

Mémoire

Auteur : Kammegne Kamdem, Bertrand

Promoteur(s) : Baart, Véronique; Ponte, Stephan

Faculté : Faculté des Sciences

Diplôme : Master en sciences physiques, à finalité spécialisée en radiophysique médicale

Année académique : 2020-2021

URI/URL : <http://hdl.handle.net/2268.2/12712>

Avertissement à l'attention des usagers :

Tous les documents placés en accès restreint sur le site MatheO sont protégés par le droit d'auteur. Par conséquent, seule une utilisation à des fins strictement privées, d'enseignement ou de recherche scientifique est autorisée conformément aux exceptions légales définies aux articles XI. 189 et XI. 190. du Code de droit économique. Toute autre forme d'exploitation (utilisation commerciale, diffusion sur le réseau Internet, reproduction à des fins publicitaires, ...) sans l'autorisation préalable de l'auteur est strictement interdite et constitutive de contrefaçon.



FACULTÉ DES SCIENCES

DÉPARTEMENT DE PHYSIQUE

Évaluation du Machine Learning pour la planification de traitement dans RayStation

Mémoire en vue d'obtention du diplôme de master en radiophysique médicale

Comité de lecture :

Michel MATHOT

Sophie NICOLAS

Ngoc Duy NGUYEN

Promoteur :

Stéphan PONTE

Co-promoteur :

Véronique BAART

Auteur :

KAMMEGNE KAMDEM Bertrant

Août 2021

Remerciements

Avant toute chose, j'aimerais remercier tous ceux qui ont contribué de près ou de loin à la réalisation ce travail de fin d'étude.

Tout d'abord, je tiens à remercier chaleureusement la radiophysicienne en chef du service de radiothérapie du CHU de Liège, Madame Véronique BAART pour m'avoir accueilli au sein du service, pour son aide précieuse et d'avoir accepté d'être la copromotrice de mémoire.

Mes remerciements les plus respectueux vont à Monsieur Stéphan PONTE pour la qualité de son encadrement, sa pédagogie, ses précieux avis et tout le temps qu'il m'a consacré. Je le remercie également pour sa disponibilité, sa gentillesse et le partage de ses connaissances tout au long de mon stage.

Je voudrais aussi remercier Manon BAIWIR, Michel MATHOT et Imad HAMMAMI pour les échanges d'idées utiles durant la période de stage. Merci également au reste du personnel du service de radiothérapie du CHU de Liège : les physiciens, les dosimétristes, les médecins et les manipulateurs (infirmières et technologues) pour la qualité de leur accueil.

Je tiens à remercier le Président du Jury et les membres du Jury d'avoir respectivement accepté de présider mon travail et faire partie de mon comité de lecture.

J'aimerais exprimer toute ma gratitude à ma famille pour m'avoir rappelé à de nombreuses reprises qu'elle croyait en moi. Je remercie également mes proches pour leur soutien physique et moral tout au long de cette formation.

Enfin, j'aimerais remercier sincèrement tout le personnel du département de physique de l'Université de Liège, principalement l'équipe pédagogique du Master en sciences physiques, à finalité spécialisée en Radiophysique Médicale.

Résumé

L'objectif de ce mémoire est de faire une évaluation de la qualité des plans de traitement VMAT optimisés avec l'intelligence artificielle notamment le Machine Learning/Deep Learning par rapport aux plans classiques dans RayStation pour une intégration en routine clinique.

La qualité des plans a été évaluée en terme de conformation au PTV et de la protection de tissus sains grâce aux indices de conformité et d'homogénéité. La complexité des plans pour les deux modalités de planification est examinée à partir de la précision du calcul de dose à l'aide de la métrique *LOIC* et de la délivrabilité du traitement par la machine grâce à la métrique *MCS_V*. Les plans sont répartis en trois groupes selon la prescription. L'analyse statistique des plans est effectuée avec le test signé de Wilcoxon.

La conformation au PTV est presque équivalente pour les deux modalités de planification du groupe 1 (20×300 cGy) avec pour valeur moyenne 0,84 et non significative. Par contre, la planification automatisée est plus conforme que la planification manuelle pour le groupe 2 (25×263 cGy). Les valeurs obtenues sont respectivement $0,839 \pm 0,066$ et $0,826 \pm 0,033$. Le plan automatisé est aussi plus conforme dans le groupe 3 (25×275 cGy). Par ailleurs, ces plans sont plus homogènes que les plans classiques pour les trois groupes. La valeur du *MCS_V* est en moyenne plus faible pour les plans automatisés et le *LOIC* est en moyenne plus élevé par rapport à la valeur du plan original. Ceci montre que les plans optimisés avec Machine Learning sont plus modulés, induisant une augmentation des unités moniteurs. Les mesures réalisées avec l'ArcCHECK ont permis de déterminer l'indice γ . Les résultats du groupe 1 ont révélé que les différences n'étaient pas statistiquement significatives pour le critère 3%/3mm mais significatives pour le critère 3%/2mm. Par contre, pour le groupe 2, le test de Wilcoxon a démontré que la différence de valeur moyenne était statistiquement significative pour le critère 3%/3mm et non significative pour le critère 3%/2mm.

L'introduction de la planification automatisée en routine clinique est favorable puisqu'on obtient une meilleure conformation et une meilleure homogénéité de la dose au PTV bien que les métriques donnent des valeurs moins bonnes mais restent acceptables. Néanmoins, les mesures avec l'ArcCHECK montrent que la machine n'aura aucune difficulté à délivrer le traitement. De plus, la durée pour réaliser un plan de traitement acceptable avec ML reste très inférieure à la durée pour un plan optimisé manuellement.

Table des matières

Remerciements	1
Résumé	2
Introduction	9
1 Architecture de l'Intelligence Artificielle	10
1.1 Description du Machine Learning et Deep Learning	10
1.2 Présentation des algorithmes basés sur des réseaux de neurones	12
1.2.1 Apprentissage supervisé	12
1.2.2 Apprentissage non supervisé	14
1.3 Algorithme basé sur les arbres de décision	15
1.3.1 Arbres de décision	15
1.3.2 Random Forest	17
1.4 Application du Machine Learning en médecine	18
1.4.1 Imagerie médicale	18
1.4.2 Radiothérapie externe	19
1.4.3 Assurance qualité patient	20
2 Matériels et méthodes	22
2.1 Matériels	22
2.1.1 L'arcthérapie	22
2.1.2 L'accélérateur linéaire Elekta Infinity Agility	22
2.1.3 Description de RayStation 10A	23
2.1.4 Indices	25
2.1.5 Métriques	25
2.1.6 Optimisation de la charge du travail en PSQA	28
2.2 Méthodes	29
2.2.1 Les patients	29
2.2.2 Configuration du TPS pour la planification	29
2.2.3 Création d'un nouveau plan avec Machine Learning	30
2.2.4 Optimisation des plans de traitement	30
2.2.5 Mesure avec l'ArcCHECK	32

2.2.6	Analyse statistique : Test de Wilcoxon	32
3	Résultats	34
3.1	Indice de conformité (CI)	34
3.2	Indice d'homogénéité (HI)	35
3.3	Couverture du PTV des plans	35
3.4	Evaluation de la dose aux organes à risque	36
3.4.1	Moyenne des doses sur chaque OAR	37
3.4.2	Moyenne des doses reçues par 2% du volume des OAR	38
3.5	Nombre d'Unités Moniteurs des plans	39
3.6	Métrique de délivrabilité : MCS_V	39
3.7	Métrique de précision : $LOIC$	40
3.8	Mesures et calculs de l'indice γ	41
4	Discussions	42
	Conclusion et Perspectives	47
A	Données	48

Table des figures

1	Vue d'ensemble des concepts courants utilisés dans le domaine de l'IA	8
1.1	Exemple de neurone.	11
1.2	Représentation d'un réseau neuronal multicouche ou perceptron multicouche. . .	11
1.3	Représentation d'un réseau neuronal multicouche ou perceptron multicouche. . .	13
1.4	Configuration d'un réseau FCN.	13
1.5	Structure des réseaux de neurones récurrents	14
1.6	Exemples de réseaux.	14
1.7	Représentation de la machine de Boltzmann.	15
1.8	Illustration générale d'un arbre de décision avec partitionnement.	16
1.9	Illustration du principe de fonctionnement du Random Forest (RF)	17
1.10	Illustration schématique du principe de volumes définis par l'ICRU.	19
2.1	Illustration de la tête de l'accélérateur Elekta Infinity Agility.	23
2.2	Organigramme de l'algorithme ARF.	24
2.3	Illustration de deux plans de traitement	27
2.4	Illustration de caractéristiques de la planification.	29
2.5	Interface de configuration d'un plan dans le TPS.	31
2.6	Illustration de la fonction <i>Dose fall-off</i> autour du PTV.	31
2.7	Positionnement de l'ArcCHECK centré aux lasers à l'isocentre du LINAC	32
3.1	Box plot de l'indice de conformité des plans.	34
3.2	Box plot de l'indice d'homogénéité HI des plans.	35
3.3	Couverture du PTV du groupe 1 (20×300 cGy).	36
3.4	Couverture du PTV du groupe 2 (25×263 cGy).	36
3.5	Dose moyenne (cGy) aux OAR pour le groupe 3 (25×275 cGy)	37
3.6	$D_{2\%}$ (cGy) aux OAR pour le groupe 3 (25×275 cGy)	38
3.7	Nombre d'UM pour les plans de traitement.	39
3.8	Box plot des métriques MCS_V des plans de traitement.	40
3.9	Box plot des métriques $LOIC$ des plans de traitement.	41
A.1	Distribution de dose pour le plan d'origine.	48
A.2	Distribution de dose pour le plan automatique.	48

Liste des tableaux

3.1	Indice de conformité du patient du groupe 3 (25×275 cGy)	35
3.2	Indice d'homogénéité du patient du groupe 3 (25×275 cGy)	35
3.3	Couverture du PTV pour le groupe 3 (25×275 cGy)	36
3.4	Moyenne des doses calculées aux OAR des patients du groupe 1 (20×300 Gy) .	37
3.5	Moyenne des doses calculées aux OAR des patients du groupe 2 (25×263 cGy)	37
3.6	Moyenne des $D_{2\%}$ aux OAR du groupe 1 (20×300 cGy)	38
3.7	Moyenne des $D_{2\%}$ aux OAR du groupe 2 (25×263 cGy)	38
3.8	Nombre d'UM du seul plan du groupe 3 (25×275 cGy)	39
3.9	La métrique MCS_V du seul plan du groupe 3 (25×275 cGy)	40
3.10	La métrique LOIC du seul plan du groupe 3 (25×275 cGy)	40
A.1	Valeurs de l'indice γ du groupe 3 (25×275 cGy)	48
A.2	Valeurs de l'indice γ du groupe 1 (20×300 cGy)	49
A.3	Valeurs de l'indice γ du groupe 2 (25×263 cGy)	49

Abréviations

AAPM American Association of Physicists in Medicine

AAV Aperture Area Variability

AE Autoencoder

ARF Adaptative Random Forest

cGy centi-Gray

CNN Convolutional Neural Network

CP Control Point

DL Deep Learning

DT Decision Tree

EF Extracting Features

ESNet Enhanced Skip Network

FCN Fully Connected Network

GAN Generative Adversarial Network

GPU Graphics Processing Unit

IA Intelligence Artificielle

IMRT Intensity-Modulated Radiation Therapy

LOIC Leaf Offset Impact on Calculation

LSTM Long Short-Term Memory

LSV Leaf Sequence Variability

MCS Modulation Complexity Score

ML Machine Learning

OAR organe à risque

OOB Out-Of-Bag

PSQA Patient Specific Quality Assurance

RBM Restricted Boltzmann Machine

RF Random Forest

RNN Recurrent Neural Network

TPS Treatment Planning System

VMAT Volumetric Modulated Arc Therapy

Introduction

La révolution de l'informatique a été accélérée pendant la période de la seconde guerre mondiale avec la volonté de comprendre comment relier le principe de fonctionnement des machines à celui des êtres organiques. L'idée était de pouvoir allier les théories mathématiques, de l'électronique et de l'automatisation avec la pratique. Warren McCulloch fut le premier à développer le modèle mathématique d'un neurone biologique. John Neumann et Alan Turing feront la transition entre les calculateurs décimaux vers les machines binaires et sont arrivés à formaliser l'architecture des ordinateurs à cette époque [1]. Le mot "Intelligence Artificielle (IA)" a été utilisé pour la première fois par le mathématicien John McCarthy marquant ainsi sa révolution. Malheureusement, les machines ne disposaient pas assez de mémoire pour répondre aux nombreuses attentes rendant impossible l'utilisation du langage de programmation. Grâce aux premiers microprocesseurs, l'IA reprend un nouvel essor mais de courte durée. Ceci est dû au fait que les données générées par les Hommes devenaient très importantes. En retour, la programmation de tel volume de données devenait complexe et demandait un effort considérable. C'est à ce moment que les machines ont été qualifiées de boîte noire. Depuis quelques années, nous assistons à un nouvel essor de l'IA principalement pour deux raisons : l'accès à des volumes massifs de données et le développement des unités graphiques de traitement ou Graphics Processing Unit (GPU) pour accélérer les calculs d'apprentissage. La conférence de Dartmouth a permis d'étendre le domaine de l'intelligence artificielle en donnant naissance au Machine Learning (ML) et l'apprentissage profond ou Deep Learning (DL) (Fig. 1).

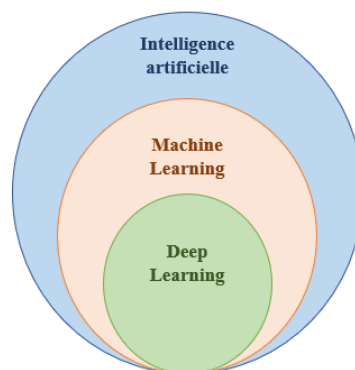


FIGURE 1 – Vue d'ensemble des concepts courants utilisés dans le domaine de l'IA

Le DL/ML est un algorithme d'apprentissage automatique basé sur des réseaux de neurones permettant à un système d'apprendre à partir de son expérience. Plusieurs techniques d'apprentissage ont été développés pour permettre à un système d'être performant. Il y'a l'apprentissage supervisé qui utilise des algorithmes nécessitant une aide extérieure, l'apprentissage par renforcement est basé sur le concept de décision et l'apprentissage non supervisé qui apprend des données caractéristiques en entrée [2]. Ainsi, plusieurs réseaux de neurones ont été développés tels que les réseaux de neurones convolutifs ou Convolutional Neural Network (CNN), les réseaux antagonistes génératifs ou Generative Adversarial Network (GAN) [3], les autoencodeurs ou Autoencoder (AE) pour la réduction des dimensions des données, les réseaux entièrement connectés ou Fully Connected Network (FCN) principalement utilisés dans la segmentation des échantillons de données et Random Forest (RF) pour la classification et la régression. Le domaine de la médecine fait partie des plus grands défis de l'intelligence artificielle pour l'aide à la décision médicale par exemple. L'IA est utilisée en médecine pour améliorer le flux de travail et des opérations cliniques, pour faciliter le diagnostic des maladies. L'IA notamment le DL ou le ML est appliquée en imagerie médicale dans la reconstruction des images en utilisant principalement l'algorithme GAN, de même dans la segmentation des organes sur la base de l'analyse des contours en utilisant différents algorithmes tels que le CNN, FCN et les réseaux de neurones récurrents. Dans le cadre de la radiothérapie, l'émergence de l'intelligence artificielle est susceptible de réduire considérablement le temps de travail dans la réalisation des plans de traitement des patients. L'IA (ML ou DL) est de plus en plus utilisé dans l'amélioration de la qualité des plans de traitement avec les techniques de radiothérapie à modulation d'intensité ou Intensity-Modulated Radiation Therapy (IMRT) comme par exemple le cas de la prostate.

Dans la suite de ce travail, nous allons au premier chapitre présenter l'architecture de l'intelligence artificielle tout en faisant le point sur les différents algorithmes implémentés dans le Machine Learning et le Deep Learning. Dans cette même optique d'idée, nous ferons le point sur l'intégration de l'apprentissage profond dans la médecine plus précisément en radiothérapie. Dans le chapitre suivant, nous présenterons la méthodologie de notre étude avec une description du processus d'optimisation de l'algorithme utilisé par le système de planification. Les résultats seront présentés au chapitre 3 et enfin des discussions au chapitre 4.

