J. C. Hong, R. Geng, B. G. Ackerson, J.-G. Liu, Z. Zhou, J. Torok, C. R. Kelsey et F.-F. Yin, “Association of pre-treatment radiomic features with lung cancer recurrence following stereotactic body radiation therapy,” *Physics in Medicine & Biology*, vol. 64, n°. 2, p. 025007, janv. 2019.

[92] K. Lee, T. Le, E. Hau, G. G. Hanna, H. Gee, S. Vinod, S. Dammak, D. Palma, A. Ong, R. Yeghiaian-Alvandi, J. Buck et R. Lim, “A Systematic Review Into the Radiologic Features Predicting Local Recurrence After Stereotactic Ablative Body Radiotherapy (SABR) in Patients With Non-Small Cell Lung Cancer (NSCLC),” *International Journal of Radiation Oncology*Biology*Physics*, vol. 113, n°. 1, p. 40–59, mai 2022.

[93] E. Huynh, T. P. Coroller, V. Narayan, V. Agrawal, Y. Hou, J. Romano, I. Franco, R. H. Mak et H. J. Aerts, “CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer,” *Radiotherapy and Oncology*, vol. 120, n°. 2, p. 258–266, août 2016.

[94] M. Astaraki, C. Wang, G. Buizza, I. Toma-Dasu, M. Lazzeroni et O. Smedby, “Early survival prediction in non-small cell lung cancer from PET/CT images using an intra-tumor partitioning method,” *Physica Medica*, vol. 60, p. 58–65, avr. 2019.

[95] S. B. Edge, A. C. S. American Joint Committee on Cancer *et al.*, *AJCC cancer staging handbook : from the AJCC cancer staging manual*. Springer, 2010, vol. 19.

- [96] J. D. Brierley, M. K. Gospodarowicz et C. Wittekind, *TNM classification of malignant tumours*. John Wiley & Sons, 2017.
- [97] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper et H. J. Aerts, “Computational Radiomics System to Decode the Radiographic Phenotype,” *Cancer Research*, vol. 77, n°. 21, p. e104–e107, nov. 2017.
- [98] Isensee Fabian, Jäger Paul, Wasserthal Jakob, Zimmerer David, Petersen Jens, Kohl Simon, Schock Justus, Klein Andre, Roß Tobias, Wirkert Sebastian, Neher Peter, Dinkelacker Stefan, Köhler Gregor, Maier-Hein Klaus, “batchgenerators - a python framework for data augmentation.”
- [99] A. Bezjak, R. Paulus, L. E. Gaspar, R. D. Timmerman, W. L. Straube, W. F. Ryan, Y. I. Garces, A. T. Pu, A. K. Singh, G. M. Videtic, R. C. McGarry, P. Iyengar, J. R. Pantarotto, J. J. Urbanic, A. Y. Sun, M. E. Daly, I. S. Grills, P. Sperduto, D. P. Normolle, J. D. Bradley et H. Choy, “Safety and efficacy of a five-fraction stereotactic body radiotherapy schedule for centrally located non-small-cell lung cancer : Nrg oncology/rtog 0813 trial,” *Journal of Clinical Oncology*, vol. 37, n°. 15, p. 1316–1325, avr. 2019.
- [100] M. Giuliani, A. S. Mathew, H. Bahig, S. V. Bratman, E. Filion, D. Glick, A. V. Louie, S. Raman, A. Swaminath, A. Warner, V. Yau et D. Palma, “SUNSET : Stereotactic Radiation for Ultracentral Non-Small-Cell Lung Cancer—A Safety and Efficacy Trial,” *Clinical Lung Cancer*, vol. 19, n°. 4, p. e529–e532, juill. 2018.
- [101] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot et E. Duchesnay, “Scikit-learn : Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, p. 2825–2830, 2011.
- [102] J. Ho, T. Tumkaya, S. Aryal, H. Choi et A. Claridge-Chang, “Moving beyond p values : Everyday data analysis with estimation plots,” *BioRxiv*, 2019.
- [103] Q. Li, J. Kim, Y. Balagurunathan, Y. Liu, K. Latifi, O. Stringfield, A. Garcia, E. G. Moros, T. J. Dilling, M. B. Schabath, Z. Ye et R. J. Gillies, “Imaging features from pretreatment ct scans are associated with clinical outcomes in nonsmall-cell lung cancer patients treated with stereotactic body radiotherapy,” *Medical Physics*, vol. 44, n°. 8, p. 4341–4349, août 2017.