Chapitre 1

Architecture de l'Intelligence Artificielle

L'IA est un domaine de l'informatique qui traite de la conception et du développement de machines intelligentes. C'est un ensemble de techniques et de théories permettant de réaliser des appareils qui seront capables, grâce à un algorithme approprié de simuler l'intelligence humaine. L'IA a été implémenté dans de nombreux secteurs d'activités et est devenue un outil incontournable pour de systèmes informatiques. Le raisonnement et l'autocorrection sont des fonctionnalités impressionnantes des machines intelligentes. L'IA est entrée depuis quelques années dans une nouvelle ère donnant lieu à de nouveaux espoirs. Il a été implémenté à la médecine pour garantir une prestation de qualité. L'intelligence artificielle a été étendue donnant naissance au ML et au DL.

1.1 Description du Machine Learning et Deep Learning

Le Machine Learning (ML) est un sous-ensemble de l'IA basé sur la capacité du système à apprendre un modèle à partir d'un ensemble d'informations. Cet apprentissage du système utilise un réseau de neurones artificiel qui imite un neurone humain (Fig. 1.1). Selon Arthur Samuel [4], l'apprentissage des machines est défini comme un algorithme permettant à un ordinateur d'analyser et d'interpréter une information sans être explicitement programmé pour produire un résultat particulier.

Les données entrées sont les plus importantes dans la structure d'un réseau neuronal. Elles sont traitées et transformées en un résultat de sortie en appliquant une fonction d'activation ou fonction seuil [5]. La fonction mathématique 1.1 permet de traiter l'information en entrée afin d'avoir un résultat en sortie dans un neurone ou b est le seuil d'activation du neurone.

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right) \quad (1.1)$$

Lorsqu'on présente un exemple d'échantillon en entrée, on fait tourner les poids w du réseau neuronal jusqu'à ce qu'ils fassent le lien entre l'entrée et la sortie. Le poids est par définition un ensemble de connexions synaptiques entre les neurones.

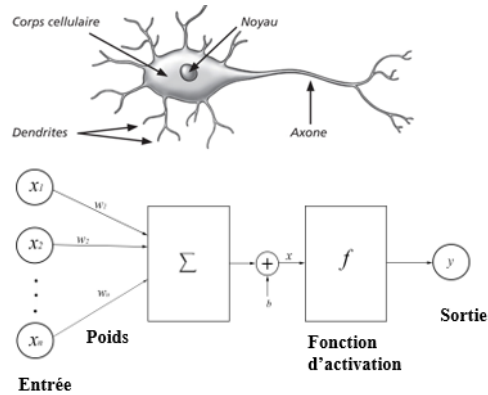


FIGURE 1.1 – Exemple de neurone.
En haute le neurone biologique et en bas le neurone artificiel.

En appliquant ce principe sur plusieurs neurones, le modèle devient complexe et difficile à mettre en œuvre. Par exemple pour une image de n -pixels, il faudra n points d'entrées à attribuer au réseau; ce qui n'est pas évident. Pour cela, on crée une étape qui permettra d'extraire les caractéristiques intermédiaires de l'image (forme, contours, etc.). Toutefois, la qualité de la reconnaissance va dépendre du travail intermédiaire nécessitant une durée importante. C'est ici qu'intervient le Deep Learning (*DL*) dont l'idée est de sauter cette étape intermédiaire. Il est fondé sur un ensemble de neurones artificiels appelés réseaux multicouches ou les résultats de la première couche serviront d'entrée pour la couche suivante et ainsi de suite (Fig. 1.2).

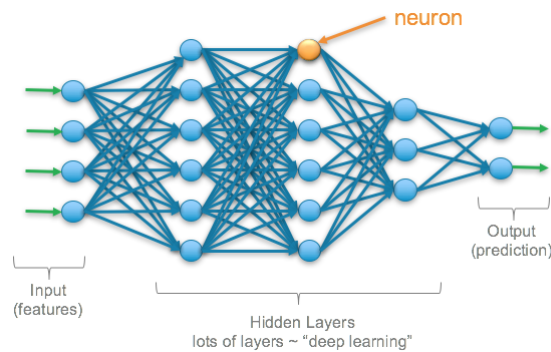


FIGURE 1.2 – Représentation d'un réseau neuronal multicouche ou perceptron multicouche.

La différence entre le DL et les autres méthodes d'apprentissages réside dans le fait que le DL peut apprendre automatiquement sans apport d'une autre fonctionnalité. La performance du DL repose en quelque sorte sur sa profondeur et sa composition.

Nous allons, dans la suite de ce travail nous focalisés sur ces deux méthodes d'apprentissage et principalement sur le Machine Learning, tout en mettant en évidence leurs implications dans la médecine et plus précisément en radiothérapie externe.

1.2 Présentation des algorithmes basés sur des réseaux de neurones

De nombreuses architectures ont été développés selon les besoins pour implémenter le DL. Ces dernières sont également classées selon deux méthodes d'apprentissage : la méthode par apprentissage supervisé et la méthode par apprentissage non supervisé.

1.2.1 Apprentissage supervisé

L'apprentissage supervisé est un type d'apprentissage qui utilise un ensemble de données d'entraînement bien étiquetés. Étiqueter les données signifie classer les données en différentes catégories. Ces données d'entraînement sont utilisées pour former les algorithmes à classer les données et à prédire de meilleure résultat. Cette méthode d'apprentissage est utilisée principalement pour faire de la classification qui permet de ranger les données par catégorie et la régression permettant de trouver le lien entre deux variables sur des données lors de son apprentissage.

a) Les réseaux de neurones convolutifs

Le réseau de neurones le plus utilisé en apprentissage profond est le réseau de neurones convolutif ou *CNN*. Il s'agit d'une forme particulière de réseau neuronal multicouche. L'architecture des connexions est inspirée de celle du cortex visuel des mammifères. Ils sont appliqués dans plusieurs domaines d'activité telles que l'analyse d'image. L'architecture typique d'un réseau convolutif se compose de plusieurs couches (Fig. 1.3).

La couche convolutive : Elle est appliquée comme un filtre de convolution permettant d'extraire les caractéristiques de l'image. Chaque filtre utilisé permet de reconnaître des motifs qui peuvent être des contours ou certaines formes explicites.

La couche de correction ReLU : Elle est appliquée après la couche de convolution et joue le rôle de fonction d'activation. Elle est définie par l'équation ci-dessous où x est l'entrée du neurone.

$$ReLU(x) = \max(0, x) \tag{1.2}$$

La couche de Pooling : Elle est la couche généralement placée entre la couche de convolution et le ReLU. Le but est de réduire le nombre de paramètres à calculer de même que la taille du réseau tout en préservant les informations les plus importantes. Ceci permettra de contrôler le surapprentissage qui est la cause principale des mauvaises performances des modèles prédictifs. Par définition, le surapprentissage est une étape qui repose principalement sur la mémorisation des données d'exemples empêchant la machine de prédire sur de nouvelles valeurs

inconnues. Les données en sortie sont introduites dans la couche entièrement connectée (Fully connected). Il existe plusieurs itérations des deux couches de convolution avec des pooling.

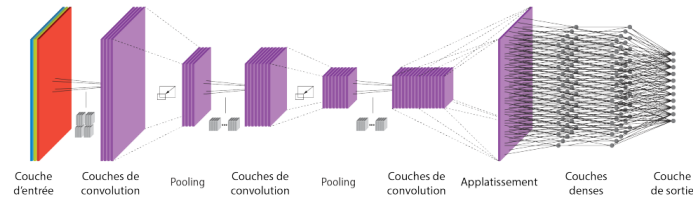


FIGURE 1.3 – Représentation d'un réseau neuronal multicouche ou perceptron multicouche.

b) Réseaux entièrement convolutifs

Le réseau entièrement convolutif ou *FCN* est un CNN modifié proposé par Shelhamer et al. [6], où les couches entièrement connectées à la fin du CNN sont remplacées principalement par des couches convolutives. La forme la plus répandue du FCN est le réseau encodeur-décodeur convolutif généralement utilisé pour la détection des contours d'images (Fig. 1.4). L'encodeur possède plusieurs couches convolutives pour encoder les images et le décodeur dispose de plusieurs couches déconvolutives servant à décoder les informations issues du codeur.

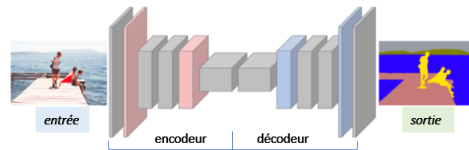


FIGURE 1.4 – Configuration d'un réseau FCN.

Différentes variantes de ce réseau ont été explorées telles que le réseau codeur-décodeur résiduel et U-net plus flexible. Prasad et al. [7] proposent un autre type de ce réseau appelé réseau à saut amélioré ou Enhanced Skip Network (ESNet) pour l'apprentissage bout en bout des graphiques. De même, Badrinarayanan et al. [8] ont créé un modèle appelé SegNet pour répondre en temps réel à la segmentation.

c) Réseaux de neurones récurrents

Les réseaux de neurones récurrents ou *Recurrent Neural Network (RNN)* sont une variété de réseaux de neurones particulièrement adaptés à plusieurs tâches comme le traitement de l'étiquetage [9]. Ils utilisent les informations de la sortie précédente pour faire une prédiction sur la donnée suivante et forment donc des boucles répétitives (Fig. 1.5). Ceci montre que, les RNN sont généralement formés par rétropropagation. Pour empêcher les erreurs de rétropropagation dans les RNN, les modèles à long terme ou *Long Short-Term Memory (LSTM)* ont été développés [10]. Ces derniers permettent aux réseaux de ne prendre en compte que certaines informations passées afin d'accorder plus d'importance à la prédiction des nouvelles informations.

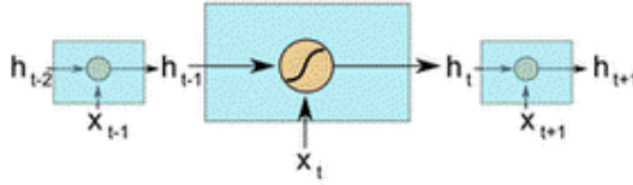


FIGURE 1.5 – Structure des réseaux de neurones récurrents

1.2.2 Apprentissage non supervisé

Aucune étiquette n'est fournie dans ce cas et l'algorithme se concentre uniquement sur la détection des données qu'il reçoit et les réorganisent. Par conséquent, au lieu d'essayer de prédire une sortie particulière pour chaque entrée, les algorithmes tentent plutôt de découvrir les similarités dans les données et les regroupent. Plusieurs algorithmes sont conçus pour cette méthode d'apprentissage.

a) Réseaux antagonistes génératifs

Les réseaux antagonistes génératifs ou *GAN* sont un type spécial du modèle de réseau de neurones. Ils ont été introduits par Goodfellow et al. [3] pour modéliser la représentation des images. Ils sont constitués de deux réseaux, l'un est le générateur (G) qui prend en entrée des signaux aléatoires (z) et le discriminateur $G(z)$ qui génère un échantillon aussi réaliste que possible (Fig. 1.6a) [11]. Le discriminateur (D) reçoit les données créées par le générateur et fait de son mieux pour les distinguer des données vraies. Lorsque le discriminateur devient incapable de faire la différence entre les données vraies et celles provenant du générateur, on dit que l'état optimal est atteint. On peut donc entraîner ces modèles en utilisant la rétropropagation et les algorithmes de dropout [12] (Fig. 1.6).

L'idée du GAN est de permettre à deux ou plusieurs réseaux de neurones de se défier et d'atteindre un équilibre pendant l'optimisation. Une variante du GAN la plus connue est le cGAN [13].

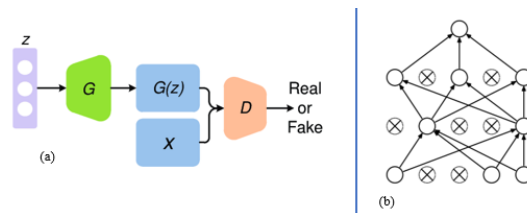


FIGURE 1.6 – Exemples de réseaux.

(a) Generative Adversarial Network (GAN) et (b) Réseau de neurone après l'application du dropout.

b) Auto-encodeurs

L'auto-encodeur (AE) est un réseau de neurones utilisé comme méthode de réduction de la dimensionnalité et d'extraction des caractéristiques. L'AE est constitué de trois compartiments. L'encodeur, qui est un réseau de neurones entièrement connectés et a pour rôle de compresser les données d'entrée afin de réduire ses dimensions. Ensuite l'espace latent appelé code et enfin Le décodeur qui a pour rôle d'analyser les données encodées afin de trouver les possibilités de reconstruire les données d'origine.

c) Machines de Boltzmann restreintes

Dans sa forme originale, la machine de Boltzmann est un réseau de neurones ayant la particularité de faire correspondre un ensemble de neurones entre eux. Elle est dite restreinte lorsque ses connexions sont strictement limitées à un sous-ensemble de toutes les connexions possibles et est constituée principalement de deux couches (Fig. 1.7) [14].

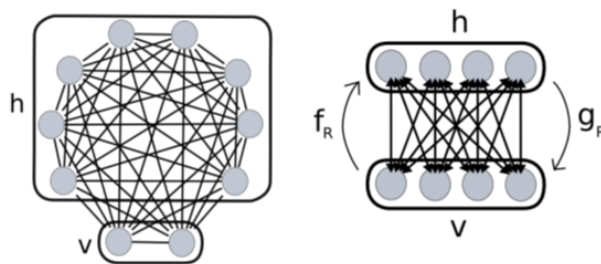


FIGURE 1.7 – Représentation de la machine de Boltzmann.

À gauche le cas original et à droite la topologie d'une machine de Boltzmann restreinte [14]. V et h sont respectivement les couches d'entrée (visibles) et les couches cachées.

À la différence des CNN, la circulation des données ne se fait plus uniquement de la couche d'entrée vers la couche de sortie mais dans les deux sens. Dans la pratique, les Restricted Boltzmann Machine (RBM) servent généralement à la réduction de dimension, la classification et la modélisation des données.

1.3 Algorithme basé sur les arbres de décision

1.3.1 Arbres de décision

Un arbre de décision ou *Decision Tree (DT)* est un outil très employé dans plusieurs domaines tels que les statistiques, la théorie des décisions ou encore l'apprentissage automatique. Dans sa formule générale, il est considéré comme un algorithme permettant d'illustrer un ensemble de choix. Il est utilisé pour résoudre les problèmes de classifications basés sur les variables qualitatives et des problèmes de régressions basés sur la prédiction des variables quantitatives.

Un DT modélise hiérarchiquement des données d'entrée pour prédire un résultat général. Cette modélisation se fait pendant la phase d'apprentissage en partitionnant sa population [15].

Le choix de la règle de partitionnement est par conséquent très important pour l'analyse et la compréhension du principe de fonctionnement des forêts aléatoires.

a) Phase de construction d'un arbre de décision

Semblable à un organigramme, un DT est composé d'un nœud racine par lequel entrent les données initiales. Cette racine va scinder les échantillons qu'elle contient en sous-groupes donnant lieu à d'autres nœuds et ainsi de suite. Une règle de partitionnement est appliquée sur chaque nœud permettant la répartition des données dans les sous-groupes et sans chevauchement avec les valeurs de réponse similaire. Les résultats du test sont représentés par chaque branche partant d'un nœud primaire vers plusieurs autres en fonction du nombre de partitions créées.

Ce processus est appliqué sur chaque nouveau nœud jusqu'à ce qu'au bout de chaque branche, il n'existe plus que des nœuds terminaux appelés feuilles. Cependant, l'arrêt de croissance d'un arbre de décision dépend de la pureté de ses feuilles c'est-à-dire que, les données présentes dans une feuille n'appartiennent qu'à une seule classe (Fig. 1.8).

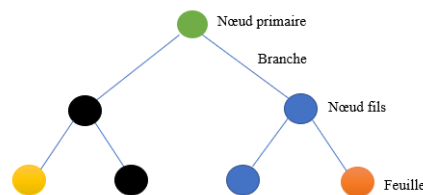


FIGURE 1.8 – Illustration générale d'un arbre de décision avec partitionnement.

Lorsque le DT a atteint sa taille maximale et devient incapable d'optimiser, Sebban et al. [16] ont proposé la technique dite d'élagage qui est un processus par lequel la taille de l'arbre est réduite en supprimant *a posteriori* les branches dans les parties terminales qui ont peu d'intérêt dans le but de maintenir les bonnes performances prédictives. Dans le cas où aucune technique d'élagage n'est appliquée, l'arbre sur-apprend les données d'apprentissage [17]. La plupart des algorithmes utilisés dans les RF n'implémentent pas la méthode d'élagage.

b) Évaluation des règles de partitionnement

Dans un arbre de décision, les règles de partitionnement sont évaluées sur la base des mesures d'impureté dans l'ensemble des données. Quel que soit le méthode utilisée pour faire ces mesures, les principes suivants doivent impérativement être respectés : l'impureté doit être minimale lorsque les données dans un nœud possèdent la même étiquette où appartiennent à la même classe et, maximale lorsque chaque étiquette est représentée par un nombre égal d'échantillons dans les données [18]. Il existe plusieurs critères de partitionnement et le plus utilisé est l'entropie de Shannon [19]. Ce critère est principalement utilisé dans les algorithmes de forêts aléatoires comme le CART (Classification And Regression Tree), ID3 développé par Ross Quinlan [18].

1.3.2 Random Forest

Une forêt aléatoire ou *RF* a été proposé pour la première fois par Tin Ho [20]. L'algorithme consistait à faire croître un ensemble d'arbres de décisions sur un ensemble de données dans l'objectif de prédire la variable cible. Toutes les fonctionnalités n'étaient pas prises en compte dans l'algorithme de Ho. Leo Breiman fait une amélioration en introduisant le concept de Bagging [21] qui est une contraction de Bootstrap Aggregating. Le bagging est une classe d'algorithme utilisé pour améliorer la classification et la régression des données dans un modèle RF. Le Bootstrap utilisé dans cet algorithme permet d'assurer la stabilité du système et d'éviter le surajustement. Cette combinaison (Bootstrap Aggregating) permet d'avoir un algorithme avec une meilleure précision que les arbres de décisions traditionnels.

a) Principe de fonctionnement

Tout arbre du RF est construit sur la base d'un échantillon Bootstrap ou un sous-échantillon des données. Par conséquent, toutes les données excluent dans le processus d'échantillonnage lors de l'apprentissage sont appelées données Out-Of-Bag (OOB). De plus, la sélection des variables à chaque nœud parmi les variables se fait aléatoirement. L'algorithme crée donc un arbre de décision pour chacun des échantillons sélectionnés. Après une optimisation individuelle, il obtient un résultat de prédiction sur chaque arbre créé. La prédiction finale provient d'une décision majoritaire de l'ensemble des arbres (Fig. 1.9)[22].

Les données OOB sont utilisées pour estimer l'erreur de prédiction de RF appelé erreur OOB. Cette dernière permet d'évaluer les performances de prédiction d'un modèle RF. Les données Bootstrap et les données OOB serviront de données d'apprentissage pendant la croissance de l'arbre dans une forêt aléatoire.

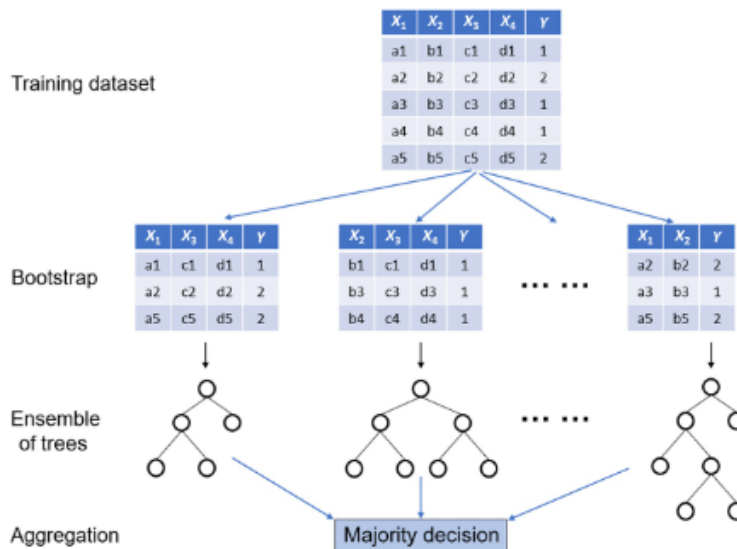


FIGURE 1.9 – Illustration du principe de fonctionnement du RF

1.4 Application du Machine Learning en médecine

Les concepts, techniques et outils de l'IA deviennent de plus en plus important dans le domaine médical. Grâce aux développements des unités de traitement graphique ou GPU, l'intelligence artificielle s'intègre parfaitement dans les outils d'aide aux personnels de la santé confrontés aux problèmes de plus en plus complexes. Ces GPU ont permis de développer des algorithmes très performants qui permettent l'adoption du ML/DL dans de nombreuses applications médicales telles que l'imagerie médicale, l'aide au diagnostic, la segmentation des organes ainsi que la planification des traitements en radiothérapie externe.

En physique médicale, la radiothérapie et l'imagerie médicale sont deux domaines les plus connus de l'IA notamment l'apprentissage automatique [23].

1.4.1 Imagerie médicale

L'imagerie médicale consiste à utiliser différents types de rayonnements dans le but de produire des images internes de l'anatomie humaine pour le diagnostic médical. Plusieurs techniques sont employées à cet effet selon la zone et/ou la maladie telles que l'imagerie par résonance magnétique, la tomодensitométrie et la tomographie par émission de positons. L'apprentissage automatique est utilisé en imagerie médicale pour la reconstruction d'images et la segmentation des organes.

a) Reconstruction d'image

Les algorithmes habituellement utilisés pour la reconstruction des images tomographiques sont classés en deux catégories [24]. D'une part, la reconstruction analytique basée sur la transformée de Radon en appliquant par exemple les approches de type rétroprojection filtrée [25]. D'autre part la reconstruction itérative fonctionnant par approximations successives afin de trouver la solution qui se rapproche au mieux de l'objet réel [26]. La nouvelle méthode de reconstruction d'images est basée sur l'apprentissage automatique [23]. Les GAN sont les plus utilisés pour la reconstruction des images [27]. D'autres algorithmes sont également utilisés tel que le CNN utilisé par Dong et al.[28].

b) Segmentation des organes

La segmentation est le processus de séparation d'une image en différentes régions d'intérêt (ROI). Elle consiste à déterminer les limites des organes ou des tumeurs à partir des images médicales reconstruites. La segmentation est très importante en radiothérapie car elle permet d'identifier les lésions afin d'éviter au mieux les tissus sains environnants pendant le traitement. Dans la pratique clinique actuelle, la délimitation du volume cible tumorale et les organe à risque (OAR) sont effectués par des médecins oncologues avec une durée non négligeable.

De plus, pour un même oncologue, il n'est pas possible de reproduire avec la même précision sa propre délimitation pour un même organe [23]. Par conséquent, la méthode de segmentation

automatique permettrait de pallier à ces problèmes de manque de précision avec un gain de temps considérable. Les algorithmes destinés à la segmentation automatique d'images sont basés sur les réseaux de neurones [29]. Ciresan et al.[30] adoptent le CNN pour la segmentation des membranes neuronales des images. Les réseaux entièrement convolutifs (FCN) ont été utilisés par Lei et al.[31] pour la segmentation des lésions cutanées et les résultats ont montré que, cette méthode de segmentation est meilleure que les méthodes traditionnelles. De même, Nie et al.[32] exploitent l'efficacité du FCN dans la segmentation des images cérébrales chez les nourrissons et Gordienko et al.[33] ont utilisé un réseau de neurone basé sur U-Net pour la segmentation des tumeurs pulmonaires.

1.4.2 Radiothérapie externe

La radiothérapie est l'une des modalités la plus utilisée pour le traitement du cancer et des maladies bénignes . Elle consiste en l'irradiation des patients grâce aux rayonnements ionisants produits à l'aide des accélérateurs linéaires. La radiothérapie externe peut être utilisée pour soulager le patient des douleurs, pour le guérir ou pour limiter la progression de la maladie. L'application des techniques liées à l'apprentissage automatique est en émergence dans le domaine de la radiothérapie.

a) Définition des volumes cibles

La commission internationale des unités et mesures de rayonnements (ICRU) a publié des définitions des volumes cibles pour la radiothérapie dans ses rapports n°50 [34] et n°62 [35]. Le volume tumoral macroscopique GTV (Gross Tumor Volume) est la partie de la maladie visualisable par toutes les données d'imagerie avec une densité tumorale la plus élevée. Le CTV (Clinical Target Volume) comprend le GTV ainsi que les extensions microscopiques de la tumeur sur les tissus environnants. Autour du CTV est défini le PTV (Planning Target Volume) avec une marge prenant en compte les mouvements du patient et les incertitudes de placement du patient (Fig. 1.10). Le PTV tient donc compte des erreurs systématiques (Σ) et des erreurs aléatoires (σ) et la marge CTV-PTV est calculée à l'aide de la formule de Van Herk [36].

$$M_{CTV-PTV} = 2.5\Sigma + 0.7\sigma \quad (1.3)$$

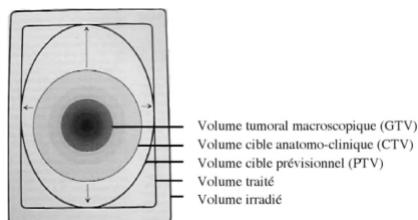


FIGURE 1.10 – Illustration schématique du principe de volumes définis par l'ICRU.

b) Planification des traitements

L'objectif de la planification de traitement est de générer pour chaque patient un plan optimal répondant aux exigences cliniques. Dans la pratique, il est recommandé que le PTV reçoive un minimum 95% de la dose prescrite sur 95% de son volume et au plus 107% de la dose tout en maintenant les doses aux organes sains à un niveau aussi bas que possible. L'optimisation d'un plan de traitement est réalisée selon trois techniques : 3D conformationnelle, l'IRMT (Intensity-modulated radiation therapy) et VMAT (Volumetric Modulated Arc Therapy). Le problème d'optimisation est résolu à l'aide d'un système de planification de traitement ou Treatment Planning System (TPS).

Actuellement, la réalisation d'un plan se fait encore manuellement prenant généralement du temps pour avoir un résultat cliniquement acceptable. Ces paramètres d'optimisation sont ajustés manuellement et la qualité du plan dépend fortement de l'expérience du dosimétriste. Il s'agit d'un processus d'essais et d'erreurs sur plusieurs itérations pour trouver le compromis entre le contrôle tumoral et les complications des organes sains environnants guidés par les objectifs cliniques [37]. Par conséquent, la réalisation des plans de traitement de façon automatisée basée sur les réseaux de neurones et les forêts aléatoires permettrait par exemple de réduire la durée de la planification. Ainsi, un modèle de réseau de neurone artificiel pour la prédiction de la distribution de dose en radiothérapie stéréotaxique a été proposé par Campbell et al.[38]. Les résultats ont démontré que cette méthode est prometteuse dans la distribution de la dose 3D. Nguyen et al.[39] ont développé une nouvelle application de l'architecture U-Net dans une étude de faisabilité pour la prédiction de la distribution de dose chez des patients atteints du cancer de la prostate et obtiennent une prédiction avec une meilleure distribution de dose. Kajikawa et al.[40] ont utilisé un réseau de neurone convolutif avec pour objectif de savoir si toutes les contraintes dosimétriques sont applicables pour un patient atteint du cancer de la prostate. En utilisant la méthode IMRT pour réaliser les plans de traitement, les résultats montrent que l'optimisation automatisée peut être appliquée pour le cas du cancer de la prostate.

Au fur et à mesure que la technique de planification améliore la qualité des plans de traitements, un programme d'assurance qualité systématique et complet doit être établi [37].

1.4.3 Assurance qualité patient

En radiothérapie, l'assurance qualité est un ensemble de procédures et d'actions structurées mis en œuvre pour évaluer les performances des machines. Cette mesure d'assurance qualité pour chaque patient est effectuée dans le but de vérifier l'exactitude de la délivrabilité du traitement. L'AQ patient ou Patient Specific Quality Assurance (PSQA) est réalisée au CHU de Liège uniquement sur des plans de traitement VMAT. Le processus de réalisation d'un PSQA est décrit au chapitre 2 paragraphe 2.2.5.

Au CHU de Liège, la technique de planification VMAT est toujours optimisée manuellement. Ce qui signifie que la planification automatisée n'est pas encore utilisée en routine clinique. L'intérêt de ce travail est d'examiner si l'introduction de la planification optimisée avec ML/RF pourrait contribuer à améliorer la qualité des plans de traitements ou si elle augmenterait la complexité des plans de traitement. L'objectif de ce travail est d'évaluer l'intégration de cette modalité de planification en routine clinique au CHU de Liège. Pour ce faire, nous allons réaliser des plans de traitement VMAT optimisés avec Machine Learning et les comparer aux mêmes plans VMAT optimisés manuellement. Par ailleurs, la qualité des plans de traitement sera évalués en terme d'indice de conformité et d'indice d'homogénéité de la dose au PTV, et la complexité des plans sera déterminée à l'aide des métriques.

Chapitre 2

Matériels et méthodes

2.1 Matériels

2.1.1 L'arcthérapie

Le Volumetric Modulated Arc Therapy (VMAT) est une technique de planification inverse de plus en plus utilisée en radiothérapie externe. Dans cette technique, les lames du MLC se déplacent pendant l'irradiation en fonction du temps et au même moment, le bras de l'accélérateur effectue des rotations autour du patient. Toutefois, la vitesse de rotation du bras de l'accélérateur est dépendant du nombre d'unités moniteurs délivrées entre chaque point de contrôle (CP). Les points de contrôle sont des points créés par discrétisation de l'arc correspondant aux différentes positions des lames du MLC et du nombre d'unité moniteur associé à chaque angle du bras de l'accélérateur linéaire. La dose délivrée est modulée à l'aide du MLC et du débit de dose (UM/min) pour tout CP.

2.1.2 L'accélérateur linéaire Elekta Infinity Agility

Un accélérateur linéaire est un dispositif utilisant des rayonnements électromagnétiques avec des fréquences de l'ordre de 3 GHz. Ces fréquences élevées permettent d'accélérer les électrons produits grâce à un filament chauffé et vont heurter une cible de tungstène située à l'entrée de la tête du LINAC (Fig 2.1) produisant un rayonnement de freinage (faisceau de photon) grâce à son numéro atomique effectif. Le faisceau créé est collimaté une première fois avec un collimateur primaire. Ce dernier est un matériau en tungstène qui peut avoir une ouverture conique ou pyramidale. Il y a par la suite un cône égalisateur placé directement après le collimateur primaire et a pour rôle de rendre le faisceau de photon homogène. Ce cône peut être retiré de la trajectoire du faisceau pour le traitement des tumeurs peu profondes et lorsqu'on veut traiter les lésions superficielles avec les électrons. Un système composé de deux chambres d'ionisations scellées est utilisé pour monitorer la dose délivrée par le faisceau en terme du nombre d'unités moniteurs et permet aussi de vérifier la symétrie du faisceau dans les deux directions. Ensuite, on a le collimateur secondaire encore appelé les mâchoires. Elles sont constituées de quatre blocs de tungstène disposés sur les deux plans (X et Y) . Ces mâchoires

sont utilisées pour définir la taille du champ, qui peut être carrée ou rectangulaire pouvant atteindre 40 cm de côté à l'isocentre. Enfin, il y a le collimateur Multilames dit aussi MLC est composé de plusieurs blocs de tungstène pouvant se déplacées indépendamment les unes des autres. Le MLC permet donc d'adapter le faisceau d'irradiation à la région d'intérêt à irradier.

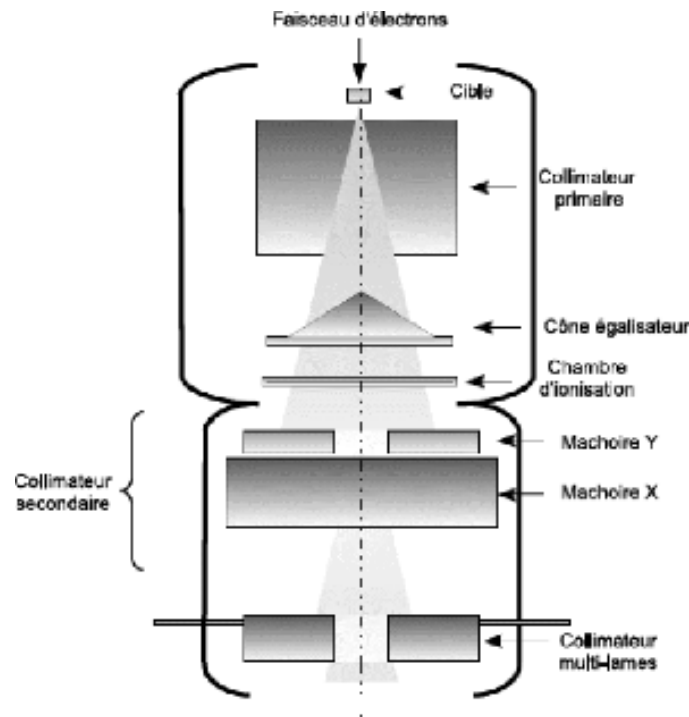


FIGURE 2.1 – Illustration de la tête de l'accélérateur Elekta Infinity Agility.

2.1.3 Description de RayStation 10A

RayStation est un TPS qui permet d'effectuer plusieurs tâches telles que le contournage des volumes et organes et le calcul de la distribution de dose. Dans sa version 10A, il dispose d'une autre fonctionnalité pour la réalisation des plans de traitement : le *Machine Learning* utilisant un algorithme appelé Random Forests (RF) tous deux décrits en détail dans le paragraphe 1.3.2 du chapitre 1 . Cet algorithme du calcul de dose prend également en compte l'effet de l'hétérogénéité du milieu. Plusieurs étapes entrent en jeu pour la création des plans pendant l'optimisation.

a) Optimisation avec l'algorithme Random Forest

Une fois les paramètres du plan introduit dans le TPS, le système génère automatiquement le plan (*Create ML plan*). Ensuite, on a la phase dite d'extraction des caractéristiques (*Extracting Features (EF)*) telle que la forme des objets dans l'image, les bords de l'image, le mouvement de l'image etc. L'EF est une étape de réduction de la dimensionnalité dans laquelle les données d'origine sont subdivisées puis réduites en sous-groupes. Cette réduction des données brutes

permet au TPS de construire un modèle avec moins d'effort. L'EF fait référence au nœud de l'arbre de décision. Une fois que l'extraction des caractéristiques est réalisée, le TPS va procéder au stockage de ces informations (*storing*) sous la forme d'une matrice de nombres dont la taille dépend du nombre de pixels initiaux. Après cette phase, l'algorithme crée un ensemble d'arbre représentant une forêt (*Loading forests*) sur lequel le calcul de la distribution des informations précédentes est effectué pour chaque nœud de l'arbre créé (*calculating node distribution for all forests*).

Par la suite, l'algorithme fait un calcul du coefficient de Bhattacharyya (*Computing Bhattacharyya coefficients*) qui permet d'évaluer le contraste entre l'image du modèle et celle étudiée. Ce coefficient mesure également la quantité de chevauchement entre deux échantillons, c'est-à-dire le chevauchement entre les plans du modèle proposé et les nouveaux plans. Le coefficient de Bhattacharyya est compris entre 0 à 1. L'algorithme RF va créer une extension appelée Adaptive Random Forest (ARF) puis, procédera à l'exécution de l'ARF sélectionné (*Performing ARF selection*) qui servira à gérer l'évolution de la classification de données (Fig 2.2) [41].

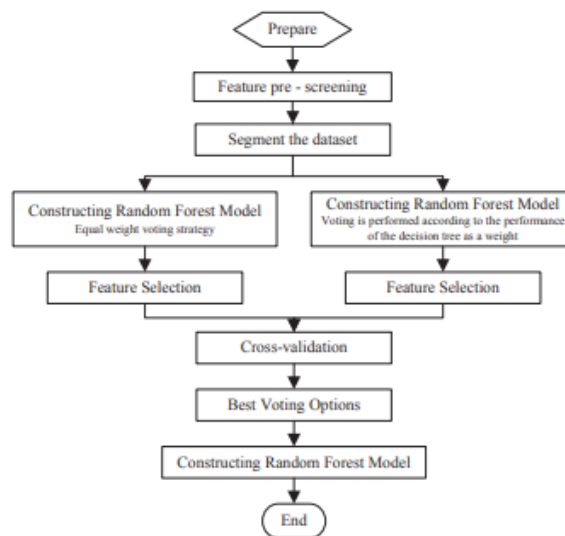


FIGURE 2.2 – Organigramme de l'algorithme ARF.

L'ARF fera donc une optimisation qui aboutira à la création du plan de traitement (*Running ARF inference*). L'étape dite *Setting up optimization problem* permet au système de générer les mimic-dose créant automatiquement des contraintes, puis l'optimisation est exécutée en deux phases (*Running ML optimization 1 et 2*).

Au cours de la planification, la distribution de la dose est visualisable à partir du DVH et des lignes isodoses. Cependant, il existe d'autres paramètres permettant d'évaluer la distribution de la dose au volume cible tels que l'indice de conformité et l'indice d'homogénéité.

2.1.4 Indices

a) Indice de conformité

Dans le TPS RayStation, l'indice de conformité est défini comme le rapport entre le volume cible couvert par l'isodose (TV_{RI}) et le volume total de l'isodose (V_{RI}) [42].

$$CI = \frac{TV_{RI}}{V_{RI}} \quad (2.1)$$

Cet indice est compris entre 0 et 1 où 1 représente une conformation optimale sans irradiation des tissus sains. Par contre, une valeur de 0 indique une absence totale de conformation. Cette situation se présente lorsque $TV_{RI} = 0$ avec $V_{RI} \neq 0$. Selon l'étude de Lomax [43], pour une irradiation conformationnelle, cet indice doit être supérieure ou égale à 0,6. Nous fixons dans le cadre de cette étude un seuil de 0,8. En général, cet indice aiderait à atteindre les objectifs cliniques pour la protection des tissus sains.

b) Indice d'homogénéité

L'indice d'homogénéité est défini dans le TPS RayStation comme le rapport entre la dose pour un volume de x et la dose pour un volume de $(100 - x)$ [42] tel que :

$$HI = \frac{D_x}{D_{(100-x)}} \quad (2.2)$$

Où D_x est la dose pour $x\%$ du volume cible. Cet indice est compris entre 0 et 1. Dans un volume cible, la dose est considérée comme parfaitement homogène si et seulement si son indice vaut 1. Dans le cas où cet indice vaut 0, on parle de la non homogénéité sur le volume considéré. De même, un seuil de 0,9 est fixé dans le cadre de cette étude pour $x = 98\%$.

2.1.5 Métriques

Dans la pratique, la radiothérapie a considérablement évolué allant de la radiothérapie 3D conventionnelle à l'arc thérapie à modulation volumétrique (VMAT) en passant par la radiothérapie à modulation d'intensité (IMRT). Le Volumetric Modulated Arc Therapy (VMAT) est une nouvelle forme d'optimisation de l'IMRT permettant de délivrer la dose de rayonnement avec une ou plusieurs rotations de la gantry à vitesse variable combiné à un collimateur MLC dynamique avec des débits de dose parfois constants. Le cas VMAT avec des débits de dose

variables offrent une meilleure délivrance de la dose que le cas avec des débits de dose constants et améliorent la conformité de la dose au volume cible tout en limitant au mieux les toxicités aux tissus sains environnants [44]. Le VMAT possède plusieurs avantages tels que la réduction importante du temps d'administration du traitement.

Ces caractéristiques pertinentes du VMAT sont parfois associées à l'augmentation de la complexité des plans de traitements et aussi en partie liée à l'utilisation de l'indice de conformité et l'indice d'homogénéité. Plus la conformité augmente, plus la complexité de la technique VMAT augmente car les points de contrôles (CP) présentent différentes formes de MLC comme en IMRT mais avec des champs plus petits. Dans sa généralité, la complexité d'un plan VMAT dépend de la morphologie du patient, des contraintes et objectifs dosimétriques, des caractéristiques de l'accélérateur linéaire (LINAC) et de l'algorithme utilisé dans le TPS.

Afin d'évaluer la complexité d'un plan, plusieurs métriques ont été proposées et classées en trois catégories [45]

- **Les métriques de fluence** : utilisées dans le but de quantifier la complexité avec pour hypothèse qu'une fluence très hétérogène reflète une grande complexité.
- **Les métriques de délivrabilité** : évaluent les aptitudes du linac à administrer le traitement comme le prévoit le calcul du TPS. Ces métriques sont liées aux paramètres mécaniques (gantry, MLC), dosimétriques (débit de dose, unités moniteurs) et dépendent de la méthode utilisée lors de la planification du traitement.
- **Les métriques de précision** : visent à quantifier la précision du calcul de la dose réalisé par le TPS en se focalisant uniquement sur la forme du MLC. L'imprécision peut également être liée à la modélisation de la machine et à l'erreur introduite par l'algorithme utilisé dans le TPS.

Les métriques de précision et de délivrabilité sont étroitement liées [45] et au CHU de Liège, seules ces deux métriques sont utilisées pour évaluer la qualité du plan calculé par le TPS.

a) Métrique de délivrabilité : MCS_V

Modulation Complexity Score (MCS) est la métrique de délivrabilité initialement proposée par McNiven et al.[46] pour la technique IMRT. Selon ces auteurs, MCS est composé de deux paramètres : le Leaf Sequence Variability (LSV) utilisé pour caractériser la variabilité de la forme des segments pour un plan spécifique et le Aperture Area Variability (AAV) qui caractérise la variation de surface de segment par rapport à l'ouverture maximale définie par l'ensemble des segments. Par définition, le MCS une somme normalisée sur tous les segments du LSV et du AAV pondérées par les unités moniteurs (UM) délivrées. Cette métrique prend en compte toutes les informations du TPS telles que l'irrégularité de la forme du champ par rapport à la position des lames adjacentes, le poids des segments et la variation de la taille de champ par rapport à une taille de champ maximale. Elle varie entre 0 pour un plan très modulé et 1 pour un plan sans modulation.

Cette métrique a été modifiée par Masi et al [47] pour être appliquée à un plan VMAT en considérant les points de contrôles ou Control Point (CP) de l'arc au lieu des segments et notée MCS_V . Un point de contrôle est par définition la discrétisation de l'arc. Le paramètre AAV se réfère ici à l'ouverture entre les lames opposées pour chaque point de contrôle. Pour la technique VMAT, les unités moniteurs varient entre les points de contrôles avoisinants. Le MCS_V pour des plans de traitement VMAT est défini comme suit :

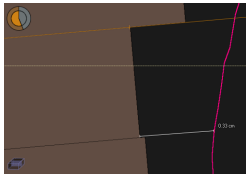
$$MCS_V = \sum_{i=1}^{I-1} \left(\frac{AAV_{cp_i} + AAV_{cp_{i+1}}}{2} \right) \times \left(\frac{LSV_{cp_i} + LSV_{cp_{i+1}}}{2} \right) \times \frac{MU_{cp_{i,i+1}}}{MU_{arc}} \quad (2.3)$$

Où $MU_{cp_{i,i+1}}$ représentent les unités moniteurs délivrées entre deux points de contrôle successifs cp_i et cp_{i+1} ou $MU_{cp_{i,i+1}} < MU_{arc}$. Comme dans la définition d'origine, le MCS_V a des valeurs comprises entre 0 et 1. Une valeur du $MCS_V = 1$ montre qu'il n'y a pas de modulation du faisceau. Ceci stipule que lors de la rotation du gantry, les lames sont statiques pour une ouverture rectangulaire fixe par exemple [47]. Par contre, avec la même ouverture balayant le champ pendant la rotation du gantry, on aura comme valeur $MCS_V < 1$ due au fait que $LSV=1$ et $AAV < 1$ [47].

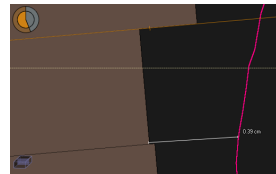
b) Métrique de précision : LOIC

Le Leaf Offset Impact on Calculation (LOIC) a été introduit par Mathot et al [48] et vise à quantifier l'influence des petits offsets du MLC. Le LOIC est une variabilité de la dose moyenne au PTV lorsqu'il existe un offset des lames par rapport à la position initiale. Dans le TPS RayStation, la position des lames est calculée selon l'équation 2.4 ; où x_c définit la position de la lame utilisée dans le calcul de la dose et x^t représente la position nominale de la lame définie pendant le traitement. Le *gain*, *curvature* et *offset* sont des paramètres d'étalonnage des collimateurs. Ils sont utilisés pour moduler la position des lames dans le TPS lors du calcul de la distribution de dose [42].

$$x_c = x^t + offset + x^t \times gain + x^t \times curvature \quad (2.4)$$



(a) Sans offset



(b) Avec offset

FIGURE 2.3 – Illustration de deux plans de traitement

Dans le service de radiothérapie du CHU de Liège, le *LOIC* est utilisé pour réduire la charge du travail de l'assurance qualité (QA) patient spécifique. Lorsque, pour un plan de traitement réalisé, on a $LOIC \leq 1.25\%$, un QA n'est pas nécessaire. Ce seuil de sensibilité totale a été déterminé dans une étude menée par Mathot et al [48] car il ne fournit aucun faux négatif. Par contre pour une valeur plus élevée ($LOIC > 3\%$), on réalise une nouvelle optimisation du plan dans le but de réduire la complexité. Dans le cas où cette complexité persiste après une replanification, une mesure très stricte est réalisée par la physique. Enfin, si la métrique est comprise entre les deux seuils, un QA patient spécifique est obligatoire.

2.1.6 Optimisation de la charge du travail en PSQA

L'évaluation de la qualité du plan en routine clinique se fait pour des plans de traitement qui n'ont pas encore été délivrés. Au CHU de Liège, ce contrôle est effectué en utilisant un fantôme nommé ArcCHECK (*Sun Nuclear* corporation, Melbourne, FL, USA). Le but est de comparer les distributions de la dose calculée dans le TPS importée au format Dicom (RTdose) à celle mesurée sur l'ArcCHECK (RTPLANS) en utilisant l'indice γ . Cet indice γ est un paramètre composite qui combine à la fois la différence entre les doses mesurée et calculée (Δd) et la Distance To Agreement (DTA) en une seule quantité. Ce calcul est effectué pour chaque point de la distribution de dose et l'écart entre la dose mesurée et calculée en un point ne doit pas excéder la valeur de 3%. Selon le TG 218 [49], le critère de validité de la DTA (Δd) à dose égale entre les points est fixé à 2 mm. L'indice γ est calculé comme suite :

$$\Gamma(r_e, r_r) = \sqrt{\frac{r(r_e, r_r)^2}{(\Delta d)^2} + \frac{\delta(r_e, r_r)^2}{(\Delta D)^2}} \quad (2.5)$$

$$\gamma = \min[\Gamma(r_e, r_r)] \quad (2.6)$$

Avec $r(r_e, r_r)$ la distance entre les points de référence (calculé) du TPS et évalués (mesuré) sur l'ArcCHECK, et $\delta(r_e, r_r)$ la différence de dose. Un plan de traitement est acceptable si et seulement si au moins 95% des points ont un indice γ inférieur à 1 en 3%,2mm selon le TG 218.

L'indice γ peut être calculé par normalisation locale ou globale. Autrement dit, l'indice γ local est normalisé sur la dose par pixel par contre l'indice γ global est normalisé à la dose maximale du plan de traitement. Compte tenu de ces différences, le service de radiothérapie du CHU de Liège utilise en routine clinique l'indice γ global pour évaluer la qualité du plan de traitement en appliquant le critère 3%,2 mm recommandé par le TG 218, avec un niveau de dose seuil de 10%.

2.2 Méthodes

2.2.1 Les patients

Pour cette étude, nous avons sélectionné dans la base de données (*MOSAIQ*) du service de radiothérapie du CHU de Liège les plans des patients atteints du cancer de la prostate. Ces derniers ont été traités avec des plans VMAT conventionnels (manuellement) réalisés dans le TPS RayStation 7.0. Au total, 34 plans de traitement ont été sélectionnés pour cette étude et regroupés en trois groupes :

- Groupe 1 : 22 plans de traitement de prostate simple avec une prescription de 6000 centi-Gray (cGy) en 20 fractions soit 300 cGy par fraction.
- Groupe 2 : 11 plans de traitement prostate simple avec une prescription de 6575 cGy en 25 fractions soit 263 cGy par fraction.
- Groupe 3 : 1 plan de traitement simple avec une prescription de 6875 cGy en 25 fractions soit 275 cGy pour chaque fraction.

La prescription et le nombre de fraction du groupe 1 sont ceux utilisés dans le modèle proposé par Raysearch. La prescription et le fractionnement du groupe 2 et 3 utilisés dans ce travail serviront à évaluer les performances du modèle sur les fractionnements utilisés dans le service. Nous avons intégré des plans initiaux ou plans de référence avec un arc et deux arcs. Par contre, les plans de traitement contenant les prothèses ont été exclus de notre étude.

2.2.2 Configuration du TPS pour la planification

Dans le TPS RayStation, il est possible de définir deux arcs pour les traitements en VMAT. Pour une seule rotation, la gantry fait une rotation de 179° à 182° dans le sens trigonométrique ou *Counterclockwise* et si une deuxième rotation est nécessaire, on utilise l'option *Oclockwise* c'est-à-dire de 182° à 179° . Dans le cadre de notre étude, nous réalisons des plans de traitement VMAT optimisés avec ML en utilisant un arc dans le sens *Counterclockwise*. L'énergie utilisée est de 6 MV avec un collimateur à 30° et l'espacement maximal entre les points de contrôles adjacents est de 3° . Chaque point de contrôle est un faisceau statique avec un seul segment c'est-à-dire une forme d'ouverture. Ces paramètres sont également ceux utilisés pour la réalisation des plans de références (Fig. 2.4).

No.	Name	Description	Isocenter [cm]				SSD [cm]		Machine	Energy [MV]	Gantry start [deg]	Gantry stop [deg]	Rotation	Coll. [deg]	Couch [deg]
			Name	R-L	I-S	P-A	To surface	To skin							
1	1r179	1r179	Photon 1	0.89	-81.66	-9.63	84.12	89.30	Test10A_2020	6	179.0	182.0	Counterclockwise	30.0	0.0

FIGURE 2.4 – Illustration de caractéristiques de la planification.

En effet, l'utilisation de 3° entre les CP au lieu d'une valeur plus grande permet d'augmenter la précision du calcul de la dose, principalement pour des plans qui nécessitent un grand déplacement de la lame du MLC [47]. Par contre, la durée d'optimisation nécessaire pour la

planification du traitement augmente quand l'écart entre les CP diminue.

De plus, les lames du MLC changent de forme continuellement entre les CP tout au long de la rotation du gantry. Par conséquent, une contrainte sur les mouvements des lames est importante en raison de son influence sur la qualité du plan. La contrainte sur les mouvements des lames du MLC est de 3 mm/degré de rotation du gantry et utilisée sur l'ensemble des plans de traitement. Cette contrainte permet d'éviter les grands mouvements du MLC entre les CP adjacents.

2.2.3 Création d'un nouveau plan avec Machine Learning

Pour tout plan, le module *Optimization* de Machine Learning permet de créer un nouveau plan de traitement et cela ouvre une boîte de dialogue dite *New Machine Learning plan*. Parmi les modèles du plan qui figurent dans le TPS, le modèle sélectionné dans la première liste déroulante est *PMH Prostate 10A* (Fig. 2.4). Les paramètres de configuration de ce dernier sont définis lors de la création du modèle par RaySearch et ne sont pas modifiables.

Les caractéristiques utilisées sont communes à tous les plans et sont définies ci-après et illustrés sur la figure 2.5 :

- Utilisation de 1 arc (179° - 182°)
- Isocentre placé dans le PTV
- Le nombre d'UM limitée à 500
- Introduction de la dose prescrite et le nombre de fraction

2.2.4 Optimisation des plans de traitement

L'optimisation des plans est réalisée dans le TPS en essayant d'avoir une bonne conformité et homogénéité de la dose au volume cible tout en gardant une bonne couverture. Le nombre d'itérations a été fixé à 25. Pendant la phase d'optimisation, le TPS crée automatiquement des contraintes afin de respecter les objectifs cliniques. Pour améliorer la qualité du plan de traitement, la fonction dite *Dose fall-off* est utilisée en mode contrainte sur le volume cible.

Le *Dose fall-Off* est une fonction qui permet de réduire la dose d'un organe à l'autre sur une distance. Il est principalement utilisé sur les contours externes et permet de réduire considérablement la dose autour du PTV pour une meilleure conformation de la dose dans ce dernier. Autrement dit, Le *Dose fall-off* demande qu'un certain niveau de dose diminue jusqu'à atteindre un niveau de dose plus faible sur une certaine distance (Fig. 2.6).

Lorsqu'un organe à risque reçoit une dose au-dessus du seuil fixé dans les objectifs cliniques pendant l'optimisation, on place des contraintes sur ce dernier tout en essayant de ne pas

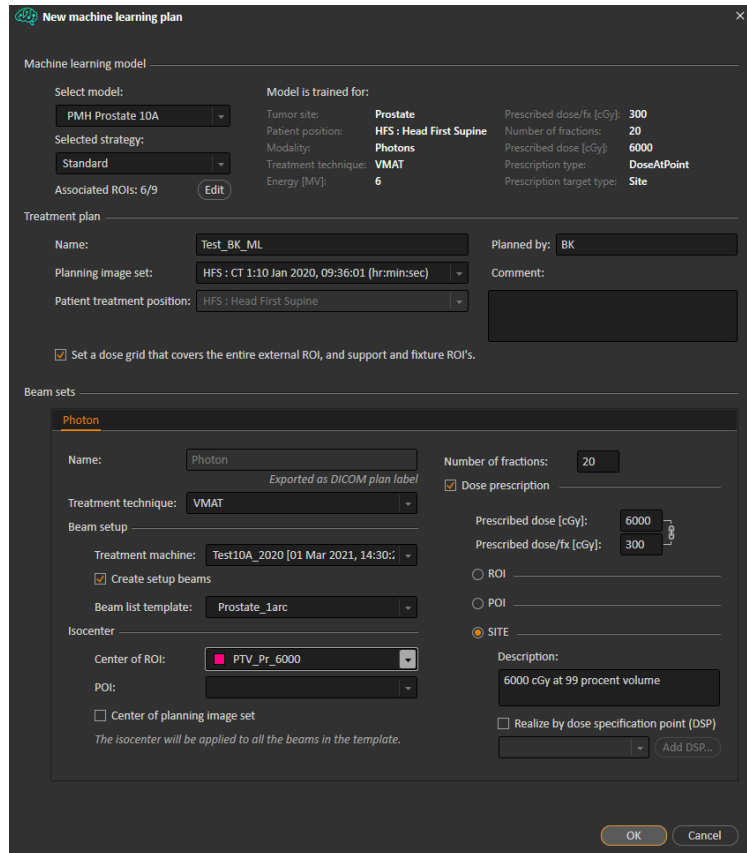


FIGURE 2.5 – Interface de configuration d'un plan dans le TPS.

dégrader la couverture et la conformité du PTV.

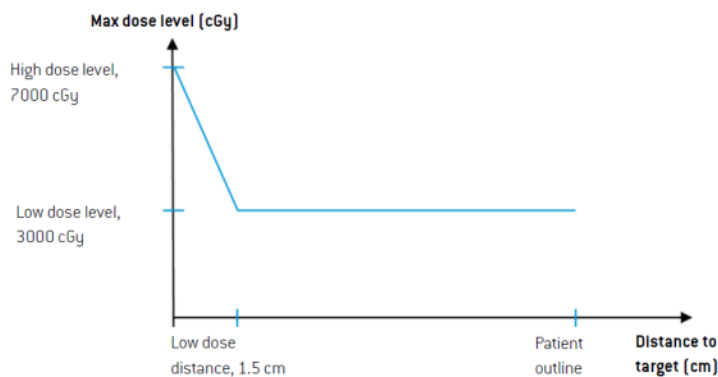


FIGURE 2.6 – Illustration de la fonction *Dose fall-off* autour du PTV. On demande ici au système de passer de 70 Gy à 30 Gy sur 1,5 cm.

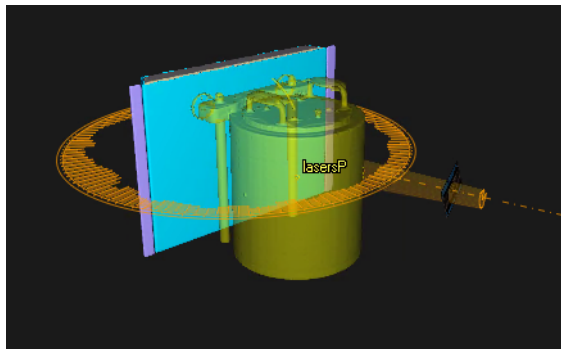
À la fin de l'optimisation de chaque plan, nous relevons les valeurs des doses moyennes aux différents OAR ainsi que les valeurs de l'indice de conformité et de l'indice d'homogénéité. De même, la couverture du PTV et la $D_{2\%}$ pour chaque OAR sont également notés. Par définition, la $D_{2\%}$ est la dose reçue par 2% du volume considéré. Les unités moniteurs sont aussi notés

ainsi que les valeurs calculées du $LOIC$ et du MCS_V . L'ensemble des valeurs ainsi obtenues serviront de point de comparaison avec les valeurs de référence.

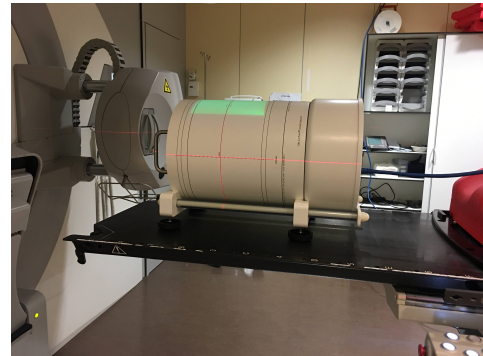
2.2.5 Mesure avec l'ArcCHECK

L'équipement utilisé pour évaluer la qualité du plan dans notre étude est l'ArcCHECK (*Sun Nuclear*, Melbourne, FL, USA). Il s'agit d'un fantôme cylindrique avec 1386 détecteurs à diodes en spirale réparties sur toute sa surface, avec un espacement de 10 mm.

La première étape du QA patient est le calcul du plan du patient copié sur l'ArcCHECK dans le TPS RayStation (Fig. 2.7a). Cette simulation permet d'obtenir les informations sur les distributions calculées de la dose. La deuxième étape consiste à placer le fantôme sur la table de traitement et aligné sur les lasers à l'isocentre du LINAC et à délivrer le plan de traitement (Fig. 2.7b). Ce dispositif permet d'obtenir les informations sur les distributions de la dose mesurée à l'aide du fantôme. Ces deux étapes sont réalisées pour chaque patient et permettront de calculer individuellement l'indice γ global. Le calcul de l'indice γ est effectué comme décrit dans le paragraphe 2.1.6.



(a) Simulation virtuelle



(b) Cas pratique

FIGURE 2.7 – Positionnement de l'ArcCHECK centré aux lasers à l'isocentre du LINAC

2.2.6 Analyse statistique : Test de Wilcoxon

Le test des rangs signés de Wilcoxon est un test non paramétrique utilisé pour la comparaison de deux groupes d'échantillons appariés. Le mot *non paramétrique* signifie que les données n'ont pas une distribution normale. Il s'agit d'un test de différences par paires. La statistique est testée sous deux hypothèses :

- **L'hypothèse H_0 est acceptée** stipule qu'on ne peut pas affirmer que les résultats d'analyse sont identiques. Autrement, plus les données servant d'analyse augmentent, plus la précision dans l'interprétation augmente et plus le risque de se tromper diminue.

- **L'hypothèse H1 est acceptée** indique que les échantillons présentent une différence statistiquement significative. Ceci permet de confirmer avec certitude que les échantillons utilisés pour l'analyse présentent une différence.

Avec ce test, on peut dire que les *CI* et *HI* sont différents mais ne donne pas d'information sur le fait que le plan est amélioré. Le niveau de signification statistique est exprimé sous la forme d'une valeur p comprise entre 0 et 1. Pour cela, on compare la valeur de p au seuil de signification qui est généralement noté α avec une valeur de 0,05. Ce seuil indique un risque de 5% que la différence soit due au hasard.

- Si $p < \alpha$, l'hypothèse nulle (H_0) est rejetée. Cette dernière stipule qu'il n'y a pas de relation entre les échantillons étudiés c'est-à-dire qu'une variable n'affecte pas l'autre. Il indique que la différence dans les résultats n'est pas due au hasard et on peut conclure que la différence entre les moyennes où les médianes des deux modalités de planification est statistiquement significative.
- Si $p > \alpha$, on ne rejette pas l'hypothèse nulle. L'hypothèse alternative indique que la différence dans les valeurs obtenues n'est pas due au hasard. Ainsi, la différence entre les moyennes ou les médianes des deux modalités de planification n'est pas statistiquement significative.

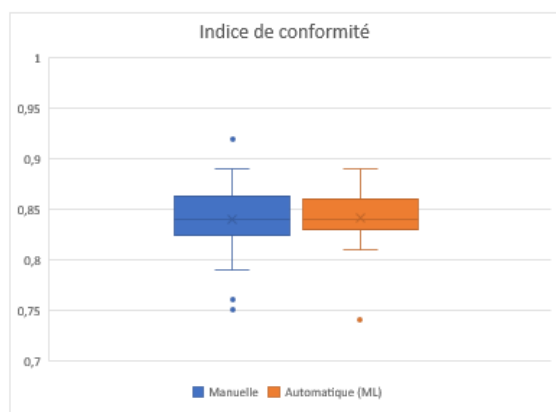
Chapitre 3

Résultats

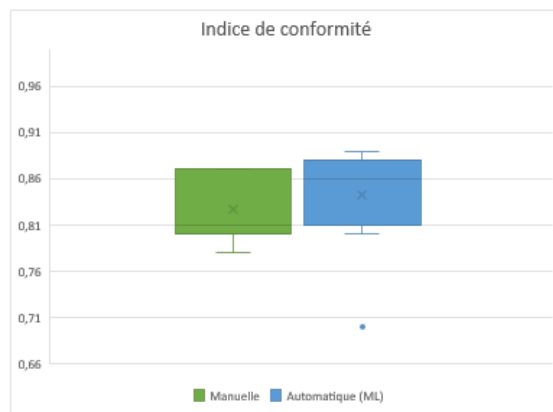
3.1 Indice de conformité (CI)

la figure 3.1a montre pour le groupe 1 que la valeur moyenne associée à l'écart type de l'indice de conformité pour la planification manuelle est de $0,840 \pm 0,039$ tandis que pour la planification automatique, elle est de $0,842 \pm 0,031$. Le test de Wilcoxon donne une valeur de $p = 0,952$. Pour le groupe 2 représenté par la figure 3.1b, la valeur moyenne de $0,826 \pm 0,0326$ pour la planification manuelle et de $0,839 \pm 0,066$ pour la planification automatique. Le test de Wilcoxon indique que les résultats ne sont pas statistiquement significatifs ($p = 0.119$). Pour le seul patient du groupe 3, le tableau 3.1 montre les valeurs pour les deux modalités obtenues.

Sur l'ensemble des Box plot, La boîte rectangulaire contient les deux quartiles du milieu, la ligne horizontale interne à la boîte représente la médiane. La croix (\times) indique la moyenne et les lignes extérieures à la boîte contiennent les quartiles supérieur et inférieur. Les valeurs abberantes sont représentées par les points plus externes.



(a) Groupe 1 (20 x 300 cGy)



(b) Groupe 2 (25 x 263 cGy)

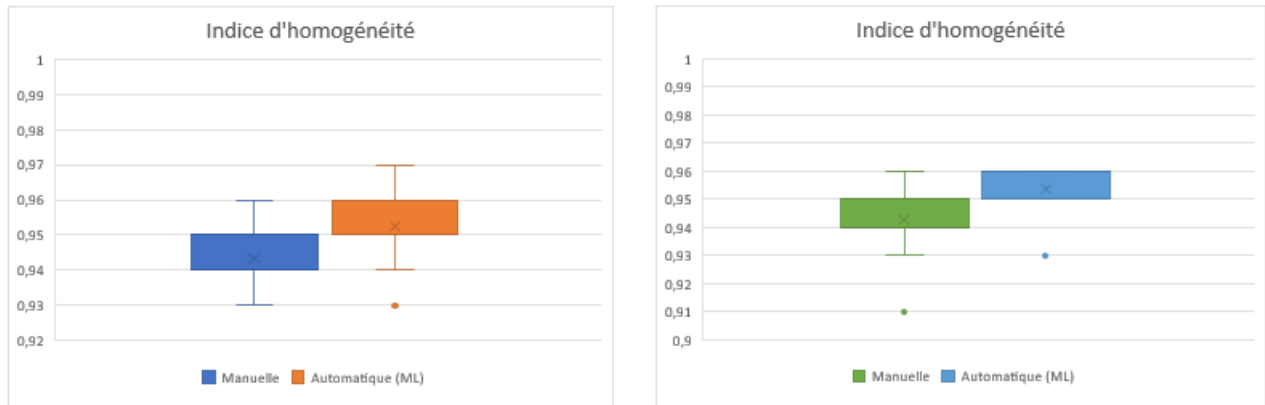
FIGURE 3.1 – Box plot de l'indice de conformité des plans.

Patient	CI manuel	CI automatique
C1	0,78	0,8

TABLE 3.1 – Indice de conformité du patient du groupe 3 (25×275 cGy)

3.2 Indice d’homogénéité (HI)

L’indice d’homogénéité du groupe 1 est représenté par la figure 3.2a et indique que la moyenne pour la planification manuelle est de $0,944 \pm 0,01$ et de $0,954 \pm 0,01$ pour la planification automatisée. L’évaluation statistique de la qualité des plans indique que les deux modalités de planification des traitements sont significativement différentes ($p = 0.015$). La figure 3.2b indique que la planification manuelle présente un indice d’homogénéité moyen de $0,943 \pm 0,014$ et pour la planification automatisée de $0,954 \pm 0,009$. Selon le test de Wilcoxon, la différence entre la planification manuelle et celle automatique n’est pas statistiquement significatif ($p = 0.085$). Les valeurs de l’indice d’homogénéité du seul patient du groupe 3 est explicitée dans le tableau 3.2.



(a) Groupe 1 (20×300 cGy)

(b) Groupe 2 (25×263 cGy)

FIGURE 3.2 – Box plot de l’indice d’homogénéité HI des plans.

Patient	HI manuel	HI automatique
C1	0,95	0,96

TABLE 3.2 – Indice d’homogénéité du patient du groupe 3 (25×275 cGy)

3.3 Couverture du PTV des plans

Tous les plans de cette étude ont atteint la couverture requise au PTV qui est d’avoir au moins 95% de la dose prescrite au volume cible tumoral. Pour le groupe 1 de la figure 3.3, le volume cible a reçu en moyenne 98,9% ($\sigma = 0,6\%$) de la dose prescrite dans la planification

manuelle et 98,2% ($\sigma = 0.8\%$) de la dose prescrite pour l'auto-planning. De même, les PTV du groupe 2 de la figure 3.4 reçoivent en moyenne 98,1% ($\sigma = 0,97\%$) et 97,69% ($\sigma = 1,05\%$) de la dose prescrite respectivement pour les planifications manuelles et automatisées avec Machine Learning. Le test de Wilcoxon indique que la différence entre les deux modalités de planification est significative pour le groupe 1 ($p = 0.0015$) et non significative pour le groupe 2 ($p = 0.24$). Les valeurs de la couverture au PTV pour le groupe 3 sont resumées dans le tableau 3.3 pour les deux modalités de planification des traitements.

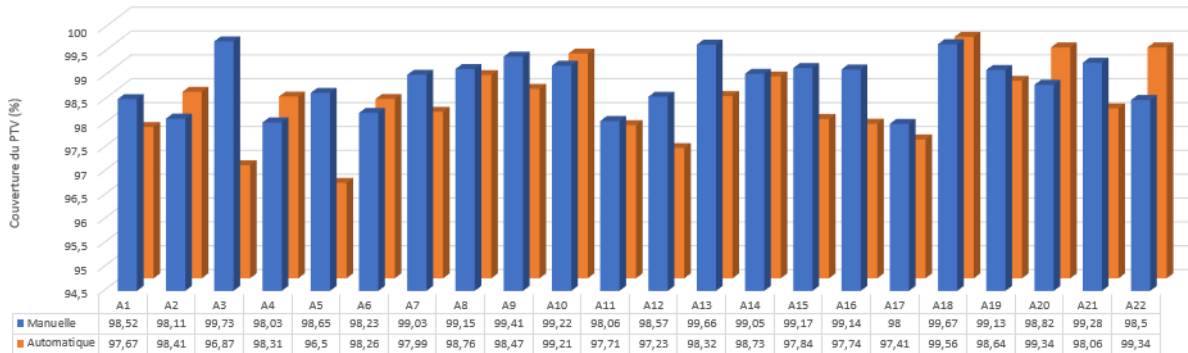


FIGURE 3.3 – Couverture du PTV du groupe 1 (20×300 cGy).

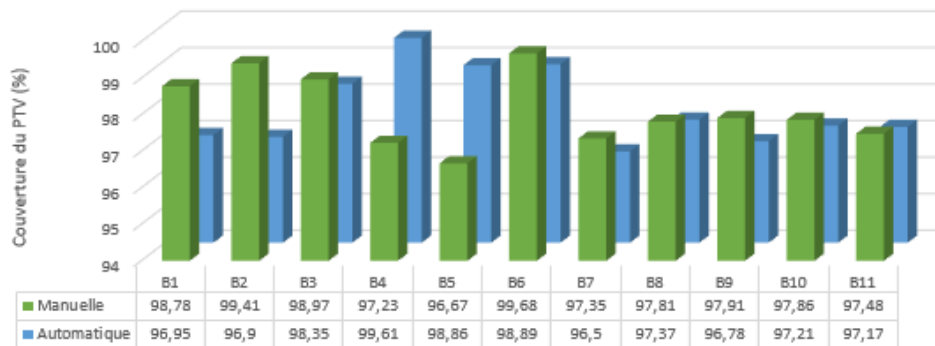


FIGURE 3.4 – Couverture du PTV du groupe 2 (25×263 cGy).

Patient	$V_{95\%}$ manuelle	$V_{95\%}$ automatique
C1	98,91%	97,55%

TABLE 3.3 – Couverture du PTV pour le groupe 3 (25×275 cGy)

3.4 Evaluation de la dose aux organes à risque

Pour éviter des complications ou des cancers radio-induits, les objectifs cliniques ont été fixés sur les organes à risque au CHU de Liège.

3.4.1 Moyenne des doses sur chaque OAR

Les valeurs de la dose moyenne de chaque organe à risque pour les 3 groupes et pour les deux modalités de planification ont été relevés dans les TPS. Ensuite, un calcul de la moyenne de ces doses a été réalisé sur chaque OAR pour le groupe 1 (Tab 3.4) et le groupe 2 (Tab 3.5). Le groupe 3 ne contenant qu'un seul patient (C1), la figure 3.5 montre la dose moyenne reçue par chaque organe à risque.

OAR	Planification manuelle		Planification automatisée		p-value
	Dose moyenne (cGy)	σ (cGy)	Dose moyenne (cGy)	σ (cGy)	
CanalAnal	2744,09	712,78	2979,73	848,67	0,08
Bladder	2459,05	719,91	2499,82	712,68	0,48
FemoralHead L	1486,18	417,97	1539,32	355,2	0,41
FemoralHead R	1424,32	410,06	1394,59	329,11	0,81
Rectum	3539,55	686,65	3061,32	725,33	0,00015
Sigmoid	442,64	360,77	441,27	388,78	0,33

TABLE 3.4 – Moyenne des doses calculées aux OAR des patients du groupe 1 (20×300 Gy)

OAR	Planification manuelle		Planification automatisée		p-value
	Dose moyenne (cGy)	σ (cGy)	Dose moyenne (cGy)	σ (cGy)	
CanalAnal	2964	770,14	3159,44	962,38	0,426
Bladder	3440,18	982,86	3272,55	1055,19	0,067
FemoralHead L	1600,91	583,98	2268,73	499,65	0.00098
FemoralHead R	1366,36	448,22	1955	504,39	0,002
Rectum	4084	360,47	3969,82	939,13	0,147
Sigmoid	772,78	711,602	953,78	914,91	1

TABLE 3.5 – Moyenne des doses calculées aux OAR des patients du groupe 2 (25×263 cGy)

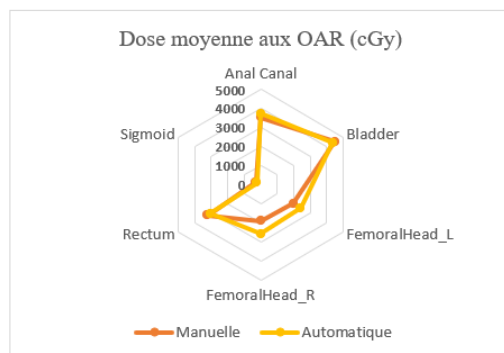


FIGURE 3.5 – Dose moyenne (cGy) aux OAR pour le groupe 3 (25×275 cGy)

3.4.2 Moyenne des doses reçues par 2% du volume des OAR

Tout comme dans le cas précédent, la $D_{2\%}$ de tous les organes à risque est relevée dans les TPS pour les deux modalités de planification. Une moyenne est calculée pour chacun d'entre eux avec les écarts types associés et récapitulés dans le tableau 3.6 pour le groupe 1 et le tableau 3.7 pour le groupe 2. Le cas du seul patient du groupe 3 est représenté sur la figure 3.6.

OAR	Planification manuelle		Planification automatisée		p-value
	Dose moyenne (cGy)	σ (cGy)	Dose moyenne (cGy)	σ (cGy)	
CanalAnal	5827,64	396,95	5826,46	322,51	0,262
Bladder	6045,59	113,75	5948,27	94,29	0.00058
FemoralHead L	2373,27	470,12	2521,46	413,58	0,408
FemoralHead R	2354,18	441,78	2334,18	359,44	0,849
Rectum	5966,09	170,87	5841,91	175,4	$4,8 \times 10^{-6}$
Sigmoid	1728,14	1834,97	1897,18	1963,94	0,949

TABLE 3.6 – Moyenne des $D_{2\%}$ aux OAR du groupe 1 (20×300 cGy)

OAR	Planification manuelle		Planification automatisée		p-value
	Dose moyenne (cGy)	σ (cGy)	Dose moyenne (cGy)	σ (cGy)	
CanalAnal	6235,22	558,02	6271	371,75	0,82
Bladder	6704,64	118,41	6525,18	77,66	0,00098
FemoralHead L	2809,36	561,33	3805	361,53	0,00098
FemoralHead R	2498,18	552,18	3397,09	379,81	0,00098
Rectum	6642,27	69,99	6535,91	74,88	0,0049
Sigmoid	2551,33	2348,03	2677	2312,85	0,734

TABLE 3.7 – Moyenne des $D_{2\%}$ aux OAR du groupe 2 (25×263 cGy)

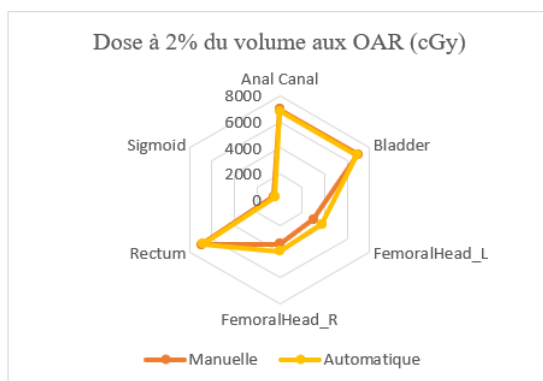
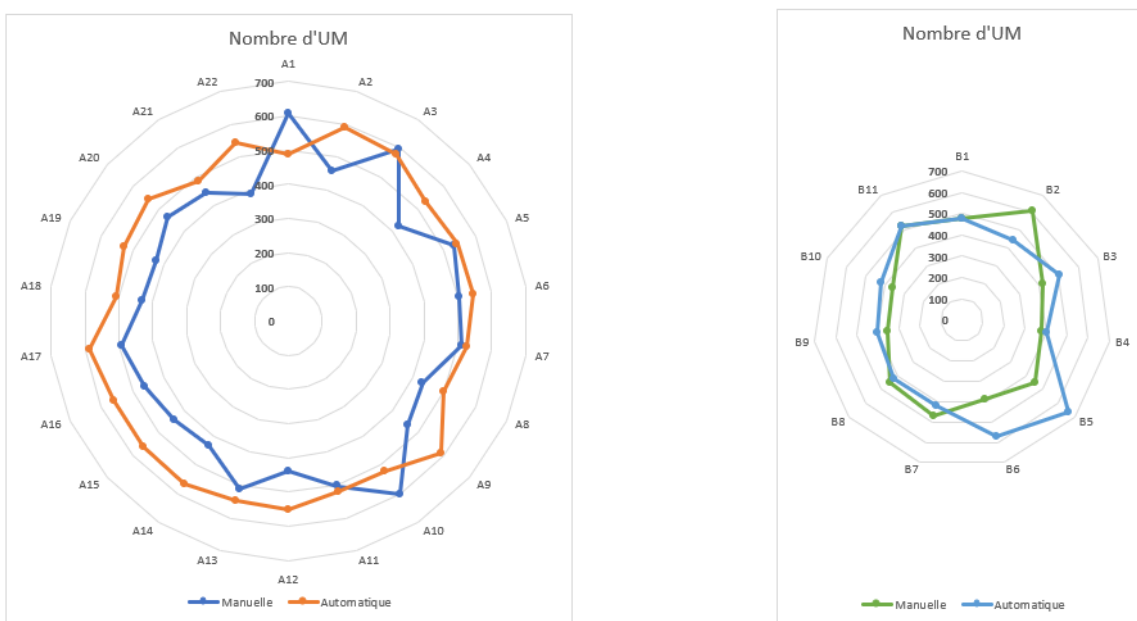


FIGURE 3.6 – $D_{2\%}$ (cGy) aux OAR pour le groupe 3 (25×275 cGy)

3.5 Nombre d'Unités Moniteurs des plans

Les unités moniteurs du groupe 1 sont représentées par la figure 3.7a. La valeur moyenne obtenue pour la planification manuelle est de $480,5 \pm 61,2$ et pour la planification automatique de $541,9 \pm 31,1$. Les résultats obtenus pour les plans du groupe 2 sont représentés sur la figure 3.7b. Ces résultats indiquent que les valeurs moyennes du nombre d'unités moniteurs sont de $443,6 \pm 76,8$ et $478,1 \pm 82,3$ respectivement pour la planification manuelle et la planification automatique. Les valeurs du groupe 3 pour les deux modalités de planification sont détaillées dans le tableau 3.8. Le test de Wilcoxon donne une valeur de $p = 0,00126$ pour les résultats du groupe 1 et de $p = 0,1536$ pour les résultats du groupes 2.



(a) Groupe 1 (20×300 cGy)

(b) Groupe 2 (25×263 cGy)

FIGURE 3.7 – Nombre d'UM pour les plans de traitement.

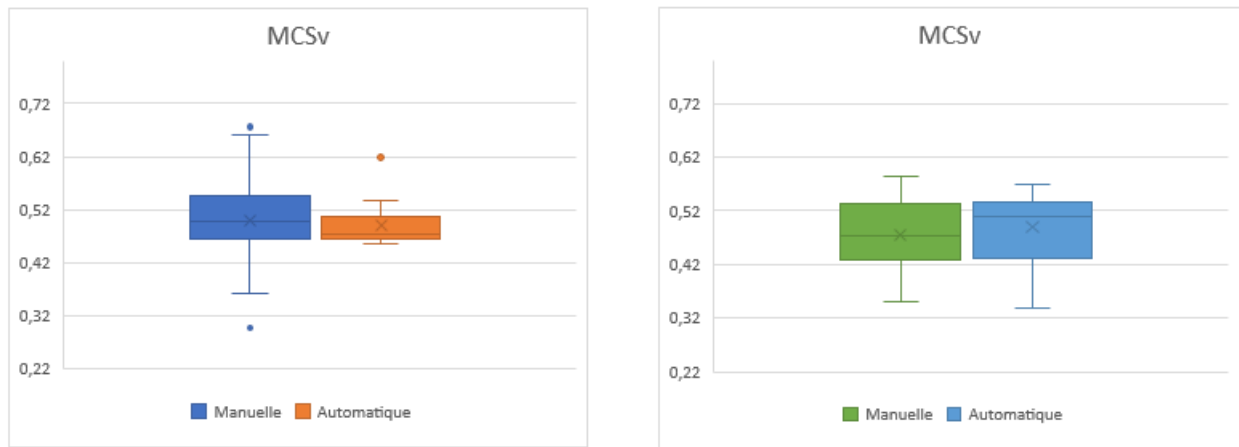
Patient	Planification manuelle	Planification automatisée
C1	500,1	544,35

TABLE 3.8 – Nombre d'UM du seul plan du groupe 3 (25×275 cGy)

3.6 Métrique de délivrabilité : MCS_V

La figure 3.8 montre les résultats de la métrique MCS_V obtenus. Les valeurs moyennes pour le groupe 1 (Fig 3.8a) est de $0,499 \pm 0,085$ pour la planification manuelle et de $0,491 \pm 0,037$ pour la planification automatique. De même sur la figure 3.8b du groupe 2, la valeur moyenne

pour le cas manuel est de $0,474 \pm 0,069$ et de $0,489 \pm 0,068$ pour l'auto-planning. L'évaluation statistique avec le test de Wilcoxon indique pour le groupe 1 une valeur de $p = 0,4222$ et une valeur de $p = 0,6103$ pour le groupe 2. Le tableau 3.9 présente les valeurs obtenues du groupe 3 pour les deux modalités de planification.



(a) Groupe 1 (20×300 cGy)

(b) Groupe 2 (25×263 cGy)

FIGURE 3.8 – Box plot des métriques MCS_V des plans de traitement.

Patient	MCS_V manuelle	MCS_V automatique
C1	0,515	0,52

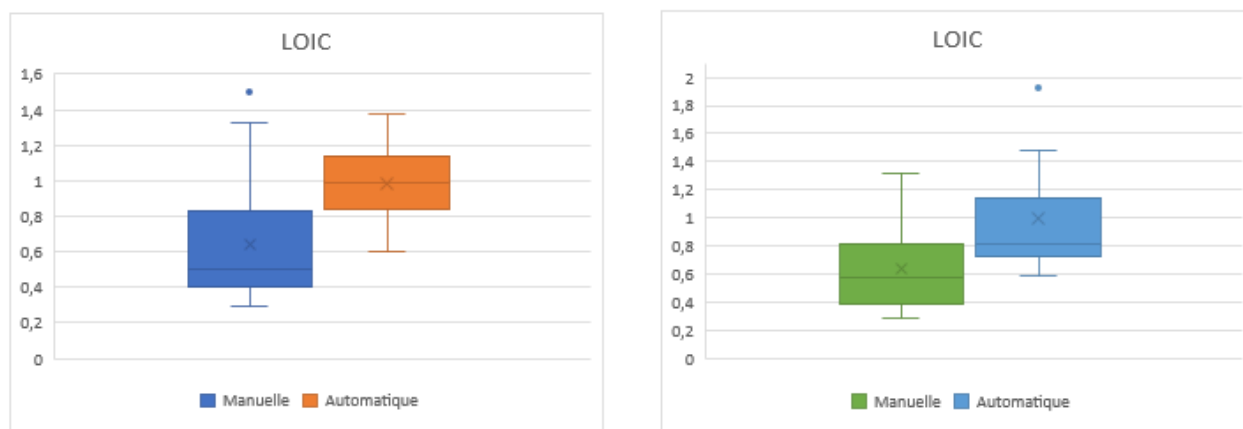
TABLE 3.9 – La métrique MCS_V du seul plan du groupe 3 (25×275 cGy)

3.7 Métrique de précision : *LOIC*

La figure 3.9 représente les métriques LOIC pour les plans du groupe 1 et du groupe 2. Les valeurs moyennes obtenues pour le groupe 1 (Fig 3.9a) sont respectivement de $0,641 \pm 0,334$ dans la planification manuelle et de $0,981 \pm 0,174$ pour l'auto-planning. Le test de Wilcoxon donne une valeur de $p = 0,0019$. Le groupe 2 présenté sur la figure 3.9b indique que les valeurs moyennes sont $0,64 \pm 0,32$ et $0,996 \pm 0,402$ respectivement pour la planification manuelle et l'auto-planning. Selon l'analyse statistique faite par le test de Wilcoxon, on a obtenu une valeur de $p = 0,0366$. Les résultats du groupe 3 sont explicités dans le tableau 3.10.

Patient	LOIC manuelle	LOIC automatique
C1	0,62	0,81

TABLE 3.10 – La métrique LOIC du seul plan du groupe 3 (25×275 cGy)



(a) Groupe 1 (20×300 cGy)

(b) Groupe 2 (25×263 cGy)

FIGURE 3.9 – Box plot des métriques *LOIC* des plans de traitement.

3.8 Mesures et calculs de l'indice γ

Les mesures avec l'ArcCHECK et le calcul de l'indice γ sont effectuées sur tous les plans de traitement et pour les deux modalités de planification (manuelle et automatique). Au total, 68 mesures et calculs sont réalisés pour évaluer les distributions de la dose afin de s'assurer que la machine est capable de délivrer le traitement. Les tableaux en annexe présentent les résultats obtenus selon la procédure du CHU de Liège (3%,3mm) et les recommandations du TG 218 de l'American Association of Physicists in Medicine (AAPM) (3%,2mm). Les figures en annexe (Fig A.1 et Fig A.2) montrent les distributions de dose mesurée avec l'ArcCHECK et calculée avec le TPS pour un patient du groupe 1 respectivement pour la planification manuelle et la planification automatique.

Pour le groupe 1, la valeur moyenne obtenue pour le critère 3%/3mm est de $99,62\% \pm 0,55\%$ pour les plans optimisés manuellement et de $99,34\% \pm 0,62\%$ pour la planification automatisée. Pour ce même groupe, le critère 3%,2mm a donné comme valeur moyenne de $96,51\% \pm 1,86\%$ et $94,91\% \pm 2,13\%$ respectivement pour la planification manuelle et la planification optimisé avec Machine Learning. L'analyse statistique révèle que la différence de valeur est statistiquement significative pour le critère 3%,2mm ($p = 0,0119$) et non significative pour le critère 3%,3mm avec $p = 0,1021$. Le groupe 2 a obtenu une valeur moyenne pour le critère $\gamma(3\%/3\text{mm})$ de $99,02\% \pm 0,67\%$ pour la planification manuelle et de $99,9\% \pm 0,23\%$ pour la planification automatisée. Avec le critère $\gamma(3\%/2\text{mm})$, ce groupe a obtenu une valeur moyenne de $95,42\% \pm 2,6$ pour les plans optimisés manuellement et $97,8\% \pm 1,89\%$ pour les plans optimisés avec Machine Learning. Le test de Wilcoxon montre que la différence pour le critère $\gamma(3\%/3\text{mm})$ est significative ($p = 0,0207$) mais non significative pour le critère $\gamma(3\%/2\text{mm})$ car $p = 0,0578$.

Chapitre 4

Discussions

L'analyse des résultats de l'indice de conformité montre que les valeurs moyennes obtenues pour les deux modalités de planification du groupe 1 sont équivalentes (CI = 0,84). En général, le Machine Learning réduit un peu la couverture. Pour les patients du groupe 2, une augmentation de la conformation au PTV est constatée pour la planification automatique. L'augmentation de la conformation au volume cible a entraîné une nette dégradation de sa couverture à l'exception du patient B4 où la couverture passe de 97,23% manuellement à 99,61% dans la planification automatisée. Le test de wilcoxon du CI révèle que la différence entre les deux modalités de planification de traitement n'étaient pas significative ($p > 0,05$) pour ces deux groupes. Le groupe 3 présente un CI assez similaire des deux autres groupes et une diminution de la couverture au PTV. Le fait d'avoir qu'un seul plan dans le groupe 3 n'a pas permis de réaliser le test statistique.

En général, une meilleure conformation indique que les tissus sains et les organes à risque sont mieux épargnés. Tous les plans ont atteint une couverture strictement supérieure à 95% de la dose prescrite sans excéder la 107%. On constate également que l'indice de conformité des plans de traitement réalisés avec Machine Learning est en majorité supérieur à la valeur seuil de 0,8. Les valeurs extrêmes de la figure 3.1 sont essentiellement plus importantes dans le cas ML mais n'influence pas la qualité du plan.

En ce qui concerne l'indice d'homogénéité, le groupe 1 présente une légère augmentation pour la planification automatisée. Autrement, la valeur moyenne était de 0,944 pour la planification manuelle et de 0,954 pour l'auto-planning. Pour le groupe 2, les valeurs moyennes étaient de 0,943 et 0,954 respectivement pour les plans VMAT manuels et les plans VMAT automatisés. Les valeurs extrêmes de l'indice d'homogénéité de la figure 3.2 sont moins marquées pour les plans automatisés. La différence obtenue est statistiquement significative pour le groupe 1 avec une valeur de $p = 0.015$. Ceci signifie que les résultats ne sont pas dus au hasard d'où le rejet de l'hypothèse nulle. Contrairement au groupe 1, les valeurs du groupe 2 n'étaient pas statistiquement significatives car $p = 0.085$. Les résultats du groupe 3 sont presque similaires à ceux des deux autres groupes. Globalement, l'indice d'homogénéité de l'ensemble des plans pour les trois groupes n'a pas été inférieur à la valeur fixée de 0,9.

Dans les tableaux 3.4 et 3.5 montrant la dose moyenne et globale reçue par les organes à risque, on remarque sur certain OAR une variation de la dose sur les plans VMAT optimisés avec ML par rapport aux plans VMAT optimisés manuellement. Pour le cas du groupe 1, on observe une augmentation de la dose pour la planification automatisée, principalement sur le canal anal de 235,6 cGy et sur la vessie (Bladder). On observe également une diminution de la dose sur la tête fémorale droite (FemoralHead-R) de 29,7 cGy, sur le rectum de 478,2 cGy et sur la sigmoïde de 1,4 cGy. L'analyse statistique avec le test de Wilcoxon a montré que les résultats étaient statistiquement significatifs uniquement sur le rectum ($p < 0,05$). Cela pourrait indiquer que le modèle force sur ces deux organes en particulier.

Pour les OAR du groupe 2, il y a une nette augmentation de la dose pour les plans VMAT automatisés. C'est le cas canal anal, la tête fémorale gauche, la tête fémorale droite et la sigmoïde avec respectivement 195,4 cGy ; 667,8 cGy ; 588,6 cGy et 181 cGy. La dose a été réduite de 167,7 sur la vessie et de 115,2 sur le rectum. Dans ce groupe 2, seules les têtes fémorales présentaient une différence statistique significative ($p < 0,05$). Pour le cas des OAR du groupe 3, on note une petite augmentation de la dose uniquement sur les têtes fémorales. L'augmentation de la dose aux OAR des groupes 2 et 3 est probablement due aux faibles performances du modèle pour un autre fractionnement que celui pour lequel il a été créé.

Dans le tableau 3.6 présentant la dose reçue par 2% du volume de chaque OAR, il y a une légère variation de la dose reçue pour les plans automatisés. On remarque une diminution de la moyenne des $D_{2\%}$ sur le canal anal de 1,1 cGy, sur la vessie de 97,3 cGy, sur la tête fémorale droite de 20 cGy et sur le rectum de 142,2 cGy. Par contre, la tête fémorale gauche et le sigmoïde reçoivent respectivement en moyenne 148,2 cGy et 169,1 cGy en plus. Selon le test de Wilcoxon, il semble que le modèle insiste sur le rectum et la vessie car la p-value est inférieure à 5%. Le tableau 3.7 du $D_{2\%}$ pour le groupe 2 révèle une réduction de la dose sur 2% du volume des OAR tels que la vessie de 179,4 cGy et le rectum de 106,4 cGy pour la planification automatique par rapport à la planification manuelle. De même, on remarque une augmentation de la $D_{2\%}$ au canal anal (35,8 cGy), sur les têtes fémorales 898,5 cGy sur la droite et 995,6 cGy sur la gauche) et sur la sigmoïde (125,7 cGy). Néanmoins, seuls le canal anal et le sigmoïde présentaient des différences non significatives statistiquement ($p > 0,05$). Les résultats du groupe 3 (Fig 3.6) montrent comme dans de la figure 3.5 une légère augmentation de la $D_{2\%}$ sur les têtes fémorales.

Globalement, les valeurs de l'indice *LOIC* ont augmenté dans pour les plans VMAT automatisés. Par définition, le MCS_V a des valeurs comprises entre 0 et 1 et le *LOIC* a une limite inférieure 0 mais pas de limite supérieure. Cependant, nous n'avons obtenu aucune valeur de l'indice MCS_V supérieure à 0,63 et plus de 80% des plans VMAT optimisés avec ML/DL avaient des valeurs comprises entre 0,33 et 0,618.

Le groupe 2 (Fig 3.9b) a le *LOIC* le plus élevé avec une valeur moyenne de $0,996 \pm 0,402$ et le *MCS_V* plus faible de $0,489 \pm 0,068$ (Fig 3.8b). Cette augmentation de la valeur des métriques est liée à l'amélioration de la conformation au PTV qui crée des arrangements complexes du MLC avec des petits segments. En d'autre terme, cela signifie que lorsque le degré de modulation augmentait (diminution du *MCS_V*), le déplacement moyen des lames augmentait également pour la plupart des plans VMAT optimisés avec Machine Learning avec des ouvertures de faisceau plus petites et irrégulières [47]. De plus, selon la valeur du *LOIC*, il y a une création de segments plus étroits qui amène à une imprécision du calcul de dose dans le TPS. Pour ce groupe 2, l'analyse statistique a donné une valeur de $p = 0,6103$ pour l'indice *MCS_V* et de $p = 0,0366$ pour l'indice *LOIC*. L'optimisation du seul plan du groupe 3 était moins modulé (Table 3.10) par rapport aux deux autres groupes. Les résultats obtenus du groupe 1 étaient intermédiaires aux groupes 2 et 3 comme présentés sur les figures 3.8a et 3.9a. Le test de Wilcoxon a révélé que la différence entre les moyennes des deux modalités de planification est statistiquement significative pour l'indice *LOIC* ($p = 0.0019$) et non significative pour l'indice *MCS_V* ($p = 0.4222$).

On remarque que, lorsque la valeur de l'indice *MCS_V* diminuait, la valeur de l'indice *LOIC* augmentait également. Ceci rendait les plans plus complexes et plus modulés avec une augmentation des unités moniteurs (UMs) pour les plans VMAT optimisés avec Machine Learning. D'autre part, l'augmentation de l'indice de conformité et de l'indice d'homogénéité pour des plans VMAT automatisés sont également l'une des causes de l'augmentation des UMs raison pour laquelle les organes à risque ont reçus des doses plus importantes. Toutefois, les UMs du groupe 1 présentaient une différence statistiquement significative ($p = 0,0015$) et non significative pour le groupe 2 ($p = 0,24$).

Au vue de l'augmentation du *MCS_V*, du *LOIC* et UMs pour les plans VMAT automatisés, les mesures avec l'ArcCHECK ont été effectuées. En principe, les mesures avec l'ArcCHECK sont faites lorsque l'indice *LOIC* est comprise entre 1,25% et 3% au CHU de Liège afin d'évaluer l'indice γ . Bien que la moyenne du *LOIC* était inférieure à 1,25%, tous les 68 plans ont été mesurés. Dans notre pratique clinique, pour les critères 3%/3mm et 3%/2mm, les niveaux de tolérance pour les taux de réussite sont fixés respectivement à 95% et 90% pour l'analyse gamma global. En moyenne pour le groupe 1, les mesures ont révélé un taux de réussite du critère $\gamma(3\%/3\text{mm})$ de $99,92\% \pm 0,55\%$ pour les plans d'origine et $99,34\% \pm 0,62\%$ pour les plans réalisés avec Machine Learning. L'analyse statistique a confirmé que la différence de valeur n'était pas significative car $p = 0,1021$. Pour ce même groupe 1 avec pour critère 3%/2mm, la moyenne de l'indice γ était de $96,51\% \pm 1,86\%$ pour les plans originaux et de $94,91\% \pm 2,13\%$ pour les plans optimisés avec Machine Learning. Le test de Wilcoxon a montré qu'il y avait une nette différence entre les deux modalités de planification avec pour valeur $p = 0.0119$. Donc, pour un critère plus sévère, on a une diminution significative de l'indice γ . Cela indique une

plus grande complexité des plans et va dans le même sens que ce que dit le LOIC et le nombre d'unités moniteurs mais reste très bon (94,91% en γ (3%/2mm)).

Le groupe 2 en revanche, la valeur moyenne de l'indice γ était le plus élevé pour le critère 3%/2mm pour les plans VMAT automatisés. La valeur moyenne obtenue pour le critère 3%/3mm était de $99,02\% \pm 0,67\%$ pour les plans d'origine et de $99,9\% \pm 0,23\%$ pour les plans automatisés. De même, pour le critère 3%/2mm de ce groupe 2, la planification manuelle a donné une valeur moyenne de l'indice γ global de $95,02\% \pm 2,6\%$ tandis que la planification automatisée donnait une valeur moyenne de $97,8\% \pm 1,89\%$. L'analyse statistique par le test de Wilcoxon a démontré que la différences de valeur moyenne était statistiquement significative pour le critère 3%/3mm avec $p = 0,0207$ et non significative pour le critère 3%/2mm car $p = 0,0578$. Pour le groupe 2, les conclusions sont moins évidentes. Le LOIC indique une plus grande complexité, mais pas le MCS_V ni les UM, ni le indice γ . Ici, la dose au rectum et à la vessie est améliorée mais elle augmente aux têtes fémorales. Cela indique que le système automatisé trouve une autre solution sans dégrader la complexité du plan.

Le seul patient du groupe 3 a donné une valeur de 100% et 96,8% respectivement pour le critère 3%/3mm et 3%/2mm des plans automatisés. Le plan d'origine pour ce groupe 3 a donné une valeur de 100% pour le critère 3%/3mm et 98,3% pour le critère 3%/2mm.

Dans l'ensemble, les valeurs de l'indice γ obtenue pour les plans VMAT optimisés avec ML sont évidemment des bons résultats montrant la fiabilité de la planification et de la délivrabilité du plan VMAT. Dans le même temps, l'adoption du critère 3%/2mm permettrait d'avoir une plage de valeurs plus importante rendant l'analyse de l'indice γ plus sensible aux différences entre la dose calculée avec l'ArcCHECK et la dose réellement délivrée.

Pour quelques patients pris au hasard, un dosimétriste avec plus d'expérience a fait une replanification et a noté le temps mis pour faire le plan. Il prenait environ 35 à 50 minutes pour réaliser un plan VMAT manuellement. Par contre avec une optimisation avec Machine Learning de ces mêmes plans de patients, le système de planification met en moyenne 15 à 20 minutes pour obtenir un plan acceptable.

Conclusion et Perspectives

L'objectif de ce mémoire était de faire une évaluation de la qualité des plans de traitement VMAT optimisés avec l'intelligence artificielle notamment le Machine Learning par rapport aux plans classiques dans le TPS RayStation. Pour ce faire, nous avons réalisé les plans de traitement, puis notés les données obtenues et réalisés une analyse statistique sur ces dernières.

Nous avons introduit différents indices (CI et HI) pour évaluer la qualité des plans et des métriques (MCS_V et $LOIC$) pour étudier la complexité de ces plans réalisés. Globalement, les indices de conformité et d'homogénéité étaient meilleurs pour les plans automatisés mais non significatifs. La couverture au PTV remplissait la contrainte fixée qui stipule que le volume cible doit recevoir au moins 95% de la dose prescrite. Néanmoins, on a constaté une augmentation significative des unités moniteurs pour les plans VMAT automatisés du groupe 1 et du groupe 2, de même que le groupe 3. Ceci est une conséquence de la modulation des plans plus importante pour l'optimisation automatisée. Cependant, la complexité a augmenté mais la valeur moyenne l'indice γ était équivalent à celui des plans d'origine pour le critère 3%/3mm du groupe 1 et 2. Pour le critère 3%/2mm du groupe 1, la valeur moyenne de l'indice γ a significativement diminué lorsqu'on passe de la planification manuelle à la planification automatisée. Par contre, la valeur moyenne de l'indice γ du groupe 2 a augmenté pour le même critère (3%/2mm) de façon non significative.

Des procédures de comparaison statistique ont été décrites et le test de Wilcoxon a été choisi. Il est par définition un test non paramétrique utilisé pour des séries de données appariées. Le choix de ce test est bien adapté pour notre étude car nous utilisons une petite taille d'échantillon (34 plans de traitement). Enfin, nous avons relevé le temps mis pour réaliser un plan de traitement VMAT. Il en ressort qu'on gagne plus de temps lorsque le plan est réalisé avec l'optimisation ML/DL.

L'introduction de la planification automatisée en routine clinique est favorable puisqu'on obtient une meilleure conformation et une homogénéité significative de la dose au PTV avec des métriques qui restent acceptables. Néanmoins, les mesures avec l'ArcCHECK montrent que la machine n'aura aucune difficulté à délivrer le traitement.

La réalisation des plans de traitement avec Machine Learning avec l'algorithme RF utilise un modèle proposé par RaySearch avec pour prescription 20×300 cGy. Nous avons implémenté ce modèle sur les plans de même prescription et sur d'autres prescriptions pour une prostate simple au CHU de Liège. Les paramètres utilisés dans l'algorithme RF tels que la dose maximale aux OAR et le poids sur les contraintes associés à chaque organe à risque, étaient par défauts ceux destinés à des plans de prescription 6000 cGy. Il serait donc intéressant à la suite de ce travail, de faire un certain nombre de réglages de ces paramètres dans le RF du TPS avec l'équipe de RaySearch pour l'adapter avec précision aux différentes prescriptions de cette étude. Puis, refaire une étude similaire après réglage et comparer les résultats obtenus à ceux de ce mémoire. Une étude aussi intéressante serait d'explorer les limites du modèle proposé par RaySearch en utilisant d'autres prostates avec SIB par exemple ou un autre fractionnement sans SIB. Ceci permettrait de ne pas créer un modèle pour chaque fractionnement ou pour chaque localisation.

Annexe A

Données

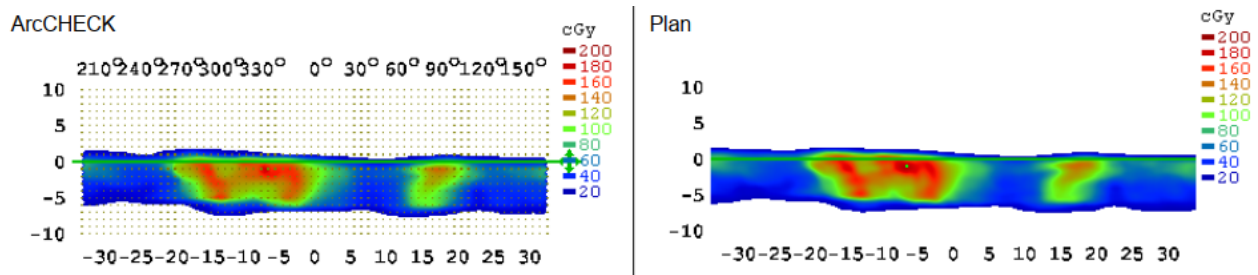


FIGURE A.1 – Distribution de dose pour le plan d’origine.
L’ArcCHECK à gauche et le TPS à droite.

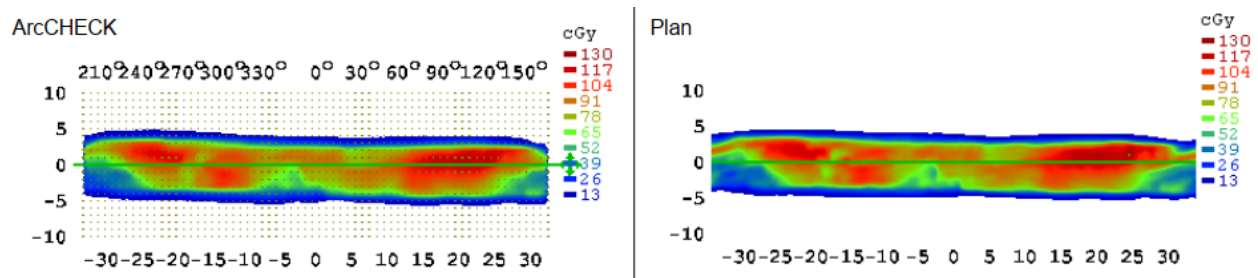


FIGURE A.2 – Distribution de dose pour le plan automatique.
L’ArcCHECK à gauche et le TPS à droite.

Patients	Planification manuelle (%)		Planification automatisée (%)	
	3% , 3mm	3% , 2mm	3% , 3mm	3% , 2mm
C1	100	98,3	100	96,8

TABLE A.1 – Valeurs de l’indice γ du groupe 3 (25×275 cGy)

Patients	Planification manuelle (%)		Planification automatisée (%)	
	3% , 3mm	3% , 2mm	3% , 3mm	3% , 2mm
A1	98,5	93,1	100	97,9
A2	99,6	99,5	99,4	95,1
A3	97,7	94,1	98,4	93,4
A4	99,8	95,3	99,8	94,6
A5	99,6	96,7	99	93,4
A6	99,8	97,3	100	98,5
A7	99,2	96,8	100	97,7
A8	99,5	96,9	99,5	94
A9	99,7	97,1	99,3	94,8
A10	99,4	94,2	99	94,3
A11	99,7	96,6	99,4	94,3
A12	100	96,7	97,9	91,6
A13	99,8	94,6	99,4	94,5
A14	100	94,3	99,4	93,5
A15	100	98,8	99,3	95,9
A16	99,9	99,2	99,5	94
A17	99,8	98,6	99,4	94,2
A18	100	99,2	100	99,4
A19	100	96,1	100	97,5
A20	99,8	96,9	97,9	92,6
A21	99,9	96,9	99,9	95,3
A22	100	94,3	99	91,5

TABLE A.2 – Valeurs de l'indice γ du groupe 1 (20×300 cGy)

Patients	Planification manuelle (%)		Planification automatisée (%)	
	3% , 3mm	3% , 2mm	3% , 3mm	3% , 2mm
B1	99,4	94	99,3	95
B2	98	94,7	100	99,2
B3	99,7	98,9	100	96,5
B4	100	97,5	100	99,8
B5	98,3	92,7	99,8	99,1
B6	98,5	90,9	100	95,1
B7	99,3	95,5	100	97,8
B8	99,2	97	100	97,9
B9	98,8	97,6	100	99,8
B10	100	98,7	99,8	99,8
B11	80,6	72,5	100	97,9

TABLE A.3 – Valeurs de l'indice γ du groupe 2 (25×263 cGy)

Bibliographie

- [1] Alan TURING et J HAUGELAND. *Computing machinery and intelligence*. 1950.
- [2] Neha SHARMA, Reecha SHARMA et Neeru JINDAL. “Machine Learning and Deep Learning Applications-A Vision”. In : *Global Transitions Proceedings 2.1* (2021), p. 24-28.
- [3] Ian J GOODFELLOW et al. “Generative adversarial networks”. In : *preprint* (2014).
- [4] Arthur L SAMUEL. “Some studies in machine learning using the game of checkers”. In : *IBM Journal of research and development 3.3* (1959), p. 210-229.
- [5] Phil KIM. “Matlab deep learning”. In : *With machine learning, neural networks and artificial intelligence 130* (2017), p. 21.
- [6] Jonathan LONG, Evan SHELHAMER et Trevor DARRELL. “Fully convolutional networks for semantic segmentation”. In : *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, p. 3431-3440.
- [7] Shyam Prasad ADHIKARI, Heechan YANG et Hyongsuk KIM. “Learning semantic graphics using convolutional encoder–decoder network for autonomous weeding in paddy”. In : *Frontiers in plant science 10* (2019).
- [8] Vijay BADRINARAYANAN, Alex KENDALL et Roberto CIPOLLA. “Segnet : A deep convolutional encoder-decoder architecture for image segmentation”. In : *IEEE transactions on pattern analysis and machine intelligence 39.12* (2017), p. 2481-2495.
- [9] Marco DINARELLI et Isabelle TELLIER. “Étude des réseaux de neurones récurrents pour étiquetage de séquences (A study of Recurrent Neural Networks for Sequence Labelling)[In French]”. In : *Actes de la conférence conjointe JEP-TALN-RECITAL 2016. volume 2 : TALN (Articles longs)*. 2016, p. 98-111.
- [10] Jürgen SCHMIDHUBER. “Deep learning in neural networks : An overview”. In : *Neural networks 61* (2015), p. 85-117.
- [11] Xin YI, Ekta WALIA et Paul BABYN. “Generative adversarial network in medical imaging : A review”. In : *Medical image analysis 58* (2019).
- [12] Zhaoqing PAN et al. “Recent progress on generative adversarial networks (GANs) : A survey”. In : *IEEE Access 7* (2019), p. 36322-36333.
- [13] Sunan CUI et al. “Introduction to machine and deep learning for medical physicists”. In : *Medical physics 47.5* (2020), p. 127-147.
- [14] Ludovic ARNOLD, Hélène PAUGAM-MOISY et Michèle SEBAG. “Optimisation de la Topologie pour les Réseaux de Neurones Profonds”. In : *17e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle-RFIA 2010*. 2010.
- [15] Kolluru Venkata Sreerama MURTHY. *On growing better decision trees from data*. 1996.

- [16] M SEBBANÜ et al. “Impact of learning set quality and size on decision tree performances”. In : *IJCSS* 1.1 (2000), p. 85.
- [17] Steven L SALZBERG. *C4. 5 : Programs for machine learning by j. ross quinlan. morgan kaufmann publishers, inc., 1993.* 1994.
- [18] J. Ross QUINLAN. “Induction of decision trees”. In : *Machine learning* 1.1 (1986), p. 81-106.
- [19] Sreerama K MURTHY. “Automatic construction of decision trees from data : A multidisciplinary survey. Data Mining and Knowledge Discovery”. In : (1998).
- [20] Tin Kam HO. “The random subspace method for constructing decision forests”. In : *IEEE transactions on pattern analysis and machine intelligence* 20.8 (1998), p. 832-844.
- [21] L BREIMAN. “Random Forests Machine Learning, vol. 45”. In : (2001).
- [22] Siddharth MISRA et Hao LI. “Noninvasive fracture characterization based on the classification of sonic wave travel times”. In : *Machine Learning for Subsurface Characterization* (2019), p. 243-287.
- [23] Hongming SHAN et al. “Synergizing medical imaging and radiotherapy with deep learning”. In : *Machine Learning : Science and Technology* 1.2 (2020).
- [24] Ge WANG et al. *Machine learning for tomographic imaging.* 2019.
- [25] Michele Arcangelo QUINTO. “Méthode de reconstruction adaptive en tomographie par rayons X : optimisation sur architectures parallèles de type GPU”. Thèse de doct. Université de Grenoble, 2013.
- [26] Joël GREFFIER. “Reconstruction itérative en scanographie : optimisation de la qualité image et de la dose pour une prise en charge personnalisée”. Thèse de doct. Université Montpellier, 2016.
- [27] Changhee HAN et al. “GAN-based synthetic brain MR image generation”. In : *IEEE-15th International Symposium on Biomedical Imaging.* 2018, p. 734-738.
- [28] Chao DONG et al. “Image super-resolution using deep convolutional networks”. In : *IEEE transactions on pattern analysis and machine intelligence* 38.2 (2015), p. 295-307.
- [29] Ashnil KUMAR et al. “Machine learning in medical imaging”. In : *Biomedical Information Technology.* 2020, p. 167-196.
- [30] Dan CIRESAN et al. “Deep neural networks segment neuronal membranes in electron microscopy images”. In : *Advances in neural information processing systems* 25 (2012), p. 2843-2851.
- [31] Lei Bi et al. “Dermoscopic image segmentation via multistage fully convolutional networks”. In : *IEEE Transactions on Biomedical Engineering* 64.9 (2017), p. 2065-2074.
- [32] Dong NIE et al. “Fully convolutional networks for multi-modality isointense infant brain image segmentation”. In : *2016-IEEE 13Th international symposium on biomedical imaging.* 2016, p. 1342-1345.
- [33] Yu GORDIENKO et al. “Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-ray analysis of lung cancer”. In : *International Conference on Computer Science, Engineering and Education Applications.* 2018, p. 638-647.
- [34] *ICRU Report 50 : Prescribing, recording, and reporting photon beam therapy.* Rapp. tech. 50. International Commission on Radiation Units et Measurements, 1993.

- [35] *Recording and reporting photon beam therapy (supplement to ICRU Report 50)*. Rapp. tech. 62. International Commission on Radiation Units et Measurements, 1999.
- [36] Marcel VAN HERK. “Errors and margins in radiotherapy”. In : *Seminars in radiation oncology*. T. 14. 1. 2004, p. 52-64.
- [37] Mingqing WANG et al. “A review on application of deep learning algorithms in external beam radiotherapy automated treatment planning”. In : *Frontiers in Oncology* 10 (2020).
- [38] Warren G CAMPBELL et al. “Neural network dose models for knowledge-based planning in pancreatic SBRT”. In : *Medical physics* 44.12 (2017), p. 6148-6158.
- [39] Dan NGUYEN et al. “A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning”. In : *Scientific reports* 9.1 (2019), p. 1-10.
- [40] Tomohiro KAJIKAWA et al. “Automated prediction of dosimetric eligibility of patients with prostate cancer undergoing intensity-modulated radiation therapy using a convolutional neural network”. In : *Radiological physics and technology* 11.3 (2018), p. 320-327.
- [41] Qiong REN, Hui CHENG et Hai HAN. “Research on machine learning framework based on random forest algorithm”. In : *AIP Conference Proceedings*. T. 1820. 1. 2017.
- [42] *RayStation 10A - User Manual*. 2020.
- [43] Loic FEUVRET et al. “Conformity index : a review”. In : *International Journal of Radiation Oncology* Biology* Physics* 64.2 (2006), p. 333-342.
- [44] David PALMA et al. “Volumetric modulated arc therapy for delivery of prostate radiotherapy : comparison with intensity-modulated radiotherapy and three-dimensional conformal radiotherapy”. In : *International Journal of Radiation Oncology* Biology* Physics* 72.4 (2008), p. 996-1001.
- [45] Sophie CHIAVASSA et al. “Complexity metrics for IMRT and VMAT plans : a review of current literature and applications”. In : *The British journal of radiology* 92.1102 (2019).
- [46] Andrea L MCNIVEN, Michael B SHARPE et Thomas G PURDIE. “A new metric for assessing IMRT modulation complexity and plan deliverability”. In : *Medical physics* 37.2 (2010), p. 505-515.
- [47] Laura MASI et al. “Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy”. In : *Medical physics* 40.7 (2013), p. 1-11.
- [48] Michel MATHOT et David DECHAMBRE. “6 VMAT complexity metrics can reduce patient QA workload”. In : *European Journal of Medical Physics* 56 (2018), p. 3-4.
- [49] Moyed MIFTEN et al. “Tolerance limits and methodologies for IMRT measurement-based verification QA : recommendations of AAPM Task Group No. 218”. In : *Medical physics* 45.4 (2018), p. 53-83.